School of intensive courses
in Novi Sad

April 4-9, 2008

Development of Computer-aided Methods in Teaching Mathematics and Science

www.im.ns.ac.yu

CIP-Каталогизација у публикацији
Библиотека Матице српске, Нови Сад

51:371.3]:004(082)

5:371.3]:004(082)

SCHOOL of Intensive Courses (2008 ; Novi Sad)
Development of computer-aided methods in teaching mathematics and science / School of Intensive Courses in Novi Sad, April 4-9, 2008. - Novi Sad : Prirodno-matematički fakultet, 2008 (Novi Sad : Departman za grafičko inženjerstvo i dizajn FTN). - IV, 499 str. : ilustr. ; 24 cm.

Tekst na srp. i engl. jeziku. - Str. 1-2: Preface / Arpad Takač. - Bibliografija uz većinu radova.

ISBN 978-86-7031-166-4

а) Математика - Настава - Рачунари - Зборници б)
Природне науке - Настава - Рачунари - Зборници

COBISS.SR-ID 230141191

Sadržaj

Preface

This publication is prepared for the participants of the ***School of Intensive Courses***, which will take place in the period 4-9 April 2008 at the Faculty of Sciences, Novi Sad. It is planned that the seven contributors teach their courses at the School. The School is organized under auspices of the one year project entitled

"Development of computer-aided methods in teaching mathematics and natural sciences".

This project (No. 06SER02/02/003) is funded by the European Union, within the Neighbourhood Programme Hungary-Serbia, and is hosted at the Faculty of Sciences, Novi Sad. The overall objective of the project is promoting and making available in several languages the existing means and methods of computer aided teaching and its practical results obtained so far.

It is important to emphasize that the project is one of the two mirror projects, the other being hosted by the University of Szeged. In Szeged, another series of courses will be held in the period 25-20 March 2008.

The aim of the courses, both in Novi Sad and Szeged, is to develop methods of computer-aided modeling in teaching mathematics and natural sciences. It is planned that the exchange of listeners and teachers between two host institutions is to enable a much closer cooperation between Faculty of Natural Sciences, Novi Sad, and University of Szeged. The intensive courses will be imbedded in further improvement of teachers' education and the curricula of university studies at all levels including PHD studies.

We, the contributors of this book, wish to thank to Prof. Miroslav Vesković, Dean of the Faculty of Sciences, Novi Sad, to Prof. Marko Nedeljkov, Director of the Department of Mathematics and Informatics in Novi Sad, and also to many other colleagues for their support, help and useful advices during the preparation of the book.

As the project manager of the Project within whose auspices the School will take part, I wish to express my thanks to Ms. Bojana Milićević, Coordinator of the Neighbourhood Programme Hungary-Serbia, to Mr. Zoran Krtinić and Mr. Relja Burzan from the Local Office of the Neighbourhood Programme Hungary-Serbia, for their support in preparation of the School, and their useful advices that helped improve the documentation relevant for the implementation of the Project.

Last but not least, my special thanks go the contributors of this book, for their continuous and successful work on developing the courses and preparation of the texts.

Novi Sad, March 2008

Dr Arpad Takači, Project Manager

Project: 06SER02/02/003

Computer-aided analysis of mathematical models

Dr Arpad Takači

Glava 1

Modeliranje i simulacija

1.1 Modeliranje

Modeliranje je način rešavanja problema koji se pojavljuju u stvarnom svetu. Primenjujemo ga kada nije moguće vršiti eksperimente u realnom okruženju (svemirska istraživanja) ili kada je skupo napraviti prototip. Pomoću modela sistema dobijamo odgovor na pitanje ”Šta ako?” i mogućnost da optimizujemo sistem pre implementacije.

Modeliranje predstavlja jedan od osnovnih procesa ljudskoga uma i izražava našu sposobnost da mislimo i zamišljamo, da koristio simbole i jezike, da komuniciramo, da vršimo generalizacije na osnovu iskustva, da se suočavamo sa neočekivanim. Ono nam omogućava da uočavamo obrasce, da procenjujemo i predviđamo, da upravljamo procesima i objektima, da izlažemo značenje i svrhu. Upravo zato, modeliranje se najčešće posmatra kao najznačajnije konceptulano sredstvo koje čoveku stoji na raspolaganju.

Model je apstrakcija realnosti u smislu da on ne može da obuhvati sve njene aspekte. Model je uprošćena i idealizovana slika realnosti. On nam omogućava da se suočimo sa realnim svetom (sistemom) na pojednostavljen način, izbegavajući njegovu komplekstnost i ireverzibilnost, kao i sve opasnosti koje mogu proisteći iz eksperimenata nad samim realnim sistemom. Model je opis realnog sistema sa svim onim karakteristikama koje su relevantne iz našeg ugla posmatranja.

Treba reći da je modeliranje ne samo nauka, nego i veština i umetnost. U stvari, kod pravljenja nekog modela je najbitniji izbor odgovarajućeg u mnoštvu potencijalnih. Dodajmo da u procesu modeliranja nije najvažnije napraviti sveobuhvatni model, već je mnogo korisnije napraviti najprostiji model koji sadrži esencijalne elemente realnog sistema koji se modelira. Drugim rečima, u procesu modeliranja treba uočiti i izdvojiti one elemente i karakteristike sistema koji su bitni za naše istraživanje i oni će biti obuhvaćeni modelom, dok će ostali elementi i karakteristike biti zanemareni. Zbog toga, model ne sadrži samo objekte i atribute realnog sistema već i određene pretpostavke o uslovima njegove validnosti. Suviše složeni ili savršeni modeli, čak iako su ostvarivi, najčešće su preskupi i neadekvatni za eksperimentisanje. Sa druge

strane, suviše pojednostavljeni modeli ne odslikavaju na pravi način posmatrani sistem, a rezultati koji se dobijaju njihovom primenom mogu da budu neadekvatni i pogrešni. Zato, treba pažljivo odabrati nivo apstrakcije tako da rezultujući model što vernije odslikava posmatrani sistem, ali i da njegova složenost i cena ne budu ograničavajući faktor.

Neformalni opis modela daje osnovne pojmove o modelu i, mada se teži njegovoj potpunosti i preciznosti, on to najčešće nije. Prilikom izgradnje neformalnog opisa, upravo radi eliminisanja pomenutih nedostataka, vrši se podela na objekte, opisne promenljive i pravila interakcija objekata.

Objekti su delovi iz kojih je model izgrađen; opisne promenljive opisuju stanja u kojima se objekti nalaze u određenim vremenskim trenucima; u opisne promenljive takođe spadaju i parametri koji opisuju konstantne karakteristike modela; pravila iteracije objekata definišu kako objekti modela utiču jedan na drugi u cilju promene njihovog stanja.

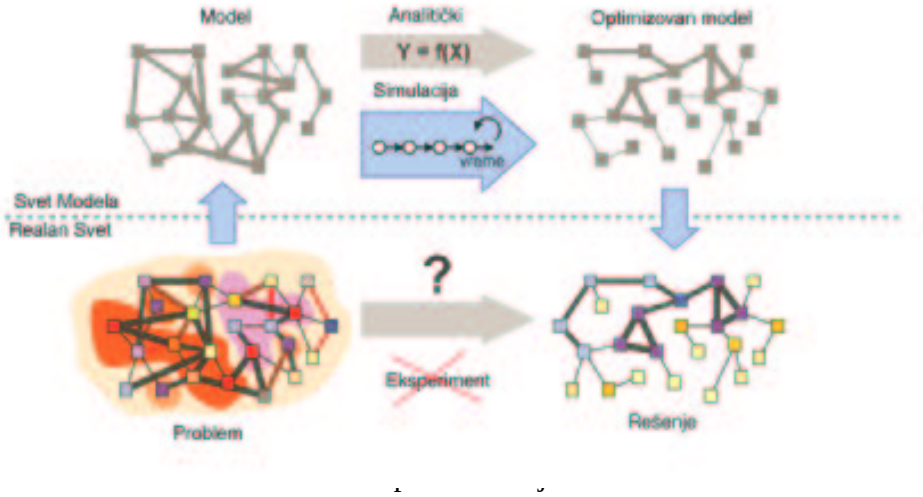
Anomalije koje se javljaju prilikom neformalnog opisa modela najčešće su nekompletan, nekonzistentan ili nejasan opis modela. Ukoliko model ne sadrži sve situacije koje mogu da nastupe, tada je opis nekompletan. Ukoliko su u opisu modela za istu situaciju predviđena dva ili više pravila čijom se primenom dobijaju kontradiktorne akcije, tada je opis nekonzistentan, a ako u jednoj situaciji treba obaviti dve ili više akcija, a pri tome nije definisan njihov redosled, tada je opis modela nejasan.

Za šire savladavanje i razumevanje teorije modeliranja i računarske simulacije, čitalac može konsultovati neke od knjiga navedenih u literaturi.

1.1.1 Analitički i simulacioni modeli

Postupak kreiranja modela delimo na korake. Prvo je potrebno identifikovati problem koji hoćemo da realizujemo u modelu. Zatim sledi apstrakcija tog problema, odnosno pojednostavljivanje problema do određenog nivoa da bi bio moguć za realizaciju, ali i da nam daje validne podatke. To je sledeći korak odnosno verifikacija modela. Na kraju, posle analize modela i optimizacije prevodimo dobijene rezultate u krajnji rezultat koji je primenjiv na problem u realnom sistemu. Napravimo razliku između dve osnovne vrste modela: analitičkih i simulacionih modela. Pre svega, u analitičkim ili statičkim modelima rezultat funkcionalno zavisi od ulaznih podataka. Takav model je jednostavan za realizaciju i za njega se može koristi tabelarni kalkulator.

Međutim, u većini slučajeva je teško pronaći analitičko rešenje za problem. Tada se primenjuje simulacija, odnosno dinamičko modeliranje. Simulacioni model se sastoji iz skupa pravila koja nam opisuju ponašanje sistema kroz vreme u odnosu na trenutno stanje. Pravila se obično zadaju u obliku jednačina, konačnih automata, dijagrama stanja i slično. Simulacija je proces izvršavanja modela. Model prelazi kroz određena stanja (kontinualna ili diskretna) u toku vremena. Iz ovog se može zaključiti da je



za složene probleme gde je promena kroz vreme značajna bolje koristiti simulaciono modeliranje.

1.1.2 Nivoi apstrakcije u modelu

Probleme koje rešavamo možemo klasifikovati po nivoima apstrakcije koju primenjujemo na realan problem.

Ako krenemo od najnižeg nivoa apstrakcije, govorimo o fizičkim modelima odnosno o modelima gde nam je od značaja tačna veličina objekata, rastojanje između njih, brzina kretanja, precizno merenje vremena i slično. Ovaj nivo se može nazvati **operacioni** ili **mikro nivo**. Primeri za ovaj nivo su modeli na mikro nivou za saobraćaj, naime kretanje pešaka, kontrolni sistemi, mehanotronika. Na mikro nivou se nalaze i modeli proizvodnje u fabrikama, ali su oni malo iznad ostalih jer kod njih koristimo tačne putanje i prosečno vreme. Logistički modeli skladišta gde modeliramo utovar i istovar se nalaze na istom nivou kao i model proizvodnje jer primenjujemo isti nivo apstrakcije.

Taktički ili **srednji nivo** predstavlja oblast srednje apstrakcije modela. Tipični primeri su sistemi opsluživanja. Ovi modeli se zasnivaju na zadatim vremenskim rasporedima, ali nekada moramo da uzmemo u obzir i kretanja u modelu, odnosno fizička pomeranja objekata. Primer za to je model urgentnog centra gde moramo uzeti u obzir arhitekturu objekta odnosno raspored prostorija, jer od toga zavisi kretanje osoblja i pacijenata. Računarske mreže i simulacija transporta se može takođe smestiti na ovaj nivo, jer opet radimo sa rasporedima, kapacitetima vozila, vremenom procesiranja, vremenom transporta. Dakle, možemo govoriti o makro modelu saobraćaja gde ne posmatramo tip vozila odnosno ne tretiramo vozila kao zasebne objekte (u slučaju računarskih mreža kao zasebne pakete podataka), nego, na primer, posmatramo zapreminu

Slika 1.2: Pristupi u simulacionom modeliranju na različitim nivoima apstrakcije.

objekata. Modeli lanaca snabdevanja se mogu modelirati na različitim nivoima apstrakcije pa ih stavljamo u srednji ili u visoki nivo apstrakcije.

Problemi na **strateškom** ili **makro nivou** su oni kod kojih primenjujemo visok nivo apstrakcije. Po pravilu su takvi problemi zavisni od globalnih trendova, globalnih povratnih informacija i agregatne vrednosti. U ovom slučaju zanemarujemo objekte kao što su ljudi, vozila, proizvodi itd.

1.1.3 Matematički modeli

Pronalazak diferencijalnog i integralnog računa od strane I. Njutna (*Isaac Newton*) i G. V. Lajbnica (*Gottfried Vilhelm Leibniz*) u drugoj polovini 17. veka je doveo do ogromnog razvoja ne samo matematike nego i njene primene u raznim oblastima, pre svega u fizici i ostalim prirodnim naukama. Ubrzo se shvatilo da se neki procesi u realnom svetu koji se menjaju u vremenu mogu opisati i time objasniti korišćenjem diferencijalnih i diferencnih jednačina.

Može se konstatovati da se korišćenje matematike u analizi i predviđanju pojava realnog sveta već davno podrazumeva u fizici i njenim brojnim primenama. Ipak, pojam matematičkog modela u današnjem smislu je nešto noviji. Na primer, populacioni modeli Maltusa i Verhulsta u 19. veku, i još više biološki modeli V. Voltere (*Vito Volterra*) i A. Lotke (*Alfred Lotka*) u dvadesetim godinama prošlog veka su pravljeni sa jasnom namerom da se makar kvalitativno objasne promene u određenom segmentu realnog sveta, imajući u vidu nemogućnost dobijanja pouzdanih kvantitativnih podataka o realnom sistemu (videti 2. glavu). Uspeh ovakvog pristupa razumevanju realnog sveta je doveo do primene matematičkih i drugih modela, a pre svega simulacionih, u raznim oblastima, uključujući i, na primer, društvene, za koje se do skora to nije moglo ni zamisliti. Naravno, buran razvoj računara zadnjih decenija 20. veka je, uz ostalo, doveo do povećanja potencijalne koristi od modela, a u isto vreme u velikoj meri smanjio i pojednostavio troškove njihove konstrukcije.

Vremenom su matematički modeli postajali komplikovaniji, što je zahtevalo dalji razvoj matematičkih metoda za njihovo rešavanje i analizu. Sjajne i duboke apstrakcije kao rezultat razvoja matematike često dozvoljavaju da se kompleksni problemi koji se pojavljuju u kreiranom matematičkom modelu shvate na novi, često jednostavniji i jasniji način. Razumljivo, ne sme se zaboraviti na suštinsku razliku u matematičkim istraživanjima sa jedne i razumevanju i rešavanju matematičkih modela sa druge strane. Dodajmo da je u slučaju modeliranja složenijih realnih sistema po pravilu potrebno kombinovati matematičke i simulacione metode.

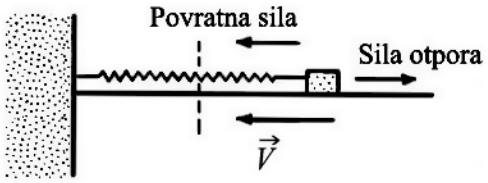
Svaka naučna disciplina pokušava da objasni i reši "svoje" fenomene realnog sveta pomoću eksperimenata, posmatranja, konstrukcije modela i, konačno, postavljanja teorije. Kako smo već napomenuli, u ovoj i sledećoj glavi nas prevashodno zanimaju matematički modeli i metode za njihovo rešavanje. U tom cilju, u ovom poglavlju ćemo analizirati nekoliko modela u prirodnim naukama, kao neke vrste uvoda u modeliranje u populacionoj dinamici, koja ce biti izložena u sledećoj glavi. U suštini, svi izloženi modeli u potpoglavlju 1.1.4 su posledica principa da je ubrzanje tela proporcionalno sa silama koje dejstvuju na njega, a u potpoglavlju 1.1.5 su posledica principa održanja mase. U zadnjem potpoglavlju 1.1.6 ove glave dajemo još nekoliko modela iz prirodnih nauka koji se takođe svode na obične diferencijalne jednačine.

1.1.4 Modeli dobijeni korišćenjem drugog Njutnovog zakona

Primer 1.1 (Sistem masa-opruga na horizontalnoj površini)
Jedan od najprostijih fizičkih sistema je onaj koji se sastoji od tela mase m obešene na oprugu. Pretpostavićemo da su dimenzije tela zanemarljive i da se ono može kretati samo u jednom horizontalnom pravcu, koji ćemo obeležiti sa x (slika 1.3). Na toj slici je sa $\vec{V}$ obeležen trenutni smer kretanja tega (dakle, ulevo).

Ako je F intenzitet spoljne sile, onda je na osnovu drugog Njutnovog zakona jednačina kretanja tela data sa

$$m\frac{d^2x}{dt^2} = F \tag{1.1}$$



Slika 1.3: Sistem masa-opruga na horizontalnoj površini.

Jasno, ako nema spoljnih sila, tada telo miruje, ili se kreće samo konstantnom brzinom (Njutnov prvi zakon). Tačku $x = 0$ ćemo izabrati tako da je u njoj posmatrani sistem u ravnoteži - na slici 1.3 to je obeleženo sa vertikalnom isprekidanom linijom. To znači da opruga nije ni nategnuta niti sabijena. Ako sada pomerimo telo u desno ili

u levo, onda se opruga pokušava vratiti u početni položaj $x = 0$. Sila koja se tu javlja je proporcionalna sa otklonom x tela od početnog položaja. Ako zanemarimo otpor trenja, ali i sve ostale otpore između podloge i tela, onda je desna strana (1.1) jednaka

$$m\frac{d^2x}{dt^2} = -k_1x. \tag{1.2}$$

Ovde je $k_1 > 0$ konstanta[1], dok znak "$-$" u (1.2) potiče od suprotnog smera restitucione sile od smera kretanja tela.

Ako sada uzmemo u obzir trenje koje se javlja između tela i ravne površine po kojoj se kreće, onda umesto (1.2) dobijamo jednačinu

$$m\frac{d^2x}{dt^2} = -k_1x - k_2\frac{dx}{dt} \tag{1.3}$$

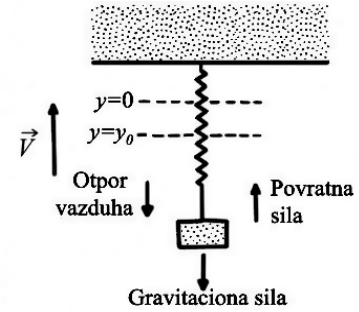
(gde je k_2 koeficijent trenja), jer je sila otpora sredine koja se javlja pri kretanju proporcionalna sa brzinom kretanja, ali je suprotnog smera.

Jednačine (1.2) i (1.3) su obične linearne diferencijalne jednačine drugog reda sa konstantnim koeficijentima. Ostavljamo čitaocu da odredi rešenje jednačine (1.3), uz početne uslove

$$x(0) = x_0, \quad \frac{dx}{dt}(0) = v(0) = v_0, \tag{1.4}$$

i to u zavisnosti od odnosa konstanti k_1, k_2 i m.

Primer 1.2 (Sistem masa-opruga sa gravitacijom)
Neka je, za razliku od prethodnog primera, opruga sa tegom mase m obešena, što znači da se ceo sistem (teg i opruga) može kretati u jednom vertikalnom pravcu, koji ćemo obeležiti sa y (slika 1.4). Na toj slici je sa $\vec{V}$ obeležen trenutni smer kretanja tega (dakle, nagore).



Slika 1.4: Sistem masa-opruga sa gravitacionom silom

Ako sa g označimo gravitacionu silu po jedinici mase, onda dolazimo do obične diferencijalne jednačine

$$m\frac{d^2y}{dt^2} = -k_1y - k_2\frac{dy}{dt} - mg. \tag{1.5}$$

[1]Ako nije drugačije naglašeno, sve konstante koje se pojavljuju u modelima su pozitivne.

Znak "$-$" u (1.5) potiče od dejstva gravitacione sile nadole, tj. u smeru suprotnom od pozitivnog smera y-ose.

Ako stavimo $p = k_2/m$ i $\omega^2 = k_1/m$, tada (1.5) postaje

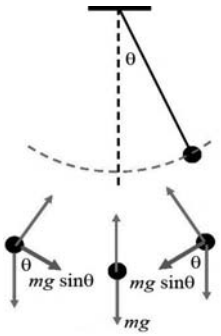
$$\frac{d^2y}{d^2t} + p\frac{dy}{dt} + \omega^2 y = -g. \tag{1.6}$$

Dobijena nehomogena obična diferencijalna jednačina drugog reda sa konstantnim koeficijentima se smenom $z = y + g/\omega^2$ svodi na

$$\frac{d^2z}{d^2t} + p\frac{dz}{dt} + \omega^2 z = 0, \tag{1.7}$$

tj. homogenu linearnu običnu diferencijalnu jednačinu drugog reda sa konstantnim koeficijentima. Ostavljamo čitaocu da uporedi ovu jednačinu sa (1.3).

Primer 1.3 (Klatno)
Klatno se sastoji od obešene kugle mase m, tanke krute šipke dužine ℓ, pri čemu masu šipke zanemarujemo (slika 1.5).



Slika 1.5: Klatno.

Pretpostavićemo da se klatno kreće samo u vertikalnoj ravni koja sadrži šipku. Kada se spoljnom silom klatno pomeri iz položaja ravnoteže, onda se pojavljuje ugao otklona, koji ćemo obeležiti sa θ. U drugom redu slike 1.5 su date tri razne mogućnosti za θ, pri čemu srednja, koja odgovara uglu $\theta = 0$, je prvi od dva položaja ravnoteže. Drugi položaj ravnoteže odgovara uglu $\theta = \pi$.

Kada se klatno izvede iz ravnoteže, smatraćemo da je jedina povratna sila gravitacija. Komponenta gravitacione sile koja je tangencijalna u odnosu na putanju klatna, tzv. restituciona sila, je na osnovu slike 1.5 jednaka $mg\sin\theta$. Za nju ćemo pretpostaviti da je proporcionalna sa brzinom kuglice, tj. sa

$$k_1\ell\frac{d\theta}{dt},$$

za neko $k_1 > 0$.

Sada iz drugog Njutnovog zakona sledi

$$m\ell\frac{d^2\theta}{d^2t}+k_1\ell\frac{d\theta}{dt}+mg\sin\theta=0. \tag{1.8}$$

Za male uglove θ, kada je $\sin\theta\approx\theta$, gornja nelinearna jednačina (1.8) prelazi u linearnu običnu diferencijalnu jednačinu drugog reda sa konstantnim koeficijentima, kao što su i prethodno dobijene (1.3) ili (1.7), ali, naravno, sa drugim koeficijentima.

1.1.5 Modeli dobijeni korišćenjem principa održanja mase

Drugi važan princip kod konstrukcije matematičkih modela u obliku obične diferencijalne jednačine je tzv. princip **održanja mase**. Taj princip kaže da je stopa po kojoj se količina supstance u nekoj sredini menja proporcionalna sa stopom po kojoj se količina te supstance povećava unošenjem u sredinu umanjenom za stopu za koju se ista smanjuje napuštanjem te sredine. Grubo, to bi se moglo opisati kao

$$\text{stopa promene} = k_1\cdot\text{ulazna stopa} - k_2\cdot\text{izlazna stopa}. \tag{1.9}$$

gde su k_1 i k_2 konstante proporcionalnosti.

Primer 1.4 (Rezervoar za vodu)
Posmatrajmo rezervoar za vodu u koji se voda uliva sa stalnom stopom po jedinici zapremine, a voda isparava sa stopom koja je proporcionalna sa $V^{2/3}$, gde je $V=V(t)$ zapremina vode u rezervoaru u momentu t. Pod ovim pretpostavkama je stopa promene $\dfrac{dV}{dt}$ jednaka (uporediti sa (1.9)):

$$\frac{dV}{dt}=k_1-k_2V^{2/3}, \tag{1.10}$$

gde su k_1 i k_2 pozitivne konstante.

Primer 1.5 (Biohemijska reakcija)
U mnogim biohemijskim reakcijama se pojavljuje supstrat i encim, čije ćemo koncentracije obeležiti sa S i E, što dovodi do jedinjenja koncentracije K.

Ovde imamo reverzibilnu reakciju, koja je proporcionalna sa stopom k_+ nastajanja novog jedinjenja ("$E+S$"), a koje se može raspasti na svoje sastojke sa stopom k_-. Ako energija reakcije postane dovoljno visoka, encim može dejstvovati na supstrat tako da se stvori novi proizvod koncentracije P. Ako se to dogodi, encim se oslobađa, čime ponovo postaje aktivan. Sada je to ireverzibilna reakcija čija je stopa proporcionalna sa C i stopom k_1. Na kraju, novostvoreno jedinjenje koncentracije P se dezaktivira u jedinjenje M ("mrtvo") proporcionalno sa svojom koncentracijom k_1 i stopom k_2.

Šematski se ovaj lanac reakcija može predstaviti na sledeći način:

$$E+S \quad\leftrightarrows\quad K\longrightarrow P+E\longrightarrow D \tag{1.11}$$

odnosno sledećeg sistema običnih diferencijalnih jednačina:

$$\begin{aligned}
\frac{dC}{dt} &= k_+(E+S)-(k_1+k_-)C \\
\frac{dP}{dt} &= k_1C-k_2P \\
\frac{dS}{dt} &= -k_+(E+S)+k_-C \\
\frac{dD}{dt} &= k_2P.
\end{aligned} \tag{1.12}$$

Primetimo da je zbir desnih strana ovog sistema običnih diferencijalnih jednačina jednak 0, pa je veličina $C+P+S+D$ konstantna, što je u skladu sa principom održanja mase. Takođe je i $E+C$ konstanta.

1.1.6 Razni modeli iz prirodnih nauka

Primer 1.6 (Rastvaranje hemikalije)
Jedna hemikalija se rastvara u vodi brzinom koja je proporcionalna proizvodu nerastvorene količine i razlike izmedju koncentracije u zasićenom rastvoru i postojećem rastvoru. Poznato je da je u 100 grama zasićenog rastvora rastvoreno tačno 60 grama. Ako je još poznato da je 40 grama te hemikalije stavljeno u 100 grama vode i da se posle 2 sata rastvorilo 10 grama, odredićemo kolika će biti količina rastvorene hemikalije posle 6 sati.

U tom cilju, Neka je $H = H(t)$ količina **nerastvorene** hemikalije u momentu t izražena u gramima. Tada je

$$\frac{dH}{dt} = RH(t)\left(\frac{60}{100}-\frac{40-H(t)}{100}\right),$$

ili

$$\frac{dH}{dt} = RH(t)\left(\frac{1}{5}+\frac{H(t)}{100}\right), \tag{1.13}$$

gde je R faktor proporcionalnosti.

Diferencijalnu jednačinu (1.13) lako rešavamo, jer razdvaja promenljive. Dobijamo da je

$$\ln\frac{H(t)}{H(t)+20} = \frac{R}{5}(t+C),$$

ili

$$H(t) = \frac{20}{e^{-R\cdot(t+C)/5}-1}. \tag{1.14}$$

Faktor R i konstantu C ćemo odrediti iz uslova $H(0) = 40$ i $H(2) = 30 = 40 - 10$. Zamenom ovih vrednosti u jednačinu (1.14) dobijamo

$$R = \frac{5}{2} \cdot \ln \frac{9}{10}, \quad C = 2\frac{\ln(3/2)}{\ln(10/9)}.$$

Sada iz (1.14) konačno dobijamo

$$H(t) = \frac{20}{\frac{3}{2} \cdot (\sqrt{10}/3)^t - 1}. \tag{1.15}$$

Dakle, tražena količina rastvorene hemikalije je

$$40 - H(6) = 40 - \frac{20}{\frac{3}{2} \cdot (\frac{1000}{729}) - 1} = 21.09 \text{ grama}.$$

Primer 1.7 (Koncentracija amina)
Iz diferencijalne jednačine

$$\frac{d[(CH_3)_3N]}{dt} = -k[(CH_3)_3N][CH_3I] \tag{1.16}$$

odredićemo $[(CH_3)_3N]$ kao funkciju od vremena, ako su početne koncentracije reagensa A_0 i B_0 dobijene na osnovu reakcije

$$(CH_3)_3N + CH_3I \xrightarrow{k} (CH_3)_3N^+ + I^-.$$

Ako sa x označimo koncentraciju amina, tada se data diferencijalna jednačina (1.16) može zapisati u sledećem obliku

$$\frac{d(A_0 - x)}{dt} = -k(A_0 - x)(B_0 - x),$$

ili

$$\frac{-1}{(A_0 - x)(B_0 - x)}dx + kdt = 0.$$

Ovo je diferencijalna jednačina koja razdvaja promenljive, čije je rešenje

$$\frac{1}{B_0 - A_0}\ln(A_0 - x) + \frac{1}{A_0 - B_0}\ln(B_0 - x) + kt = C.$$

Ako pretpostavimo da je $x(0) = 0$, onda dobijamo

$$C = \frac{\ln(B_0/A_0)}{A_0 - B_0},$$

odakle je

$$kt = \frac{1}{A_0 - B_0} \ln\left(\frac{B_0(A_0 - x)}{A_0(B_0 - x)}\right),$$

odnosno

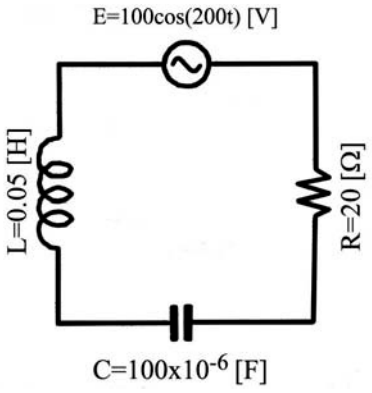
$$\ln[(CH_3)_3N] = (A_0 - B_0)kt + \ln[CH_3I] + \ln\frac{A_0}{B_0}.$$

Odavde se lako nalazi $[(CH_3)_3N]$.

Primer 1.8 (Zatvoreno električno kolo)
Posmatrajmo zatvoreno električno kolo sa elektromotornom silom $E = E(t)$, otpornikom R, kondenzatorom C i kalemom L (slika 1.6).

Ako sa $I = I(t)$ obeležimo jačinu struje, a sa $Q = Q(t)$ količinu elektriciteta u kolu, tada je

$$I(t) = \frac{dQ(t)}{dt}. \tag{1.17}$$



Slika 1.6: Električno kolo.

Poznato je da važi relacija

$$L \cdot \frac{dI}{dt} + R \cdot I + \frac{Q}{C} = E(t). \tag{1.18}$$

Koristeći jednačine (1.17) i (1.18), odredićemo jačinu struje $I(t)$ ako je $L = 0.05$ henrija, $R = 20$ oma i $C = 100$ mikrofarada uz elektromotornu silu

$$E(t) = 100\cos(200t),$$

i početne uslove

$$I(0) = Q(0) = 0.$$

Na osnovu (1.17), jednačina (1.18) se može napisati u obliku

$$L \cdot \frac{d^2Q}{dt^2} + R \cdot \frac{dQ}{dt} + \frac{Q}{C} = E(t),$$

odnosno u našem slučaju

$$\frac{d^2Q}{dt^2} + 400 \cdot \frac{dQ}{dt} + 2 \cdot 10^5 Q = 2 \cdot 10^3 \cos(200t). \tag{1.19}$$

Dobili smo nehomogenu diferencijalnu jednačinu drugog reda sa konstantnim koeficijentima, čije je rešenje

$$\begin{aligned} Q(t) &= 0.01 \cdot \cos(200t) + 0.005 \cdot \sin(200t) \\ &\quad + e^{-200t}(C_1 \cdot \cos(400\,t) + C_2 \cdot \sin(400\,t)). \end{aligned} \tag{1.20}$$

Konstante C_1 i C_2 se nalaze iz početnih uslova: $C_1 = -0.01$ i $C_2 = -0.0075$. Dakle,

$$\begin{aligned} Q(t) &= 0.01 \cdot \cos(200t) + 0.005 \cdot \sin(200t) \\ &\quad + e^{-200t}((-0.01) \cdot \cos(400\,t) + (-0.0075) \cdot \sin(400\,t)). \end{aligned}$$

Odavde sledi da je tražena jačina struje $I(t)$ data sa

$$I(t) = \cos(200t) - 2\sin(200t) + e^{-200t}(-\cos(400\,t) + 5.5\sin(400\,t)).$$

Primer 1.9 (Kretanje broda)
Brod težine Q_b sa posadom težine Q_p kreće se pravolinijski po površini mirne vode sa brzinom v_0 u trenutku isključivanja motora. Odredićemo zakon kretanja broda, ako je otpor vode proporcionalan brzini.

Neka je $v = v(t)$ brzina broda u momentu t. Tada je premu uslovu zadatka

$$\frac{1}{g}(Q_b + Q_p)\frac{dv}{dt} = -kv,$$

gde je g gravitaciona konstanta, a $k > 0$ konstanta proporcionalnosti. Opšte rešenje ove jednačine je

$$v(t) = C\exp(-\frac{kg}{Q_b + Q_p}t).$$

Kako je početna brzina broda $v(0) = v_0$, to je konačno rešenje

$$v(t) = v_0\exp(-\frac{kg}{Q_b + Q_p}t).$$

Primer 1.10 (Ravnotežno stanje provođenja toplote)
Posmatraćemo homogenu cilindričnu dugačku šipku. Radi jednostavnosti, pretpostavićemo da je tražena raspodela temperature ista na celom poprečnom preseku, ali se temperatura može menjati od jednog poprečnog preseka do drugog.

Neka se osa simetrije šipke dužine $b > 0$ poklapa sa $x-$osom i neka je početak šipke u tački $x = 0$. Pretpostavimo, dalje, da se temperatura šipke na početku ($x = 0$) održava na stalnoj temperaturi T_1, a da se na kraju šipke ($x = b$) održava na stalnoj temperaturi T_2.

Poznato je iz fizike da nakon dovoljno dugog vremena temperatura unutar šipke dostiže ravnotežno stanje, ili, drugim rečima, ne menja se sa daljim protokom vremena. Može se pokazati da ravnotežna raspodela temperature, koju označavamo sa $y = y(x)$, i koja zavisi samo od mesta x, $0 < x < b$, zadovoljava diferencijalnu jednačinu

$$y''(x) = -\frac{1}{k}F(x), \quad 0 < x < b, \tag{1.21}$$

sa graničnim uslovima

$$y(0) = T_1, \quad y(b) = T_2, \tag{1.22}$$

gde je F stopa zagrevanja po jedinici zapremine, a k pozitivna konstanta. Negativan znak u relaciji (1.21) označava da se toplota širi od toplijih ka hladnijim delovima.

Odredićemo ravnotežnu raspodelu temperature u šipki dužine 1, ako je šipka izolovana sa strane i nema drugih izvora toplote unutar šipke, a krajevi šipke se održavaju na stalnim temperaturama T_1 i T_2.

Gornje prepostavke povlače da je u relaciji (1.21) $F(x) = 0$, odnosno da treba rešiti problem

$$y'' = 0, \quad 0 < x < 1, \quad y(0) = T_1, \quad y(1) = T_2. \tag{1.23}$$

Opšte rešenje ove diferencijalne jednačine je

$$y = C_1 + C_2 x.$$

Primenom graničnih uslova, dobijamo

$$y(0) = C_1 = T_1, \qquad y(1) = C_1 + C_2 = T_2,$$

odakle je $C_1 = T_1$, a $C_2 = T_2 - T_1$. Prema tome, rešenje problema (1.23) je

$$y = T_1 + (T_2 - T_1)x.$$

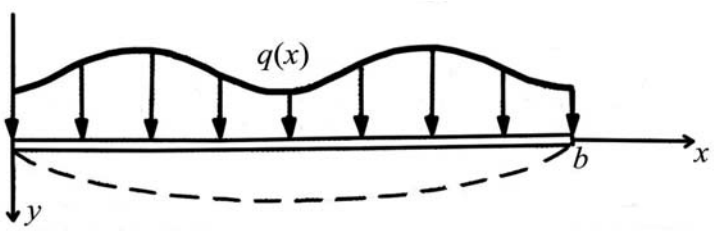
Važno je primetiti da se mnogi fizički problemi svode na granične probleme tipa (1.21), (1.22).

Primer 1.11 (Odstupanje od ravnotežnog položaja elastične žice)
Posmatraćemo homogenu žicu učvršćenu na krajevima $x = 0$ i $x = b$ na koju deluje vertikalna sila intenziteta $q(x)$ po jedinici dužine. Pretpostavlja se da na žicu deluje jaka sila T koja ostaje konstantna za mala odstupanja y (slika 1.7).

Može se pokazati da, pod određenim pretpostavkama, odstupanje y žice od ravnotežnog položaja zadovoljava jednačinu ravnoteže žice

$$y'' = -\frac{q(t)}{T}, \quad 0 < x < b,$$

koja se zove **Poasonova jednačina**.



Slika 1.7: Odstupanje žice od ravnotežnog položaja.

Na krajevima $x = 0$, $x = b$ žica može biti pričvršćena i tada je odstupanje y na krajevima jednako nuli. Ako su krajevi slobodni, tada je $y' = 0$, za $x = 0$ i $x = b$.

Kao prvi zadatak, odredićemo oblik žice pričvršćene na krajevima $x = 0$ i $x = \pi$, na koju deluje konstantna sila q_0 po jedinici dužine. U drugom zadatku ćemo odrediti maksimalno odstupanje žice kao i silu koja je potrebna da maksimalno odstupanje ne pređe veličinu $0.3q_0$.

Prvi zadatak odgovara sledećem graničnom problemu:

$$y'' = -\frac{q_0}{T}, \quad 0 < x < \pi \quad y(0) = 0, \quad y(\pi) = 0.$$

Opšte rešenje gornje diferencijalne jednačine je

$$y = -\frac{q_0}{2T}x^2 + C_1 x + C_2.$$

Iz graničnih uslova se dobija da je $C_1 = \dfrac{q_0\pi}{2T}$, a $C_2 = 0$. Prema tome dobija se da je odstupanje žice parabolično, tj.

$$y = \frac{q_0}{2T}x(\pi - x). \tag{1.24}$$

Da bi rešili drugi zadatak, primetimo prvo da parabola (1.24) dostiže maksimum y_{max} za $x = \pi/2$, čija je vrednost

$$y_{max} = \frac{q_0\pi^2}{8T}.$$

Sada rešavanjem po T sledi da nejednakost $y_{max} \leq 0.3q_0$ važi za

$$T \geq \frac{\pi^2}{2.4} \approx 4.112$$

Primer 1.12 (Rast populacije bakterija)
Neka je $N(t)$ broj jedinki jedne populacije bakterija u momentu t i pretpostavimo da je stopa rasta populacije bakterija, odnosno $\dfrac{dN}{dt}$, proporcionalna sa $N(t)$, pri čemu ćemo konstantu proporcionalnosti označiti sa R_0.

Izračunaćemo koliko je vremena potrebno da bi se populacija bakterija povećala 10 puta, ako se zna da se jedna populacija bakterija duplira za 24 sata. Uzećemo da je je ova populacija u momentu $t = t_0 = 0$ imala N_0 jedinki.

Neka je $N = N(t)$ broj jedinki posmatrane populacije u momentu t (koje merimo u satima). Treba rešiti početni problem

$$\frac{dN}{dt} = R_0 N(t), \quad N(0) = N_0, \tag{1.25}$$

posle čega ćemo R_0 naći iz uslova dupliranja populacije. Rešenje problema (1.25) je

$$N(t) = N_0 e^{R_0 t}.$$

Zamenjujući da je $t = 24$, dobijamo da je

$$N(24) = N_0 e^{24 \cdot R_0},$$

a prema uslovu zadatka je $N(24) = 2N_0$, te je

$$R_0 = \frac{1}{24} \ln 2 = 0.029.$$

Sada tražimo t_1 sa osobinom da je

$$N(t_1) = 10 \cdot N(0) = 10 \cdot N_0.$$

Rešenje je

$$t_1 = 24 \frac{\ln 10}{\ln 2} = 79.726 \text{ sati } \approx \text{ 3 dana, 7 sati i 44 minuta.}$$

Prethodni primer je iz populacione dinamike, kojom ćemo se baviti u sledećoj, drugoj glavi.

1.2 Simulacija

Simulaciju čini skup metoda koje imitiraju ponašanje realnog sistema i obuhvataju proces izgradnje apstraktnih modela za neke sisteme ili podsisteme realnog sveta i obavljanje većeg broja eksperimenata nad njima. U slučaju kada se ti eksperimenti odvijaju na računaru govorimo o računarskom modeliranju i simulaciji. Uopšte, možemo konstatovati da popularnost simulacije raste sa razvojem računara i softvera za simulaciju.

Kao i sve metode za analizu sistema i simulacija se oslanja na model sistema napravljenog da oponaša realan sistem. Proučavanje realnog sistema se svodi na merenje performansi sistema, poboljšanje rada sistema ili kreiranje novog sistema. Upotreba

modela nekog sistema omogućava analizu rezultata simulacije ili na praćenje ponašanja modela u toku simulacije da bi se bolje razumelo kako sistem radi. Izraz modeliranje i simulacija izražava složenu aktivnost koja uključuje tri elementa: realni sistem, model i računar.

Pod realnim sistemom podrazumevamo uređen, međuzavisan skup elemenata koji formiraju jedinstvenu celinu i deluju zajednički kako bi ostvarili zadati cilj ili funkciju, bez obzira da li se radi o prirodnom ili veštačkom sistemu, i takođe, bez obzira da li taj sistem u posmatranom trenutku postoji ili se njegovo postojanje planira u budućnosti. Realni sistem je izvor podataka o ponašanju, a ovi se podaci javljaju u obliku zavisnosti $X(t)$, gde je X bilo koja promenljiva koja interesuje istraživača, a t je vreme mereno u odgovarajućim jedinicama.

Model, kao i svaki realni sistem, ima svoje objekte koji se opisuju atributima ili promenljivima. On je apstraktni prikaz sistema i daje njegovu strukturu, njegove komponente i njihovo uzajamno delovanje. Kada se eksperimenti odvijaju na računaru govorimo o računarskom modeliranju i simulaciji. Tada, pod modelom se može podrazumevati skup instrukcija (program) koji služi da se generiše ponašanje simuliranog sistema. Ponašanje modela ne mora da bude u potpunosti jednako ponašanju simuliranog sistema, već samo u onom domenu koji je od interesa.

Računar, kao treća komponenta ove aktivnosti, predstavlja uređaj sposoban za izvršavanje instrukcija modela.

Prilikom modeliranja uspostavlja se veza između realnog sveta i modela; ta veza se naziva relacija modeliranja i odnosi se na validnost modela. Validnost ili valjanost modela opisuje koliko verno jedan model predstavlja simulirani sistem. Proces utvrđivanja stepena slaganja podataka o realnom sistemu sa podacima modela naziva se validacija modela i na osnovu procesa validacije se odlučuje o upotrebljivosti rezultata simulacije, izmeni modela ili podataka, itd.

Eksperimentisnje sa realnim sistemom ima određene prednosti pod uslovom da u toku eksperimenta postoji sigurnost da neće doći do značajnih promena u sistemu koje bi uticale na njegovu stabilnost. Tada nije potrebno praviti model sistema i voditi računa da li model realno predstavlja sistem. U većini slučajeva je teško, skupo ili nemoguće eksperimentisati nad realnim sistemom. Posledice takvog eksperimenta mogu da budu katastrofalne (tragičan primer je nuklearna elektrana Černobilj, gde je posle eksperimenta sa reaktorom došlo do eksplozije). Promena rasporeda mašina u fabrici, promena broja radnika na šalteru banke, otvaranje nove piste na aerodromu, reorganizacija rada hitne pomoći, ili sistem nastave i ocenjivanja u prosveti, ili globalni ekonomski sistem su samo neki od slučajeva gde nije poželjno eksperimentisati na realnom sistemu. Naravno, to ne znači da ti sistemi ne treba da se vremenom reformišu i modernizuju.

Istraživanja su pokazala da je najpopularniji metod u operacionim istraživanjima simulacija. Pored već pomenute mogućnosti da se nosi sa složenim modelima, popularnost simulacije raste zbog razvoja računara. Od vremena korišćenja programiranja niskog nivoa koje je sklono greškama na računarima koji su bili skupi i neeefikasni, došlo se do brzih i jeftinih računara koji efikasno izvršavaju složene modele uz pomoć

softvera koji je jednostavan za upotrebu, brz i fleksibilan.

Poslovna primena simulacije osamdesetih godina se razvijala zahvaljujući pojavi personalnih računara. Takođe, u ovom periodu se razvija sa simulacijom i animacija. Iako se još uvek koristila uglavnom za analizu sistema koji ne funkcionišu kako je predviđeno, mnoge kompanije su zahtevale da se uradi simulacija pre investicije sredstava. Svoj pravi oblik i primenu simulacija dobija tek devedesetih godina kada i manje kompanije počinju da koriste sve dostupnije alate za simulaciju. Softver za simulaciju postaje jednostavan za upotrebu, brz, sa mogućnosti integrisanja sa drugim softverskim paketima.

Najvažniji razlozi za primenu modeliranja i simulacije sistema su kada je eksperiment nad realnim sistemom može da bude skup ili nemoguć, zatim ako analitički model nema analitičko rešenje, ili ako je realni sistem suviše složen da bi se opisao analitički.

Pored ovih, u nastavku dajemo još nekoliko značajnih razloga za korišćenje modeliranja i simulacije:

- izgradnja modela i sama simulacija ponekad imaju za cilj da se shvati funkcionisanje postojećeg sistema čija je struktura nepoznata i ne može joj se prići;
- prilikom iznalaženja optimalnog funkcionisanja nekog sistema, uobičajeno je da se menjaju razni parametri, što je često neizvodljivo sa realnim sistemom, bilo zato što takvog sistema uopšte nema ili bi takav eksperiment bio preskup;
- ponekad je potrebno simulirati uslove pod kojima dolazi do razaranja sistema;
- vreme se pri simulaciji može sažeti ili produžiti;
- ponekad je potrebno zaustaviti dalje odvijanje eksperimenta, kako bi se ispitale vrednosti svih promenljivih u tom trenutku.

1.2.1 Podela simulacionih modela

Postoje dva osnovna načina podele simulacionih modela, i to prema vrsti promenljivih u modelu, a druga prema načinu na koji se stanja u modelu menjaju u vremenu.

Prva podela je na osnovu vrsti promenljivih u modelu na **determinističke** i **stohastičke modele**. Deterministički modeli su oni u kojima je novo stanje sistema potpuno određeno prethodnim stanjem. Sa druge strane, stohastički modeli su oni čije se ponašanje ne može unapred sa sigurnošću predvideti, ali se često mogu odrediti verovatnoće promena stanja sistema. Generalno je za stohastičke modele karakteristično postojanje slučajnih promenljivih u sistemu.

Druga podela je prema načinu na koji se stanja u modelu menjaju u vremenu na **diskretne modele**, **kontinualne modele** i **kontinualno-diskretne modele**.

Kod diskretnih modela se stanja sistema menjaju samo u pojedinim diskretnim momentima, dok se kod kontinualnih modela stanja sistema menjaju kontinualno u vremenu. Kontinualno-diskretni modeli sadrže i kontinualne i diskretne promenljive.

1.2.2 Simulacija diskretnih sistema

U diskretnim modelima stanje sistema se menja samo u pojedinim tačkama u vremenu. Te promene stanja entiteta sistema se nazivaju događaji. Simulacija diskretnih događaja bavi se modeliranjem sistema koji se mogu predstaviti skupom događaja i simulacija treba da opiše svaki diskretan događaj, krećući se od jednog do drugog događaja pri čemu se pomera vreme simulacije.

Vreme simulacije se može meriti na dva načina: ili se izabere minimalna jedinica vremena ili se mere intervali u kojima dolazi do promene stanja sistema. Između dva događaja stanje sistema se ne menja. Ova simulacija se najčešće koristi za analizu dinamičkih sistema sa stohastičkim karakteristikama. Banka, samoposluga, telefonska centrala, ali i drugi sistemi masovnog opsluživanja su primeri sistema sa diskretnim događajima.

Kod modela sistema sa diskretnim događajima, pored koncepata koji opisuju strukturu (objekti, relacije između objekata, atributi objekata), uvedeni su i koncepti za opis dinamike. To su:

- događaj – diskretna promena stanja entiteta u sistemu;
- aktivnost – skup događaja koji menjaju stanje jednog ili više entiteta, pri čemu se trajanje aktivnosti u nekim slučajevima može unapred definisati, a u nekim zavisi od ispunjenja određenih uslova u modelu;
- proces – niz uzastopnih, logički povezanih događaja kroz koje prolazi neki privremeni objekat, odnosno to je hronološki uređena sekvenca koja opisuje jednu pojavu od nastajanja do uništavanja.

1.2.3 Prednosti i nedostaci simulacije

Bavljenje simulacionim metodama i tehnikama zahteva poznavanje njenih prednosti i nedostataka. Glavne **prednosti** korišćenja simulacije su:

- mogućnost višestrukog korišćenja istog modela;
- simulacioni podaci se najčešće mogu mnogo jeftinije dobiti od odgovarajućih podataka iz realnog sistema;
- analitički modeli uglavnom zahtevaju više pojednostavljenih pretpostavki koje ih čine matematički prilagodljivim, dok simulacioni modeli nemaju ovakva ograničenja;
- simulacioni eksperiment se može ponoviti više puta.

Osnovni **nedostaci** korišćenja simulacije su sledeći:

- simulacioni modeli mogu biti skupi i mogu zahtevati značajno vreme za izgradnju i validaciju;

- potrebno je izvođenje većeg broja simulacionih eksperimenata kako bi dobili odgovarajući uzorak, a to pored memorije računara, može zahtevati dosta vremena i napora;

- na osnovu simulacije često nije moguće uočiti zavisnost izlaznih promenljivih od ulaznih promenljivih,

- simulacija nije optimizacija, pa se pomoću nje optimalna rešenja retko mogu pronaći;

- validacija modela je složena i zahteva dodatne eksperimente.

Simulacioni proces je struktura rešavanja stvarnih problema pomoću simulacionog modeliranja. On se može prikazati u obliku niza koraka koji opisuju pojedine faze rešavanja problema ovom metodom (životni ciklus simulacije). Struktura simulacionog procesa nije strogo sekvencijalna, već je moguć i povratak na prethodne korake procesa, zavisno od rezultata dobijenih u pojedinim fazama procesa. Broj faza i redosled njihovog obavljanja zavisi od svake konkretne situacije, ali je ipak moguće navesti jedan opšti, uređen skup procedura.

1.2.4 Tehnike u simulacionim modelima

Glavni pristupi izrade modela su sistemska dinamika, diskretni događaji, sistemi zasnovani na agentima i dinamički sistemi.

Sistemska dinamika i diskretni događaji su tradicionalan pristup kreiranju modela i pomoću njih realizujemo većinu kontinualnih procesa, dok su sistemi zasnovani na agentima razvijeni relativno skoro.

Većina modela zasnovanih na agentima mogu se razviti iz postojećih modela sistemske dinamike i modela diskretnih događaja. Pri tome dobijamo bolju realizaciju složenog ponašanja objekata. Zavisnost između objekata je predstavljena finije i time dobijamo preciznije rezultate na kraju simulacije. Modeli zasnovani na agentima uglavnom rade u diskretnom vremenu. Oblast dinamičkih sistema je uglavnom rezervisana za realizaciju modela na fizičkom nivou, odnosno modela koji imaju nizak nivo apstrakcije. Na visokom nivou apstrakcije se koristi sistemska dinamika. Diskretni događaji se koriste u niskom i srednjem nivou. Modeli zasnovani na agentima su primenljivi na svim nivoima. Na najnižem nivou pomoću agenata mogu se predstaviti na primer pešaci ili vozila, na srednjem nivou klijenti, dok na najvišem nivou agent predstavlja, na primer, kompaniju.

Sistemska dinamika je pristup modeliranju koji prikazuje dinamiku promena u kompleksnim sistemima. Metod za kreiranje modela razvio je inženjer Dž. V. Forester (*Jay W. Forester*) pedesetih godina prošlog veka. Oblasti primene ovog pristupa je modeliranje populacionih, socijalnih, ekoloških i ekonomskih sistema. U ovim modelima dešavanja u realnom svetu su prikazana kroz skupove (na primer: materijala,

ljudi, novca, znanja, itd.), tokove između skupova i informacija koje određuju vrednosti tokova.

Slika 1.8: Model sistemske dinamike.

Kod kreiranja modela potrebno je definisati ponašanje određenog skupa. Modeli sistemske dinamike su na veoma visokom nivou apstrakcije, što znači da se zanemaruje ponašanje pojedinačnih objekata jednog skupa već se ono definiše za ceo skup srodnih objekata. Dakle, za vreme izvršavanja modela ne možemo razlikovati pojedinačne objekte. Onaj ko razvija model mora da razmišlja o globalnoj strukturi modela, zavisnostima u njemu i mora da obezbedi dovoljno pouzdane podatke za model.

Obratimo sada pažnju na samu prirodu ponašanja kontinualnih sistema. Razlog za složeno ponašanje kontinalnih sistema leži u povratnom uticaju (*feedback*). Povratna petlja se odnosi na situaciju gde X utiče na Y, ali i Y istovremeno utiče na X kroz niz uzroka i pojava. Nemoguće je sistem posmatrati kroz samo jednu vezu, jer se na taj način ne može predvideti ponašanje sistema. Vrednosti promenljive se vraćaju povratnom petljom i utiču na njih u sledećem koraku. Analiza povratnih petlji daje uvid u moguća ponašanja sistema. Bitno je da li povratna petlja ima pozitivan ili negativan uticaj. Ako pratimo uticaj povratne petlje, možemo brojati znake direktnog uticaja između promenljivih. Posebno, ako je funkcija f_i pozitivna, uticaj između q_{i-1} i q_i je pozitivan, što znači da će pozitivna vrednosti q_{i-1} povećati vrednost od q_i. Ako je, međutim, znak funkcije f_i negativan, pozitivna vrednost q_{i-1} će smanjiti vrednost od q_i. Pozitivna povratna petlja je ona sa parnim brojem negativnih uticaja, a negativna povratna petlja je ona sa neparnim brojem negativnih uticaja. Povezivanje skupova objekata preko povratnih petlji omogućava prikazivanje složenosti i nelinearnosti u odnosima između skupova.

Matematički model sistemske dinamike po pravilu se posmatra kao sistem diferencijalnih jednačina. U ovim modelima ne definiše se sledeće stanje direktno, već se koristi diferencijabilna funkcija radi menjanja promenljive. U jednom vemenskom

koraku za dato stanje i ulazne vrednosti zna se samo za koliko se promenilo trenutno stanje. Od ove informacije može se izračunati stanje u svakoj tački. Obično se kontinualan sistem opisuje korišćenjem više promenljivih, pa su onda i izvodi funkcije od dve ili više promenljivih. Ako su $q_1, q_2, ..., q_n$ promenljive, a $x_1, x_2, ..., x_m$ ulazne promenljive, kontunualni model se može izraziti kao sledeći sistem diferencijalnih jednačina prvog reda:

$$\begin{aligned}
\frac{dq_1(t)}{dt} &= f_1(q_1(t), q_2(t), ..., q_n(t), x_1(t), x_2(t), ..., x_m(t)) \\
\frac{dq_2(t)}{dt} &= f_2(q_1(t), q_2(t), ..., q_n(t), x_1(t), x_2(t), ..., x_m(t)) \\
&\vdots \\
\frac{dq_n(t)}{dt} &= f_n(q_1(t), q_2(t), ..., q_n(t), x_1(t), x_2(t), ..., x_m(t)).
\end{aligned}$$

Dakle, računaju se izvodi po svim promenljivima q_i u funkcijama f_i koje imaju te promenljive i ulazni vektor kao argument.

Osnovni problem simulacije ovakvih sistema je u tome što je računar diskretna mašina. Pitanje je kako dobiti dinamičko ponašanje? U simulaciji na računaru neophodan je sledeći rezultat u vremenu t_{i+1} na intervalu $[t_i, t_{i+1}]$. Pretpostavlja se da model ima kontinualno ponašanje na ovom intervalu i da su promenljive neprekidne na ovom intervalu. Računar mora na osnovu vrednosti u t_i da proceni vrednost u t_{i+1}, iako ne zna šta se dešava na zadatom intervalu, pa izbor metoda numeričke integracije može predstavljati problem. Na sreću, moderni paketi za računarsku simulaciju po pravilu sami biraju najpogodniji metod. Jasno, uvek treba biti na oprezu i treba tražiti znake koji pokazuju da metod ne daje dobre rezultate. Na primer, neočekivani rezultati simulacije su verovatno zato što je integracioni metod nestabilan, a ne zbog zanimljivog ponašanje modela.

Ovaj način kreiranja modela podržava nekoliko komercijalno dostupnih softvera, kao što su *VenSim*, *PowerSim*, *iThink* i *ModelMaker*.

Modeliranje dinamičkih sistema predstavlja izazov pošto je potrebno napraviti precizan matematički model koji opisuje sistem. Primenom niskog nivoa apstrakcije dobija se model blizak realnom sistemu koji se modelira. Stanje dinmičkog sistema opisuje skup realnih brojeva i u skladu sa načinom kreiranja modela mala promena stanja sistema izaziva malu promenu u vrednosti u skupa brojeva. Pravila ponašanja sistema opisuju sledeće stanje koje sistem zauzima na osnovu trenutnog stanja. Ova pravila su deterministička, što znači da je u datom vremenskom intervalu samo jedno buduće stanje moguće nakon trenutnog stanja sistema.

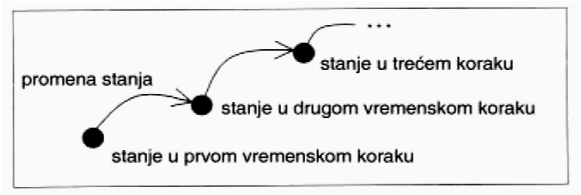
Za neki sistem smatramo da je potpuno opisan ako je određena relacija koja prevodi trenutno stanje sistema u neko buduće stanje. Ta relacija je diferencijalna ili diferencna jednačina. Iteracijama sa malim vremenskim korakom je moguće odrediti sva buduća stanja sistema, i taj se postupak naziva rešavanje ili integracija sistema. Kada se sistem reši, za zadatu počentu vrednost moguće je odrediti buduća stanja sistema. Sva

stanja sistema daju trajektoriju (putanju) sistema. Numeričke metode za rešavanje se koriste najviše i sa njima je korišćenjem računara moguće rešiti veliki broj raznovrsnih problema.

Za neke dinamičke sisteme je dovoljno poznavati trajektoriju sistema. Međutim, neki sistemi sa složenom strukturom daju trajektorije koje zahtevaju malo više analize. Ako se uzme u obzir da je sistem aproksimacija realnog sistema, numerička rešenja koja se dobijaju mogu biti nedovoljno precizna. Zato se uvodi pojam stabilnosti sistema, gde se određuje skup početnih uslova za koje će trajektorije biti približno jednake. Određivanje tipa putanje je potrebno za bolje razumevanje sistema. Neki sistemi imaju tačke gde se tip trajktorije zanačajno menja. Na primer, sistem posle promene vrednosti nekog parametra iz periodičnog prelazi u haotično ponašanje. Klasifikovanje putanja to tipu je dovelo do podele dinamičkih sistema. Najzanimljivije za proučavanje su putanje koje pokazuju da sistem ima haotično ponašanje jer tada putanja izgleda kao da je dobijena na slučajan način. Proučavanje dinamičkih sistema sa takvim putanjama dovelo je do razvoja teorije haosa, statističke mehanike i pojma determinističkog haosa.

Upoređivanjem dinamičkih sistema sa sistemskom dinamikom vidimo da je razlika u tome što promenljive ovde predstavljaju fizičke osobine objekata kao što su brzina, ubrzanje, masa itd. To omogućuje matematički složenije modele u odnosu na sistemsku dinamiku. Softver koji se koristi za opis dinamičkog sistema može se koristiti i za kreiranje modela sistemske dinamike. Predstavnici komercijalno dostupnih softvera su *MATLAB*, *VisSim*, *LabView* i *Easy5*.

Simulacija diskretnih događaja je način kreiranja modela čije će ponašanje biti praćeno kroz vreme. Postoje formalni metodi za kreiranje simulacionih modela koji obezbeđuju validnost modela. Tokom eksperimentalne faze modeli se izvršavaju i dobijaju se rezultati simulacije.



Slika 1.9: Model diskretnih događaja.

Formalizam kod diskretnih modela podrazmeva izvršavanje modela u koracima. Sledeće stanje modela zavisi od trenutnog stanja i od trenutnog uticaja okoline.

U digitalnim sistemima časovnik određuje sledeći diskretni korak. Bitna primena diskretnih modela je i kod aproksimacije kontinualnih sistema. Tada se određuje vremenska osnova (sekund, minut, sat ili godina) i na osnovu uočenih stanja sistema generiše ponašanje. Pretpostavka je da se svaki diskretni korak dobija umnožavanjem osnovnog koraka.

Najlakši način za definisanje ponašanja sistema je tabela. Pretpostavimo da sistem ima konačan broj stanja i ulaza. Jednostavno, možemo napisati sve kombinacije u

tabeli.

Trenutno stanje	Ulaz	Sledeće stanje	Izlaz
0	0	0	0
0	1	1	1
1	0	0	0
1	1	1	1

Tabela 3.1.

U tabeli postoje dva stanja 1 i 0, i dva ulaza 1 i 0. Ukupno četiri kombincije i svakoj kombinaciji je dodeljen jedan izlaz i sledeće stanje koje sistem zauzima. Tabelu je moguće intrepretirati na sledeći način: Ako je stanje q u vremenu t i ulaz je x u vremenu t, onda će stanje u vremenu $t+1$ biti $\delta(q,x)$ i izlaz y u vremenu t će biti $\lambda(q,x)$, pri čemu je δ funkcija promene stanja, a λ izlazna funkcija. Funkcija promene stanja se odnosi na prve tri kolone tabele, dok izlazna funkcija zavisi od prve dve i poslednje kolone. Korišćenjem ove dve funkcije tabela se može zapisati kao $\delta(q,x) = x$ i $\lambda(q,x) = x$.

Sledeće stanje i trenutni izlaz su zadati trenutnim ulazom. Funkcije je bolje koristiti od tabele pošto daju opštije značenje i nije potrebno pisati tabelu kada sistem ima puno stanja. Naravno funkcije se koriste i kada nisu poznata sva stanja sistema.

Oblast primene modela diskretnih događaja je velika i postoje brojni primeri primene na sistemima masovnog opsluživanja i proizvodnih procesa. Simulacija se najviše primenjuje kod razvijanja novog sistema koji zahteva velike investicije. Kod sistema koji već postoje simulacija se može koristiti za testiranje promena u dizajnu sistema i za proveru stabilnosti sistema u slučaju kvarova ili nekih nepredviđenih okolnosti.

Kreiranje modela se zasniva na konceptu objekta, resursa i dijagrama koji opisuju kretanje objekata i deljenje resursa. Metod je razvio Dž. Gordon (*Jeffrey Gordon*) šesdesetih godina. On je u saradnji sa IBM-om kreirao prvi simulacioni jezik pod nazivom *GPSS* (akronim od *General Purpose Simulation System*). Objekti u *GPSS* modelu nazivaju se transakcije. To su pasivni objekti koji mogu predstavljati ljude, delove, dokumente, poruke itd. Objekti putuju kroz dijagram gde čekaju određeno vreme u redu na obradu. Ponašanje modela je ugrađeno u elemente dijagrama toka. Resursi se zauzimaju i oslobađaju po potrebi od strane objekata. Trend u razvoju softvera za diskretnu simulaciju ide ka kreiranju grafičkih okruženja za modeliranje. U tom slučaju nije potrebno učiti specifičan programski jezik, pa je kreiranje modela moguće od strane većeg broja korisnika koji nisu informatičke struke.

Najveći broj komercijalno dostupnog softvera podržava gore opisani metod kreiranja modela. Neki od njih su za opštu, dok su neki razvijeni za posebnu primenu, na primer softveri za modeliranje u oblasti medicine ili tehnici. Modeliranje diskretnih sistema se uzima kao definicija za globalni algoritam za procesiranje objekata, obično sa stohastičkim elementima. Spomenućemo u ovoj grupi programe kao što su *Arena*,

Extend, *SimProcess* i *AutoMod*.

Postoji veliki potencijal za primenu simulacije diskretnih događaja. Nažalost, ograničena primena simulacije je zbog zastarelog i komplikovanog softvera za modeliranje i simulaciju što otežava komercijalnim korisnicima upotrebu.

1.2.5 Modeli sa agentima

Kreiranje modela korišćenjem agenta je relativno nova oblast. Drugi naziv za ovaj pristup modeliranju je tzv. ”modeliranje odozgo na gore”. Taj naziv dolazi od najbitnije osobine koju imaju modeli sa agentima-decentralizovanost. Ponašanje složenih modela sistemske dinamike se definiše globalno, odnosno za ceo skup objekata. Kod modela zasnovanih na agentima ponašanje se definiše na nivou agenta koji dalje u interakciji sa drugim agentima, u zadatom okruženju, kroz simulaciju generišu globalno ponašanje modela.

Pojava agentnog modeliranja se povezuje sa poznatim naučnicima fon Nojmanom i Ulamom kreiranjem fon Nojmanove mašine i automata. Veliki napredak u tom pravcu je bila ”Igra života” - prvi agentni model matematičara Dž. Konveja (*John Conway*). Igra života sadrži osnovne principe agentnog modeliranja i sastoji se od pojedinačinh objekata za koje su definisana jednostavna i precizna pravila. Objekti utiču jedni na druge i na osnovu toga se dobija ponašanje sistema u definisanom dvodimenzionalnom prostoru.

Logičan prelazak sa modela sistemske dinamike na agentne modele je uslovljen povećanjem mogućnosti računara. Sada nije problem definisati nekoliko hiljada agenata i izvršavati ih paralelno da bi se simuliralo ponašanje određene populacije na nekom prostoru.

Oblasti primene agentnih modela su logistika i optimizacija proizvodnje, modeliranje ponašanja potrošača, socijalna simulacija, simulacija saobraćaja i slično. Alati za agentno modeliranje mogu se koristiti i za optimizaciju modela i testiranje stabilnosti modela. Stabilnost modela se odnosi na praćenje promena koje se dešavaju u globalnom ponašanju modela posle promene ponašanja pojedinačnih agenata.

Agentno modeliranje ima veliki potencijal za dalji razvoj. Naime, u kombinaciji sa veštačkom iteligencijom mogućnosti se značajno povećavaju. Intezivno se radi na projektima tzv. veštačkog života.

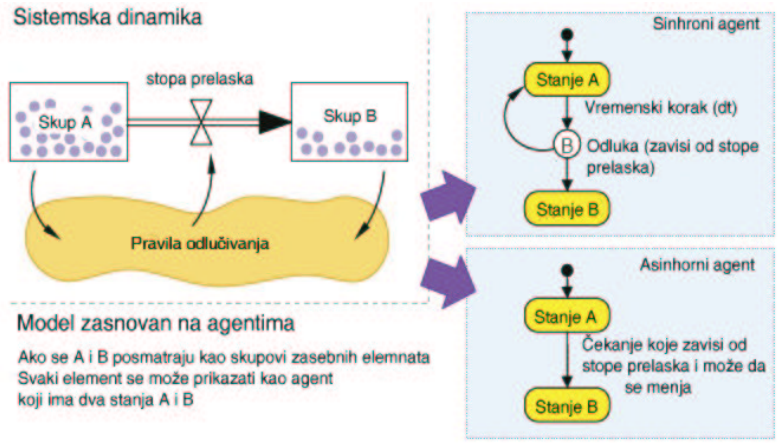
Sa spoznajom mogućnosti agentnog modeliranja se povećava broj komericajlno dostupnih alata koji uglavnom nastaju iz do skoro korišćenog akademskog softvera. Kako su agentni modeli složeni i uglavnom zahtevaju dobro poznavanje programskih jezika i programiranja za izradu modela, sve veći broj kompanija koje proizvode softver za agentno modeliranje nude i usluge kreiranja modela na zahtev klijenata.

1.2.6 Odnos između tehnika u simulacionim modelima

Ograničimo razmatranje samo na modele koji sadrže veliki broj objekata sa različitim oblicima ponašanja. U tom slučaju modeli sa agentima omogućavaju uopšteniji

pristup odnosno kompleksniju strukturu i dinamiku modela i konstrukciju modela bez znanja o globalnim zavisnostima u modelu. Dakle, ako nije poznato kako se dešavaju uticaji u modelu, ili kako se elementi modela ponašaju u tačno određenim situacijama, onda je bolje koristiti agentni pristup. Dodajmo da čak i u slučajevima kada model sistemske dinamike daje zadovoljavajuće rezultate, mnogo je lakše kreirati model zasnovan na agentima.

Na jednostavnom primeru ćemo pokazati kako se konstruiše precizniji model sa agentima na osnovu modela sistemske dinamike. Za početak, pomoću dijagrama stanja treba definisati ponašanje agenta. Dijagram stanja predstavlja automat sa nekoliko dodataka. Sam princip je predložio D. Harel (*David Harel*) i princip je prihvaćen kao deo standardnog *UML* (akronim od *Unified Modeling Language*) kasnije proširen do *UML-RT* (od *Unified Modeling Language for Real Time*). Dijagrami stanja omogućavaju da se grafički prikažu stanja kroz koje agent prolazi tokom svog postojanja. Agent može imati više dijagrama stanja koji rade paralelno i postoji interakcija između njih. To se koristi kada je potrebno opisati nekoliko aspekata života agenta kao na primer kod ljudi porodica i posao. Dijagrami stanja će kasnije biti detaljnije objašnjeni.

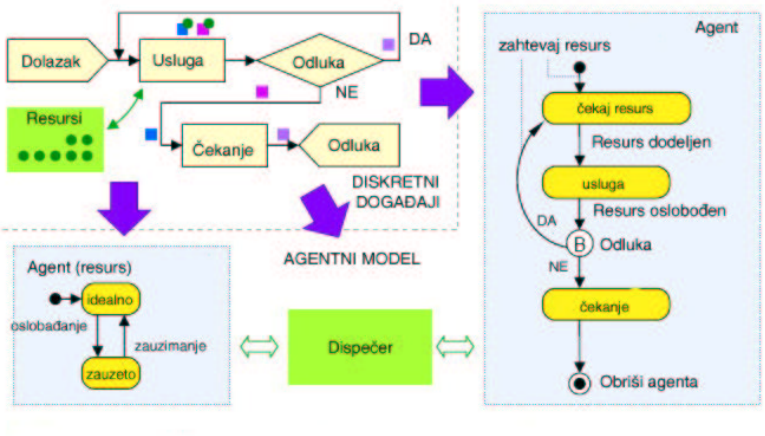


Slika 1.10: Prevođenje modela sistemske dinamike u model sa agentima.

Model sistemske dinamike posmatra kao niz skupova i tokova između njih. Pravila odlučivanja kontrolišu tokove između skupova. Skupovi koji se posmatraju mogu se zamisliti kao rezervoari sa vodom. Voda iz rezervoara A prelazi u rezervoar B sa nekom stopom prelaska. Rezervoar se mora predstaviti tako da sadrži pojedinačne objekte pa se dalje može posmatrati kao kutija napunjena lopticama. Loptice će postati agenti. Ako se izdvoji jedna loptica i posmatra njeno kretanje u jednom trenutku loptica će preći iz kutije A u kutiju B. Vreme kada će se to desiti zavisi od toga kako je tok između kutija definisan. Model zasnovan na agentima koji opisuje ovakvo ponašanje koristi dijagram stanja sa dva stanja koja predstavljaju kutije A i B. Promena između stanja u dijagramu se može realizovati na različite načine. Sinhroni prelaz je kada se odluka o prelazu dešava jednom za neki zadati vremenski interval. Asinhroni prelaz je

kada se vreme prelaza posebno računa za svakog agenta u odnosu na neke parametre.

U modelima diskretnih događaja već se nalaze pojedinačni entiteti što olakšava prelaz na model sa agentima, jer naravno ti objekti postaju agenti. Problem je u činjenici da su objekti u diskretnim događajima pasivni. Sva pravila po kojima se model ponaša nalaze se u blokovima dijagrama pomoću kojih je model opisan. Dakle, cilj je opisati problem sa tačke gledišta objekta i decentralizovati što veći broj pravila, odnosno prebaciti skup pravila iz blokova dijagrama u dijagram stanja agenta. Naravno, ovo sve ima smisla ako se želi dobiti model sa više individualnog ponašanja objekata u modelu.



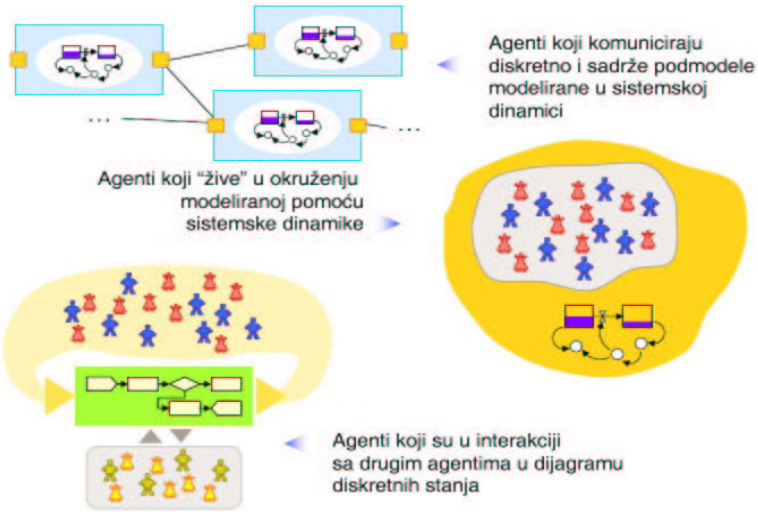
Slika 1.11: Prevođenje modela diskretnih događaja u model sa agentima.

Posmatra se jednostavan sistem opsluživanja gde entiteti ulaze u sistem dobijaju određenu uslugu jednom ili više puta i pri tome provedu u tom stanju određeno vreme i zatim napuštaju sistem. Entiteti postaju agenti. Ulazak u sistem može se izjednačiti sa kreiranjem agenta. Nakon kreiranja, agent šalje zahtev da bude uslužen, ali nekada mora da čeka da se resursi oslobode odnosno da dođe na red što može da traje unapred nepoznato vreme.

Kada se resursi oslobode, model prelazi u stanje primanja usluge. Kada završi odlučuje da li će napustiti sistem ili će zatražiti uslugu ponovo odnosno opet preći u stanje čekanja da se oslobode resursi. Dijagram stanja za resurs ima dva stanja: zauzet i slobodan. Komunikacija između ta dva dijagrama stanja može se implementirati u same agente. Oni komuniciraju direktno između sebe i rukuju sa resursom. Drugi način je uvođenje dispečera koji predstavlja deo okruženja u kojim se agenti nalaze, pri čemu dispečer indirektno implementira komunikaciju između agenata i upravlja resursima.

Korišćenje modeliranja sa agenatima omogućava dobijanje modela koji su bliži realnom sistemu, nego kada se koristi modeliranje pomoću diskretnih događaja ili sistemska dinamika. To nikako ne znači da treba potpuno izbaciti te pristupe modeliranju. U nekim slučajevima pomoću njih se dobijaju dovoljno pouzdani rezultati, pa nema potrebe za metodom sa agentima. Na žalost, u nekim slučajevima je jako teško opisati

ponašanje agenata, pa se dobijaju agentni modeli koji su neefikasni.



Slika 1.12: Primeri kombinovanja pristupa u kreiranju složenih modela.

Realna potreba za preciznim modelima zadovoljena je kombinovanjem tehnika u procesu izrade modela. Programi koji to omogućavaju komercijalno su dostupni i treba ih iskoristi na najbolji način. Primeri za hibridne model su brojni i većina realnih problema zahteva kombinovanje tehnika. Ako se posmatra mreža snabdevanja, procesi unutar kompanije se modeliraju preko sistemske dinamike, a komunikacija između kompanija se realizuje diskretnim događajima. Na primer, ako se posmatra bolnica, pacijenti se mogu predstaviti kao jedan skup agenata, bolničko osoblje kao drugi skup agenata, dok se komunikacija između dva skupa odvija preko blok dijagrama stanja (diskretni događaji). Još jedan primer je naselje, gde stanovnici koriste kuće, imaju posao i koriste infrastrukturu. U tom se slučaju stanovnici prikazuju pomoću agenata, a naselje pomoću sistemske dinamike.

Svi dostupni alati za izradu modela su grupisani oko određene tehnike modeliranja, ali ih je najviše dostupno za modeliranje diskretnih događaja; verovatno najbolji program tog tipa je *Arena*. Postoji samo nekoliko alata za modeliranje sistemske dinamike. U oblasti dinamičkih sistema dominira *MATLAB Simulink*. Što se tiče agenata, do skora nije postojao komercijalno dostupan softver. Sve što je postojalo su biblioteke za *C++* ili *JAVA* jezik koje su razvijane eksperimentalno u okviru akademskih institucija.

Program *AnyLogic* je softverski paket koji je razvila grupa naučnika i programera koji se ne bave modeliranjem. Zato nije korišćen nijedan specifičan pristup kreiranju modela nego se težilo ka kreiranju alata koji će moći da se koristi za modeliranje složenih realnih problema. Budući da je *AnyLogic* alat koji je objektno-orijentisan, moguće je dosta lako kreiranje pouzdanih modela u vizuelnom okruženju, ali je ostavljena mogućnost korišćenja *JAVA* jezika za definisanje i implementaciju specifičnih struktura.

Slika 1.13: Alati za simulaciono modeliranje

1.2.7 Simulacioni jezici

Postoji mnogo programskih jezika koji su napravljeni isključivo za potrebe simulacije. Izbor simulacionog jezika najviše zavisi od prirode problema koji želimo da simuliramo

Najznačajnija prednost simulacionih jezika je to što analitičaru omogućuju da se efikasnije bavi istraživanjima nego kad upotrebljava jezike opšte namene. Većina simulacionih jezika ima mogućnost generisanja slučajnih promenljivih za različite raspodele, upravljanje simulacionim vremenom, rutine za uništavanje simulacionih događaja, upravljanje redovima, skupljanje podataka, sumiranje i analiziranje podataka, formulisanje i štampanje rezultata, detekciju i ispravljanje grešaka, interaktivnu grafičku simulaciju.

Simulacioni jezik mora imati mogućnost vođenja dokumentacije o događajima, procesima i aktivnostima, mora imati mogućnost modifikacije programa, jednostavnu verifikaciju programa i kontrolu dinamičkog odvijanja procesa.

Glava 2

Populaciona dinamika

2.1 Uvod

Potreba za poznavanjem broja stanovnika grada, regije, države ili čitavog sveta, kao i promena broja te populacije je verovatno stara koliko i ljudska civilizacija. U današnjem smislu, modeliranje rasta populacije je staro oko dve stotine godina i vezano je za engleskog sveštenika Maltusa. Njegov model rasta, tzv. Maltusov model, se može smatrati početkom važne oblasti matematičkog modeliranja - populacione dinamike. Danas je to veoma razvijena i kompleksna nauka, koja traži dobro poznavanje raznih matematičkih disciplina, pre svega diferencijalnih i diferencnih jednačina, zatim verovatnoće i statistike, i, naravno, odgovarajućih oblasti iz biologije, hemije, itd.

U okviru ove knjige, morali smo da izaberemo one oblasti koje obuhvataju glavne delove ove teorije, ali ne traže suviše duboko poznavanje na primer biologije, niti previše komplikovane matematičke metode. U principu, ograničićemo se na determinističke modele, sa izuzetkom jednog jednostavnog stohastičkog modela. Analiziraćemo modele rasta jedne (poglavlje 2.2), odnosno dve populacije (poglavlje 2.3), pri čemu se dobijaju diferencijalne i diferencne jednačine, odnosno odgovarajući sistemi takvih jednačina. Za praćenje ove glave, bar što se matematike tiče, dovoljno je savladati standardni kurs iz običnih diferencijalnih jednačina i poznavanje nekih jednostavnih elemenata linearne algebre (u potpoglavlju 2.2.4) i verovatnoće (u potpoglavlju 2.2.5).

Za potpunije i dublje razumevanje populacionih modula, čitaoc može, na primer, konsultovati i sledeće knjige: [2], [5], [6], [8], [13], [17].

2.2 Populacioni modeli sa jednom vrstom

2.2.1 Uvod

Početkom 19. veka se ljudska populacija, već dovoljno velika, stalno povećavala, pa se moglo postaviti pitanje da li će se dugoročno ljudi moći prehraniti ako se taj trend rasta nastavi. Da bi dao odgovor na ovo pitanje, Maltus je krenuo od nekoliko

pretpostavki koje su dovele do jednog veoma uprošćenog i grubog globalnog modela rasta svetske populacije. Radi jednostavnosti, on je pretpostavio da se broj stanovnika na svetu menja na isti način, ne uzimajući u obzir očevidne razlike između ljudi kao što su podela prema starosti, polu, podneblju, ekonomskom i kulturnom razvoju, itd.

Model koji je uveo, tzv. **Maltusov model**, ima pre svega istorijski značaj, jer je to prvi matematički model rasta jedne populacije u okviru nekog ekosistema. Ta populacija može biti stanovništvo jednog grada, regije, države ili čak čitavog sveta, ali može biti i neka životinjska ili biljna populacija unutar jednog ekosistema.

U sledeća dva potpoglavlja, 2.2.2 i 2.2.3, ćemo izložiti kako se može doći do tog modela, i to kako diskretnog, tako i kontinualnog. U suštini, u oba slučaja polazimo od formulacije problema kakvu je možda dao i sam Maltus. U potpoglavlju 2.2.4 analiziramo Leslijev model rasta sa podelom populacije prema uzrastu, a u potpoglavlju

2.2.2 Diskretni Maltusov model

Kao u svakom procesu modeliranja, moraju se prvo uvesti osnovne veličine, a zatim odrediti glavne činioce koji utiču na posmatranu populaciju. Obeležimo sa $x = x(t)$ broj jedinki te populacije u momentu t. U praksi, nije potrebno znati $x(t)$ za sve t, već samo za vrednosti $m \cdot \Delta t$, $m = 0, 1, 2, ...$, gde je Δt jedinica vremena. (Za ljudsku populaciju bi Δt moglo biti godinu dana, a za globalni model svetske populacije ili bar države i 10 godina.) Što se činilaca tiče, jasno je da ih po pravilu ima toliko da ih je teško i nabrojati; mi ćemo pretpostaviti da se populacija menja samo zbog rađanja i umiranja. Dakle, zanemarićemo, uz ostalo, i individualne razlike među jedinkama, kao i migraciju. Tada je promena populacije u jedinici vremena, tzv. **stopa rasta populacije po jedinici vremena**, data sa

$$\frac{\Delta x}{\Delta t} = \frac{x(t+\Delta t) - x(t)}{\Delta t}. \tag{2.1}$$

Međutim, od ove stope korisnija je **specifična stopa rasta populacije**, u oznaci $R(x,t)$, tj. stopa rasta populacije po jedinici vremena i jedinici populacije, data sa

$$R(x,t) = \frac{1}{x(t)} \frac{x(t+\Delta t) - x(t)}{\Delta t}. \tag{2.2}$$

Ako, kao što je to učinio Maltus, dodatno pretpostavimo da je stopa $R(x,t)$ iz (2.2) konstanta, recimo

$$R(x,t) \equiv R_0, \quad \text{za sve } x \text{ i } t, \tag{2.3}$$

onda dolazimo do **diskretnog Maltusovog modela rasta jedne populacije**

$$x(t+\Delta t) = x(t)(1 + R_0 \Delta t), \tag{2.4}$$

Ako stavimo $t_m = m \cdot \Delta t$ i $x_m = x(t_m)$ za $m = 0, 1, 2, \ldots$, onda se (2.4) svodi na sledeću diferencnu jednačinu sa konstantnim koeficijentima:

$$x_{m+1} = x_m(1 + R_0 \Delta t) \tag{2.5}$$

Ova diferencna jednačina se lako rešava ako znamo populaciju u početnom momentu $t = 0$, tzv. početnu populaciju. Dakle, ako je

$$x(0) = x_0, \tag{2.6}$$

onda je rešenje (2.5) dato sa

$$x_m = x_0(1 + R_0\Delta t)^m, \quad m = 0, 1, 2, \ldots$$

Ako je $R_0 > 0$ (na primer, za ljudsku populaciju je $R_0 \approx 0.02$), to se skiciranjem niza $(x_m)_{m\in\mathbb{N}}$ vidi da se sa povećanjem m veoma brzo povećava i x_m. To je pre dva veka veoma zabrinulo Maltusa, jer je njegov model predviđao da će se ljudska populacija mnogo brže povećavati nego neophodna hrana za tu populaciju. Naravno, on nije mogao predvideti da će u prvoj polovini 19. veka doći do daleko intenzivnijeg privrednog razvoja nego ikada pre toga u istoriji, što je kasnije sa pravom nazvano industrijska revolucija.

Sa gledišta modeliranja, ovo pokazuje da treba biti veoma oprezan kod korišćenja modela na "duge staze", jer se promenom parametara, ili čak pojavom novih, realni sistem i te kako može promeniti i za relativno kraće, a pogotovo za duže vreme. Ustvari, Maltusov model može dobro poslužiti kao dokaz da se u procesu modeliranja jednom dobijeni modeli moraju stalno popravljati i usavršavati, kako bi što vernije opisali posmatrani realni sistem.

Primer 2.1 Oko 1965. godine, ljudska populacija je procenjena na 3.34 milijarde. Ako je R_0 iz (2.3) zaista 0.02, onda se ona duplira za približno 35 godina, što se slaže sa podacima za 2000 godinu (oko 6 milijardi). Međutim, ako bi ovaj trend rasta sveukupne ljudske populacije nastavio, onda bi 2500. godine na svetu bilo oko 200 milijardi ljudi. Ostavljamo čitaocu da sam zaključi o vrednosti Maltusovog modela na "duge staze".

Primer 2.2 Prethodni primer deluje apsurdno. Ali, za biljku *Microtus Arvallis Pall* je eksperimentalno pokazano da se njihov broj mesečno povečava za 40%.

Primer 2.3 Tvorac logističkog modela (videti potpoglavlje 2.2.6) Verhulst je oko 1830. godine pokušao da proceni populacije Belgije i Francuske kroz 100 godina. Zanimljivo je da je veličinu populacije svoje zemlje prilično potcenio, ali je zato dosta dobro pogodio populaciju Francuske. Verhulstova greška u slučaju Belgije po svemu sudeći potiče od danas opšte poznate činjenice da je razvoj te zemlje, pa i povećanje njenog stanovništva, bitno ubrzala eksploatacija njenih afričkih kolonija.

2.2.3 Kontinualni Maltusov model

Ako se radi o ljudskoj ili nekoj životinjskoj populaciji, jasno je da je funkcija $x(t)$ u prethodnim jednačinama prirodan broj za sve $t > 0$, pa je kao takva deo po deo

konstantna funkcija sa mnogobrojnim prekidima prve vrste. Strogo gledano, takva funkcija nema izvod u tim prekidima. Međutim, pošto se brojčano velike populacije po pravilu za kraće vreme relativno ne menjaju previše mnogo (čitalac može uzeti za primer stanovništvo nekog većeg grada ili države), to je razumno apoksimirati funkciju $x = x(t)$ sa (bar) neprekidno diferencijabilnom funkcijom. Onda, umesto (2.2), ima smisla uvesti tzv. **trenutnu stopu rasta populacije** $R(x,t)$ te populacije, datu sa

$$R(x,t) = \frac{1}{x}\frac{dx}{dt}. \tag{2.7}$$

Ako ponovo uzmemo da je $R(x,t) \equiv R_0$, onda umesto (2.4) dobijamo običnu diferencijalnu jednačinu

$$\frac{dx}{dt} = R_0 x \tag{2.8}$$

uz početni uslov (2.6).

Problem (2.8), (2.6) je **kontinualni Maltusov model rasta jedne populacije**. Rešenje tog problema je

$$x(t) = x_0 e^{R_0 t}, \quad t \geq 0. \tag{2.9}$$

Primer 2.4 (Model sa migracijom) Neka je $x_j = x(t_j)$, $j = 0,1,\ldots,$ populacija u momentu $t_j = j\tau$, $\tau > 0$, a μ_j, $j = 0,1,\ldots,$ migracija u vremenskom intervalu (t_j, t_{j+1}), za diskretan model. Neka su, dalje, $x = x(t)$ i $\mu = \mu(t)$ $(t > 0)$ populacija i migracija za kontinualan model rasta populacije.

Pokazati da su odgovarajući modeli dati sa:

$$x_{j+1} - x_j = (R_0 \cdot x_j + \cdot\mu_j)\tau, \quad \text{odnosno} \quad \frac{dx}{dt} = R_0 \cdot x(t) + \mu(t)$$

i rešiti ih.

2.2.4 Leslijev model sa podelom populacije prema starosti

Engleski biolog Lesli (*Leslie*), jedan od pionira populacione dinamike, je oko 1940. godine korišćenjem matrica konstruisao model (koji danas nosi njegovo ime) da bi opisao populacionu dinamiku ženskih jedinki neke populacije.

Za većinu vrsta, broj ženskih jedinki je jednak broju muških jedinki pa ćemo i mi to pretpostaviti u nastavku. Model može da se primeni na ljudsku populaciju, populaciju insekata, riba, kao i celokupnu životinjsku populaciju.

Mi živimo u nelinearnom svetu, a kako je ovaj model primer linearnog diskretnog dinamičkog sistema, možemo očekivati da će davati neprecizne, ili čak netačne rezultate, ako ga primenimo na neku populaciju tokom dužeg vremenskog perioda. Međutim, Leslijev model koji ćemo izložiti u ovom poglavlju, daje neke veoma interesantne i dobre rezultate kada se primenjuje na populacije tokom kraćeg vremenskog perioda. Kao što je uobičajeno kod ovakvih, relativno jednostavnih modela, ignorisaćemo bolesti, uticaj životne sredine, npr. zagađenje, kao i sezonske uticaje.

Pretpostavimo da su ženske jedinke podeljene u n starosnih klasa, tako da ako je N teoretski maksimum kada su u pitanju godine starosti ženskih jedinki neke vrste,[1] onda svaka starosna klasa sadrži periode od N/n podjednakih vremenskih intervala, dana, meseci ili godina. Populaciju posmatramo u regularnim diskretnim vremenskim intervalima od kojih je svaki jednak dužini jedne starosne klase. Tako je k-ti vremenski period dat sa $t_k = kN/n$.

Definišimo $x_i^{(k)}$ kao broj ženskih jedinki i-te starosne klase nakon k-tog vremenskog perioda. Neka je b_i broj novorođenih ženskih jedinki od strane neke ženske jedinke u toku i-te starosne klase, a c_i broj ženskih jedinki koje su nastavile da žive i u $(i+1)$-oj starosnoj klasi. Da bi model bio postavljen na realnim osnovama, treba da su ispunjeni sledeći uslovi:

$$b_i \geq 0,\ \ i = 1,2,\ldots,n \ \ \text{i} \ \ 0 < c_i \leq 1,\ \ i = 1,2,\ldots,n-1. \tag{2.10}$$

Jasno je da neki b_i moraju biti pozitivni da bi osigurali da su se neka rođenja desila, a da $c_i > 0$ za sve i, inače ne bi bilo ženskih jedinki u $(i+1)$-oj klasi.

Napisaćemo sada sistem linearnih jednačina, čije su nepoznate brojevi $x_i^{(k)}$ (broj ženskih jedinki u i-toj starosnoj klasi u vremenu t_k). U tom cilju, primetimo da je u prvoj starosnoj klasi posle k vremenskih perioda broj $x_1^{(k)}$ jednak broju novorođenih ženskih jedinki u svih n starosnih klasa u vremenskom intervalu (t_{k-1}, t_k), tako da je

$$x_1^{(k)} = b_1 x_1^{(k-1)} + b_2 x_2^{(k-1)} + \ldots + b_n x_n^{(k-1)}.$$

Broj ženskih jedinki u $(i+1)$-oj starosnoj klasi u vremenu t_k je jednak broju ženskih jedinki u i-toj starosnoj klasi u vremenu t_{k-1} koja nastavlja da živi i do $(i+1)$-ve starosne klase, tj.

$$x_{i+1}^{(k)} = c_i x_i^{(k-1)}.$$

Ovako dobijene jednačine možemo zapisati u sledećoj matričnoj formi:

$$\begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \\ \cdot \\ \cdot \\ \cdot \\ x_n^{(k)} \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & b_3 & \cdot & \cdot & \cdot & b_{n-1} & b_n \\ c_1 & 0 & 0 & \cdot & \cdot & \cdot & 0 & c_n \\ 0 & c_2 & 0 & \cdot & \cdot & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & c_{n-1} & 0 \end{bmatrix} \begin{bmatrix} x_1^{(k-1)} \\ x_2^{(k-1)} \\ x_3^{(k-1)} \\ \cdot \\ \cdot \\ \cdot \\ x_n^{(k-1)} \end{bmatrix} \tag{2.11}$$

ili, u skraćenom obliku, kao

$$X^{(k)} = LX^{(k-1)}. \tag{2.12}$$

Matrice kolone u (2.11) su dimenzije $n \times 1$, a kvadratna matrica L, koja se obično zove **Leslijeva matrica**, je dimenzije $n \times n$.

[1] Na primer, u slučaju ljudske populacije mogli bi uzeti $N = 100$.

Pretpostavimo da je $X(0)$ vektor koji predstavlja početni broj ženskih jedinki u svakoj od n starosnih klasa. Tada je:

$$X^{(1)} = LX^{(0)}, \quad X^{(2)} = LX^{(1)} = L^2X^{(0)}, \ldots, \quad X^{(k)} = L^kX^{(0)}, \quad k = 1,2,\ldots \tag{2.13}$$

Na osnovu toga, ako znamo početnu starosnu distribuciju i Leslijevu matricu L iz (2.12), moguće je utvrditi starosnu distribuciju ženskih jedinki u bilo kom kasnijem vremenskom intervalu.

Pre nego što nastavimo prethodnu analizu, podsetimo se da su broj λ i vektor kolona $\vec{v}$ respektivno **sopstvena vrednost** i odgovarajući **sopstveni vektor** neke matrice M ako je

$$M\vec{v} = \lambda \cdot \vec{v}.$$

Da bismo istražili ponašanje dobijenog sistema, pokazuje se da je neophodno razmotriti sopstvene vrednosti i sopstvene vektore matrice L. Te veličine mogu da se iskoriste da bi se odredila populaciona distribucija u budućnosti uzimajući u obzir starosne klase, o čemu govori sledeća važna teorema.

Teorema 2.5 *Neka je Leslijeva matrica L definisana kao u (2.12), i pretpostavimo da pored uslova (2.10) važi da su najmanje dva sukcesivna koeficijenta b_i strikno pozitivna. Tada*

1. *matrica L ima jedinstvenu sopstvenu vrednost λ_1, koja je ili pozitivna ili ima algebarsku višestrukost jednaku 1;*

2. *sopstveni vektor $\vec{v}_1$, koji odgovara sopstvenoj vrednosti λ_1 ima pozitivne komponente;*

3. *za svaku drugu sopstvenu vrednost λ_i matrice L, koja je različita od λ_1, važi da je $|\lambda_i| < \lambda_1$.*

Pozitivna sopstvena vrednost λ_1 iz tačke 3 se zove **striktno dominantna**. Tada je

$$LX = \lambda_1 \cdot X,$$

gde je $X = \vec{v}_1 \neq \vec{0}$ sopstveni vektor iz tačke 2 gornje teoreme. Ako sada radi jednostavnosti uzmemo da je $x_1^0 = 1$, onda je

$$\vec{v}_1 = \left[\begin{array}{ccccc} 1 & c_1/\lambda_1 & c_1c_2/\lambda_1^2 & \ldots & c_1c_2\cdots c_{n-1}/\lambda_1^{n-1} \end{array} \right]^* \tag{2.14}$$

gde znak $*$ na kraju poslednje jednakosti označava adjungovanu matricu (tj. $\vec{v}_1$ je matrica kolona).

Pretpostavimo još da L ima n linearno nezavisnih sopstvenih vektora, $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$, kojima redom odgovaraju proste sopstvene vrednosti $\lambda_1, \lambda_2, \ldots, \lambda_n$. Ako je početna raspodela populacije data sa vektorom $X^{(0)} = X_0$, onda postoje konstante $b_1, b_2, \ldots, b_n$ takve da je

$$X^{(0)} = X_0 = b_1\vec{v}_1 + b_2\vec{v}_2 + \ldots + b_n\vec{v}_n.$$

Na osnovu jednačine (2.13) je onda

$$\begin{aligned} X^{(k)} &= L^k(b_1\vec{v_1} + b_2\vec{v_2} + \ldots + b_n\vec{v_n}) \\ &= b_1\lambda_1^k\vec{v_1} + b_2\lambda_2^k\vec{v_2} + \ldots + b_n\lambda_n^k\vec{v_n} \\ &= \lambda_1^k\left(b_1\vec{v_1} + b_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\vec{v_2} + \ldots + b_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\vec{v_n}\right). \end{aligned}$$

Kako je po pretpostavci λ_1 dominantna sopstvena vrednost, to je

$$\left|\frac{\lambda_i}{\lambda_1}\right| < 1 \text{ za } \; i = 2,3,\ldots,n,$$

što povlači

$$\lim_{k\to\infty}\left(\frac{\lambda_i}{\lambda_1}\right)^k = 0, \quad i = 2,\ldots,n.$$

Odavde sledi da je za veliko k

$$X^{(k)} \approx b_1\lambda_1^k\vec{v_1}.$$

Ova približna jednakost znači da posle dužeg vremena možemo očekivati stabilizaciju starosne distribucije, koja vremenom postane proporcionalna vektoru $\vec{v_1}$ iz (2.14).

Obično se vektor $\vec{v_1}$ normalizuje tako da je zbir njegovih komponenata jednak 1. Tada elementi tako normalizovanog vektora $\vec{v_1}$ daju prognozirane procente ženskih jedinki u svakoj od n starosnih grupa (naravno, pre toga pomnožene sa 100). Primetimo da ako je $\lambda_1 = 1$, onda se populacija stabilizuje, a ako je $\lambda_1 < 1$ (resp. $\lambda_1 > 1$), onda se vremenom populacija smanjuje (resp. povećava) sa faktorom λ_1.

2.2.5 Jedan probabilistički model

Neka je $P_N(t)$ verovatnoća da u momentu $t > 0$ ima tačno N jedinki jedne populacije. Pretpostavićemo da u modelu nema umiranja, a da za kratko vreme Δt može da se rodi najviše jedna jedinka, i da je verovatnoća da će za vreme Δt jedna jedinka roditi novu jednaka $\lambda\Delta t$ gde je $\lambda > 0$.

Ispitaćemo ovaj probabilistički model tako što ćemo odrediti $P_N(t)$ za $N = 1,2,\ldots$, i pokazati da je očekivana vrednost populacije $E(t)$ u momentu t zapravo veličina $x(t)$ iz rešenja Maltusovog modela datog sa (2.9), uz zamenu broja λ sa R_0.

Pretpostavićemo još da je početna populacija jednaka N_0, a budući da u ovom modelu nema umiranja, to znači da je

$$P_N(0) = \begin{cases} 1, & N = N_0; \\ 0, & N > N_0. \end{cases} \tag{2.15}$$

Sada ćemo odrediti običnu diferencijalnu jednačinu koju zadovoljava funkcija $P_N(t)$. U tom cilju, primetimo da pod navedenim pretpostavkama posmatrana populacija može dostići veličinu N u momentu $t+\Delta t$ na dva načina: ili je u momentu t već bila jednaka N, što znači da tokom vremenskog intervala $[t,t+\Delta t]$ nije rođena nova jedinka, ili je u momentu t bila jednaka $N-1$, pa je tokom intervala $[t,t+\Delta t]$ jedna od tih $N-1$ jedinki rodila još jednu, i to sa verovatnoćom $\lambda\Delta t$. Kako su ta dva događaja disjunktna, to je

$$P_N(t+\Delta t) = (1-\lambda\Delta t)^N P_N(t) + \binom{N-1}{1}\lambda\Delta t(1-\lambda\Delta t)^{N-2}P_{N-1}(t). \tag{2.16}$$

Izraz $(1-\lambda\Delta t)^N$ (uz $P_N(t)$) se može razviti po binomnoj formuli kao

$$(1-\lambda\Delta t)^N = 1 - N\lambda\Delta t + \binom{N}{2}\lambda^2\Delta^2 t + r_1(\Delta t). \tag{2.17}$$

Lako je videti da ostatak $r_1(\Delta t)$ ima osobinu da je $\lim\limits_{\Delta t\to 0}\dfrac{r_1(\Delta t)}{\Delta t} = 0$.

Izraz $\binom{N-1}{1}\lambda\Delta t(1-\lambda\Delta t)^{N-2}$ (uz $P_{N-1}(t)$) se na sličan način može napisati kao

$$\binom{N-1}{1}\lambda\Delta t(1-\lambda\Delta t)^{N-2} = (N-1)\lambda\Delta t + r_2(\Delta t)\Delta t \tag{2.18}$$

gde za ostatak $r_2(\Delta t)$ važi $\lim\limits_{\Delta t\to 0} r_2(\Delta t) = 0$.

Ako sada izraze (2.17) i (2.18) zamenimo u jednačinu (2.16), prebacimo $P_N(t)$ na levu stranu i dobijenu jednačinu podelimo sa Δt, to dolazimo do jednačine

$$\frac{P_N(t+\Delta t)-P_N(t)}{\Delta t} = \Big(-\lambda N P_N(t) + \frac{r_1(\Delta t)}{\Delta t}\Big) + \Big(\lambda(N-1)P_{N-1}(t) + r_2(\Delta t)\Big). \tag{2.19}$$

Prelaskom na graničnu vrednost kad $\Delta t \to 0$, dobijamo diferencijalnu jednačinu

$$\frac{dP_N}{dt} = \lambda(N-1)P_{N-1}(t) - \lambda N P_N(t), \quad N \in \mathbb{N}. \tag{2.20}$$

Budući da u modelu nema umiranja, to je verovatnoća $P_{N-1}(t)$ jednaka 0 za sve $t > 0$. Sada jednačina (2.20) za $N = N_0$, dobija oblik

$$\frac{dP_{N_0}}{dt} = -\lambda N_0 P_{N_0}(t)$$

Korišćenjem početnog uslova iz (2.15) za $N = N_0$, dobijamo verovatnoću $P_{N_0}(t)$:

$$P_{N_0}(t) = e^{-\lambda N_0 t}, \quad t \geq 0.$$

Sa nešto više truda, korišćenjem matematičku indukcije, se pokazuje da je

$$P_{N_k}(t) = \frac{N_0(N_0+1)\cdots(N_0+k-1)}{k!}e^{-\lambda N_0 t}(1-e^{-\lambda t})^k, \quad t \geq 0, \quad k = 1,2,\ldots \tag{2.21}$$

Iz rešenja (2.21) sledi da važi

$$\frac{dE}{dt} = \lambda E(t), \tag{2.22}$$

gde je $E = E(t)$ očekivana vrednost populacije u momentu t. Iz (2.22), uz korišćenje uslova (2.15) za $N = N_0$, se dobija da je

$$E(t) = N_0 e^{\lambda t}, \quad t > 0.$$

Dodajmo da se jednačina (2.22) može izvesti i bez poznavanja rešenja (2.21), uz korišćenje obične diferencijalne jednačine (2.20).

Konačno, važno je primetiti da se dobijena očekivana vrednost poklapa sa rešenjem jednačine (2.9) uz početni uslov $x(0) = x_0$ (deterministički Maltusov model), jasno uz zamenu $E(t)$, λ i N_0 sa, respektivno, $x(t)$, R_0 i x_0.

2.2.6 Logistički (Verhulstov) model

Maltusov model rasta populacije, kako diskretni, tako i kontinualni, predviđa eksponencijalni rast populacije, pa čak i za male pozitivne vrednosti konstante R_0 i za relativno kratko vreme dovode do nerealnih, pa možemo reći i apsurdnih rezultata, kako smo mogli videti iz prethodno datih primera. Glavni nedostatak Maltusovog modela je pretpostavka (2.3), tj. da je specifična stopa rasta populacije konstantna. Zbog toga je već tridesetih godina 19. veka belgijski naučnik Verhulst predložio model rasta jedne populacije, danas poznat kao **logistički** ili **Verhulstov model**. Zanimljivo je da su isti model "otkrili" i američki naučnici Perl i Rid skoro vek kasnije.

Umesto (2.3), Verhulst je predložio zavisnost specifične stope rasta $R(x,t)$ iz (2.7) od same populacije x (ali ne i od vremena t), odnosno da je

$$R = R(x).$$

pri čemu je funkcija R opadajuća po x. Naime, mnogi primeri iz životinjskog sveta, ali i iz ljudske istorije, sugerisali su da sa rastom populacije dolazi do usporavanja njenog rasta. Najjednostavnija opadajuća funkcija koju možemo iskoristiti za ovaj model je linearna funkcija oblika

$$R(x) = a - bx, \tag{2.23}$$

gde su a i b pozitivne konstante. Tako dolazimo do početnog problema

$$\begin{aligned} \frac{dx}{dt} &= x(a - bx) \\ x(0) &= x_0 \end{aligned} \tag{2.24}$$

koji ćemo zvati **kontinualni logistički modeli rasta populacije**. Za ljudsku populaciju je nađeno da je

$$a = 0,029 \quad \text{i} \quad b = 2,695 \cdot 10^{-12}, \tag{2.25}$$

tj. a je mnogo veće od b, što pišemo $a >> b$.

Reč "logistika" u nazivu modela (2.24) potiče od sledećeg razmatranja. Ako pretpostavimo da rast populacije zavisi od hrane i obeležimo sa H_d i H_p dostupnu odnosno potrebnu količinu hrane u okviru posmatranog ekosistema za populaciju veličine x, onda možemo uzeti da je

$$R(x) = \alpha(H_d - H_p) \tag{2.26}$$

gde je α pozitivna konstanta. Ako sada pretpostavimo da je H_d konstanta, a da je H_p proporcionalna sa populacijom x, tj. $H_p = \beta x$, gde je β pozitivna konstanta, onda iz (2.26) sledi

$$\frac{1}{x}\frac{dx}{dt} = \alpha(H_d - \beta x), \tag{2.27}$$

što se za $a = \alpha H_d$ i $b = \alpha\beta$ poklapa sa diferencijalnom jednačinom iz (2.24). Ostavljamo čitaocu da proveri da je rešenje početnog problema (2.24) dato sa

$$x(t) = \frac{a/b}{1 + \dfrac{a - bx_0}{bx_0}e^{-at}}. \tag{2.28}$$

Grafik ove funkcije, **logistička kriva**, je data na koricama skripti, kao i na slici **??** u 5. glavi (plava kriva).

Uzećemo da je $0 < x_0 < a/b$. Tada je specifična stopa rasta populacije $R = R(x) = a - bx$ pozitivna, pa iz (2.24) sledi da populacija $x = x(t)$ raste sa vremenom t. Međutim, za razliku od Maltusovog, u posmatranom logističkom modelu je $x_0 < x(t) < a/b$, tj. za sve t je populacija od gore ograničena sa veličinom a/b, što neposredno sledi iz (2.28). Budući da je

$$\lim_{t\to\infty} x(t) = \frac{a}{b}, \tag{2.29}$$

to je a/b najmanje gornje ograničenje populacije ("supremum"). Geometrijski, (2.29) znači da grafik funkcije $x = x(t)$ ima horizontalnu asimptotu kada $t \to \infty$.

Slučaj $x_0 > a/b$ se u praksi mnogo ređe dešava. U stvari, tada ovaj model predviđa opadanje populacije, jer je tada njena specifična stopa rasta $R = R(x) = a - bx$ negativna. Kod ljudske populacije, to je po pravilu posledica vanrednih situacija, kao što su ratovi, ekonomske krize, velike epidemije, posebno loši klimatski uslovi i slično, koji dovode do velikog manjka hrane i ostalih osnovnih potreba. Naravno, za takve posebne slučajeve neophodni su drugi, bitno drugačiji modeli, koje ovde nećemo razmatrati.

Primer 2.6 (Diskretni logistički model)
U ovom potpoglavlju ćemo konstruisati diskretni logistički model. U tom cilju, neka je x veličina populacije, a Δt tzv. vreme diskretizacije. To znači da je naš zadatak da postavimo i analiziramo diferencnu jednačinu, u kojoj su nepoznati članovi niza $x_n = x(n \cdot \Delta t)$ za $n = 0, 1, 2, \ldots$ U nastavku ćemo smatrati da je početna vrednost $x(0) = x_0$ poznata.

Ako sada izvod u diferencijalnoj jednačini (2.24) zamenimo sa količnikom

$$\frac{x(t+\Delta t)-x(t)}{\Delta t},$$

a vreme t sa $n\cdot\Delta t$, onda dolazimo do nelinearne diferencne jednačine prvog reda

$$x_{n+1}-x_n(t)=\Delta t\cdot x_n(a-bx_n),\quad n=0,1,2,..., \tag{2.30}$$

koju možemo uzeti za **diskretni logistički model**.

Ispitaćemo ponašanje x_n iz (2.30) za velike n. Ako stavimo

$$x_n=\frac{a}{b}+\varepsilon y_n,\quad n=0,1,2,..., \tag{2.31}$$

gde je $\varepsilon>0$ mali parametar, a y_n, $n=0,1,2,...$, veličine koja daje odstupanje x_n od a/b do na faktor ε, to posle sređivanja dobijamo diferencnu jednačinu

$$(y_{n+1}-y_n)=-\Delta t\cdot(ay_n+b\varepsilon y_n^2),\quad n=0,1,2,..., \tag{2.32}$$

Ako još u (2.32) zanemarimo član sa ε, onda dobijamo linearnu diferencnu jednačinu prvog reda

$$y_{n+1}=y_n(1-a\Delta t),\quad n=0,1,2,..., \tag{2.33}$$

čije je rešenje

$$y_n=y_0(1-a\Delta t)^n,\quad n=0,1,2,...,$$

Ako prepostavimo da je $a\Delta t<1$, onda je $\lim\limits_{n\to\infty}y_n=0$, što povlači da je

$$\lim_{n\to\infty}x_n=\frac{a}{b}$$

(uporediti sa (2.29)).

Dodajmo da su modeli ovog tipa posebno pogodni za biljne ili životinjske vrste koje se razmnožavaju samo u određenim vremenskim intervalima, npr. jednom godišnje.

Primer 2.7 (Smitov model rasta populacije) Videli smo u potpoglavlju 2.2.6 da se logistički model rasta populacije (2.24) može dobiti na osnovu relacije (2.26). Naime, tu je uvedena pretpostavka da je specifična stopa rasta proporcionalna sa razlikom između dostupne hrane H_d i potrebne hrane H_p, pri čemu je još uzeto da je H_p proporcionalna sa veličinom populacije x.

U ovom modelu, međutim, Smit (*Smith*) je pretpostavio da je veća količine hrane potrebna u razvojnom periodu populacije. Dakle, umesto jednakosti $H_p=\beta x$, u Smitovom modelu se uzima da je

$$H_p=\beta x+\gamma\frac{dx}{dt},$$

pa umesto logističke jednačine dobijamo

$$\frac{dx}{dt}=\frac{\alpha x(H_d-\beta x)}{1+\alpha\gamma x}.$$

Ako stavimo $a = \alpha H_d$, $b = \alpha\beta$ i $c = \alpha\gamma$, dobijamo sledeću običnu diferencijalnu jednačinu, koja karakteriše Smitov model:

$$\frac{dx}{dt} = \frac{x(a-bx)}{1+cx}$$

Ostavljamo čitaocu da proveri da se ponašanje Smitovog modela ne razlikuje bitno od logističkog; posebno, ravnotežni položaji su isti kao kod logistčkog modela. Ipak, može se konstatovati da Smitov model nešto precizniji od logistčkog ako se radi o "mladoj" populaciji.

Primer 2.8 Iz konstanti u (2.25) sledi da je veličina a/b približno jednaka $10,76$ milijardi. Dakle, ovaj logististički model predviđa da je **maksimalni kapacitet** ljudske populacije nešto ispod 11 miljardi stanovnika.

Primer 2.9 U jednom svom eksperimentu, nemački biolog G. F. Gause je stavio pet jedinki *Paramecium Caudatum*-a u epruvetu u odgovarajućoj sredini, i kasnije merenjem utvrdio da je stopa rasta ove biljke 231% na dan! U stvari, broj jedinki x te populacije se ponašao u skladu sa logističkom jednačinom (2.24), uz $a = 2.309$ i $b = 6.157\,3 \times 10^{-3}$, tj.

$$x(t) = \frac{375}{1+74e^{-2.309t}}.$$

Posle četiri dana došlo je do zasićenja, jer je *Paramecium Caudatum* stigao do svog maksimalnog kapaciteta u spomenutim uslovima.

Primer 2.10 Posmatrajmo sledeći početni problem:

$$\frac{dx}{dt} = ax + bx^2, \quad x(0) = x_0,$$

gde su konstante a i b pozitivne. Koristeći faznu ravan, lako je videti da rešenje ovog problema postaje beskonačno u *konačnom* momentu t_e, koje se zove i vreme eksplozije (odredite taj momenat!).

Iako se jednačina u ovom primeru razlikuje samo u jednom znaku (ispred x^2), od logističkog, vidimo da ovaj problem *ne može* biti matematički model rasta neke populacije.

2.2.7 Modeli sa kašnjenjem

Videli smo u potpoglavlju 2.2.6 da logistički model (2.24) može dosta dobro da opiše rast jedne populacije, pri čemu se iz rešenja u (2.28) vidi da populacija ne može nadmašiti tzv. kapacitet populacije a/b. Međutim, u slučaju "mlade" populacije koja se relativno brzo razvija, nije retkost da spomenuti kapacitet bude i premašen. Ova pojava pokazuje, sa jedne strane, očit nedostatak logističkog modela, ali, sa druge strane, nagoveštava da je veličina neke današnje populacije zapravo posledica događaja

u prošlosti. Na osnovu prethodno dobijenih populacionih modela, to se matematički može modelirati, na primer, kao neke od sledećeih **jednačina sa kašnjenjem**:

$$\frac{dx}{dt} = R_0 x(t-\tau), \tag{2.34}$$

$$\frac{dx}{dt} = x(t)(a-bx(t-\tau)), \tag{2.35}$$

$$\frac{dx}{dt} = x(t-\tau_1)(a-bx(t-\tau_2)), \tag{2.36}$$

gde su τ, τ_1 i $\tau_2 > 0$ kašnjenja, R_0 konstanta iz Maltusovog modela (2.8), dok su a i b konstante iz logističkog modela (2.24).

U jednačinama (2.34)-(2.36) se pojavljuju tzv. diferencijalno-diferentne jednačine, čije bi nas rešavanje i analiza suviše udaljili od teme ove knige. Zbog toga ćemo se baviti diskretnim modelom koji odgovora jednačini (2.35).

Za početak, dodatno ćemo pretpostaviti da se posmatrana populacija razmnožava u jednakim vremenskim intervalima dužine Δt (recimo godinu dana), pa je razumno uzeti da je dužina τ (kašnjenje) jednako priraštaju vremena Δt, tj. $\tau = \Delta t$. Dalje, analogno kao u primeru 2.6, stavićemo $t = n \cdot \Delta t$, $n = 0,1,2,\ldots$ i

$$x_n = x(t) = x(n \cdot \Delta t).$$

Posle zamene izvoda u jednačini (2.35) sa

$$\frac{x(t+\Delta t)-x(t)}{\Delta t},$$

dobijamo nelinearnu diferencnu jednačinu drugog reda

$$x_{n+1} - x_n = x_n(a - bx_{n-1}), \quad n = 1,2,\ldots \tag{2.37}$$

Kao i u primeru 2.6, i u ovom ćemo zameniti x_n kao u (2.31). Zamenom u (2.37), sređivanjem i, konačno, zanemarivanjem članova koji sadrže ε, dolazimo do sledeće linearne diferencne jednačine drugog reda

$$y_{n+2} - y_{n+1} + a\Delta t \cdot y_n = 0, \quad n = 0,1,2,\ldots \tag{2.38}$$

Rešenje jednačine (2.38) je, kako se čitalac lako može uveriti, oblika $y_n = k^n$, gde je k rešenje **karakteristične jednačine**

$$k^2 - k + a\Delta t = 0.$$

tj.

$$k_{1,2} = \frac{1 \pm \sqrt{1-4a\Delta t}}{2}. \tag{2.39}$$

Ako je podkorena veličina $1-4a\Delta t$ različita od nule, onda sledi da je opšte rešenje diferencne jednačine (2.38) oblika

$$y_n = C_1 k_1^n + C_2 k_2^n, \quad n = 0,1,2,\ldots \tag{2.40}$$

gde su C_1 i C_2 proizvoljne konstante.[2]

U zavisnosti od znaka podkorene veličine $1-4a\Delta t \neq 0$, imamo dva bitna slučaja (slučaj $a\Delta t = 1/4$ za nas nije od interesa):

1. Ako je $a\Delta t < 1/4$ (najvažniji slučaj), onda su rešenja (2.39) realna i različita, pa ako su poznate početne vrednosti y_0 i y_1, onda iz (2.40) sledi da je traženo rešenje difrencne jednačine (2.38) dato sa

$$y_n = \frac{k_2 y_0 - y_1}{k_2 - k_1} k_1^n + \frac{y_1 - k_1 y_0}{k_2 - k_1} k_2^n \tag{2.41}$$

   Iz pretpostavke $a\Delta t < 1/4$ sledi $\sqrt{a\Delta t} < 1/2$, što povlači sledeće nejednakosti:

$$k_1 = 1/2 + \sqrt{a\Delta t}/2 > 1/2 \quad \text{i} \quad k_1 < 1/2 + 1/2 = 1,$$

$$k_2 = 1/2 - \sqrt{1-4a\Delta t}/2 > 0 \quad \text{i} \quad k_2 < k_1.$$

   Tako dobijamo produženu nejednakost

$$0 < k_2 < k_1 < 1,$$

   koja sa (2.41) povlači $\lim\limits_{n\to\infty} y_n = 0$, ili

$$\lim_{n\to\infty} x_n = \frac{a}{b}. \tag{2.42}$$

2. Ako je $a\Delta t > 1/4$, onda su rešenja iz (2.39) konjugovano-kompleksna:

$$k_{1,2} = \frac{1}{2} \pm \frac{i}{2}\sqrt{4a\Delta t - 1}.$$

   Brojevi k_1 i k_2 se mogu napisati u eksponencijalnom obliku kao $k_1 = \rho e^{i\phi}$, odnosno $k_2 = \rho e^{-i\phi}$, gde je moduo ρ jednak

$$\rho = |k_1| = |k_2| = \sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{\sqrt{4a\Delta t - 1}}{2}\right)^2} = \sqrt{a\Delta t}, \tag{2.43}$$

   a argument ϕ jednak

$$\phi = \operatorname{arctg}(\sqrt{4a\Delta t - 1}). \tag{2.44}$$

[2] Ako je čitalac imao prilike da se upozna sa rešavanjem i posebno oblikom rešenja običnih linearnih diferencijalnih jednačina sa konstantnim koeficijentima, nesumnjivo će zapaziti njihovu sličnost sa gornjim postupkom i oblikom rešenja iz (2.40).

Na osnovu de Moavrove formule

$$[R(\cos\Phi + i\sin\Phi)]^n = R^n[\cos(n\Phi) + i\sin(n\Phi)], \quad n = 0,1,2,\ldots\ (R > 0,\ \Phi \in \mathbb{R})$$

i rešenja (2.40), sledi da je opšte rešenje diferencne jednačine (2.38) oblika

$$y_n = \rho^n(C_1\sin(n\phi) + C_2\cos(n\phi)), \quad n = 0,1,2,\ldots \tag{2.45}$$

gde su ρ i ϕ dati sa (2.43) i (2.44) respektivno, a C_1 i C_2 proizvoljne konstante. Ako su y_0 i y_1 poznate veličine, onda je traženo rešenje jednačine (2.38) dato sa

$$y_n = \rho^n\left(\frac{y_1 - y_0\cos\phi}{\sin\phi}\sin n\phi + y_0\cos n\phi\right)$$

U zavisnosti od veličine modula $\rho = \sqrt{a\Delta t}$, imamo da su rešenja (2.45), a time i početne diferencne jednačine (2.37), stabilna ako je $\rho < 1$, odnosno nestabilna ako je $\rho > 1$. U zadnjem slučaju možemo reći da je kašnjenje dovelo do nestabilnosti rešenja.

2.2.8 Ravnotežne populacije

U prethodnom poglavlju smo videli da je veličina a/b veoma značajna za logistički model (2.24). Sa druge strane, iz te obične diferencijalne jednačine dobijamo da je a/b upravo vrednost za koju je izraz $x(a-bx)$ na desnoj strani jednak nuli; usput, to je tačno i za 0. Očevidno je da populacija $x = x(t)$ može uzeti vrednosti 0 ili a/b jedino ako je početna populacija x_0 jednaka nekoj od te dve vrednosti. Međutim, u oba ta slučaja je onda $x(t)$ konstanta, tj. ili je $x(t) \equiv 0$, ili je $x(t) \equiv a/b$, za sva vremena t, tj. u ta dva (u praksi sasvim nerealna) slučaja ili nema te populacije, ili je ona stalno jednaka kapacitetu te populacije.

Uopšte, nule desne strane u (2.24) su veoma važne za našu dalju analizu.

Ravnotežna populacija je ona populacija za koju je stopa rasta jednaka nuli, tj.

$$\frac{dx}{dt} = 0. \tag{2.46}$$

Prema prethodnom, ako za neko t_r važi $x = x_r$ (gde je x_r ravnotežna populacija), onda je

$$x(t) = x_r, \quad t \geq t_r.$$

Ravnotežna populacija x_r je **stabilna** ako važi

$$\lim_{t\to+\infty} x(t) = x_r. \tag{2.47}$$

U suprotnom, x_r je **nestabilna** ravnotežna populacija. Mi već znamo da logistička jednačina (2.24) ima dve ravnotežne populacije, naime

$$x_{1r} = 0 \quad \text{i} \quad x_{2r} = \frac{a}{b},$$

od kojih je druga, tj. a/b, prema (2.29), stabilna. Ostavljamo čitaocu da proveri da je $x1r = 0$ nestabilna ravnotežna populacija, koristeći metod dat u nastavku.

Analiza stabilnosti ravnotežnih populacija se može uraditi pomoću **metoda perturbacije**, čak i bez poznavanja eksplicitnog rešenja. Pokazaćemo to na primeru ravnotežne tačke $x_r = a/b$ za logističku jednačinu. Ako stavimo

$$x(t) = x_r + \varepsilon x_1(t) = \frac{a}{b} + \varepsilon x_1(t),$$

gde je $x_1(t)$ odstupanje od x_r sa multiplikativnim faktorom $\varepsilon > 0$ (”mali parametar”). Tako zamenom u jednačinu (2.24) dobijamo sledeću dobijamo običnu diferencijalnu jednačinu po $x_1(t)$:

$$\begin{aligned} \varepsilon\frac{dx_1}{dt} &= \left(\frac{a}{b} + \varepsilon x_1(t)\right)\left(a - b\left(\frac{a}{b} + \varepsilon x_1(t)\right)\right) = -\left(\frac{a}{b} + \varepsilon x_1(t)\right) b\varepsilon x_1(t) \\ &= -\varepsilon(a + \varepsilon b x_1(t))x_1(t). \end{aligned}$$

Posle skraćivanja sa ε, pa zanemarivanja veličina koje množi ε, dobijamo običnu diferencijalnu jednačinu

$$\frac{dx_1}{dt} = -ax_1(t)$$

čije je rešenje

$$x_1(t) = Ce^{-at}$$

za neku multiplikativnu konstantu C.

Dakle, odstupanje $x(t)$ od a/b sa protokom vremena ($t \to \infty$) eksponencijalno teži 0, što pokazuje da je a/b stabilna ravnotežna populacija.

2.3 Populacioni modeli sa dve vrste

2.3.1 Uvod

U prethodnom poglavlju smo analizirali rast jedne populacije unutar nekog ekosistema. Dobijeni modeli su u suštini veoma jednostavni, jer smo pri prilikom njihove konstrukcije zanemarili većinu više ili manje značajnih faktora. Implicitna pretpostavka u prethodnim analizama je da rast populacije zavisi ili samo od dostupne hrane (eksponencijalni ili Maltusov model), ili, dodatno, od same populacije (logistički ili Verhulstov model).

U ovom poglavlju ćemo izložiti dva deterministička populaciona modela u kojima se pojavljuje interakcija dve populacije, pri čemu je ona dominantna u odnosu na sve ostale faktore. Kasnije ćemo videti da se takvi i slični modeli mogu, bar sa gledišta običnih diferencijalnih jednačina, na isti način analizirati.

Prvi od takvih modela sa dve vrste je tzv. model ”lovac-žrtva” (ili ”predator-plen”), i razvili su ga zajedno dvadesetih godina prošlog veka veliki italijanski matematičar

V. Voltera (*Vito Volterra*) i biolog A. Lotka (*Alfred Lotka*), pa se otada modeli u kojima se jedna populacija hrani drugom (ili je na neki drugi način ugrožava), često nazivaju modeli tipa **Lotka-Voltera**. Prvobitno, spomenuta dvojica naučnika su analizirala promenu broja određenih vrsta riba u Jadranskom moru i tako došli do "svog" modela, koji je u stvari sistem nelinearnih diferencijalnih jednačina prvog reeda.

U literaturi se obično spominju "ajkule" kao lovci (predatori) i "ribe" kao žrtve (plen), iako su Lotka i Voltera posmatrali dve vrste riba, od kojih se jedna vrsta hranila drugom, a koja se opet hranila planktonom. Na osnovu podataka koji su skupljani skoro čitav vek, pokazalo se da se interakcija između jedne vrste risova i jedne vrste zečeva u Severnoj Americi sa matematičkog gledišta može na skoro isti način opisati. Vremenom su pronađene i druge interakcija ovog tipa, a i uvedene su mnogobrojne generalizacije početnog modela Lotka-Voltera.

Drugi je model **dve populacije u takmičenju**. U tom se modelu radi o dve populacije koje se bore za istu hranu, ili su direktno međusobno neprijateljske. Videćemo da njihov odnos i stabilnost (odn. nestabilnost) takvog sistema bitno zavisi od početnih vrednosti tih populacija, kao i od odnosa odgovarajućih parametara.

2.3.2 Modeli tipa Lotka-Voltera

Posmatrajmo mali ekosistem sa dve populacije, čije ćemo veličine označiti sa x i y, respektivno. Kao u matematičkim modelima sa jednom populacijom, pretpostavićemo da mera promene odgovarajuće populacije zavisi samo od veličina tih populacija. Obeležimo sa $x = x(t)$ i $y = y(t)$ veličine dve populacije u momentu $t \geq 0$, i pretpostavimo da se populacija x isključivo hrani populacijom y, dok se rast populacije y opisuje logističkom jednačinom (2.24). Tada, bez interakcije između te dve populacije (na primer, ako se radi o velikom području sa relativno malo jedinki obe populacije), dolazimo do sledećeg sistema običnih diferencijalnih jednačina:

$$\frac{dx}{dt} = -kx, \tag{2.48}$$

$$\frac{dy}{dt} = y(a - by) \tag{2.49}$$

Konstante a, b i k u ove dve jednačine su, ako nije drugačije pretpostavljeno, pozitivne. U nastavku to ćemo pretpostaviti i za sve ostale konstante koje se pojavljuju u jednačinama.

Primetimo da je u slučaju prve populacije x, tj. populacije predatora (lovaca), specifična stopa rasta jednaka negativnom broju $-k$, što u nedostatku jedinki druge populacije y, tj. populacije žrtvi, brzo dovodi do izumiranja x-eva.

Ako sada pretpostavimo da postoji gore opisana interakcija između ovih populacija, onda je odgovarajući matematički model dat sa sledećim sistemom običnih di-

ferencijalnih jednačina:

$$\frac{dx}{dt} = x(-k+\lambda y), \tag{2.50}$$

$$\frac{dy}{dt} = y(a-by-\gamma x) \tag{2.51}$$

(uporediti sa sistemom (2.48), (2.49)), gde su λ i γ konstante interakcije. U stvari, gornji sistem običnih diferencijalnih jednačina bi se dobio ako pretpostavimo da se specifične stope rasta populacija x i y povećavaju odnosno smanjuju sa prisustvom one druge populacije.

Jednačine (2.50), (2.51) čine sistem nelinearnih običnih diferencijalnih jednačina, koji se, po pravilu ne može rešiti eksplicitno. Zbog toga je taj sistem potrebno analizirati u tzv. **faznoj ravni** xy, u kojoj su ove dve populacije tzv. **faze** posmatranog ekosistema, a te dve obične diferencijalne jednačine su **jednačine fazne ravni**. Rešenje jednačine fazne ravni se naziva **trajektorija**; u suštini, to je skup tačaka $(x(t),y(t))$ u xy ravni koje se dobijaju za sve vrednosti t iz nekog vremenskog intervala. Analiza trajektorija u faznoj ravni može nam dosta reći o ponašanju veličina tih populacija sa protokom vremena. Važno je reći da se dve trajektorije u faznoj ravni ili poklapaju, ili su disjunktne, tj. ne seku se ni u jednoj tački.

Kao i slučaju populacionog modela rasta jedne populacije, tako i u slučaju dve populacije definišemo tzv. **ravnotežne populacije**. One se za sistem (2.50), (2.51) dobijaju rešavanjem po x i y sistema jednačina

$$\begin{aligned} x(-k+\lambda y) &= 0, \\ y(a-by-\gamma x) &= 0. \end{aligned} \tag{2.52}$$

Lako je videti da su rešenja sistema (2.52) sledeće tri ravnotežne tačke:

1. $x = 0, y = 0$;
2. $x = 0, y = \dfrac{a}{b}$;
3. $x = \dfrac{a}{\gamma} - \dfrac{bk}{\gamma\lambda}, y = \dfrac{k}{\lambda}$.

Opštu analizu stabilnosti ravnotežnih tačaka u faznoj ravni za sisteme kao što je (2.50), (2.51) izložićemo u poglavlju 2.4. Primenom te analize dobija se da su prve dve ravnotežne tačke nestabilne.

Sa gledišta analize odnosa populacija lovaca i žrtava, od interesa je samo treća ravnotežna tačka, tj.

$$(x_r, y_r) = \left(\frac{a}{\gamma} - \frac{bk}{\gamma\lambda}, \frac{k}{\lambda}\right), \tag{2.53}$$

a i ona samo ako se nalazi u prvom kvadrantu xy ravni. Lako je videti da je za to potreban i dovoljan sledeći uslov:

$$\frac{a}{b} > \frac{k}{\lambda}, \tag{2.54}$$

Zadnji uslov je po pravilu zadovoljen, jer je u ekosistemima sa gore opisanom interakcijom konstanta b mnogo manja od ostale tri.

Sad ćemo pokazati da je pod spomenutim uslovom (2.54) treća ravnotežna tačka stabilna. Zapravo, analiziraćemo ponašanje trajektorija u blizini tačke (x_r, y_r). U tom cilju, stavimo

$$\begin{aligned} x(t) &= x_r + \varepsilon x_1(t) \\ y(t) &= y_r + \varepsilon y_1(t), \end{aligned} \tag{2.55}$$

gde su x_1 i y_1 odstupanja u vremenu t, do na multiplikativni faktor $\varepsilon > 0$ ("mali parametar"), po x- odnosno y-osi od tačke (x_r, y_r). Ako zamenimo (2.55) u sistem (2.50), (2.51), pa uprostimo dobijene jednačine, onda dobijamo sistem običnih diferencijalnih jednačina

$$\begin{aligned} \frac{dx_1(t)}{dt} &= (x_r + \varepsilon x_1(t))\lambda y_1 \\ \frac{dy_1(t)}{dt} &= -(y_r + \varepsilon y_1(t))(by_1 + \gamma x_1) \end{aligned} \tag{2.56}$$

Ovo je nelinearan sistem običnih diferencijalnih jednačina koji sadrži i mali parametar $\varepsilon > 0$. Zanemarivanjem članova koji u sebi sadrže ε, sistem (2.56) postaje

$$\begin{aligned} \frac{dx_1(t)}{dt} &= \lambda x_r y_1 \\ \frac{dy_1(t)}{dt} &= -\gamma y_r x_1 - b y_r y_1 \end{aligned} \tag{2.57}$$

Dobili smo linearan sistem običnih diferencijalnih jednačina po nepoznatim funkcijama x_1 i y_1. Diferenciranjem prve od gornje dve jednačine po t i zamenom u nju izvoda $\frac{dy_1(t)}{dt}$ iz druge, posle sređivanja dolazimo do sledeće jednačine:

$$\frac{d^2x_1(t)}{dt^2} + by_r\frac{dx_1(t)}{dt} + \gamma\lambda x_r y_r x_1(t) = 0 \tag{2.58}$$

(x_r i y_r iz (2.53)). Ovo je obična diferencijalna jednačina drugog reda po $x_1(t)$ sa konstantnim koeficijentima, čije se opšte rešenje oblika

$$x_1(t) = C_1 e^{r_1 t} + C_2 e^{r_2 t} \tag{2.59}$$

gde su C_1 i C_2 konstante koje zavise od početnih uslova, a r_1 i r_2 rešenja po r karakteristične jednačine

$$r^2 + by_r r + \gamma\lambda x_r y_r = 0.$$

Dakle,

$$r_{1,2} = \frac{-by_r \pm \sqrt{b^2y_r^2 - 4\gamma\lambda x_r y_r}}{2} = \frac{1}{2}\left(-\frac{bk}{\lambda} \pm \frac{1}{\lambda}\sqrt{b^2k^2 - 4k\lambda(a\lambda - bk)}\right).$$

Analiziraćemo sada izraz pod kvadratnim korenom. Ako prihvatimo da je "logistička" konstanta $b > 0$ mnogo manja od svih ostalih, onda je izraz pod kvadratnim korenom negativan, pa je veličina $\sqrt{b^2k^2 - 4k\lambda(a\lambda - bk)}$ imaginarna. Odavde sledi da su onda r_1 i r_2 konjugovano-kompleksni brojevi sa negativnim realnim delom $-\frac{bk}{2\lambda}$. To znači da rešenje $x_1(t)$ jednačine iz (2.59) teži 0 kad $t \to \infty$. Takođe je i

$$\lim_{t\to\infty} y_1(t) = 0,$$

jer je iz (2.59) i druge jednačine u (2.57)

$$y_1(t) = \frac{1}{\lambda x_r} C_1 r_1 e^{r_1 t} + C_2 r_2 e^{r_2 t}. \tag{2.60}$$

Na osnovu (2.56) sada možemo zaključiti da se tačka $(x_1(t), y_1(t))$ za velike vrednosti t nalazi blizu ravnotežne tačke (x_r, y_r), tj. ona je stabilna ravnotežna tačka.

Na modele tipa Lotka-Voltera vratićemo se u poglavlju 2.4, u okviru opšte analize stabilnosti sistema običnih diferencijalnih jednačina u okolini ravnotežnih tačaka. Uz ostalo, videćemo da su za $b > 0$ trajektorije u xy-ravni konvergentne (a time i stabilne) spirale, koje konvergiraju ka ravnotežnoj tački (2.53).

Za kraj ovog potpoglavlja pogledajmo jednu varijaciju modela Lotka-Voltera datog sa (2.50), (2.51).

Primer 2.11 (Leslijev model) Engleski naučnik Lesli, čiji smo model rasta jedne populacije sa podelom po starosti analizirali u potpoglavlju 2.2.4, je, uz ostalo, predložio i sledeći matematički model tipa Lotka-Voltera:

$$\frac{dx}{dt} = x(k - \lambda\frac{x}{y}), \tag{2.61}$$

$$\frac{dy}{dt} = y(a - by - \gamma x). \tag{2.62}$$

Primetimo da se umesto proizvoda λy u (2.50), sada u (2.61) pojavio količnik $\lambda x/y$, dok se jednačine (2.51) i (2.62) poklapaju. Leslijev model bi se mogao ovako tumačiti: ako ima mnogo predatora (lovaca) na jednu žrtvu, onda se broj predatora smanjuje ili oni čak izumiru, a ako u odnosu na jednog predatora ima mnogo žrtava, onda broj predatora raste. Dakle, u ovom modelu je bitan odnos predatora i žrtava.

2.3.3 Model dve populacije u takmičenju

Kao u prethodnom, i u ovom poglavlju ćemo posmatrati mali ekosistem sa dve populacije, koje ćemo opet označiti sa x i y, respektivno. Za razliku od modela tipa Lotka-Voltera, u kojima je jedna populacija bila plen druge, u ovom ćemo pretpostaviti da su u pitanju dve populacije koje se bore za isti prostor ili hranu. Ako ostale faktore u ovakvoj interakciji zanemarimo, onda dobijamo model **dve populacije u takmičenju**.

Za početak, da bismo konstruisali traženi matematički model, pretpostavićemo da se rast obe populacije može modelirati po logističkoj jednačini (2.24). Ako odgovarajuće koeficijente označimo redom sa a i b za populaciju x, odnosno sa c i δ, a populaciju y, onda dobijamo sledeći sistem običnih diferencijalnih jednačina:

$$\frac{dx}{dt} = x(a - bx), \tag{2.63}$$

$$\frac{dy}{dt} = y(c - \delta y) \tag{2.64}$$

(Umesto očekivanog d, ovde smo koristili grčko slovo δ, i to radi razlikovanja proizvoda δy od uobičajene oznake dy za diferencijal od y.)

Analogno jednačini (2.51), pretpostavićemo da se specifične stope rasta obe populacije smanjuju za negativni sabirak jednak proizvodu odgovarajućih konstanti (obeležićemo ih redom sa k_1 i k_2) i suparničke populacije. Tako dolazimo do sistema običnih diferencijalnih jednačina:

$$\frac{dx}{dt} = x(a - bx - k_1 y), \tag{2.65}$$

$$\frac{dy}{dt} = y(c - \delta y - k_2 x) \tag{2.66}$$

Analiza sistema (2.65), (2.66) je dosta komplikovana zbog prisustva šest konstanti. Zbog toga ćemo se ograničiti na važan slučaj da se radi o sličnim populacijama, u smislu da se rast obe populacije, u slučaju nepostojanja interakcije, može opisati istom logističkom jednačinom. Drugim rečima, uzećemo da je $a = c$ i $b = \delta$ kako u (2.63) i (2.64), tako i u (2.65) i (2.66). Tako dobijamo sledeći sistem običnih diferencijalnih jednačina:

$$\frac{dx}{dt} = x(a - bx - k_1 y), \tag{2.67}$$

$$\frac{dy}{dt} = y(a - by - k_2 x) \tag{2.68}$$

Pretpostavićemo još da važi produžena nejednakost

$$k_2 > k_1 > b. \tag{2.69}$$

Prvu nejednakost u (2.69) možemo protumačiti tako da populacija x "više" smeta populaciji y nego obratno, dok druga znači da su smanjenja stopa rasta obe populacije zbog opisane interakcije veća nego smanjenja zbog logističke prirode rasta obe populacije (u slučaju izostanka interakcije).

Ravnotežne tačke nalazimo kao rešenja sistema algebarskih jednačina:

$$\begin{aligned} x(a - bx - k_1 y) &= 0, \\ y(a - by - k_2 x) &= 0. \end{aligned} \tag{2.70}$$

Lako je videti da ova jednačina ima četiri rešenja, i to

1. $x = 0, y = 0$;
2. $x = 0, y = a/b$;
3. $x = a/b, y = 0$;
4. $x = \dfrac{ak_1 - ab}{k_1k_2 - b^2}, y = \dfrac{ak_2 - ab}{k_1k_2 - b^2}$;

Ostavljamo čitaocu da proveri (direktno, ili korišćenjem analize stabilnosti ravnotežnih tačaka, koja je data u poglavlju 2.4), da je prva tačka nestabilan ravnotežni položaj, a druga i treća stabilni ravnotežni položaji. U zadnja dva spomenuta slučaja dolazi do izumiranja jedne ili druge populacije, što, kako se čitalac lako može uveriti crtanjem odgovarajućih trajektorija, zavisi od veličina (početnih) populacija u momentu $t = 0$.

Sada ćemo ispitati četvrtu ravnotežnu tačku, (x_r, y_r) datu sa

$$(x_r, y_r) = \left(\frac{ak_1 - ab}{k_1k_2 - b^2}, \frac{ak_2 - ab}{k_1k_2 - b^2} \right), \tag{2.71}$$

koja je najinteresantnija kako sa biološkog, tako i sa matematičkog stanovišta. U tom cilju, napisaćemo $x(t)$ i $y(t)$ kao u relaciji (2.55), stim što je sada u pitanju ravnotežna tačka (x_r, y_r) iz (2.71). Posle zamene tih vrednosti u sistem (2.67), (2.68) i sređivanja, dobijamo sledeći sistem običnih diferencijalnih jednačina:

$$\frac{dx_1}{dt} = (x_r + \varepsilon x_1(t))(-bx_1 - k_1y_1), \tag{2.72}$$

$$\frac{dy_1}{dt} = (y_r + \varepsilon y_1(t))(-by_1 - k_2x_1) \tag{2.73}$$

Ako zanemarimo članove koji sadrže mali parametar ε, dobijamo sledeći linearni sistem običnih diferencijalnih jednačina sa konstatnim koeficijentima:

$$\begin{aligned} \frac{dx_1(t)}{dt} &= -bx_rx_1 - k_1x_ry_1 \\ \frac{dy_1(t)}{dt} &= -k_2y_rx_1 - by_ry_1. \end{aligned} \tag{2.74}$$

Na ovaj sistem običnih diferencijalnih jednačina ćemo se vratiti u sledećem poglavlju.

2.4 Stabilnost linearnog sistema običnih diferencijalnih jednačina

2.4.1 Analiza stabilnosti linearnog sistema običnih diferencijalnih jednačina

Modeli tipa Lotka-Voltera, dati sistemom (2.50), (2.51), i modeli dve populacije u takmičenju, dati sistemom (2.65), (2.66), su bili analizirani u poglavlju 2.3. Videli

smo da se oni međusobno veoma razlikuju, kako u biološkom, tako i u matematičkom smislu. Posebno je to došlo do izražaja u ponašanju trajektorija u faznoj ravni u neposrednoj blizini ravnotežnih tačaka (x_r, y_r), datih respektivno sa (2.53) i (2.71), jer je prva bila stabilna, a druga nestabilna ravnotežna tačka odgivarajućeg sistema.

Značajno je, međutim, da se posle uvođenja x_1 i y_1 sa (2.55), zamene u sistem (2.50), (2.51) (sistem Lotka-Voltera), i konačno zanemarivanja "malih"članova (tj. onih koji teže 0 kada $\varepsilon \to 0+$), dobija linearni sistem običnih diferencijalnih jednačina (2.57) po x_1 i y_1. Analogno se dobija linearni sistem po x_1 i y_1 običnih diferencijalnih jednačina (2.74), koji opisuje ponašanje trajektorija u okolini tačke (2.71) (dve slične populacije u takmičenju). Ostavljamo čitaocu da proveri da se do linearnog sistema običnih diferencijalnih jednačina dolazi i u ostalim slučajevima analize ponašanja oko ravnotežnih tačaka.

Dakle, radi analize stabilnosti ravnotežnih tačaka populacionih modela sa dve vrste u jednom ekosistemu, potrebno je analizirati stabilnost ponašanja opšteg linearnog sistema običnih diferencijalnih jednačina. Ova analiza je prostija, bar što se tiče oznaka, ako tačku (x_r, y_r) zamenimo sa koordinatnim početkom $O(0,0)$. Tada dolazimo do sledećeg sistema dve linearne obične diferencijalne jednačine po funkcijama $x_1 = x_1(t)$ i $y_1 = y_1(t)$:

$$\begin{aligned} \frac{dx_1(t)}{dt} &= Ax_1 + By_1 \\ \frac{dy_1(t)}{dt} &= Cx_1 + Dy_1, \end{aligned} \tag{2.75}$$

pri čemu su A, B, C i D konstante, od kojih je bar jedna različita od nule.

Pretpostavićemo da je $C \neq 0$. Onda se x_1 iz druge jednačine sistema (2.75) može izraziti preko funkcije y_1 i njenog prvog izvoda $\frac{dy_1}{dt}$, čime dolazimo do sledeće obične diferencijalne jednačine drugog reda sa konstantnim koeficijentima po y_1:

$$\frac{d^2y_1}{dt^2} - (A+D)\frac{dy_1}{dt} + (AD - BC)y_1 \quad = \quad 0. \tag{2.76}$$

Ako sada stavimo

$$P := (A+D) \quad \text{i} \quad Q := (AD - BC), \tag{2.77}$$

onda je karakteristična jednačina za (2.76) oblika

$$r^2 - Pr + Q \quad = \quad 0,$$

čija su rešenja

$$r_{1,2} \quad = \quad \frac{P \pm \sqrt{P^2 - 4Q}}{2}. \tag{2.78}$$

Ako je podkorena veličina u (2.78) pozitivna, tj. $P^2-4Q>0$, onda su rešenja sistema (2.75) oblika

$$\begin{aligned} y_1(t) &= C_1e^{r_1t}+C_2e^{r_2t} \\ x_1(t) &= C_1\frac{r_1-D}{C}e^{r_1t}+C_2\frac{r_2-D}{C}e^{r_2t}, \end{aligned} \tag{2.79}$$

gde su C_1 i C_2 proizvoljne konstante. Ako je, međutim, $P^2-4Q<0$, onda su rešenja sistema (2.75) oblika

$$\begin{aligned} y_1(t) &= e^{tP/2}\Big(C_1\sin(\alpha t)+C_2\cos(\alpha t)\Big) \\ x_1(t) &= \frac{1}{C}e^{tP/2}\Big[\Big((\frac{P}{2}-D)C_1-C_2\alpha\Big)\sin(\alpha t)+\Big(C_1\alpha+C_2(\frac{P}{2}-D)\Big)\cos(\alpha t)\Big], \end{aligned} \tag{2.80}$$

gde je stavljeno $\alpha=\frac{1}{2}\sqrt{4Q-P^2}$.

Ostaje da se izvede analiza stabilnosti tačke $O(0,0)$ kada $t\to\infty$ u zavisnosti od brojeva P i Q iz (2.77).

U slučaju da je diskriminanta $P^2-4Q>0$, onda je O **stabilna ravnotežna tačka** ako (i samo ako) su oba rešenja iz (2.78) negativni brojevi. Očevidno, to će važiti ako je tačka (P,Q) u drugom kvadrantu, ali "ispod" parabole $Q=P^2/4$.

Ako je diskriminanta $P^2-4Q<0$, onda je O **stabilna ravnotežna tačka** ako je realni deo oba rešenja negativan, tj. ako je tačka (P,Q) u drugom kvadrantu, ali "iznad" parabole $Q=P^2/4$.

U svim ostalim slučajevima (za P i Q različite od nule) je O **nestabilna ravnotežna tačka** za sistem (2.75). Ova analiza je predstavljena na slici 2.1.

Dodajmo da je pored gore navedenih glavnih slučajeva, od interesa i analiza stabilnosti na pozitivnom delu Q-ose. Naime, ona razdvaja oblast stabilnosti (za $P<0$) od oblasti nestabilnosti (za $P>0$), što može da stvori probleme u slučaju nedovoljno preciznih merenja konstanti P i Q.

Analizom jednačine fazne ravni

$$\frac{dy_1}{dx_1} = \frac{Cy_1+Dx_1}{Ax_1+By_1} \tag{2.81}$$

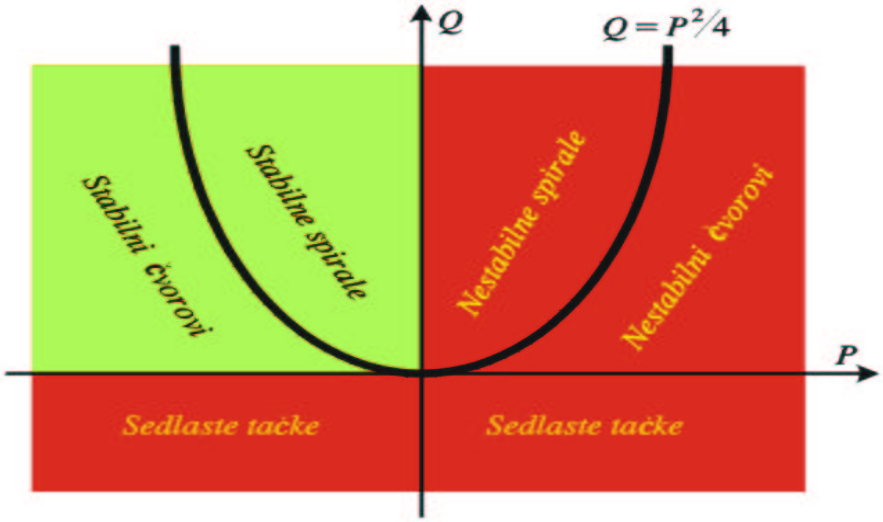
kompletiraćemo prethodnu analizu stabilnosti.

1. Ako je $P^2-4Q>0$ i $Q<0$, onda je O **sedlasta tačka**, koja je uvek nestabilna.

2. Ako je $P^2-4Q>0$ i $Q>0$, onda je O **čvor**, koji, u zavisnosti od P, može biti stabilan (za $P<0$), odnosno nestabilan (za $P>0$) ravnotežni položaj.

3. Ako je $P^2-4Q<0$ i $P\neq 0$, onda mora biti $Q>0$. U tom se slučaju u blizini O pojavljuju ili **spirale** (pa, kao u slučaju 2, stabilnost tačke O zavisi od znaka broja P), a ako su koreni čisto imaginarni, mogu se pojaviti i **zatvorene trajektorije** koje obuhvataju tačku O.

Gornji rezultati su predstavljeni na slici 2.1. Zelenom bojom (odn. svetlo-sivom) je obojena oblast u kojoj je ravnotežna tačka $O(0,0)$ stabilna, a crvenom bojom (odn. tamno-sivom) oblast u kojoj je nestabilna.



Slika 2.1: Oblasti stabilnosti odnosno nestabilnosti u PQ-ravni.

2.4.2 Primena analize stabilnosti sistema na populacione modele sa dve vrste

Primenićemo rezultate iz prethodnog potpoglavlja 2.4.1 na oba populaciona modela iz poglavlja 2.3. Tačnije, analiziraćemo stabilnost ravnotežne tačke (x_r, y_r) date u (2.53) u slučaju modela Lotka-Voltera, odnosno stabilnost tačke (x_r, y_r) date u (2.71) u slučaju modela dve populacije u takmičenju.

U potpoglavlju 2.3.2 videli smo da smena (2.55) sa x_r i y_r iz (2.53) i uz zanemarivanje članova koji sadrže faktor ε dovodi do sistema (2.57). U nastavku pretpostavljamo da važi uslov (2.54), što je potreban uslov da tačka (x_r, y_r) bude u prvom kvadrantu. Koristeći oznake za P i Q iz (2.77), dobijamo da je

$$\begin{aligned} &P = 0 - by_r = -by_r, \qquad Q = \lambda\gamma x_r y_r \quad \text{ i} \\ &P^2 - 4Q = (-by_r)^2 - 4(\lambda\gamma x_r y_r) = b^2 y_r^2 - 4\lambda\gamma x_r y_r. \end{aligned} \tag{2.82}$$

U zavisnosti od parametra b, imamo dva slučaja.

1. $b > 0$. U potpoglavlju 2.2.6, u kome smo analizirali logistički model rasta jedne populacije, videli smo da je "logistička" konstanta b u (2.24) mnogo manja od ostalih, na primer od a, λ ili γ (videti i relaciju (2.25)). Zbog toga je razumno pretpostaviti da je veličina $P^2 - 4Q$ u (2.82) negativna.

Budući da je i $P < 0$, a $Q > 0$, to je tačka (P,Q) u drugom kvadrantu slike 2.1 "iznad" parabole $Q = P^2$, što znači da su u ravni x_1y_1 trajektorije u okolini nule stabilne spirale. To povlači da su trajektorije u faznoj ravni xy u okolini ravnotežne tačke

$$(x_r, y_r) = \left(\frac{a}{\gamma} - \frac{bk}{\gamma\lambda}, \frac{k}{\lambda}\right)$$

spirale koje teže ka toj tački.

2. $b = 0$. Iz pretpostavke $b = 0$ sledi da je $P = 0$, $Q > 0$ i $P^2 - 4Q < 0$. Dakle, radi se o graničnom slučaju između oblasti stabilnosti i nestabilnosti, pa prema slici 2.1 sledi da su trajektorije u okolini tačke $O(0,0)$ krive koje sa malim pomeranjima parametra b mogu postati bilo stabilne bilo nestabilne spirale.

   U faznoj xy ravni su rešenja u blizini ravnotežne tačke $(x_r, y_r) = (a/\gamma, k/\lambda)$ linearne kombinacije sinusnih i kosinusnih funkcija sa istom periodom, što se slaže sa mnogobrojnim eksperimentalnim podacima za ove (najjednostavnije) modele tipa Lotka-Voltera.

Sada ćemo na sličan način analizirati ponašanje sistema (2.74) u okolini tačke $O(0,0)$, što će nam omogućiti analizu ponašanja modela dve populacije u takmičenju, potpoglavlje 2.3.3), oko tačke

$$(x_r, y_r) = \left(\frac{ak_1 - ab}{k_1k_2 - b^2}, \frac{ak_2 - ab}{k_1k_2 - b^2}\right).$$

Kako smo videli u potpoglavlju 2.3.3, zamenom (2.55) u sistem (2.67), (2.68) sređivanjem pa zanemarivanjem članova koji sadrže ε dolazimo do sistema (2.74).

Na osnovu uslova (2.56), dobijamo da je

$$P = -b(x_r + y_r) < 0, \qquad Q = x_r y_r (b^2 - k_1k_2) < 0 \quad \text{i}$$

$$P^2 - 4Q = b^2(x_r + y_r)^2 + 4x_r y_r (k_1k_2 - b^2) > 0,$$

pa na osnovu rezultata potpoglavlja 2.4.1 sledi da je u faznoj ravni x_1y_1 tačka $(0,0)$ nestabilna ravnotežna tačka (u stvari, **sedlasta tačka**). Iz toga sledi da je u ravni xy tačka (x_r, y_r) takođe nestabilna ravnotežna tačka.

Glava 3

Some mathematical aspects of traffic flow

3.1 Introduction

In this chapter we present mathematical models appearing in the so called Traffic Flow theory. Here we analyze only its most common type, namely the road traffic. As a matter of fact, nowadays practically everybody has some knowledge and personal experience about the traffic. Some of its aspects give much conformity, but we all know that, even in relatively small cities, one has everyday some traffic problems, like, e.g., traffic jams, parking problems and last but not least, accidents. Now, most drivers try to reduce travel time, and avoid accidents, thus there is a difficult and complicated task for engineers to design, control and manage the whole road system to help the drivers achieve such goals.

Hopefully, adequate mathematical models might help to understand some effects appearing in traffic, and help the engineers in constructing a road system with smooth and safe traffic. In the next few pages, we shall introduce the three main variables appearing in the traffic flow theory, find some of the main relations between them and try to solve some of them.

Let us note that the movement of a pollutant in a flow field is similarly modeled. Next, the movement of a steady stream of cars stopped with red light and then released with a green one shows similarity with propagation of waves. Finally, let us mention that traffic jams propagate very much like shock waves we hear whenever a supersonic jet flies over us.

For further reading and more complicated models than you will find in this text, one can consult, e.g., [4], [6], [7], [11].

3.1.1 The three main variables: velocity, density and flux

In every modeling process, one has to introduce the main variables and notions and then find and solve the main laws in it. However, in the case of the traffic flow theory, we can not expect to introduce any general theory which could cover all of its aspects, because each driver has its driving habits, sometimes very different than most other drivers have. Even such a single driver (say, driving too slow or too fast compared to others), might make big problems and change the whole traffic.

Whether we like or not, we have to accept that our models will, in the best case, describe some kind of average behaviors of cars, as seen by an observer. Essentially, there are two possible approaches here: to observe the phenomena of the motion of the stream of cars, or to observe that of individual cars. Being more appropriate for describing fluid motion, we shall use the first approach. Thus we shall mostly analyze a uniform, steady single lane traffic, with little possibilities for passing a slow car. Actually, this is a road situation which allows us to get a rather general model.

The first and perhaps most obvious variable we introduce in this text is the **traffic velocity** (or simply *velocity*), denoted in this text by u, and measured in (say) km/h. As said before, we observe cars in a single lane, so it is reasonable to assume that the velocity u depends on the position $x \in \mathbb{R}$ of the car in a moment $t \geq 0$. Thus we can write

$$u = u(x,t).$$

By definition, if in a moment t there is a car at the place x of the road, moving with a speed U then $u(x,t) = U$. In the other case, when there is no car at the place x and at the moment t, then $u(x,t) = 0$. Clearly, the velocity may vary between 0 and the maximum allowable speed $u_{\max}$, at least when the drivers see no police around.

The second variable is the **traffic density** (or simply *density*), denoted by ρ, measuring the number of cars on a unit road length, say on 1 km. In other words,

$$\rho = \frac{N}{x},$$

where N is the number of cars on the measured distance. Soon we shall find out that the density will be main variable in the forthcoming models. Of course, the density also depends on the place x and the time t, i.e.,

$$\rho = \rho(x,t).$$

Let us analyze for a moment two extreme cases appearing in traffic. The first appears when there are only few cars (or no cars at all) in the observed segment of the road; clearly, then we can conclude that the density, at least in that segment, is equal to 0. The other extreme case appears quite often in the rush hour, but also when there is a queue of stopped cars bumper to bumper by the red light (and thus waiting for the green light). Assuming for a moment that all vehicles are of the same length, say L, we

obtain that on such a segment of the road, the density is maximal, and we shall denote it by $\rho_{\max}$. Clearly, it holds $\rho_{\max} = 1/L$. This analysis also shows that the density ρ is always between 0 and $\rho_{\max}$.

The third variable we want to introduce is the number of cars passing at the point x in a moment t in a unit time, say in one hour, called **traffic flux** (shortly *flux*), or *traffic flow*, and denoted by F. Thus

$$F = \frac{N}{t},$$

where N is the number of cars that passed at a point during a time interval of length $t > 0$. As in the case of speed and density, the flux is also a function of x and t:

$$F = F(x,t).$$

As a matter of fact, the flux gives an information about the performance of the road system, which means that one of the important tasks for road engineers is enable traffic that has maximal flux.

3.1.2 The first equation: flux equals the product of speed and density

Let us show the relation announced in the title of this section, i.e., the equation

$$F = u \cdot \rho,$$

or, more precisely,

$$F(x,t) = u(x,t) \cdot \rho(x,t), \tag{3.1}$$

which we shall prove for all x and t.

To that end, assume first that there is a steady stream of cars moving at a constant velocity u_0 producing a constant density ρ_0 on then road. Now, if an observer finds that during a time period T, N cars pass in front of him, then the flux is equal to

$$F = \frac{N}{T}. \tag{3.2}$$

In time T, under our assumptions, every car moves for a distance of $u_0 \cdot T$, which implies that the density should be equal to

$$\rho_0 = \frac{N}{T \cdot u_0}. \tag{3.3}$$

Eliminating N from (3.2) and (3.3), we obtain (3.1), at least for constant velocity and constant density. The general case follows similarly, by analyzing the traffic in a small time Δt.

3.2 Conservation of cars

We are now in a position to formulate a deterministic model for traffic flow, namely

$$\frac{\partial \rho}{\partial t} + \frac{\partial F}{\partial x} = 0. \tag{3.4}$$

In order to prove this equation, let us take an arbitrary interval $[a,b]$ of the road for which we only assume that there are no entrances nor exits - in other words, the number of cars is conserved. Then, in view of the definition of the traffic density, we have that the number N of cars in the interval $[a,b]$, at a moment t is equal to

$$N(t) = \int_a^b \rho(x,t)\,dx. \tag{3.5}$$

Assuming that the density is a continuously differentiable function in t (a reasonable assumption at least in a steady traffic), we can differentiate in t under the integral sign in (3.5), and obtain

$$\frac{dN}{dt} = \frac{d}{dt}\int_a^b \rho(x,t)\,dx = \int_a^b \frac{\partial \rho}{\partial t}\,dx. \tag{3.6}$$

From the definition of the flux it follows that the derivative $\dfrac{dN}{dt}$ on the left hand side of (3.6) is equal to the difference $F(a,t) - F(b,t)$, or

$$\frac{dN}{dt} = F(a,t) - F(b,t). \tag{3.7}$$

The Newton-Leibniz formula allows us to rewrite this formula in a form

$$\frac{dN}{dt} = -\int_a^b \frac{\partial F}{\partial x}\,dt. \tag{3.8}$$

Now by combining (3.6) and (3.8), we obtain the equality

$$\int_a^b \left(\frac{\partial \rho}{\partial t} + \frac{\partial F}{\partial x}\right) dx = 0. \tag{3.9}$$

Since the interval $[a,b]$ has been chosen arbitrarily, and the partial derivatives appearing in the left-hand side of last equation were assumed to be continuous, it now follows that the last integrand is identically equal to zero. This brings us (3.4).

3.2.1 Velocity vs. density relations

Once we obtained the Conservation Law (3.4) for traffic flow (actually, one of the fundamental laws of nature), we see a problem with it. Namely, equation (3.4) is a differential equation with two unknowns, the density ρ and the flux F. In order to obtain

another equation, we need to find a mathematical relation between velocity and density; such a relation should originate from our experience, observation or, preferably, both. Such a relation is usually called **constitutive equation**.

To that end, let us note that a reasonable driver might go rather fast on an empty road (but not faster than the allowed limit speed $u_{\max}$), and does slow down once he finds himself in a heavier traffic. Thus we might assume that the velocity u of a car is a decreasing (or, at, least non-increasing) function of the density ρ. Hence our constitutive equation will be of the form

$$u = u(\rho), \quad 0 \le \rho \le \rho_{\max}, \tag{3.10}$$

where we assume that for all ρ

$$\frac{du}{d\rho} \le 0, \tag{3.11}$$

and, as explained above, the following conditions hold:

$$u(0) = u_{\max} \quad \text{and} \quad u(\rho_{\max}) = 0. \tag{3.12}$$

Perhaps the simplest relationship satisfying (3.11) and (3.12) is the following linear one:

$$u(\rho) = u_{\max}\Big(1 - \frac{\rho}{\rho_{\max}}\Big), \quad 0 \le \rho \le \rho_{\max}. \tag{3.13}$$

Using experimental observations, one can get much better models, e.g.,

$$u(\rho) = -c \ln \frac{\rho}{\rho_{\max}}. \tag{3.14}$$

Of course, equation (3.14) has sense only for positive ρ, hence we have to assume that the previous equation holds for $\rho \in [\rho_0, \rho_{\max}]$, where ρ_0 is a conveniently chosen positive density. For $\rho \in [0, \rho_0]$ we simply put

$$u(\rho) = u(\rho_0).$$

Let us add that the constant c turns out to be the velocity corresponding to the maximum flow (see the next subsection!).

3.2.2 Flux vs. density relation

Using (3.1) and the assumption (3.10) in the previous section, we get the following relation:

$$F(\rho) = \rho \cdot u(\rho). \tag{3.15}$$

As mentioned before, the road engineers want the flux to be as large as possible, hence they are looking for an "optimal" density, usually denoted by ρ_{opt}, which enables maximal flux.

In order to find the maximum flux, we derive (3.15) in ρ, getting

$$F'(\rho) \quad = \quad u(\rho)+\rho u'(\rho), \tag{3.16}$$

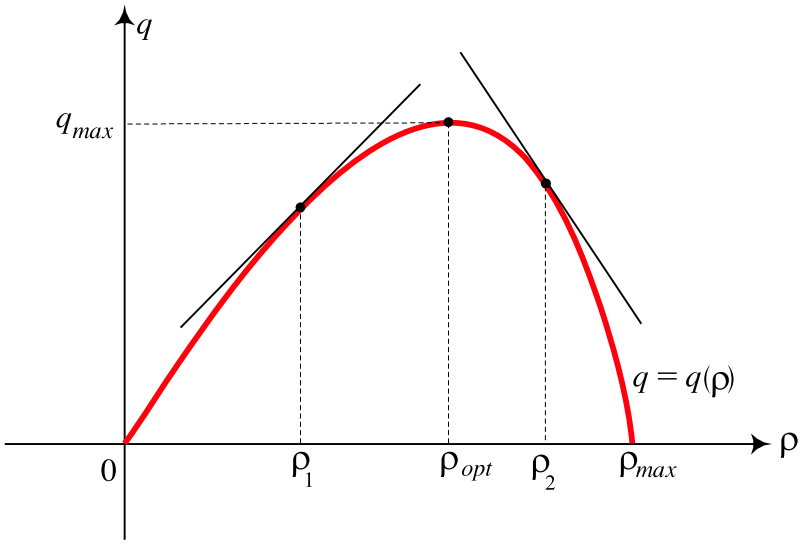
and then seek the value of ρ for which the last derivative is equal to 0.

Returning to (3.15), we see that the conservation of cars law, (3.4), can be written in a form

$$\frac{\partial \rho}{\partial t}+F'(\rho)\frac{\partial \rho}{\partial x} \quad = \quad 0. \tag{3.17}$$

Further on, we shall refer to (3.17) as the **traffic equation**.

In Picture 3.1 we see an example of a flux vs. density relation; as a matter of fact, the flux is zero both at $\rho = 0$ and $\rho = \rho_{max}$, and has a maximum at some point $\rho = \rho_{opt}$. (The importance of tangent lines on this picture will be explained in Section 3.3.2.)



Picture 3.1. Graph of flux as a function of density.

In particular, if $u(\rho)$ is given by (3.13), then we have the **parabolic model**:

$$F(\rho) \quad = \quad u_{\max}\left(\rho-\frac{\rho^2}{\rho_{\max}}\right), \quad 0 \le \rho \le \rho_{\max}. \tag{3.18}$$

In this (rather oversimplified) case, the maximum flow $F_{\max}$ is equal to $u_{\max}\rho_{\max}/4$, and is obtained for $\rho_{opt} = \rho_{\max}/2$.

3.3 Density waves

In this section we shall analyze a model of a steady traffic, whose solution will show that the traffic has a wave like behavior.

3.3.1 Solution of a model of nearly constant density

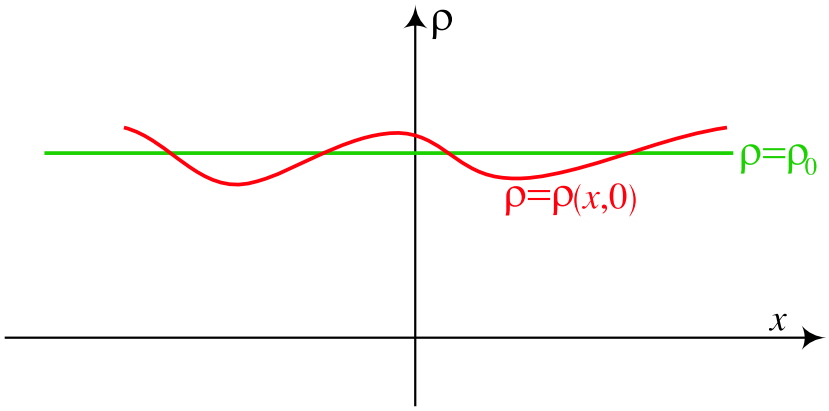
In this section we analyze a rather often appearing traffic situation, when on the whole road the traffic density is nearly equal to a constant ρ_0. In this case, the density can be modeled as

$$\rho(x,t) \quad = \quad \rho_0 + \varepsilon\rho_1(x,t), \quad x \in \mathbb{R},\ t \geq 0, \tag{3.19}$$

where the absolute value of the displacement $\varepsilon\rho_1(x,t)$ from ρ_0 is much less than the traffic density ρ_0. In (3.19), ε is a positive "small "parameter, and ρ_1 is a function to be determined. Note that assuming that the known initial density $\rho(x,0)$ is of the form

$$\rho(x,0) \quad = \quad \rho_0 + \varepsilon\phi(x), \quad x \in \mathbb{R}. \tag{3.20}$$

then $\rho_1(x,0) = \phi(x)$ for all $x \in \mathbb{R}$. The initial density is presented on Picture 3.2.



Picture 3.2. The initial density at time $t = 0$.

In order to find the solution of this model, i.e., the function ρ_1 from (3.19), we substitute it into (3.17) and thus obtain the following equation with the unknown function ρ_1:

$$\frac{d\rho_1}{dt} + F'(\rho_0 + \varepsilon\rho_1(x,t)) \cdot \frac{d\rho_1}{dx} \quad = \quad 0, \tag{3.21}$$

compare to (3.4).

Next, we replace the function F' (evaluated at $\rho_0 + \varepsilon\rho_1(x,t)$) with its first order Taylor polynomial (plus a remainder $r(\varepsilon)$ at the point ρ_0 in powers of $\varepsilon\rho_1(x,t)$), and obtain:

$$F'(\rho_0 + \varepsilon\rho_1(x,t)) \quad = \quad F'(\rho_0) + \varepsilon\rho_1(x,t) \cdot F''(\rho_0) + r(\varepsilon). \tag{3.22}$$

Here the remainder $r(\varepsilon)$ is a function of ε such that

$$\lim_{\varepsilon \to 0} \frac{r(\varepsilon)}{\varepsilon} \quad = \quad 0.$$

allowing us to neglect such terms in the continuation. Thus by replacing (3.22) into (3.21) takes us to the following **first order partial differential equation**:

$$\frac{\partial \rho_1}{\partial t} + F'(\rho_0) \cdot \frac{\partial \rho_1}{\partial x} \quad = \quad 0. \tag{3.23}$$

Denoting

$$c \quad = \quad F'(\rho_0), \tag{3.24}$$

we rewrite (3.23) in a form

$$\frac{\partial \rho_1}{\partial t} + c \cdot \frac{\partial \rho_1}{\partial x} \quad = \quad 0. \tag{3.25}$$

We remark that (3.23) is a special case of (3.15), namely when $F'(\rho)$ is a constant c.

Equation (3.25) can be most easily solved by introducing new variables x_1 and t_1 as follows:

$$x_1 \quad = \quad x - ct \quad \text{and} \quad t_1 = t. \tag{3.26}$$

Calculating the partial derivatives of x_1 and t_1 in both x and t, and then replacing them into (3.25), yields perhaps the simplest possible partial differential equation in $\rho_1 = \rho_1(x_1, t_1)$:

$$\frac{\partial \rho_1}{\partial t_1} \quad = \quad 0. \tag{3.27}$$

This equation implies that ρ_1 depends only on x_1, i.e., there is a function $\psi = \psi(x_1)$ such that

$$\rho_1(x,t) \quad = \quad \psi(x_1),$$

or, in view of (3.26),

$$\rho_1(x,t) \quad = \quad \psi(x - ct).$$

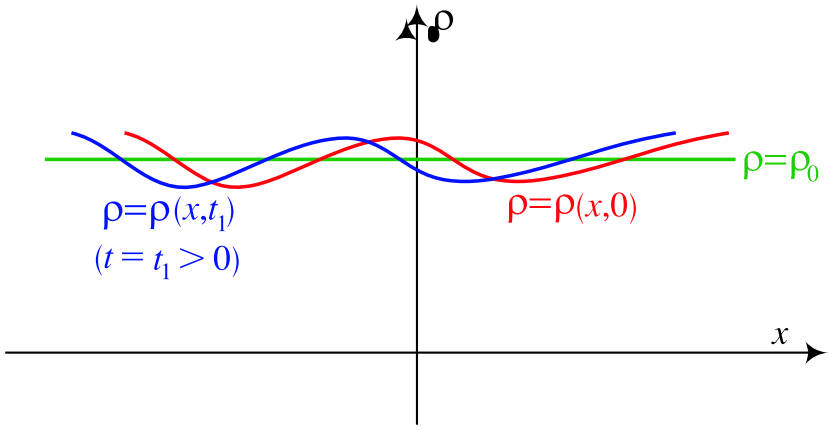
Putting $t = 0$ one sees at once from (3.20) that the function ψ is equal to the given function ϕ, hence we finally get

$$\rho(x,t) \quad = \quad \rho_0 + \varepsilon\phi(x - ct), \quad x \in \mathbb{R},\ t \geq 0. \tag{3.28}$$

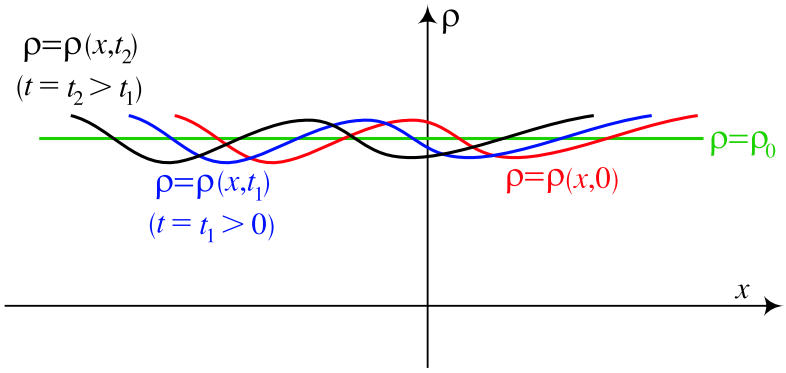
A visualization of the obtained solution in (3.28) is given in Pictures 3.3 and 3.4.[1] As in Picture 3.2, on these two pictures we also have the initial density $\rho(x,0)$, colored in red. On Picture 3.3, besides the red curve, there is also a blue one presenting the density at a later time t_1. Finally, on Picture 3.4 there are three curves, the red and the blue as before, and the third curve in black, presenting the density at a time $t_2 > t_1$.

One should notice that the blue and the black curve (obtained at times $t = t_1$ and $t = t_2$, respectively), are obtained by shifting the initial red curve to the left. In fact, these three pictures should explain the reader why it is said that the density behaves like a wave.

We return to this analysis in the next section.



Picture 3.3. The density at times 0 and $t_1 > 0$.



Picture 3.4. The density at times $0, t_1$ and $t_2 > t_1$.

[1] If you have the CD with this book, then you can see the colors mentioned in the text.

3.3.2 The density wave and characteristics

The obtained solution (3.28), together with the analysis in the previous section, shows that the traffic density propagates as a wave (here called **density wave**), and with velocity c from (3.24).

As we shall see later, the density wave might propagate in the same direction like the cars, but also might very well propagate in the opposite one. Geometrically, the velocity c is the slope of the (ρ, F) curve constructed at its point $(\rho_0, F(\rho_0))$, see Picture 3.1.

We shall return to the notion of density wave once we learn another important notion, namely the **characteristics**. The latter are curves, sometimes also called **Cauchy characteristics**, after the great French mathematician Augustin Cauchy (1789-1857), who invented them for solving first order partial differential equations. The main property of the characteristic curve (of a partial differential equation) is that along it this equation reduces to an ordinary one.

In our case, the density ρ will remain constant on the characteristics of the form $x - ct = C$ (where C=const), see equation (3.28). In fact, here the characteristics are straight lines, with slopes each equal to the constant c from (3.24).

An interpretation of this analysis is that if an observer moves with velocity c:

$$\frac{dx}{dt} = c, \tag{3.29}$$

then

$$\frac{d\rho}{dx} = 0. \tag{3.30}$$

In other words, the density ρ is a constant along the mutually parallel characteristics of the form

$$x - ct = C,$$

for different constants C.

3.3.3 A model of nonconstant density

Let us turn to a somewhat more involved model, namely when the traffic on the road is nonuniform. Firstly, remember that, starting from equation (3.23), we obtained that an uniform traffic moves approximately as a density wave. Now, in the nonuniform case, we shall assume that an observer moves in a known way, say determined by a function $x = x(t)$. Then we have

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \frac{dx}{dt} \cdot \frac{\partial \rho}{\partial x}. \tag{3.31}$$

Comparing this equation with the traffic equation (3.17), we obtain that the traffic density is constant along the curve $x = x(t)$, see (3.30), if

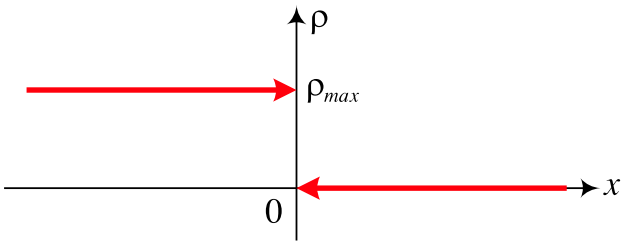
$$\frac{dx}{dt} = \frac{dF(\rho)}{d\rho} = F'(\rho). \tag{3.32}$$

Generally, $F'(\rho)$ may vary on different sections of the road, hence the wave density might also have different values on different characteristics.

3.4 Modeling the traffic behind red and green lights

Our task in this section is to model a too well known undesirable situation to all drivers, namely when there is a line of cars standing behind the red light, and waiting for the light turn green.

If we choose the origin of the x-axis at the traffic light, then we obtain the maximum density $\rho_{\max}$ behind the light, and it is reasonable to assume that there is zero density after the light.



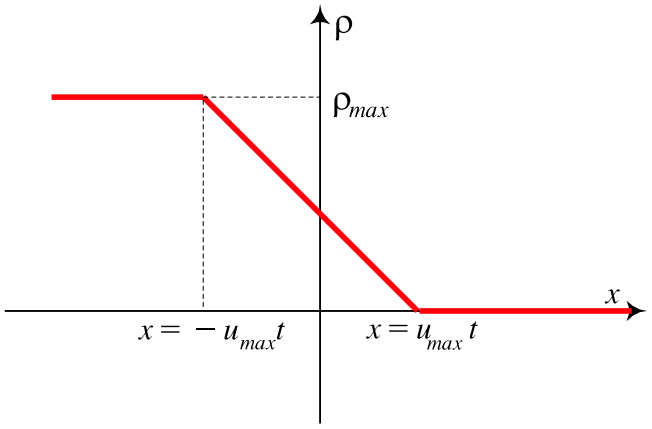
Picture 3.5. Maximal density behind, and zero density after the red light.

Thus the initial density, seen on Picture 3.5, is equal to

$$\rho(x,0) = \begin{cases} \rho_{\max}, & x \le 0; \\ 0, & x > 0. \end{cases} \tag{3.33}$$

The density after the light turns green is decreasing behind the light, but increasing after it. If the velocity vs. density relation is given by (3.13), then the density behind the light at some moment t, perhaps not too far from 0, is presented in Picture 3.6 below.

Once the light turns green, the front car starts (with some small delay, depending on the car and on the driver), and since the road in front of it is empty, we can assume that its speed soon becomes $u_{\max}$. Next, our driver's experience tells us that each car behind the leading one starts with certain delay, which depends on the car's initial position.



Picture 3.6. The density after the light turns green

We already know that a car in the line will start once it is reached by the backwards moving density wave, that started at the moment $t = 0$ and at the traffic light $x = 0$. In view of (3.15), the velocity of this wave is equal to

$$F'(\rho_{\max}) \;=\; u(\rho_{\max}) + \rho_{\max} u'(\rho_{max}) \;=\; \rho_{\max} u'(\rho_{max}),$$

which is the slope of the characteristics starting from the negative part of the x-axis. If the flux F is given by (3.18), then this slope is equal to

$$F'(\rho_{\max}) \;=\; -u_{\max}.$$

The density on each of these characteristics, colored in red on the Picture 3.7, is equal to $\rho = \rho_{\max}$. Thus the density in the "left" region $\mathcal{L}$, where

$$\mathcal{L} \;=\; \Big\{(x,t)\Big| x < -u_{\max}t\Big\}, \tag{3.34}$$

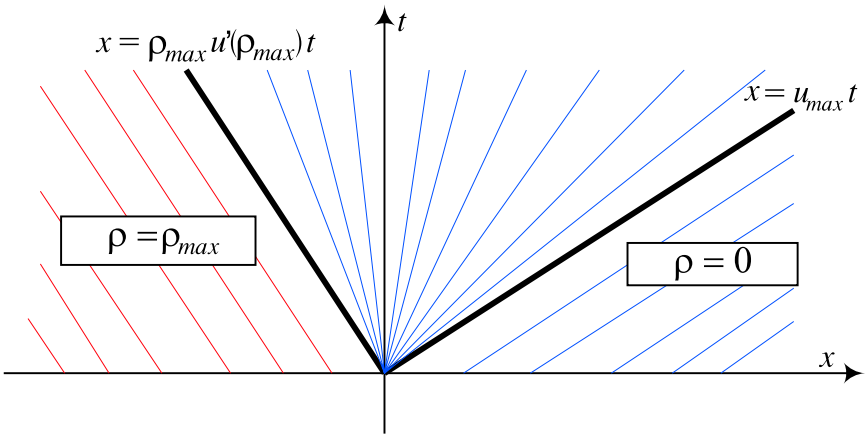
is equal to $\rho_{\max}$.

As noted before, the density after the light decreases, and since the cars are accelerating, we might assume that in a small time they will go with the maximum allowed speed $u_{\max}$. In view of (3.13), we obtain that the density in the "right" region $\mathcal{R}$ (see Picture 3.7) is equal to 0, where

$$\mathcal{R} \;=\; \Big\{(x,t)\Big| x > u_{\max}t\Big\}. \tag{3.35}$$

3.4.1 Rarefaction waves

There is still a part of the upper half plane ($t \geq 0$), where we have to calculate the density, namely a wedged region $\mathcal{W}$ with vertex at the origin, just between the regions $\mathcal{L}$ and $\mathcal{R}$, and containing the positive part of the t-axis (see again Picture 3.7).



Picture 3.7. The characteristics in the xt-plane

Firstly, note that the borders of $\mathcal{W}$ are the following two half-lines, both starting from the origin:

$$x = \rho_{\max} u'(\rho_{\max}) t, \quad t \geq 0,$$

(which is equal to $x = -u_{\max} t$ if (3.13) is assumed), and

$$x = u_{\max} t.$$

The region $\mathcal{W}$ is thus defined by

$$\mathcal{W} = \left\{ (x,t) \,\middle|\, -\rho_{\max} u'(\rho_{\max}) t < x < u_{\max} t \right\}. \tag{3.36}$$

In regions like $\mathcal{W}$ it might happen that there are no characteristics at all. For that reason, the method of characteristics will be replaced by the so called **rarefaction waves**. Essentially, this is a way of constructing the density on $\mathcal{W}$, satisfying the initial condition from (3.33). Unfortunately, this analysis would take us too far, so we have to omit it. Rather, let us just say that a rarefaction wave is a function of the form

$$u(x,t) = \phi((x-a)/t),$$

where ϕ is a nonconstant function and a a real number. For our purposes, it turns out that it is enough to take simply $a = 0$.

In view of the importance of characteristics, we are now seeking for lines having the property that the density in $\mathcal{W}$ is constant on them. Now, it is desirable that the density, for every fixed $t > 0$, on the lines we are looking for is continuously decreasing with x, more precisely, from $\rho_{\max}$ to 0. Thus we assume that there is an infinite

number of half-lines (we again call them, somewhat non-regularly, characteristics) in the wedged region, all starting from the origin. Each of these half-lines carries another density, starting from $\rho_{\max}$ and ending at 0.

Let us calculate the density at an arbitrary point (x,t) of the wedged region $\mathcal{W}$. For simplicity, we assume that the constitutive equation is the one from (3.13). Then the (blue) half-line through (x,t) starting from the origin has the slope equal to x/t, which should be equal to $F'(\rho)$, where the flux F is given by (3.18) and ρ has to be found.

Thus we have the equation

$$\frac{x}{t} \;=\; u_{\max}\Big(1-\frac{2\rho}{\rho_{\max}}\Big),$$

which gives us the density at the point (x,t) from the $\mathcal{W}$:

$$\rho(x,t)=\frac{\rho_{\max}}{2}\Big(1-\frac{x}{u_{\max}t}\Big). \tag{3.37}$$

Note that the slopes of the half lines vary continuously from $\rho_{\max}\cdot u'(\rho_{\max})$ (a negative number!) through 0 up to $u_{\max}$, as desired. The two boundary characteristics, dividing the upper half plane $\{(x,t)|\,t\geq 0\}$ in three regions, are thus $x=\rho_{\max}\cdot u'(\rho_{\max}t)$ and $x=u_{\max}t$.

For simplicity, we now assume that the flux $F=F(\rho)$ is given by (3.18). Let us find the maximal flux and the corresponding "optimal" density, usually denoted by ρ_{opt}. To that end, we first find that the derivative of F in ρ is equal to

$$F'(\rho) \;=\; u_{\max}\Big(1-\frac{2\rho}{\rho_{\max}}\Big).$$

Now the equation $F'(\rho)=0$ has only one solution, namely

$$\rho \;=\; \rho_{\text{opt}} \;=\; \frac{\rho_{\max}}{2}.$$

Since $F''(\rho)=-2u_{\max}/\rho_{\max}$ is a negative constant, hence also $F''(\rho_{\text{opt}})<0$, it follows that the function F has a local maximum at the point ρ_{opt}. One easily sees that at ρ_{opt} the function F attains in fact its absolute maximum on the segment $[0,\rho_{\max}]$.

But from (3.37) it follows that

$$1-\frac{x}{u_{\max}t} \;=\; 1$$

which is true only for $x=0$, i.e., at any time t the biggest flux is just at the traffic light.

3.4.2 Trajectory of a car

This model would not be complete without solving the following exercise:

Find the trajectory of the car that was at the position $-x_0<0$ in the initial moment

$t = 0$, when the light turned green. Assume that the constitutive equation is given by (3.13).

In order to solve the given exercise, we first note that the car's velocity dx/dt is equal to the velocity u from (3.13):

$$\frac{dx}{dt} = u_{\max}\left(1 - \frac{\rho}{\rho_{\max}}\right). \tag{3.38}$$

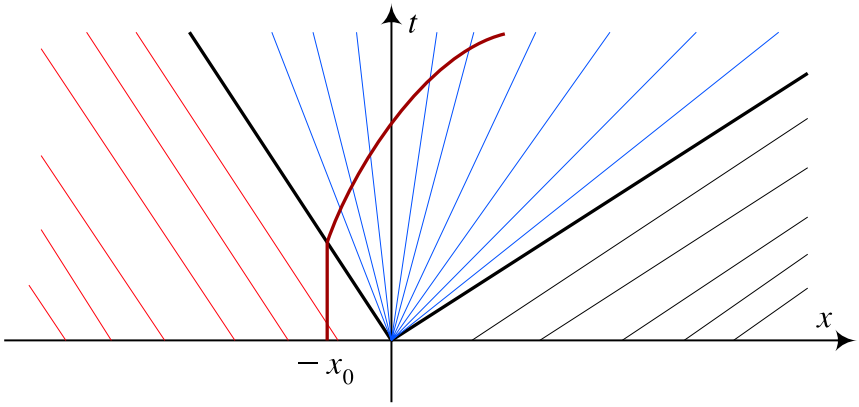
After inserting ρ from (3.37) into (3.38) we obtain the following first order ordinary differential equation of Euler type:

$$\frac{dx}{dt} = \frac{u_{\max}}{2} + \frac{x}{2t}.$$

We leave to the reader to check that the car, originally at the position $-x_0$, started to move at the moment $t_0 = x_0/u_{\max}$, and, after time $t \geq t_0$, arrives to the position $x(t)$, given by

$$x(t) = u_{\max}t - 2\sqrt{x_0 u_{\max} t}.$$

The trajectory of this car is shown on Picture 3.8.

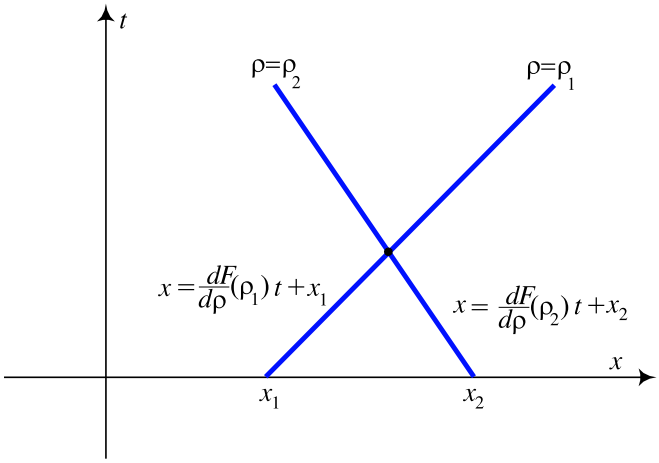


Picture 3.8. Trajectory of a car.

3.5 Shock waves

In Subsection 3.3.1 we used the assumption that the signal speed in the conservation law (3.25) is a constant, given by (3.24). This leads to a solution (3.28) propagating on mutually parallel characteristics given by (3.29) as a wave, here called density wave.

Now there is a natural question what happens if c is not a constant like in (3.24), but rather it holds $c = c(x,t)$, i.e., if the signal speed depends on x and/or t, or even on both. Actually, this situation of nonconstant c, may give us a nonlinear conservation law, and one of its effects is that there might appear two characteristics that intersect in some point, see Picture 3.9 below.



Picture 3.9. Two characteristics intersect.

As can be seen from (3.30), the densities are constants on characteristics, but it may very well happen that at the intersection point we have two (!) densities, meaning that the solution breaks down at that point. More precisely, an analysis of of the slope u_x in direction x shows that it might become infinite as $t \to t_b$, t_b being the time corresponding to the intersection of two (usually rather close and nearly parallel) characteristics.

This event is called **gradient catastrophe**; we shall soon learn that it often causes a jump discontinuity in the solution, called **shock wave**.

One can prove that the earliest breaking time is the minimum of the expression

$$t_b = \frac{-1}{\frac{d}{dx_0} c(\rho_0(x_0))}, \tag{3.39}$$

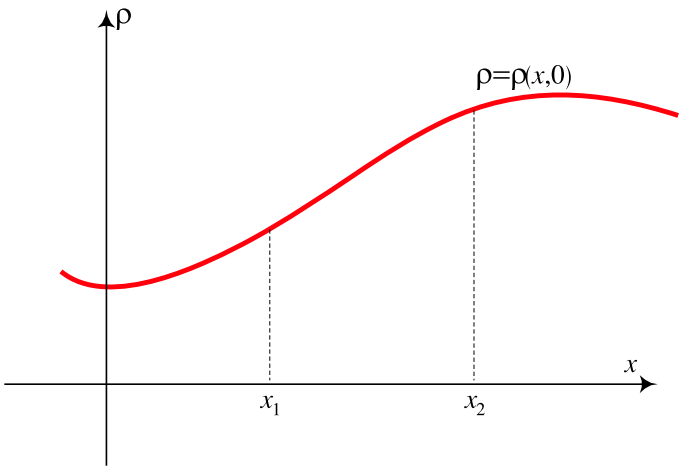
where x_0 is chosen so that it produces the earliest blowup time, while the function $c = c(\rho)$ is the one from the conservation law

$$\rho_t + c(\rho)\rho_x = 0, \tag{3.40}$$

$$\rho(x,0) = \rho_0(x). \tag{3.41}$$

The shock waves appear, e.g., in gas dynamics. There, changes in pressure and density of air propagate and sometimes can even be heard: ”sound waves” and ”sonic boom”. However, when amplitudes of fluctuations of pressure and density are large, then they can be modeled as being discontinuous - the **shock wave**.

Shock waves in the Traffic Flow theory appear if the initial density is increasing on some interval(s), as can be seen on Picture 3.10. (See also Picture 3.1, where the slopes of the tangent lines in ρ_1 and $\rho_2 > \rho_1$ indicate that the corresponding characteristics might intersect at some point.)



Picture 3.10. The initial density increases on some segment.

3.5.1 The Rankine-Hugeniot condition

The gradient catastrophe stops the solution of the problem (3.40) via the method of characteristics, thus it important to find an extension of the solution $\rho = \rho(x,t)$ beyond the breaking point t_b. Clearly, we have to admit solutions of (3.40) that are not necessarily smooth (continuously differentiable), but rather only piecewise smooth. The discontinuity arises along a curve in the xt-plane, called ”shock-path” (or simply *shock*), which should separate two families of characteristics. On the shock path the solution $\rho = \rho(x,t)$ has a jump discontinuity, but otherwise it is a smooth function.

Let us find this curve $x = x_s(t)$, or, more precisely, recognize it as the only right choice among many candidates. To that end, let us return to Section 3.2, where we obtained the differential form of the conservation law. The proof there depended on the equalities in (3.6), which assumed that the differentiation in t can be moved to the integrand. However, this is impossible in our case, since our function ρ is not smooth anymore.

To remedy this situation, we shall somewhat alter the derivation of the conservation law, used in Section 3.2. Firstly, we choose a sufficiently large segment $[a,b]$ of the

x-axis which contains the projection to the x-axis of the curve $x_s = x_s(t)$ for all $t \geq 0$.

With this choice of $[a,b]$, we can split the integral over $[a,b]$ into the following two:

$$\frac{d}{dt}\int_a^b \rho(x,t) \;=\; \frac{d}{dt}\int_a^{x_s(t)^-} \rho(x,t)\,dx + \frac{d}{dt}\int_{x_s(t)^+}^b \rho(x,t)\,dx.$$

We know from Section 3.2, equation (3.7), that the left-hand side of the last equation is equal to the derivative of the difference of the fluxes at a and at b, $F(a,t) - F(b,t)$. From (3.42) after differentiation in t we obtain

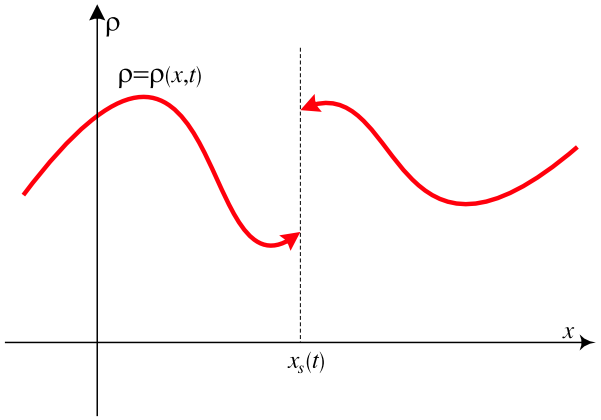
$$\int_a^{x_s(t)^-} \Big(\rho_t(x,t)\,dx + \rho(x_s(t)^-,t)\Big)\frac{dx_s}{dt} + \int_{x_s(t)^+}^b \Big(\rho_t(x,t)\,dx - \rho(x_s(t)^+,t)\Big)\frac{dx_s}{dt} \;=\; F(a,t) - F(b,t).$$

Letting now $a \to x_s^-$ and $b \to x_s^+$, and solving the obtained equation in the velocity of the shock dx_s/dt, we obtain the so-called **Rankine-Hugeniot condition**:

$$\frac{dx_s}{dt} \;=\; \frac{F(x_s^+,t) - F(x_s^-,t)}{\rho(x_s^+,t) - \rho(x_s^-,t)} \;=\; \frac{[F]}{[\rho]}. \tag{3.42}$$

where, e.g., $[F] = F(x_s^+,t) - F(x_s^-,t)$.

Now we obtained that if there is a jump in the initial density, then, as can be seen from Picture 3.11, the shock wave is immediate, and its **velocity** is given by (3.42).



Picture 3.11. The initial density has a discontinuity.

3.6 Model of a uniform traffic stopped by a red light or an obstacle

In this section we analyze an everyday traffic situation, namely, when a uniform stream of cars on a one lane road is stopped by an obstacle, say by a red light, railway line or an accident. We assume that the initial density on a one-lane road is

$$\rho(x,0) \;=\; \rho_0, \quad 0 < \rho_0 < \rho_{\max},$$

and suppose that the traffic is stopped for a certain period of time $T > 0$. It is a natural question to ask what happens with the traffic density after time T, i.e., once the red light turns green (or the obstacle has been removed).

According to the analysis in Subsection 3.5.1, we know that after the light turns red two immediate shocks appear at the traffic light (i.e., at the origin of the xt-plane). On Picture 3.4 these shocks are denoted by x_{sr} and x_{sl}.

For simplicity, in this section we assume that the linear relation between the velocity u and the density ρ from (3.13) holds. Before we continue our analysis of the given model, let us calculate the velocity of the shock if the initial density has the form

$$\rho(x,0) \;=\; \begin{cases} \rho_1, & x \le 0; \\ \\ \rho_2, & x > 0. \end{cases} \tag{3.43}$$

where $0 < \rho_1 < \rho_2 < \rho_{\max}$. (Actually, this is a special case of the function presented on Picture 3.11, since the function in (3.43) is piecewise continuous.)

From (3.42) and (3.13) we get

$$\begin{aligned} \frac{dx_s}{dt} &= \frac{F(\rho_2)-F(\rho_1)}{\rho_2-\rho_1} \\ &= \frac{u_{\max}\cdot\left(\rho_2-\frac{\rho_2^2}{\rho_{\max}}\right)}{\rho_2-\rho_1} - \frac{u_{\max}\cdot\left(\rho_1-\frac{\rho_1^2}{\rho_{\max}}\right)}{\rho_2-\rho_1} \\ &= u_{\max}\Big(1-\frac{\rho_1+\rho_2}{\rho_{\max}}\Big). \end{aligned} \tag{3.44}$$

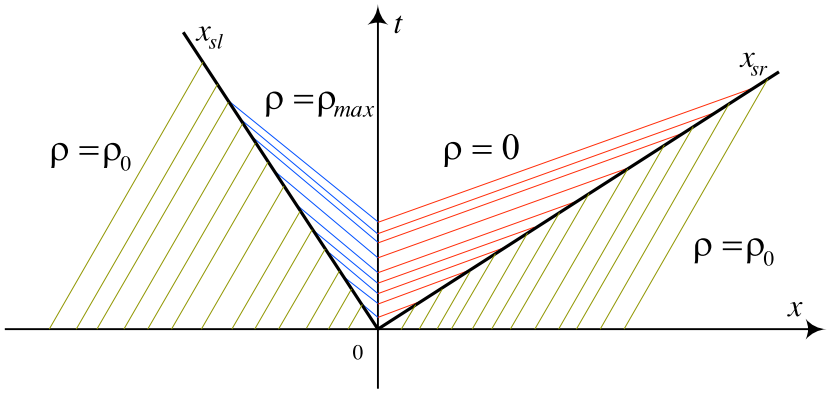
Applying equation (3.44), we obtain that the left-hand shock x_{sl} has the equation

$$\frac{dx_s}{dt} \;=\; u_{\max}\Big(1-\frac{\rho_0+\rho_{\max}}{\rho_{\max}}\Big) \;=\; -u_{\max}\frac{\rho_0}{\rho_{\max}}, \tag{3.45}$$

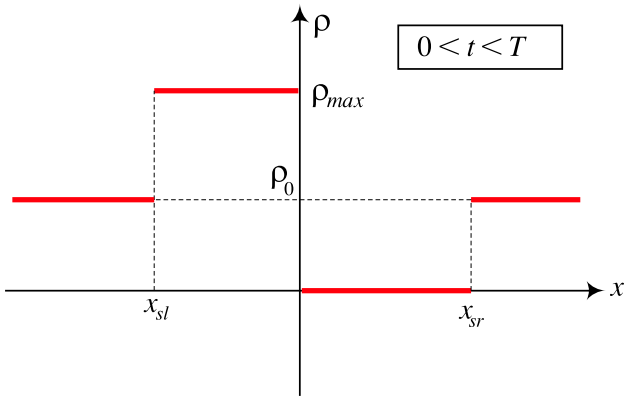
while the right-hand shock x_{sr} has the equation

$$\frac{dx_s}{dt} \;=\; u_{\max}\Big(1-\frac{0+\rho_0}{\rho_{\max}}\Big) \;=\; u_{\max}\Big(1-\frac{\rho_0}{\rho_{\max}}\Big). \tag{3.46}$$

Now on Picture 3.12 we see the positions of the two shocks at some moment t, $t \in (0,T)$, while on Picture 3.13 one can see the value of the density.

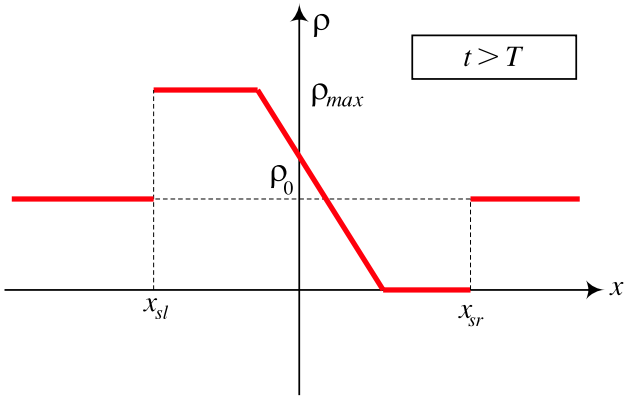


Picture 3.12. The two shocks and four families of characteristics at some time $t \in (0,T)$.

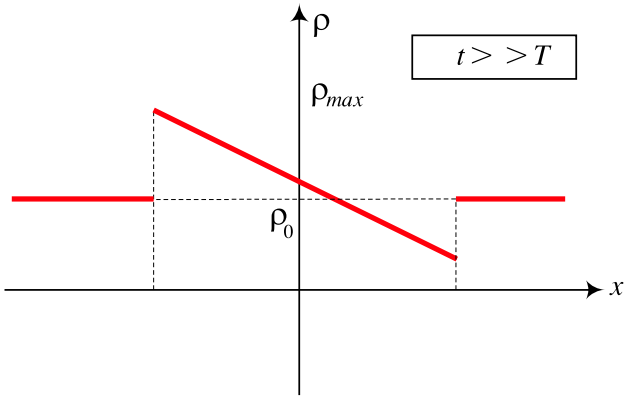


Picture 3.13. The density at some time $t \in (0,T)$, just after the light turned red.

Once the light turns green ($t > T$), the line of the cars eventually dissipates and the strengths of both shocks decrease, as can be seen from Picture 3.14. In fact, one can expect a rarefaction wave in this case.



Picture 3.14. The density at some time $t > T$.



Picture 3.15. The shock intensity decreases.

Assuming that (3.13) holds (the parabolic model), then, applying equation (3.44), the velocity of a shock is equal to

$$\frac{dx_s}{dt} = u_{\max}\left(1-\frac{\rho_0+\rho_f}{\rho_{\max}}\right). \tag{3.47}$$

Here ρ_f is the density in the fanlike region:

$$\rho_f = \frac{\rho_{\max}}{2}\left(1-\frac{x}{u_{\max}(t-T)}\right). \tag{3.48}$$

Thus we obtain the ODE

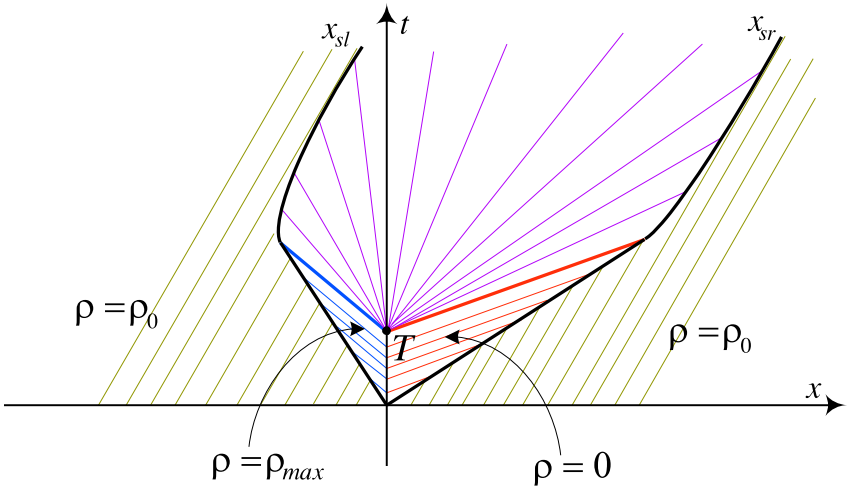
$$\frac{dx_s}{dt} = u_{\max}\left(\frac{1}{2}-\frac{\rho_0}{\rho_{\max}}\right)+\frac{x_s}{2(t-T)}. \tag{3.49}$$

Clearly, both the right and the left shock satisfy a corresponding initial condition. Solving the Euler type ODE from (3.49) for $t > T$, one finds the equation of the shock

is given by

$$x_s(t) \;=\; B\sqrt{t-T} + u_{\max}\left(1 - \frac{2\rho_0}{\rho_{\max}}\right)(t-T). \tag{3.50}$$

Finally, on Picture 3.16 one can see that the left- and the right-hand side shocks are both contained of a segment (see equations (3.45) and (3.46)) and the curve given by (3.50), of course with different constants B.



Picture 3.16. The density at some time $t > T$.

Note that on this picture we see also a rarefaction wave, as could have been guessed from Pictures 3.6 and 3.7.

3.6.1 Avoiding a crash with the last car in a queue

Careful drivers try to follow as much as possible the changes in traffic on the road. In particular, this is important if the driver is approaching a queue of cars, stopped for some reason ahead. Now if driver is at some optimal speed, here denoted by u_{opt}, it is likely that he will avoid accidents.

Mathematically, this means that the flux depends not just on the density ρ, but also on its partial derivative in x, $\dfrac{\partial \rho}{\partial x}$. Hence the flux is equal to

$$F \;=\; \rho u(\rho) - K\frac{\partial \rho}{\partial x} \tag{3.51}$$

for some constant $K > 0$. Thus we obtain the following partial differential equation:

$$\frac{\partial \rho}{\partial t} + \frac{\partial F}{\partial x} \;=\; \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u(\rho)) - K\frac{\partial^2 \rho}{\partial x^2} \;=\; 0. \tag{3.52}$$

Let $x = 0$ correspond to the end of the queue of cars, waiting for green light or a train to pass the crossing. Then the initial density is given by

$$\rho(x,0) \;=\; \begin{cases} \rho_{\max}, & x > 0; \\ 0, & x < 0, \end{cases} \tag{3.53}$$

(compare to (3.33).

Attention: In this model u is not necessarily the velocity but simply a function of ρ! Moreover, for a change, we assume that u is given by (3.14).

Now, because the density begins as steady (for $x \neq 0$), the solution should differ from $\rho(x,0)$ from (3.53) only in a narrow transition layer near $x = 0$.

Thus we neglect the term $\dfrac{\partial \rho}{\partial t}$ from (3.52), and seek the solution of

$$\frac{\partial}{\partial x}(\rho u(\rho)) - K\frac{\partial^2 \rho}{\partial x^2} \;=\; 0 \tag{3.54}$$

with the conditions

$$\begin{array}{l} \rho(\infty) = \rho_{\max}, \quad \rho'(\infty) = 0, \\ \rho(-\infty) = 0, \quad \rho'(-\infty) = 0, \end{array} \tag{3.55}$$

The solutions for u and ρ are, respectively,

$$\frac{u(x)}{u_{\max}} = \begin{cases} 1, & x < 0; \\ \exp(-u_{\text{opt}} x/K), & x \geq 0, \end{cases} \tag{3.56}$$

$$\frac{\rho(x)}{\rho_{\max}} = \begin{cases} \exp(u_{\max}(x/K - 1/u_{\text{opt}})), & x < 0; \\ \exp(-u_{\max} e^{-u_{\text{opt}} x/K}/u_{\text{opt}}), & x \geq 0. \end{cases} \tag{3.57}$$

Finally, let us add that the last model allows us to approximately find the constant K:

$$K \approx 0.1 \text{ m}^2/\text{h}.$$

Bibliografija

[1] J. Caldwell, D. K. S. Ng, *Mathematical Modelling: Case Studies and Projects*, Kluwer Academic Publishers, 2004.

[2] H. Caswell, *Matrix Population Models: construction, analysis, and interpretation*, Sinauer Associates, Inc. Publishers, 2001.

[3] V. Čerić, *Simulacijsko modeliranje*, Školska knjiga, Zagreb, 1993.

[4] N. D. Fawkes, J. J. Mahony, *An Introduction to Mathematical Modelling*, John Willey and Sons, Second Edition, 1996.

[5] H. I. Freedman, *Deterministic Mathematical Models in Population Ecology* Marcel Dekker, 1980.

[6] R. Haberman, *Mathematical Models*, Fourth Edition, Prentice-Hall 1977.

[7] R. Haberman, *Applied Partial Differential Equations*, Fourth Edition, Prentice-Hall 1998.

[8] M. Haddin, *Modelling and Quantitative Methods in Fisheries*, Chapman & Hall/CRC, 2001.

[9] J. A. Hamilton, Jr., D. A. Nash, U. W. Pooch. *Distributed Simulation*, CRC Press, 1997.

[10] D. Kelton, R. P. Sadowski, D. T. Sturrock, *Simulation with Arena*, Third Edition, McGraw-Hill, 2004.

[11] R. Knobel, *An Introduction to the Mathematical Theory of Waves*, American Mathematical Society, 2000.

[12] S. Lynch, *Dynamical Systems with Applications using MATLAB*, Birkhäuser, 2003.

[13] D. Mooney, R. Swift, *A Course in Mathematical Modeling* (Classroom Resource Material), MAA, 1999,

[14] E. Pap, A. Takači, Đ. Takači, *Partial Differential Equations through Examples and Exercises*, Kluwer Acadamic Publishers, 1997.

[15] B. Radenković, M. Stanojević, A. Marković, *Računarska simulacija*, Univerzitet u Beogradu, FON i Saobraćajni fakultet, Beograd, 1999.

[16] S. Ross, *Simulation*, Third Edition, Academic Press, 2002.

[17] M. Smith, *Models in Ecology*, Cambridge University Press, 1974.

[18] T. J. Schriber, *An Introduction To Simulation Using GPSS/H*, John Wiley & Sons, New York, 1991.

[19] B. P. Zeigler, H. Praehofer, T. G. Kim, *Theory of Modelling and Simulation*, Second Edition, Academic Press, 2000.

[20] *AnyLogic User Manual*, XJ Technologies, 2005.

[21] *Simulink, Model-Based and System-Based Design*, The Mathworks, 2002.

Project: 06SER02/02/003

Continuous time models

Dr Stevan Pilipović:

PART I

Generalized Functions and Operations (an introduction)

Teodor M. Atanacković, Stevan Pilipović
University of Novi Sad

Introduction

Introductory lectures contain basic definitions and assertions of the distribution theory, structural properties, operations in $\mathcal{D}'$ (emphasizing differentiation) and regularization. They are devoted to students having some knowledge of functional analysis. Everything in these lectures is very well known for many years but since this sophisticated theory is usually used without understanding its foundations,

The main feature of various generalized function theories is to explain real models in cases when the classical analysis could not give satisfactory justifications of corresponding mathematical models. Especially, the problem of differentiation leads to the notion of a weak derivative and thus to the new concept of a function called generalized function. Results of Sobolev (around 1937.) were in this sense the most important although, at that time, Dirac's calculus already existed for around fifteen years. But the monograph "Theorie des distributions I, II" (1950/51) of L. Schwartz is considered as the most important contribution to the generalized function theory and L. Schwartz is considered as the inventor and the main contributor of distribution theory. He was the first who published a systematized distribution theory on the basis of the theory of locally convex spaces which was very much developed in the end of fortes. There are many approaches but the concept of Schwartz is now days generally accepted and in fact today it is a language of modern mathematics and physics.

1 Space of basic functions

We will always denote by Ω an open set in $\mathbb{R}^n$. Recall, the support of a continuous function $f : \Omega \to \mathbb{C}$ is the adherent set for the set $\{x \in \Omega; f(x) \neq 0\}$. It is denoted by $\mathrm{supp}f$. If $x \in \mathrm{supp}f$, then there does not exist a neighborhood of x where $f = 0$. Since $\overline{A \bigcup B} = \overline{A} \bigcup \overline{B}$ it follows $\mathrm{supp}(f + g) \subset \mathrm{supp}f \bigcup \mathrm{supp}g$, where f and g are continuous.

$\mathcal{C}^m(\Omega), m \in \mathbb{N}_0$ or $m = \infty$, is the set of functions defined on Ω with all continuous derivatives up to order m. $\mathcal{C}_0^m(\Omega)$ is the subset of $\mathcal{C}^m(\Omega)$ consisting of functions with compact supports in Ω.

Clearly, $\mathcal{C}_0^m(\Omega) \subset \mathcal{C}_0^m(\mathbb{R}^n)$. With the usual addition and multiplication by scalars, $\mathcal{C}^m(\Omega)$ and $\mathcal{C}_0^m(\Omega)$ are vector spaces over $\mathbb{C}$. Elements of $\mathcal{C}^\infty(\Omega)$ are called smooth functions.

Example 1.1 The function

$$f(x) = \begin{cases} 0, & |x| \geq 1 \\ \exp(\frac{1}{|x|^2-1}), & |x| < 1 \end{cases} \quad (|x|^2 = x_1^2 + ... + x_n^2),$$

is an element of $\mathcal{C}_0^\infty(\mathbb{R}^n)$ and $\mathrm{supp}f = Z(0,1)$, where $Z(0,1) \subset \mathbb{R}^n$ is the closed

ball with the center at 0 and radius 1. (Open ball is denoted by $L(x_0, r)$.) Put $g(x) = c^{-1}f(x),\ x \in \mathbb{R}$, where $c = \int\limits_{\mathbb{R}^n} f(x)dx$. The function:

$$\omega_\varepsilon(x) = \epsilon^{-n} g(x\epsilon^{-1}), \varepsilon > 0, \quad x \in \mathbb{R}, \tag{1.1}$$

will be often used. It has the following properties:

$$\int_{\mathbb{R}^n} \omega_\varepsilon(x)dx = 1, \text{supp}\omega_\varepsilon(x) = \{x; |x| \leq \varepsilon\} \text{ and } \omega_\varepsilon(x) \geq 0.$$

With $\varepsilon = \dfrac{1}{n}, n \in \mathbb{N}$, we use the notation

$$\omega_{1/n}(x) = \delta_n(x), \quad x \in \mathbb{R}. \tag{1.2}$$

Denote by $\mathcal{C}_0^m(K), m \in \mathbb{N}_0$ or $m = \infty$, where K is a compact set in Ω ($K \subset\subset \Omega$), a subspace of $\mathcal{C}_0^m(\Omega)$ which elements are supported by K (supp$\phi \subset K$). We define in $\mathcal{C}_0^\infty(K)$, a sequence of norms $\{p_{K,m};\ m \in \mathbb{N}_0\}$:

$$p_{K,m}(\phi) = \sum_{|j| \leq m} \sup_{x \in K} |\phi^{(j)}(x)|.$$

With this norms, $\mathcal{C}_0^\infty(K)$ is a locally convex space; A basis of neighborhoods of zero is constituted by sets

$$V_{K,m,\nu} \equiv \{\psi \in \mathcal{C}_0^\infty(K); p_{K,m}(\psi) < \frac{1}{\nu}\}, \nu \in \mathbb{N}, m \in \mathbb{N}_0.$$

Vector space $\mathcal{C}_0^\infty(K)$ endowed with the quoted topology is the space of test functions $\mathcal{D}(K)$.

It follows that a sequence (ψ_k) in $\mathcal{D}(K)$ converges to $\psi \in \mathcal{D}(K)$ if and only if $(\psi_k^{(\alpha)})$ converges to $\psi^{(\alpha)}$ as $k \to \infty$, uniformly on K, for every $\alpha \in \mathbb{N}_0^n$.

Thorem 1.1 *a) If K is a compact subset of Ω, then there exists $\psi \in \mathcal{C}_0^\infty(\Omega)$ with the properties $0 \leq \psi \leq 1$ and $\psi(x) = 1$ in a neighborhood of K.*

b) If je $f \in \mathcal{C}^m(\Omega), m \in \mathbb{N}_0$ or $m = \infty$, then for every compact set $K \subset \Omega$ there exists an $f_K \in \mathcal{C}_0^m(\Omega)$ such that $f(x) = f_K(x)$ in a neighborhood of K.

$\mathcal{D}(K)$ is a projective limit of a sequence of normed spaces $(\mathcal{C}_0^\infty(K); p_{K,m})$, $m \in \mathbb{N}_0$ (see the appendix). We will explain this. The inclusion mappings

$$i_s^m, m \geq s, i_s^m : (\mathcal{C}_0^\infty(K), p_{K,m}) \to (\mathcal{C}_0^\infty(K), p_{K,s})$$

are continuous, since $p_{K,0}(\phi) \leq p_{K,1}(\phi) \leq ....$ The sequence of normed spaces with the mappings i_s^m makes a projective spectar. The topology in $\mathcal{D}(K)$ is the

coarsest topology such that the mappings $i^m : \mathcal{D}(K) \to (\mathcal{C}_0^\infty(K), p_{K,m}),\ m \in \mathbb{N}_0$, are continuous.

We have that $\mathcal{D}(K)$ is a Fréchet space with the metric

$$d(f,g) = \sum_{m=0}^{\infty} 2^{-m} \frac{p_{K,m}(f-g)}{1+p_{K,m}(f-g)}.$$

Since every Fréchet space is barreled, it follows that $\mathcal{D}(K)$ is also barreled (see the appendix).

If sets K_1 and K_2 are compact and $K_1 \subset K_2$, then $\mathcal{C}_0^\infty(K_1) \subset \mathcal{C}_0^\infty(K_2)$ and that the topology in $\mathcal{D}(K_1)$ is equal to the topology induced by $\mathcal{D}(K_2)$ on $\mathcal{D}(K_1)$. For an open set $\Omega \subset \mathbb{R}^n$ there exists a sequence (K_ν) of compact sets such that $K_\nu \subset K_{\nu+1}$ i $\bigcup_{\nu=1}^{\infty} K_\nu = \Omega$. Then the sequence of locally convex spaces $(\mathcal{D}(K_\nu))$ with respect to inclusion mappings $i_s^m : \mathcal{D}(K_m) \to \mathcal{D}(K_s), m \leq s$, constitutes a strict inductive spectar. Supply $\mathcal{C}_0^\infty(\Omega)$ by the topology of a strict inductive limit. This is the finest locally convex toplogy in $\mathcal{C}_0^\infty(\Omega)$ such that the identity mappings $i^m : \mathcal{D}(K_m) \to \mathcal{C}_0^\infty(\Omega)$ are continuous (see the appendix). The vector space $\mathcal{C}_0^\infty(\Omega)$ endowed with the above topology is the basic Schwartz space $\mathcal{D}(\Omega)$.

Note that in the definition of toplogy in $\mathcal{D}(\Omega)$ this topology does not depend on a sequence (K_ν) of compact sets with the properties quoted above.

If we take from every space $\mathcal{D}(K_\nu)$ a circled convex neighborhood of zero $U_\nu \in \mathcal{B}_\nu$, where $\mathcal{B}_\nu$ is a basis of neighborhoods of zero in $\mathcal{D}(K_\nu),\ \nu \in \mathbb{N}$, then the convex hull (envelope) U of the set $V = \bigcup_{\nu\in\mathbb{N}} U_\nu$:

$$U = \{\phi \in \mathcal{C}_0^\infty(\Omega); \phi = \sum_{j=1}^{n} \beta_j \psi_j, \psi_j \in U_{\nu_j}, \beta_j \geq 0, \beta_1 + ... + \beta_n = 1\}$$

is a circled convex neighborhood of zero in the topology of $\mathcal{D}(\Omega)$. In this way we can form the basis of neighborhoods in $\mathcal{D}(\Omega)$.

Thorem 1.2 *(i) A linear mapping T of the space $\mathcal{D}(\Omega)$ into a locally convex space E is continuous if and only if it is continuous on $\mathcal{D}(K)$ for every compact set $K \subset \Omega$.*

(ii) A set $A \subset \mathcal{D}(\Omega)$ is bounded if and only if there exists a compact set $K \subset \Omega$ such that $A \subset \mathcal{D}(K)$ and it is bounded in $\mathcal{D}(K)$.

(iii) A sequence $(\phi_\nu) \in \mathcal{D}(\Omega)^{\mathbb{N}}$ converges in $\mathcal{D}(\Omega)$ if and only if there exists a compact set K such that $(\phi_\nu) \in \mathcal{D}(K)^{\mathbb{N}}$ and that it converges in $\mathcal{D}(K)$

(iv) $\mathcal{D}(\Omega)$ is sequentially complete.

(Moreover, it is well known that a strict inductive limit of complete spaces is complete.)

Example 1.2 Let $f(x)$ be as in Example 1.1. Let $g_\nu(x) = \frac{1}{\nu} f(x),\ x \in \mathbb{R}^n,\ \nu \in \mathbb{N}$. All the members of the sequence have supports contained in the closed ball $Z(0.1)$; The sequence (g_ν) converges in $\mathcal{D}(Z(0,1))$ to zero and so it converges in $\mathcal{D}(\mathbb{R}^n)$. On the other hand the sequence $k_\nu(x) = \frac{1}{\nu} f(\frac{x}{\nu}),\ x \in \mathbb{R}^n,\ \nu \in \mathbb{N}$, does not converge in $\mathcal{D}(\mathbb{R}^n)$, because $\text{supp} k_\nu = Z(0,\nu)$, and there does not exist a compact set K containing all sets supp k_ν.

Since the inductive limit of barreled spaces is barreled, we have:

Thorem 1.3 *A closed and bounded set in $\mathcal{D}(\Omega)$ is compact. In particular, $\mathcal{D}(\Omega)$ is a Montel space (see the appendix).*

Note that the space $\mathcal{D}(\Omega)$ is not metrisabile.

2 Space of distributions

A continuous linear functional on $\mathcal{D}(\Omega)$ is called distribution. The space of distributions is denoted by $\mathcal{D}'(\Omega)$. We denote distributions as functions: $f, g, ...$;

$$f : \phi \mapsto \langle f, \phi \rangle = (f, \overline{\phi}),$$

where $\overline{\phi}$ is the complex conjugation for ϕ.

Example 2.1 Dirac distribution $\delta(\cdot - x_0) \in \mathcal{D}'(\Omega),\ x_0 \in \Omega$, is defined by

$$(\delta(x - x_0), \overline{\phi}(x)) = \langle \delta(x - x_0), \phi(x) \rangle = \phi(x_0), \phi \in \mathcal{D}(\Omega).$$

The following two theorems simplify very much the work with distributions.

Thorem 2.1 *A linear functional $f : \mathcal{D}(\Omega) \to \mathbb{C}$ belongs to $\mathcal{D}'(\Omega)$ if and only if if for every sequence $(\phi_\nu) \in \mathcal{D}(\Omega)^{\mathbb{N}}$ converging to zero in $\mathcal{D}(\Omega)$ it follows that (f, ϕ_ν) converges to zero in the set of complex numbers.*

Thorem 2.2 *Necessary and sufficient condition that a linear functional $f : \mathcal{D}(\Omega) \to \mathbb{C}$ is a distribution is that for every compact set $K \subset \Omega$ there exist constants $C > 0$ and $m \in \mathbb{N}$ such that for every $\phi \in \mathcal{D}(K)$*

$$| < f, \phi > | \leq C p_{K,m}(\phi). \tag{2.1}$$

Previous theorems are consequences of the fact that $\mathcal{D}(\Omega)$ is a bornological space (i.e. that a seminorm bounded on a bounded set in $\mathcal{D}(\Omega)$, is continuous on $\mathcal{D}(\Omega)$; see the appendix).

Recall that a family $\{A_i, i \in I\}$ of subsets of Ω is locally finite if for every $x \in \Omega$ there exists a neighborhood of $x, W(x)$, such that $W(x)$ has a non void intersection with at most finite number of sets $A_i, i \in I$.

Thorem 2.3 *A linear functional on $\mathcal{D}(\Omega)$ is a distribution if and only if there exists a family $R = \{\rho_\alpha;\ \alpha \in \mathbb{N}_0^n\}$ of continuous functions on Ω, such that the family of supports $\{\text{supp}\rho_\alpha;\ \alpha \in \mathbb{N}_0^n\}$ is locally finite and*

$$| < u, \phi > | \leq \sum_{\alpha \in \mathbb{N}_0^n} \sup_{x \in \mathbb{R}^n} |\rho_\alpha(x) \partial^\alpha \phi(x)|, \phi \in \mathcal{C}_0^\infty(\Omega).$$

Let $\mathcal{R}$ be a family of all families: $R = \{\rho_\alpha; \alpha \in \mathbb{N}_0^n\}$ in $\mathcal{C}^0(\Omega)$ such that the family $\{\sup \rho_\alpha; \alpha \in \mathbb{N}_0^n\}$ is locally finite. Theorem 2.3 implies that the topology of $\mathcal{D}(\Omega)$ is defined by the uncountable family of seminorms

$$P_R(\phi) := \sum_{\alpha \in \mathbb{N}_0^n} \sup_x |\rho_\alpha \partial^\alpha \phi|,$$

where $R \in \mathcal{R}$. (With this family $\mathcal{C}_0^\infty(\Omega)$ becomes a locally convex space with the dual $\mathcal{D}'(\Omega)$.)

Let $f \in \mathcal{D}'(\Omega)$. Assume that there exists $p \in \mathbf{N}_0$ such that (2.1) holds with $m = p$ and every compact set $K \subset \Omega$. Then the smallest p with this property is called the order of f.

Example 2.2 Let $j \in \mathbb{N}_0^n\ \Omega \ni 0$. Then $\delta^{(j)}; \phi \mapsto (-1)^{|j|} \phi^{(j)}(0)$ is a distribution of order $|j|$.

Let, now T be defined by

$$T : \phi \to \sum_{j=1}^{\infty} \phi^{(j)}(j), \phi \in \mathcal{D}(\mathbb{R}).$$

This is a distribution of infinite order.

Different topologies in $\mathcal{D}'(\Omega)$. The weak topology is defined by the family of seminorms: $\|f\|_\phi = | < f, \phi > |, \phi \in \mathcal{D}(\Omega)$. The topology of compact convergence is defined by the family of seminorms $\|f\|_K = \text{supp}\{| < f, \phi > |; \phi \in K\}$, where K is a compact subset od $\mathcal{D}(\Omega)$. The strong topology is defined by the family of seminorms: $\|f\|_B = \sup\{\langle |f, \phi\rangle |; \phi \in B\}$, where B is a bounded set in $\mathcal{D}(\Omega)$. They induce Hausdorff topologies. They are denoted by τ_w, τ_c, τ_b, and we have: $\tau_w \leq \tau_c \leq \tau_b$.

Thorem 2.4 *Topologies τ_c i τ_b in $\mathcal{D}'(\Omega)$ are equal.*

Clearly, τ_c is strictly finer than τ_w.

Thorem 2.5 *a) $\mathcal{D}'(\Omega)$ is sequentially complete with respect to the weak topology.*

b) Let (f_n) be a Cauchy sequence in $(\mathcal{D}'(\Omega), \tau_w)$ (in the sense of weak topology). Then it is a convergent one in the sense of strong topology.

In particular the convergence of sequences in the weak and strong topologies in $\mathcal{D}'(\Omega)$ coincide.

Moreover, $\mathcal{D}(\Omega)$ is bornological and it follows that $\mathcal{D}'(\Omega)$ is complete with respect to the strong topology in $\mathcal{D}'(\Omega)$. Also, as $\mathcal{D}(\Omega)$ is a Montel space, we have that $\mathcal{D}(\Omega)$ is reflexive i.e. the set $\mathcal{D}''(\Omega) = (\mathcal{D}'(\Omega))'$, of continuous linear functionals on $\mathcal{D}'(\Omega)$ with respect to the strong topology in $\mathcal{D}'(\Omega)$, equals $\mathcal{D}(\Omega)$ (semi reflectivity) and that the identity mapping $\mathcal{D}(\Omega) \to \mathcal{D}''(\Omega)$ is continuous with respect to the strong topology in $\mathcal{D}''(\Omega)$ (reflexivity).

It is easy to check that the sequence (δ_n) from Example 1.1 converges weakly to $\delta-$ distribution. Thus, $\delta_n \to \delta$ in the sense of strong topology.

Thorem 2.6 *Let $B' \subset \mathcal{D}'(\Omega)$. The following assertions are equivalent:*

(i) B' is bounded in the weak topology. (ii) B' is bounded in the strong topology. (iii) B' is uniformly continuous subset of $\mathcal{D}'(\Omega)$.

Regular distributions and Radon measures. It is easy to show that if $f \in L^1_{loc}(\Omega)$ then,

$$\phi \mapsto < f, \phi >= \int_{R^n} f(x)\phi(x)dx, \phi \in \mathcal{D}(\Omega),$$

defines an element of $\mathcal{D}'(\Omega)$. It is called regular distribution and it is denoted by $\tilde{f}$.

Thorem 2.7 *a) If a sequence of functions (f_n) iz $L^1_{loc}(\Omega)$ converges to 0 in $L^1_{loc}(\Omega)$, then this sequence determines a sequence of distributions (f_n) which converges to 0 in $\mathcal{D}'(\Omega)$.*

b) If $f, g \in L^1_{loc}(\Omega)$ and if for every $\phi \in \mathcal{D}(\Omega)$, $\langle \tilde{f}, \phi \rangle = \langle \tilde{g}, \phi \rangle$, then f equals g almost everywhere in Ω.

Thus, $L^1_{loc}(\Omega)$ is isomorphic to a subspace of $\mathcal{D}'$, called the space of regular distributions. So, $\mathcal{D}(\Omega)$ is a subspace of $\mathcal{D}'(\Omega)$. Moreover it is a dense subspace of $\mathcal{D}'(\Omega)$.

Note that Dirac δ-distribution given in Example 2.1 is not a regular distribution.

If $u \in L^1_{loc}(\mathbb{R}^n)$, we put

$$u_\varepsilon(x) = (u * \omega_\varepsilon)(x) := \varepsilon^{-n} \int\limits_{\mathbb{R}^n} u(y) g\Big(\frac{x-y}{\varepsilon}\Big) dy =$$

$$\int\limits_{\mathbb{R}^n} u(x - \varepsilon y) g(y) dy, x \in \mathbb{R}^n. \tag{2.2}$$

where g and ω_ε are smooth functions from Example 1.1. It is clear that (2.2) is a smooth function The operation $*$ is called the convolution.

If $u \in L^1_{loc}(\Omega)$, then (2.2) does not have sense, in general. In that case we consider $x \in \Omega_{-\varepsilon}$, where $\Omega_{-\varepsilon} = \{x \in \Omega;\ d(x, (\mathbb{R}^n \setminus \Omega)) > \varepsilon\}$, and by

$$\begin{aligned} u_\varepsilon(x) &= (u * \omega_\varepsilon)(x) = \varepsilon^{-n} \int\limits_{\Omega} u(y) g\Big(\tfrac{x-y}{\varepsilon}\Big) dy \\ &= \int\limits_{|y| \le 1} u(x - \varepsilon y) g(y) dy,\ x \in \Omega_{-\varepsilon} \end{aligned} \tag{2.3}$$

is defined a smooth function on $\Omega_{-\varepsilon}$.

In the first and second part of the next theorem the convolution is understood in the sense of (2.2) because we define u and $u_\nu,\ \nu \in \mathbb{N}$, to be equal zero out of Ω. The sequence $(u_{\nu,\nu})$ in the second part is in $\mathcal{C}^\infty(\mathbb{R}^n)$. In the fourth part convolution is understood in the sense of (2.3).

Thorem 2.8 *(i) If u is an integrable function vanishing out of a compact set $K \subset \Omega$ and if $\varepsilon \le d(K, (\mathbb{R}^n \setminus \Omega))$, then $u_\varepsilon \in \mathcal{C}^\infty_0(\Omega)$.*

(ii) If a sequence (u_ν) in $L^p(\Omega),\ p \in [1, \infty)$, in $L^p(\Omega)$-norm converges to $u \in L^p(\Omega)$, then $u_{\nu,\nu} = u_\nu * \delta_\nu \in \mathcal{C}^\infty(\mathbb{R}^n)$ for every $\nu \in \mathbb{N}$ and $u_{\nu,\nu} \to u$ u $L^p(\Omega)$.

(iii) If (u_ν) is a sequence of continuous functions in $\mathbb{R}^n$ which converges almost uniformly to u, then $(u_{\nu,\nu})$ is a sequence in $\mathcal{C}^\infty(\mathbb{R}^n)$ which converges almost uniformly to u.

(iv) If (u_ν) is a sequence in $L^p_{loc}(\Omega),\ p \in [1, \infty)$, which converges to $u \in L^p_{loc}(\Omega)$ (in the sense of $L^p_{loc}(\Omega)$ convergence), then $(u_{\nu,\nu}) \in \mathcal{C}^\infty(\Omega_{-1/\nu})$ and in the sense of $L^p_{loc}(\Omega)$ convergence converges to u.

Thus we have:

Thorem 2.9 *$\mathcal{C}^\infty_0(\Omega)$ is dense in: (i) $L^p(\Omega),\quad p \ge 1$; (ii) $\mathcal{C}^0_0(\Omega)$, (iii) $L^p_{loc}(\Omega),\ p \ge 1$.*

Since $L^p(\Omega), \mathcal{C}^0(\Omega)$, and $L^p_{loc}(\Omega)$ are subspaces of $L^1_{loc}(\Omega)$, they can be identified with the corresponding subspaces of the space of regular distributions.

Now we introduce the space of **Radon measures**. Let (K_ν) be a sequence of compact subsets of Ω such that $K_\nu \subset K_{\nu+1}$ i $\bigcup_\nu K_\nu = \Omega$. We define in $\mathcal{C}^0_0(K_\nu),\ \nu \in \mathbb{N}$, the norms

$$P_{K_\nu}(\phi) = \sup_{x \in K_\nu} |\phi(x)|, \qquad \phi \in \mathcal{C}^0_0(K_\nu).$$

The sequence of normed spaces $(\mathcal{C}^0_0(K_\nu),\ P_{K_\nu})$, with respect to identity mappings $i^\nu_{\nu+1} : (\mathcal{C}^0_0(K_\nu), P_{K_\nu}) \to (\mathcal{C}^0_0(K_{\nu+1}), P_{K_{\nu+1}})$, constitutes a strict inductive spectar.

We denote by $\mathcal{K}_c(\Omega)$ the respective strict inductive limit. (One can easily show that the construction does not depend on a chosen sequence of compact sets). Then we have the definition:

Radon (complex) measure on Ω is a continuous linear functional on $\mathcal{K}_c(\Omega)$.

Thorem 2.10 *A necessary and sufficient condition for a linear functional $f : \mathcal{K}_c(\Omega) \to \mathbb{C}$ to be continuous is that for every compact set $K \subset \Omega$ there exists $C > 0$ such that $|\langle f, \phi\rangle| \leq C_{P_K}(\phi)$ for every $\phi \in \mathcal{C}^0_0(K)$.*

We will use the following simple assertion:

If $\mathcal{C}^\infty_0(\Omega)$ is dense in a locally convex space V and if the topology in $\mathcal{D}(\Omega)$ is finer than the topology which V induces on $\mathcal{C}^\infty_0(\Omega)$, then V', the dual of V, is algebraically isomorphic to a subspace of $\mathcal{D}'(\Omega)$ (and we make the corresponding identification).

Thus we have that $\mathcal{C}^0_0(\Omega)$ is dense in $\mathcal{K}_c(\Omega)$ and since the inclusion $\mathcal{D}(\Omega) \to \mathcal{K}_c(\Omega)$ is continuous, the space of Radon measures is isomorphic to a subspace of $\mathcal{D}'(\Omega)$. On the basis of Theorem 2.10 we have that Radon measures are distributions of zero order.

Example 2.3 Dirac δ- distribution is the Radon measure.

Let f be a complex valued function on Ω. Then by

$$\mu(\phi) := \sum_{x \in \Omega} f(x)\phi(x),\ \phi \in \mathcal{C}^0_0(\Omega),$$

with the assumption that these series converge unconditionally, is defined a measure called atomic measure; in short,

$$\mu(t) = \sum_{x \in \Omega} f(t)\delta(t - x),$$

where $\delta(t - x)$ is Dirac $\delta-$distribution concentrated at x.

The unconditional convergence means that for every $\varepsilon > 0$ there exists a finite set $J_0(\varepsilon) \subset \Omega$, such that for every finite set set $J \subset \Omega$ such that $J_0(\varepsilon) \subset J$, it follows $\Big| \sum_{x \in J} f(x)\phi(x) - c \Big| < \varepsilon$, where $c \in \mathbb{C}$ is the sum of this series.

Example 2.4 Lebesgue measure is defined by the Riemann integral:

$$\mu(\phi) = \int_Q \phi(t)dt, \qquad \phi \in \mathcal{C}_0^0(\Omega),$$

where Q is a ball containing the support of ϕ. One can show that the Lebesgue measure is the only one Radon measure on $\mathbb{R}^n$ (up to the product with a positive constant) such that for any $\phi \in \mathcal{C}_0^0(\mathbb{R}^n)$ and any $h \in \mathbb{R}^n : \mu(\phi(\cdot - h)) = \mu(\phi)$.

Example 2.5 If f is a continuous function on Ω and ν is a given Radon measure on Ω, then by

$$\mu(\phi) = \nu(f \cdot \phi) \tag{2.4}$$

is defined a Radon measure. We denote μ by $f \cdot \nu$. Moreover, one can show that μ is well defined by (2.4) if f is a locally integrable function with respect to ν. Then it is said that $\mu = f \cdot \nu$ has the density f or that μ has a "derivative" f with respect to ν.

Example 2.6 Lebesgue-Stieltjes measure is defined similarly as the Lebesgue measure (Example 2.4) but with the Lebesgue -Stieltjes integral instead of the Lebesgue integral (with a corresponding function of locally bounded variation).

Sheaf properties. Distributions f and g in $\mathcal{D}'(\Omega)$ are equal on an open set $\Omega_1 \subset \Omega$ if and only if their restrictions on Ω_1 (distributions applied on test functions supported by Ω_1) are equal on Ω_1. It is said that they are equal in a neighborhood of $x \in \Omega$ if they are equal in an open neighborhood of this point.

Thorem 2.11 *Let $f, g \in \mathcal{D}'(\Omega)$ such that they are equal in the neighborhood of every point of Ω. Then $f = g$ in Ω.*

The proof of this theorem is based on the following finite partition of unity:

Let $\Omega_1, ..., \Omega_k$ be open sets and let K be a compact set such that $K \subset \bigcup_{i=1}^k \Omega_i = \Omega$ and $K \cap \Omega_i \neq \emptyset$, $i = 1, ..., k$. Then there exist $\phi_i \in \mathcal{C}_0^\infty(\Omega_i)$, $i = 1, ..., k$, such that $\phi_i \geq 0$; $\sum_{i=1}^k \phi_i \leq 1$ and $\sum_{i=1}^k \phi_i(x) = 1$, in a neighborhood of K in Ω.

Now we can define the support of a distribution: The support of $f \in \mathcal{D}'(\Omega)$, $\mathrm{supp} f$, is the set of points in Ω having a neighborhood where f is not equal to zero.

In other words, the support of $f \in \mathcal{D}'(\Omega)$ is the complement of the union of all open sets in Ω where $f = 0$. It is a closed set.

The next notion is of a great importance in the theory of partial differential equations.

The singular support of $f \in \mathcal{D}'(\Omega)$, sing supp$f$, is the set of points in Ω which do not have a neighborhood where f is a smooth function (i.e. f is defined as a regular distribution by a smooth function).

In other words the singular support is the complement of the union of all open sets where f is a smooth function.

Partition of unity. A family $\{U_\alpha;\ \alpha \in A\}$ of subsets of a topological space X is a cover of $W \subset X$ if $W \subset \bigcup\limits_{\alpha\in A} U_\alpha$. It is an open cover if every U_α is an open set. Subfamily of $\{U_\alpha\}$ which also covers W is called a sub cover and $\{V_\beta;\ \beta \in B\}$ is called a refinement of $\{U_\alpha\}$ if it covers W and for every $\beta \in B$ there exists $\alpha \in A$ such that $V_\beta \subset U_\alpha$. Recall, a family $\{A_i;\ i \in I\}$ is locally finite in X if for every $x \in X$ there exists its neighborhood ω_x such that $\omega_x \cap A_i \neq \emptyset$ only for a finite number of indices $i \in I$.

We will assume that a topological space X is paracompat (every open cover of X has a locally finite sub cover).

Partition of unity on an open set $\Omega \subset X$ is a family $\{\phi_i;\ i \in I\}$ of continuous functions on Ω such that:

(a) The family of supports $(\text{supp}\phi_i;\ i \in I\}$ is locally finite.

(b) $\sum\limits_{i\in I} \phi_i(x) = 1$ for every $x \in \Omega$ and $\phi_i(x) \geq 0$ for every $x \in \Omega$, $i \in I$.

(Note for every x the sum is finite.)

The partition of unity $\{\phi_i;\ i \in I\}$ corresponds to the cover $\{U_\alpha;\ \alpha \in A\}$ if for every $i \in I$ there exists α_i such that $\text{supp}\phi_i \subset U_{\alpha_i}$.

Since the set $\mathbb{R}^n$ is paracompact we have:

Thorem 2.12 *Let $\Omega \subset \mathbb{R}^n$ and $\{U_\alpha;\ \alpha \in A\}$ be an open cover of Ω. Then there exist a countable partition of unity $\{\phi_i;\ i \in \mathbb{N}\}$ which corresponds to $\{U_\alpha\}$ such that functions ϕ are smooth and* $\text{supp}\phi_i$ *are compact sets, for every $i \in \mathbb{N}$.*

Now one can prove the (second sheaf) property:

Thorem 2.13 *Let $\{\Omega_\alpha;\ \alpha \in A\}$ be a family of open set and let $f_\alpha \in \mathcal{D}'(\Omega_\alpha)$ $\alpha \in A$. Assume that for every $\alpha, \beta \in A$ such that $\Omega_\alpha \cap \Omega_\beta \neq \emptyset$ there holds: $f_\alpha = f_\beta$ on $\Omega_\alpha \cap \Omega_\beta$. Then, there exists one and only one distribution f on $\Omega = \bigcup\limits_\alpha \Omega_\alpha$ such that $f = f_\alpha$ on Ω_α.*

Without explanations, note that the sheaf of distributions is a fine sheaf but not flabby.

3 Operations in $\mathcal{D}'(\Omega)$

Operations on the whole $\mathcal{D}'(\Omega)$ are called regular operations while those operations defined only on appropriate distributions (proper subset of $\mathcal{D}'(\Omega)$) are called irregular ones.

Change of variables. Let

$$x = Ay + b, \qquad y \in \Omega_1, \tag{3.1}$$

where A is a regular linear transformation of Ω_1 onto Ω and $b \in \mathbb{R}^n$. (Ω and Ω_1 are open sets in $\mathbb{R}^n$). Let $f \in \mathcal{D}'(\Omega)$. Then $f(Ay+b) \in \mathcal{D}'(\Omega_1)$ is defined by

$$\langle f(Ay+b), \phi(y)\rangle := \langle f(x), \frac{\phi(A^{-1}(x-b))}{|detA|}\rangle,\ \phi \in \mathcal{D}(\Omega_1). \tag{3.2}$$

Since $\phi(y) \mapsto \phi(A^{-1}(x-b))$ is a mapping from $\mathcal{D}(\Omega_1)$ into $\mathcal{D}(\Omega)$, it follows that $f(Ay+b) \in \mathcal{D}'(\Omega_1)$.

If in (3.1) $A = I$ (I is the identity matrix), $b = -h = -(h_1, ..., h_n)$, then (3.2) defines the translation: $\langle f(t-h), \phi(t)\rangle := \langle f(t), \phi(t+h)\rangle$.

If $A = -I,\ b = 0$, (3.2) defines the transposition: $\langle f(-t), \phi(t)\rangle = \langle f(t), \phi(-t)\rangle$.

If $A = aI,\ a \neq 0,\ b = 0$, (3.2) defines the homotety:

$$\langle f(at), \phi(t)\rangle := \langle f(t), \frac{1}{|a|^n}\phi(\frac{t}{a})\rangle.$$

It is said that $f \in \mathcal{D}'(\mathbb{R}^n)$ is homogeneeous of order λ if for every homotety $y = ax,\ a > 0$: $f(ax) = a^\lambda f(x)$ i.e.

$$\langle f(x), \phi(\frac{x}{a})\rangle = a^{\lambda+n}\langle f(x), \phi(x)\rangle,\ \phi \in \mathcal{C}_0^\infty(\mathbb{R}^n),\ a > 0.$$

This follows from

$$\langle f(ax), \phi(x)\rangle = \frac{1}{a^n}\langle f(x), \phi(\frac{x}{a})\rangle;\ \langle f(ax), \phi(x)\rangle = a^\lambda\langle f(x), \phi(x)\rangle.$$

For example, $\delta(x)$ is homogeneous of order $-n$.

If A is an orthogonal matrix ($A^{-1} = A^T$) and $b = 0$, then (3.2) defines the rotation:

$$\langle f(At), \phi(t)\rangle := \langle f(t), \phi(A^T t)\rangle \quad (|detA| = 1).$$

For example, $\delta(x)$ is invariant with respect to the rotation around zero.

In general, let $y_1, ..., y_n$, be smooth functions defined in $\Omega \subset \mathbb{R}^n$ such that

$$\alpha : x \mapsto y = (y_1(x), ..., y_n(x)) \tag{3.3}$$

is the bijection of Ω onto $\Omega_1 \subset \mathbb{R}^n$ (then at every point $x \in \Omega$ the Jacobian

$$J = \frac{\partial(y_1, ..., y_n)}{\partial(x_1, ..., x_n)} := \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_n} \end{vmatrix}$$

is different from zero). Mapping (3.3) has the smooth inverse $\alpha^{-1} : \Omega_1 \ni y \mapsto x \in \Omega$ and the mapping $\mathcal{D}(\Omega) \to \mathcal{D}(\Omega_1)$

$$\phi(x) \mapsto \phi(\alpha^{-1}(y))|J^{-1}(y)|, \quad y \in \Omega_1, \ x \in \Omega \tag{3.4}$$

is continuous. Thus, for $f \in \mathcal{D}'(\Omega_1)$,

$$\langle f(\alpha(x)), \phi(x)\rangle := \langle f(y), \phi(\alpha^{-1}(y))|J^{-1}(y)|\rangle, \ \phi \in \mathcal{D}(\Omega)$$

defines a mapping $\mathcal{D}'(\Omega_1) \to \mathcal{D}'(\Omega)$. It is adjoined to (3.4) : $f(y) \mapsto f(\alpha(x))$. It is called the pull back and denoted by α^* ($\alpha^* f(x) = f(\alpha(x))$). This mapping is continuous with respect to weak, respectively strong, topologies in $\mathcal{D}'(\Omega_1)$ and $\mathcal{D}'(\Omega)$.

Note, if a distribution $f(y)$ is defined by $f \in L^1_{loc}(\Omega_1)$, then $\alpha^* f$ is defined by the locally integrable function $f(\alpha(\cdot))$.

The change of variables can be defined as an irregular operation in cases when the above assumptions do not hold for α. Let us show this.

Let $\alpha : \Omega \to \Omega_1$ be a smooth mapping and $f \in \mathcal{D}'(\Omega_1)$. Let δ_n be as in (1.2) and $f_n = f * \delta_n$. If $\lim\limits_{n\to\infty} f_n(\alpha(x))$ exists in $\mathcal{D}'(\Omega)$, then

$$f(\alpha(x)) := \lim_{n\to\infty} f_n(\alpha(x)) \quad \text{in } \mathcal{D}'(\Omega). \tag{3.5}$$

$$(\ \langle f(\alpha(x)), \phi\rangle = \lim_{n\to\infty} \langle f_n(\alpha(x)), \phi\rangle.\ )$$

Example 3.1 Using (3.5), one can show:

$$\delta(x^2 - a^2) = \frac{1}{2a}(\delta(x - a) + \delta(x + a)).$$

Value at the point. Let $f \in \mathcal{D}'(\mathbb{R}^n), x_0 \in \mathbb{R}^n$ and $y = \varepsilon x + x_0$ ($A = \varepsilon I$, $\varepsilon \in \mathbb{R}$, I is the identity matrix). If there is a constant C such that for

$$\lim_{\varepsilon\to 0}\langle f(x_0 + \varepsilon x), \phi(x)\rangle = \langle C, \phi\rangle, \ \phi \in \mathcal{C}_0^\infty(\mathbb{R}^n), \tag{3.6}$$

then it is said that$f$ has a value C at the point $x = x_0; \Lambda; f(x)_{x=x_0} = C$.

Note that (3.6) holds if and only if (3.6) holds for every sequence $\varepsilon_n \to 0,\ n \to \infty$.

The value at the point is an irregular operation. Delta distribution $\delta(x)$ does not have the value at $x_0 = 0$. Also the Heaviside function (distribution)

$$H(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}$$

does not have a value at $x_0 = 0$. But if $\tilde{f}(x)$ is a regular distribution determined by a continuous function $f(x)$ such that $f(x_0) = C$, then the corresponding regular distribution$\tilde{f}(x)$ has the value C at $x = x_0$. This is a consequence of the Lebesgue theorem.

Let $f(x, y) \in \mathcal{D}'(\mathbb{R}^n \times \mathbb{R}^m),\ x_0 \in \mathbb{R}^n$. If for every $\phi \in \mathcal{C}_0^\infty(\mathbb{R}^n)$ and every $\psi \in \mathcal{C}_0^\infty(\mathbb{R}^m)$ there exists $g(y) \in \mathcal{D}'(\mathbb{R}^m)$ such that

$$\lim_{\varepsilon \to 0} \langle f(x_0 + \varepsilon x, y), \phi(x)\psi(y) \rangle = \int_{\mathbb{R}^n} \phi(x) dx \langle g(y),\ \psi(y) \rangle,$$

then it is said that $g(y)$ is the value of $f(x, y)$ over the manifold $\{x_0\} \times \mathbb{R}^m$, $f(x, y)|_{x=x_0} = g(y)$.

Multiplication of a distribution by a smooth function. Let $a \in \mathcal{C}^\infty(\Omega)$. Then the mapping:

$$\mathcal{D}(\Omega) \to \mathcal{D}(\Omega),\ \phi \mapsto a\ \phi,$$

is continuous and the adjoined mapping $\mathcal{D}'(\Omega) \to \mathcal{D}'(\Omega)$:

$$\langle a(x)f(x), \phi(x) \rangle := \langle f(x), a(x)\phi(x) \rangle,\ \ \phi \in \mathcal{D}(\Omega),$$

is continuous with respect to weak, respectively strong, topologies in $\mathcal{D}'(\Omega)$.

If for example f is a measure, then we can assume weaker condition on a. We can assume that $a \in L^1_{loc}(\Omega)$ with respect to the measure f (cf. Example 3.3). Also, if f is a distribution of order m, then $a(x)f(x)$ is defined for $a \in \mathcal{C}^m(\Omega)$.

In general the product of two distributions is an irregular operation. It can be defined through the product of their regularized sequences or in case of "good position" of their wave fronts. This will not be explained in this short presentation.

Differentiation of distributions. Let $f \in \mathcal{D}'(\Omega)$. Then $\dfrac{\partial}{\partial x_i} f = f^{(e_i)}$ is defined by

$$\langle f^{(e_i)}(x), \phi(x) \rangle := -\langle f(x),\ \frac{\partial \phi(x)}{\partial x_i} \rangle,\ i = 1, ..., n. \tag{3.7}$$

Clearly, if $(\phi_\nu) \to 0$, then $\left(\frac{\partial \phi_\nu}{\partial x_i}\right) \to 0,\ \nu \to \infty,$ in $\mathcal{D}(\Omega)$. This means that with (3.7) is defined a mapping $\mathcal{D}'(\Omega) \to \mathcal{D}'(\Omega)$ i.e., every distribution has all partial derivatives:

$$\langle \partial^\alpha f, \phi \rangle = (-1)^{|\alpha|} \langle f, \phi^{(\alpha)} \rangle$$

and the order of differentiation does not change the value $\partial^\alpha f$ in $\mathcal{D}'(\Omega)$.

Thorem 3.1 *Let $\alpha \in \mathbb{N}_0$. Then the differentiation ∂^α is a continuous and linear mapping $\mathcal{D}'(\Omega) \to \mathcal{D}'(\Omega)$ with respect to the weak (strong) topology in $\mathcal{D}'(\Omega)$.*

Example 3.2 Let us show $H'(x) = \delta$. For every $\phi \in \mathcal{D}(\mathbb{R})$, we have

$$\langle H'(x), \phi(x) \rangle = -\langle H(x), \phi'(x) \rangle = -\int_0^\infty \phi'(x)dx = \phi(0) = \langle \delta(x), \phi(x) \rangle.$$

Example 3.3 We have

$$\langle \delta^{(j)}(x - x_0), \phi(x) \rangle = (-1)^{|i|} \phi(x_0), \quad \phi(x) \in \mathcal{C}_0^\infty(\mathbb{R}).$$

Example 3.4 If we apply Theorem 3.1 on a (classically) convergent series, after differentiation we obtain a series which converges in $\mathcal{D}'(\Omega)$ although it does not converge in the classical sense.

Let us show that the trigonometric series $\sum_{\nu=-\infty}^{\infty} b_\nu e^{i\nu x}, \qquad x \in \mathbb{R}$, converges in $\mathcal{D}'(\mathbb{R})$ if there exist $M > 0$ and $k \in \mathbb{N}_0$ such that

$$|b_\nu| \leq M|\nu|^k, \qquad \nu \neq 0. \tag{3.8}$$

We have

$$b_0 + \tilde{g}^{(p)} = \sum_{\nu=-\infty}^{\infty} b_\nu e^{i\nu x},\ x \in \mathbb{R},$$

where $p \in \mathbb{N}_0,\ p \geq k + 2$, and $g(x)$ is a continuous periodic function defined by

$$g(x) = \sum_{\substack{\nu=-\infty \\ \nu \neq 0}}^{\infty} \frac{b_\nu e^{i\nu x}}{(i\nu)^p}, \quad x \in \mathbb{R}^n. \tag{3.9}$$

($\tilde{g}$ denotes a regular distribution determined by g.) In fact from (3.8) it follows $\left|b_\nu/(i\nu)^p\right| \leq M/\nu^2,\ \nu \neq 0$. So, $g(x)$ is a continuous function and by differentiation (3.9) p-times, we have

$$b_0 + \tilde{g}^{(p)} = \sum_{\nu=-\infty}^{\infty} b_\nu e^{i\nu x}.$$

Example 3.5 One can easily prove the Leibnitz rule for $a(x)f(x),\ a \in \mathcal{C}^\infty(\Omega),\ f(x) \in \mathcal{D}'(\Omega)$:

$$\frac{\partial}{\partial x_j}(a(x)f(x)) = \frac{\partial}{\partial x_j}a(x)f(x) + a(x)\frac{\partial}{\partial x_j}f(x).$$

or for general $k \in \mathbb{N}^n$:

$$\partial^k(a(x)f(x)) = \sum_{|\alpha|\le k} \binom{k}{\alpha} a^{(\alpha)}(x) f^{(k-\alpha)}(x). \tag{3.10}$$

We will generalize (3.10): Let $p(y),\ y \in \mathbb{R}^n$, be a polynomial of order $m \in \mathbb{N}$, $p(y) = \sum\limits_{|\alpha|\le m} c_\alpha y^\alpha$ $(y^\alpha = y_1^{\alpha_1}...y_n^{\alpha_n} \in \mathbb{R})$. Then we put $p(\partial) = \sum\limits_{|\alpha|\le m} c_\alpha \partial^\alpha$ and have (Hörmander's formula) for $f \in \mathcal{D}'(\mathbb{R}^n)$ and $a \in \mathcal{C}_0^\infty(\mathbb{R}^n)$:

$$p(\partial)(af) = \sum_{|\alpha|\le m} \frac{1}{\alpha!}\partial^\alpha a\ p^{(\alpha)}(\partial)f.$$

Relations of classical and distributional derivatives. Let a function $f(x),\ x \in \mathbb{R}$, have the continuous derivative in the open set $\mathbb{R}\setminus\bigcup_{j=1}^k\{x_i\}, \qquad x_j \in \mathbb{R},\ \ j = 1, ..., k$. Assume that the jump of f at $x = x_j$ is $s_j = f(x_j+0) - f(x_j-0)$. Since f defines a regular distribution $\tilde{f}$ in $\mathcal{D}'(\mathbb{R})$, we have

$$\langle \frac{d}{dx}\tilde{f}, \phi\rangle = -\int_{-\infty}^{\infty} f(x)\phi'(x)dx = \sum_j \phi(x_j)s_j + \int_{-\infty}^{\infty} f'(x)\phi(x)dx,$$

where f' is a function equal to the ordinary derivative of $f(x)$ for $x \neq x_j,\ 1 \le j \le k$. This means

$$\frac{d}{dx}\tilde{f} = \tilde{f'} + \sum_j s_j\delta_{x_j}$$

where $\tilde{f'}$ is a distribution defined by f'.

Thorem 3.2 *If $f \in \mathcal{C}^p(\Omega)$, then the regular distribution $\tilde{f}$ satisfies:*

$$D^i\tilde{f} = (\tilde{D^i f}) \quad \textit{for}\ \ |i| \le p\ \ (\Omega \subset \mathbb{R}^n,\ i \in \mathbb{N}^n).$$

Example 3.6 Let $f \in \mathcal{C}^1(\overline{\Omega})$ where $\overline{\Omega} = \Omega \cup \partial\Omega$ is a closed bounded set with the boundary $S = \partial\Omega$ consisting of a finitely many closed $(n-1)$- dimensional surfaces of $\mathcal{C}^1$-class which do not intersect each other. Put $f = 0$ out of $\bar{\Omega}$. By partial integration we have

$$\langle \frac{\partial}{\partial x_j}\tilde{f}, \psi\rangle = -\int_\Omega f(x)\frac{\partial}{\partial x_j}\psi(x)dx =$$

$$= \int_S f(x)\psi(x)\cos(\nu, x_j)dS + \int_\Omega \frac{\partial f}{\partial x_j}\psi(x)dx, \quad \psi \in \mathcal{C}_0^\infty(\mathbb{R}^n),$$

where ν is the outer normal on S (on $\mathcal{C}^1-$ parts of S) and (ν, x_j) - is the angle between the positive axes x_j and the normal ν and dS is a part of the surface. This implies

$$\frac{\partial}{\partial x_j}\tilde{f} = \Big(\frac{\tilde{\partial f}}{\partial x_j}\Big) + s \ \text{ on } \ \mathbb{R}, \tag{3.11}$$

where s is a distribution defined by $(s, \psi) = \int_S f(x)\cos(\nu, x_j)\psi(x)dS$.

If $f \in \mathcal{C}^2(\Omega)\cap\mathcal{C}^1(\overline{\Omega})$ and $f = 0$ out of $\bar{\Omega}$ then (3.16) and $\dfrac{\partial}{\partial\nu} = \sum\limits_{j=1}^{n}\cos(x_j, \nu)\dfrac{\partial}{\partial x_j}$ imply Green's formula:

$$\langle\Delta\tilde{f}, \phi\rangle = \langle(\tilde{\Delta f}), \phi\rangle + \int_S \frac{\partial}{\partial\nu}f(x)\phi dS - \int_S f(x)\frac{\partial}{\partial\nu}\phi dS, \ \phi \in \mathcal{C}_0^\infty(\Omega), \tag{3.12}$$

where $\Delta = \sum\limits_{j=1}^{n}\dfrac{\partial^2}{\partial x_j^2}$ is the Laplace operator.

Note that in (3.11) and (3.12) $\Big(\frac{\tilde{\partial f}}{\partial x_j}\Big)$ i $(\tilde{\Delta f})$ denotes regular distributions determined by locally integrable functions which are respectively equal to the ordinary partial derivatives $\dfrac{\partial}{\partial x_j}f$ i Δf where they exist in the classical sense.

Connections of differentiations in the classical and distributional sense is given in the next theorem.

Thorem 3.3 *Let u and f be continuous functions in Ω and $\dfrac{\partial}{\partial x_j}u = f$ in the sense of distributions: $\dfrac{\partial}{\partial x_j}\tilde{u} = \tilde{f}$, where $\tilde{u}$ and $\tilde{f}$ are corresponding regular distributions. Then $\dfrac{\partial}{\partial x_j}u = f$ in the classical sense.*

Example 3.7. This example illustrates the difference between the classical and the distributional derivatives. Let f be the characteristic function for the interval $[0, 1]$. Then $u(x, t) = f(x - ct), \ x, t \in \mathbb{R}$, is a distributional (we call it the weak) solution to the differential equation

$$u_{tt} - c^2u_{xx} = 0.$$

Since $f \in L^1(\mathbb{R})$, Theorem 2.10 implies that there exists a sequence $(f_n) \in \mathcal{C}_0^\infty(\mathbb{R})^{\mathbb{N}}$ converging in L^1 to f. Thus

$$u_n(x, t) := f_n(x - ct) \to u(x, t) \quad \text{as} \ \ n \to \infty, \ \ n \in \mathbb{N} \ \ (x, t \in \mathbb{R}),$$

in the sense of distributions. By previous theorem, $f_n(x-ct),\ n\in\mathbb{N}$, is the classical and weak solution to the given equation, and this implies that $u(x,t)$ is a weak (but not classical) solution of this equation.

Thorem 3.4 *Let $f\in L^p(\Omega),\ 1\le p<\infty,\ \delta_n$ given in (1.2) and let $f^{(\alpha)}\in L^p(\Omega)$ in the sense of distributions.*

(i) If $y\in\Omega$ and $d(y,\mathbb{R}^n\setminus\Omega)>\dfrac{1}{k}$, then

$$(f^{(\alpha)}*\delta_k)(y)=(f*\delta_k)^{(\alpha)}(y).$$

(ii) If $\Omega_1\subset\Omega$ and $(\Omega_1)_\varepsilon\subset\Omega$ for some $\varepsilon>0$, then

$$\|(f*\delta_k)^{(\alpha)}-f^{(\alpha)}\|_{L^p(\Omega_1)}\to 0\quad as\ \ k\to\infty.$$

Example 3.8. Let Ω be open so that $\bar\Omega$ is as in Example 3.6. Let it be divided with a smooth $(n-1)$-dimensional surface Γ of C^1-class into two domains Ω_1 and Ω_2 such that $\Omega=\Omega_1\cup\Omega_2\cup\Gamma$.

If $u\in\mathcal{C}(\Omega)$ and $u|_{\Omega_k}\in\mathcal{C}^1(\Omega_k),\ k=1,2$, then u has the distributional partial derivatives (regular distributions) and

$$\frac{\partial}{\partial x_j}\tilde u=\sum_{k=1}^{2}\Big(\widetilde{\frac{\partial}{\partial x_j}(u|_{\Omega_k})}\Big). \tag{3.13}$$

Let $\phi\in\mathcal{C}_0^\infty(\Omega_k),\ k=1,2$. By Example 3.6

$$\int\limits_{\Omega_k}u(x)\frac{\partial}{\partial x_j}\phi(x)dx=-\int\limits_{\Omega_k}\Big(\frac{\partial}{\partial x_j}u(x)\Big)\phi(x)dx-\int\limits_{\Gamma}u(x)\phi(x)\cos(\nu,x_j)dr.$$

Since the direction of the normal ν on Γ is taken so that for $k=1,2$, it is directed out of Ω_k, we obtain

$$-\Big\langle\frac{\partial}{\partial x_j}u,\phi\Big\rangle=\Big\langle u,\frac{\partial}{\partial x_j}\phi\Big\rangle=\int\limits_{\Omega_1}u\frac{\partial}{\partial x_j}\phi\,dx+\int\limits_{\Omega_2}u\frac{\partial}{\partial x_j}\phi\,dx,$$

$$\langle\frac{\partial}{\partial x_j}u,\phi\rangle=\int\limits_{\Omega}v(x)\phi(x)dx,$$

where $v\in\mathcal{L}^\infty$ so that $v|_{\Omega_k}=(\dfrac{\partial}{\partial x_j}u)|_{\Omega_k}$), and this proves (3.13).

Structural theorem. One can easily prove that $f=C$ is the only distributional and classical solution to equation $f'=0$ on $\Omega=(a,b)$. Also we have if

$f \in \mathcal{D}'(\Omega)$, then for every $k \in \mathbb{N}_0$ there exists $g \in \mathcal{D}'(\Omega)$ such that $g^{(k)}(x) = f(x)$ and $g(x)$ is uniquely determined up to a polynomial of order $\leq k-1$.

Now we will give a structural theorem for distributions which shows that $\mathcal{D}'(\Omega)$ is the smallest extension of $L^\infty_{loc}(\Omega)$ where every element has all the derivatives.

Thorem 3.5 *Let $f(x) \in \mathcal{D}'(\Omega)$ and $\Omega_1 \subset\subset \Omega \subset \mathbb{R}^n$.*

(i) There exist $g(x) \in L^\infty(\Omega_1)$ and $r \in \mathbb{N}$ such that in the sense of distributions

$$f(x) = \frac{\partial^{n\cdot r}}{\partial x_1^r ... \partial x_n^r} g(x).$$

($\Omega_1 \subset\subset \Omega$ means theta $\overline{\Omega}_1$ is a compact subset of Ω).

(ii) For every neighborhood Ω_2 of $\overline{\Omega}_1$ in Ω there exist a continuous function F in $\mathbf{R}^n$ and $s \in \mathbb{N}$ so thatsupp $F \subset \Omega_2$ such that

$$f = \frac{\partial^{n\cdot s}}{(\partial x_1)^s ... (\partial x_n)^s} F(x).$$

4 Regularization

Division of distributions by smooth functions. Let $a \in \mathcal{C}^\infty(\Omega)$ and $g \in \mathcal{D}'(\Omega)$. Consider

$$a(x)f(x) = g(x) \text{ in } \mathcal{D}'(\Omega). \tag{4.1}$$

If $a(x)$ is different from zero everywhere in Ω, this is a trivial problem since $1/a \in \mathcal{C}^\infty(\Omega)$ and $f = \frac{1}{a}g$.

Let $n = 1$ and let the zeros of a be isolated and of finite order. This means that for every zero point x_0 ($a(x_0) = 0$) there exists $k \in \mathbb{N}$ such that $x \mapsto (x - x_0)^{-k} a(x_0)$ is finite in a neighborhood of x_0. So let $\{\alpha_k;\ k \in \mathbb{N}\}$ be isolated points of a; $\{U_i,\ i \in \mathbb{N}\}$ be a family of open sets which cover $\mathbb{R}$ such that every zero lies in at most one open set of this family. Denote by $\{\phi_i;\ i \in \mathbb{N}\}$ the partition of unity which corresponds to $\{U_i\}$. If a solution of (4.1) exists it is clear that

$$\sum_{i\in\mathbb{N}} (af, \phi_i\phi) = \sum_{i\in\mathbb{N}} (g, \phi_i\phi),\ \phi \in \mathcal{C}_0^\infty(\Omega),$$

i.e. that for every $i \in \mathbb{N}$ there exists a unique solution f_i in $\mathcal{D}'(U_i)$ to

$$a_i(x)f_i(x) = g_i(x) \quad \text{in } \Omega,$$

where a_i and g_i are the corresponding restrictions of of a and g on U_i. Conversely, if we know the existence of solution to $a_i f_i = g_i, i \in \mathbb{N}$, then using the above partition of unity we have that $\sum_{i\in\mathbb{N}} \phi_i g_i/a_i$ is a solution of (4.1).

Thus we will assume that a has only one zero of finite order. Moreover, assume that $x = 0 \in \Omega$ is a zero of order m: $a(x) = x^m b(x), x \in \Omega$, where $b \in \mathcal{C}^\infty(\Omega),\ b(x) \neq 0$ in Ω. After dividing by b this equation becomes

$$x^m f(x) = h(x),\ h \in \mathcal{D}'(\Omega). \tag{4.2}$$

Thorem 4.1 *Let $m \in \mathbb{N}$. (i) A necessary and sufficient condition that $f_0(x)$ satisfies $x^m f_0(x) = 0$ in $\mathcal{D}'(\Omega)$ is that*

$$f_0(x) = \sum_{\nu=0}^{m-1} C_\nu \delta^{(\nu)}(x),\ C_\nu \in \mathbb{C}. \tag{4.3}$$

(ii) Equation (4.2) has a solution in $\mathcal{D}'(\Omega)$.

Thus the solution of (4.2) if exists, is not unique: two solutions differ by a function of the form (4.3).

Let $H = \{\phi;$ there exists $\psi \in \mathcal{D}(\Omega), \phi = x^m \psi\}$. By

$$\langle f_1, \phi \rangle := \langle h, \psi \rangle,\ \ \phi = x^m \psi,\ \phi \in H,$$

is defined a continuous linear functional on H. By the Hahnn-Banach theorem (see the appendix) it follows that there exists a unique $\tilde{f} \in \mathcal{D}'(\Omega)$ such that $\tilde{f}_{|H} = f_1$. This $\tilde{f}$ satisfies (4.2).

Any other solution is of the form $\tilde{f}(x) + \sum_{\nu=0}^{m-1} C_\nu \delta^{(\nu)}$. If we consider (4.1) only on $\mathbb{R} \setminus \{0\}$, then the solution is unique:

$$f(x) = 1/x^m,\ x \in \mathbb{R} \setminus \{0\}.$$

This function does not have a locally integrable extension on $\mathbb{R}$. Note that there exists a distribution $\tilde{f}$ such that $f(x) = \tilde{f}(x),\ \ x \in \mathbb{R} \setminus \{0\} : \langle f, \phi \rangle = \langle \tilde{f}, \phi \rangle$ for every $\phi \in \mathcal{C}_0^\infty,\ \ \phi(x) = 0$ in a neighborhood of zero.

Thus we come to the definition of regularization:

Let $f \in L^1_{loc} \setminus \{0\}$. A distribution $\tilde{f} \in \mathcal{D}'(\Omega \setminus \{0\})$, which satisfies

$$\langle \tilde{f}(x), \phi(x) \rangle = \int\limits_{\mathbb{R}^n} f(x)\phi(x)dx,\ \phi \in \mathcal{C}_0^\infty(\Omega \setminus \{0\})$$

($\phi(x) = 0$ in a neighborhood of x_0) is called the regularization of a function f (pseudo function).

A regularization of $f \in L^1_{loc}(\Omega \setminus \{x_0\})$ is a distribution in $\mathcal{D}'(\Omega)$ such that its restriction on $\Omega \setminus \{x_0\}$ equals regular distribution on $\Omega \setminus \{x_0\}$ determined by f, if it exists. Theorem 4.1 implies that the regularization is not unique, if exists.

By partition of unity, we reduce the problem of regularization for a function with discrete set of singularities to one point singularity.

Regularization is not always possible:

Example 4.1 Let $h \in L^1_{loc}(\mathbb{R} \setminus \{0\})$ such that $h(x) \geq H(|x|),\ x \in \mathbb{R} \setminus \{0\}$ where $H(|x|)$ increases faster than any power of $\dfrac{1}{|x|}$ as $|x| \to 0$. Then there does not exist the regularization of h. For example $h(x) = e^{\frac{1}{x}} \in \mathcal{D}'((0,\infty))$ does not have an extension on $\mathbf{R}$.

Thorem 4.2 *If $f \in L^1_{loc}(\Omega \setminus \{0\})$ and for some $m \in \mathbb{N}$, $f|x|^m \in L^1_{loc}(\Omega)$, then f has a regularization in Ω :*

$$\langle \tilde{f}(x), \phi(x)\rangle := \int\limits_{\Omega} f(x)\Big[\phi(x) - (\phi(0) + \sum_{i=1}^{m-1} \frac{1}{i!} d^i\phi(0))H(\varepsilon - |x|)\Big]dx,$$

$\phi \in \mathcal{C}^\infty_0(\Omega)$, *where*

$$d^i\phi(0) = \Big(x_1\frac{\partial}{\partial x_1} + ... + x_n\frac{\partial}{\partial x_n}\Big)^i \phi(x)|_{x=0},$$

$$H(\varepsilon - |x|) = \begin{cases} 1 & za\ \ |x| < \varepsilon \\ 0 & za\ \ |x| \geq \varepsilon, \end{cases}$$

ε is chosen so that $Z(0,\varepsilon) \subset \Omega$.

Regularizations by analytic continuation. Let Λ be an open set in $\mathbb{R}^n$ and $t \mapsto h_t(x)$ be a mapping $\Lambda \to \mathcal{D}'(\Omega)$ (a family of distributions). For every $\phi \in \mathcal{C}^\infty_0(\Omega)$, by $t \mapsto \langle h_t(x), \phi(x)\rangle$ is defined a function $\Lambda \to \mathbb{C}$.

The mapping $\Lambda \to \mathcal{D}'(\Omega),\ t \mapsto h_t(x)$ has a limit as $t \mapsto t_0$ ($t_0 \in \overline{\Lambda},\ \ t \in \Lambda$) if there exists $g \in \mathcal{D}'(\Omega)$ such that $h_t \to g$ in the sense of strong topology in $\mathcal{D}'(\Omega)$. Since the strong and the weak sequential convergences are equivalent, this is equivalent with

$$\lim_{t \to t_0} \langle h_t(x), \phi(x)\rangle = \langle g, \phi\rangle,\ \phi \in \mathcal{C}^\infty_0(\Omega).$$

Then we say that h_t has a limit g at t_0.

A family h_t is continuous at the point $t_0 \in \Lambda$ if $g = h(t_0)$. (Again this is equivalent with:

$$\langle h_t, \phi\rangle \to \langle h_{t_0}, \phi\rangle,\ \phi \in \mathcal{C}^\infty_0(\Omega),\ \ \text{as}\ \ t \to t_0\ \ \text{in}\ \ \Lambda.$$

In a similar way we introduce the differentiation of $t \mapsto h_t$ and, as above, we have that

"$\partial^k_t h_t$ exists in $\mathcal{D}'(\Omega)$ if and only if for every $\phi \in \mathcal{D}(\Omega)$ there exist the $k-$th derivative of the function $\Lambda \ni t \mapsto \langle h_t, \phi\rangle \in \mathbb{C}$."

Thorem 4.3 *a) Let $f(x) \in \mathcal{D}'(\mathbb{R}^n)$, then $\{f(x-h);\ h \in \mathbb{R}^n\}$ is a continuous family of distributions and*

$$f^{(e_j)}(x) = \lim_{r\to 0} \frac{f(x+re_j)-f(x)}{r},$$

in the sense of convergence in $\mathcal{D}'(\mathbb{R}^n)$.

b) If a family h_t has a partial derivative ∂_{t_j} at t_0, then for every $m \in \mathbb{N}_0^n$:

$$\frac{\partial}{\partial t_j}\Big|_{t=t_0}(\partial_x^m h_t) = \partial_x^m\Big(\frac{\partial}{\partial t_j}\Big|_{t=t_0} h_t\Big).$$

In the case that Λ is an open set in $\mathbb{C}$ and that for every $\phi \in \mathcal{C}_0^\infty(\Omega)$

$$\lambda \mapsto \langle h_\lambda(x), \phi(x)\rangle$$

is an analytic function in Λ, then we say that $h_\lambda(x)$ is an anlytic family or an analytic distribution (with respect to λ.)

Thorem 4.4 *(i) A family $\{h_\lambda(x);\ \lambda \in \Lambda\} \subset \mathbb{C}$ is analytic if and only if for every $\lambda_0 \in \Lambda$*

$$\lim_{\lambda\to\lambda_0} \frac{h_\lambda - h_{\lambda_0}}{\lambda-\lambda_0} \text{ exists in } \mathcal{D}'(\Omega).$$

(ii) For every $m \in \mathbb{N}^n$ and $k \in \mathbb{N}$

$$\frac{d^k}{d\lambda^k}(\partial_x^m h_\lambda) = \partial_x^m\Big(\frac{d^k}{d\lambda^k} h_\lambda\Big), \quad \lambda \in \Lambda.$$

(iii) For every $\lambda_0 \in \Lambda$ and λ in an open neighbourhood of λ_0 there holds

$$h_\lambda = \sum_{k=0}^{\infty} \frac{(\lambda-\lambda_0)^k}{k!}\Big(\frac{d^k}{d\lambda^k} h_\lambda\Big)\Big|_{\lambda=\lambda_0} \text{ (Taylor expansion).}$$

Let $f_\lambda \in \mathcal{D}'(\Omega), \quad \lambda \in \Lambda$, be an analytic distribution and let for every $\phi \in \mathcal{C}_0^\infty(\Omega)$ the analytic function $\langle f_\lambda, \phi\rangle$ have an analytic continuation on an open set $\Lambda_1 \supset \Lambda$. Then:

Thorem 4.5 *For every $\lambda_1 \in \Lambda_1, \quad f_{\lambda_1} \in \mathcal{D}'(\Omega)$, where $\langle f_{\lambda_1}, \phi\rangle$ is the analytic continuation of $\lambda \mapsto \langle f_\lambda, \phi\rangle$ in the point $\lambda_1 = \lambda$ (for every $\phi \in \mathcal{C}_0^\infty(\Omega)$).*

Now we will explain the regularization through the analytic continuation by examples.

Example 4.2. Let x_+^λ be defined on $\mathbb{R}$ by

$$x_+^\lambda = \begin{cases} x^\lambda & x > 0 \\ 0 & x \leq 0. \end{cases}$$

For $Re\lambda > -1,\ x_+^\lambda \in L^1_{loc}(\mathbb{R})$ defines a regular distribution

$$\langle x_+^\lambda, \phi \rangle = \int_0^\infty x^\lambda \phi(x) dx. \tag{4.4}$$

By Thorem 4.4:

$$\frac{d}{d\lambda}\langle x_+^\lambda, \phi(x)\rangle = \langle \frac{d}{d\lambda} x_+^\lambda, \phi\rangle = \int_0^\infty x^\lambda \ell n x \phi(x) dx.$$

Thus, $x_+^\lambda$is an analytic distribution on $\{\lambda \in \mathbb{C};\ Re\lambda > -1\}$. If $\langle x_+^\lambda, \phi\rangle$ in (4.4) is written in the form

$$\int_0^1 x^\lambda(\phi(x) - \phi(0))dx + \int_1^\infty x^\lambda \phi(x) dx + \frac{\phi(0)}{\lambda + 1}, \tag{4.5}$$

(since $\frac{\phi(0)}{\lambda+1} = \int_0^1 \phi(0) x^\lambda dx$) it follows that the first member is an analytic function for $Re\lambda > -2$, that the second member is an analytic function for every $\lambda \in \mathbb{C}$ and that the third member is an analytic function for $\lambda \neq -1$. This means that $x_+^\lambda,\ Re\lambda > -1$ is by (4.5) analytically continued to $Re\lambda > -2,\ \lambda \neq -1$. Continuing this procedure, we get x_+^λ is analytically continued by

$$\langle x_+^\lambda, \phi\rangle = \int_0^\infty x^\lambda \phi(x) dx = \int_0^1 x^\lambda [\phi(x) - \phi(0) - ...$$

$$-... - \frac{x^{n-1}}{(n-1)!}\phi^{(n-1)}(0)]dx + \int_1^\infty x^\lambda \phi(x) dx$$

$$+ \sum_{k=1}^n \frac{\phi^{(k-1)}(0)}{(k-1)!(\lambda + k)}$$

into the domain $Re\lambda > -(n+1),\ \lambda \neq -1, -2, ..., -n$, where we use

$$\sum_{k=1}^n \frac{\phi^{(k-1)}(0)}{(k-1)!(\lambda+k)} = \int_0^1 x^\lambda(\phi(0) + x\phi'(0) + ... + \frac{x^{n-1}}{(n-1)!}\phi^{(n-1)}(0))dx.$$

We can show

$$(x_+^\lambda)' = \lambda x_+^{\lambda-1} \qquad \lambda \neq -1, -2, ...$$

For $Re\lambda > 0$ it is clear and from the unicity of the analytic continuation, it follows for $\lambda \neq -1, -2, ...$

Note that if $-\lambda \in \mathbb{N}$ then the regularization is made by the use of Taylor expansion.

Using the regularization of x_+^λ we can formulate the regularization for x_-^λ, $-\lambda \notin \mathbb{N}$, where

$$x_-^\lambda = \begin{cases} (-x)^\lambda & x < 0 \\ 0 & x \geq 0. \end{cases}$$

For $Re\lambda > -(n+1)$, $\lambda \neq -1, -2, ..., -n$:

$$\langle x_-^\lambda, \phi \rangle = \int_0^1 x^\lambda [\phi(-x) - \phi(0) - ... - (-1)^{n-1} \frac{x^{n-1}}{(n-1)!} \phi^{(n-1)}(0)] dx$$

$$+ \int_1^\infty x^\lambda \phi(-x) dx + \sum_{k=1}^n \frac{(-1)^{k-1} \phi^{(k-1)}(0)}{(k-1)!(\lambda+k)}.$$

Example 4.3 Let $|x|^\lambda = x_+^\lambda + x_-^\lambda$ i $|x|^\lambda sgnx = x_+^\lambda - x_-^\lambda, Re\lambda > -1$. Then

$$\langle |x|^\lambda, \phi(x) \rangle = \int_0^1 x^\lambda [\phi(x) + \phi(-x) - 2(\phi(0) + \frac{x^2}{2!}\phi''(0) + ...$$

$$+ \frac{x^{2m-2}}{(2m-2)!} \phi^{(2m-2)}(0))] dx + \int_1^\infty x^\lambda [\phi(x) + \phi(-x)] dx \tag{4.6}$$

$$+ 2 \sum_{k=0}^{m-1} \frac{\phi^{(2k)}(0)}{(2k)!(\lambda+2k+1)}$$

for $Re\lambda > -(2m+1)$ i $\lambda \neq -1, -3, ..., -(2m-1)$.

$$\langle |x|^\lambda sgnx, \phi \rangle = \int_0^1 x^\lambda [\phi(x) - \phi(-x) - 2(x\phi'(0)$$

$$+ \frac{x^3}{3!}\phi^{(3)}(0) + ... + \frac{x^{2m-1}}{(2m-1)!} \phi^{(2m-1)}(0))] dx \tag{4.7}$$

$$+ \int_1^\infty x^\lambda [\phi(x) - \phi(-x)] dx + 2 \sum_{k=0}^{m-1} \frac{\phi^{(2k)}(0)}{(2k+1)!(\lambda+2k+2)}$$

for $Re\lambda > -(2m+2)$, $\lambda \neq -2, -4, ..., -2m$.

Since even negative numbers are in the domain of analytic continuation for $|x|^\lambda$ and $|x|^{-2k} = x^{-2k}$, from (4.6) we get regularizations for $x^{-2k},\ k \in \mathbb{N}$. Similarly (4.7) implies regularizations for $x^{-(2k+1)},\ \ k \in \mathbb{N}$.

One can see that $x^{-(2k+1)}$ or x^{-2k} solutions to (4.2) if m is even or odd.

Example 4.4 Let $x \in \mathbb{R}^n$ i $f(x) = |x|^\lambda = r^\lambda$. If $Re\lambda > -n$ then f defines a regular distribution

$$\langle r^\lambda, \phi\rangle = \int\limits_{\mathbb{R}^n} t^\lambda \phi(x)dx.$$

By differentiation with respect to λ for $\lambda > -n$, we can enter into the integral. It follows that $|x|^\lambda$, $Re\lambda > -n$, defines an analytic distribution. Now we will make the analytic continuation. Put $x = r\ \dot{x}$ where $\dot{x}$ is a variable on the unite sphere S^{n-1}. We have

$$\begin{aligned}\int\limits_{\mathbb{R}^n} |x|^\lambda \phi(x)dx &= \int\limits_0^\infty \int\limits_{S^{n-1}} r^\lambda \phi(r\ \dot{x}) r^{n-1} dr d\ \dot{x} \\ &= \int\limits_0^\infty r^{\lambda+n-1} \int\limits_{S^{n-1}} \phi(r\ \dot{x}) d\ \dot{x}\ .\end{aligned}$$

The integral $\dfrac{1}{|S^{n-1}|} \int\limits_{S^{n-1}} \phi(r\ \dot{x}) d\ \dot{x}$, where $|S^{n-1}|$ is the surface area of the sphere S^{n-1}, called the average (mean value) of ϕ on the sphere with the radius r, is denoted by $S_\phi(r)$. Thus

$$\int\limits_{\mathbb{R}^n} |x|^\lambda \phi(x)dx = |S^{n-1}| \int\limits_0^\infty r^{\lambda+n-1} S_\phi(r)dr.$$

Since the support of ϕ is contained in a sphere of a bounded radius, the function $r \mapsto S_\phi(r)$ is bounded. By definition, $S_\phi(r)$ is a smooth function of r if $r > 0$. Taylor expansion of $\phi(x)$ gives

$$\begin{aligned}S_\phi(r) = \frac{1}{|D^{n-1}|} \int\limits_{S^{n-1}} \Big[\phi(0) + \sum_{j=1}^n \frac{\partial\phi(0)}{\partial x_j} x_j \\ + \frac{1}{2} \sum_{i,j=1}^n \frac{\partial^2\phi(0)}{\partial x_i \partial x_j} x_i x_j + ... + \frac{1}{k!} d^k \phi(\xi)\Big] dx,\end{aligned}$$

where $x_j = rw_j$ and w_j is a product of corresponding "sin" and "cos" functions, $1 \le j \le n$. Integration of addends containing an even number of factors, say $2m$, x_j gives the addend of the form $a_m r^{2m}$. By integration of addends containing an odd number of factors x_j, one gets the value zero.

So we obtain $S_\phi(r) = \phi(0) + a_1 r^2 + a_2 r^4 + ... + a_k r^{2k} + o(r^{2k})$, and this implies that S_ϕ has all the derivatives at zero. Thus $S_\phi(r) \in \mathcal{C}_0^\infty(\mathbb{R})$ and for $Re\lambda > -n$,

$$\langle |x|^\lambda, \phi \rangle = \langle |S^{n-1}|\ r_+^{\lambda+n-1},\ \ S_\phi(r) \rangle, \tag{4.8}$$

where the right hand side $|S^{n-1}| r_+^{\lambda+n-1}$ is applied to the test function in $\mathcal{C}_0^\infty(\mathbb{R})$. By Example 4.2 it follows that (4.8) can analytically be continued in the complex plane without points $\lambda = -n, -n-1, ...$.

References

[1] J. Baros-Neto, An introduction to the theory of distributions, Marcel Dekker, Inc., New York (1973).

[2] E. J. Beltrami, M. R. Wohlers, Distributions and the Boundary Values of Analytic Functions, Acad. Press, New York (1966).

[3] H. Bremerman, Distributions, Complex Variables and Fourier Transforms, Addison-Wesley, Reading, Mass. (1965).

[4] W. F. Donoghue, Distributions and Fourier Transforms, Acad. Press, New York (1969).

[5] R. E. Edwards, Functional Analysis Theory and Applications, Holt, Rinehart and Winston, New York (1965).

[6] A. Friedman, Generalized Functions and Partial Differential Equations, Prentice-Hall, Inc., Englwood Cliffs (1963).

[7] I. M. Gel'fand, G. E. Shilov, Generalized Functions-Properties and operations, Vol. 1, Acad. Press, New York (1964).

[8] I. M. Gel'fand, G. E. Shilov, Generalized Functions-Spaces of Fundamental and Generalized Functions, Vol. 2, Acad. Press, New York (1968).

[9] I. M. Gel'fand, G. E. Shilov, Generalized Functions-Theory of Differential Equations, Vol. 3, Acad. Press, New York (1967).

[10] I. M. Gel'fand, N. Ya. Vilenkin, Generalized Functions - Applications of Harmonic Analysis, Acad. Press, New York, (1964).

[11] I. M. Gel'fand, M. I. Graev, N. Ya. Vilenkin, Generalized Functions - Integral Geometry and Representation Theory, Vol. 5, New York, (1966).

[12] L. Hörmander, Linear Partial Differential Operators, Springer, Berlin, (1963).

[13] J. Horvath, Topological Vector Spaces and Distributions, Vol. I, Addison-Wesley, Reading, Mass. (1966).

[14] L. Schwartz, Théorie des distributions, 1, 2 vols., Hermann, Paris (1950-1951).

[15] S. L. Sobolev, Méthode nouvelle á resoudre probleme de Cauchy pour les équations linéaires hyperboliques normales, Mat. Sb., 1 (43) (1935), 39-72.

[16] F. Treves, Topological Vector Spaces, Distributions and Kernels, Acad. Press, New York (1967).

[17] F. Treves, Basic Linear Partial Differential Equations, Acad. Press, New York 1975.

[18] V. S. Vladimirov, Generalized Functions in Mathematical Physics, Mir, Moscow (1979).

PART II

Applications of Fractional Calculus in Mechanics

Teodor M. Atanacković, Stevan Pilipović
University of Novi Sad

1 Some basic properties of fractional integrals and derivatives

The story of fractional integrals and derivatives started with Leibnitz who in his letters to L'Hôspital[1] in 1695 and Wallis in 1697 made remarks on the possibility of calculating $d^n y/dx^n$ when n is $\frac{1}{2}$. There are many possible generalizations of a notion of a derivative of a function that would lead to the answer of the question: what is $d^n y/dx^n$ when n is $\frac{1}{2}$? We start from the Cauchy formula for $n-$fold primitive of a function $f(t)$ given as

$$J^n f(t) = \frac{1}{(n-1)!}\int_0^t (t-\tau)^{n-1} f(\tau)\, d\tau,\ t>0, \quad n\in\mathbb{N} \tag{1.1}$$

where it was assumed that $f(t)=0$, for $t<0$. By noting that $(n-1)! = \Gamma(n)$, where $\Gamma(n)$ is the Euler gamma function, the fractional integral of order α is defined as

$$J^\alpha f(t) = \frac{1}{\Gamma(\alpha)}\int_0^t (t-\tau)^{\alpha-1} f(\tau)\, d\tau,\ t>0, \quad \alpha\in\mathbb{R}^+ \tag{1.2}$$

This is the Reimann-Liouville fractional integral. Note that J^α satisfies the semigroup property $J^\alpha J^\beta = J^{\alpha+\beta}, \alpha,\beta\geq 0$. Thus the fractional integral is commutative $J^\alpha J^\beta = J^\beta J^\alpha$. Introducing the function

$$\Phi_\alpha(t) = \frac{t^{\alpha-1}}{\Gamma(\alpha)},\ t>0, \quad \alpha>0, \tag{1.3}$$

the integral (1.2) may be written in the form of convolution as

$$J^\alpha(t) = \Phi_\alpha(t) * f(t) = \int_0^t \Phi_\alpha(t-\tau) f(\tau)\, d\tau. \tag{1.4}$$

Let $\mathcal{L}(f)(s) = \int_0^\infty e^{-st} f(t)\, dt = \widehat{f}(s)$, for $s\in\mathbb{C}$ be the Laplace transform of f. Then, from the well known property of the Laplace transform of convolution, and $\mathcal{L}\left(\frac{t^{\alpha-1}}{\Gamma(\alpha)}\right)(s) = 1/s^\alpha$, we have

$$\mathcal{L}(J^\alpha f)(s) = \frac{\widehat{f}(s)}{s^\alpha}. \tag{1.5}$$

The fractional derivative of the order α may be defined as the left inverse of J^α. Formally suppose that $\alpha\in\mathbb{R}^+$ and let m be an integer such that $m-1<\alpha<m$. Then the Reimann-Liouville fractional derivative of the order α is defined as

$$D^\alpha f(t) = \begin{cases} \frac{d^m}{dt^m}\left[\frac{1}{\Gamma(m-\alpha)}\int_0^t \frac{f(\tau)}{(t-\tau)^{\alpha+1-m}}d\tau\right], & m-1<\alpha<m \\ \frac{d^m}{dt^m} f(t) & \alpha=m \end{cases} \tag{1.6}$$

[1] Note that l'Hospital's name is commonly seen spelled both "l'Hospital" and "l'Hôpital", the two being equivalent in French spelling.

For example by using (1.2) and (1.6) we obtain

$$\begin{aligned} J^{\alpha}t^{\beta-1} &= \frac{\Gamma(\beta)}{\Gamma(\gamma+\alpha)}t^{\beta-1+\alpha}, \\ D^{\alpha}t^{\beta-1} &= \frac{\Gamma(\beta)}{\Gamma(\beta-\alpha)}t^{\beta-\alpha-1}, \qquad \alpha>0,\ t>0. \end{aligned} \tag{1.7}$$

Note also that the $\alpha-th$ derivative of a constant function $C=const.$ is not zero

$$D^{\alpha}C=\frac{Ct^{-\alpha}}{\Gamma(1-\alpha)},\ \alpha\geq 0, \quad t>0, \tag{1.8}$$

Of course if $\alpha\in\mathbb{N}$ then $D^{\alpha}C=0$ due to the poles of gamma function at the points $0,-1,-2,...$ In our application we shall use in most cases (1.7) with $m=1$.

Note also that

$$D^{\alpha}f(t)\equiv 0, \quad \text{if} \quad f(t)=\frac{1}{t^{1-\alpha}}. \tag{1.9}$$

Thus, the function $t^{\alpha-1}$ plays the same role for fractional derivative $D^{\alpha}f$ as a constant in usual differentiation.

The definition (1.6) is just a special case of a more general definition. Thus, suppose that $f(t)$ is an absolutely continuous function in the interval $t\in[a,b]$ and suppose that $0<\alpha<1$. Then

$$\begin{aligned} D_{a+}^{\alpha}f(t) &= \frac{1}{\Gamma(1-\alpha)}\frac{d}{dt}\int_a^t\frac{f(\tau)}{(t-\tau)^{\alpha}}d\tau, \\ D_{b-}^{\alpha}f(t) &= -\frac{1}{\Gamma(1-\alpha)}\frac{d}{dt}\int_t^b\frac{f(\tau)}{(t-\tau)^{\alpha}}d\tau, \end{aligned} \tag{1.10}$$

are called left an right handed fractional derivatives of the order α, respectively. To determine the Laplace transform of a fractional derivative we may use definition so that for $a=0, 0<\alpha<1$

$$\mathcal{L}\left(D^{\alpha}f\right)(s)=s^{\alpha}\widehat{f}(s)-\left[\frac{1}{\Gamma(1-\alpha)}\int_0^t\frac{f(\tau)}{(t-\tau)^{\alpha}}d\tau\right]_{t=0^+}. \tag{1.11}$$

If the function is locally integrable the term in brackets vanishes, so that

$$\mathcal{L}\left(D^{\alpha}f\right)(s)=s^{\alpha}\widehat{f}(s). \tag{1.12}$$

The relation (1.12) could be used to define fractional derivative. We mention the law of exponents valid for integer order integrals and derivatives

$$\begin{aligned} J^nJ^mf(t) &= J^mJ^nf(t)=J^{m+n}f(t), \\ D^nD^mf(t) &= D^mD^nf(t)=D^{m+n}f(t). \end{aligned} \tag{1.13}$$

It could be shown that for fractional derivatives, in general

$$D^{\alpha}D^{\beta}f(t)\neq D^{\beta}D^{\alpha}f(t)\neq D^{\alpha+\beta}f(f). \tag{1.14}$$

There are special classes of functions for which the law of exponents holds. The Leibnitz rule for fractional derivatives does not hold. It could be shown that[2]

$$D^{\alpha}\left[f(t)g(t)\right]=\sum_{k=0}^{\infty}\left(\begin{array}{c}\alpha\\ k\end{array}\right)\left[D^{k}g(t)\right]\left[D^{\alpha-k}f(t)\right],\quad \alpha>0. \tag{1.15}$$

There is another important property known as integration by part formula (used in formulating variational principles for fractional order differential equations. It states (see [24] p. 46)

$$\int_{a}^{b}z(t)D_{a+}^{\gamma}y(t)dt=\int_{a}^{b}y(t)D_{b-}^{\gamma}z(t)dt. \tag{1.16}$$

We note that for validity of (1.16) no restrictions on the values of $y(t)$ and $z(t)$ at the end points $t=a$ and $t=b$ are imposed. Finally we state another property of fractional derivative, $0<\alpha<1$, of a function $f(t)$ as

$$\begin{aligned}\frac{d^{\alpha}}{dt^{\alpha}}f(t) &= f^{(\alpha)}\equiv\frac{d}{dt}\frac{1}{\Gamma(1-\alpha)}\int_{0}^{t}\frac{f(\tau)d\tau}{(t-\tau)^{\alpha}}=\frac{d}{dt}\frac{1}{\Gamma(1-\alpha)}\int_{0}^{t}\frac{f(t-\tau)d\tau}{\tau^{\alpha}}\\ &= \frac{f(0)t^{-\alpha}}{\Gamma(1-\alpha)}+\frac{1}{\Gamma(1-\alpha)}\int_{0}^{t}\frac{f^{(1)}(t-\tau)d\tau}{\tau^{\alpha}}\\ &= \sum_{n=0}^{\infty}\binom{\alpha}{n}\frac{t^{n-\alpha}}{\Gamma(n+1-\alpha)}f^{(n)}(t),\end{aligned} \tag{1.17}$$

where in $(1.17)_4$ the binomial coefficients are $\binom{\alpha}{n}=\frac{(-1)^{n-1}\alpha\Gamma(n-\alpha)}{\Gamma(1-\alpha)\Gamma(n+1)}=\frac{(-1)^{n}(-\alpha)_{n}}{n!}$ and $(z)_{n}=z(z+1)...(z+n-1),\quad n=1,2,...,\ (z)_{0}\equiv 1$. From $(1.17)_4$ it is obvious that any constitutive equation that takes into account $\alpha-th$ derivative of a function, takes, all integral derivatives at a fixed time t, each one with the weighting factor equal to $\binom{\alpha}{n}\frac{t^{n-\alpha}}{\Gamma(n+1-\alpha)}$.

There is another important definition of the fractional derivative, due to Caputo [8] and Caputo and Mainardi [9]. Thus, the Caputo derivative $D_{*}^{\alpha}f(t)$ of a function $f(t)$, of the order α, with $m-1<\alpha<m$ is

$$D_{*}^{\alpha}f(t)=\left\{\begin{array}{ll}\left[\frac{1}{\Gamma(m-\alpha)}\int_{0}^{t}\frac{f^{(m)}(\tau)}{(t-\tau)^{\alpha+1-m}}d\tau\right], & m-1<\alpha<m\\ \frac{d^{m}}{dt^{m}}f(t) & \alpha=m\end{array}\right. . \tag{1.18}$$

[2]Note the apparent lack of symmetry in the derivatives of the two functions. The left hand side of (1.15) does not depend on the order of the functions $f(t)$ and $g(t)$, while on the right hand side there are only *integer* derivatives of $g(t)$ and *noninteger* derivatives(integrals) of $f(t)$! It could be shown that the two functions $f(t)$ and $g(t)$ can be interchanged without changing the value of the fractional derivative.

We shall need another definition of a fractional derivative. It is introduced in the form of fractional order differences. Thus, the Grünwald-Letnikov fractional derivative of the order α is defined as

$$f_{a+}^{\alpha)}(t)=\lim_{h\to 0}\frac{1}{h^{\alpha}}\sum_{j=0}^{\left[\frac{t=a}{h}\right]}(-1)^{j}\binom{\alpha}{j}f(t-jh),\quad x>a \tag{1.19}$$

It could be shown that $f_{a+}^{\alpha)}(t)$ exists for certain class of function and that it is equal to Reimann-Liouville fractional derivative and Marchaud fractional derivative and reads (see [24] p. 386)

$$\begin{aligned} f_{a+}^{\alpha)}(t) &= f^{(\alpha)}(t)=\frac{f(t)}{\Gamma(1-\alpha)(t-a)^{\alpha}}+\frac{\alpha}{\Gamma(1-\alpha)}\int_{a}^{t}\frac{f(t)-f(\tau)}{(t-\tau)^{1+\alpha}}d\tau, \\ 0 &< \alpha<1. \end{aligned} \tag{1.20}$$

1.1 Fermat's theorem for Riemann-Liouville and Caputo fractional derivative

Starting from ([24]) p. 111, we have

$$\begin{aligned} ({}_0D_t^{\alpha}y)(t) &= \frac{d}{dt}\frac{1}{\Gamma(1-\alpha)}\int_0^t\frac{y(\tau)}{(t-\tau)^{\alpha}}d\tau=\frac{d}{dt}\frac{1}{\Gamma(1-\alpha)}\int_0^t\frac{y(t-\tau)}{\tau^{\alpha}}d\tau \\ &= \frac{1}{\Gamma(1-\alpha)}\frac{y(a)}{t^{\alpha}}+\frac{1}{\Gamma(1-\alpha)}\int_0^t\frac{y^{(1)}(t-\tau)}{\tau^{\alpha}}d\tau \\ &= \frac{1}{\Gamma(1-\alpha)}\frac{y(0)}{t^{\alpha}} \\ &\quad +\frac{1}{\Gamma(1-\alpha)}\int_0^t y^{(1)}(t-\tau)\left(\alpha\int_{\tau}^{t}\xi^{-1-\alpha}d\xi+\frac{1}{t^{\alpha}}\right)d\tau \\ &= \frac{y(t)}{\Gamma(1-\alpha)t^{\alpha}}+\frac{\alpha}{\Gamma(1-\alpha)}\int_0^t\frac{y(t)-y(t-\tau)}{\tau^{1+\alpha}}d\tau \\ &= \frac{y(t)}{\Gamma(1-\alpha)t^{\alpha}}+\frac{\alpha}{\Gamma(1-\alpha)}\int_0^t\frac{y(t)-y(t-\tau)}{\tau^{1+\alpha}}d\tau,\quad t>0,\ \ 0<\alpha<1. \end{aligned} \tag{1.21}$$

Similarly for Caputo derivative, we have

$$\begin{aligned} ({}_0^{*}D_t^{\alpha}y)(t) &= ({}_0D_t^{\alpha}y)(t)-\frac{y(0)}{\Gamma(1-\alpha)t^{\alpha}} = \frac{y(t)-y(0)}{\Gamma(1-\alpha)t^{\alpha}} \\ &\quad +\frac{\alpha}{\Gamma(1-\alpha)}\int_0^t\frac{y(t)-y(x)}{x^{1+\alpha}}dx,\quad t>0,\ \ 0<\alpha<1. \end{aligned} \tag{1.22}$$

Now suppose that $y(t)$ is an increasing positive function with a maximum at $t^* \in (0,t)$. Then $y(t^*) - y(x) \geq 0, x \in [0,t^*]$. From (1.22) we conclude that

$$\begin{aligned} ({}_0D_t^{\alpha} y)(t^*) &\geq \frac{y(t^*)}{\Gamma(1-\alpha)(t^*)^{\alpha}} > 0, \\ ({}_0^*D_t^{\alpha} y)(t^*) &\geq \frac{y(t^*) - y(0)}{\Gamma(1-\alpha)(t^*)^{\alpha}} > 0. \end{aligned} \tag{1.23}$$

The above results may be used to prove the following theorem:

Theorem Suppose that an integrable on $[A,B]$ function $u(t)$ attains a maximum at a point $x^* \in [A,B]$ and there exist $\delta > 0$ such that $u(t)$ satisfies Hölder condition with the exponent $h > \alpha$ in the interval $\{t : x^* - \delta \leq t \leq x^*\}$. Then, for any $\alpha \in [0,1]$ and $a \neq x^*$ we have

$$\begin{aligned} ({}_aD_{x^*}^{\alpha} y)(t) &\geq \frac{u(x^*)}{\Gamma(1-\alpha)(x^*)^{\alpha}}, \\ ({}_0^*D_t^{\alpha} y)(t^*) &\geq \frac{u(x^*) - u(0)}{\Gamma(1-\alpha)(x^*)^{\alpha}}. \end{aligned} \tag{1.24}$$

Thus, at a maximum, fractional derivatives either satisfy (1.24) or do not exist.

It could be easily shown that for a minimum of a function, the inequalities in (1.24) reverse sign.

1.2 An Expansion formula

It is well known that for the analytic function $f(t)$ (see [24], p. 35 and p.278) the left fractional derivative$({}_0D_t^{\alpha} f)(t), 0 < \alpha < 1$, defined as

$$\begin{aligned} ({}_0D_t^{\alpha} f)(t) &\equiv \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_0^t \frac{f(\tau)}{(t-\tau)^{\alpha}} d\tau \\ &= \frac{1}{\Gamma(1-\alpha)} \left[\frac{f(0)}{t^{\alpha}} + \int_0^t \frac{f^{(1)}(\tau)}{(t-\tau)^{\alpha}} d\tau \right]. \end{aligned} \tag{1.25}$$

is expandable in a power series involving integer order derivatives as (see [24], p. 278)

$$({}_0D_t^{\alpha} f)(t) = \sum_{n=0}^{\infty} \binom{\alpha}{n} \frac{t^{n-\alpha}}{\Gamma(n+1-\alpha)} f^{(n)}(t), \tag{1.26}$$

where $\binom{\alpha}{n} = \frac{(-1)^{n-1}\alpha\Gamma(n-\alpha)}{\Gamma(1-\alpha)\Gamma(n+1)} = \frac{\Gamma(\alpha+1)}{n!\Gamma(\alpha-n+1)} = \frac{\sin(n-\alpha)\pi}{\pi} \frac{\Gamma(\alpha+1)\Gamma(n-\alpha)}{n!}$.

We shall derive another expansion formula for $({}_0D_t^{\alpha} f)(t)$. Let $V_n\left(f^{(p)}\right), n \in \mathbb{N}$, denote the $n-th$ moment of the function $f^{(p)}$,where $f^{(p)}, p \in \mathbb{N}$ is the $p-th$ (integer) derivative of f, i.e.,

$$V_n\left(f^{(p)}\right)(t) = \int_0^t f^{(p)}(\tau)\tau^n d\tau, \quad n \in \mathbb{N}, \;\; t \geq 0. \tag{1.27}$$

In what follows it is assumed that $f \in AC^2([0,b])$ and $0 < \alpha < 1$. Then, (1.25) is true on $(0,b]$. By partial integration in (1.25) we obtain

$$({}_0D_t^{\alpha} f)(t) = \frac{1}{\Gamma(1-\alpha)} f(0)\, t^{-\alpha} + \frac{1}{\Gamma(2-\alpha)} f^{(1)}(0)\, t^{1-\alpha} + \frac{1}{\Gamma(2-\alpha)} \int_0^t (t-\tau)^{1-\alpha} f^{(2)}(\tau)\, d\tau, \quad 0 < t \le b. \tag{1.28}$$

We now make use of the binomial formula

$$(1+z)^{\gamma} = \sum_{p=0}^{\infty} \left(\begin{array}{c} \gamma \\ p \end{array} \right) z^p = \sum_{p=0}^{\infty} \frac{(-1)^p \Gamma(p-\gamma)}{\Gamma(-\gamma)\, p!} z^p, \quad |z| < 1. \tag{1.29}$$

The expression (1.29) holds also for $z = 1$ if and only if $\gamma > -1$ and $z = -1, \gamma \neq 0$ if and only if $\gamma > 0$.

Also it is well known that

$$\left| \left(\begin{array}{c} \gamma \\ p \end{array} \right) \right| \le C \frac{1}{p^{1+\gamma}}, \ \ \gamma \neq -1, -2, ... \text{ and } p \to \infty. \tag{1.30}$$

With (1.29) the expression for $({}_0D_t^{\alpha} f)(t)$ becomes $(\gamma = 1 - \alpha)$

$$({}_0D_t^{\alpha} f)(t) = \frac{1}{\Gamma(1-\alpha)} f(0)\, t^{-\alpha} + \frac{1}{\Gamma(2-\alpha)} f^{(1)}(0)\, t^{1-\alpha} + t^{1-\alpha} \int_0^t f^{(2)}(\tau) \left(\sum_{p=1}^{\infty} \frac{\Gamma(p-1+\alpha)}{\Gamma(\alpha-1)\, p!} \left(\frac{\tau}{t} \right)^p \right) d\tau, \quad t > 0. \tag{1.31}$$

Because of (1.30) we can integrate series in (1.31) term by term. Also by using the relation

$$\int_0^t f^{(2)}(\tau)\, \tau^p d\tau = t^p f^{(1)}(t) - p \int_0^t f^{(1)}(\tau)\, \tau^{p-1} d\tau, \ \ p \ge 1$$

in (1.31) we obtain

$$({}_0D_t^{\alpha} f)(t) = \frac{1}{\Gamma(1-\alpha)} f(0)\, t^{-\alpha} + \frac{t^{1-\alpha}}{\Gamma(2-\alpha)} \sum_{p=1}^{\infty} \frac{\Gamma(p-1+\alpha)}{\Gamma(\alpha-1)\, p!} \left(f^{(1)}(t) - \frac{p}{t^p} V_{p-1}\left(f^{(1)} \right)(t) \right) + \frac{t^{1-\alpha}}{\Gamma(2-\alpha)} f^{(1)}(t), \quad t > 0 \ . \tag{1.32}$$

It is easy to derive a similar expression for the right fractional derivative by using the operator Q : If $f \in C([0,b])$, or $f \in C((0,b])$then $(Qf)(t) = f(b-t)$, where $f(b-t) \in C([0,b])$ and $f(b-t) \in C([0,b))$, respectively. We know that

${}_tD_b^\alpha(\cdot) = Q\left({}_0D_t^\alpha(\cdot)\right)Q$. If we apply the operator Q to (1.32) taking into account the properties of Q, we have for $f \in AC^2([0,b])$ and $0 < \alpha < 1$

$$\begin{aligned}
\left({}_tD_b^\alpha f\right)(t) &= \frac{1}{\Gamma(1-\alpha)} f(b)(b-t)^{-\alpha} \\
&+ \frac{(b-t)^{1-\alpha}}{\Gamma(2-\alpha)} \sum_{p=1}^{\infty} \frac{\Gamma(p-1+\alpha)}{\Gamma(\alpha-1)\,p!} \\
&\times \left(-f^{(1)}(b-t) - \frac{p}{(b-t)^p} V_{p-1}\left(-f^{(1)}(b-t)\right)\right) \\
&- \frac{(b-t)^{1-\alpha}}{\Gamma(2-\alpha)} f^{(1)}(b-t), \quad t > 0 \ . \qquad (1.33)
\end{aligned}$$

Integrating by parts in (1.32) and rearranging the result, we obtain

$$\begin{aligned}
\left({}_0D_t^\alpha f\right)(t) &= \frac{f(t)\,t^{-\alpha}}{\Gamma(1-\alpha)} + \frac{1}{\Gamma(2-\alpha)} \left\{ f^{(1)}(t) \left[1 + \sum_{p=1}^{\infty} \frac{\Gamma(p-1+\alpha)}{\Gamma(\alpha-1)\,p!}\right] t^{1-\alpha} \right. \\
&\left. - \sum_{p=2}^{\infty} \frac{\Gamma(p-1+\alpha)}{\Gamma(\alpha-1)\,(p-1)!} \left(\frac{f(t)}{t^\alpha} + \frac{\widetilde{V_p}}{t^{p-1+\alpha}}\right)\right\}, \qquad (1.34)
\end{aligned}$$

where

$$\widetilde{V_p} = -(p-1)\int_0^t \tau^{p-2} f(\tau)\, d\tau, \quad p = 2,3,.... \qquad (1.35)$$

Note that the moments, $\widetilde{V_p}, p = 1,2,...$ are solutions to the following system of differential equations

$$\widetilde{V_p}^{(1)}(t) = -(p-1)\,t^{p-2} f(t), \quad \widetilde{V_p}(0) = 0, \quad p = 2,3,... \qquad (1.36)$$

In application we shall use (1.34) with finite number of terms that is

$$\begin{aligned}
\left({}_0D_t^\alpha f\right)(t) &= \frac{f(t)\,t^{-\alpha}}{\Gamma(1-\alpha)} + \frac{1}{\Gamma(2-\alpha)} \left\{ f^{(1)}(t) \left[1 + \sum_{p=1}^{N} \frac{\Gamma(p-1+\alpha)}{\Gamma(\alpha-1)\,p!}\right] t^{1-\alpha} \right. \\
&\left. - \sum_{p=2}^{N} \frac{\Gamma(p-1+\alpha)}{\Gamma(\alpha-1)\,(p-1)!} \left(\frac{f(t)}{t^\alpha} + \frac{\widetilde{V_p}}{t^{p-1+\alpha}}\right)\right\}, \qquad (1.37)
\end{aligned}$$

with N suitably chosen.

Another approximation of the fractional derivative may be obtained as follows. First, by substituting $z = -1$ in (1.29) we obtain

$$1 + \sum_{p=1}^{\infty} \frac{\Gamma(p-1+\alpha)}{\Gamma(\alpha-1)\,p!} = 0, \qquad (1.38)$$

so that (1.34) becomes

$$\begin{aligned}\left({}_0D_t^{\alpha}f\right)(t) &= \frac{f(t)}{t^{\alpha}}\left[\frac{1}{\Gamma(1-\alpha)}-\frac{1}{\Gamma(\alpha-1)\Gamma(2-\alpha)}\sum_{p=2}^{\infty}\frac{\Gamma(p-1+\alpha)}{(p-1)!}\right]\\ &\quad-\frac{1}{\Gamma(\alpha-1)\Gamma(2-\alpha)}\sum_{p=2}^{\infty}\frac{\Gamma(p-1+\alpha)}{(p-1)!}\frac{\widetilde{V_p}}{t^{p-1+\alpha}}. \qquad (1.39)\end{aligned}$$

Again, by using a finite number of terms in (1.39) we have

$$\begin{aligned}\left({}_0D_t^{\alpha}f\right)(t) &\approx \frac{f(t)}{t^{\alpha}}\left[\frac{1}{\Gamma(1-\alpha)}-\frac{1}{\Gamma(\alpha-1)\Gamma(2-\alpha)}\sum_{p=2}^{N}\frac{\Gamma(p-1+\alpha)}{(p-1)!}\right]\\ &\quad-\frac{1}{\Gamma(\alpha-1)\Gamma(2-\alpha)}\sum_{p=2}^{N}\frac{\Gamma(p-1+\alpha)}{(p-1)!}\frac{\widetilde{V_p}}{t^{p-1+\alpha}}. \qquad (1.40)\end{aligned}$$

Note that the first sum on the right hand side could be written as

$$S(\alpha,N)=\sum_{p=2}^{N}\frac{\Gamma(p-1+\alpha)}{(p-1)!}=\frac{\Gamma(N+\alpha)}{(N-1)!}-\frac{\Gamma(1+\alpha)}{\alpha} \qquad (1.41)$$

Then, (1.40) becomes

$$\begin{aligned}\left({}_0D_t^{\alpha}f\right)(t) &\approx \frac{f(t)}{t^{\alpha}}\left[\frac{1}{\Gamma(1-\alpha)}-\frac{\dfrac{\Gamma(N+\alpha)}{(N-1)!}-\dfrac{\Gamma(1+\alpha)}{\alpha}}{\Gamma(\alpha-1)\Gamma(2-\alpha)}\right]\\ &\quad-\frac{1}{\Gamma(\alpha-1)\Gamma(2-\alpha)}\sum_{p=2}^{N}\frac{\Gamma(p-1+\alpha)}{(p-1)!}\frac{\widetilde{V_p}}{t^{p-1+\alpha}}, \qquad (1.42)\end{aligned}$$

with $\widetilde{V_p}$ satisfying (1.36). Equations (1.37),(1.42),(1.36) represent the basic relation that may be used in numerical solutions to fractional differential equations.

Remark. A numerical procedure based on different expression for the fractional derivative can be used. Namely, using the relation

$$\Gamma(\alpha)=\int_0^{\infty}z^{\alpha-1}\exp(-z)\,dz \qquad (1.43)$$

together with

$$\Gamma(\alpha)\Gamma(1-\alpha)=\frac{\pi}{\sin(\pi\alpha)} \qquad (1.44)$$

we obtain

$$\begin{aligned}
\left({}_0D_t^{\alpha} f\right)(t) &= \frac{1}{\Gamma(1-\alpha)}\frac{f(0)}{t^{\alpha}} + \frac{\sin(\pi\alpha)}{\pi}\int_0^t \Gamma(\alpha)\frac{f^{(1)}(\tau)}{(t-\tau)^{\alpha}}d\tau \\
&= \frac{1}{\Gamma(1-\alpha)}\frac{f(0)}{t^{\alpha}} + \frac{\sin(\pi\alpha)}{\pi}\int_0^t\int_0^{\infty} z^{\alpha-1}\exp(-z)\frac{f^{(1)}(\tau)}{(t-\tau)^{\alpha}}dzd\tau \\
&= \frac{1}{\Gamma(1-\alpha)}\frac{f(0)}{t^{\alpha}} + \frac{\sin(\pi\alpha)}{\pi}\int_0^t\int_0^{\infty} f^{(1)}(\tau)\exp(-z)\left[\frac{z}{t-\tau}\right]^{\alpha}\frac{dz}{z}d\tau.
\end{aligned} \tag{1.45}$$

Introducing new variable y by the relation $z = \left[(t-\tau)y^2\right]$ so that $dz = 2(t-\tau)ydy$ we obtain

$$\begin{aligned}
\left({}_0D_t^{\alpha} f\right)(t) &= \frac{1}{\Gamma(1-\alpha)}\frac{f(0)}{t^{\alpha}} + \frac{\sin(\pi\alpha)}{\pi}\int_0^t\int_0^{\infty} f^{(1)}(\tau)\exp\left(-(t-\tau)y^2\right)y^{2\alpha}\frac{2}{y}dyd\tau \\
&= \frac{1}{\Gamma(1-\alpha)}\frac{f(0)}{t^{\alpha}} + \frac{2\sin(\pi\alpha)}{\pi}\int_0^{\infty} y^{2\alpha-1}\left[\int_0^t f^{(1)}(\tau)\exp\left(-(t-\tau)y^2\right)d\tau\right]dy.
\end{aligned}$$

Again introducing the new dependent variable

$$\phi(y,t) = y^{2\alpha-1}\int_0^t f^{(1)}(\tau)\exp\left(-(t-\tau)y^2\right)d\tau \tag{1.46}$$

the last relation becomes

$$\left({}_0D_t^{\alpha} f\right)(t) = \frac{1}{\Gamma(1-\alpha)}\frac{f(0)}{t^{\alpha}} + \frac{2\sin(\pi\alpha)}{\pi}\int_0^{\infty}\phi(y,t)\,dy. \tag{1.47}$$

Note that from (1.46) we obtain

$$\frac{d}{dt}\phi(y,t) = y^{2\alpha-1}f^{(1)}(t) - y^2\phi(y,t) \tag{1.48}$$

Next the integral in (1.47) is approximated by using Laguerre's integration formula (taking N points $y_i, i = 1, ..., N$) that is $\int_0^{\infty}\phi(y,t)\,dy \approx \sum_{i=1}^{N} w_i\exp(y_i)\phi_i$. Therefore,

$$\left({}_0D_t^{\alpha} f\right)(t) \approx \frac{f(0)}{\Gamma(1-\alpha)t^{\alpha}} + \frac{2\sin(\pi\alpha)}{\pi}\sum_{i=1}^{N} w_i\exp(y_i)\phi_i, \tag{1.49}$$

where y_i and w_i are abscissas and weights of the Laguerre integration. The relation (1.48) now becomes

$$\begin{aligned}
-\phi_1^{(1)}(t) + y_1^{2\alpha-1}f^{(1)}(t) - y_1^2\phi_1(t) &= 0, \\
-\phi_2^{(1)}(t) + y_2^{2\alpha-1}f^{(1)}(t) - y_2^2\phi_2(t) &= 0, \\
-\text{.....} &= 0, \\
-\phi_N^{(1)}(t) + y_N^{2\alpha-1}f^{(1)}(t) - y_N^2\phi_N(t) &= 0.
\end{aligned} \tag{1.50}$$

It can be shown that the system (1.49),(1.50) is equivalent to the replacement, ab initio, of fractional dissipation element with classical Maxwell chain of springs and linear viscous elements. Moreover, the relaxation times in Maxwell model may be chosen arbitrary, while in (1.49) the values y_i^2 are fixed by the location of Laguerre integration points. Note that functions $\phi_i, i = 1, ..., N$ are solutions of the system (1.50) in which coefficients are constants while in (1.34) the moments $\widetilde{V}_i, i = 1, ..., N$ are solution of "rheonomic" system (1.36).

1.3 Viscoelastic body with fractional derivatives

In continuum mechanics, in general, and in Theory of viscoelasticity in particular the determination of restrictions on the coefficients in constitutive equations that follow from the second Law of Thermodynamics constitutes a major field of research. There are several possibilities to obtain those restrictions in the case (as our case is) of isothermal uniaxial state of stress. Thus one may impose the condition: *i*) that the work that is dissipated in *any* deformation process is positive, i.e., $\int_{\varepsilon_0}^{\varepsilon_1} \sigma d\varepsilon \geq 0$ where σ and ε denote the stress and strain (ε_0 is the strain at the beginning of the deformation process and ε_1 is the strain at the end of the deformation process. It is assumed that $\varepsilon_0 = \varepsilon_1$), *ii*) one may impose the condition of thermodynamic stability of the equilibrium state, and finally *iii*) use the condition that the tangent of the mechanical loss angle is positive. We demonstrate next two of those procedures. We analyze here the linear fractional viscoelasticity only.

Consider a standard viscoelastic body (Zener model), that in a uniaxial isothermal deformation has stress strain relation of the form

$$\tau_\sigma \sigma^{(1)} + \sigma = E\tau_\varepsilon \varepsilon^{(1)} + E\varepsilon, \tag{1.51}$$

where σ and ε denote the stress and strain at time t, respectively and $(\cdot)^{(1)} = \frac{d}{dt}(\cdot)$ denotes the first derivative with respect to time, $\tau_\sigma, \tau_\varepsilon$ are constants called relaxation times, and E is a constant called the modulus of elasticity. The second law of thermodynamics, implies that in (1.51) the following restrictions on the constants must be satisfied

$$E > 0, \quad \tau_\sigma > 0, \quad \tau_\varepsilon > \tau_\sigma. \tag{1.52}$$

With equations of the type (1.51) it is not always possible to explain the experiments. Indeed, there is a class of viscoelastic materials which is better described by the constitutive equation (modified Zener model),

$$\sigma + b\sigma^{(\beta)} = E_0\varepsilon + E_1\varepsilon^{(\alpha)}, \tag{1.53}$$

where $\sigma^{(\beta)}$ and $\varepsilon^{(\alpha)}$ are fractional derivatives; $0 < \alpha < 1, 0 < \beta < 1$. For classical and modified Zener models of viscoelasticity theory, see also [23].

By invoking the second law of thermodynamics for sinusoidal strain and sinusoidal stress (the tangent of the mechanical loss angle approach) the following restrictions on the constants α, β, b, E_0 and E_1 are obtained:

$$E_0 \geq 0, \quad E_1 > 0, \quad b > 0, \quad \frac{E_1}{b} \geq E_0, \quad \alpha = \beta. \tag{1.54}$$

Except for the equality in $(1.54)_1$ these conditions are identical to (1.52), whenever a comparison is possible, i.e., for $\alpha = \beta = 1$.

There are other propositions for the constitutive relations that lead to (1.53) as well. Thus the relation

$$\sigma(t) = D_*^{(\alpha)}\varepsilon = \frac{E\tau^\alpha}{\Gamma(1-\alpha)}\int_0^t \frac{\varepsilon^{(1)}(u)}{(t-u)^\alpha}du, \tag{1.55}$$

was analyzed, where E and τ are constants. $D_*^{(\alpha)}\varepsilon$ is the *Caputo fractional derivative of the order* α. Note that $D_*^{(\alpha)}\varepsilon$ and $\varepsilon^{(\alpha)}$ are connected by $D_*^{(\alpha)}\varepsilon = \varepsilon^{(\alpha)} - \frac{\varepsilon(0)}{\Gamma(1-\alpha)t^\alpha}$. An element with the constitutive equation given by (1.55) is called a springpot. In the context of (1.55) it is possible to identify the inelastic part ε_{in} of the strain by

$$D_*^{(\alpha)}\varepsilon_{in} = \frac{1}{\tau^\alpha}(\varepsilon - \varepsilon_{in}), \tag{1.56}$$

and define a standard solid with the constitutive relation

$$\sigma(t) = E_{eq}\varepsilon + E(\varepsilon - \varepsilon_{in}). \tag{1.57}$$

By interpreting a fractional damping in terms of a continuous superposition of Maxwell elements in parallel it was shown that the condition $E > 0$ guarantees that the entropy inequality is satisfied. The parallel connection of springpot elements of the type (1.55) is exactly the rheological interpretation of the model (1.53) with $\alpha = \beta$. References to other works treating thermodynamic admissibility of viscoelastic models with fractional derivatives can also be given.

Here we follow [3] and formulate a fractional derivatives type model for one dimensional deformations of a viscoelastic body. The constitutive equation that we construct is of the form (1.53) with $\alpha = \beta$ and with an additional term. It reads

$$\sigma + \tau_\sigma\sigma^{(\gamma)} = E\left[\varepsilon + \tau_\varepsilon\varepsilon^{(\gamma)}\right] + E(1 - \frac{\tau_\varepsilon}{\tau_\sigma})[d^{(\gamma)} + \frac{1}{\tau_\sigma}d]. \tag{1.58}$$

The last term, by which (1.58) is different from (1.53), results from the desire to make viscoelastic constitutive equation compatible with an internal variable theory which lends itself for a *clear-cut* exposition of the thermodynamic conditions imposed by the second law. In (1.58) the functional d is given as

$$\begin{aligned} d &= \int_0^t e^{-\frac{1}{\tau_\sigma}(t-\tau)}(\varepsilon(\tau) \\ &\quad + \int_0^\tau e_{\gamma,\gamma}(u;\frac{1}{\tau_\sigma})[-\frac{1}{\tau_\sigma}\varepsilon(\tau-u) - \varepsilon^{(1)}(\tau-u)]du)d\tau, \end{aligned} \tag{1.59}$$

where $e_{\gamma,\gamma}(t;\lambda)$ is defined by (1.84) below. The functional d is chosen so that (1.58) satisfies the entropy inequality for *all* deformations $\varepsilon(t)$. The explicit form (i.e. (1.58) solved for $\sigma(t)$) of the constitutive equation is given by (1.111) below. The functional $d(\varepsilon)$ is equal to zero when $\gamma = 1$ and decreases strongly

with time. Its influence is important only at the beginning of the motion. We shall examine the restrictions which the entropy inequality *and* equilibrium stability conditions imply on a constitutive equation of the type (1.58) and we shall conclude that the presence of $d(\varepsilon)$ in (1.58) is essential, if one wants to satisfy the entropy inequality for all deformations $\varepsilon(t)$ (see Remark 1).

1.3.1 The internal variable theory

We recall the description of the constitutive relation (1.51) and of the thermodynamic stability conditions (1.52) in the context of an internal variable theory.

Consider a rod in the uniaxial isothermal deformation. The length is L in the undeformed state and $l(t)$ during the deformation. The rod is loaded by the force P and F is the cross-sectional area in the undeformed state. Thus stress[3] and strain are given by

$$\sigma(t) = \frac{P}{F} \quad \text{and} \quad \varepsilon(t) = \frac{l}{L} - 1. \tag{1.60}$$

We describe a state of the body by two variables: the strain $\varepsilon(t)$ and an internal variable $\xi(t)$. The equilibrium state of the unloaded body corresponds to

$$\varepsilon = 0, \quad \xi = 0. \tag{1.61}$$

Thus the internal energy U, the entropy S and the free energy $U - TS$ are all functions of ε, or l and ξ and we may write the Gibbs equation for the free energy $U - TS$ in the form

$$\frac{d(U - TS)}{dt} = \sigma V \frac{d\varepsilon}{dt} - \Theta \frac{d\xi}{dt}. \tag{1.62}$$

T is the temperature, assumed to be constant. Θ is the "force" associated with the internal variable ξ so that $\Theta \frac{d\xi}{dt}$ is the power of the force Θ, and V is the volume of the body which is assumed to be constant. We assume that both σ and Θ are linear in the variables ε and ξ so that

$$\sigma = E_\infty \varepsilon + \beta \xi, \quad \Theta = \gamma \varepsilon + \delta \xi, \tag{1.63}$$

where E_∞, β, γ and δ are constants. From (1.63) we conclude that ξ is proportional to the difference between the instantaneous and equilibrium stress $E_\infty \varepsilon$. Note that with (1.63) the force P is given as

$$P = F(E_\infty \varepsilon + \beta \xi). \tag{1.64}$$

The integrability for the free energy requires

$$V\beta = -\gamma. \tag{1.65}$$

Therefore, with (1.65) the equation (1.63) becomes

$$\sigma = E_\infty \varepsilon + \beta \xi, \quad \Theta = -V\beta\varepsilon + \delta \xi. \tag{1.66}$$

[3] The stress σ is the load referred to the cross-sectional area of the unloaded rod.

We return to the Gibbs equation (1.62) in which we replace $\frac{dU}{dt}$ and $\frac{d\xi}{dt}$ by the equations of balance of energy and of internal variable, viz.

$$\frac{dU}{dt} = \dot{Q} + \sigma V \frac{d\varepsilon}{dt} \quad \text{and} \qquad \frac{d\xi}{dt} = P_\xi. \tag{1.67}$$

$\dot{Q}$ is the heating and P_ξ is the production of ξ. Thus we obtain an equation of balance of entropy in the form

$$\frac{dS}{dt} - \frac{\dot{Q}}{T} = \frac{\Theta}{T} P_\xi \geq 0, \tag{1.68}$$

where we have indicated that the entropy production is non-negative. The inequality (1.68) may be satisfied by setting $P_\xi = \alpha\Theta$ with a non-negative coefficient α. Therefore

$$\xi^{(1)} = \alpha\Theta \qquad \alpha \geq 0. \tag{1.69}$$

Elimination of ξ and Θ between the three equations (1.66) and (1.69) provides

$$\sigma + \frac{1}{-\delta\alpha}\sigma^{(1)} = \left(E_\infty + V\frac{\beta^2}{\delta}\right)\varepsilon + \frac{E_\infty}{-\alpha\delta}\varepsilon^{(1)}. \tag{1.70}$$

which is of the form (1.51), if we identify the coefficients in (1.51) with the coefficients of the internal variable theory as follows

$$\tau_\sigma = \frac{1}{-\delta\alpha}, \quad E = \left(E_\infty + V\frac{\beta^2}{\delta}\right), \quad \tau_\varepsilon = \frac{E_\infty}{-\alpha\delta}\frac{1}{E_\infty + \frac{V\beta^2}{\delta}}. \tag{1.71}$$

Thus the viscoelastic constitutive equation (1.51) is a consequence of the internal variable theory. It results upon elimination of all explicit reference to the internal variable field.

The internal variable theory lends itself for an easy evaluation of thermodynamic stability conditions. Indeed, elimination of $\dot{Q}$ between (1.67) and (1.68) provides the inequality

$$\frac{d\,(U - TS - Pl)}{dt} \leq l\frac{dP}{dt}. \tag{1.72}$$

Thus for a constant force the Gibbs free energy $(U - TS - Pl)$ tends to a minimum and it assumes that minimum in equilibrium. Therefore the matrix of the second derivatives of the Gibbs free energy must be positive definite. Thus by (1.62) an (1.66) we must have

$$\begin{bmatrix} E_\infty & \beta V \\ \beta V & -\delta \end{bmatrix} \quad \text{pos. def.} \tag{1.73}$$

or

$$E_\infty > 0, \quad -E_\infty\delta - \beta^2 V > 0, \tag{1.74}$$

so that $E_\infty > 0$ and $\delta < 0$ or $E_\infty + \frac{\beta^2 V}{\delta} > 0$. By using this in (1.71) we obtain

$$\tau_\sigma > 0, \quad E > 0, \quad \tau_\varepsilon > \tau_\sigma, \tag{1.75}$$

i.e. the conditions (1.52).

1.3.2 The fractional derivative model

Consider now the constitutive equation of the type (1.58). We write it as

$$\sigma + \tau_\sigma \sigma^{(\gamma)} = E\left[\varepsilon + \tau_\varepsilon \varepsilon^{(\gamma)}\right] + g(\varepsilon). \tag{1.76}$$

where $g(\varepsilon)$ is given functional. If $g(\varepsilon) \equiv 0$ equation (1.76) reduces to (1.53) with $\alpha = \beta$. We want to include this type of equation into the internal variable framework of the previous section. We repeat the relevant system of equations (1.66),(1.67),(1.68)

$$\begin{aligned} \sigma &= E_\infty \varepsilon + \beta\xi, \quad \Theta = -V\beta\varepsilon + \delta\xi, \\ \xi^{(1)} &= P_\xi, \quad \frac{dS}{dt} - \frac{1}{T}\frac{dQ}{dt} = \frac{\Theta}{T} P_\xi \geq 0, \end{aligned} \tag{1.77}$$

that must be satisfied by $\sigma(t)$ and $\varepsilon(t)$ satisfying (1.76). Suppose that the γ^{th} derivative of ξ is given by

$$\xi^{(\gamma)} = \alpha\Theta + X, \tag{1.78}$$

where X is a functional that will be specified later. Equation (1.78) is a generalization of (1.69) and is of central importance in the analysis that follows. Note that with $X = 0$ the equation (1.78) leads to a constitutive relation of the type (1.53) with $\alpha = \beta$. Equation (1.69) is obtained if we take $\gamma = 1$ and $X = 0$. The special case of (1.78) with $X = 0, \gamma \neq 1$ was used for an internal variable description of the fractional derivative model. In [**?**] equation (1.78) with $X = 0$ is used in the context of finite deformations of a viscoelastic body.. Our intention is to choose X so that $(1.77)_4$ is satisfied. Now from $(1.66)_2$ and (1.78) it follows

$$\xi^{(\gamma)} + (-\alpha\delta)\,\xi = K\varepsilon + X, \tag{1.79}$$

where $K = -\alpha V \beta$. Suppose that X is given as

$$X = x^{(\gamma)} + (-\alpha\delta)\,x, \tag{1.80}$$

where x is another function. With (1.80) equation (1.79) could be written as

$$(\xi - x)^{(\gamma)} + (-\alpha\delta)(\xi - x) = K\varepsilon. \tag{1.81}$$

The solution to the fractional order differential equation (1.81) reads:

$$\begin{aligned} \xi - x &= K\frac{E_{\gamma,\gamma}(\alpha\delta t^\gamma)}{t^{1-\gamma}} \frac{1}{\Gamma(1-\alpha)} \left(\int_0^t \frac{(\xi(\tau) + x(\tau))\,d\tau}{(t-\tau)^\alpha}\right)_{t=0} \\ &\quad + K\int_0^t \frac{E_{\gamma,\gamma}(\alpha\delta(t-\tau)^\gamma)}{(t-\tau)^{1-\gamma}} \varepsilon(\tau)\,d\tau, \end{aligned} \tag{1.82}$$

where $E_{\alpha,\beta}(t)$ is the Mittag-Lefller function defined by

$$E_{\alpha,\beta}(t) = \sum_{n=0}^{\infty} \frac{t^n}{\Gamma(\alpha n + \beta)}, \tag{1.83}$$

Note that $E_{1,1}(t) = e^t$. With $E_{\alpha,\beta}(t)$ thus given, a new function $e_{\alpha,\beta}$ can be defined by

$$e_{\alpha,\beta}(t;\lambda) \equiv \frac{E_{\alpha,\beta}(-\lambda t^\alpha)}{t^{1-\beta}}, \tag{1.84}$$

This function possesses the following properties

$$\begin{aligned} e_{1,1}(t;\lambda) &= e^{-\lambda t}, \quad (-1)^n \frac{d^n}{dt^n}\left(e_{\alpha,\beta}(t;\lambda)\right) \geq 0, \quad n = 0,1,2,.., \\ e_{\alpha,\beta}(t;\lambda) &= \int_0^t e^{-rt} \frac{1}{\pi} \frac{\lambda \sin[(\beta-\alpha)] + r^\alpha \sin(\beta r)}{r^{2\alpha} + 2\lambda r^\alpha \cos(\alpha\pi) + \lambda^2} r^{\alpha-\beta} dr. \end{aligned} \tag{1.85}$$

With (1.84) and by use of the fact that at $t = 0$ both ξ and x are equal to zero (1.82) becomes

$$\begin{aligned} \xi &= K \int_0^t e_{\gamma,\gamma}(t-\tau;-\alpha\delta)\,\varepsilon(\tau)\,d\tau + x \\ &= K \int_0^t e_{\gamma,\gamma}(\tau;-\alpha\delta)\,\varepsilon(t-\tau)\,d\tau + x. \end{aligned} \tag{1.86}$$

We now specify x in the form

$$\begin{aligned} x &= K\left\{\int_0^t e^{-\lambda(t-\tau)}(\varepsilon(\tau) \right. \\ &\left. + \int_0^\tau e_{\gamma,\gamma}(u;-\alpha\delta)\left[\Lambda\varepsilon(\tau-u) - \varepsilon^{(1)}(\tau-u)\right]du\right)d\tau\right\}, \end{aligned} \tag{1.87}$$

where λ and Λ are constants. Combining (1.86) and (1.87) we obtain

$$\begin{aligned} \xi &= K\left\{\int_0^t e_{\gamma,\gamma}(\tau;-\alpha\delta)\,\varepsilon(t-\tau)\,d\tau + \int_0^t e^{-\lambda(t-\tau)}(\varepsilon(\tau) \right. \\ &\left. + \int_0^\tau e_{\gamma,\gamma}(u;-\alpha\delta)\left[\Lambda\varepsilon(\tau-u) - \varepsilon^{(1)}(\tau-u)\right]du\right)d\tau\right\}. \end{aligned} \tag{1.88}$$

Next we use (1.88) to determine $\xi^{(1)}$, and thus P_ξ (see $(1.77)_3$). The first derivative of ξ becomes

$$\begin{aligned} \xi^{(1)} &= K\left\{e_{\gamma,\gamma}(t;-\alpha\delta)\,\varepsilon(0) + \int_0^t e_{\gamma,\gamma}(\tau;-\alpha\delta)\,\varepsilon^{(1)}(t-\tau)\,d\tau \right. \\ &+ \varepsilon(t) + \int_0^t e_{\gamma,\gamma}(u;-\alpha\delta)\left[\Lambda\varepsilon(t-u) - \varepsilon^{(1)}(t-u)\right]du \\ &-\lambda\int_0^t e^{-\lambda(t-\tau)}(\varepsilon(\tau) \\ &\left. + \int_0^\tau e_{\gamma,\gamma}(u;-\alpha\delta)\left[\Lambda\varepsilon(\tau-u) - \varepsilon^{(1)}(\tau-u)\right]du\right)d\tau\right\}, \end{aligned} \tag{1.89}$$

and since $\varepsilon(0)=0$

$$\begin{aligned}\xi^{(1)} &= K\left\{\varepsilon(t)+\Lambda\int_0^t e_{\gamma,\gamma}(\tau;-\alpha\delta)\,\varepsilon(t-\tau)\,d\tau-\lambda\int_0^t e^{-\lambda(t-\tau)}(\varepsilon(\tau)\right.\\ &\quad\left.+\int_0^\tau e_{\gamma,\gamma}(u;-\alpha\delta)\,[\Lambda\varepsilon(\tau-u)-\varepsilon^{(1)}(\tau-u)]du)d\tau\right\}\\ &= \alpha\left\{-\beta V\varepsilon(t)-\beta V\Lambda\int_0^t e_{\gamma,\gamma}(t-\tau;\alpha\delta)\,\varepsilon(\tau)\,d\tau\right.\\ &\quad+\beta V\lambda\int_0^t e^{-\lambda(t-\tau)}(\varepsilon(\tau)\\ &\quad\left.+\int_0^\tau e_{\gamma,\gamma}(u;-\alpha\delta)\,[\Lambda\varepsilon(\tau-u)-\varepsilon^{(1)}(\tau-u)]du)d\tau\right\}. \qquad (1.90)\end{aligned}$$

Suppose that Λ and λ are selected so that

$$-\beta V\Lambda=K\delta=(-\alpha V\beta)\,\delta,\quad \beta V\lambda=K\delta=-\alpha V\beta\delta, \qquad (1.91)$$

or

$$\Lambda=(\alpha\delta),\quad \lambda=-(\alpha\delta). \qquad (1.92)$$

Then (1.90) becomes (use $\Theta=-V\beta\varepsilon+\delta\xi$ and (1.88))

$$\begin{aligned}\xi^{(1)} &= \alpha\left\{-\beta V\varepsilon(t)+\delta K\left[\int_0^t e_{\gamma,\gamma}(\tau;\alpha\delta)\,\varepsilon(t-\tau)\,d\tau+\int_0^t e^{-\lambda(t-\tau)}\right.\right.\\ &\quad\left.\left.\times(\varepsilon(\tau)+\int_0^\tau e_{\gamma,\gamma}(u;-\alpha\delta)\,[\Lambda\varepsilon(\tau-u)-\varepsilon^{(1)}(\tau-u)]du)d\tau\right]\right\}\\ &= \alpha\{-\beta V\varepsilon(t)+\delta\xi(t)\}=\alpha\Theta. \qquad (1.93)\end{aligned}$$

Thus, with

$$\alpha>0, \qquad (1.94)$$

the condition $(1.77)_4$ is satisfied. Note also that (1.93) guarantees that the constitutive equation for the internal variable ξ is local in time, since the right hand side of (1.93) is *a function* of $\varepsilon(t)$ and $\xi(t)$.

We determine now the form of the constitutive equation ($\sigma-\varepsilon$ relation) that follows from (1.79). Combining $(1.77)_1$,(1.79),(1.80) and (1.87) we have

$$\left(\frac{\sigma-E_\infty\varepsilon}{\beta}\right)^\gamma+(-\alpha\delta)\left(\frac{\sigma-E_\infty\varepsilon}{\beta}\right)=(-\alpha V\beta)\,\varepsilon+x^{(\gamma)}+(-\alpha\delta)\,x, \qquad (1.95)$$

where

$$\begin{aligned}x &= -\alpha V\beta\left\{\int_0^t e^{\alpha\delta(t-\tau)}(\varepsilon(\tau)\right.\\ &\quad\left.+\int_0^\tau e_{\gamma,\gamma}(u;-\delta)\,[\alpha\delta\varepsilon(\tau-u)-\varepsilon^{(1)}(\tau-u)]du)d\tau\right\}. \qquad (1.96)\end{aligned}$$

From (1.95) we obtain

$$\sigma^{(\gamma)} - E_\infty \varepsilon^{(\gamma)} - \alpha\delta\sigma + \alpha\delta E_\infty \varepsilon = -\alpha V\beta^2 \varepsilon + \beta x^{(\gamma)} - \alpha\delta\beta x, \tag{1.97}$$

or

$$\frac{1}{-\alpha\delta}\sigma^{(\gamma)} + \sigma = \frac{E_\infty}{-\alpha\delta}\varepsilon^{(\gamma)} + [E_\infty + \frac{V\beta^2}{\delta}]\varepsilon + \frac{\beta}{-\alpha\delta}x^{(\gamma)} + \beta x. \tag{1.98}$$

With

$$\tau_\sigma = \frac{1}{-\alpha\delta}, \quad E = E_\infty + \frac{V\beta^2}{\delta}, \quad \tau_\varepsilon = \frac{E_\infty}{-\alpha\delta}\frac{1}{E} = \tau_\sigma \frac{E_\infty}{E} \tag{1.99}$$

(1.98) becomes

$$\tau_\sigma \sigma^{(\gamma)} + \sigma = E[\tau_\varepsilon \varepsilon^{(\gamma)} + \varepsilon] + E(1 - \frac{\tau_\varepsilon}{\tau_\sigma})[d^{(\gamma)} + \frac{1}{\tau_\sigma} d], \tag{1.100}$$

where

$$\begin{aligned} d &= \int_0^t e^{-\frac{1}{\tau_\sigma}(t-\tau)}(\varepsilon(\tau) \\ &\quad + \int_0^\tau e_{\gamma,\gamma}(u; \frac{1}{\tau_\sigma})[-\frac{1}{\tau_\sigma}\varepsilon(\tau - u) - \varepsilon^{(1)}(\tau - u)]du)d\tau. \end{aligned} \tag{1.101}$$

Note that for $\gamma = 1$ we have

$$e_{1,1}(t; \frac{1}{\tau_\sigma}) = e^{-\frac{1}{\tau_\sigma}t}, \tag{1.102}$$

so that

$$\begin{aligned} d &= \int_0^t e^{-\frac{1}{\tau_\sigma}(t-\tau)}(\varepsilon(\tau) + \int_0^\tau e^{-\frac{1}{\tau_\sigma}u}[-\frac{1}{\tau_\sigma}\varepsilon(\tau - u) - \varepsilon^{(1)}(\tau - u)]du)d\tau \\ &= \int_0^t e^{-\frac{1}{\tau_\sigma}(t-\tau)}(\varepsilon(\tau) + \int_0^\tau \frac{d}{du}\left(e^{-\frac{1}{\tau_\sigma}u}\varepsilon(\tau - u)\right)du)d\tau \\ &= \int_0^t e^{-\frac{1}{\tau_\sigma}(t-\tau)}(\varepsilon(\tau) - \varepsilon(\tau))d\tau = 0, \end{aligned} \tag{1.103}$$

and (1.100) reduces to (1.51).

Remark 1 *The constitutive equation (1.100) reduces to (1.53), if we take* $d = 0$. *We show that in this case there is a deformation process* $\varepsilon(t)$ *such that* $(1.77)_4$ *is violated. If* $d = 0$ *then,* $x = 0$ *so that (1.88) becomes*

$$\xi = K\int_0^t e_{\gamma,\gamma}(\tau; -\alpha\delta)\,\varepsilon(t - \tau)\,d\tau. \tag{1.104}$$

Also, instead of (1.89) we get

$$\begin{aligned} \xi^{(1)} &= K\left(e_{\gamma,\gamma}(t; -\alpha\delta)\,\varepsilon(0) + \int_0^t e_{\gamma,\gamma}(\tau; -\alpha\delta)\,\varepsilon^{(1)}(t - \tau)\,d\tau\right) \\ &= K\int_0^t e_{\gamma,\gamma}(\tau; -\alpha\delta)\,\varepsilon^{(1)}(t - \tau)\,d\tau. \end{aligned} \tag{1.105}$$

Therefore $P_\xi = K\int_0^t e_{\gamma,\gamma}(\tau;-\alpha\delta)\,\varepsilon^{(1)}(t-\tau)\,d\tau$. *Also, with* $d = x = 0$ *the function* Θ *reads*

$$\begin{aligned}\Theta(t) &= -V\beta\varepsilon + \delta\xi \\ &= -V\beta\varepsilon(t) - \alpha V\beta\delta\int_0^t e_{\gamma,\gamma}(\tau;-\alpha\delta)\,\varepsilon(t-\tau)\,d\tau, \qquad (1.106)\end{aligned}$$

so that

$$\begin{aligned}\frac{\Theta}{T}P_\xi &= \frac{1}{T}\left(-V\beta\varepsilon(t) - \alpha V\beta\delta\int_0^t e_{\gamma,\gamma}(\tau;-\alpha\delta)\,\varepsilon(t-\tau)\,d\tau\right)\times \\ &\quad (-\alpha V\beta)\int_0^t e_{\gamma,\gamma}(\tau;-\alpha\delta)\,\varepsilon^{(1)}(t-\tau)\,d\tau. \qquad (1.107)\end{aligned}$$

It is obvious from (1.107) that we can always choose $\varepsilon(t)$ *so that (1.77)*$_4$ *is violated.*

The restrictions that follow from the stability of the equilibrium states are satisfied since we started with (1.77) and since $(1.77)_1$ guarantees that the integrability conditions for $U-TS$ are satisfied. Therefore the convexity of $U-TS$ as a function of the variables ε,ξ leads, again, to (1.74),(1.75).

To analyze (1.100) we apply the Laplace transform ($\mathcal{L}f = \int_0^\infty e^{-st}f(t)\,dt = \bar{f}(s)$) so that

$$\tau_\sigma s^\gamma\bar{\sigma} + \bar{\sigma} = E\bar{\varepsilon}[\tau_\varepsilon s^\gamma + 1] + [E(1-\frac{\tau_\varepsilon}{\tau_\sigma})(s^\gamma + \frac{1}{\tau_\sigma})]\bar{d}, \qquad (1.108)$$

where it is assumed that $\sigma(t)$ and $\varepsilon(t)$ are bounded[4] for $t\to+0$. From (1.101) it follows

$$\begin{aligned}\bar{d} &= \frac{1}{s+\frac{1}{\tau_\sigma}}\mathcal{L}\left(\varepsilon(\tau) + \int_0^\tau e_{\gamma,\gamma}(u;\frac{1}{\tau_\sigma})[-\frac{1}{\tau_\sigma}\varepsilon(\tau-u) - \varepsilon^{(1)}(\tau-u)]du\right) \\ &= \frac{1}{s+\frac{1}{\tau_\sigma}}\left(\bar{\varepsilon} + \mathcal{L}\left(\int_0^\tau e_{\gamma,\gamma}(u;\frac{1}{\tau_\sigma})[-\frac{1}{\tau_\sigma}\varepsilon(\tau-u) - \varepsilon^{(1)}(\tau-u)]du\right)\right) \\ &= \frac{1}{s+\frac{1}{\tau_\sigma}}\left(\bar{\varepsilon} + \mathcal{L}\left(e_{\gamma,\gamma}(t;\frac{1}{\tau_\sigma})\right)\mathcal{L}\left([-\frac{1}{\tau_\sigma}\varepsilon(t) - \varepsilon^{(1)}(t)]\right)\right) \\ &= \frac{1}{s+\frac{1}{\tau_\sigma}}\left(1 + \frac{1}{s^\gamma+\frac{1}{\tau_\sigma}}\left(-\frac{1}{\tau_\sigma} - s\right)\right)\bar{\varepsilon} \\ &= \left(\frac{1}{s+\frac{1}{\tau_\sigma}} - \frac{1}{s^\gamma+\frac{1}{\tau_\sigma}}\right)\bar{\varepsilon} = \left(\frac{s^\gamma - s}{\left(s+\frac{1}{\tau_\sigma}\right)\left(s^\gamma+\frac{1}{\tau_\sigma}\right)}\right)\bar{\varepsilon}, \qquad (1.109)\end{aligned}$$

[4] The Laplace transform of $f^{(\gamma)}$ is given as $\mathcal{L}\left(f^{(\gamma)}\right) = s^\gamma\bar{f} - \left[\frac{1}{\Gamma(1-\gamma)}\int_0^t\frac{f(\tau)}{(t-\tau)^\gamma}\right]_{t=0}$. The term in brackets vanishes if $\lim_{t\to+0}f(t)$ is bounded (see [20]).

where $\mathcal{L}\left(e_{\gamma,\gamma}\left(t;1/\tau_\sigma\right)\right)=1/\left(s^\gamma+\frac{1}{\tau_\sigma}\right)$ and $\varepsilon\left(0\right)=0$ was used. From (1.109) and (1.108) we get

$$\begin{aligned}\bar{\sigma} &= \bar{\varepsilon}\frac{E[\tau_\varepsilon s^\gamma+1]+E\left(1-\frac{\tau_\varepsilon}{\tau_\sigma}\right)[s^\gamma+\frac{1}{\tau_\sigma}]\frac{s^\gamma-s}{\left(s+\frac{1}{\tau_\sigma}\right)\left(s^\gamma+\frac{1}{\tau_\sigma}\right)}}{\tau_\sigma s^\gamma+1}\\ &= E\bar{\varepsilon}\frac{\tau_\varepsilon s^\gamma+1+\left(1-\frac{\tau_\varepsilon}{\tau_\sigma}\right)\frac{s^\gamma-s}{\left(s+\frac{1}{\tau_\sigma}\right)}}{\tau_\sigma s^\gamma+1}. \end{aligned} \tag{1.110}$$

The solution $\sigma\left(t\right)$ can be obtained by finding the inverse Laplace transform of (1.110) or from $(1.77)_1$. By using (1.88),(1.92) and (1.99) we may write the solution of (1.100) for σ as

$$\begin{aligned}\sigma\left(t\right) &= E\frac{\tau_\varepsilon}{\tau_\sigma}\varepsilon+\frac{1}{\tau_\sigma}E(1-\frac{\tau_\varepsilon}{\tau_\sigma})\left\{\int_0^t e_{\gamma,\gamma}(\tau;\frac{1}{\tau_\sigma})\varepsilon\left(t-\tau\right)d\tau+\int_0^t e^{-\frac{1}{\tau_\sigma}(t-\tau)}\right.\\ &\left.\times\left(\varepsilon\left(\tau\right)+\int_0^\tau e_{\gamma,\gamma}(u;\frac{1}{\tau_\sigma})[-\frac{1}{\tau_\sigma}\varepsilon\left(\tau-u\right)-\varepsilon^{(1)}\left(\tau-u\right)]du\right)d\tau\right\}. \end{aligned} \tag{1.111}$$

Note that the constitutive equation (1.100) with d given by (1.101) is equivalent to (1.51). This can be concluded from (1.93) and (1.69) or from (1.110).

1.3.3 Creep and stress relaxation

To examine the asymptotic behavior of the solution to (1.100) or (1.111) and to compare it to the asymptotic behavior of the solutions to (1.53) for $\alpha=\beta$ we consider the special kind of applied stress (strain). Thus, suppose that

$$\sigma\left(t\right)=\left\{\begin{array}{cc}0 & t\leq 0\\ \sigma_0 & t>0\end{array}\right. . \tag{1.112}$$

Then $\mathcal{L}\left(\sigma\right)=\sigma_0/s$, so that (1.110) leads to

$$\bar{\varepsilon}=\frac{\sigma_0}{E}\frac{1}{s}\frac{\tau_\sigma s^\gamma+1}{[\tau_\varepsilon s^\gamma+1]+\left(1-\frac{\tau_\varepsilon}{\tau_\sigma}\right)\frac{s^\gamma-s}{\left(s+\frac{1}{\tau_\sigma}\right)}}. \tag{1.113}$$

Now, if $\lim\limits_{t\to\infty}\varepsilon\left(t\right)$ exists, it is given as $\lim\limits_{t\to\infty}\varepsilon\left(t\right)=\lim\limits_{s\to 0}s\bar{\varepsilon}\left(s\right)$ (see [**?**]). Use of this in (1.113) leads to

$$\lim_{t\to\infty}\varepsilon\left(t\right)=\frac{\sigma_0}{E}. \tag{1.114}$$

Also, if $\lim\limits_{t\to 0}\varepsilon\left(t\right)=\varepsilon\left(+0\right)$ exists, it is given as $\varepsilon\left(+0\right)=\lim\limits_{s\to\infty}s\bar{\varepsilon}\left(s\right)$. Therefore

$$\lim_{t\to 0}\varepsilon\left(t\right)=\varepsilon\left(+0\right)=\frac{\sigma_0}{E}\frac{\tau_\sigma}{\tau_\varepsilon}<\frac{\sigma_0}{E}. \tag{1.115}$$

Next suppose that

$$\varepsilon(t) = \begin{cases} 0 & t \leq 0 \\ \varepsilon_0 & t > 0 \end{cases} . \tag{1.116}$$

From (1.110) it follows that

$$\bar{\sigma} = E\varepsilon_0 \frac{1}{s} \frac{\tau_\varepsilon s^\gamma + 1 + \left(1 - \frac{\tau_\varepsilon}{\tau_\sigma}\right) \frac{s^\gamma - s}{\left(s + \frac{1}{\tau_\sigma}\right)}}{\tau_\sigma s^\gamma + 1}. \tag{1.117}$$

Again $\lim\limits_{t\to\infty} \sigma(t) = \lim\limits_{s\to 0} s\bar{\sigma}(s)$ so that

$$\lim_{t\to\infty} \sigma(t) = E\varepsilon_0. \tag{1.118}$$

Also, for $t \to 0$ we have

$$\lim_{t\to 0} \sigma(t) = \sigma(+0) = \lim_{s\to\infty} s\bar{\sigma}(s) = E\varepsilon_0 \frac{\tau_\varepsilon}{\tau_\sigma} > E\varepsilon_0. \tag{1.119}$$

We prove next that $\sigma(t)$ is a decreasing function. First we decompose $\sigma(t)$ as

$$\sigma(t) = E\varepsilon_0 \frac{\tau_\varepsilon}{\tau_\sigma} + \sigma_R(t). \tag{1.120}$$

Then from (1.117) it follows that

$$\begin{aligned} \bar{\sigma}_R &= -E\varepsilon_0 \frac{\tau_\varepsilon}{\tau_\sigma} \left(\frac{1}{\tau_\sigma} - \frac{1}{\tau_\varepsilon}\right) \left\{ \frac{1}{s} \frac{1}{s^\gamma + \frac{1}{\tau_\sigma}} + \frac{s^{\gamma-1}}{s^\gamma + \frac{1}{\tau_\sigma}} \frac{1}{s + \frac{1}{\tau_\sigma}} \right. \\ &\quad \left. - \frac{1}{s^\gamma + \frac{1}{\tau_\sigma}} \frac{1}{s + \frac{1}{\tau_\sigma}} \right\}. \end{aligned} \tag{1.121}$$

Inversion of (1.121) leads to

$$\begin{aligned} \sigma_R(t) &= -E\varepsilon_0 \frac{\tau_\varepsilon}{\tau_\sigma} \left(\frac{1}{\tau_\sigma} - \frac{1}{\tau_\varepsilon}\right) \left\{ \int_0^t e_{\gamma,\gamma}(\tau; \frac{1}{\tau_\sigma}) d\tau \right. \\ &\quad \left. + \int_0^t e^{-\frac{1}{\tau_\sigma}(t-\tau)} \left[E_\gamma(-\frac{1}{\tau_\sigma}\tau^\gamma) - e_{\gamma,\gamma}(\tau; \frac{1}{\tau_\sigma}) \right] d\tau \right\}, \end{aligned} \tag{1.122}$$

where $E_\gamma(t) = E_{\gamma,1}(t)$. With (1.120) and (1.122) we have

$$\begin{aligned} \sigma(t) &= E\varepsilon_0 \frac{\tau_\varepsilon}{\tau_\sigma} - E\varepsilon_0 \frac{\tau_\varepsilon}{\tau_\sigma} \left(\frac{1}{\tau_\sigma} - \frac{1}{\tau_\varepsilon}\right) \left\{ \int_0^t \left(1 - e^{-\frac{1}{\tau_\sigma}(t-\tau)}\right) e_{\gamma,\gamma}(\tau; \frac{1}{\tau_\sigma}) d\tau \right. \\ &\quad \left. + \int_0^t e^{-\frac{1}{\tau_\sigma}(t-\tau)} E_\gamma(-\frac{1}{\tau_\sigma}\tau^\gamma) d\tau \right\}. \end{aligned} \tag{1.123}$$

Since $E_\gamma(-\frac{1}{\tau_\sigma}t^\gamma) = e_{\gamma,1}(t; \frac{1}{\tau_\sigma})$, there follows that $E_\gamma(-\frac{1}{\tau_\sigma}t^\gamma)$ is a monotone (positive) function. Therefore from (1.123) we conclude that $\sigma^{(1)}(t) < 0$, i.e. $\sigma(t)$ is a decreasing function.

Remark 2 *If the term d in (1.100) is neglected we obtain a solution to (1.53) with $\alpha = \beta$. By setting the corresponding terms to zero we have, instead of (1.123)*

$$\sigma_{d=0} = E\varepsilon_0 \left(\frac{\tau_\varepsilon}{\tau_\sigma} - \frac{\tau_\varepsilon}{\tau_\sigma} \left(\frac{1}{\tau_\sigma} - \frac{1}{\tau_\varepsilon} \right) \int_0^t e_{\gamma,\gamma}(\tau; \frac{1}{\tau_\sigma}) d\tau \right). \tag{1.124}$$

Let $\Delta = \sigma - \sigma_{d=0}$. From (1.123),(1.124) it follows that $\Delta(0) = 0$. To determine $\lim_{t\to\infty} \Delta(t)$ note that from (1.117) it follows that

$$\bar{\Delta} = E\varepsilon_0 \left(1 - \frac{\tau_\varepsilon}{\tau_\sigma} \right) \frac{1}{s} \frac{s^\gamma - s}{(\tau_\sigma s^\gamma + 1)\left(s + \frac{1}{\tau_\sigma} \right)}. \tag{1.125}$$

By using (1.125) we conclude that $\lim_{s\to 0} s\bar{\Delta}(s) = 0$. Therefore $\lim_{t\to\infty} \Delta(t) = 0$ and thus, the difference between $\sigma(t)$ determined by (1.100) and $\sigma_{d=0}$ is zero at $t = 0$ and tends to zero when $t \to \infty$.

1.3.4 Conclusions

1. By using the internal variable approach we formulate a constitutive equation for a viscoelastic body in isothermal uniaxial deformation. It is assumed that the state of the body is described by two variables, strain (ε), and an internal variable (ξ). The time evolution (balance equation) of the internal variable is assumed to be in the form

$$\xi^{(\gamma)} = \alpha\Theta + X, \tag{1.126}$$

and X is chosen so that

$$\xi^{(1)} = P_\xi = \alpha\Theta. \tag{1.127}$$

The functional X is given by (1.80). Note that $\xi^{(1)}$ is local in time when expressed in terms of $\varepsilon(t)$ and $\xi(t)$. However when $\xi^{(1)}$ is expressed in terms of $\varepsilon(t)$ *only,* it is not local in time (see (1.88)).

2. The constitutive equation that follows from (1.126) is of fractional derivative type, see (1.100),(1.101). In the form solved for stress the constitutive equation is given by (1.111). Equation (1.111) satisfies the entropy inequality for all deformations $\varepsilon(t)$ and the stability condition of the equilibrium state (minimum of Gibbs free energy in dead loading).

3. To compare (1.111) with the generalized Zener model (in an arbitrary deformation process) we solve (1.53) with $\alpha = \gamma, \beta = \gamma, b = \tau_\sigma, E = E_0, E_1 = E\tau_\varepsilon$ for stress. The Laplace transform of (1.53) leads to

$$\begin{aligned} \bar{\sigma}_z &= \bar{\varepsilon} \frac{E[\tau_\varepsilon s^\gamma + 1]}{\tau_\sigma s^\gamma + 1} = E \frac{\tau_\varepsilon}{\tau_\sigma} \frac{s^\gamma + \frac{1}{\tau_\varepsilon}}{s^\gamma + \frac{1}{\tau_\sigma}} \bar{\varepsilon} \\ &= E \frac{\tau_\varepsilon}{\tau_\sigma} \left[\bar{\varepsilon} + \left(\frac{1}{\tau_\varepsilon} - \frac{1}{\tau_s} \right) \frac{1}{s^\gamma + \frac{1}{\tau_\sigma}} \bar{\varepsilon} \right]. \end{aligned} \tag{1.128}$$

By use of $\mathcal{L}^{-1}\left(1/\left(s^{\gamma}+\frac{1}{\tau_{\sigma}}\right)\right)=e_{\gamma,\gamma}(t;\frac{1}{\tau_{\sigma}})$ and of the convolution theorem it follows that

$$\sigma_{z}(t)=E\frac{\tau_{\varepsilon}}{\tau_{\sigma}}\varepsilon+\frac{1}{\tau_{\sigma}}E(1-\frac{\tau_{\varepsilon}}{\tau_{\sigma}})\int_{0}^{t}e_{\gamma,\gamma}(\tau;\frac{1}{\tau_{\sigma}})\varepsilon(t-\tau)\,d\tau. \tag{1.129}$$

Equation (1.129) may be obtained from (1.111), if we neglect the terms containing the factor $e^{-\frac{1}{\tau_{\sigma}}(t-\tau)}$ under the integral sign. The difference between $\sigma(t)$ and $\sigma_{z}(t)$ in a stress relaxation test is zero for $t=0$ and when $t\rightarrow\infty$ (see Remark 2). Thus (1.111) has the same asymptotic behavior as the generalized Zener model (1.53) or (1.129). The most important difference between (1.129) and (1.111), with the restriction on the coefficients (1.52), is that (1.111) satisfies the entropy inequality and the condition of stability of equilibrium for all deformation processes $\varepsilon(t)$.

References

[1] Atanackovic, T. M., *Stability theory of elastic rods*. World Scientific, River Edge 1997.

[2] Atanackovic, T. M., A model for the uniaxial isothermal deformation of a Viscoelastic body. *Acta Mechanica*, **159**, 77-86 (2002).

[3] Atanackovic, T. M., A modified Zener model of a viscoelastic body. *Continuum Mech. Thermodyn.,* **14**, 137-148 (2002).

[4] T. M. Atanackovic, M. Budincevic, S. Pilipovic, On a Fractional distributed-order oscillator. J. *Phys. A: Math. Gen.* **38**, 6703-6713 (2005).

[5] Atanackovic, T. M., and Stankovic, B., Stability of an Elastic rod on a Fractional Derivative type of a Foundation. *Journal of. Sound and Vibration*, **227**, 149-161 (2004).

[6] T. M. Atanackovic, S. Pilipovic, D. Zorica, Diffusion wave equation with two fractional derivatives of different order. *J. Phys. A: Math. Theor.* **40** (2007).

[7] Brogliato, B., 1999, *Nonsmooth dynamics*. Springer, London.

[8] Caputo, M., Linear Models of dissipation whose Q is almost frequency independent, Part II., *J. R. Astr. Soc.,* **13**, 529-539 (1967).

[9] Caputo, M., and Mainardi, F., Linear Models of dissipation in anelastic solids. *Rev. Nuovo Cimento*, (Ser. II) **1**, 161-198 (1971).

[10] Cattaneo C 1948 *Atti del Semin. Matem. Univ. di Modena* **3** 83

[11] Chen, W. F., and Atsuta, T., *Theory of Beam-Columns I,* New York, McGraw-Hill, 1976.

[12] Doetsch, G., *Handbuch der Laplace-Transformation, I and II,* Birkhäuser, Basel 1950,1955.

[13] Erdély A Magnus W Oberhettinger F and Tricomi F G 1954 *Tables of Integral Transforms Volume I* (New York: McGraw-Hill Book Company)

[14] R. Gorenflo, and F. Mainardi, Fractional calculus: Integral and Differential Equations of Fractional order. In *Fractals and Fractional Calculus in Continuum Mechanics,* A. Carpinteri and F. Mainardi Editors. Springer, Wien, 1997, 223-276.

[15] Gorenflo, R., Luchko, Yu., and Rogozin, S., Mittag-Leffler type functions: notes on growth properties and distributions of zeros. Preprint A-04/97 Fachbereich Mathematik und Informatik, Freie Universität, Berlin, 1997.

[16] Gorenflo, R. and Mainardi, F., On Mittag-Leffler-type functions in fractional evolution processes. *Journal of computational and applied mathematics*, **118**, 283-299 (2000).

[17] Jantarat, J. Palamara, J. E. A., Lindner,C., and Messer, H. H., Time dependent properties of a human root dentin. *Dental Materials,* **18**, 468-493 (2002).

[18] Kilchevski, N. A., 1976, *Dynamic contact of solid bodies,* Naukova Dumka, Kiev.

[19] Mainardi, F., Luchko, Yu., and Pagnini, G., The Fundamental Solution of The Space-Time fractional Diffusion Equation. *Fractional Calculus and Applied Analysis,* **4**, 153-192 (2001).

[20] Oldham, K. B., and Spanier, J., *The fractional Calculus.* Academic Press, New York 1974.

[21] Petrovic, Lj., Spasic, D. T., and Atanackovic, T. M., On a Mathematical Model of a Human root Dentin. *Dental Materials* (in press).

[22] Podlubny, I., *Fractional differential equations.* Academic Press, San Diego 1999.

[23] Rabotnov, Yu. N. Elements of hereditary Solid Mchnaics. Mir, Moscow, 1980.

[24] Samko, S. G., Kilbas, A. A., and Marichev, O. I., *Fractional Integrals and Derivatives.* Gordon and Breach, Amsterdam 1993.

[25] Seyranian, A. P., *Modern Problems of Structural Stability*, editors A. P. Seyranian and I. Elishakoff, Springer, Wien, 2002.

[26] Vladimirov, V. S. *Generalized Function in Mathematical Physics.* Moscow, Mir 1979.

Project: 06SER02/02/003

Dynamical geometry

Dr Dragoslav Herceg

1. Introduction

There are many DGS (Dynamic Geometry Software) and CAS (Computer Algebra System) software packages today, aimed at teaching geometry. The most well known DGS are Cabri, Cinderella, Euklides, Dynageo, Geometer‘s Sketchpad, Geonext, Zirkel und Lineal, Ruler and Compass:

- Cabri geometrie II+ (www.cabri.com),
- Cinderella (www.cinderella.de),
- DynaGeo (www.dynageo.de),
- The Geometer's Sketchpad (www.keypress.com/sketchpad),
- Geonext(www.geonext.de)
- Zirkel und Lineal (www.mathsrv.kueichstaett.de/MGF/homes/grothmann).

CAS software perform algebraic and numerical computations, and provide symbolic representation as well. Therefore they are well suited for teaching algebra and calculus. The most well known CAS are Derive, Mathematica, Maple, MuPad, MathCAD:

- Derive (www.chartwellyorke.com/derive.html)
- Mathematica (www.wolfram.com)
- Maple (www.maplesoft.com)
- MuPad (www.sciface.com/)
- MathCad (www.ptc.com/appserver/mkt/products/home.jsp?k=3901)

In today's classroom there is a need for a tool which can provide combined representations of algebra, geometry and calculus. We have found such a tool in GeoGebra.

„There is no true understanding in mathematics for students who do not incorporate into their cognitive architecture the various registers of semiotic representations used to do mathematics.“ (Duval 1999)

1.1. Awards

GeoGebra has been translated into 38 languages and has won many awards.

- EASA 2002: European Academic Software Award (Ronneby, Sweden)
- Learnie Award 2003: Austrian Educational Software Award (Vienna, Austria)
- digita 2004 : German Educational Software Award (Cologne, Germany)
- Comenius 2004: German Educational Media Award (Berlin, Germany)
- Learnie Award 2005: Austrian Educational Software Award for "Spezielle Relativitätstheorie mit GeoGebra" (Vienna, Austria)
- Trophées du Libre 2005: International Free Software Award, category Education (Soisson, France)
- eTwinning Award 2006: 1st prize for "Crop Circles Challenge" with GeoGebra (Linz, Austria) Learnie Award 2006: Austrian Educational Software Award for "Wurfbewegungen mit GeoGebra" (Vienna, Austria)
- AECT Distinguished Development Award 2008: Association for Educational Communications and Technology (Orlando, USA)

1.2. What is GeoGebra?

GeoGebra is a free dynamic mathematics software for elementary and middle schools that joins geometry, algebra and calculus. It is developed by Markus Hohenwarter at the University of Salzburg (until 2007), and at the Florida Atlantic University since then.

GeoGebra is:

- a dynamic mathematics software
- for students and teachers
- dynamic **geo**metry, al**gebra** and calculus
- open source (free of charge)

On the one hand, GeoGebra is a dynamic geometry system. You can do constructions with points, vectors, segments, lines, conic sections as well as functions and change them dynamically afterwards.

On the other hand, equations and coordinates can be entered directly. Thus, GeoGebra has the ability to deal with variables for numbers, vectors and points, finds derivatives and integrals of functions and offers commands like *Root* or *Extremum*.

These two views are characteristic of GeoGebra: an expression in the algebra window corresponds to an object in the geometry window and vice versa.

Advanced authors can add functionality to GeoGebra drawings by implementing JavaScript programs in Web pages.

1.3. What does GeoGebra offer?

- Points, vectors, segments, polygons, lines, all conic sections and functions in x
- Dynamic constructions using the mouse
- Coordinates, equations, vectors, numbers and commands (keyboard)
- Intuitive notation c: $(x - 3)^2 + (y + 2)^2 = 25$
- Easy-to-use interface
- Multilingual menus, commands, help
- Bidirectional combination of dynamic geometry and computer algebra
- High portability: runs on Windows, Linux, Solaris, MacOS X
- Dynamic worksheets using HTML and Java applets

1.4. GeoGebra for teachers

- Introduction to GeoGebra.
- Tools of GeoGebra.
- Creation of dynamic worksheets (HTML).
- GeoGebra as a presentation tool.
- Constructions including points, vectors, segments, polygons, lines, all conic sections and functions in x dynamic constructions (mouse), coordinates, equations, vectors, numbers and commands (keyboard).
- Visualization two vectors with their sum and difference and the vectors measured rectangular coordinates.

- Virtual Manipulative for Linear Equations.
- Graphing in plane. One can either use the Cartesian Grapher to graph y as a function of x, or use the Parameterized Grapher to graph x and y as functions of t.
- Review the basic ideas from trigonometry.

1.5. GeoGebra for students

- Introduction to GeoGebra.
- Tools of GeoGebra.
- Creation of dynamic worksheets (HTML).
- Applets are designed for use in calculus courses.
- Continuity. The Epsilon Delta Applet designed for a visual exploration of the delta-epsilon definition of continuity of functions in one variable. The user highlights an epsilon and delta band around a proposed limit of a function at a point. It is easy to zoom in or out, and the applet has a nice collection of pre-set examples.
- How the Taylor polynomials can be used to approximate functions. The base point and degree are controlled by sliders.
- Interpolation. Interpolations polynomial for a function and corresponding error.
- Integration. Lower- und Upper Sum of a Function.
- Numerical integration. Simpson's rule
- Numerical solution of equations in one variable.

2. Introuduction to GeoGebra

2.1. Why should GeoGebra be used in a classroom?

GeoGebra's dynamic behavior enables the teacher to draw a construction and alter it in real time. The construction can later be saved to a file, played back step by step, and textual explanations can be added. Therefore, GeoGebra can be used as a presentation tool.
Students can do mathematical experiments, which helps them discover mathematics on their own. Teachers can prepare problems in GeoGebra for students to solve, and guide them along the way. Many teachers have uploaded their teaching materials on the GeoGebra wiki (www.geogebra.at/wiki).

GeoGebra implements the didactic principles:

- Student activity orientation principle
- Interaction of representations principle (iconic and symbolic representations)
- Genetic principle (experimental learning of mathemtics)
- Activity based learning
- Iconic and Symbolic representations (Bruner)

2.2. Obtaining GeoGebra

GeoGebra is available for free download at www.geogebra.org. There are two versions to choose from:

- Webstart - use GeoGebra without having to install
- Download - GeoGebra Geometry installation package

3. Tools of GeoGebra

GeoGebra has many tools. Some of them will be demonstrated in this section. From version 3.0, GeoGebra has user-defined tools, which are easy to create and can perform complex construction tasks at a click of the mouse.

3.1. Examples

3.1.1. Triangle with Angles

Select mode *New point* in the toolbar. Click three times on the drawing pad to create the three vertices *A*, *B,* and *C* of the triangle.
Afterwards, select mode *Polygon* and successively click on points *A*, *B*, and *C*. To close triangle *P* click on the starting point *A* again. In the algebra window the area of the triangle is shown.
In order to get all the angles of a triangle, choose mode *Angle* in the toolbar and click on the triangle.
Now, choose *Move* mode and drag the vertices to modify the triangle dynamically. If you don't need the algebra window and coordinate axes, hide them by using the *View* menu.

3.1.2. Linear Equation y = m x + b

We will now concentrate on the meaning of *m* and *b* in the linear equation $y = mx + b$ by trying different values for *m* and *b*. To do this we might enter the following lines in the input field at the screen's bottom and press the *Enter* key at the end of each line.

```
m = 1
b = 2
y = m x + b
```

Now we can change *m* and *b* using the input field or directly in the algebra window by right clicking one of the numbers and selecting *Edit*. Try the following values for *m* and *b*.

```
m = 2
m = -3
b = 0
b = -1
```

Also, you can change *m* and *b* very easily using

- the arrow keys
- sliders: right click on *m* or *b* and select *Show / hide object*

In a similar way we might investigate the equations of conic sections such as

- ellipses: `x^2/a^2 + y^2/b^2 = 1`
- hyperbolas: `b^2 x^2 – a^2 y^2 = a^2 b^2` or
- circles: `(x - m)^2 + (y - n)^2 = r^2`

3.1.3. Centroid of Three Points A, B, and C

We are now going to construct the centroid of three points by entering the following lines in the input field and pressing the *Enter* key at the end of each line. Of course, you

can also use the mouse to do this construction using the corresponding modes in the toolbar.

```
A = (-2, 1)
B = (5, 0)
C = (0, 5)
M_a = Midpoint[B, C]
M_b = Midpoint[A, C]
s_a = Line[A, M_a]
s_b = Line[B, M_b]
S = Intersect[s_a, s_b]
```

Alternatively you can compute the centroid directly as `S1 = (A + B + C) / 3` and compare both results using the command `Relation[S, S1]`.

Subsequently we can experiment whether `S = S1` is true for other positions of *A*, *B*, and *C* as well. We do this by selecting mode *Move* with the mouse and dragging the points.

Split Line Segment AB at a Ratio of 7 : 3

As GeoGebra allows us to calculate with vectors, this is an easy task. Type the following lines into the input field and press the *Enter* key after each line.

```
A = (-2, 1)
B = (3, 3)
s = Segment[A, B]
T = A + 7/10 (B - A)
```

Another way of doing this could be

```
A = (-2, 1)
B = (3, 3)
s = Segment[A, B]
v = Vector[A, B]
T = A + 7/10 v
```

In a next step we could introduce a number *t*, e.g. by using a *Slider* and redefine point *T* as `T = A + t v`. By changing *t* you can see point *T* moving along a straight line which could now be entered in parametric form: `g: X = T + s v`

3.1.4. Set of Linear Equations in Two Variables

Two linear equations in *x* and *y* can be interpreted as two straight lines. The algebraic solution is the intersection point of these two lines. Just type the following lines into the input field and press the *Enter* key after each line.

```
g: 3x + 4y = 12
h: y = 2x - 8
S = Intersect[g, h]
```

To change the equations you can right click one of them and select *Edit*. Using the mouse you can drag the lines in *Move* or rotate them about a point using *Rotate around point*.

3.1.5. Tangent to a Function of x

GeoGebra offers a command for the tangent to a function *f(x)* at $x = a$. Type the following lines into the input field and press the *Enter* key after each line.

```
a = 3
f(x) = 2 sin(x)
t = Tangent[a, f]
```

By animating number *a* the tangent slides along the graph of function *f*.

Here is another way of getting the tangent to a function *f* in a certain point *T*.

```
a = 3
f(x) = 2 sin(x)
T = (a, f(a))
t: X = T + s (1, f'(a))
```

This gives us point *T* on the graph of *f* whereby tangent *t* is given in parametric form.

By the way, you can create the tangent of a function geometrically too:

- Select mode *New point* and click on the graph of function *f* to get a new point *A* that lies on function *f*.
- Select mode *Tangents* and click successively on function *f* and on point *A*.

Now, select *Move* and drag point *A* along the function with your mouse. In this way you can observe the tangent changing dynamically too.

3.1.6. Investigation of Polynomial Functions

With GeoGebra you can investigate roots, local extreme, and inflection points of polynomial functions. Type the following lines into the input field and press the *Enter* key after each line.

```
f(x) = x^3 - 3 x^2 + 1
R = Root[f]
E = Extremum[f]
I = InflectionPoint[f]
```

In *Move* you can drag function *f* with the mouse now. In this context, the first two derivatives of *f* could be interesting too. You get them by typing the following command into the input field and hitting the *Enter* key after each line.

```
Derivative[f]
Derivative[f, 2]
```

3.1.7. Integrals

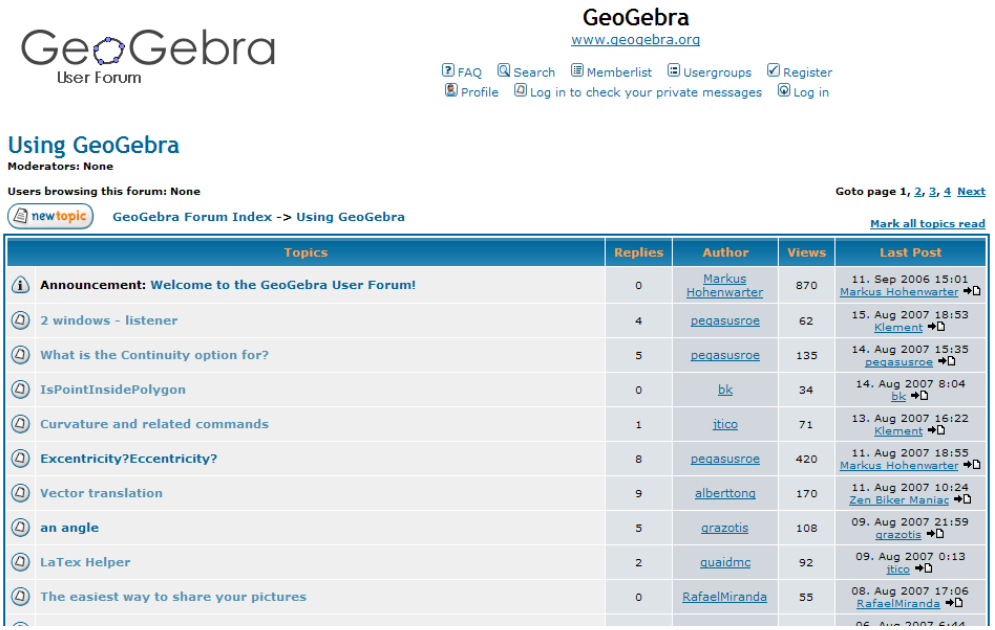
To introduce integrals, GeoGebra offers the possibility to visualize lower and upper sums of a function as rectangles. Type the following lines into the input field and press the *Enter* key after each line.

```
f(x) = x^2/4 + 2
a = 0
b = 2
n = 5
L = LowerSum[f, a, b, n]
U = UpperSum[f, a, b, n]
```

By modifying *a*, *b*, or *n* you can see the impact of these parameters on the upper and lower sum. In order to change the increment of number *n* to 1 you can right click on number *n* and select *Properties*.

The definite integral can be shown using the command `Integral[f, a, b]`, while the antiderivative *F* is created using `F = Integral[f]`.

3.2. GeoGebra Homepage



GeoGebra User Forum

GeoGebra
www.geogebra.org

FAQ Search Memberlist Usergroups Register
Profile Log in to check your private messages Log in

Using GeoGebra

Moderators: None

Users browsing this forum: None

Goto page 1, 2, 3, 4 Next

newtopic GeoGebra Forum Index -> Using GeoGebra

Mark all topics read

Topics	Replies	Author	Views	Last Post
Announcement: Welcome to the GeoGebra User Forum!	0	Markus Hohenwarter	870	11. Sep 2006 15:01 Markus Hohenwarter
2 windows - listener	4	pegasusroe	62	15. Aug 2007 18:53 Klement
What is the Continuity option for?	5	pegasusroe	135	14. Aug 2007 15:35 pegasusroe
IsPointInsidePolygon	0	bk	34	14. Aug 2007 8:04 bk
Curvature and related commands	1	jtico	71	13. Aug 2007 16:22 Klement
Excentricity?Eccentricity?	8	pegasusroe	420	11. Aug 2007 18:55 Markus Hohenwarter
Vector translation	9	alberttong	170	11. Aug 2007 10:24 Zen Biker Maniac
an angle	5	grazotis	108	09. Aug 2007 21:59 grazotis
LaTex Helper	2	quaidmc	92	09. Aug 2007 0:13 jtico
The easiest way to share your pictures	0	RafaelMiranda	55	08. Aug 2007 17:06 RafaelMiranda

GeoGebra User Forum

GeoGebra
www.geogebra.org

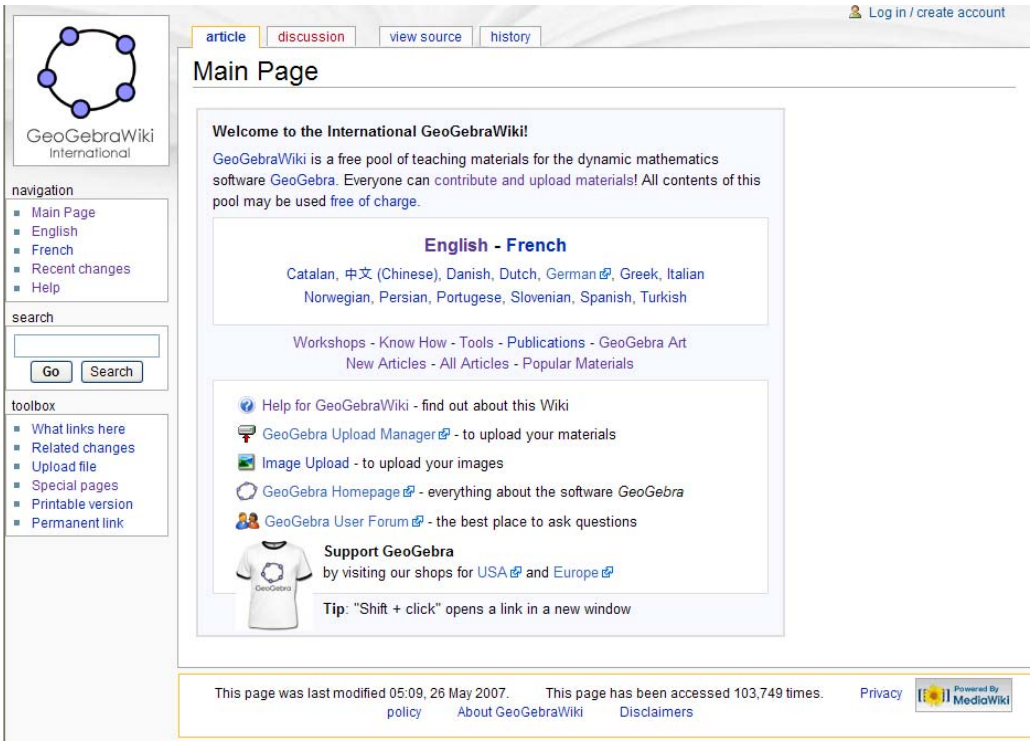
FAQ Search Memberlist Usergroups Register
Profile Log in to check your private messages Log in

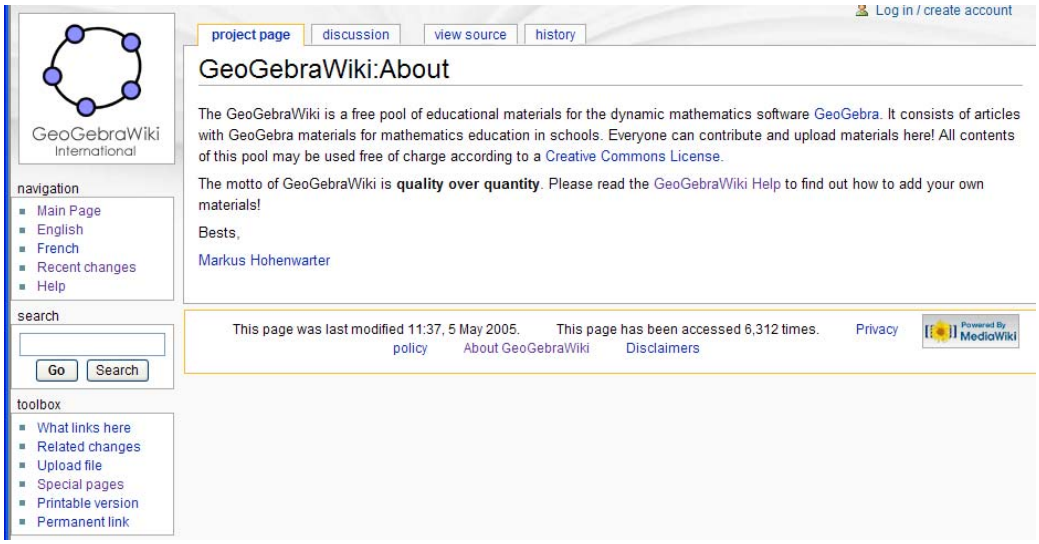
Change Image with JavaScript?

newtopic postreply GeoGebra Forum Index -> Technological Questions

View previous topic :: View next topic

Author	Message
rfant	Posted: 01. Jul 2007 21:15 Post subject: Change Image with JavaScript? quote
Joined: 14 Oct 2006 Posts: 19 Location: Dodge City, Kansas	Can the background image be changed via JavaScript? or Is there an easy way to have multiple images loaded but only have one showing at a time? Thanks, Robert
Back to top	profile pm www
simes	Posted: 02. Jul 2007 12:04 Post subject: change images quote
Joined: 22 Aug 2005 Posts: 3	Is it enough? http://www.geogebra.org/en/upload/files/english/sime_hr/forum/images.html





3.3. References

- GeoGebra – the dynamic mathematics software
  www.geogebra.at
- GeoGebraWiki – the interactive pool of materials
  www.geogebra.at/wiki
- GeoGebra User Forum
  www.geogebra.at/forum
- Broward County Teachers that are proficient in GeoGebra:
  nsfmsp.fau.edu/geogebra
- User Forum www.geogebra.org/forum
- www.geogebra.org – get the free mathematics software GeoGebra
- www.geogebra.org/talks/200610_fctm.zip - presentation, dynamic worksheet, and other materials

4. Numerical Mathematics with GeoGebra

4.1. Introduction

We have prepared a suite of motivational examples which illustrate numerical methods for equation solving. Fixed point iteration, Newton's method, secant method and regula falsi method are implemented as GeoGebra tools. Our experience in teaching of numerical mathematics in "Jovan Jovanović Zmaj" high school in Novi Sad is presented. We have tested pupil proficiency in numerical equation solving with and without use of a computer and the results are presented.

In this paper we present solving equations in GeoGebra. For that purpose, we have developed several GeoGebra tools, as well as a collection of examples, which are available at http://www.im.ns.ac.yu/personal/hercegd/cadgme2007.

Numerical equation solving in high school consists of 14 lessons, which cover the following units:

- Localization of roots
- Approximation of a root
- Bisection method
- Newton-Raphson method
- Secant method and regula falsi method
- Fixed point iteration

Without a computer, these lessons are presented on a whiteboard and by solving only the basic examples. The teacher must carefully choose the problems and solve them completely before class. In the class, the teacher can only present those and similar problems, which can be solved by using well-known properties of elementary mathematical functions.

In order to maximize the positive effects of a lesson, it is essential that the pupils engage actively with their learning. Teaching that accomplishes the maximum effect is interactive, direct and collaborative, and, most of all, interesting. We hope to have achieved this by combining several approaches and tools, which we developed in GeoGebra and Mathematica.

In this paper we present three basic numerical methods for equation solving: bisection method, Newton-Raphson method and regula falsi.

4.1.1. The classroom and the software

We teach numerical mathematics in a classroom with 18 PC workstations and one pupil at each workstation. The workstations are networked and have Internet access, as well as USB ports and a CD/DVD drive. The teacher’s computer is connected to a data projector and is also used by pupils for demonstrating their work in front of the class. This computer is also a classroom file server. The whiteboard is used for writing down key steps in equation solving. In that way important information is retained, even when the projected picture changes.

The teacher uploads the examples and tools to the file server, thus allowing the pupils to use them during class and exams, when appropriate.

Pupils are allowed to talk to each other and use the Internet, as long as it is about solving the assigned problems. It is important to keep the pupils on track, and sometimes we achieve this by blocking their access to the Internet or specific sites.

We keep record of pupil attendance and classroom activity, which influences their grades. The grading process is transparent, i.e. the pupils are allowed to track their progress and they always know their current standings.

In class we used GeoGebra 3.0 and Mathematica 5. While Mathematica satisfies most of our computing needs, it lacks interactivity, which GeoGebra supplements in more than one way. Besides, GeoGebra is much simpler to use, and it is available in many languages, including pupils' native Serbian. Of no lesser importance is the fact that GeoGebra is free, while Mathematica is not.

GeoGebra 3.0, which has appeared only recently, brings some important improvements, such as user-defined tools, which have enabled us to implement tools numerical methods for equation solving.

4.1.2. Localization of roots

Localization of roots without a computer can be a tedious job. Teachers then usually choose simple examples. For example, to solve the equation $f(x)=0$, where

$$f(x)=\ln(x+3)-2\sin x,$$

we can begin by graphing the function. This graph is not easy to produce without a computer, or at least a calculator. It is much easier to draw a graph of the two functions

$$g(x)=\ln(x+3) \text{ and } h(x)=2\sin x,$$

and to look for intersection points, since $f(x)=0$ is equivalent to $g(x)=h(x)$. The pupils can easily graph the elementary functions $g(x)$ and $h(x)$ (Figure 1).

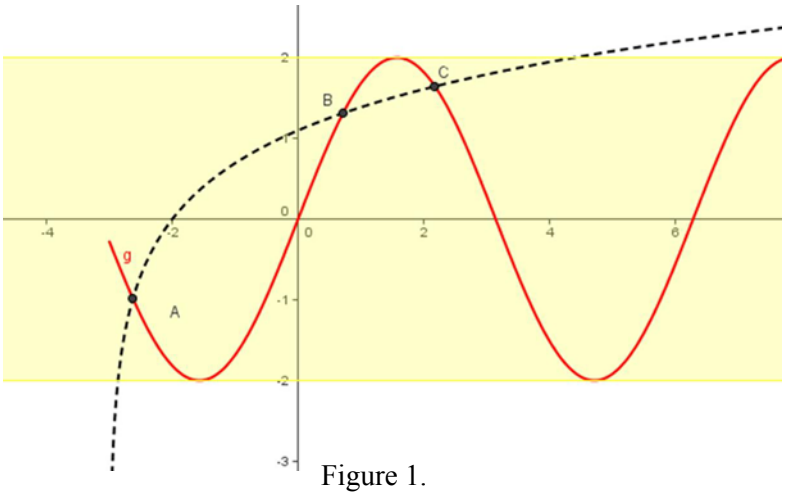


Figure 1.

Determining the intervals which contain zeros can be a tedious job. It is also time consuming and requires a rather long series of steps, where possibility of error increases with each step. The pupils usually lack the needed skill and experience, and are easily discouraged at the first failure, which quickly leads to a loss of interest on their part.

With a computer, pupils can experiment with varying the interval bounds and find where some of the roots are. To find out how many zeros a function has and where they all are requires a detailed analysis of the function.

In GeoGebra, we can easily find and count the zeros of a function by panning and zooming. GeoGebra enables us to analyze more complex examples, like these:

$$2\log x-\frac{1}{2}x+1=0,\qquad x-\sin x-\frac{1}{4}=0,\qquad x^2-20\sin x=0,$$

$$2x\ln x = 1, \quad 2^x - 4x = 0, \quad x^2 - \sin(\pi x) = 0.$$

Let us consider the equation $x^2 - 20\sin x = 0$. In Figure 2 it is not clear whether there are two or three real roots in the interval $(-8,8)$.

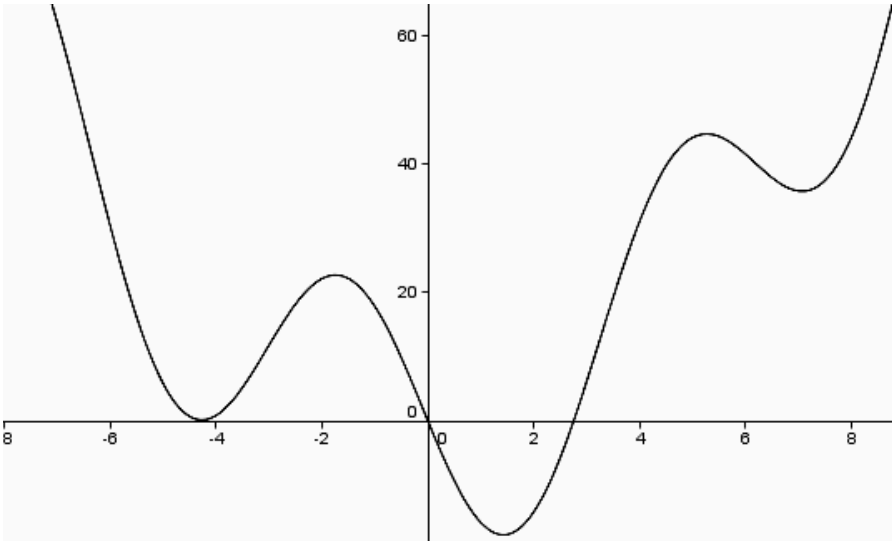


Figure 2. How many zeros does the function have?

By zooming in at the interval $(-5.6,-3.6)$ we see that the curve does not touch the x-axis, therefore we conclude that there are only two real roots in the interval $(-8,8)$.

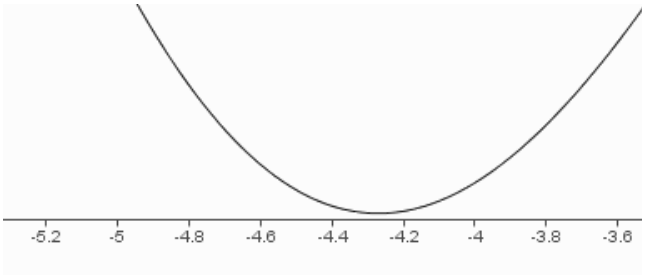


Figure 3. Zooming in at the possible zero

Using GeoGebra, we can look into more complex examples such as:

$$e^{\sin x} - 1 = 0, \quad 2\sin(\cos x) = 1$$

Simply by moving the function graph by mouse, teacher can modify the problem to get a function with no roots or one with an infinite number of roots. The pupils are encouraged to experiment and investigate new cases and to draw conclusions.

Similarly, we can observe a function which depends on several parameters. For example, the function

$$f(x) = a\, e^{b\sin(c\,x)} + d$$

has four real parameters a, b, c and d. The parameters are controlled by sliders, and by changing their values many interesting exercises are created. We can look for all roots which belong to a given interval, a maximum root, etc.

By zooming in we can read the approximate value for a zero from the intersection point of the graph and the x-axis.

The tool "New point" can be used to read the coordinates from the graph, simply by moving the point along the curve. At the intersection of the function with the x-axis we can read

the approximate value of a real zero of the function. We can then use the intersection tool to set a point there. The point's x coordinate is the approximate value of a zero.

Similarly, a root can be localized at the intersection of two graphs. Here we can emphasize parts of the drawing using different colors and line styles.

We can introduce parameters into the problems, and assign unique tests to each pupil or group of pupils. That way, we can produce an infinite number of examples.

4.2. GeoGebra tools for numerical methods

One of our goals was to facilitate understanding of principles of common iterative methods for numerical equation solving. Recent development of GeoGebra and introduction of user defined tools has made this possible. While in earlier versions of GeoGebra it was neccessary to manually construct tangents, secant lines and midpoints of intervals, now we can simply define a tool and let the pupils use it. The pupils who do not understand geometrical interpretation of iterative methods can still use the tools.

We have developed several GeoGebra tools which perform one iteration of the following iterative methods:

- Bisection method,
- Regula-falsi (secant) method,
- Newton-Raphson method.

The tools are based on geometrical representation of these iterative methods. By repeatedly applying a tool at the result of the previous iteration, we can simulate the working of an iterative method.

4.2.1. Bisection Method

One of the first numerical methods developed for finding roots of a nonlinear equation $f(x) = 0$ was the bisection method. This is one of the simplest methods and is based on the following theorem.

Theorem. An equation $f(x) = 0$, where $f(x)$ is a real continuous function in $[a,b]$, has at least one root between a and b if $f(a)f(b) < 0$.

To find a root of $f(x) = 0$ using this method, the first thing to do is to find an interval $[a,b]$ such that $f(a)f(b) < 0$. Bisect this interval to get a point $(c, f(c))$. Choose one of a or b so that the sign of $f(c)$ is opposite to the sign of ordinate at that point. Use this as the new interval and proceed until you get the root within desired accuracy.

The Bisection tool performs one step of the bisection method. It requires the following input: a function which has real zeros and two points *a* and *b* on the *x*-axis. It produces a new point *c* on the *x*-axis with the coordinates $\left(\frac{b-a}{2}, 0\right)$ and it marks the part of the interval which does not contain the zero with a red line.

Example. We shall apply the bisection tool to the equation $2\sin(\cos x) = 1$. First, we need to type in the definition f(x)=2 sin(cos(x))-1 in GeoGebra's input box. This displays a graph of the function $f(x)$. We obtain the "exact" solution of the equation with the Intersect tool, by finding the intersection point A of the curve with the *x*-axis. Then we select the Bisection tool from the toolbar. In order to perform one step of the bisection method, we need to click on the function in the geometry window, and then click two times on the *x*-axis, defining the points B and C, which will mark the initial interval (Figure 4). The Bisection tool generates a midpoint D, which will be used in the next step. .

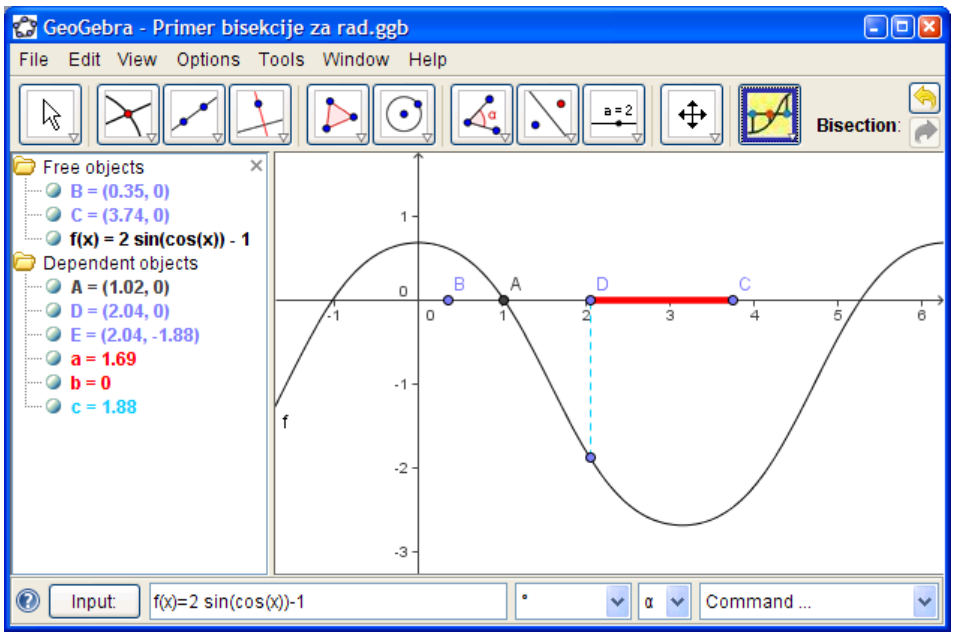


Figure 4.

We continue by again clicking on the function, and, this time, by clicking on the points B and D, since the interval [B,D] contains the root of the equation.

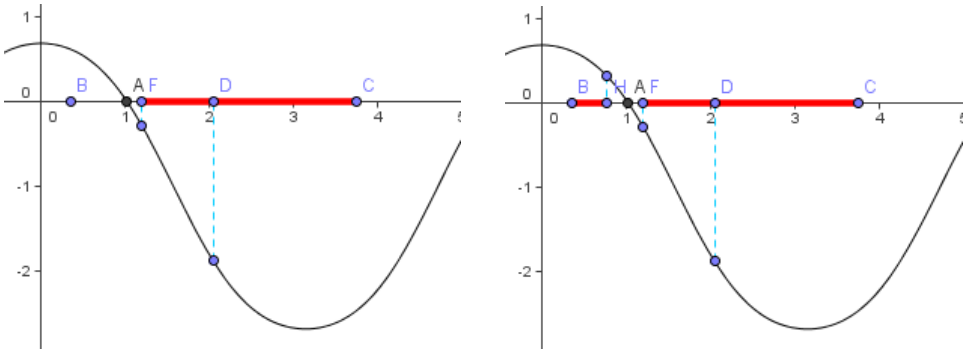


Figure 5. The second and third step of the bisection method

Figure 6 shows the interval which contains a zero after five steps of the bisection method. We need to use the zoom in tool in order to see it.

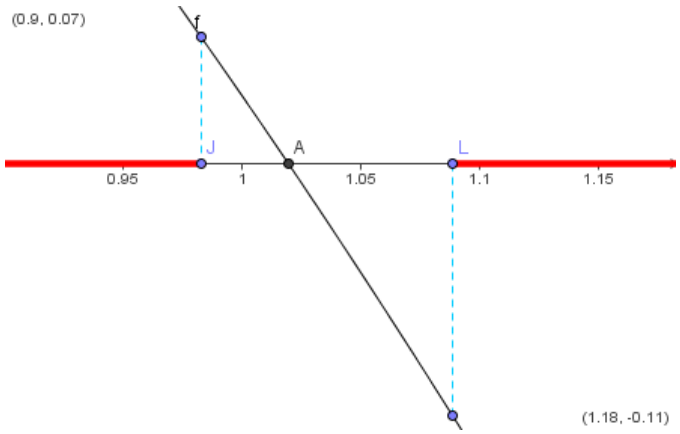


Figure 6. The interval after five iterations

New points, obtained as a result of the bisection tool, are the approximations of the root. In Figure 7, these points are named D, F, H, J, L in the algebra window.

When the interval which contains a zero becomes too small to see, we need to use the zoom in tool in order to get a better view of the interval. Since GeoGebra can display up to five

decimal digits, we can get a rather precise approximation of the root. This applies to other numerical methods as well.

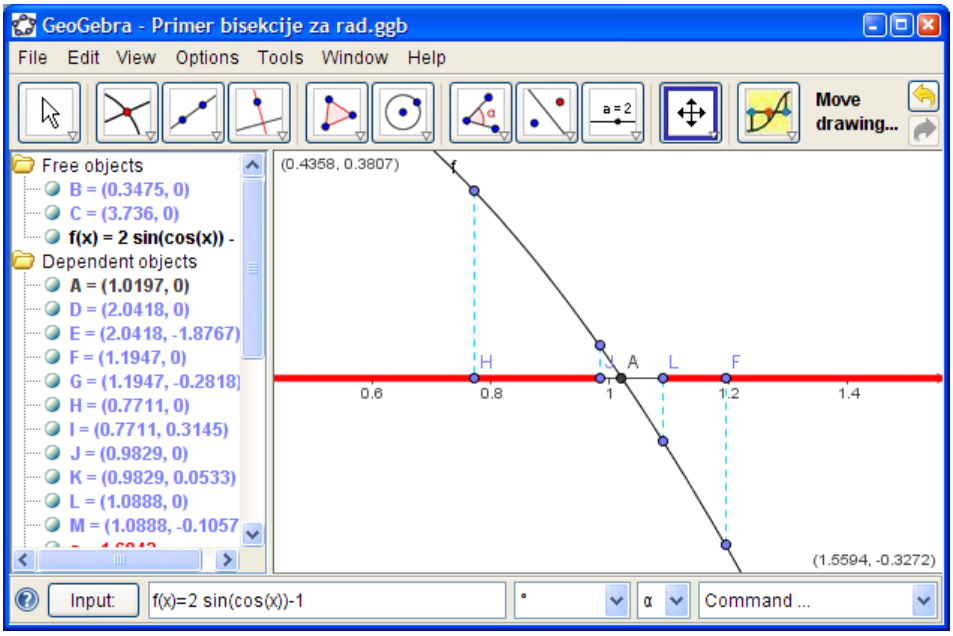


Figure 7. Approximations shown with 5 digits

4.2.2. Newton-Raphson

The Newton-Raphson method of solving the nonlinear equation $f(x)=0$ is given by the recursive formula

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

One of the drawbacks of the Newton-Raphson method is that the derivative of the function needs to be evaluated for every iteration. With availability of symbolic mathematics software such as Mathematica and GeoGebra, this process has become more convenient. However, it still can be a laborious process.

For Newton-Raphson method only one initial approximation of the root is needed to get the iterative process started to find the root of an equation. This method is based on the principle that if the initial guess of the root of $f(x)=0$ is at x_0, then if one draws the tangent to the curve at $f(x_0)$, the point x_1 where the tangent crosses the x-axis is an improved estimate of the root (Figure 8). Tangent line is given by

$$f'(x_0)(x-x_0)+f(x_0).$$

The solution x_1 of equation

$$f'(x_0)(x-x_0)+f(x_0)=0$$

is given by

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

So starting with an initial guess x_0 one can find the next approximatons $x_1, x_2, \ldots$, until the root within a desirable tolerance is found.

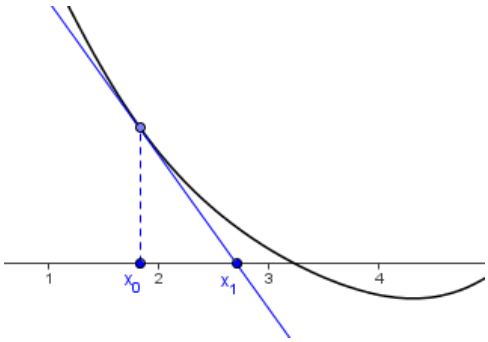


Figure 8. Geometrical representation of the Newton-Raphson method

The steps to apply Newton-Raphson method to find the root of an equation $f(x)=0$ are:

1. Evaluate $f'(x)$ symbolically
2. Use an initial guess of the root, x_0, to estimate the new value of the root x_1.
3. Repeat step 2, using the value x_i to obtain a new value x_{i+1}.

The NewtonRaphson tool performs one step of the Newton-Raphson method. It requires the following input: a function and an initial point on the *x*-axis. It then produces a tangent to the function and a new approximation in the intersection point of the tangent and the *x*-axis.

Example. Let us consider the equation $x+1.5-\sin(x+1.5)+0.06=0$. We start by typing in f(x) = x + 1.5 - sin(x + 1.5) + 0.06 into GeoGebra's input box. . This displays a graph of the function $f(x)$. We obtain the "exact" solution of the equation by intersecting the curve with the *x*-axis. Now let us select the NewtonRaphson tool from the toolbar. First we click on the function, then on the *x*-axis near the point (1,0). The tool then generates point C, which is the next approximation of the root of the equation. Repeating the procedure, we can perform several steps of the Newton-Raphson method (Figure 9).

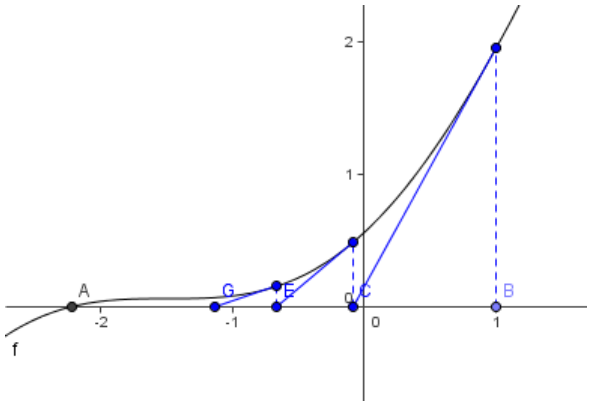


Figure 9. Three steps of the Newton-Raphson method

4.2.3. Regula falsi method

Methods such as bisection method and the false position method of finding roots of a nonlinear equation $f(x)=0$ require bracketing of the root by two guesses. These methods are always convergent since they are based on reducing the interval between the two guesses to zero in on the root.

In the regula falsi method, we start with two initial points, x_0 and x_1, such that $f(x_0)f(x_1)<0$ so that $f(x)=0$ has a solution α between x_0 and x_1. We assume that α is

the unique solution to $f(x)=0$ between x_0 and x_1. The new approximation x_2 is the point of intersection of the straight line passing through $(x_0, f(x_0))$ and $(x_1, f(x_1))$ with the x-axis:

$$x_2 = x_1 - \frac{f(x_1)(x_1 - x_0)}{f(x_1) - f(x_0)}.$$

If $f(x_2)=0$ then $\alpha = x_2$ and we stop. If $f(x_0)f(x_2)<0$, then we leave x_0 unchanged and continue to the next iteration; otherwise, we set $x_0 = x_1$ and continue to the next iteration in the same way.

In case f' and f'' have fixed signs in an interval containing α, which is the situation of interest to us here, the point x_0 ultimately remains fixed. Therefore, in such a case, the regula falsi method becomes a fixed-point method at some point during the iteration process. Without loss of generality, we will assume that x_0 remains fixed. In this case the regula falsi method is given by

$$x_{k+1} = x_k - \frac{f(x_k)(x_k - x_0)}{f(x_k) - f(x_0)}.$$

The Secant tool performs one step of the regula falsi method. It requires the following input: a function which has real zeros and two points on the x-axis. The tool produces a secant line between the given points and an intersection point with the x-axis. It also generates the point x_{k+1}, which is used as a starting point for a new iteration.

Example. Figure 10 shows three steps of the regula falsi method applied to the equation $e^{\sin(x-3)} - 0.5x = 0$. The Secant tool is used similarly to the Bisection and NewtonRaphson tools. Point A is the root of the equation. Points B and C determine the initial interval.

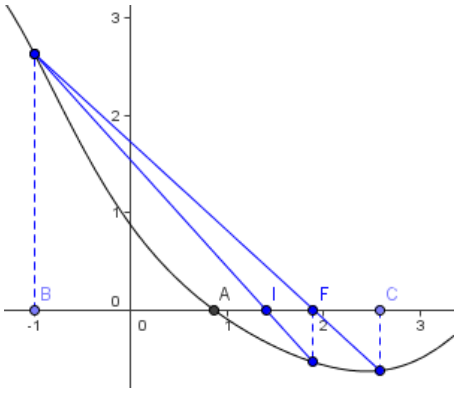


Figure 10. First three steps of the regula falsi method

4.3. Conclusion

The advantages introduced by a computer are not only in quicker calculation and drawing. The computer does all the tedious work, which leaves the teacher and the pupils with enough time to discuss the problem, try out multiple ideas and approaches to solving, and, finally, compare and analyze them. The method of solving a problem is as important as its solution. Use of a computer is particularly important when working with pupils who have difficulties understanding all the aspects of solving a mathematical problem. They are freed from uninspiring and time-consuming solving by hand, so they have more time to learn the important points.

GeoGebra has revolutionarized the way we teach numerical mathematics. It enabled us to do geometrical constructions and animate them easily. It is simple, powerful, and, most

importantly, it is free. Compared to other symbolic mathematics software, such as Mathematica, the price/usability ratio of GeoGebra is without a match.

Testing results show that the pupils who were instructed to use GeoGebra have achieved better scores. However, not all the pupils were independent enough to complete the tests without teacher's help and guidance. This confirms the fact that no mathematical software can replace the teacher.

4.4. References

[1] Herceg, D., Herceg, Đ., *Numerička matematika*, Stylos, Novi Sad, 2003.

[2] Uhl, J., Davis, W., *Is the Mathematics We Do the Mathematics We Teach?*, Contemporary Issues in Mathematics Education, MSRI Publications, Vol. 36 (1999), 67-74.

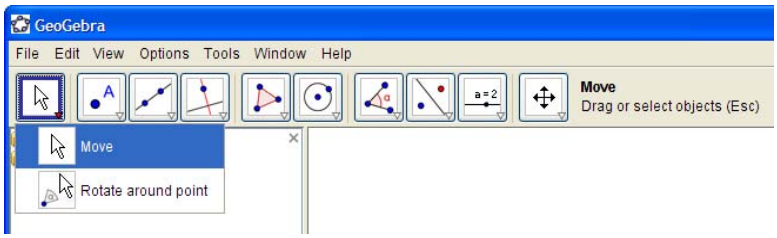
[3] Hohenwarter, M., *GeoGebra - didaktische Materialien und Anwendungen fur den Mathematikunterricht*, Dissertation, Naturwissenschaftliche Fakultät der Universität Salzburg, Salzburg, 2006.

5. GeoGebra tools reference

5.1. Geometric Input

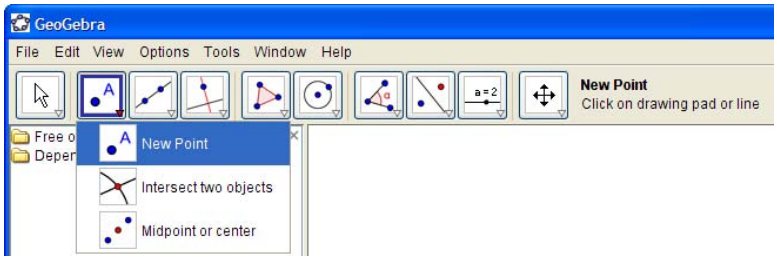
5.2. Modes

Button	Name	Description
	Move	In this mode you can drag and drop free objects with the mouse. If you select an object by clicking on it in  *Move* mode, you may delete it by pressing the *Del* key move it by using the arrow keys. Pressing the *Esc* key activates the *Move* mode too. By holding the *Ctrl* key you can select several objects at the same time. Another way of selecting multiple objects is by pressing and holding the left mouse key in order to specify a selection rectangle. You may then move the selected objects by dragging one of them with the mouse. The selection rectangle can also be used to specify a part of the graphics window for printing, exporting pictures, and for dynamic worksheets.
	Rotate around point	Select the centre point of the rotation first. Afterwards you may rotate free objects around this point by dragging them with the mouse.



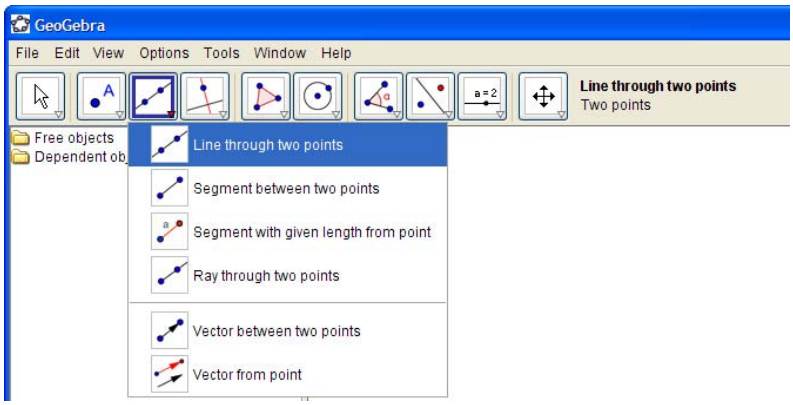
5.3. Points

	New point	Clicking on the drawing pad creates a new point. The coordinates of the point are fixed when the mouse button is released. By clicking on a segment, straight line, polygon, conic section, function, or curve you create a point on this object. Clicking on the intersection of two objects creates this intersection point.
	Intersect two objects	Intersection points of two objects can be produced in two ways. If you…mark two objects *all intersection points* are created (if possible). click on an intersection of the two objects only this *single intersection point* is created. For segments, rays, or arcs you may specify whether you want to *allow outlying intersection points*. This can be used to get intersection points that lie on the extension of an object. For example, the extension of a segment or a ray is a straight line.
	Midpoint or centre	Click on ... two points to get their midpoint, one segment to get its midpoint, a conic section to get its centre.



5.4. Line, segment, ray, vector

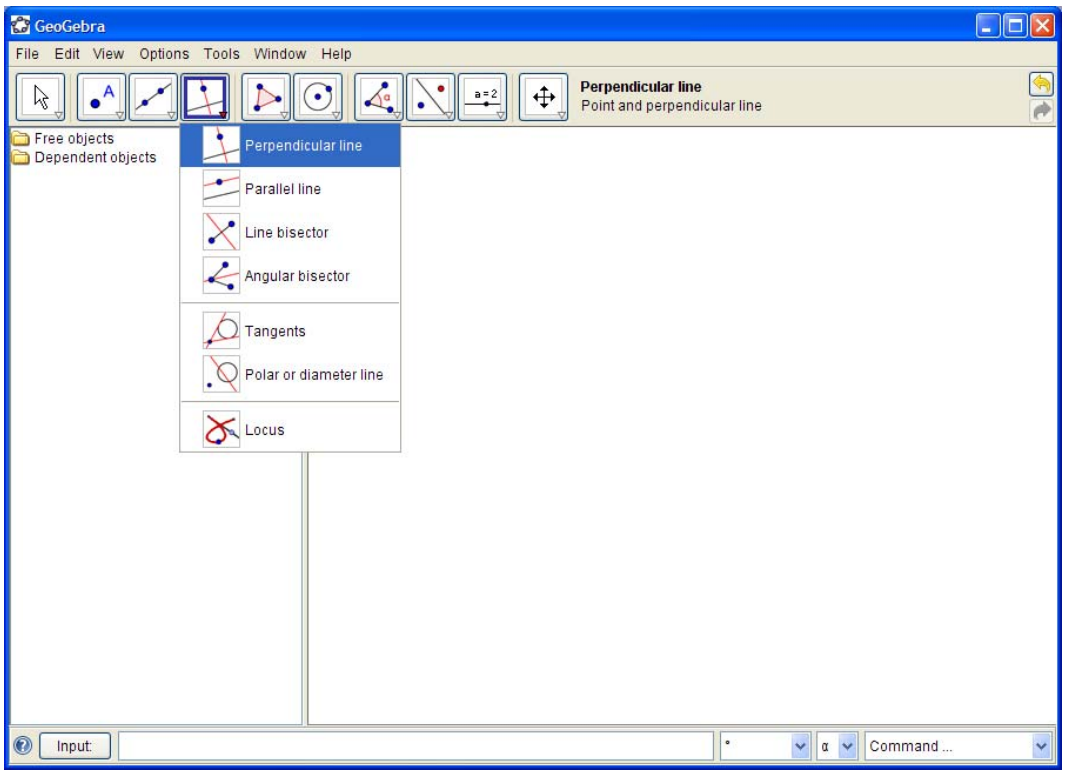
	Line through two points	Marking two points *A* and *B* fixes a straight line through *A* and *B*. The line's direction vector is (*B* - *A*).
	Segment between two points	Marking two points *A* and *B* fixes a segment between *A* and *B*. In the algebra window the segment's length is displayed.
	Segment with given length from point	Click on a point *A* that should be the starting point of the segment. Specify the desired length *a* of the vector in the appearing window.
	Ray through two points	Marking two points *A* and *B* creates a ray starting at *A* through *B*. In the algebra window you see the equation of the corresponding line.
	Vector between two points	Mark starting point and end point of the vector.
	Vector from point	Mark a point *A* and a vector *v* to create point *B* = *A* + *v* and the vector from *A* to *B*.



5.5. Perpendicular and parallel line, bisector, tangents, locus

	Perpendicular line	Marking a line *g* and a point *A* yields a straight line through *A* perpendicular to line *g*. The line's direction is equivalent to the perpendicular vector of g.
	Parallel line	Marking a line *g* and a point *A* defines a straight line through *A* parallel to *g*. The line's direction is the direction of line *g*.
	Line bisector	The line bisector of a line segment is stated by a segment *s* or two points *A* and *B*. The line's direction is equivalent to the perpendicular vector of segment *s*

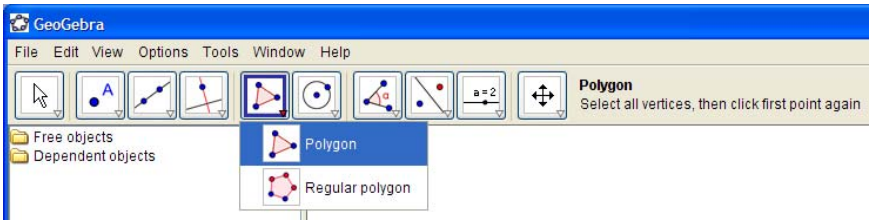
		or *AB*.
	Angle bisector	Angle bisectors can be defined in two ways. Marking three points *A*, *B*, *C* produces the angle bisector of the enclosed angle, where *B* is the apex. Marking two lines produces their two angle bisectors. Note: The direction vectors of all angle bisectors have length 1.
	Tangents	Tangents to a conic can be produced in two ways. Marking a point *A* and a conic *c* produces all tangents through *A* to *c*. Marking a line *g* and a conic *c* produces all tangents to *c* that are parallel to *g*. Marking a point *A* and a function *f* produces the tangent line to *f* in $x = x(A)$.
	Polar or diameter line	This mode creates the polar or diameter line of a conic section. You can either Mark a point and a conic section to get the polar line. Mark a line or a vector and a conic section to get the diameter line.
	Locus	Mark a point *B* that depends on another point *A* and whose locus should be drawn. Then, click on point *A*. Note: Point *B* has to be a point on an object (e.g. line, segment, circle).



5.6. Polygon

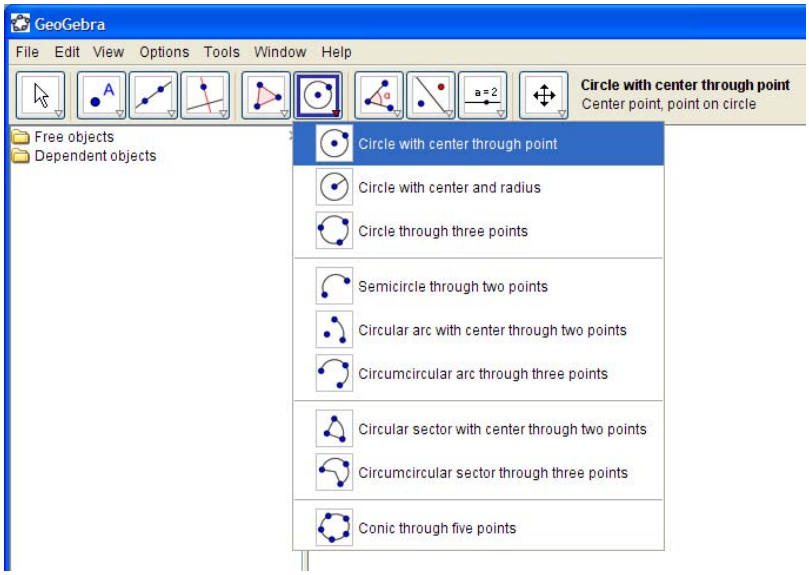
	Polygon	Mark at least three points which will be the vertices of the polygon. Then, click the first point again in order to close the polygon. In the algebra window you see the polygon's area.
	Regular polygon	Marking two points *A* and *B* and typing a number *n* into the text field of the appearing dialog gives you a

		regular polygon with n vertices (including points A and B).



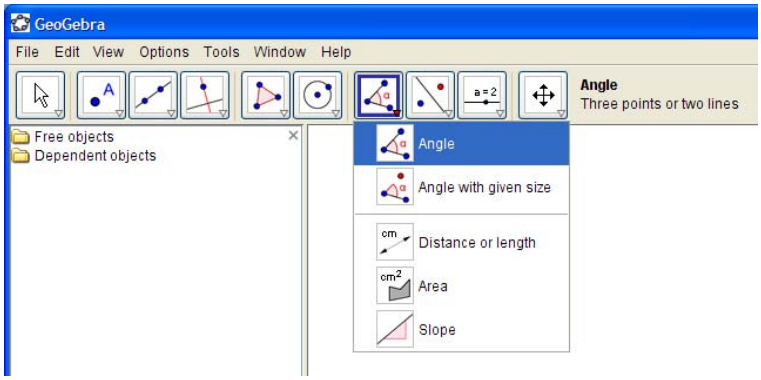
5.7. Conic section

	Circle with centre through point	Marking a point M and a point P defines a circle with centre M through P. This circle's radius is the distance MP.
	Circle with centre and radius	After marking a centre point M you are asked to enter the radius in the text field of the appearing window.
	Circle with centre and radius	After marking a centre point M you are asked to enter the radius in the text field of the appearing window.
	Semicircle	Marking two points A and B produces a semicircle above the segment AB.
	Circular arc with centre through two points	Marking three points M, A, and B produces a circular arc with centre M, starting point A and endpoint B. Note: Point B does not have to lie on the arc.
	Circumcircular arc through three points	Marking three points produces a circular arc through these points.
	Circular sector with centre through two points	Marking three points M, A, and B produces a circular sector with centre M, starting point A and endpoint B. Note: Point B does not have to lie on the sector.
	Circumcircular sector through three points	Marking three points produces a circular sector through these points.
	Conic through 5 points	Marking five points produces a conic section through them. Note: If no four of these five points lie on a line, the conic section is defined.



5.8. Angle, distance, area, slope

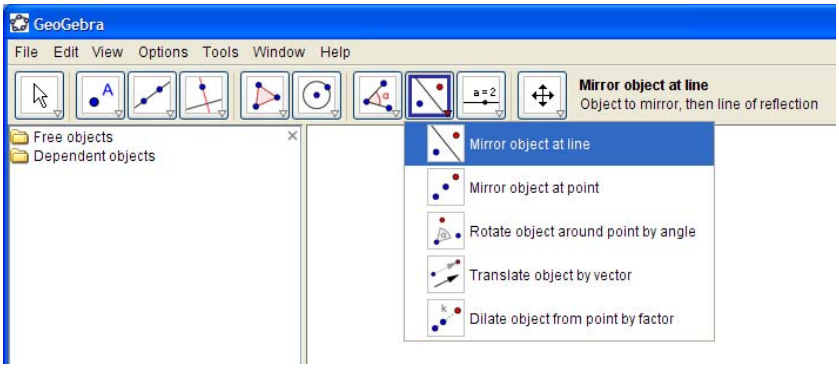
	Angle	This mode creates … angle between three points angle between two segments angle between two lines angle between two vectors all interior angles of a polygon
	Angle with given size	Mark two points *A* and *B* and type the angle's size into the text field of the appearing window. This mode produces a point *C* and an angle α, where α is the angle ABC.
	Distance or length	This mode yields the distance of two points, of two lines, or a point and a line. It can also give you the length of a segment or the circumference of a circle.
	Area	This mode gives you the area of a polygon, circle, or ellipse as a dynamic text in the geometry window.
	Slope	This mode gives you the slope of a line as a dynamic text in the geometry window.



5.9. Transformations

	Mirror object at	At first, mark the object to be mirrored. Afterwards, click

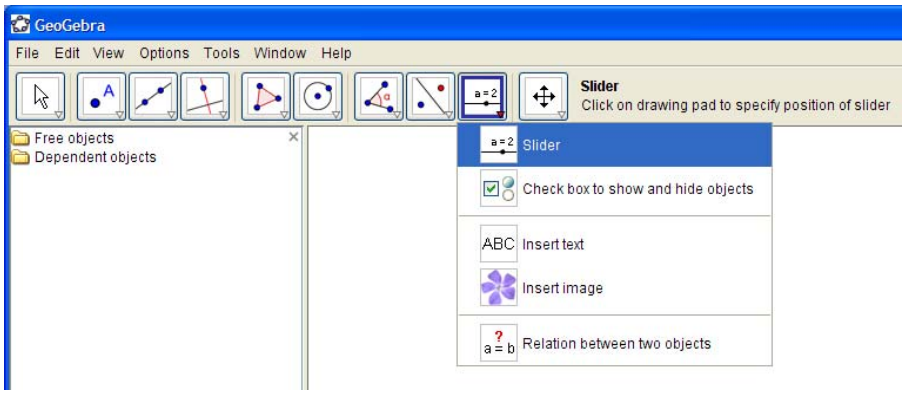
	point	on the point that should be the mirror.
	Mirror object at line	At first, mark the object to be mirrored. Afterwards, click on the line that should be the mirror.
	Rotate object around point by angle	At first, mark the object to be rotated. Then, click on the point that should be the rotation centre. Afterwards a window appears where you may specify the rotation angle.
	Translate object by vector	At first, mark the object to be translated. Afterwards, click on the translation vector.
	Dilate object from point	At first, mark the object to be dilated. Then, click on the point that should be the dilation centre. Afterwards, a window appears where you may specify the dilation factor.



5.10. Slider, text, image

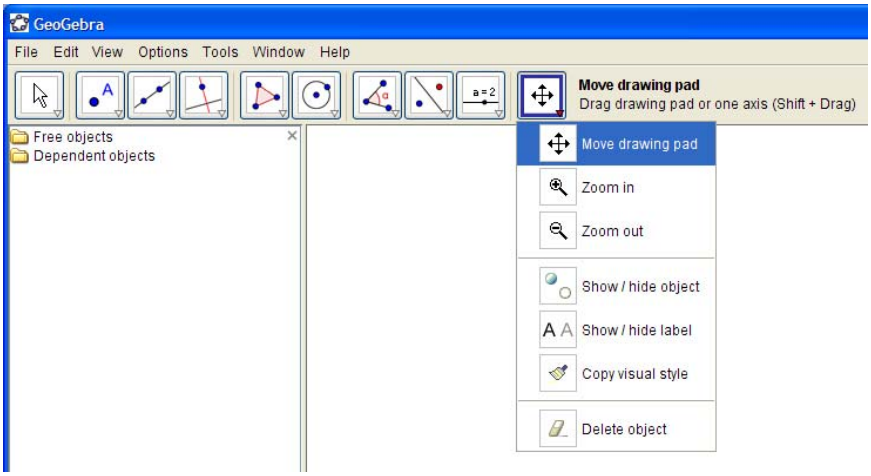
	Slider	In GeoGebra a slider is nothing but the graphical representation of a free number or angle. Click on any free place on the drawing pad to create a slider for a number or an angle. The appearing window allows you to specify the name, interval *[min, max]* of the number or angle, as well as the alignment and width of the slider (in pixel). You can easily create a slider for any existing free number or angle by showing this object. The position of a slider may be absolute on the screen or relative to the coordinate system.
	Check box to show and hide objects	Clicking on the drawing pad creates a check box (Boolean variable) in order to show and hide one or several objects. In the appearing window you can specify which objects should be affected by the check box.
	Text	With this mode you can create static and dynamic texts or LaTeX formulas within the geometry window. Clicking on the drawing pad creates a new text field at this location. Clicking on a point creates a new text field whose position is relative to this point. Afterwards, a dialog appears where you may enter the text.
	Insert image	This mode allows you to add an image to your construction. Clicking on the drawing pad specifies the lower left corner of the image. Clicking on a point specifies this point as the lower left corner of the image. Afterwards, a file-open

		dialog appears where you may choose the image file to insert.
? a = b	Relation	Mark two objects to get information about their relation.



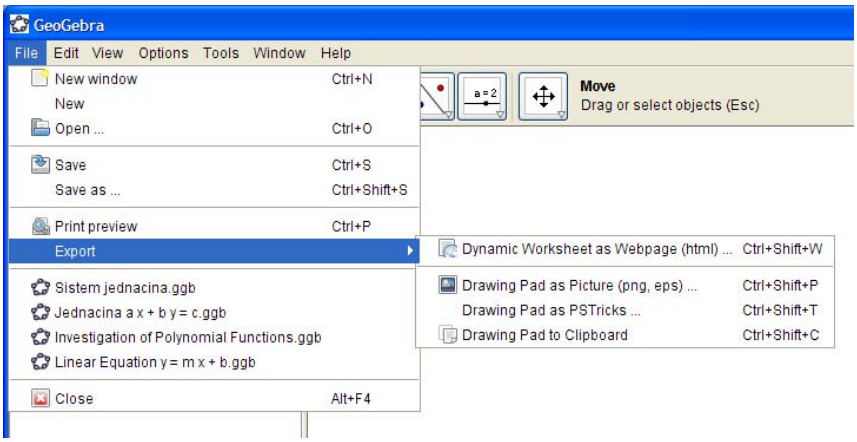
5.11. General modes

	Move drawing pad	Drag and drop the drawing pad to move the origin of the coordinate system. You can also move the drawing pad by pressing the *Shift* key (PC: also *Ctrl* key) and dragging it with the mouse. In this mode you can also scale each of the axes by dragging it with the mouse. Scaling the axes is also possible in every other mode by pressing and holding the *Shift* key (PC: also *Ctrl* key) while dragging the axis.
	Zoom in	Click on any place of the drawing pad to zoom in.
	Zoom out	Click on any place of the drawing pad to zoom out.
	Show / hide object	Click on an object to show or hide it. All objects that should be hidden are highlighted. Your changes will be applied as soon as you switch to any other mode in the toolbar.
A A	Show / hide label	Click on an object to show or hide its label.
	Copy visual style	This mode lets you copy visual properties (e.g. colour, size, line style) from one object to several others. To do so, first choose the object whose properties you want to copy. Afterwards click on all other objects that should adopt these properties.
	Delete object	Click on any object you want to delete.

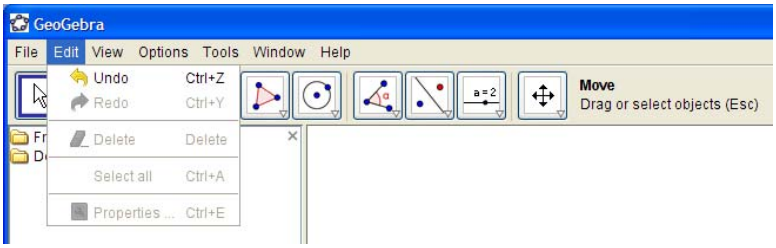


5.12. The Menus

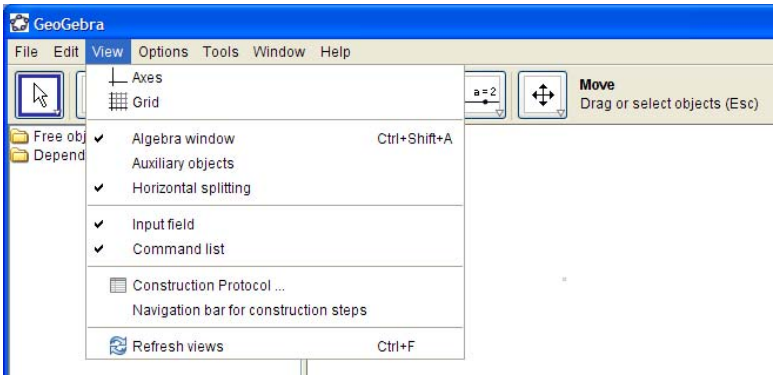
Export drawing



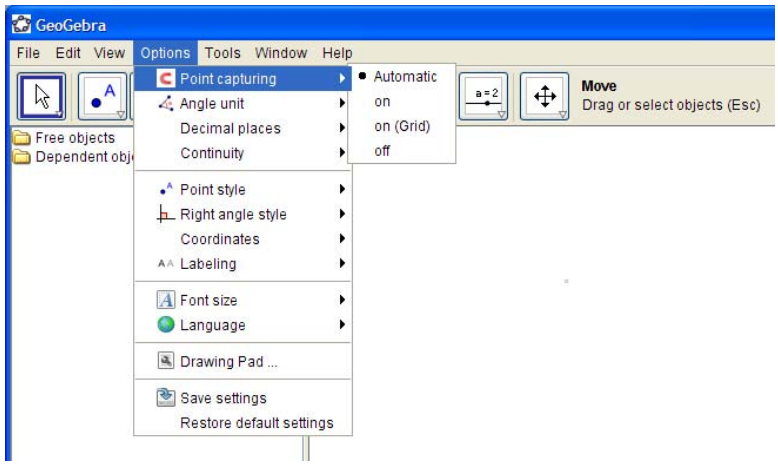
Undo the last operation



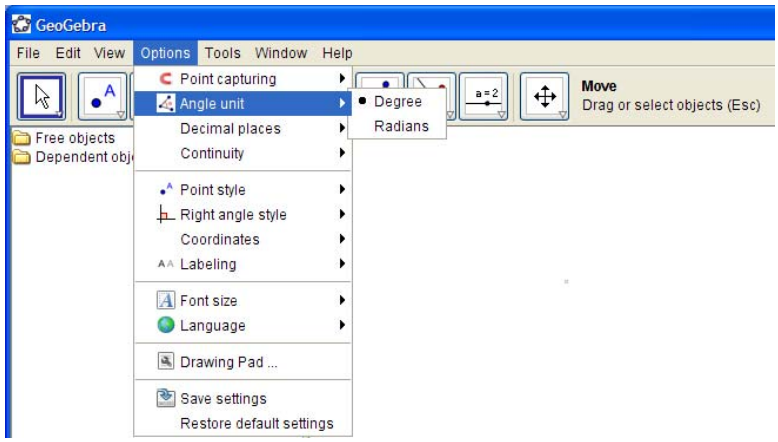
View options



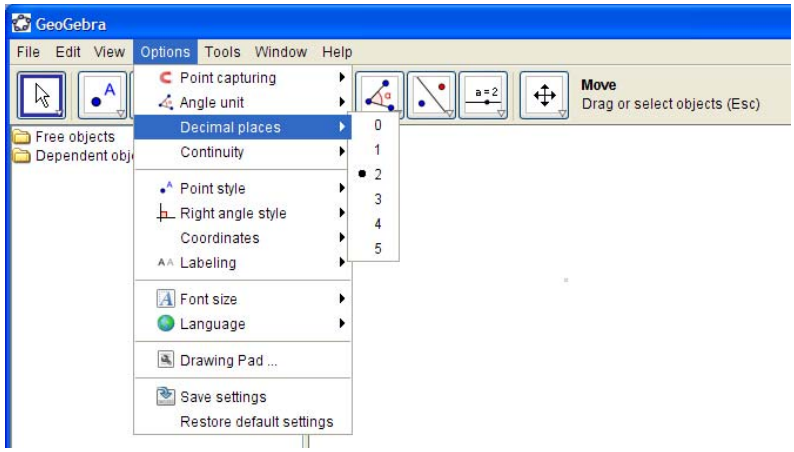
Point capturing



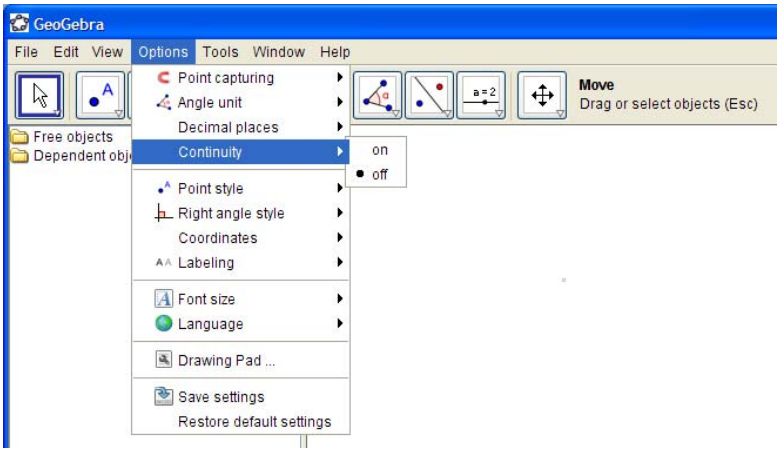
Angle units



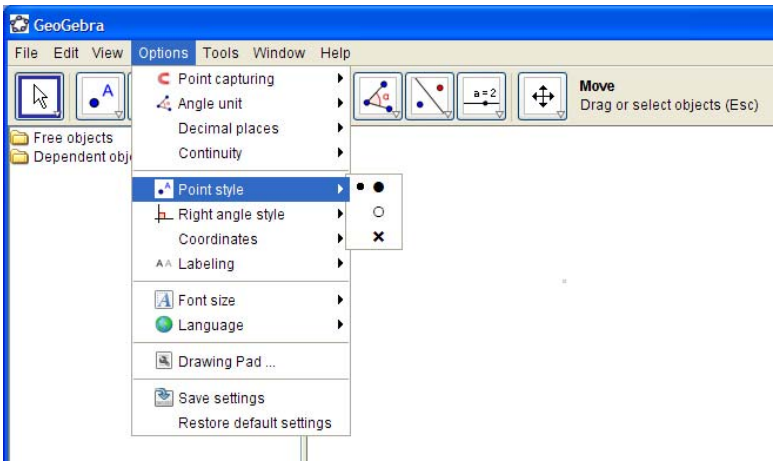
Decimal places



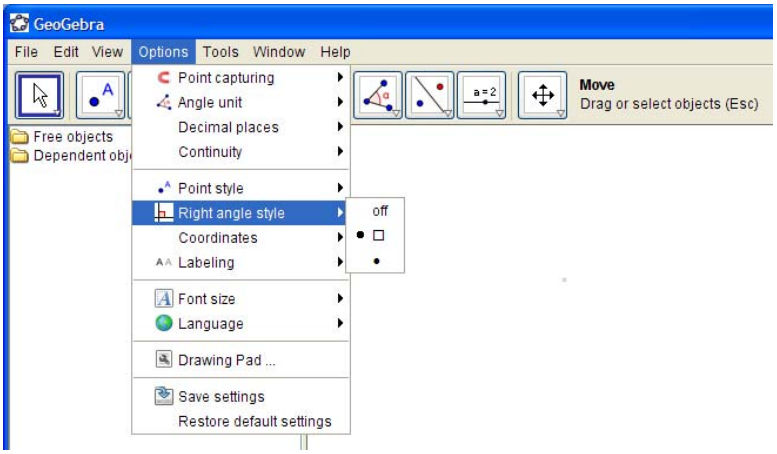
Contiguous movement



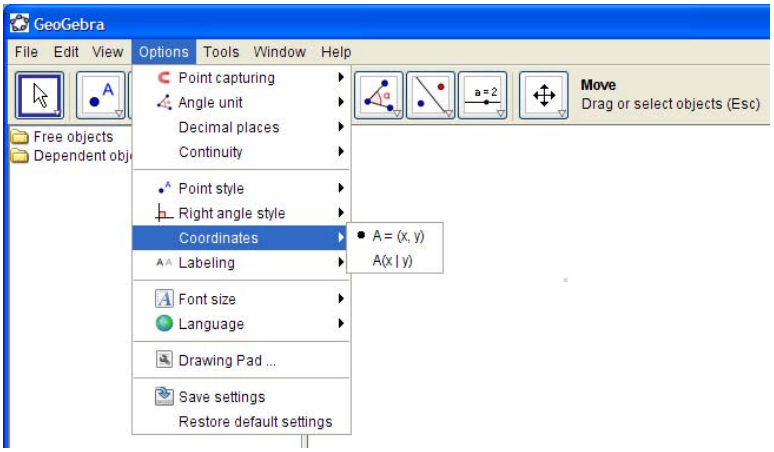
Point style



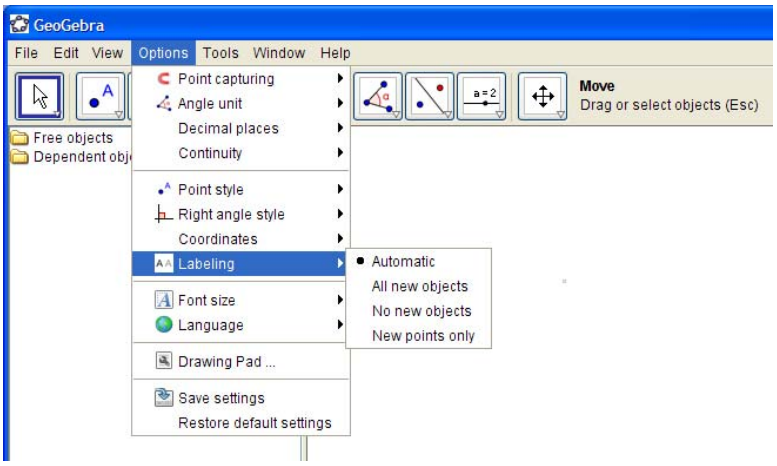
Right angle style



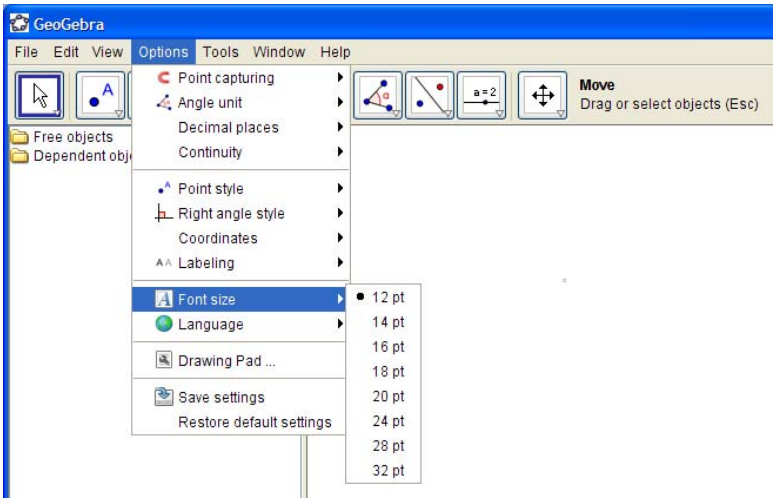
Coordinates



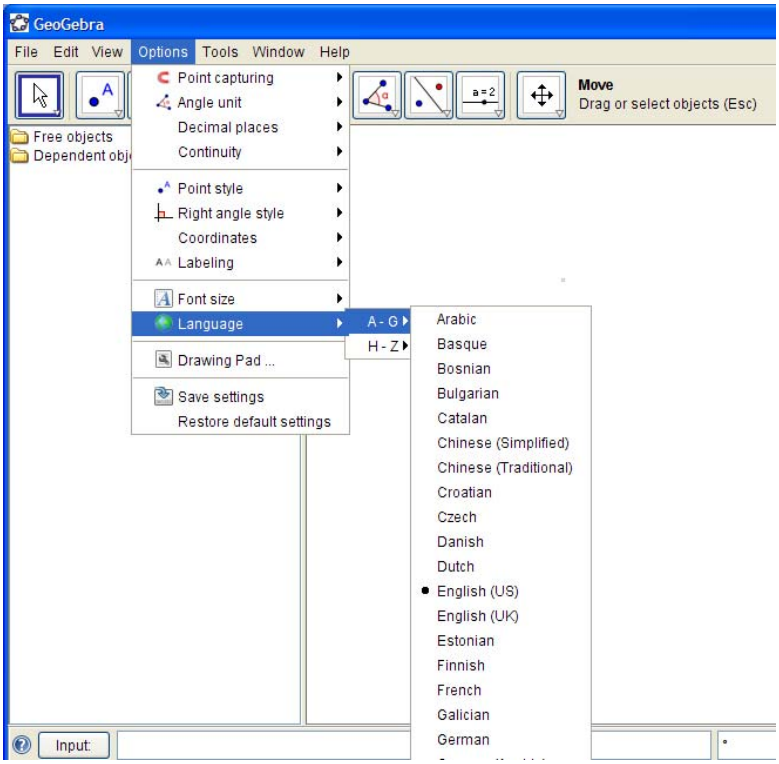
Labeling of objects



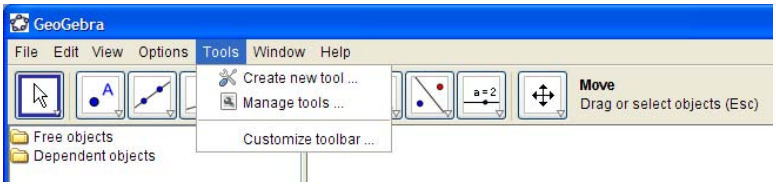
Text size



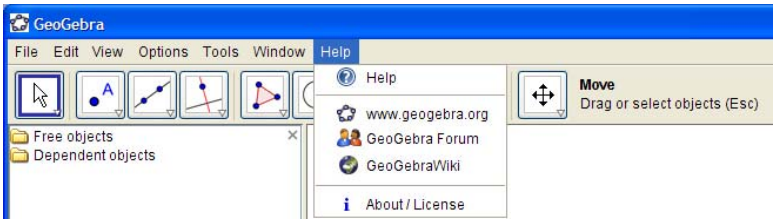
Language selection



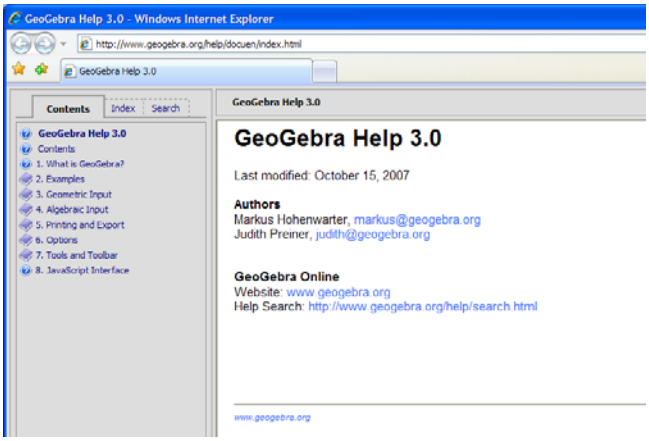
User tools



Help



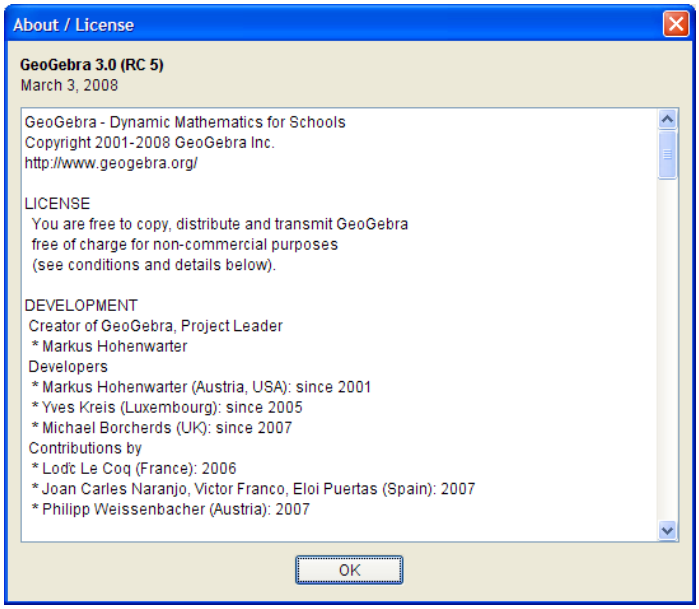
Help contents



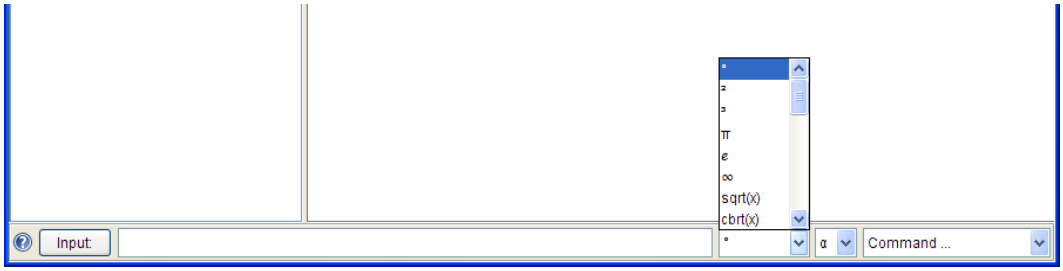
GeoGebra on the Web

GeoGebraWiki

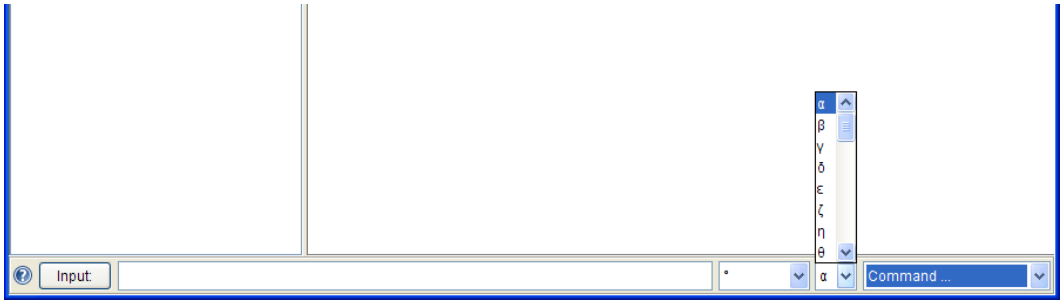
About GeoGebra



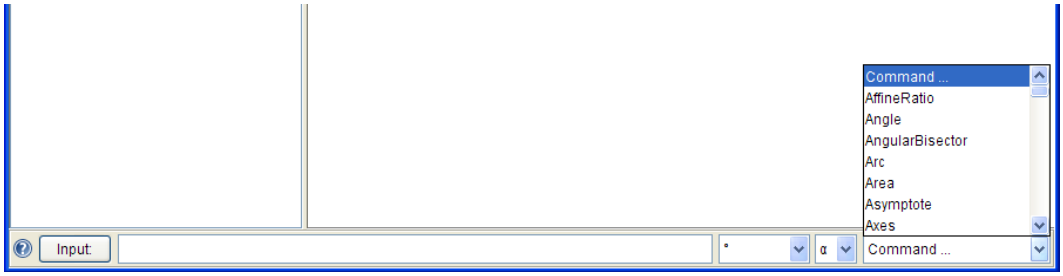
Special input



Greek alphabet



GeoGebra commands



5.13. TRANSLATIONS

* Arabic: Brahim Boulakbech, Haboubi Abdessalem (Tunisia)
* Basque: Gonzalo Elcano Vizcay (Spain)

* Bosnian: Maja Hrbat (Bosnia and Herzegovina)
* Catalan: Jorge Sánchez, Jaume Bartrolí, Pep Bujosa, Antoni Gomŕ, Roser Sebastián (Spain)
* Chinese: Fu-Kwun Hwang, Chen-Hui Lin, Pegasus Roe, Joe Chen (Taiwan)
* Croatian: Sime Suljic, Ela Rac (Croatia)
* Czech: Marie Pokorna, Pavel Sokol (Czech Republic)
* Danish: Steen Grođe (Denmark)
* Dutch: Beatrijs Versichel, Ivan De Winne, Pedro Tytgat (Belgium)
* English: Markus Hohenwarter, Judith Preiner (Austria), Yves Kreis (Luxembourg), Michael Borcherds (Great Britain)
* Estonian: Jane Albre (Estonia)
* Finnish: Hannu Korhonen, Kirsi Malinen (Finland)
* French: Noel Lambert (France)
* Galician: Jesús García Otero (Spain)
* German: Markus Hohenwarter, Judith Preiner (Austria), Yves Kreis (Luxembourg)
* Greek: Nicholas Mousoulides, Constantinos Christou (Cyprus), Spiros Mavrogiannis, Manolis Koutlis, Fergadiotis Athanasios (Greece)
* Hebrew: Guy Hed (Israel)
* Hungarian: Sulik Szabolcs (Hungary)
* Italian: Enrico Pontorno, Alessandra Tomasi, Palmira Ronchi, Simona Riva (Italy)
* Japanese: Akihito Wachi (Japan)
* Macedonian: Linda Fahlberg-Stojanovska (FYR Macedonia)
* Norwegian: Sigbjřrn Hals (Norway)
* Persian: Saeed Aminorroaya (Iran)
* Polish: Marzanna Miasko, Malgorzata Paliga, Ewa Piwek (Poland)
* Portuguese (Brazil): Humberto Bortolossi, Herminio Borges Neto, Alana Paula, Luciana de Lima, Araújo Freitas, Alana Souza de Olivieira (Brazil)
* Portuguese (Portugal): Jorge Geraldes, António Ribeiro (Portugal)
* Russian: Anatoly Scherbakov (Russia)
* **Serbian: Djordje and Dragoslav Herceg (Serbia)**
* Slovak: Peter Csiba (Slovakia)
* Slovenian: Stanislav Senveter (Slovenia)
* Spanish: Liliana Saidon (Argentina)
* Turkish: Erol Karakirik (Turkey)
* Vietnamese: Nguyen Thanh Trung, Quang Nguyen (Vietnam)

5.14. Keyboard shortcuts

5.14.1. Windows

- **Ctrl + N –** new window
- **Ctrl + O –** file
- **Ctrl + S –** save the current file
- **Ctrl + Z –**undo
- **Ctrl + Y –** redo
- **Ctrl + F –** refresh
- **Ctrl + A –** Show / hide algebra window
- **Ctrl +**key + or kay – or arrows **–** contiguous animation of a point on an object
- **Shift +** key + or kay – or arrows **–** non contiguous animation of a point on an object
- **Esc** – mowing

5.14.2. Mouse

- **Ctrl + left mouse button** – mowing drawing pad
- **Ctrl + mouse wheel** – zooming
- **Ctrl + click on multiple objects in the property window** – selection of multiple objects
- **Shift +click on two objects in the property window** – selection of two objects and all between them
- **Double left click on an object in geometry window –** select property window
- **Double left click on an object in algebra window –** redefine
- **Pressing right mouse button and mowing the mouse –** select the zoom area
- **Click right mouse botton** – select many of drawing pad
- **Click right mouse botton on an object** – select many of object

5.14.3. Input field

- **Ctrl + A –**select all
- **Ctrl + X –** cut selection
- **Ctrl + C –** copy selection
- **Ctrl + V –** paste selection

GeoGebra – динамичка геометрија и алгебра

1 Увод

Већ смо раније нагласили, глава 2, главне разлоге зашто смо се определили за GeoGebr-у. Овај пакет пружа на једноставан и веома приступачан начин припремање и самосталан рад ученика из две веома важне области школске математике, из геометрије и алгебре. Креирање анимација и једноставне илустрације школског садржаја омогућавају наставницима да за кратко време и веома ефикасно упознају ученике са основним математичким појмовима и знањима.

Програм GeoGebra је математички софтвер који повезује геометрију и алгебру. Развио га је Markus Hohenwarter са Универзитета у Салцбургу за поучавање математике у школама. Бесплатан је и доступан на више од 15 језика. Може је имати и користити сваки ученик. Инсталација је крајње једноставна.

Овај програм обједињује многе особине које имају многи системи, али ниједан овако комплетно као GeoGebra. Описујући GeoGebra-у, дајемо практично принцип рада многих других система.

Описаћемо само онај део GeoGebra-е који нам је потребан за обраду материјала од петог до осмог разреда основне школе.

У основној школи, посебно од петог до осмог разреда, настава математике је конципирана тако да се садржаји геометрије и алгебре преплићу и прожимају, тако да је GeoGebra овде посебно погодна. Уместо два одвојена програма, један за геометрију а други за алгебру, имамо само један – GeoGebra-у.

Радне површине креиране у програму GeoGebra могу се преносити у html и word документе. Поред тога конструкције се могу понављати по вољи корак по корак и аутоматски и ручно.

2 Динамичка геометрија

Елементарна геометрија равни, планиметрија, може се веома успешно обрађивати у настави помоћу динамичких геометријских система (Dynamic Geometry Systems, DGS), Schumann [85], Sträßer [89, 90]). Примери оваквих система су

- Cabri geometre II+ (www.cabri.com),
- Cinderella (www.cinderella.de),
- DynaGeo (www.dynageo.de),
- The Geometer's Sketchpad (www.keypress.com/sketchpad),
- Geonext(www.geonext.de)
- Zirkel und Lineal (www.mathsrv.kueichstaett.de/MGF/homes/grothmann).

Код ових система, алата, могуће је помоћу миша конструисати геометријске фигуре и динамички их мењати. При томе ове фигуре се неће деформисати, тј. односи геометријских објеката остају очувани, на пример паралелне и нормалне праве остају паралелен и нормалне итд. Објекти су тачке, праве кругови, конусни пресеци и графици функција.

Уобичајено је да се у системима динамичке геометрије објекти могу посматрати аналитички, преко својих координата и једначина. Обрнуто, да се задају координате и једначине и да се потом појави графичка презентација датих објеката која би се могла још и мењати директно помоћу миша у овим системима није могуће.

3 Системи рачунарске алгебре

Системи рачунарске алгебре, (Computer Algebra Systems, CAS), пружају могућност за аналитичку обраду геометрије, Davenport [11], Fuchs [27]. Познати и цењени ове групе алата су

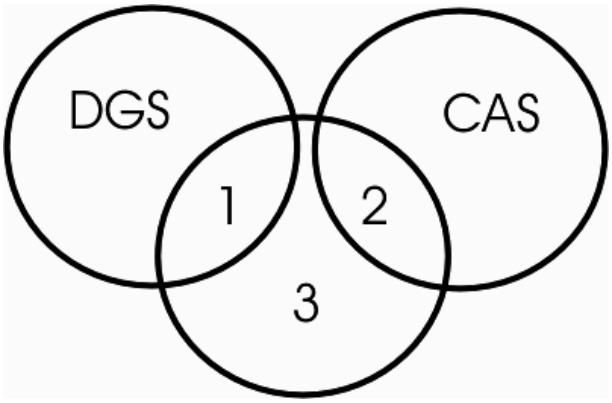
- Derive,
- Maple
- Mathematica.

Код ових алата могуће је координате и једначине геометријских објеката визуелизирати, тј. могуће је графички представити и посматрати промене код алгебарских објеката посматрати на њиховим графичким приказима. Ова графичка представљања, такозвани Plots, не могу се мењати помоћу миша. Задавање алгебарских објеката није увек једноставно, пошто синтакса ових система често има веома мало заједничког са уобичајеном школском нотацијом. Ови системи могу, између осталог, да решавају алгебарске системе, да одређују, на пример, конусне пресеке. CAS алати су веома моћни алати опште намене, али не дозвољавају у области аналитичке геометрије никакве директне динамичке промене.

4 GeoGebra

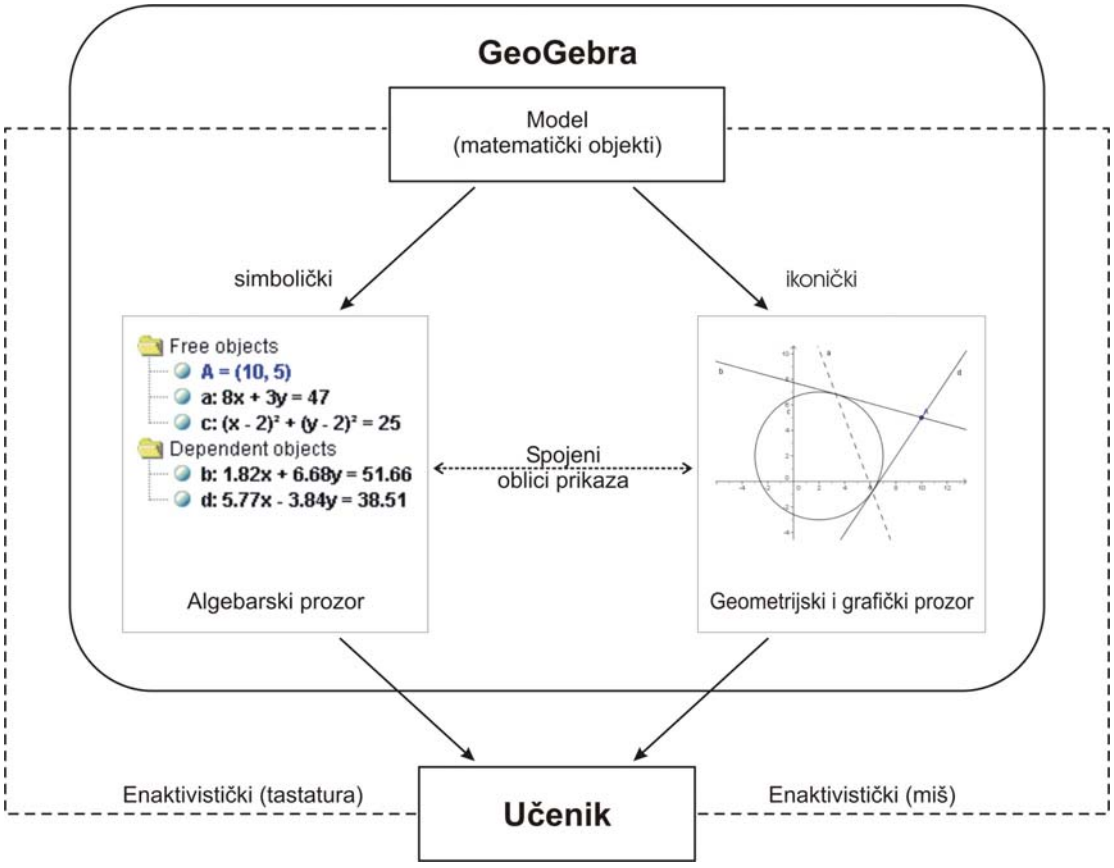
Очигледно, било би добро имати један алат који би садржавао све предности DGS и CAQ алата. У Schumann [85], налазимо прва упутства за једну комбинацију овог типа. За праве и равни ово се појављује први пут у алату „3D-Geometer" Klemenz [55]. Ова основна идеја разрађена је кроз алат „GeoGebra" Hohenwarter [46] (од Geometrie und Algebra) пренета на конусне пресеке и адекватно софтверски реализована. геометрија и алгебра се појављују у „партнерском" односу, Fuchs, Vasashelyi [29].

- Област 1: Конструкција кружнице $\Rightarrow$ показивање његове једначине,
- Област 2: Задавање једначине кружнице $\Rightarrow$ приказивање кружнице као статичне слике,
- Област 3: Задавање једначине кружнице $\Rightarrow$ приказивање кружнице као динамичке слике (померање слике је могуће).



Слика 1.

Активности засноване на учењу помоћу иконичког и симболичког представљања објеката, према Bruner-у [4], слика 2.



Слика 2.

С једне стране, GeoGebra је DGS алат. Директно интерактивно се могу конструисати тачке, вектори, линије и конусни пресеци и померањем се могу динамички мењати. Поред кругова могу се цртати елипсе, хиперболе и параболе. Конструкције тангенти и полара спадају такође у основне функције.

С друге стране, у GeoGebra-и могуће је једначине и координате директно задавати. GeoGebra познаје и експлицитне и имплицитне једначине правих и конусних пресека, параметарско представљање правих као и поларне и Декартове координате тачака и вектора. Будући да GeoGebra рачуна са бројевима, угловима, векторима, тачкама, правама и конусним пресецима, може се рећи да је GeoGebra нумерички CAS алат. Због тога GeoGebra нуди више геометријских наредби: одређивање средишта дужи, жижа и темена конусних пресека. Поред тога даје коефицијент правца, вектор правца и нормални вектор једне праве, главне осе и прешник конусног пресека.

GeoGebra је написана у Java-и. Тиме је омогућено да се користи независно од тога шта је у употреби: Windows, Linux, MacOS X или Unix. Поред тога GeoGebra се може покренути директно преко Internet Browser-a, на пример, Internet Explorer или Netscape.

GeoGebra се заснива на пројективној и еуклидској геометрији. Полиномна поједностављења се заснивају на Parser алгоритмима а геометријске алгоритме је развио аутор GeoGebra-е Markus Hohenwarter.

GeoGebra је резултат дипломског рада Markus Hohenwarter-a, који је рађен на Institut für Didaktik der Naturwissenschaften der Universität Salzburg. Даљи рад на програму GeoGebra аутор наставља кроз израду докторске дисертације из дидактике математике на

Universität Salzburg. Овај рад финансира као пројекат аустријска академија наука Österreichischen Akademie der Wissenschaften.

Програм GeoGebra је награђен са више награда за образовни софтвер:

- EASA 2002: European Academic Software Award (Ronneby, Sweden)
- Learnie Award 2003: Austrian Educational Software Award (Vienna, Austria)
- digita 2004: German Educational Software Award (Cologne, Germany)
- Comenius 2004: German Educational Media Award (Berlin, Germany)
- Learnie Award 2005: Austrian Educational Software Award for Andreas Lindner (Vienna, Austria)
- Trophées du Libre 2005: International Free Software Award, category Education (Soisson, France)

Програм GeoGebra је бесплатан и лако доступан. Довољно је повезати се преко интернета на http://www.geogebra.at. Поред програма GeoGebra

могуће је повући и бројне примере и текстове о GeoGebra-и. Многи ученици и наставници шаљу своје примере и примедбе, тако да аутор имајући у виду и повратне информације, ради на даљем побољшању овог популарног програма. GeoGebra повезује DGS i CAS алате на један нови начин што настави математике даје веома лепе и широке могућности.

GeoGebra је доступна на више језика: енглески, немачки, аустријски, италијански, француски, шпански, каталонски, португалски, холандски, дански, мађарски, словеначки, хрватски, кинески, српски, руски.

Да би илустровали могућности GeoGebra-е, даћемо неколико примера. Примери су изабрани тако да прикажемо неке садржаје наставе математике од петог до осмог разреда основне школе.

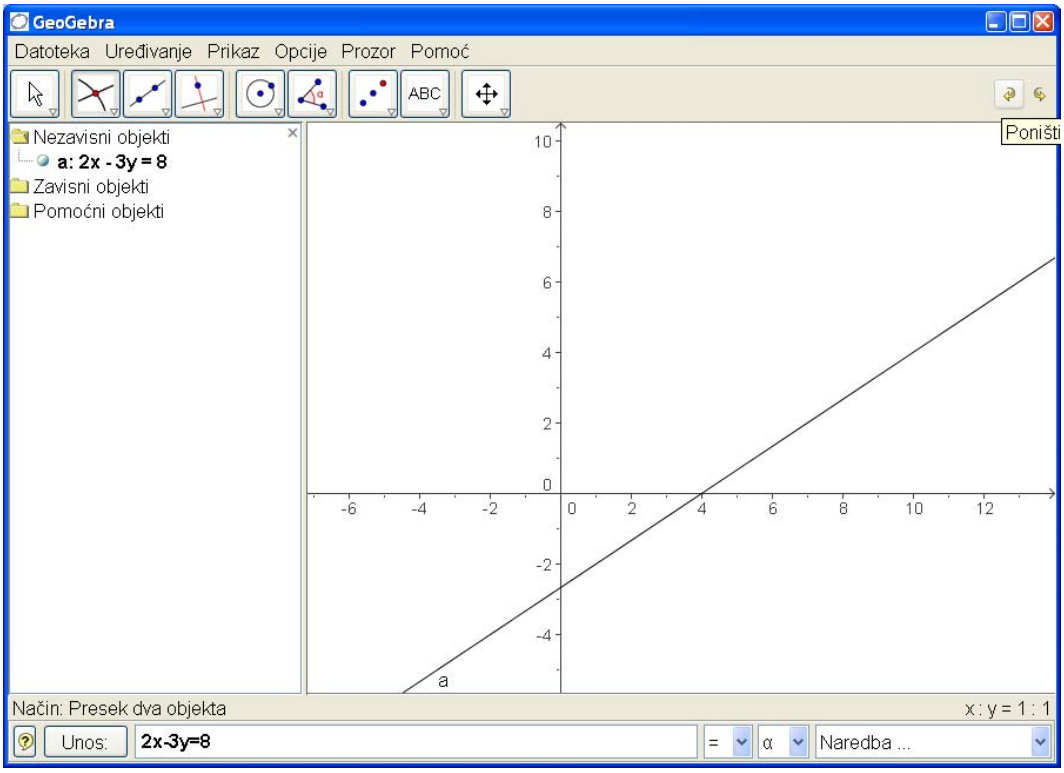
Слика 3.

4.1 Пример

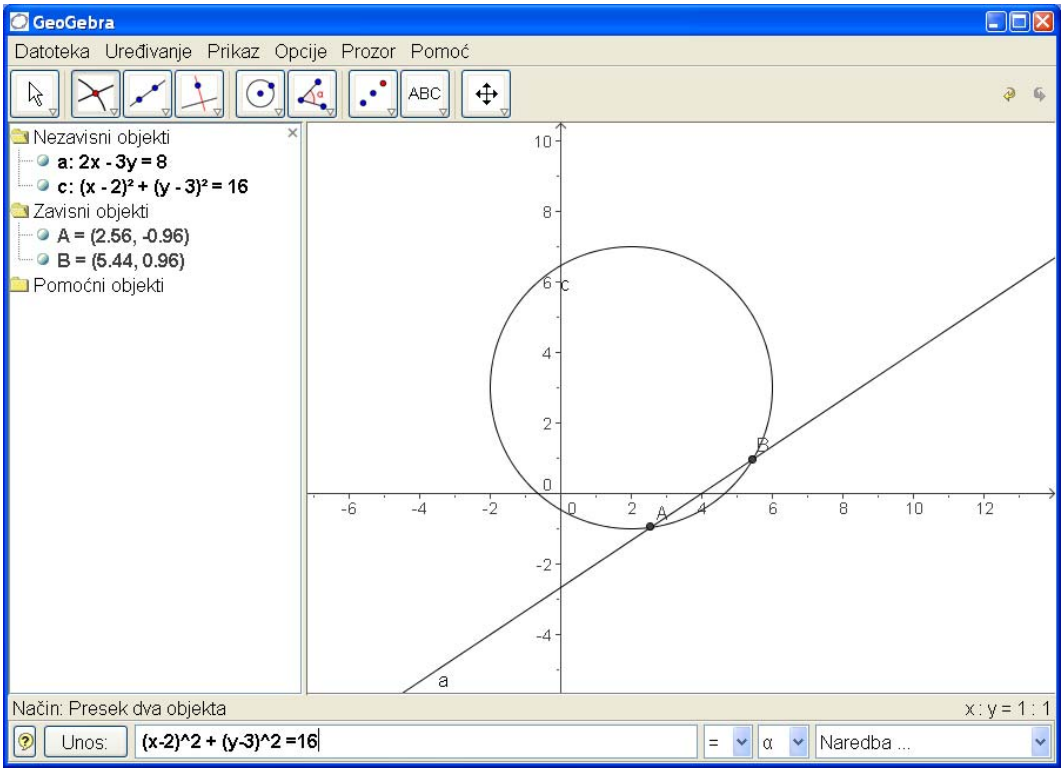
При прављењу програма GeoGebra посебна пажња је била посвећена једноставној употреби свих наредби и опција, што је јако добро за ученике, како би добили вољу да експериментишу и самостално раде. На пример, једначина праве и круга могу се задати једноставно у школској нотацији:

- права $a: 2x-3y=8$
- кружница $c:(x-2)^2+(y-3)^2=16$.

Управо се овако уносе ови објекти у GeoGebra-у! Само један покрет је довољан да се одреди пресек праве и кружнице – тачка. Истовремено се у алгебарском прозору појављује пресечна тачка са својим координатама. Другим покретом добија се друга пресечна тачка праве и кружнице и њене координате у алгебарском прозору.



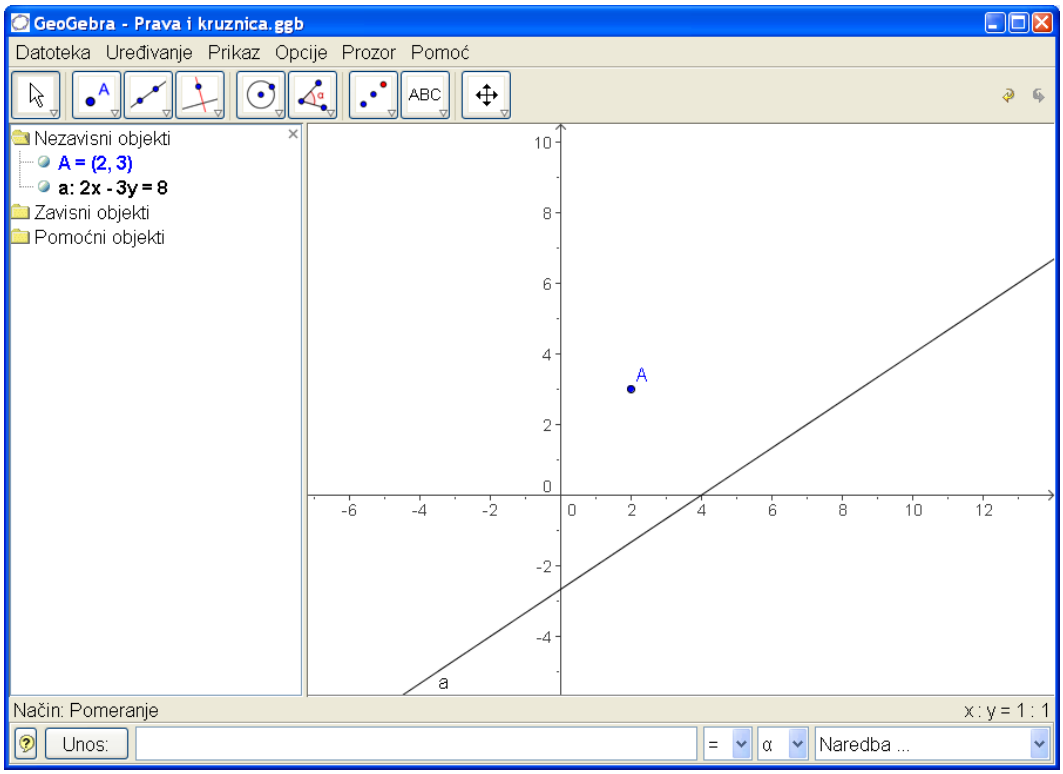
Слика 4.



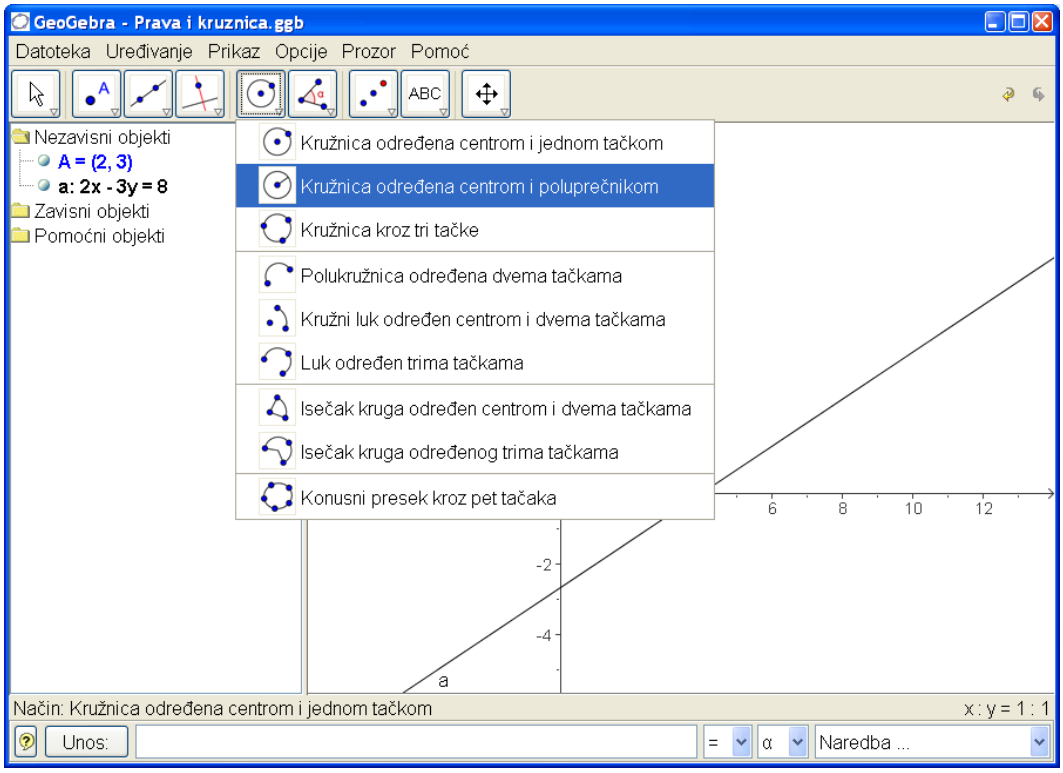
Слика 5.

Ученици у основној школи не обрађују једначину кружнице. Због тога кружницу можемо задати преко геометријског прозора. Изабраћемо тачку

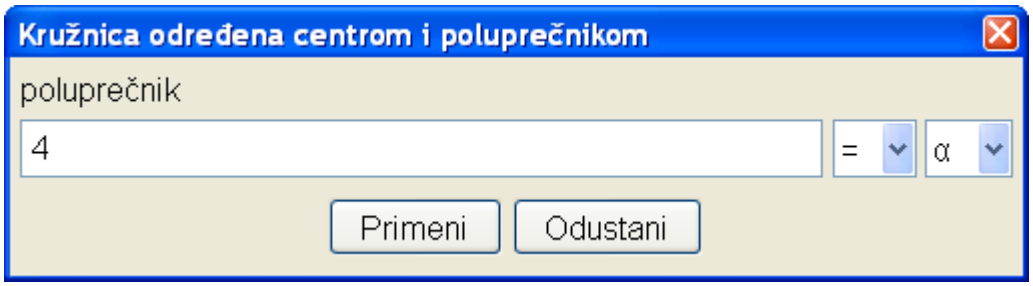
(2, 3) и са полупречником 4 нацртати кружницу. Тада ће се у алгебарском прозору појавити једначина кружнице, слике 6-9.



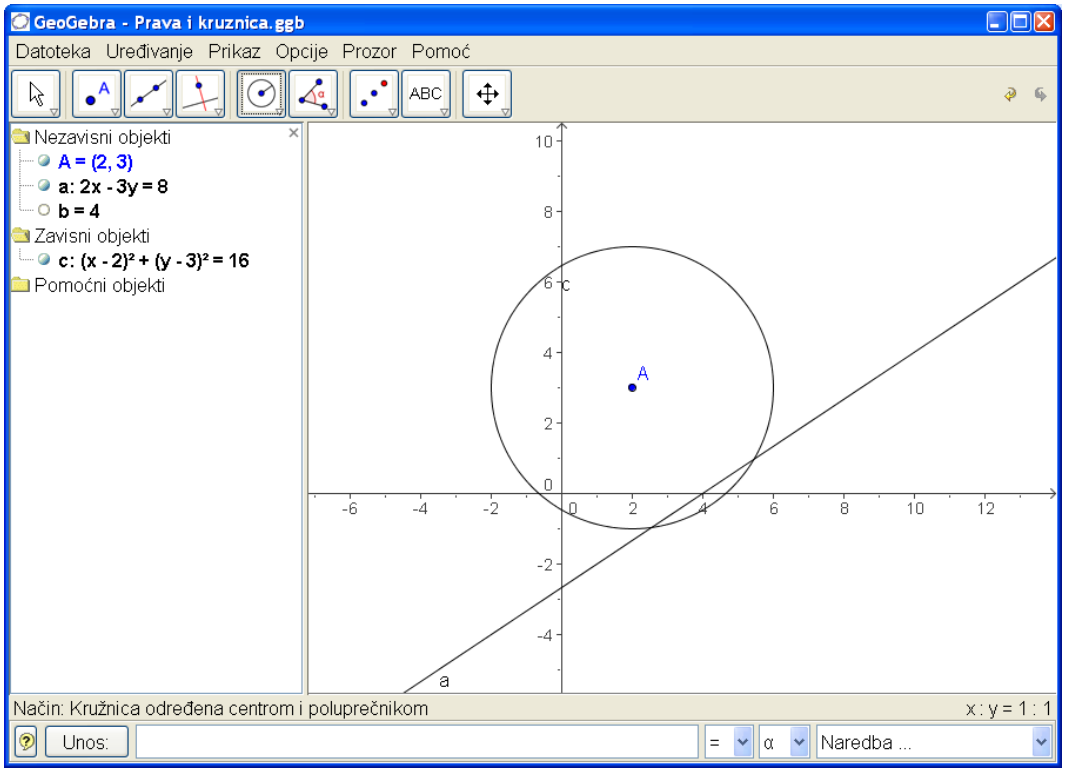
Слика 6.



Слика 7.



Слика 8.



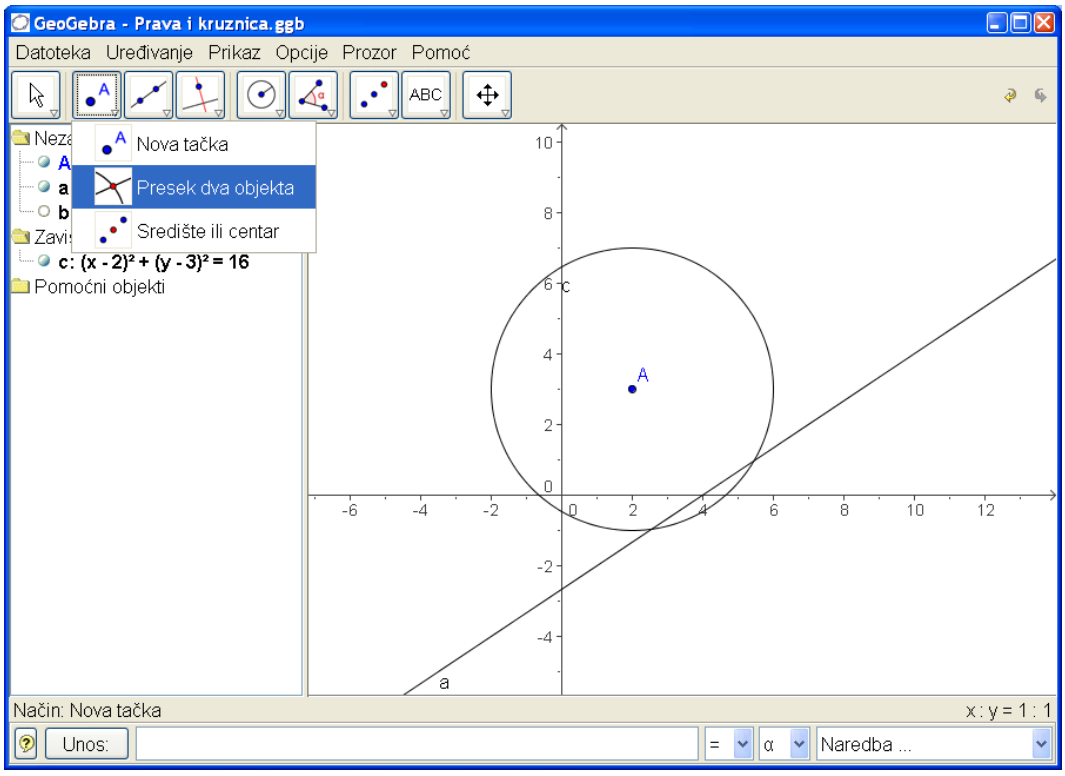
Слика 9.

Померањем праве и кружнице, једноставно их селектујемо помоћу миша и повлачимо их на жељено место, можемо демонстрирати да права и кружница могу имати две, једну или ниједну заједничку тачку. При томе се у алгебарском прозору појављују одговарајуће једначине праве, кружнице и координате пресечних тачака ако их има, слике 9-12.

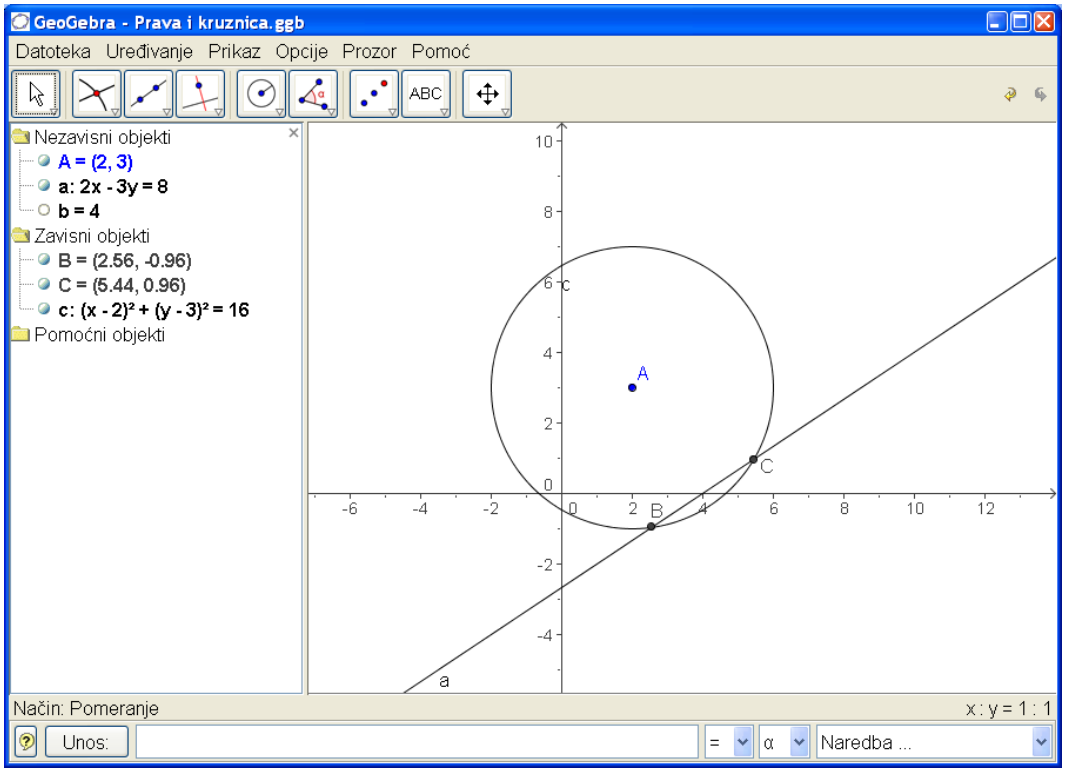
Ако се права селектује десним дугметом миша, добиће се мени помоћу којег можемо изабрати неке опције за промену праве или неких њених особина: дебљина и боја графика, промена имена итд. На пример, ако изаберемо својства добијамо следеће опције, слика 13.

Ако изаберемо експлицитни облик једначине, слика 14, наше праве $y = kx + d$ можемо померајући праву мишем посматрати промену величине d. Коефицијент правца праве остаје исти као и раније. Овај геометријски експеримент може пратити и алгебарски у којем би мењали коефицијент правца праве у алгебарском прозору. На графику би се одмах виделе промене које тако настају, тј. појавио би се график изабране праве. Јасно је

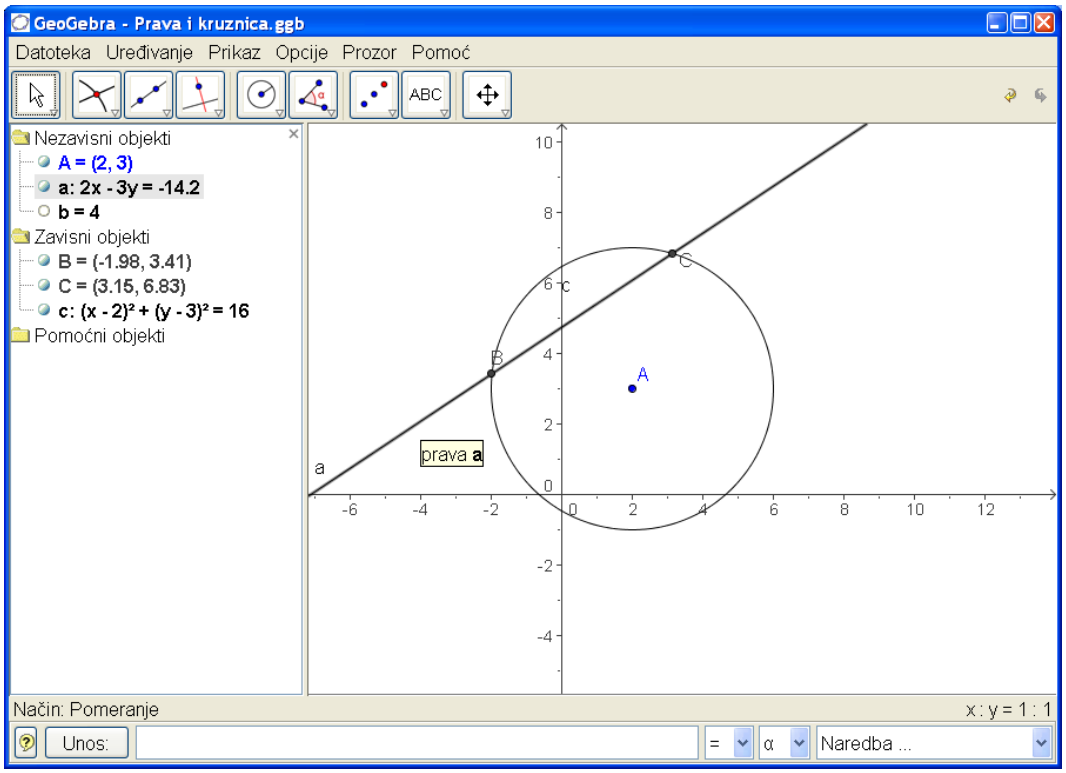
да праву коју смо прву нацртали можемо задржати и задати нову у експлицитном облику са слободним чланом истим као код прве праве.



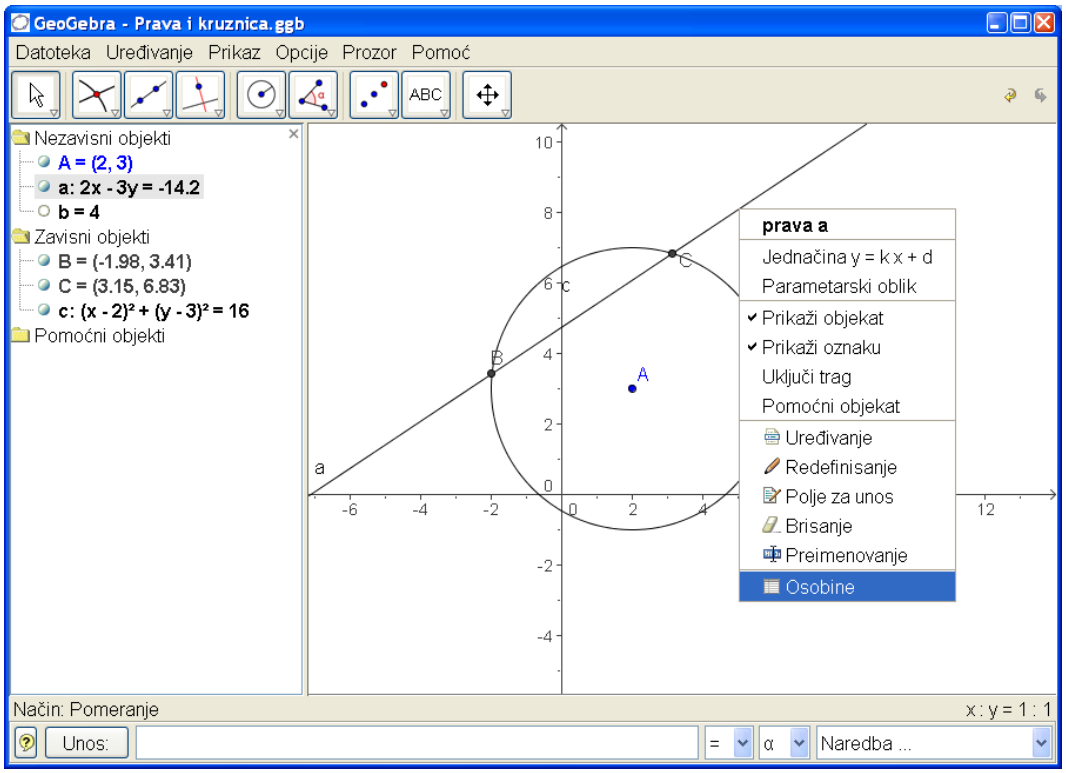
Слика 10.



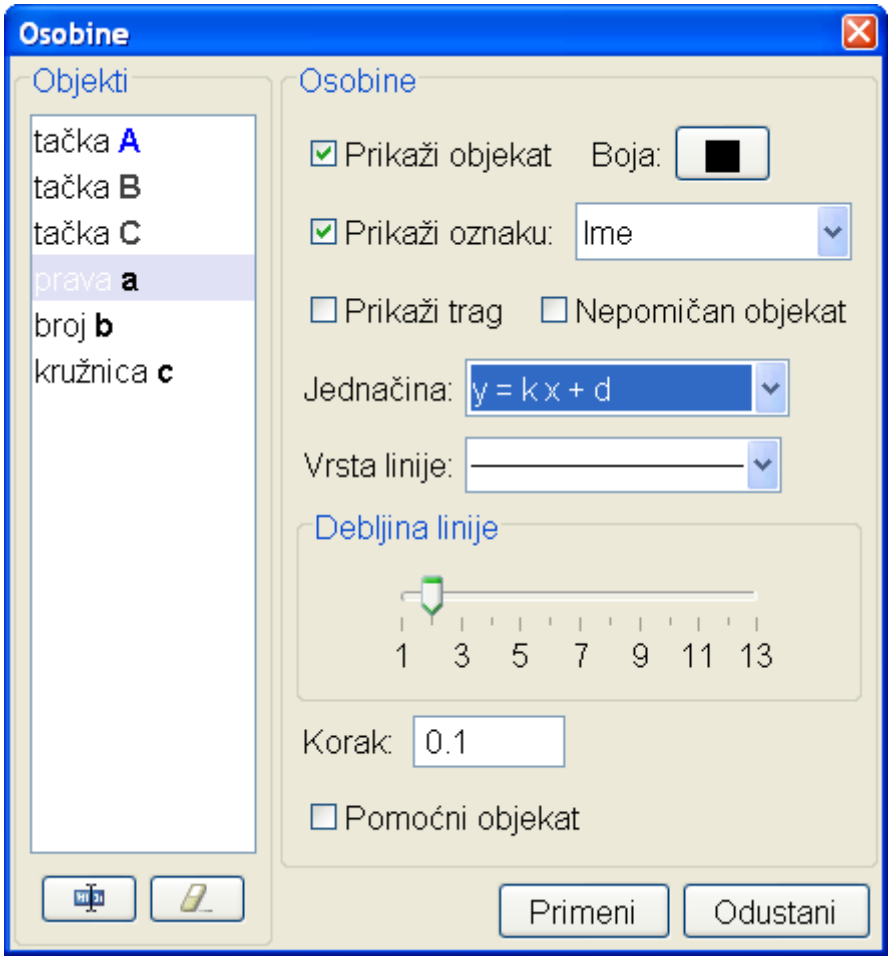
Слика 11.



Слика 12.



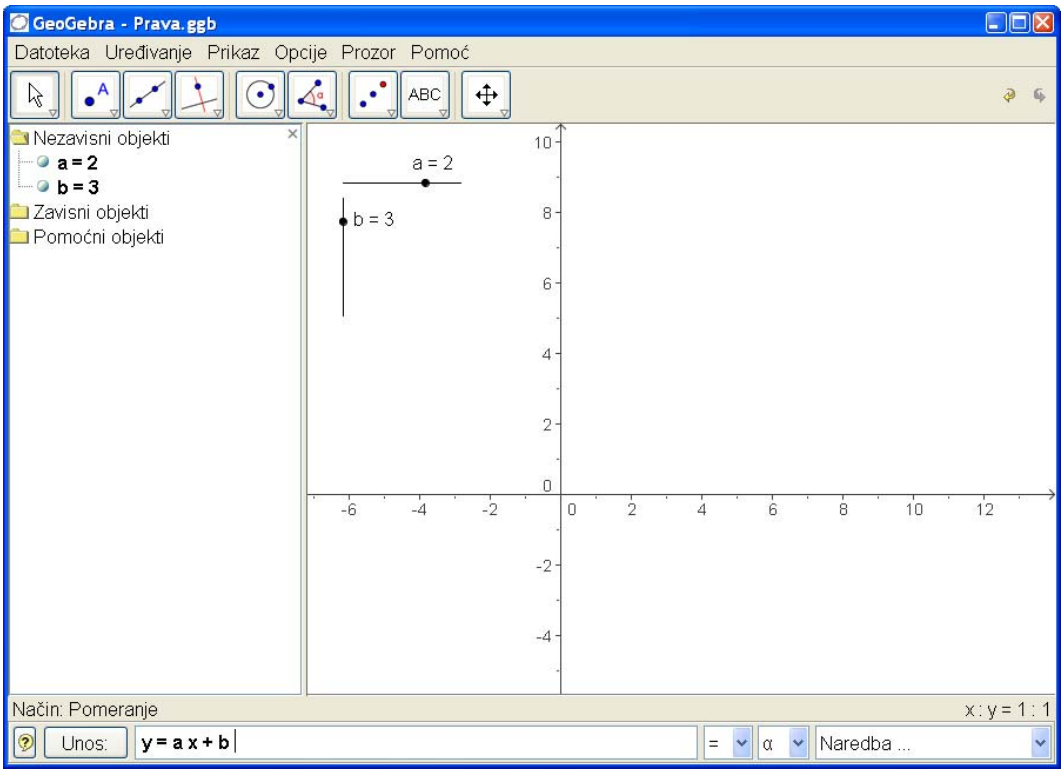
Слика 13.



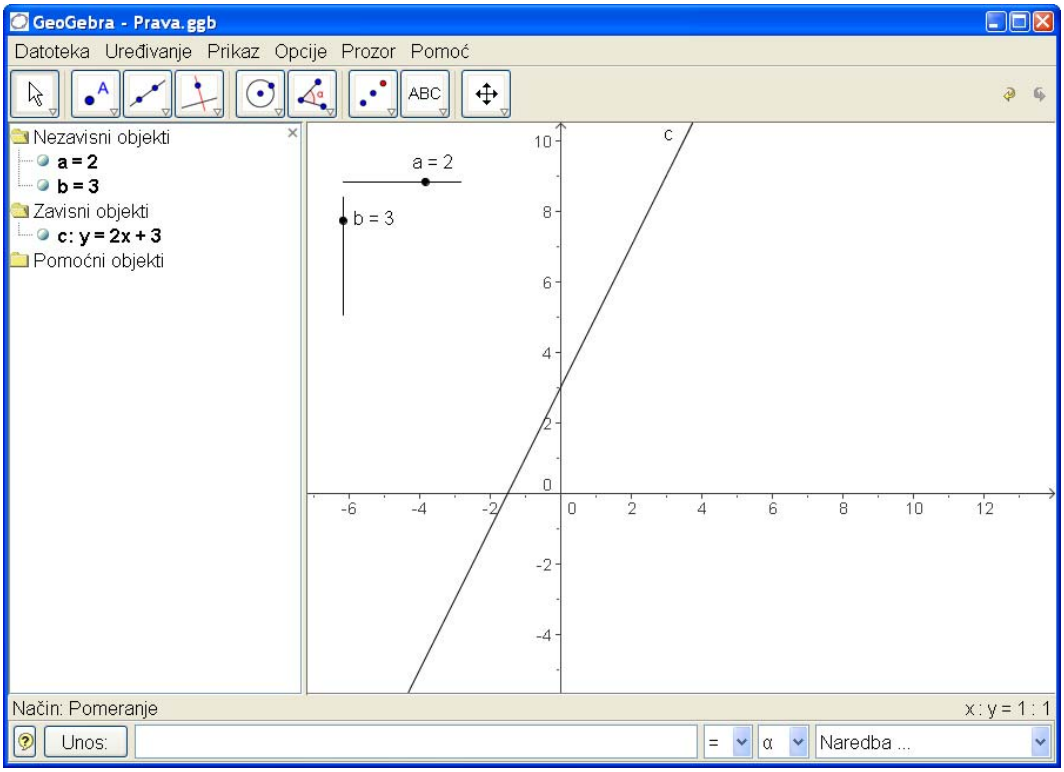
Слика 14.

4.2 Пример

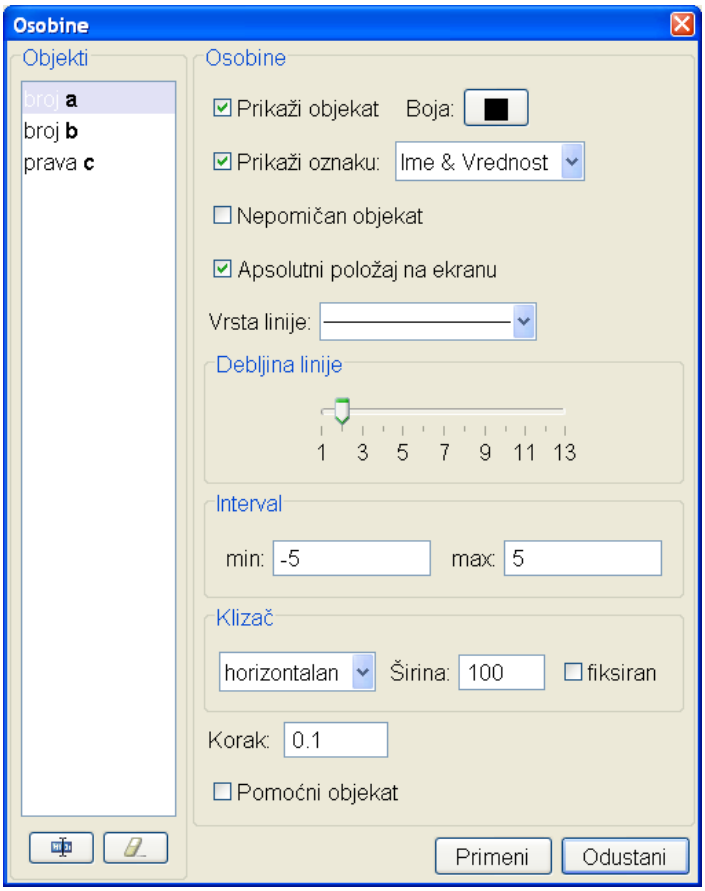
Праву у експлицитном облику $y = ax + b$ можемо задати на тај начин што ћемо прво изабрати два клизача. Први ће одређивати вредност за a а други вредност за b. На слици 15 први клизач је постављен хоризонтално а други вертикално. После тога се у прозору за унос откуца $y = ax + b$ и после уношења те наредбе добија се график праве са вредностима за a и b које су дефинисане положајем клизача, слике 15 и 16. Ако мишем померамо тачку на једном клизачу мењаћемо одговарајући параметар што ће се одмах видети померањем графика. Померањем тачке на другом клизачу изазивамо опет промену графика. На тај начин се веома очигледно демонстрира природа коефицијената у једначини праве. Интервали и корак промене сваког параметра дефинисаног клизачима мкогу се мењати. Селектујући клизач десним дугметом миша добијамо могућност да мењамо многа његова својства, слика 17. Тако можемо и да их обојимо, што мпже допринети лакшем праћењу експеримента.



Слика 15.



Слика 16.



Слика 17.

4.3 GeoGebra у настави

4.3.1 Увод

Динамичко јединство геометрије и алгебре у GeoGebra-и омогућавају ученицима једноставан експериментални прилаз математици. Они могу као сопствени учитељи самостално да напредују, индивидуално и откривачки да раде и уче. Кроз примере који следе предложићемо како може да се организује час математике уз употребу програма GeoGebra. Свакако, ово је само један покушај, настао из релативно мало експерименталних часова. Уз чешћу употребу рачунара и овог програма и бољу обученост, пре свега наставника, организација часа може бити и боља и занимљивија од предложене. Реакције ученика, повратне информације, такође ће утицати на будуће организације часова.

4.3.2 GeoGebra као алат за математичке експерименте

Пустимо ученике да сами откривају математичке ствари и односе помоћу GeoGebra-е. Дајмо им задатак који треба да реше и да на једном листу хартије или фолији помоћу GeoGebra-е испишу решење. Настава се не

мора одржавати у рачунарској учионице. Довољан је један PC са GeoGebra-ом, да би се организовао један групни рад.

4.3.3 Неке могућности

- Формулишите свој задатака што је могуће јасније и отвореније, како би ученици добили више простора да траже своје путеве за самостално решавање постављеног задатка.Учење је индивидуални процес тако да то можемо захтевати од ученика.
- Повежите индивидуални рад са тимским радом. Ако ученик ради у пару са другим учеником или у малој групи, често се дешава да своје идеје супротставља туђим и тада његове идеје могу да приме сасвим друге облике.
- Дозволите ученику да своја предосећања, идеје и резултате одштампа на папиру или да их препише у свеску. Такође је могуће да из GeoGebra штампате конструкцију и њен протокол.
- Овакав писани материјал може послужити као основа за дискусију у одељењу о идејама и резултатима. Треба дозволити да групе окупљене око задатка изложе своје идеје и резултате пред целим одељењем и да их образлажу и бране.
- За време рада са GeoGebra-ом потребно је да наставник буде доступан сваком ученику као саветник и помагач у примени одређених наредби GeoGebra-е. Тиме постижемо могућност да ученик у миру миже да размишља о математичком проблему и да тражи сопствене путеве решавања постављеног задатка.

4.4 GeoGebra као алат за презентацију

GeoGebra се може користи као динамички пројектор да би се материја изложила целом одељењу или да би се извео неки експеримент. За то је потребан један лаптоп или PC и пројектор. Излагање може почети са празним прозором GeoGebra-е или са унапред спремљеним конструкцијама. У другом случају потребно је приказати конструкциони протокол да би се корак по корак приказала и образложила конструкција.

4.4.1 Неке могућности

- Ученике при презентацији треба увлачити у математичку дискусију и њихове идеје треба одмах помоћу GeoGebra-е испробавати.
- Треба омогућити ученицима да своје идеје и сами испробавају пред одељењем користећи GeoGebra-у.
- Понудите ученицима да направе и неке реферате помоћу GeoGebra-е. Пошто је GeoGebra бесплатна, сваки ученик може понети своју копију програма кући.

4.4.2 Полазне тачке за постављање задатака

Ако припремамо задатке које ћемо решавати помоћу GeoGebra-е имамо различите могућност:

- *Постављање отвореног проблема:* припремамо ученике за математички експеримент тако што ћемо проблем и питања тако формулисати да ученици имају могућности да сами откривају и траже индивидуалне путеве за решавање постављеног задатка.
- *Конструкција слике:* остављамо ученике да сами понове конструкцију коју им показујемо као готову слику. Различити конструкциони протоколи које ученици добијају могу послужити за упоређивање поступака које су користили ученици.
- *Конструкциони протокол:* ученицима се задаје конструкциони протокол који они треба да понове. Поједини конструкционе кораке можемо "сакрити" и пустити ученике да их "нађу".

Ове могућности можемо комбиновати. Такође, можемо и сами додавати неке нове модалитете задавања проблема, где и наставникова креативност долази до изражаја.

4.4.3 Примери отворених питања

Следећи примери илуструју могућност постављања отворених питања и треба да послуже наставнику за креирање сличних и потпуно нових.

1. Систем линеарних једначина са две непознате,
2. Оштар или туп угао,
3. Тангенте на кружницу,
4. Троугао и око њега описана кружница,
5. Линеарна једначина $y = kx + d$.

4.4.4 Пример 1.

Задатак: Одреди графички решење следећег система линеарних једначина:

$$g : 4x = -8,$$
$$h : x - 2y = 3.$$

Пробај у свесци да решиш дати систем.

Промени једначину g тако да нови систем нема решење, тј. да је његов скуп решења празан. Шта то значи геометријски? Напиши запажања и резултат у свеску.

Пробај да одредиш још неки систем две једначине са две непознате тако да је скуп решења тог система празан. Можеш ли рећи којим поступком можеш наћи више таквих система? Напиши запажања и резултат у свеску.

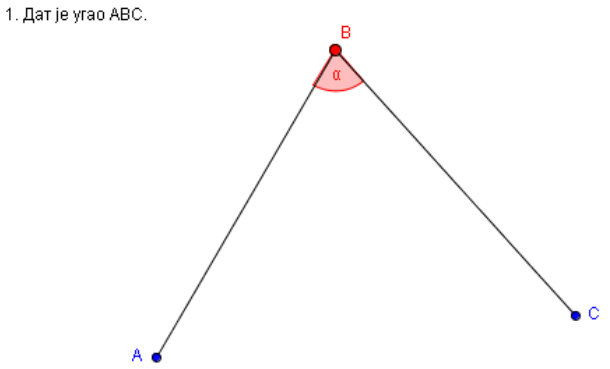
4.4.5 Пример 2.

4.4.5.1 Да ли је угао оштар или туп?

У XIII веку арапски математичар Абу Хасан је објашњавао како се брзо и сигурно може утврдити да ли је дати угао $\measuredangle ABC$ оштар, туп или прав. Абу Хасан предлаже да се нацрта круг пречника AC , па према томе да ли је тачка B ван круга, у кругу или на кружници, доноси закључак да ли је

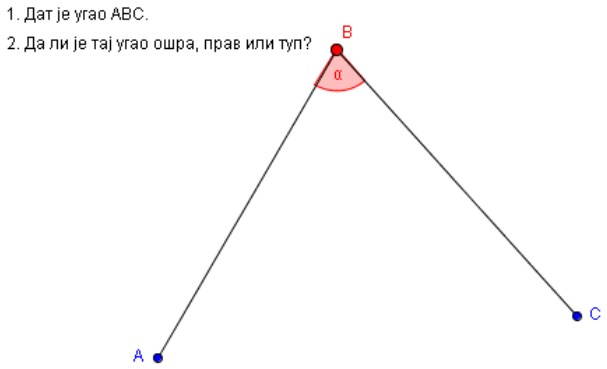
угао $\measuredangle ABC$ оштар, туп или прав. Докажимо или оповргнимо Абу Хасанову конструкцију.

1. Отворимо фајл "Da li je ugao oštar ili tup.ggb" и поставимо прву ситуацију, као на слици 1.
   Комбинација CTRL+"клик" на текст отвара одговарајући фајл: Da li je ugao oštar ili tup? У случају неких проблема покрените Uputstvo.



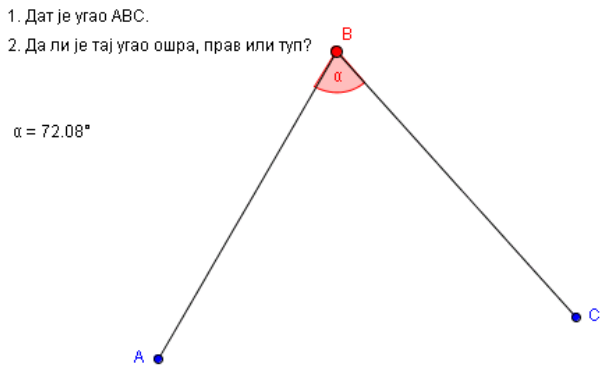
Слика 18.

2. Једним кликом на навигатору отварамо текст са питањем "Да ли је угао оштар илои туп?"



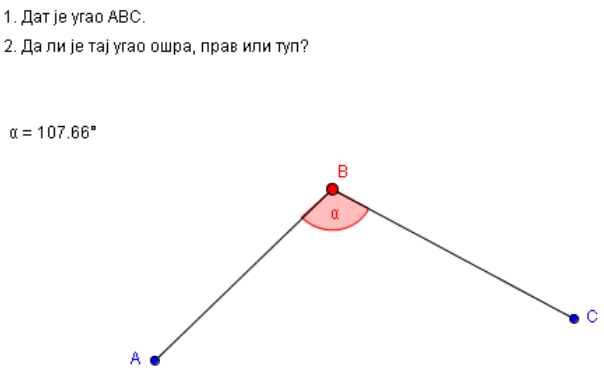
Слика 19.

3. Нови клик на навигатору избацује вредност угла $\alpha = \measuredangle ABC$.



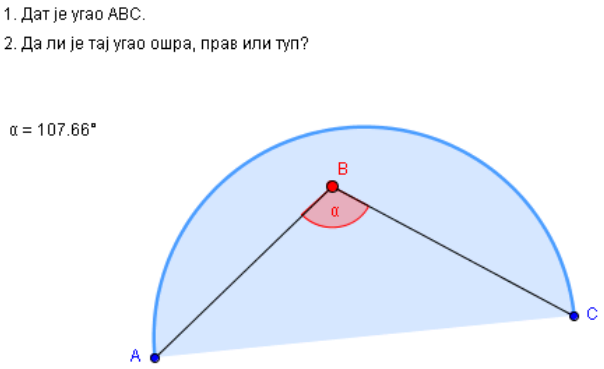
Слика 20.

4. Ученике замолимо да померањем тачке B и читањем вредности угла α покушају да утврде када је угао оштар а када туп.



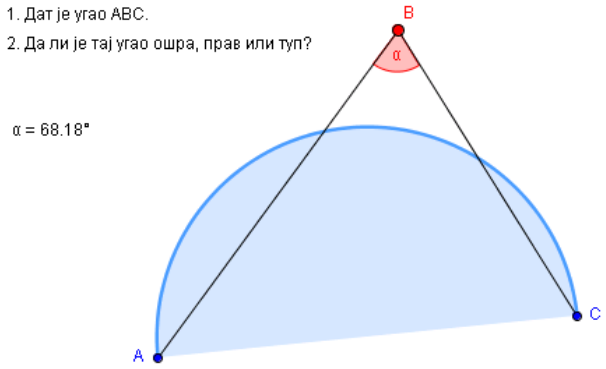
Слика 21.

5. После извесног времена, када ученици саопште своје идеје и запажања, продискутујемо предлоге за решење постављеног задатка. Ако се нико не јави са неким предлогом, наставник може сам анализирати постављени задатак и на тај начин покренути и ученике да бисмо учинили следећи корак. Тај корак се састоји у конструкцији полукруга са пречником AC.



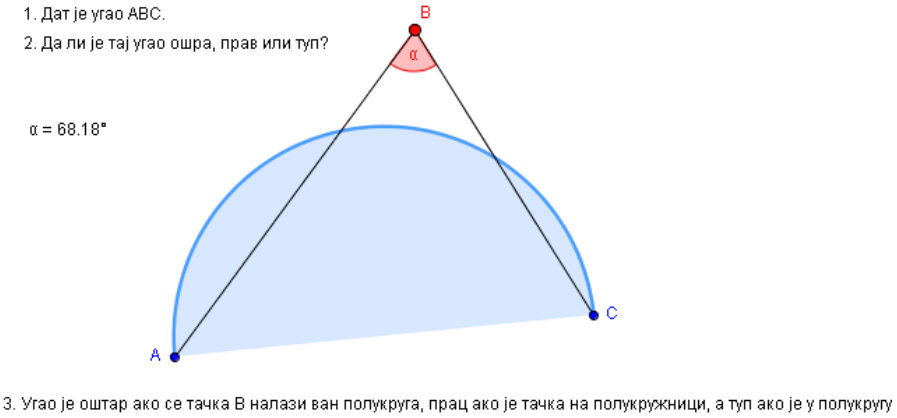
Слика 22.

6. Поново замолимо ученике да померају тачку B и саопште своја запажања. Очекујемо да ће читајући вредност угла α и гледајући положај тачке B закључити да је угао $\alpha = \measuredangle ABC$ оштар када је тачка B ван круга и да је туп када је она у кругу.



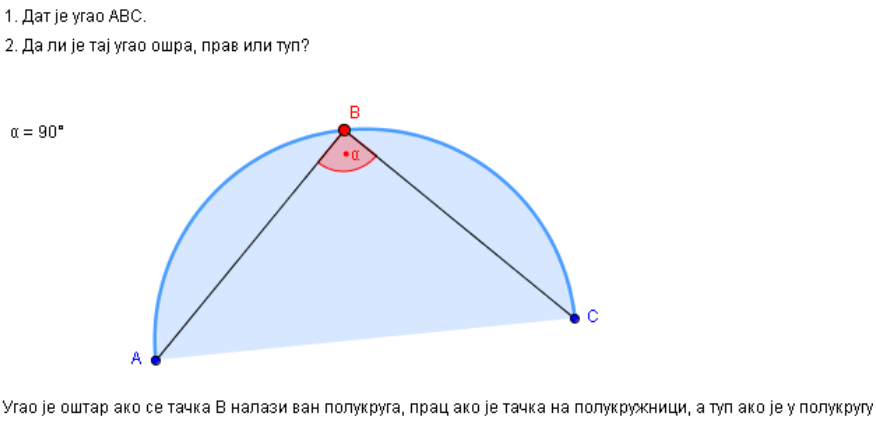
Слика 23.

7. Тај закључак ћемо записати као тврђење које ћемо и доказати.



Слика 24.

8. Питање које остаје да се размотри односи се на положај тачке B на кружници. Лаганим померањем клизача, можемо постићи да је тачка B на кружници и да је истовремено угао $\alpha = \measuredangle ABC = 90^\circ$.



.

Слика 25.

4.4.6 Теме за дискусију

1. Да ли смо добро поставили редослед слика?
2. Да ли смо добро груписали слике?
3. Да ли је потребно више текста?
4. Да није можда сувише текста?
5. Како би сада изгледао доказ тврђења које смо изрекли?
6. Зашто нисмо нацртали цео круг и затим померали тачку B?
7. Како објаснити опредељење да је довољно посматрати само полукруг?

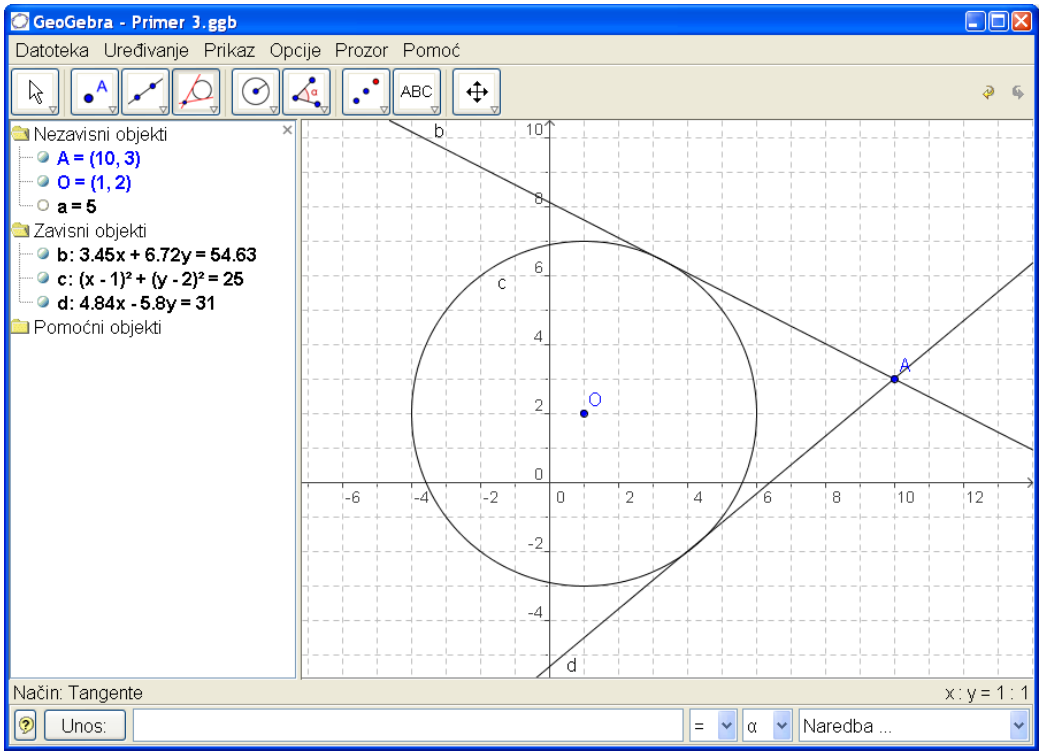
4.4.7 Пример 3.

Задатак: Конструиши покмоћу GeoGebra-е кружницу са центром у тачки $O=(1,2)$ и полупречником $r=5$. На ту кружницу постави тангенте из тачке $A=(10,3)$.

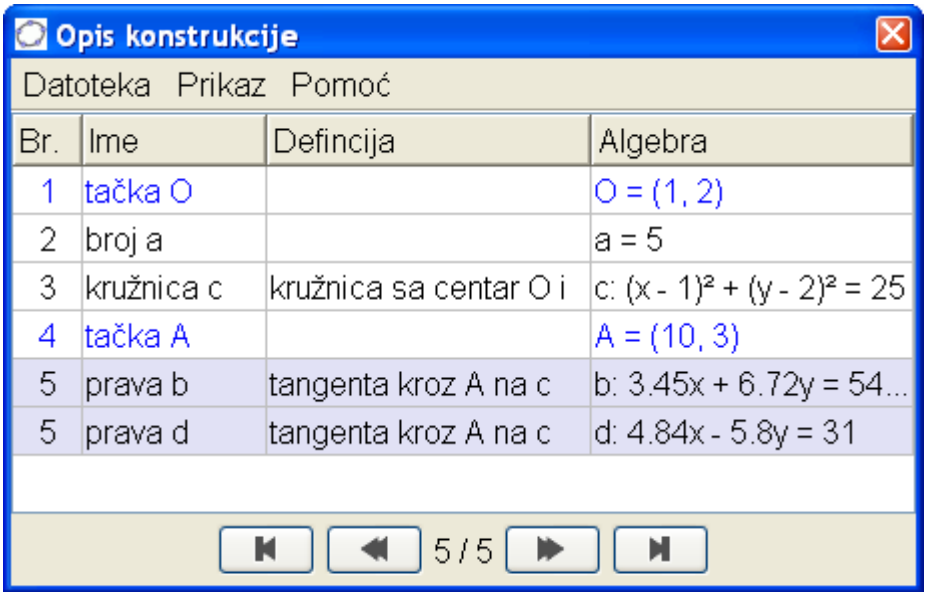
Шта се дешава када се тачка A помера мишем?

Како утиче положај тачке A на тангенте?

Напиши запажања и резултат у свеску.



Слика 26.



Слика 27.

4.4.8 Пример 4.

Посматраћемо линеарну једначину

$y = kx + d$

и проучавати значење параметара k и d варирајући различите вредности за k и d. То можемо постићи тако што ћемо у пољу за унос на дну прозора уписивати вредности за ове параметре (после сваког реда притисне се enter).

- $k = 1$
- $d = 3$
- $y = kx + d$

Параметре можемо мењати у алгебарском прозору (десни клик миша, *Уређивање*) или у пољу за унос:

- $k = -2$
- $k = 3$
- $d = -2$
- $d = -1$

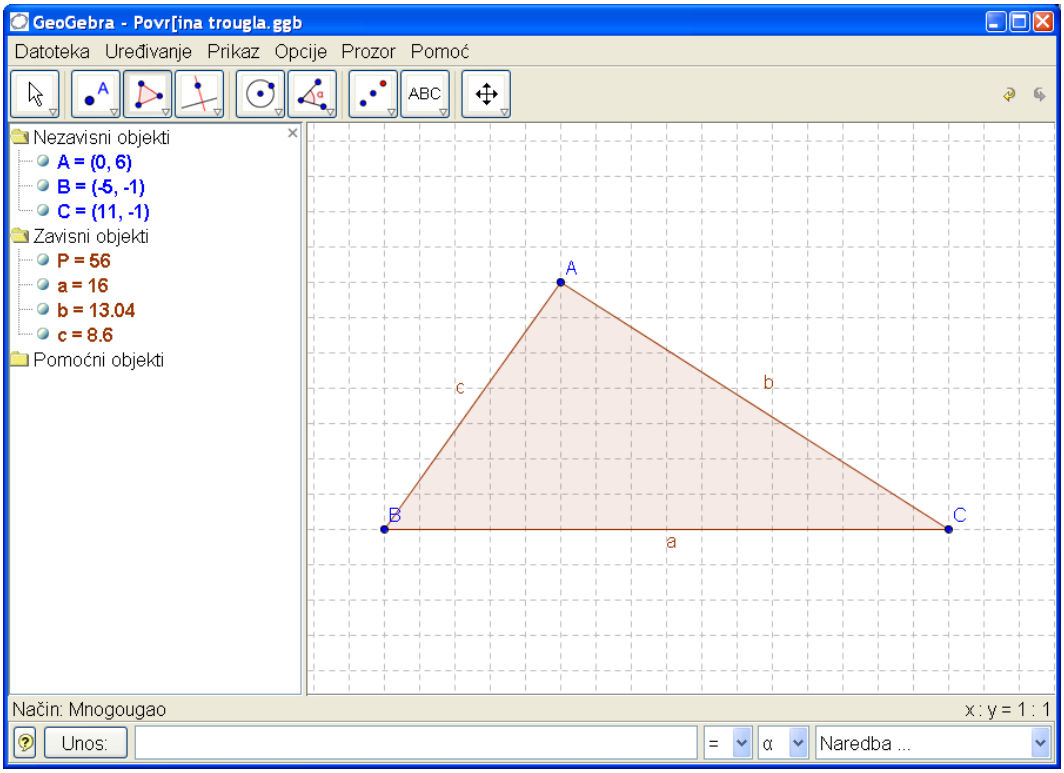
Параметре можемо мењати и једноставно, помоћу клизача или сттрелицама на тастатури.

4.5 GeoGebra као алат за припрему задатака

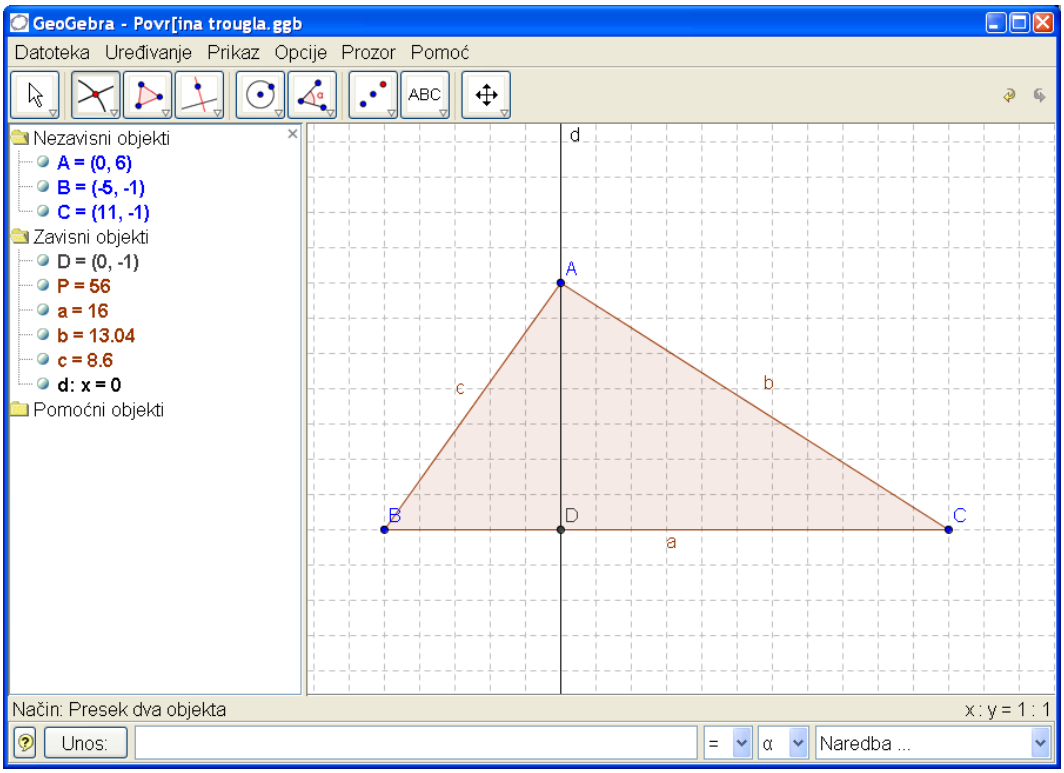
Наставник може да користи програм GeoGebra и за припрему задатака, школских, домаћих, за тестове, контролне или писмене. Од помоћи су му алгебарски и геометријски прозор, текст уз који може да даје и неке израчунате величине.

4.5.1 Пример 5.

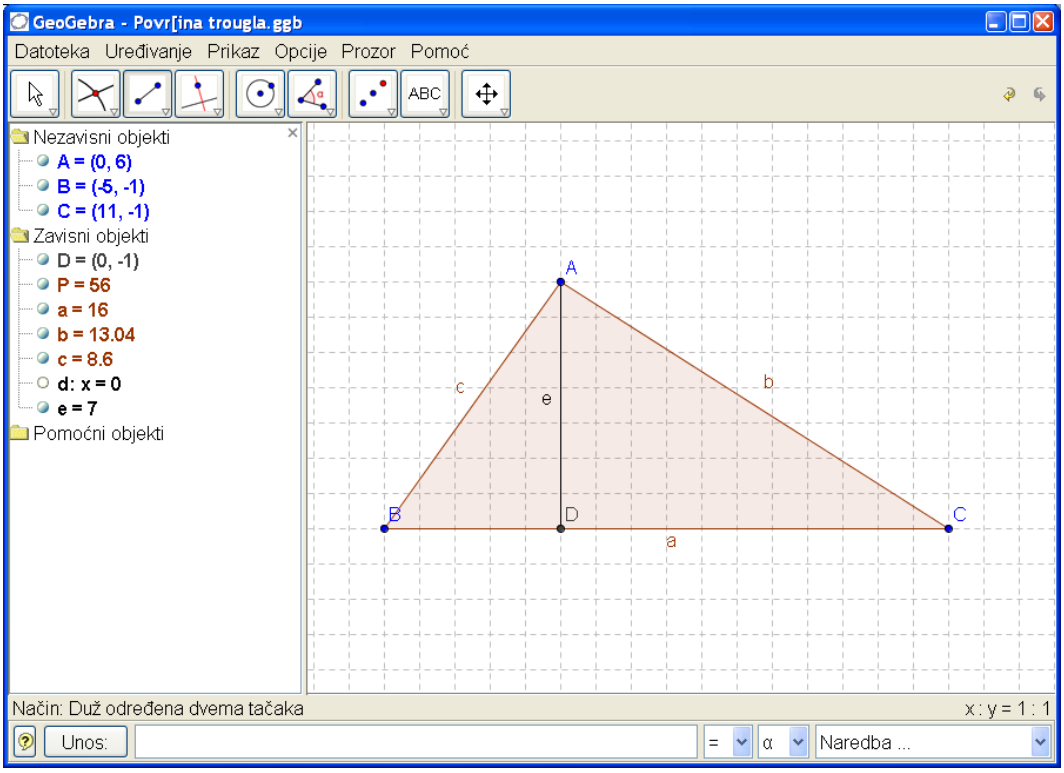
Нацрта се троугао $\triangle ABC$. На слици се појаве обележена темена и странице, слика 28. Из врха троугла A спусти се нормала на страницу a, слика 29. Затим се одреди висина e троугла из темена A, слика 30. У алгебарском прозору појави се ознака троугла P и његова површина, $P = 56$. У алгебарском прозору се појављују дужине страница троугла и његове висине e. Ако се затвори алгебарски прозор, слика 31, нестају величине везане за троугао. Сада се преко текста могу задати неки елементи троугла, а од ученика се може тражити да нађе неке друге величине. На пример, задају се страница $a = 16$, $b = 11.4$ и висина $e = 7$, а тражи се страница c или површина троугла или његов обим. Померајући слободно једно или више темена троугла наставник добија многобројне варијанте истог задатка. Тако сваки ученик може добити свој задатак. Решење сваке комбинације наставник има сакривен у алгебарском прозору. Ако га отвори, може их приказати, слике 32 и 33.



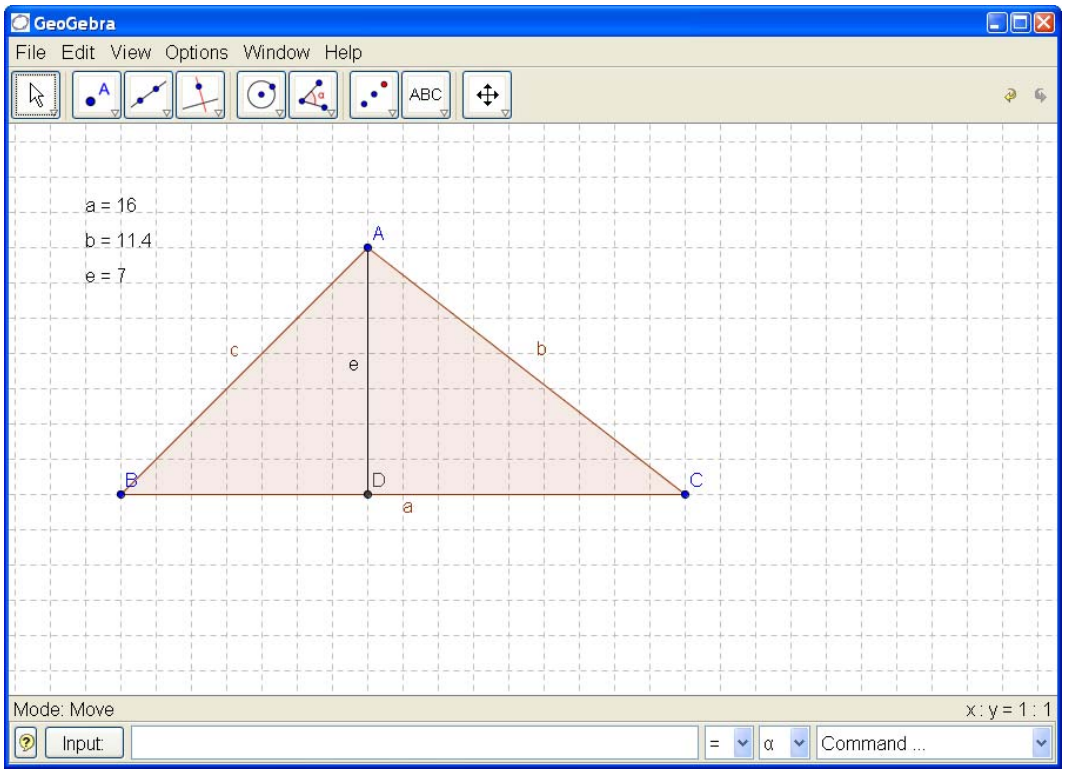
Слика 28.



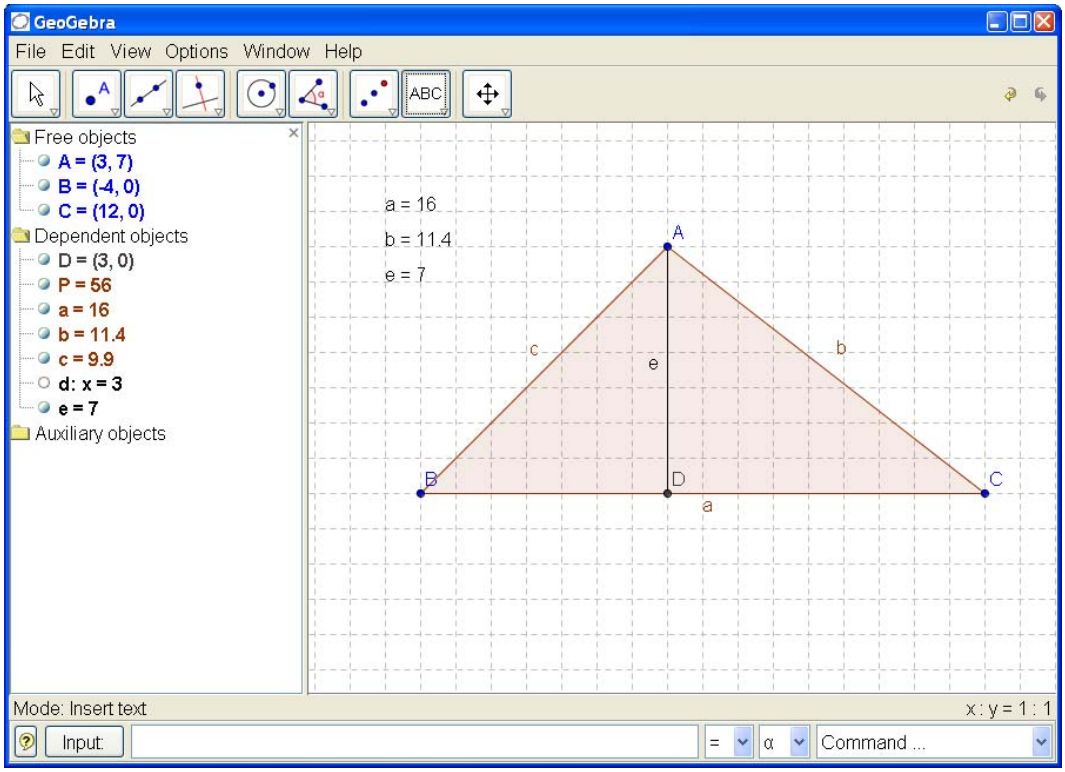
Слика 29.



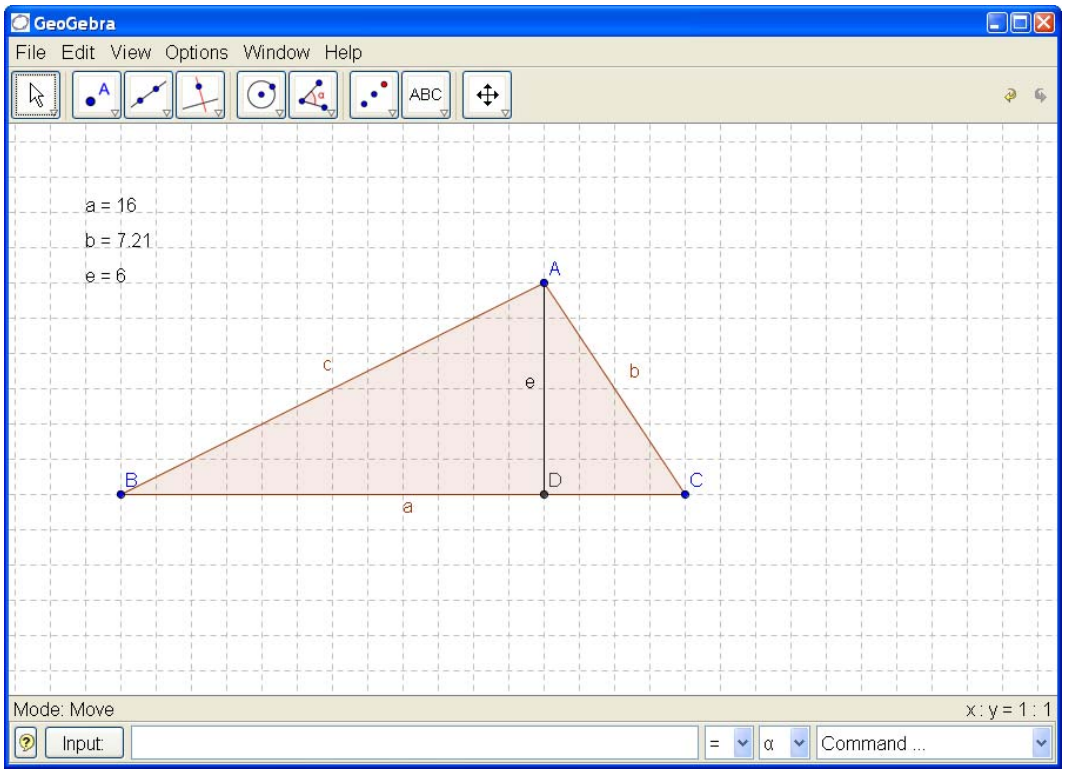
Слика 30.



Слика 31.



Слика 32.



Слика 33.

Код сложенијих задатака, где се резултати не појављују непосредно, као у претходном примеру, моу се решења добити комбинацијом наредби које су на располагању. У том случају наставник, а и ученици, морају прво да доведу решење задатка до рачунских радњи. Рачунање се тада може радити и етапно или само једном на крају.

GeoGebra пружа могућност за креирање сложенијих слика на основу којих, уз дате бројчане податке, ученици треба да решавају одређене задатке. За успешан цртеж на располагању су различите врсте линија, различите величине тачака, различите боје и њихове нијансе.

Наставник може припремљене сложеније цртеже да ученицима „подели“ електронском поштом, на неком носиоцу података или само на папиру. Решења задатака, домаће задатке, ученици треба да пишу у свеске, али их треба подстицати да задатке могу предати на неком носиоцу података, преко електронске поште или на посебном папиру. Као домаћи задатак може се захтевати да ученици варирају наставникову тему и сами састављају сличне задатке.

Лепа ученичка решења могу се приказати свим ученицима у одељењу. Добро би било да то изведе сам ученик који је решио задатак, а остали ученици и наставник треба све то да прокоментаришу.

На тај нечин се ученици стимулишу на самосталан рад, на изношење свог рада на дискусију и образлагање поступака и резултата.

4.6 GeoGebra на часу математике

Sama organizacija časa na kojem se nastava odvija uz upotrebu računara biće analizirana i sa tehničke i sa metodičke strane.

На часу који је предвиђен за обраду међусобног односа праве и кружнице могу се комбиновати рад на папиру, односно табли и рачунару. На почетку наставник или неко од ученика понови дефиницију кружнице, а затим нацрта кружницу на табли. Нека је центар кружнице тачка O и полупречник r.

Наставник: „Ако у истој равни у којој је кружница нацртамо праву p, какав може да буде њихов међусобни положај“?

Ученик: „Права може да сече кружницу, да је додирује или да нема заједничких тачака са кружницом!“

Наставник: „Од чега то зависи?“

Ученик: „Однос кружнице и праве зависи од растојања центра кружнице од праве.“

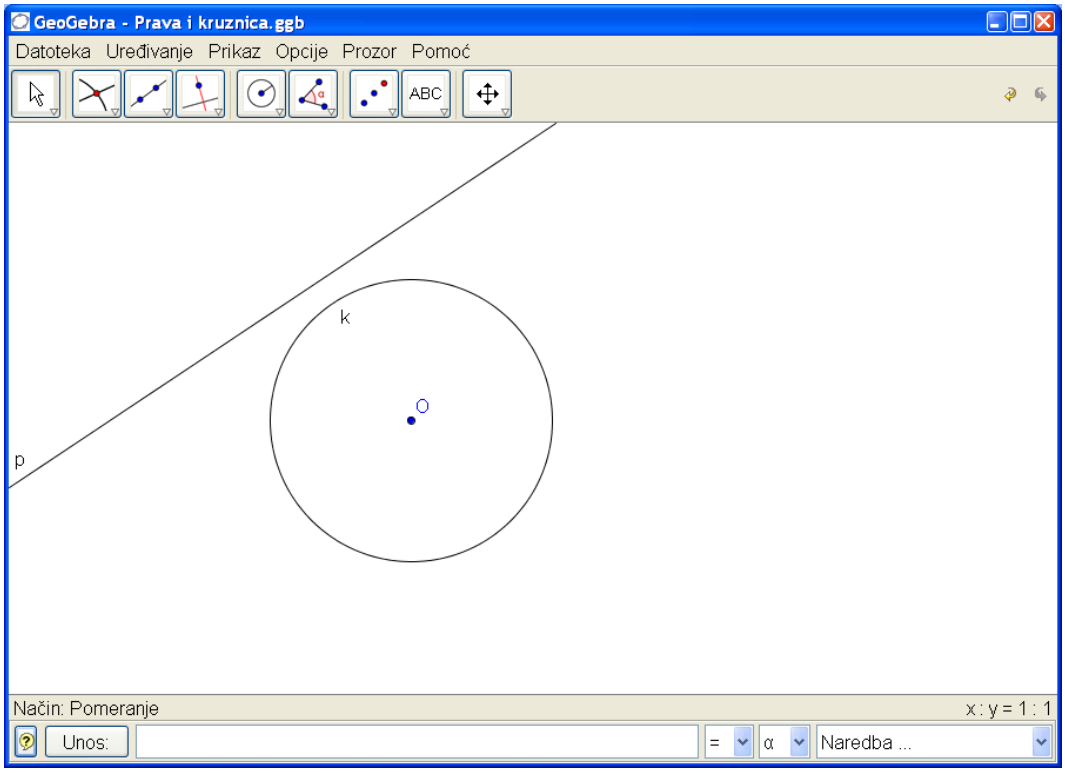
Наставник: „Како ћемо одредити то растојање?“

На овом месту потребно је обновити конструкцију нормале из дате тачке на дату праву. Уколико је мало времена за то или наставник оцени да понављање није потребно, конструкција се може избећи. У том случају се помоћу лењира поставља тражена нормала. Растојање центра кружнице од подножја нормале је дужина дужи одређена тим тачкама.

Ученике наставник сада може упутити на експерименте цртањем више правих, од којих неке секу, неке додирују а неке немају заједничких тачака са кружницом.

После краће дискусије и излагања ученика орезултатима својих експеримената, наставник може приступити рачунару и приказати неке од конструкција које су добили поједини ученици. Овај део рада би могао да уради и неки ученик који је овладао коришћењем програма GeoGebra толико да може нацрта кружницу и праву.

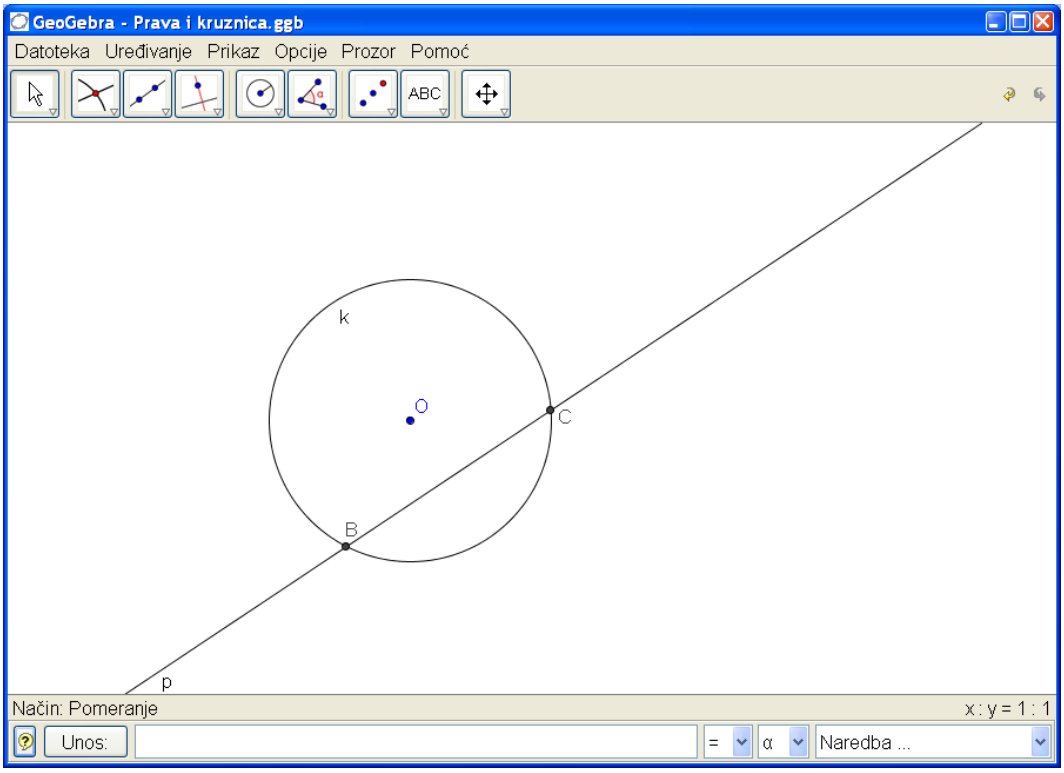
Слике могу ићи следећим редоследом. На првој слици се нацрта права и кружница са изабраним центром и датим полупречником, слика 17.



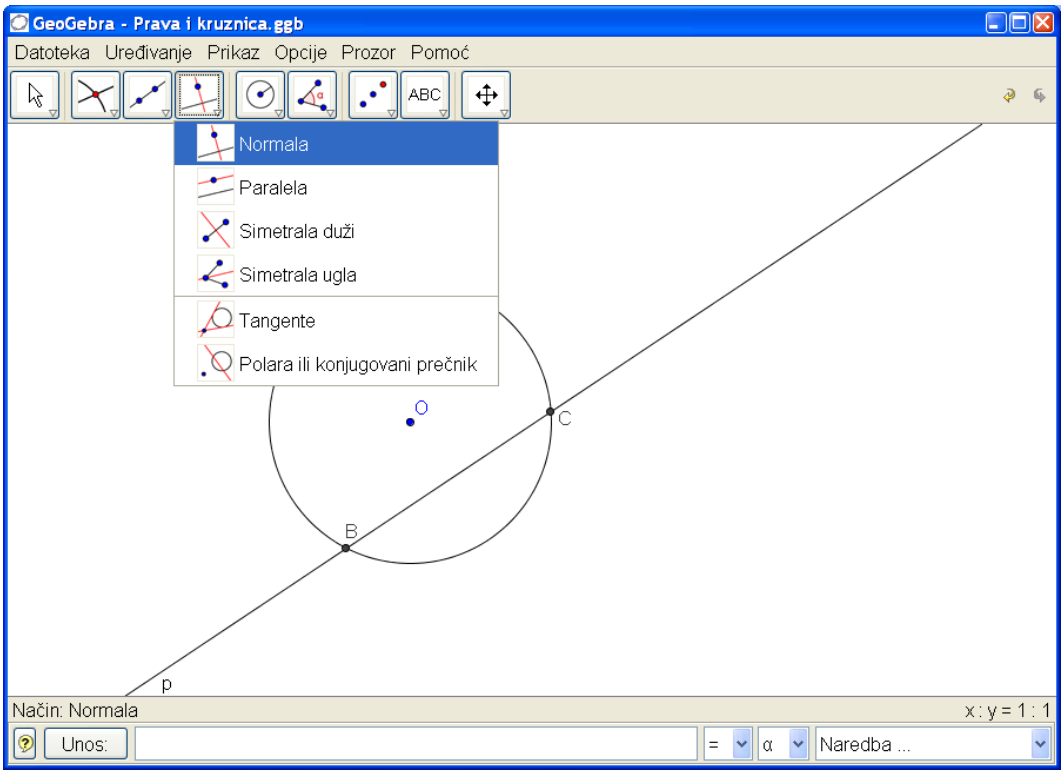
Слика 34.

Овде је важно приметити да померање праве паралелно првобитном положаја можемо остварити селекцијом било које тачке праве и повлачењем праве, слика 35. Селекција се изводи кликом на леви тастер миша. Наравно, треба демонстрирати и да се кружница може померати тако што се селектује и помера центар. Померањем објеката може постићи да се секу, додирују или да немају заједничких тачака. Избором опције за конструкцију нормале, слика 36, из центра кружнице на праву, добијамо праву која дату праву сече у тачки коју одређујемо опцијом пресек, слика 37. Затим нормалу учинимо невидљивом, слика 38, и обележимо дуж одређену центром и подножјем нормале, слика 39. Померањем праве мењамо и дужину уочене дужи. На тај начин приказујемо како се међусобни однос кружнице и праве мења а са тим и дужина посматране дужи.

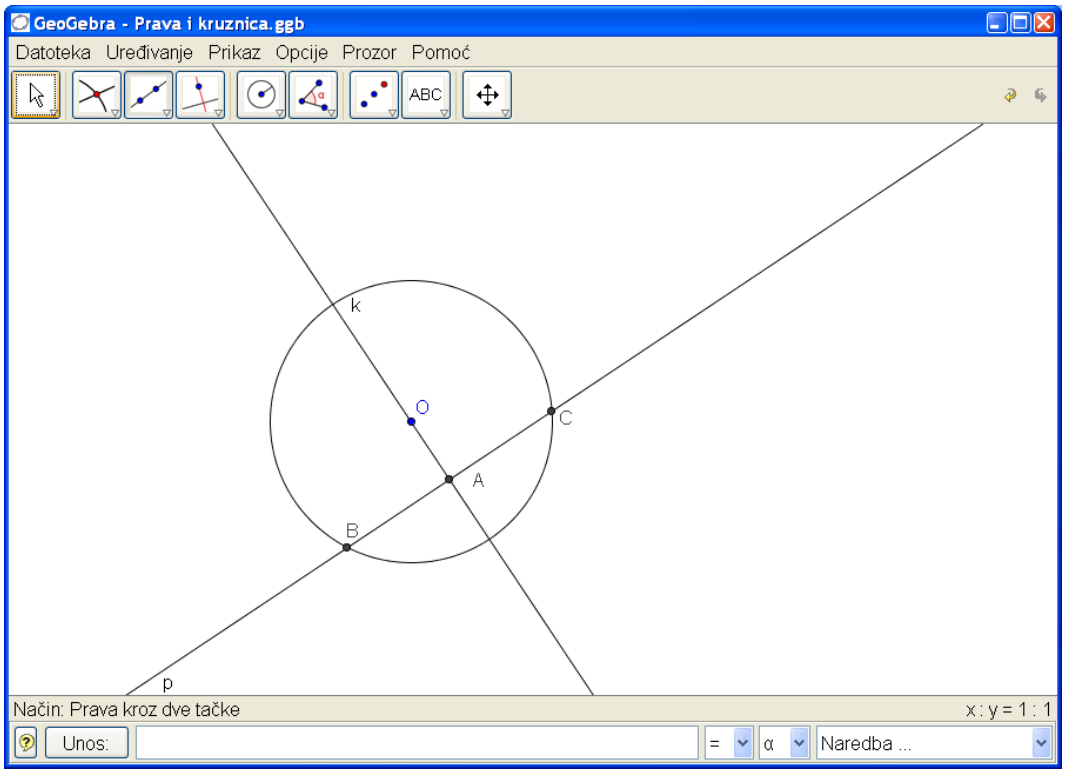
Кључни моменат наступа када мало дебљом дужи нацртамо полупречник кружнице одређен центром и пресеком нормале и кружнице, слика 40. Уз опцију „Дозволи пресек у продужетку“, слика 41, добијамо могућност да полупречник увек видимо, слика 42. Померајући или праву или кружницу и пратећи однос растојања центра кружнице од праве и полупречника кружнице наводимо ученике да изведу правилан закључак.



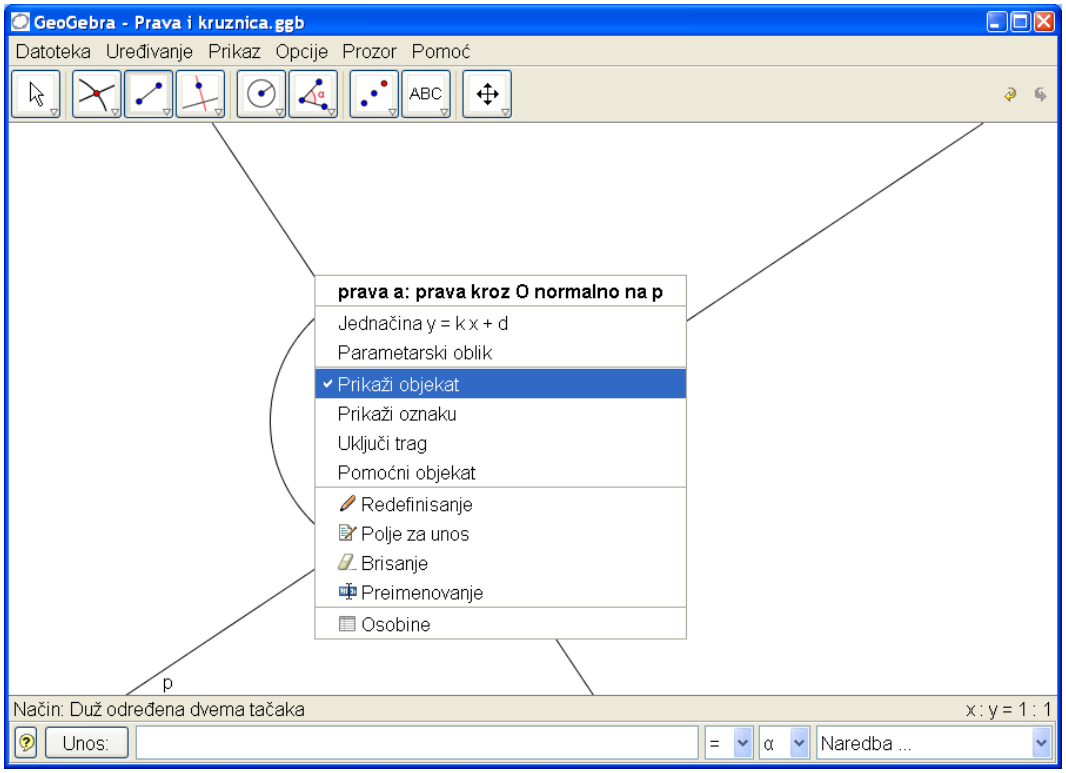
Слика 35.



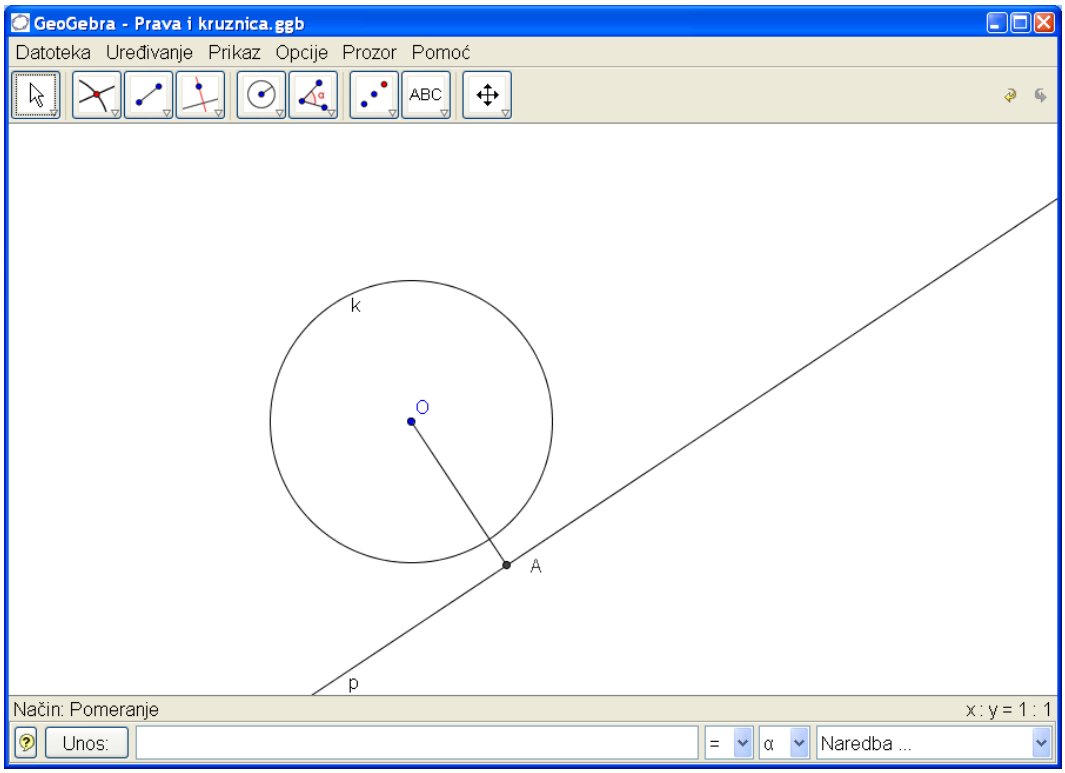
Слика 36.



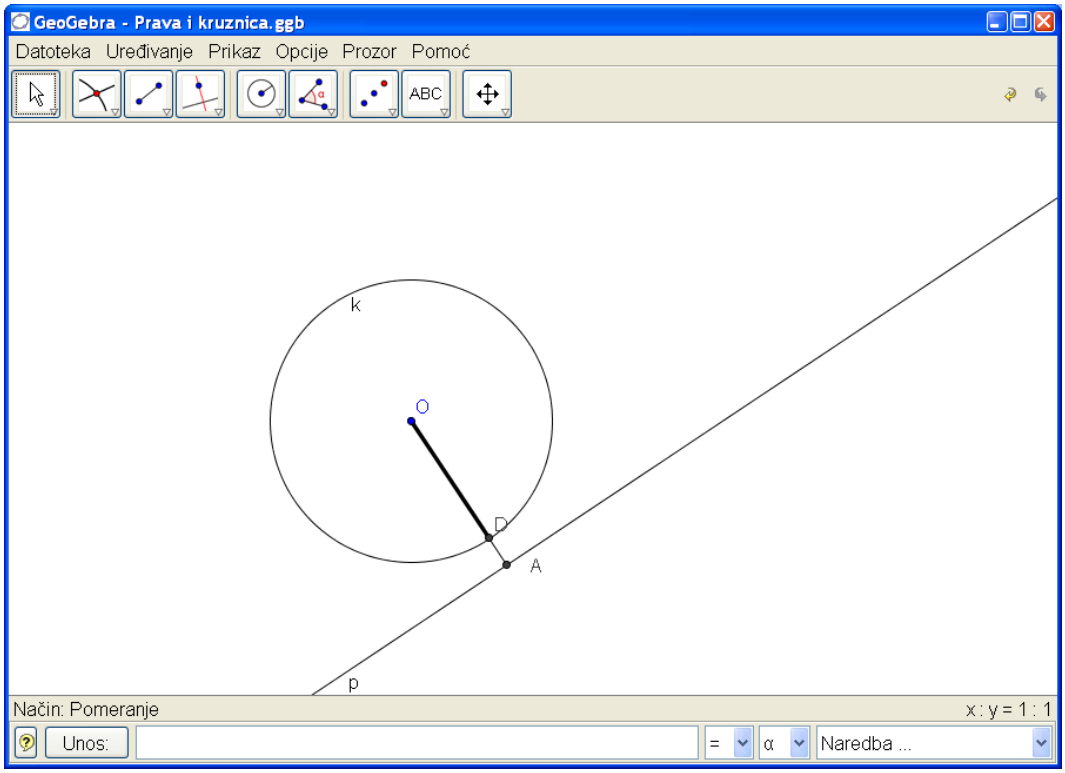
Слика 37.



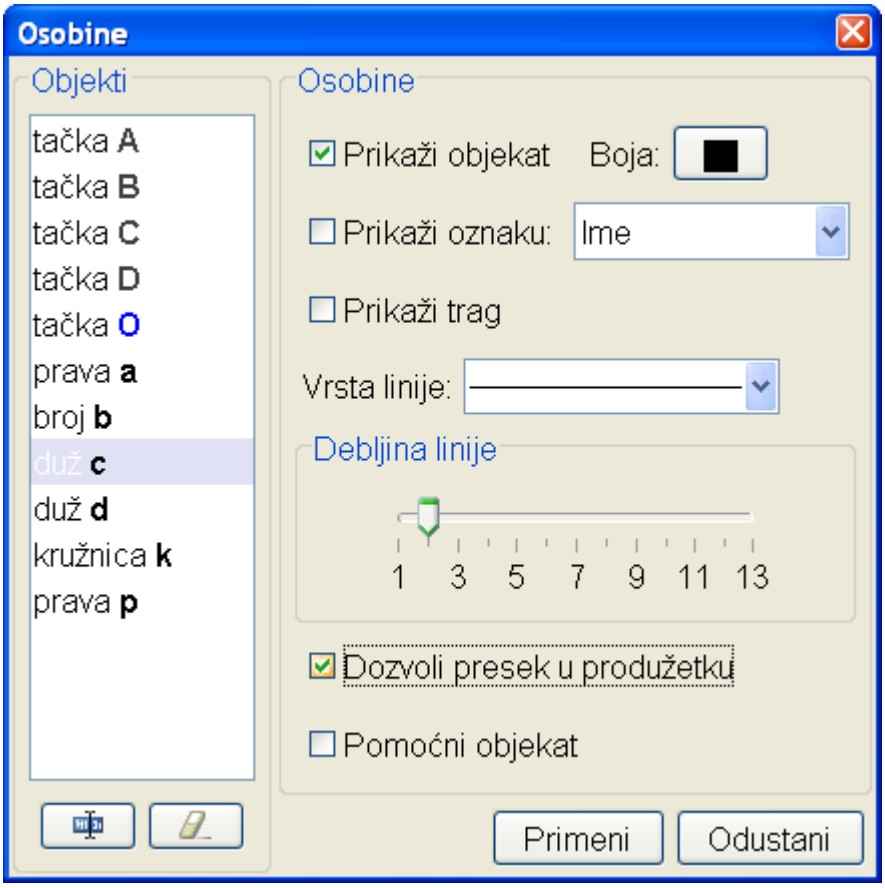
Слика 38.



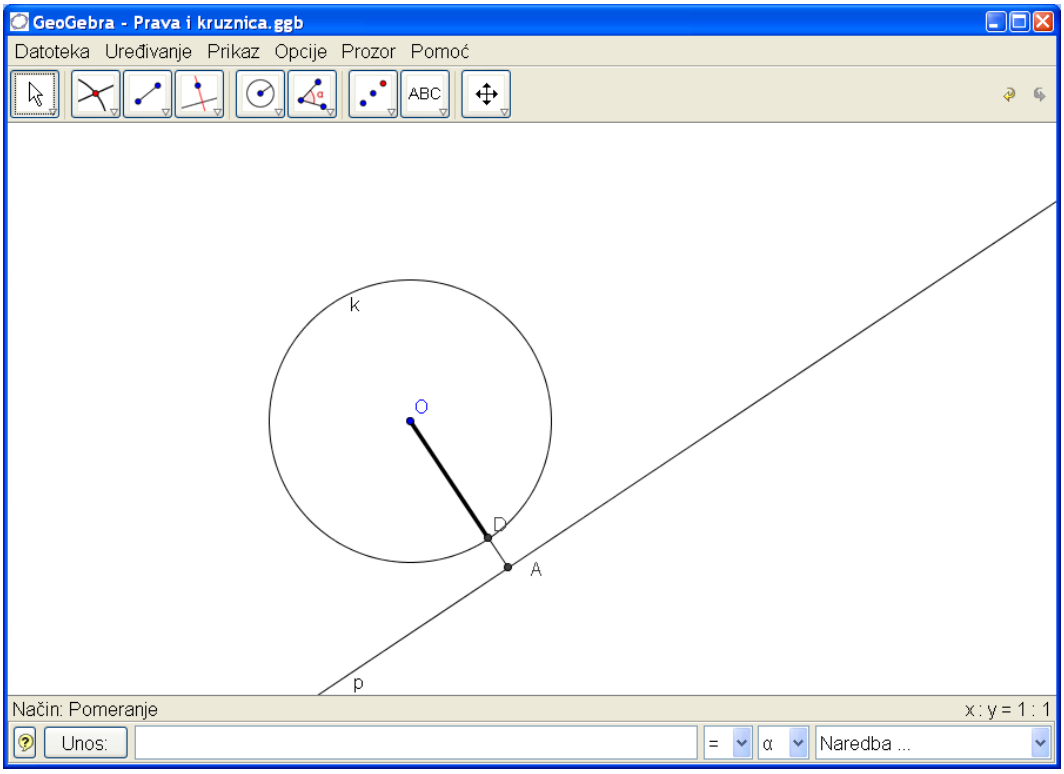
Слика 39.



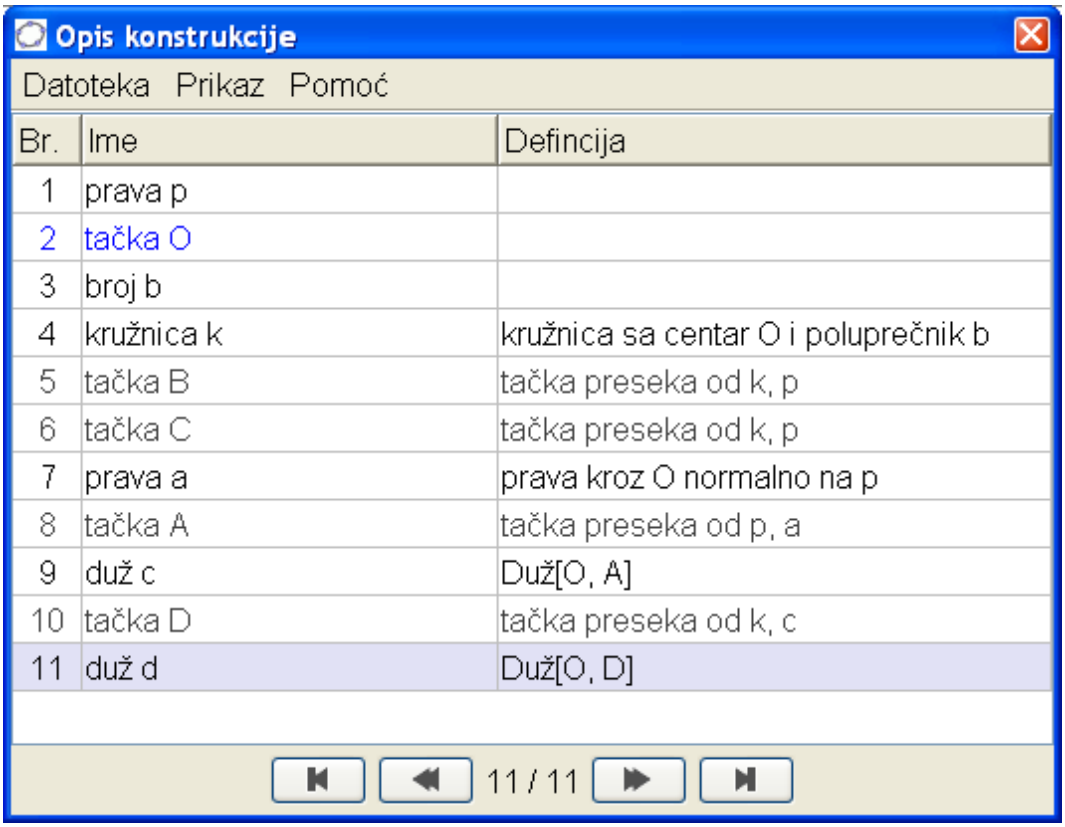
Слика 40.



Слика 41.



Слика 42.



Слика 43.

Опис конструкције, слика 43, садржи све наше кораке.

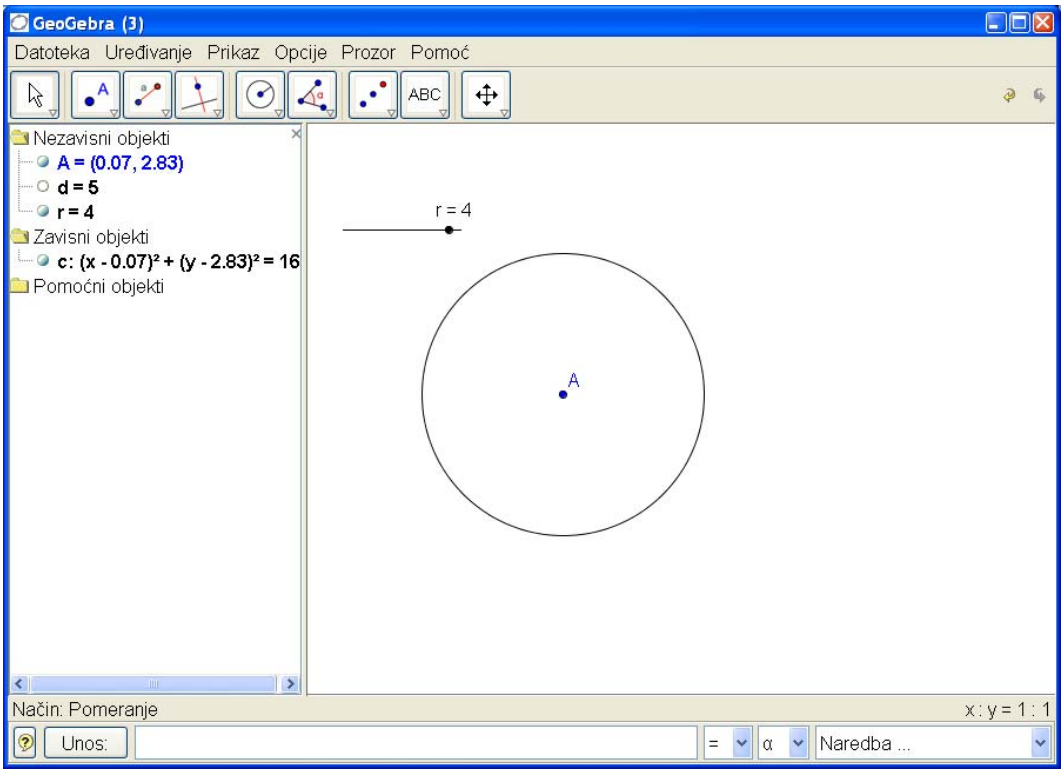
Друга могућност за стварање ситуације за експериментисање и довођење ученика до правилног закључка је следећа. Прво се зада полупречник кружнице

$$r = 4$$

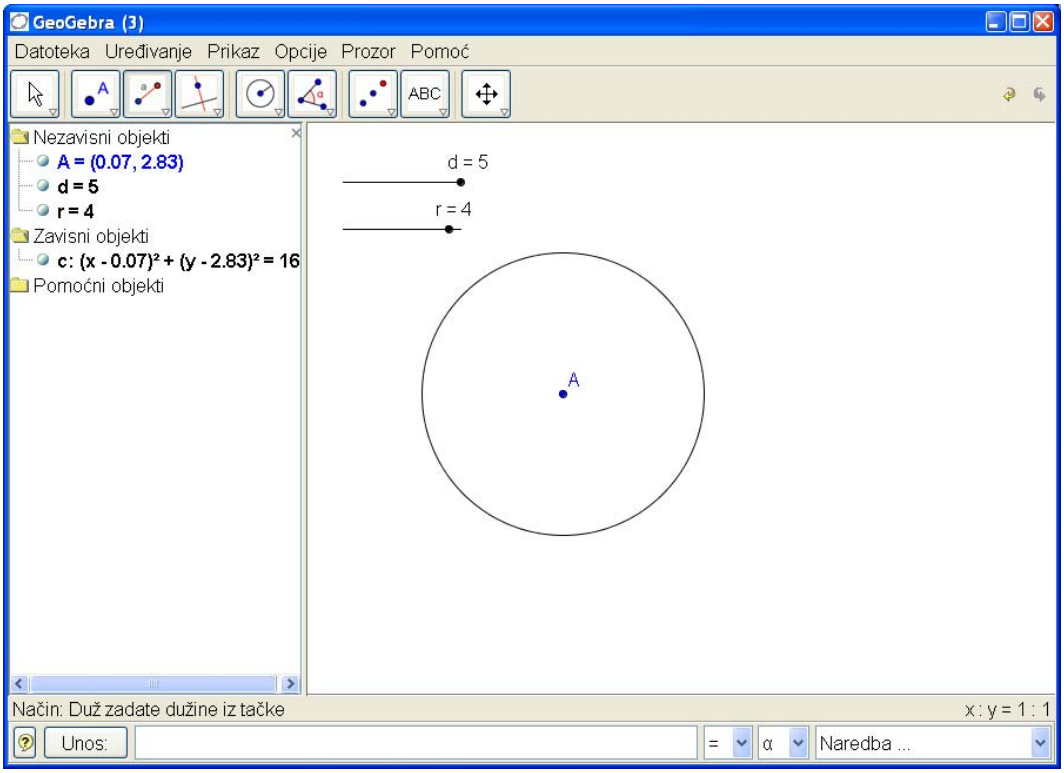
као клизач. Затим се нацрта кружница са центром у произвољној тачки A са полупречником r, слика 44. Онда се изабере клизач

$$d = 5,$$

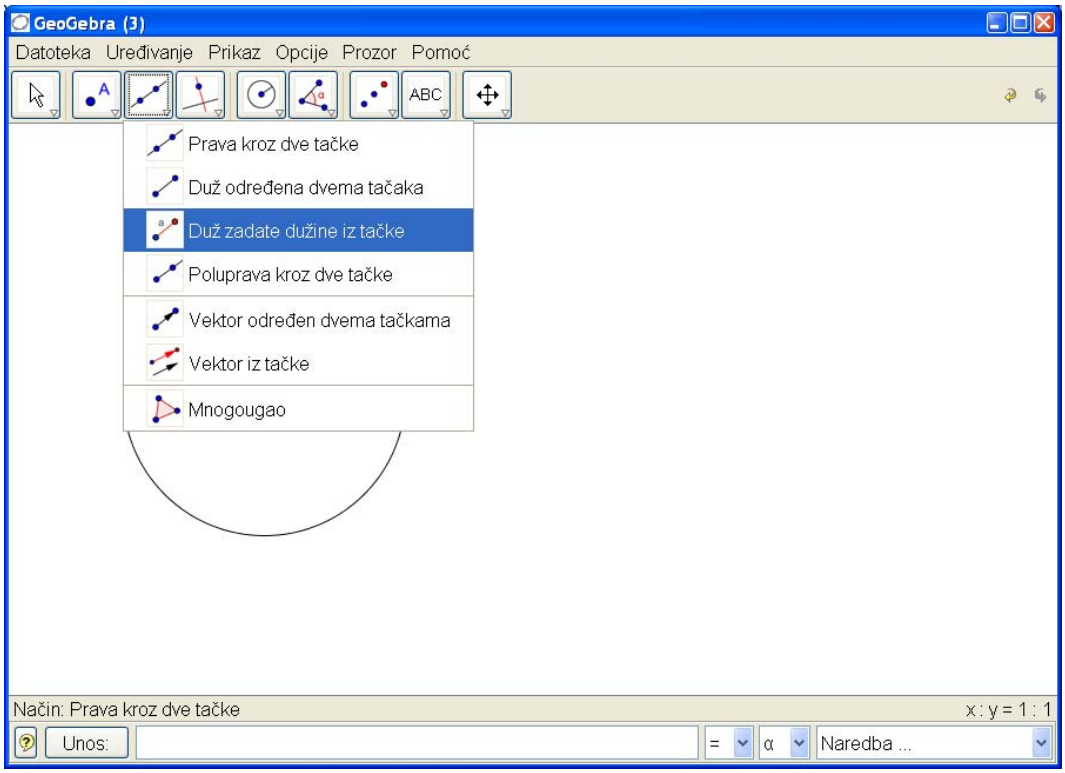
слика 45, а помоћу њега дуж AB дужине d, слике 46 и 48. Потом се поставља нормала из тачке B на дуж AB, слика 47. Померањем клизача r и d мењамо полупречник кружнице и растојање центра од праве. Свакако, овде можемо поново померати кружницу мењајући положај њеног центра. На овом месту ученици могу дискутовати и предлагати закључке о међусобном положају кружнице и праве.



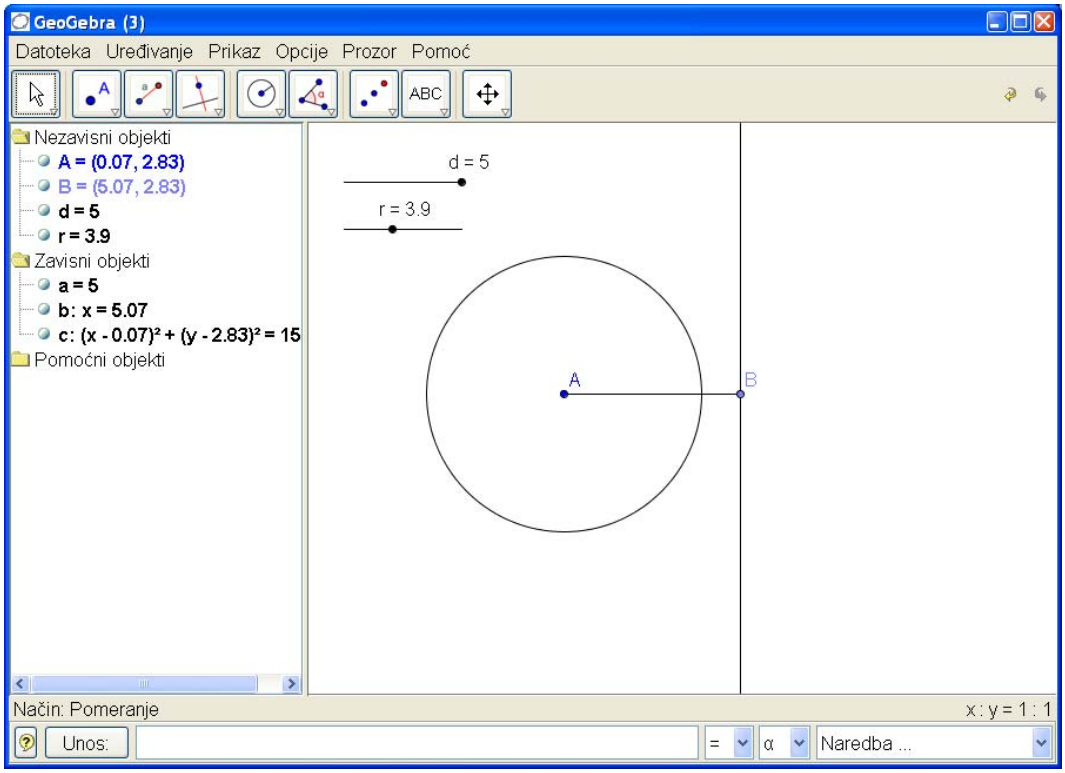
Слика 44.



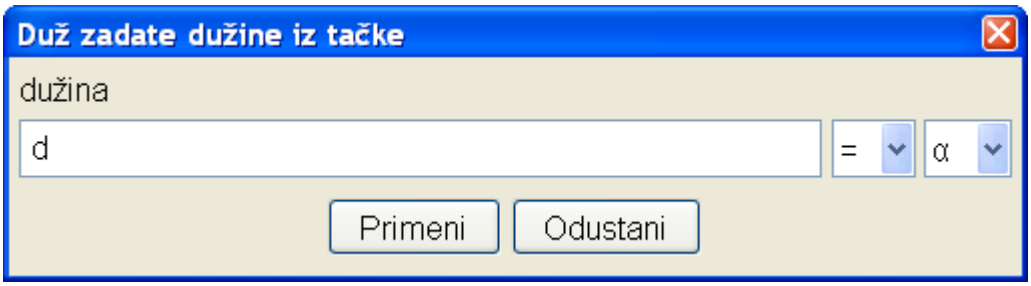
Слика 45.



Слика 46.

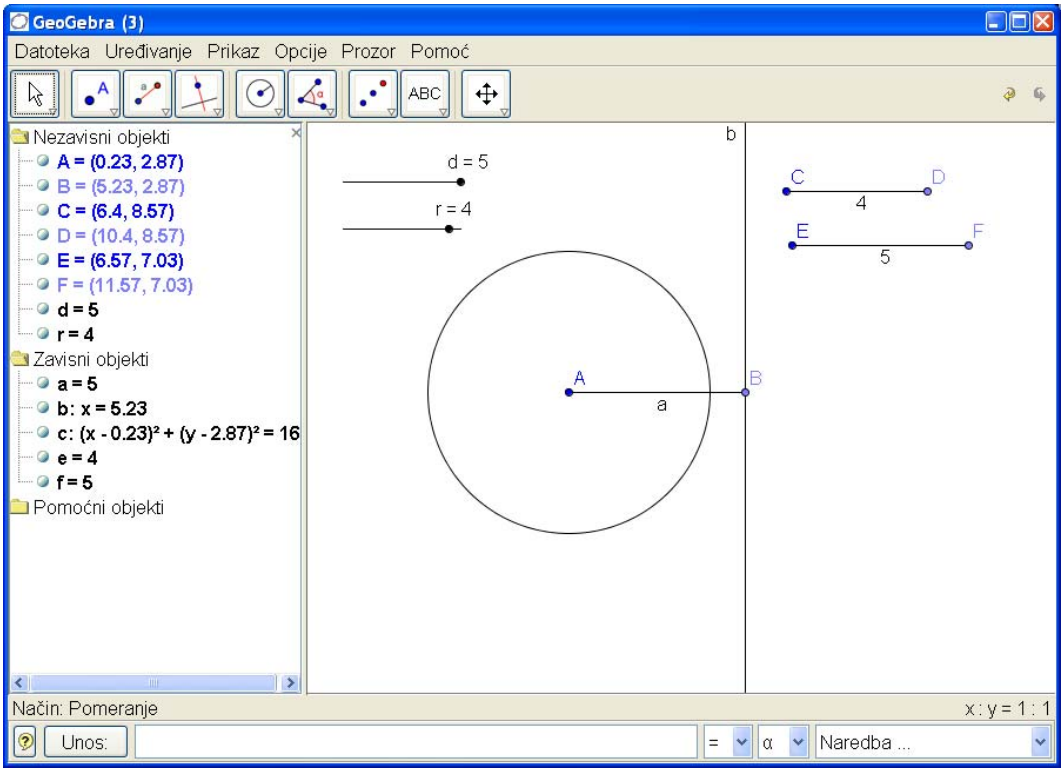


Слика 47.

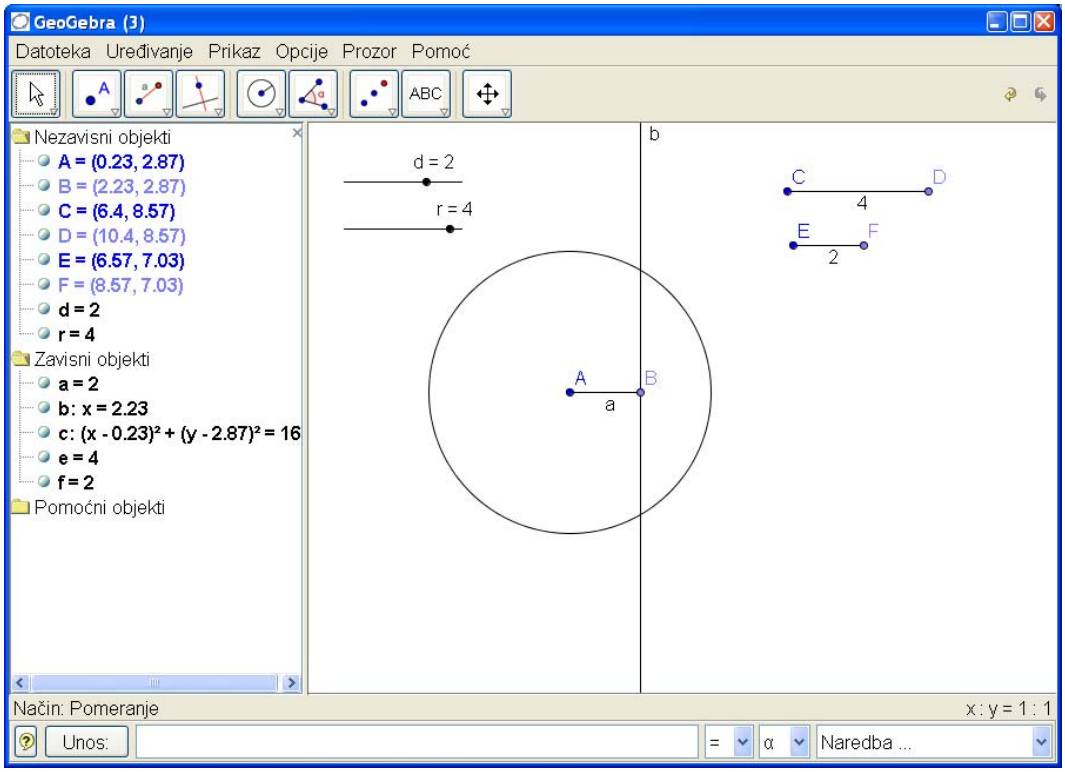


Слика 48.

Дискусија о међусобном односу кружнице и праве може бити благо усмерена ако се на радној површини додају две дужи. Прва из неке тачке C са дужином r и друга из неке тачке E са дужином d. На слици 49 су то дужи CD и EF. Ако изаберемо особину ових дужи да се појављује само њихова дужина, мењајући на клизачима полупречник и растојање центра од праве, мењаће се и дужине ових дужи. Сада само уз претходну акцију треба посматрати и однос кружнице и праве. Закључак се брзо намеће и ученици могу лако да га изведу.



Слика 49.



Слика 50.

Претходна два поступка приказујемо и на CD-у као мале филмове.

На већини часова наставник може слично поступити. Његова припрема часа састоји се у пчанирању питања и конструкција које ће изводити на часу, самосатаљно или уз помоћ ученика. Да би час текао пожељним темпом наставник треба да има потуно јасан план и довољну сигурност у раду са GeoGebra-ом. Прекидање тока конструкције уз питања упућена ученицима:

„Шта даље?“,

„Како даље?“,

„Шта закључујемо из овога?“

итд. требало би да „увуку“ ученике у дискусију и градиво.

За задатке које би ученици самостално решавали наставник може да постави неку од ситуација, на табли и на рачунару, а затим да од ученика тражи решење. На пример, наставник нацрта кружницу и тражи од ученика да нацрта праву која сече кружницу. Ученик може да бира две тачке за праву на различите начине. Питање које наставник може да постави може да гласи: „Изабери једну тачку тако да независно од избора друге тачке (наравно, различите од прве тачке) те две тачке одређују праву која сигурно сече кружницу“. Тако се подсећамо на то да кружница дели раван на унутрашњу и спољашњу област и да је један од критеријума за утврђивање која област је унутрашња управо пресек праве и границе те обласи (овог пута кружнице).

4.7 KISS Принцип

KISS је акроним од "Keep it small and simple!", што би се могло превести као "Направи једноставно и прегледно". Овај принцип се среће у информатици и био је водећа идеја при прављењу GeoGebra-е. Као образовни софтвер она треба да омогући једноставно манипулисање како би осталоместа за ученике да откривају математику и да експериментишу. Образовни софтвер тражи поред математичког знања и знање и вештину његовог коришћења. Ове прпреке би требале због тога да буду што је могуће мање. Због тога се GeoGebra оријентисала на алгебарско задавање блиско школској нотацији. Праву можемо задати као

$$p : 3x + 4y = 2,$$

а функцију као

$$f\left(x\right) = 3x - 2.$$

Geogebra наредбе даје на изабраном језику. До сада је припремљена на више од 38 језика.

Project: 06SER02/02/003

Mathematical and visualization software packages

Djurdjica Takači

DERIVATIVES OF HIGER ORDER FOR FUNCTIONS FROM R^n to R^m

S. Pilipović, A Takači, Dj. Takači

The purpose of this note is to give an elementary and unified approach to the notion of the derivative $F^{(p)}$, $p \in \mathbf{N}$, of a function F from $\mathbf{R}^n$ into $\mathbf{R}^m$. It is supposed that the students only have the elementary knowledge from the differential calculus and linear algebra. In the standard textbooks this notion is served to students on the basis of their knowledge of normed space of continuous linear mappings from one into another normed space.

The distance in Euclidean space $\mathbf{R}^n$ is denoted by $\|\ \|_{\mathbf{R}^n}$ or by $\|\ \|$ and the usual orthonormal basis by $\mathbf{e}_1,...,\mathbf{e}_n$. A function $A : \mathbf{R}^{n_1} \times ... \times \mathbf{R}^{n_s} \to \mathbf{R}^n$ is s-linear if it is linear with respect to any variable when all others $s-1$ variables are fixed. Recall the well known facts from the linear algebra: *(LA1) A mapping* $A : \mathbf{R}^{n_1} \times ... \times \mathbf{R}^{n_s} \to \mathbf{R}^n$ *is* $s-$ *linear if and only if it is of the form*

$$A(\mathbf{x}_1, \dots, \mathbf{x}_s) = \sum_{\substack{i_1=1,\dots,n_1 \dots \\ i_s=1,\dots,n_s}} \alpha^{i_1 \dots i_s} x_1^{i_1} \dots x_s^{i_s}$$

where $\alpha^{i_1 \dots i_s} \in \mathbf{R}$, $i_1 = 1,..., n_1,..., i_s = 1,..., n_s$.

(LA2) The mapping $A : \mathbf{R}^n \to \mathbf{R}^m$

$$(x^1, \dots, x^n) \mapsto (A^1(x^1, \dots, x^n), A^2(x^1, \dots, x^n), \dots, A^m(x^1, \dots, x^n))$$

is linear if and only if the mappings $A^i : \mathbf{R}^n \to \mathbf{R}$, $i = 1,...,m,$ *are linear.*

In the sequel Ω will denote an open set in $\mathbf{R}^n$ and f a function from Ω to $\mathbf{R}$. If there exists the limit

$$\lim_{h^i \to 0} \frac{f(\mathbf{x}+h^i e_i) - f(\mathbf{x})}{h^i}, \quad (\mathbf{x} + h^i e_i \in \Omega,\ \mathbf{x} \in \Omega)$$

then it is called the partial derivative of f at x and it is denoted by $\frac{\partial f(x)}{\partial x^i}$.

The partial derivatives of higher order are defined as the partial derivatives of respected partial derivatives of lower order. Note, in general, $\partial^2 f(\mathbf{x}) / \partial x^i \partial \mathbf{x}^j$ and $\partial^2 f(x)/\partial x^j \partial x^i$, $i \neq j$, (if exist) are not equal.

Definition 1: *A function* f *is differentiable at* $\mathbf{x} \in \Omega$ *if there exists a linear mapping* $A : \mathbf{R}^n \to \mathbf{R}$ *and a mapping* $\omega : \mathbf{R}^n \to \mathbf{R}$ *such that*

$$f(\mathbf{x}+\mathbf{h})-f(\mathbf{x}) = A(\mathbf{h})+\omega(\mathbf{h}), \quad \mathbf{h}\in\mathbf{R}^n, \quad \mathbf{x}+\mathbf{h}\in\Omega,$$

where ω *has the property* $\omega(\mathbf{h})/\|\mathbf{h}\|\to 0$ *as* $\|\mathbf{h}\|\to 0$.
Then A is called the first derivative of f at x and it is denoted by $f'(x)$.
In this case

$$f'(\mathbf{x})(\mathbf{h}) = A(\mathbf{h}) = \sum_{i=1}^{n}\alpha^i h^i, \ h\in\mathbf{R}^n,$$

where $\alpha^i = \partial f(\mathbf{x})/\partial x^i, \ i=1,...,n.$
Let f be differentiable at any point of Ω. Fix $\mathbf{h}\in\mathbf{R}^n$. With

$$\mathbf{x}\mapsto f'(\mathbf{x})(\mathbf{h}), \quad \mathbf{x}\in\Omega,$$

a function from Ω into $\mathbf{R}$ is defined.
Definition 2: *It is said that* f *is two times differentiable at* $\mathbf{x}_0\in\Omega$ *if for every* $\mathbf{h}\in\mathbf{R}^n$ *the function* $f'(\cdot)(\mathbf{h})$ *is differentiable at* x_0.
In this case for every $\mathbf{h}\in\mathbf{R}^n$ *there exists a linear mapping* $A(\mathbf{h}): \mathbf{R}^n\to\mathbf{R}$ *and a mapping* $\overline{\omega}_{\mathbf{h}} : \mathbf{R}^n\to\mathbf{R}$ *such that*

$$f'(\mathbf{x}_0+\mathbf{k})(\mathbf{h})-f'(x_0)(\mathbf{h}) = A(\mathbf{h})(\mathbf{k})+\overline{\omega}_{\mathbf{h}}(\mathbf{k}), \ \mathbf{k}\in\mathbf{R}^n, \ \mathbf{x}_0+\mathbf{k}\in\Omega, \quad (1)$$

and $\overline{\omega}_{\mathbf{h}}(\mathbf{k})/\|\mathbf{k}\|\to 0$ *as* $\|\mathbf{k}\|\to 0$.
We use the notation $(\alpha^1(\mathbf{h}),...,\alpha^n(\mathbf{h}))$ for the vector which determines the linear mapping $A(\mathbf{h})$ in (1).
The space of functions $\omega : \mathbf{R}^n\to\mathbf{R}$, is denoted by $\mathcal{F}$.
Proposition 1: *The mappings*

$$\mathbf{R}^n\to\mathbf{R}^n, \ \mathbf{h}\mapsto A(\mathbf{h}) = (\alpha^1(\mathbf{h}),...,\alpha^n(\mathbf{h})), \quad \mathbf{h}\in\mathbf{R}^n, \quad (2)$$

$$\mathbf{R}^n\to\mathbf{F}, \ \mathbf{h}\mapsto\overline{\omega}_{\mathbf{h}}, \ \mathbf{h}\in\mathbf{R}^n, \quad (3)$$

defined via (1) are linear.
Proof. First, we prove the linearity of (2). Let $\alpha\in\mathbf{R}$, $\mathbf{k}\in\mathbf{R}^n$, $\mathbf{x}+\mathbf{k}\in\Omega$.. From (1) it follows

$$f'(\mathbf{x}_{0(4)}+\mathbf{k})(\alpha\mathbf{h})-f'(\mathbf{x}_0)(\alpha\mathbf{h}) = A(\alpha\mathbf{h})(\mathbf{k})+\overline{\omega}_{\alpha\mathbf{h}}(\mathbf{k}), \quad (4)$$

$$\alpha(f'(\mathbf{x}_0+\mathbf{k})(\mathbf{h})-f'(\mathbf{x}_0)(\mathbf{h}) = \alpha(A(\mathbf{h})(\mathbf{k}))+\overline{\omega}_{\mathbf{h}}(\mathbf{k})), \quad (5)$$

which imply $A(\alpha\mathbf{h})(\mathbf{k}) = \alpha\overline{\omega}_{\alpha\mathbf{h}}(\mathbf{k})$. Since $\alpha\overline{\omega}_{\mathbf{h}}(\mathbf{k})/\|\mathbf{k}\|\to 0$ and

$\overline{\omega}_{\alpha\mathbf{h}}(\mathbf{k})\ \|\mathbf{k}\| \to 0$ as $\|\mathbf{k}\| \to 0$, by putting $\mathbf{k} = k^j\mathbf{e}_j$, $j = 1,...,n$, it follows

$$\frac{A(\alpha(\mathbf{h})(k^j\mathbf{e}_j) - \alpha A(\mathbf{h})(k^j\mathbf{e}_j)}{k^j} \to 0 \quad \text{as} \quad k^j \to 0.$$

This, implies $\alpha^j(\alpha\mathbf{h}) = \alpha\alpha^j(\mathbf{h})$, $j = 1,...,n$, which gives $A(\alpha\mathbf{h}) = \alpha A(\mathbf{h})$. Let $\mathbf{h}_1, \mathbf{h}_2 \in \mathbf{R}^n$ be fixed. From (1) it follows

$$f'(\mathbf{x}_0 + \mathbf{k})(\mathbf{h}_1 + \mathbf{h}_2) - f'(\mathbf{x}_0)(\mathbf{h}_1 + \mathbf{h}_2) = A(\mathbf{h}_1 + \mathbf{h}_2)(\mathbf{k}) + \overline{\omega}_{\mathbf{h}_1+\mathbf{h}_2}(\mathbf{k}) \quad (6)$$

$$\sum_{i=1}^{2}(f'(\mathbf{x}_0 + \mathbf{k})(\mathbf{h}_i) - f'(\mathbf{x}_0)(\mathbf{h}_i) = A(\mathbf{h}_1)(\mathbf{k}) + \overline{\omega}_{\mathbf{h}_1}(\mathbf{k}) + A(\mathbf{h}_2)(\mathbf{k}) + \overline{\omega}_{\mathbf{h}_2}(\mathbf{k}). \quad (7)$$

Since

$$\overline{\omega}_{\mathbf{h}_1+\mathbf{h}_2}(\mathbf{k})/\|\mathbf{k}\| \to 0,\ \overline{\omega}_{\mathbf{h}_1}(\mathbf{k})/\|\mathbf{k}\| \to 0,$$
$$\text{and}\ \overline{\omega}_{\mathbf{h}_2}(\mathbf{k})/\|\mathbf{k}\| \to 0 \ \text{and}\ \|\mathbf{k}\| \to 0,$$

by putting $\mathbf{k} = k^j\mathbf{e}_j$, $j = 1,...,n$, in (6) and (7) and by making the difference, it follows

$$A(\mathbf{h}_1 + \mathbf{h}_2) = A(\mathbf{h}_1) + A(\mathbf{h}_2).$$

For the proof of linearity of the mapping (3) one have to use again (4), (5), (6), (7) and the linearity of A. This implies that for every $\mathbf{k} \in \mathbf{R}^n$ the mapping $\mathbf{h} \mapsto \overline{\omega}_{\mathbf{h}}(\mathbf{k})$ $(\mathbf{R}^n \to \mathbf{R})$ is linear. Thus, we have

$$\overline{\omega}_{\mathbf{h}}(\mathbf{k}) = \sum_{i=1}^{n}\omega^i(\mathbf{k})h^i, \quad \omega^i(\mathbf{k})/\|\mathbf{k}\| \to 0 \qquad \|\mathbf{k}\| \to 0.$$

(LA2) implies that $A(\mathbf{h})(\mathbf{k}) = \sum_{j=1}^{n}\alpha^j(\mathbf{h})k^j$, $\mathbf{k} \in \mathbf{R}^n$, where

$$(\alpha^1(\mathbf{h}), \alpha^2(\mathbf{h}), \ldots, \alpha^n(\mathbf{h})) = (\sum_{i=1}^{n}\alpha^{1i}h^i, \sum_{i=1}^{n}\alpha^{2i}h^i, \ldots, \sum_{i=1}^{n}\alpha^{ni}h^i) \cdot \mathbf{h} \in \mathbf{R}^n.$$

Thus, for $\mathbf{h}, \mathbf{k} \in \mathbf{R}^n$,

$$f'(\mathbf{x}_0 + \mathbf{k})(\mathbf{h}) - f'(\mathbf{x}_0)(\mathbf{h}) = \sum_{i,j=1}^{n}\alpha^{ji}k^jh^i + \sum_{i=1}^{n}\omega^i(\mathbf{k})h^i.$$

Proposition 2: *Let f be differentiable on Ω and two times differentiable at $\mathbf{x}_0 \in \Omega$.*

Then f *has all the partial derivatives of second order at* $\mathbf{x}_0$ *and*

$$\partial^2 f(\mathbf{x}_0)/\partial x^j \partial x^i = \alpha^{ji}, i = 1,\ldots,n,\ j = 1,\ldots,n.$$

Proof. From

$$f'(\mathbf{x}_0 + \mathbf{k})(\mathbf{h}) - f'(\mathbf{x}_0)(\mathbf{h}) =$$

$$\sum_{i=1}^{n}\left(\frac{\partial f(\mathbf{x}_0 + \mathbf{k})}{\partial x^i}h^i - \frac{\partial f(\mathbf{x}_0)}{\partial x^i}h^i\right) = \sum_{i,j=1}^{n}\alpha^{ji}k^j h^i + \sum_{i=1}^{n}\omega^i(\mathbf{k})h^i,$$

by putting $\mathbf{h} = \mathbf{e}_j$, $\mathbf{k} = k^j\mathbf{e}_j$ and by letting $k^j \to 0$ it follows
$\alpha^{ji} = \partial^2 f(\mathbf{x}_0)/\partial x^j,\ i, j = 1,...,n.$

Definition 3: *If* f *is differentiable on* Ω *and two times differentiable at* $\mathbf{x}_0 \in \Omega$, *then the bilinear mapping from* $\mathbf{R}^n \times \mathbf{R}^n$ *into* $\mathbf{R}$ *of the form*

$$(\mathbf{h},\mathbf{k}) \to A(\mathbf{h})(\mathbf{k}) = \sum_{i,j=1}^{n}\alpha^{ji}k^j h^j = \sum_{i,j=1}^{n}\frac{\partial^2 f(\mathbf{x}_0)}{\partial x^j \partial x^i}k^j h^i, \quad \mathbf{h},\mathbf{k} \in \mathbf{R}^n$$

is called the second derivative of f *at* $\mathbf{x}_0$ *and it is denoted by* $f''(\mathbf{x}_0)(\mathbf{h},\mathbf{k})$.

For every $\mathbf{h},\mathbf{k} \in \mathbf{R}^n$ we have $(f'(\mathbf{x}_0)(\mathbf{h}))(\mathbf{k}) = f''(\mathbf{x}_0)(\mathbf{h},\mathbf{k})$.
Let f be two times differentiable on Ω and let for any $\mathbf{h} \in \mathbf{R}^n$, $\mathbf{k} \in \mathbf{R}^n$ the function $\mathbf{x} \mapsto f''(\mathbf{x})(\mathbf{h},\mathbf{k})$, $\mathbf{x} \in \Omega$, be differentiable at $\mathbf{x}_0 \in \Omega$. This means that there exist linear the mapping $A(\mathbf{h},\mathbf{k})$ represented by the vector $(\alpha^1(\mathbf{h},\mathbf{k}),...,\alpha^n(\mathbf{h},\mathbf{k}))$ and a mapping $\bar{\omega}_{\mathbf{h},\mathbf{k}}$ from $\mathbf{R}^n$ into $\mathbf{R}$ such that

$$f''(\mathbf{x}_0 + \mathbf{r})(\mathbf{h},\mathbf{k}) - f''(\mathbf{x}_0)(\mathbf{h},\mathbf{k}) = A(\mathbf{h},\mathbf{k})(\mathbf{r}) + \tilde{\tilde{\omega}}_{\mathbf{h},\mathbf{k}}(\mathbf{r})$$
$$= \sum_{p=1}^{n}\alpha^p(\mathbf{h},\mathbf{k})r^p + \tilde{\tilde{\omega}}_{\mathbf{h},\mathbf{k}}(\mathbf{r}), \qquad (8)$$

$$\tilde{\tilde{\omega}}_{(\mathbf{h},\mathbf{k})}(\mathbf{r})/\|r\| \to 0, \quad \text{as} \quad \|r\| \to 0. \qquad (9)$$

By using the same ideas as in the proofs of Propositions 1 and 2 one can prove the following one.

Proposition 3: *The mapping*

$$\mathbf{R}^n \times \mathbf{R}^n \to \mathbf{R}^n,\ (\mathbf{h},\mathbf{k}) \mapsto (\alpha^1(\mathbf{h},\mathbf{k}),...,\alpha^n(\mathbf{h},\mathbf{k})), \qquad (11)$$

$$\mathbf{R}^n \times \mathbf{R}^n \to \mathbf{F}, \quad (\mathbf{h},\mathbf{k}) \mapsto \tilde{\tilde{\omega}}_{\mathbf{h},\mathbf{k}}, \tag{12}$$

are bilinear.

Proof. We only give the sketch of the proof that (10) is bilinear. Let $\alpha \in \mathbf{R}$, $\mathbf{r} \in \mathbf{R}^n$, $\mathbf{x}+\mathbf{r} \in \Omega$ and $\mathbf{h}$ and $\mathbf{k}$ be fixed. There holds

$$f''(\mathbf{x}_0 + \mathbf{r})(\alpha\mathbf{h},\mathbf{k}) - f''(\mathbf{x}_0)(\alpha\mathbf{h},\mathbf{k}) = A(\alpha\mathbf{h},\mathbf{k})(\mathbf{r}) + \tilde{\tilde{\omega}}_{\alpha\mathbf{h},\mathbf{k}}(\mathbf{r})$$

$$\alpha(f''(\mathbf{x}_0 + \mathbf{r})(\alpha\mathbf{h},\mathbf{k}) - f''(\mathbf{x}_0)(\mathbf{h},\mathbf{k})) = \alpha(A(\mathbf{h},\mathbf{k})(\mathbf{r}) + \tilde{\tilde{\omega}}_{\mathbf{h},\mathbf{k}}(\mathbf{r})),$$

where

$$\tilde{\tilde{\omega}}_{\alpha\mathbf{h},\mathbf{k}}(\mathbf{r})/\|\mathbf{r}\| \to 0, \ \alpha\tilde{\tilde{\omega}}_{\mathbf{h},\mathbf{k}}(\mathbf{r})/\|\mathbf{r}\| \to 0 \quad \|\mathbf{r}\| \to 0.$$

By using that the second derivative in (14) bilinear and by putting $\mathbf{r} = r^p \mathbf{e}_p$ in (12) and (13) it follows

$$\frac{(\alpha^p(\alpha\mathbf{h},\mathbf{k}) - \alpha\alpha^p(\mathbf{h},\mathbf{k}))(r^p)}{r^p} \to 0, \quad r^p \to 0,$$

which gives $\alpha^p(\alpha\mathbf{h},\mathbf{k}) = \alpha^p(\mathbf{h},\mathbf{k}), \quad p = 1,\ldots,n.$
In a similar way we prove

$$\alpha^p(\mathbf{h_1} + \mathbf{h_2},\mathbf{k}) = \alpha^p(\mathbf{h_1},\mathbf{k}) + \alpha^p(\mathbf{h_2},\mathbf{k}), \quad p = 1,\ldots,n.$$

The proof of the linearity with respect to $\mathbf{k}$ is quite the same. Thus, (10) and (11) imply

$$A(\mathbf{h},\mathbf{k})(\mathbf{r}) = \sum_{i,j,p=1}^{n} \alpha^{pji} r^p k^j h^i, \quad \tilde{\tilde{\omega}}_{\mathbf{h},\mathbf{k}}(\mathbf{r}) = \sum_{i,j=1}^{n} \omega^{ji}(\mathbf{r}) k^j h^i,$$

and (8) has the form

$$f''(\mathbf{x}_0 + \mathbf{r})(\mathbf{h},\mathbf{k}) - f''(\mathbf{x}_0)(\mathbf{h},\mathbf{k}) = \sum_{i,j,p=1}^{n} \alpha^{pji} r^p k^j h^i + \sum_{i,j=1}^{n} \omega^{ji}(\mathbf{r}) k^j h^i.$$

Proposition 4: If f is two times differentiable on Ω and there times differentiable at $\mathbf{x}_0 \in \Omega$, then f has all the derivatives of third order at $\mathbf{x}_0$ and

$$\partial^3 f(\mathbf{x}_0)/\partial x^p \partial x^j \partial x^i = \alpha^{pji}, \quad p, j, i = 1,\ldots,n.$$

Proposition 5: Let f satisfy all the conditions of Proposition 4. The third derivative at $\mathbf{x}_0$ is the three-linear mapping

$$\mathbf{h},\mathbf{k},\mathbf{r}) \mapsto f'''(\mathbf{x}_0)(\mathbf{h},\mathbf{k},\mathbf{r}) = \sum_{i,j,p=1}^{n} \frac{\partial^3 f(\mathbf{x}_0)}{\partial x^p \partial x^j \partial x^i} r^p k^j h^i.$$

In an appropriate way we define the $s-$ th derivative of f as the $s-$ linear mapping

$$(\mathbf{h}_1,\ldots,\mathbf{h}_s) \mapsto f^{(s)}(\mathbf{x}_0)(\mathbf{h}_1,\ldots,\mathbf{h}_s) = \sum_{i_1=1,\ldots,i_s=1}^{n} \alpha^{i_s,\ldots,i_1} h_s^{i_s}\ldots h_1^{i_1}$$

$$= \sum_{i_1=1,\ldots,i_s=1}^{n} \frac{\partial^s f(\mathbf{x}_0)}{\partial x^{i_s}\ldots\partial x^{i_1}} h_s^{i_s}\ldots h_1^{i_1}.$$

Let $F : \Omega \to \mathbf{R}^n,$ be a function defined by

$$\Omega \ni \mathbf{x} = (x^1,\ldots,x^n) \mapsto F(\mathbf{x}) = (f^1(\mathbf{x}),\ldots,f^m(\mathbf{x}))$$
$$= (\mathbf{y}^1,\ldots,\mathbf{y}^m) \in \mathbf{R}^m$$

Definition 4: *The function f is differentiable at $\mathbf{x} \in \Omega$ if there exist a linear mapping $A : \mathbf{R}^n \to \mathbf{R}^m$ and a mapping $\omega : \mathbf{R}^n \to \mathbf{R}^m$ such that*

$$F(\mathbf{x}+\mathbf{h}) - F(\mathbf{x}) = A\mathbf{h} + \omega(\mathbf{h}), \quad \mathbf{h} \in \mathbf{R}^n, \quad \mathbf{x}+\mathbf{h} \in \Omega,$$

where ω has the property

$$\|\omega(\mathbf{h})\|_{\mathbf{R}^m} / \|\mathbf{h}\|_{\mathbf{R}^n} \to 0 \qquad \|\mathbf{h}\|_{\mathbf{R}^n} \to 0.$$

One can easily see that F is differentiable at $\mathbf{x}$ if and only if $f^1,\ldots,f^m$ are differentiable at $\mathbf{x}$ and

$$A = F'(\mathbf{x}) = (f^{1\prime}(\mathbf{x}),\ldots,f^{m\prime}(\mathbf{x})),$$

$$F'(\mathbf{x})(\mathbf{h}) = (f^{1\prime}(\mathbf{x})(\mathbf{h}),\ldots,f^{m\prime}(\mathbf{x})(\mathbf{h})), \quad \mathbf{h} \in \mathbf{R}^n.$$

H is more convergent to write A in the matrix form

$$F'() = \begin{pmatrix} f^{1'}(\mathbf{x}) \\ f^{2'}(\mathbf{x}) \\ f^{m'}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \frac{\partial f^1}{\partial x^1}(\mathbf{x}) & \frac{\partial f^1}{\partial x^n}(\mathbf{x}) \\ \frac{\partial f^m}{\partial x^1}(\mathbf{x}) & \frac{\partial f^m}{\partial x^n}(\mathbf{x}) \end{pmatrix}$$

If F is differentiable on Ω, then for fixed $\mathbf{h}\in\mathbf{R}^n$, $\mathbf{x}\mapsto F'(\mathbf{x})(\mathbf{h})$, $\mathbf{x}\in\Omega$, is a mapping from Ω into $\mathbf{R}^m$. It is said that F is two times differentiable at $\mathbf{x}_0\in\Omega$ if for every $\mathbf{h}\in\mathbf{R}^n$ the function $F'(\cdot)(\mathbf{h})$ is differentiable at $\mathbf{x}_0$. This is equivalent to the fact that everyone functions $f'(\cdot)(\mathbf{h}),\ldots, f'(\cdot)\mathbf{h}$ are differentiable at $\mathbf{x}_0$. In this case we have

$$F'(\mathbf{x}_0+\mathbf{k})(\mathbf{h})-F'(\mathbf{x}_0)(\mathbf{h})=A(\mathbf{h})(\mathbf{k})+\overline{W}_{\mathbf{h}}(\mathbf{k}),\quad \mathbf{k}\in\mathbf{R}^n,\quad \mathbf{k}\in\Omega$$

where $A(\mathbf{h})$ is a linear mapping $\mathbf{R}^n\to\mathbf{R}^m$ and $\overline{\omega}_{\mathbf{h}})$ is a mapping $\mathbf{R}^n\to\mathbf{R}^m$ such that

$$\|\overline{\omega}_{\mathbf{h}}(\mathbf{k})\|_{\mathbf{R}^n}/\|\mathbf{k}\|_{\mathbf{R}^n}\to 0,\ \|\mathbf{k}\|_{\mathbf{R}^n}\to 0.$$

As in the case $m=1$, one can prove that the mappings

$$\mathbf{h}\to A(\mathbf{h}),\quad \mathbf{h}\in\mathbf{R}^n,\qquad\qquad \mathbf{h}\to\overline{\omega}_{\mathbf{h}},\quad \mathbf{h}\in\mathbf{R}^n,$$

are linear.
Without repeating all the arguments we have

$$(F'(\mathbf{x}_0)(\mathbf{h}))'=((f^{1\,\prime}(\mathbf{x}_0)(\mathbf{h}))',\ldots,(f^{m\,\prime}(\mathbf{x}_0)(\mathbf{h}))')$$

$$(F'(\mathbf{x}_0)(\mathbf{h}))'=((f^{1\,\prime}(\mathbf{x}_0)(\mathbf{h}))'(\mathbf{k}),\ldots,(f^{m\,\prime}(\mathbf{x}_0)(\mathbf{h}))'(\mathbf{k})),\quad \mathbf{k}\in\mathbf{R}^n.$$

Since for $\mathbf{h},\mathbf{k}\in\mathbf{R}^n$ and $p=1,\ldots,m.$

$$(f^{p\,\prime}(\mathbf{x}_0)(\mathbf{h}))'(\mathbf{k})=\sum_{j,i=1}^{n}\frac{\partial^2 f^p(\mathbf{x}_0)}{\partial x^j\partial x^i}k^j h^i,$$

it follows

$$(F'(\mathbf{x}_0)(\mathbf{h}))'(\mathbf{k})=(f^{1\,\prime\prime}(\mathbf{x}_0)(\mathbf{h},\mathbf{k})),\ldots,f^{m\,\prime\prime}(\mathbf{x}_0)(\mathbf{h},\mathbf{k})).$$

Thus the second derivative of F at $\mathbf{x}\in\Omega$ is

$$F''(\mathbf{x})=(f^{1\,\prime}(\mathbf{x}),\ldots,f^{m\,\prime\prime}(\mathbf{x}))$$

and

$$F''(\mathbf{x})(\mathbf{h},\mathbf{k})=(f^{1\,\prime\prime}(\mathbf{x})(\mathbf{h},\mathbf{k}),\ldots,f^{m\,\prime\prime}(\mathbf{x})(\mathbf{h},\mathbf{k})),\ \mathbf{h},\mathbf{k}\in\mathbf{R}^n.\quad F''(\mathbf{x}),\ \mathbf{x}\in\Omega,$$

is a bilinear mapping from $\mathbf{R}^n\times\mathbf{R}^n$ into $\mathbf{R}^m$.
The higher derivatives of F are defined in an analogous way.

On the visualization of the derivative and the differential of functions

Stevan Pilipović, Djurdjica Takači, Arpad Takači

Introduction

In this paper we present a visualization of the partial derivatives and the differential of a function of two variables, but we first give a visualization of the derivative and the differential of a function of one variable, by using the programme package *GeoGebra* and *Scientific Workplace.*

Visualization of difference quotient

Let us consider the function $f : D \to R$ and its difference quotient

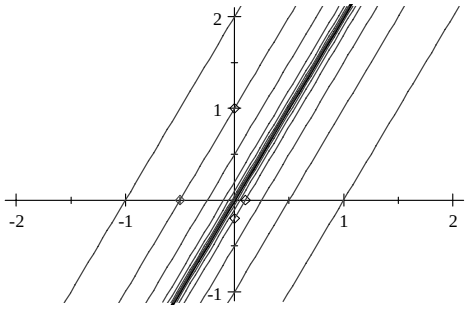
$$k(h, x) = \frac{f(x+h) - f(x)}{h}.$$

The function k is the function of two variables. Further on, we shall change the parameter h, and follow the coresponding function k, by using programme packages *Scientific WorkPlace* 3 i 5.5 and *GeoGebra.*

Example 1. The difference quotient for the function $f(x) = x^2$ has the form

$$k(x, h) = \frac{(x+h)^2 - x^2}{h}.$$

The graphs of the functio k, for diggerent values of h, are shown on Figure 1.



Slika 1

Example 1. The difference quotient for the function $f(x) = \sin x$ has the form

$$k(x,h) = \frac{\sin(x+h) - \sin x}{h}.$$

The graphs of the functio k, for diggerent values of h, are shown on Figure 2.

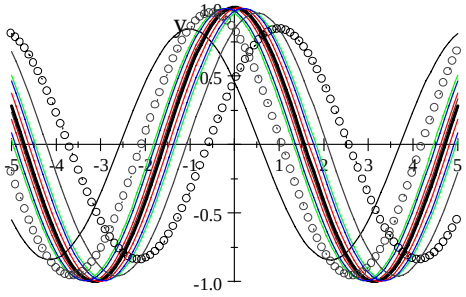


Figure 2.

In the programme packages *Geogebra* we can follow the same visualization by using slidersas on Figure 3.

.

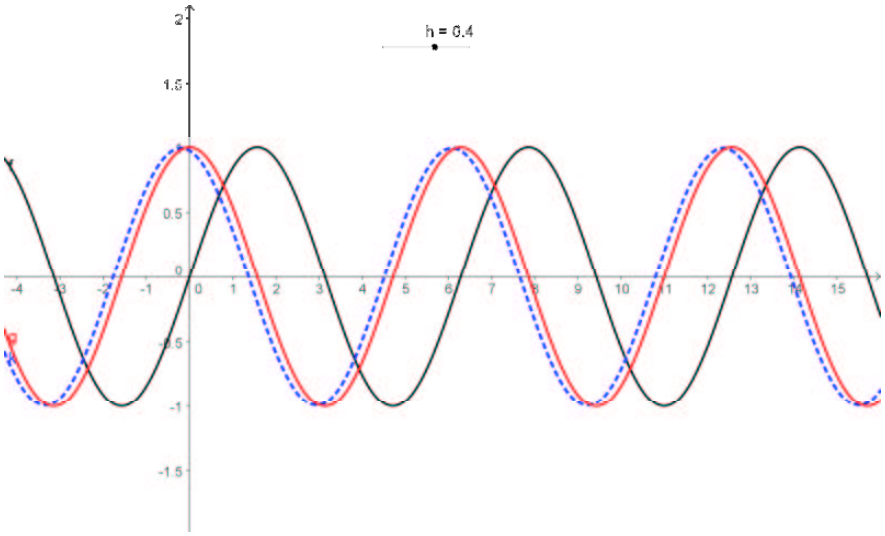


Figure 3.

E:\KURS--APRIL\
zaKNJ\Nastava1.ggb

. or [..\Nastava1.ggb](..\Nastava1.ggb) ili

In this programe packages one can change function and obtain the correspoding graphs.

Fourier Series

Đurđica Takaci

Introduction

Let f be a $2\pi-$ periodic piecewise continuous function on the interval $[-\pi,\pi]$. The trigonometric series

$$A_0+\sum_{n=1}^{\infty}(A_n\cos nx+B_n\sin nx),$$

is called the *Fourier series* of function f, if the coefficients $A_n,\ n=0,1,\ldots,\ B_n,\ n=1,2,\ldots,$ are given by

$$A_0=\tfrac{1}{2\pi}\int_{-\pi}^{\pi}f(x)dx,$$

$$A_n=\tfrac{1}{\pi}\int_{-\pi}^{\pi}f(x)\cos(nx)dx, n=1,2,\ldots,$$

$$B_n=\tfrac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin(nx)dx, n=1,2,\ldots.$$

The coefficients $A_n,\ n=0,1,\ldots,\ B_n,\ n=1,2,\ldots,$ are called Fourier coefficients.

The Fourier series of a periodic piecewise continuous function f on an interval $[-\pi,\pi]$, with piecewise continuous first derivative f', converges at any point $x\in[\pi,\pi]$ (converges pointwise for all values x). Then we have

$$\frac{f(x^+)+f(x^-)}{2}=A_0+\sum_{n=1}^{\infty}(A_n\cos nx+B_n\sin nx),$$

where $A_n,\ n=0,1,2,...,\ B_n,\ n=1,2,...,$ are given by (4coef).

Suppose the series $\sum_{n=1}^{\infty}(|A_n|+|B_n|)$ converges, where A_n and B_n are the Fourier coefficients. Then the Fourier series converges uniformly on every finite interval .

The Fourier series of a continuous $2\pi-$ periodic function f, with piecewise

continuous first derivative f', converges uniformly on every finite interval.
The change of variables $x = \pi t / \ell$, and $f(x) = f(\pi t / \ell) = g(t)$, imply the Fourier series for the function g as

$$A_0 + \sum_{n=1}^{\infty}(A_n \cos\frac{n\pi t}{\ell} + B_n \sin\frac{n\pi t}{\ell}),$$

where the coefficients have the forms

$$A_0 = \frac{1}{2\ell}\int_{-\ell}^{\ell} g(x)dx,$$

$$A_n = \frac{1}{\ell}\int_{-\ell}^{\ell} g(x)\cos\frac{n\pi x}{\ell}dx, n = 1,2,\ldots,$$

$$B_n = \frac{1}{\ell}\int_{-\ell}^{\ell} g(x)\sin\frac{n\pi x}{\ell}dx, n = 1,2,\ldots.$$

On the visualization of the coeficients of Fourier series

In this section we use the programme packages:

- *Scientific Workplace version 3, and 5,*
- *Geogebra*

in order to determine the coefficients of the coresponding Fourier series and its partial sums for different values of n.

The Fourier series for the function $f(x) = x, \quad x \in [-\pi,\pi]$, $f(x + 2\pi) = f(x)$

The function $f(x) = x, \quad x \in [-\pi,\pi]$, $f(x + 2\pi) = f(x)$ is *piecewise continuous* on any interval $[a,b]$, because this interval can be divided this interval on subintervals $[2z\pi, 2(z+1)\pi)], \quad z \in N,$

- inside each of which f is continuous,
- the left-hand and right-hand limits exists at each point on the subintervals including their end points.

The left-hand and right-hand limits are defined, respectively, by

$$f(x^-) = \lim_{t\to x^-} t = x = f(x^+) = \lim_{t\to x^+} t, \quad x \in [2z\pi, 2(z+1)\pi)], \; z \in N,$$

and the function f is continuous at the point x since $f(x^-) = f(x^+)$
In fact the piecewise continuous function f can be written as:

$$f(x) = \begin{cases} x + 4\pi & \text{if} \quad -5\pi \le x < -3\pi \\ x + 2\pi & \text{if} \quad -3\pi \le x < -\pi \\ x & \text{if} \quad -\pi \le x < \pi \\ x - 2\pi & \text{if} \quad \pi \le x < 3\pi \\ x - 4\pi & \text{if} \quad 3\pi \le x < 5\pi \\ x - 6\pi & \text{if} \quad 5\pi \le x < 7\pi \end{cases}$$

with its graph given on Figure 1.

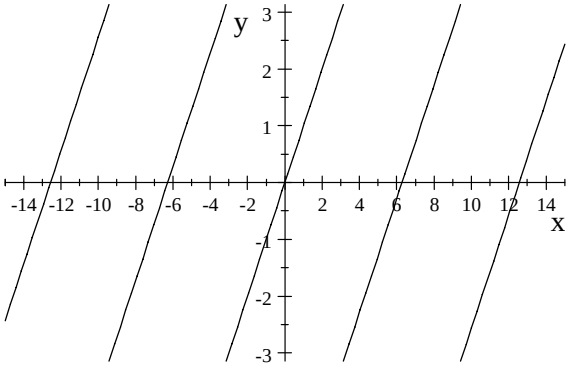


Figure 1.

The function f is odd and the coefficients $A_n = 0, \quad n = 1, 2, \ldots$. By usual clasical calculation, using the partial integration we get:

$$\begin{aligned} B_n &= \frac{2}{\ell} \int_0^{\ell} x \sin \frac{n\pi x}{\ell} dx = \frac{2}{\ell} \left(-x \frac{\ell}{n\pi} \cos \frac{n\pi x}{\ell} \Big|_0^{\ell} + \frac{\ell}{n\pi} \int_0^{\ell} \cos \frac{n\pi x}{\ell} dx \right) \\ &= (-1)^{n+1} \frac{2\ell}{n\pi}. \end{aligned}$$

By using the programme package Sci3 we get

$$\frac{1}{\pi} \int_{-\pi}^{\pi} x \sin nx dx = -\frac{2}{\pi} \frac{-\sin \pi n + \pi n \cos \pi n}{n^2}, \qquad \frac{1}{\pi} \int_{-\pi}^{\pi} x \sin nx dx = -2 \frac{\cos \pi n}{n} = \frac{2(-1)^{n+1}}{n}.$$

The Fourier series for the function $f(x)=x, \quad x\in[-\pi,\pi]$, $f(x+2\pi)=f(x)$ is

$$f(x) = 2\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\sin nx.$$

Let us consider the sequences $(A_n)_{n\in N}$ of partial sums given by series in the last relations:

$$A(n) = 2\sum_{k=1}^{n}\frac{(-1)^{k+1}}{k}\sin kx$$

The partial sum $A(12)$ obtain in SI3 is given on Figure

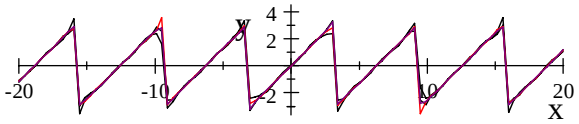


Figure 2.

Generally, the Fourier series for the function $f(x)=x, \quad x\in[-\ell,\ell]$, $f(x+2\ell)=f(x)$ can be written as

$$f(x)=\frac{2\ell}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\sin\frac{n\pi x}{\ell}, \qquad x\in(-\ell,\ell).$$

In programme package Geiger the sixth partial sum of Furie series is ilustated on interval $(-2\pi,6\pi)$.

f(x),PI-x-.ggb (E:\KURS--APRIL\ f(x),PI-x-.ggb) and the exported figure is:

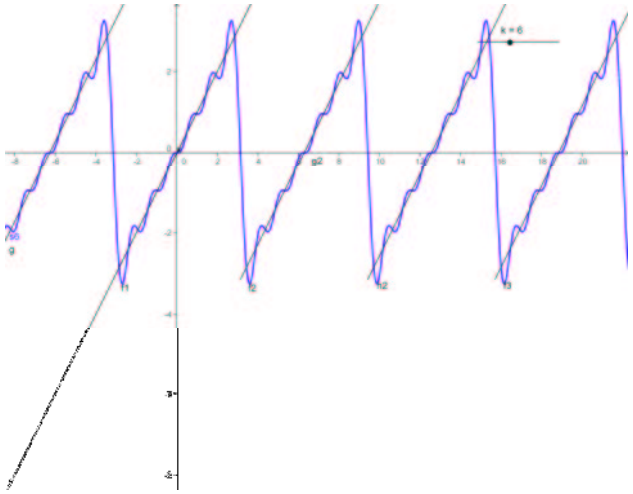


Figure 3.

The Fourier series for the function f, for $\ell = 1$ can be written as

$$f(x) = \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin(n\pi x).$$

The Fourier series for the functions $f(x) = x^2, \quad x \in [-\ell, \ell],$ $f(x + 2\ell) = f(x).$

In fact the piecewise continuous function f can be written as:

$$f(x) = \begin{cases} (x + 6\pi)^2 & \text{if} \quad -7\pi \le x < -5\pi \\ (x + 4\pi)^2 & \text{if} \quad -5\pi \le x < -3\pi \\ (x + 2\pi)^2 & \text{if} \quad -3\pi \le x < -\pi \\ x^2 & \text{if} \quad -\pi \le x < \pi \\ (x - 2\pi)^2 & \text{if} \quad \pi \le x < 3\pi \\ (x - 4\pi)^2 & \text{if} \quad 3\pi \le x < 5\pi \\ (x - 6\pi)^2 & \text{if} \quad 5\pi \le x < 7\pi \end{cases}$$

and its graph is given on Figure 4

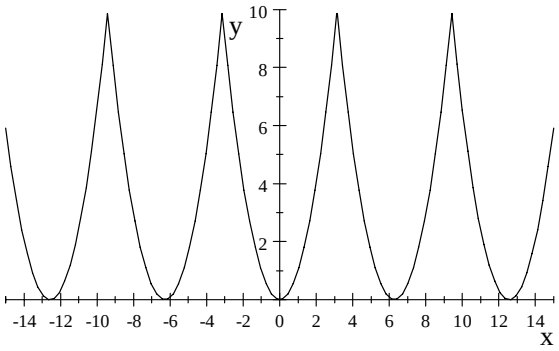


Figure 4.

The function f is even and therefore it holds $B_n = 0,\ n = 1,2,\ldots.$ The coefficients A_n, $n = 0,1,\ldots,$ can be written as

$$A_0 = \frac{1}{2\ell}\int_0^{\ell} x^2 dx = \frac{2\ell^2}{3},$$

$$A_n = \frac{2}{\ell}\int_0^{\ell} x^2 \cos\frac{n\pi x}{\ell}\,dx = \frac{(-1)^n 4\ell^2}{n^2\pi^2}, \quad n \in \mathbf{N}.$$

and the Furier series is of the form:

$$f(x) = \frac{\ell^2}{3} + \frac{4\ell^2}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\cos\frac{n\pi x}{\ell}, \quad x \in \mathbf{R}.$$

If we put $\ell = \pi$ then we have

$$x^2 = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\cos nx, \quad x \in [-\pi,\pi].$$

The sequences $(C_n)_{n\in N}$ of partial sums given by last series are:

$$C(n) = \frac{4}{3} + \frac{16}{\pi^2}\sum_{k=1}^{n}\frac{(-1)^n}{k^2}\cos\frac{k\pi x}{2}, \text{ for } \ell = 2$$

In programme package *GeoGebra* the second partial sum of Furije series is ilustated on interval $(-2\pi, 6\pi)$.

f(x),PI-x^2-.ggb (E:\KURS--APRIL\ f(x),PI-x^2-.ggb) and the followin picture is exported.

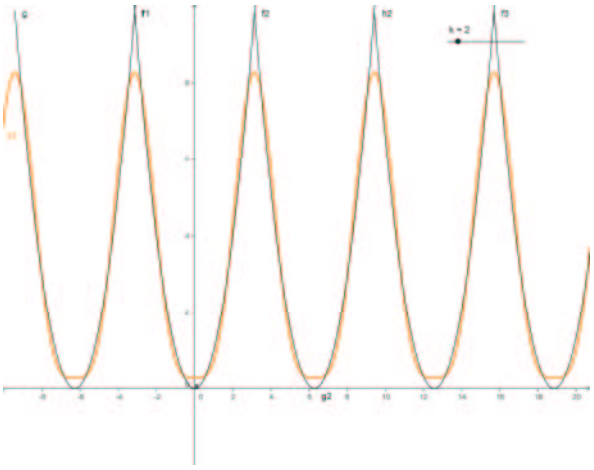


Figure 5.

The Fourier series for the functions $f(x)=|x| \quad x\in[-\pi,\pi]$, $f(x+2\pi)=f(x)$

The function f is even and therefore it holds $B_n=0,\ n=1,2,\ldots.$ The coefficients A_n, $n=0,1,\ldots,$ can be written as

$$A_0=\frac{2}{\pi}\int_0^{\pi}|x|\,dx=\frac{2}{\pi}\int_0^{\pi}x\,dx=\pi,$$

$$A_n=\frac{2}{\pi}\int_0^{\pi}|x|\cos nx\,dx=\frac{2}{\pi}\int_0^{\pi}x\cos nx\,dx=\frac{2}{\pi n^2}(\cos n\pi-1),\quad n\in N.$$

From $\cos n\pi=(-1)^n,\quad n\in\mathbf{N},$ it follows

$$A_n=\frac{2}{\pi n^2}(-1)^n-1=\begin{cases}-\frac{4}{n^2\pi}, & n=1,3,5,\ldots\\ 0, & n=2,4,6,\ldots.\end{cases}$$

In programme package *Geogebra* the secod partial sum of Furije series is shown on the interval $(-2\pi,6\pi)$.

E:\KURS--APRIL\
f(x),PI-abs(x).ggb

f(x),PI-abs(x).ggb ili ( ) and the fllowing pictures is exported.

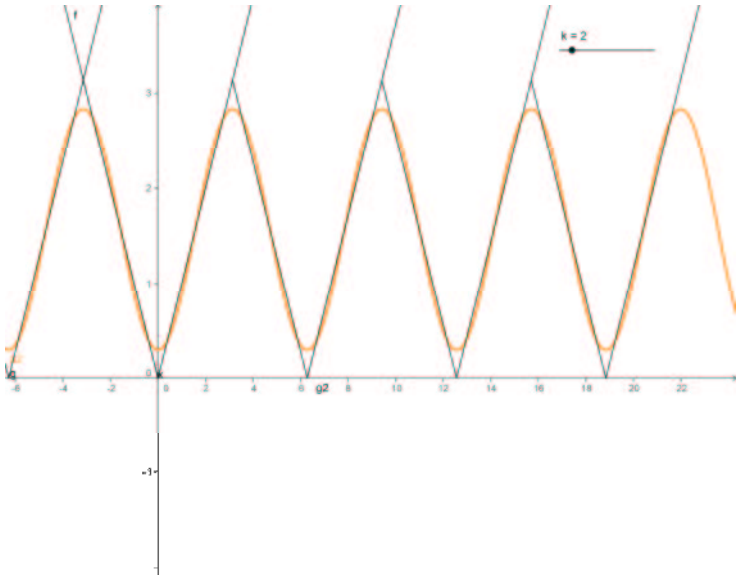


Figure 6.

Remark: In programme pakckageu *Geogebra* the functions can be chaged and by using sliders the corresponding partial sums can be obtained.

The Fourier Method of Separation of Variables

Đurdjca Takači

EIGENVALUES AND EIGENFUNCTIONS

In this part we shal first explain eigenvalues and eigenfunctions of the Sturm-Liouville problem.

We shall show that functions $\cos\frac{n\pi x}{\ell}$, $\sin\frac{n\pi x}{\ell}$ are the eigenfunctions of the Sturm-Liouville problem

$$f''+\lambda f=0, \qquad f(-\ell)=f(\ell), \qquad f'(-\ell)=f'(\ell). \tag{1}$$

By using SWP or classixaly we can show that exact solution of differential equation $f''+\lambda f=0$, has the form:

$$f(x)=C_1\cos x\sqrt{\lambda}+C_2\sin x\sqrt{\lambda}.$$

The coditions $f(-\ell)=f(\ell), \quad f'(-\ell)=f'(\ell),$ imply

$$\begin{gathered}
C_1\cos\ell\sqrt{\lambda}-C_2\sin\ell\sqrt{\lambda}=C_1\cos\ell\sqrt{\lambda}+C_2\sin\ell\sqrt{\lambda}\\
C_1\sqrt{\lambda}\sin\ell\sqrt{\lambda}+\sqrt{\lambda}C_2\cos\ell\sqrt{\lambda}=-C_1\sqrt{\lambda}\sin\ell\sqrt{\lambda}+\sqrt{\lambda}C_2\cos\ell\sqrt{\lambda}\\
C_2\sin\sqrt{\lambda}\ell=0\\
\sqrt{\lambda}C_1\sin\sqrt{\lambda}\ell=0
\end{gathered}.$$

If $C_1=C_2,$ we obtain the trivial solution and therefore we have to determine such values of parameter λ which allow nontrivial solutions of the given problem and then to find the solution. These special values λ are called *eigenvalues* and the solutions of the considered problem are called *eigenfunctions*, and they are obtained as the solutions of the equation $\sin\sqrt{\lambda}\ell=0.$

The Fourier Method of Separation of Variables

We shall explain the method of separation variables for obtaining the solution of partial differential equations.

The linear partial differential equation

$$A(x)\frac{\partial^2 u}{\partial x^2}+B(x)\frac{\partial u}{\partial x}+C(x)u-D(y)\frac{\partial^2 u}{\partial y^2}-E(y)\frac{\partial u}{\partial y}-H(y)u=0,$$

where $0 < x < \ell,\ y > 0,$ and either $D(y) > 0,$ or $D(y) = 0,\ E(y) > 0,$ for hyperbolic and parabolic equations.
We assume that the solutions of equation can be written in the form

$$u(x, y) = X(x) \cdot Y(y),$$

meaning that each factor depends on only one variable and therefore we have:

$$\frac{\partial^2 u}{\partial x^2} = X''(x)Y(y), \qquad \frac{\partial^2 u}{\partial y^2} = X(x)Y''(y),$$

$$\frac{\partial u}{\partial x} = X'(x)Y(y), \qquad \frac{\partial u}{\partial y} = X(x)Y'(y).$$

Substituting these expressions into (4sv1) we obtain

$$\tfrac{1}{X(x)}\left(A(x)X''(x) + B(x)X'(x) + C(x)X(x)\right)$$

$$= \tfrac{1}{Y(y)}\left(D(y)Y''(y) + E(y)Y'(y) + H(y)Y(y)\right).$$

In previous equation the left-hand side contains only functions depending on x and the right-hand side contains only functions depending on $y,$ meaning that left-hand side do not depend on y and the right-hand side do not depend on $x.$ This can happen only if both sides are equal to a common constant $-\lambda.$ So we obtain two ordinary differential equations

$$\tfrac{1}{X(x)}\left(A(x)X''(x) + B(x)X'(x) + C(x)X(x)\right) = -\lambda,$$

$$\tfrac{1}{Y(y)}\left(D(y)Y''(y) + E(y)Y'(y) + H(y)Y(y)\right) = -\lambda,$$

where λ is a separation constant.

On the approximate solution of partial differential equation by using computer

Let us consider the following problem:

$$a^2 \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial t^2} = 0, \qquad 0 < x < \ell, \;\; t > 0, \tag{2}$$

where a is a constant, with boundary conditions

$$u(0,t) = 0, \qquad u(\ell,t) = 0, \quad t > 0, \tag{3}$$

and initial conditions

$$u(x,0) = f(x), \qquad \frac{\partial u(x,0)}{\partial t} = g(x), \qquad 0 < x < \ell. \tag{4}$$

The previous problem characterizes free oscillations of taut string with fixed ends with zero displacement.
The soluti of the problem will be constructed by using the method of separation of variables. From

$$u(x,t) = X(x) \cdot T(t),$$

and

$$\frac{X''(x)}{X(x)} = -\lambda, \qquad \frac{T''(t)}{a^2 T(t)} = -\lambda,$$

we get two differential equations:

$$X''(x) + \lambda X(x) = 0, \qquad T''(t) + \lambda a^2 T(t) = 0.$$

Using the boundary conditions (3) it follows

$$X(0)T(t) = 0, \quad X(\ell)T(t) = 0,$$
$$X(0) = 0, \quad X(\ell) = 0.$$

The problem

$$X''(x) + \lambda X(x) = 0, \qquad X(0)T(t) = 0, \;\; X(\ell)T(t) = 0$$

is *Sturm-Liouville problem* and we have to determine *eigenvalues* and *eigenfunctions.*
If $\lambda = -k^2 < 0,$ then the separated solution is

$$X(x) = C_1 e^{kx} + C_2 e^{-kx}.$$

Using boundary conditions (3) we have

$$X(0) = C_1 + C_2 = 0, \quad C_1 e^{k\ell} + C_2 e^{-k\ell} = 0,$$

wherefrom it follows $C_1 = C_2 = 0.$
Clearly this gives us $X(x) = 0$, for all $x \in (-\ell, \ell),$ hence $u(x,t) = 0,$ fo $x \in (-\ell, \ell),\ t > 0.$ Then the given problem has no solution if at least one of the functions f and g are nonzero.
If $\lambda = 0$, then from $X(x) = C_1 + C_2 x,$ and from (3) we obtain the same conclusion.
If $\lambda = k^2 > 0,$ then $X(x) = C_1 \cos kx + C_2 \sin kx,$ and required boundary conditions (3) lead to

$$X(0) = C_1 = 0, \qquad X(\ell) = C_2 \sin k\ell = 0,$$

wherefrom, it follows that $C_1 = 0.$ If $C_2 = 0,$ we obtain the trivial solution of X again. Let us take $C_2 \neq 0.$ Then the second equation is equal zero if $\sin k\ell = 0,$ which is true for $k = \frac{n\pi}{\ell},\ n = 1, 2, \ldots.$.
So the eigenvalues of the considered problem are

$$\lambda = \lambda_n = \frac{n^2 \pi^2}{\ell^2}, \qquad n \in \mathbf{N}, \tag{5}$$

and the corresponding eigenfunctions have the forms

$$X_n(x) = \sin \frac{n\pi x}{\ell}, \qquad n \in \mathbf{N}, \tag{6}$$

where we took $C_2 = 1.$
The solution of the ordinary differential equation $T''(t) + \lambda a^2 T(t) = 0$, for λ given by (5), has the form

$$T(t) = A_n \cos \frac{na\pi t}{\ell} + B_n \sin \frac{na\pi t}{\ell}, \qquad n \in \mathbf{N}, \tag{7}$$

where A_n and B_n are arbitrary constants.
Multiplying (6) and (7) we obtain the solution of the considered problem

$$u_n(x,y) = X(x)\cdot T(t) = \left(A_n \cos\frac{na\pi t}{\ell} + B_n \sin\frac{na\pi t}{\ell} \right) \sin\frac{n\pi x}{\ell}, \qquad n \in \mathbf{N}.$$

The solutions of the problem (2), (3), can be considered in the form of as an infinite series as (superposition principle)

$$u(x,t) = \sum_{n=1}^{\infty} \left(A_n \cos\frac{na\pi t}{\ell} + B_n \sin\frac{na\pi t}{\ell} \right) \sin\frac{n\pi x}{\ell}. \tag{8}$$

The solution u, expressed in the form (8) has to satisfy the initial conditions, and therefore for $t = 0$ we obtain two Fourier series

$$u(x,0) = f(x) = \sum_{n=1}^{\infty} A_n \sin\tfrac{n\pi x}{\ell};$$

$$\tfrac{\partial u(x,0)}{\partial t} = g(x) = \sum_{n=1}^{\infty} \tfrac{na\pi}{\ell} B_n \sin\tfrac{n\pi x}{\ell}.$$

The coefficients can be determined from

$$A_n = \frac{2}{\ell}\int_0^{\ell} f(x)\sin\frac{n\pi x}{\ell}dx, \quad n = 1,2,\ldots.$$

$$\frac{na\pi}{\ell}B_n = \frac{2}{\ell}\int_0^{\ell} g(x)\sin\frac{n\pi x}{\ell}dx, \quad n = 1,2,\ldots.$$

Let us take $\ell = \pi, \quad f(x) = 0, \quad g(x) = x, \quad a = 1,$ then the coefficients can be evaluated as:

$$A_n = \tfrac{2}{\pi}\int_0^{\pi} 0\sin nx dx = 0,$$

$$B_n = \tfrac{2}{na\pi}\int_0^{\pi} x\sin nx dx = \tfrac{2(-1)^{n+1}}{n^2}.$$

Then the solution of the considered problem (2), (3), (4) has the form:

$$u(x,t) = \sum_{n=1}^{\infty} \frac{2(-1)^{n+1}}{n^2}\sin nt \sin nx. \tag{9}$$

Let us denote by

$$S(n)=\sum_{k=1}^{n}\frac{2(-1)^{k+1}}{k^2}\sin kx\sin kx$$

the " $n-th$ " partial sum of the previous series. Then each term of sequeces of partial sums can be treated as the approximate solution of the considered problem.

In Figure 1 it is ilustrated the graph of the first term $S(1)$ as. This graph is obtained by using parogramme package Scientific Workpalce 5.5.

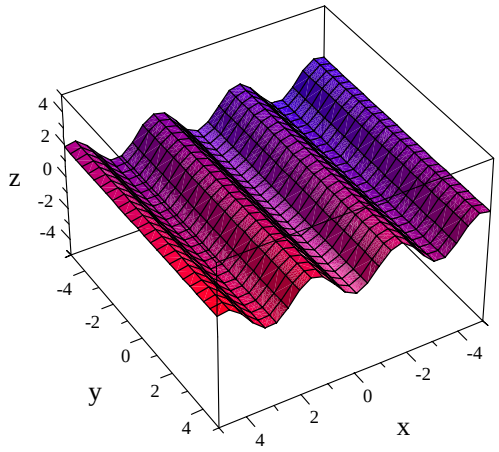


Figure 1.

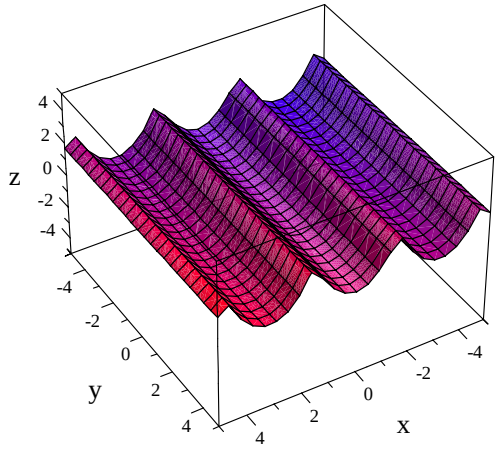


Figure 2.

In Figure the graph of the term $S(5)$.

If we take $\ell=\pi,\quad f(x)=2x,\quad g(x)=x,\quad a=1,$ then we have

$$A_n = \frac{2}{\pi}\int_0^{\pi} 2x\sin nxdx = \frac{4(-1)^{n+1}}{n},$$

$$B_n = \frac{2}{na\pi}\int_0^{\pi} x\sin nxdx = \frac{2(-1)^{n+1}}{n^2}.$$

Then the solution of the considered problem (2), (3), (4) has the form:

$$u(x,t) = \sum_{n=1}^{\infty}\left(\frac{4(-1)^{n+1}}{n}\cos nt + \frac{2(-1)^{n+1}}{n^2}\sin nt\right)\sin nx. \qquad (9)$$

The " $n-th$ " partial sum of the previous series is of the form:

$$A(n) = \sum_{k=1}^{n}\left(\frac{4(-1)^{k+1}}{n}\cos kt + \frac{2(-1)^{k+1}}{n^2}\sin kt\right)\sin kx,$$

and each term of sequeces of partial sums can be treated as the approximate solution of the considered problem.

The first term $A(1) = 2(\sin x)(2\cos t + \sin t)$ and its graph on Figure 3 is also obtained by Scientific Workpalce 5.5.

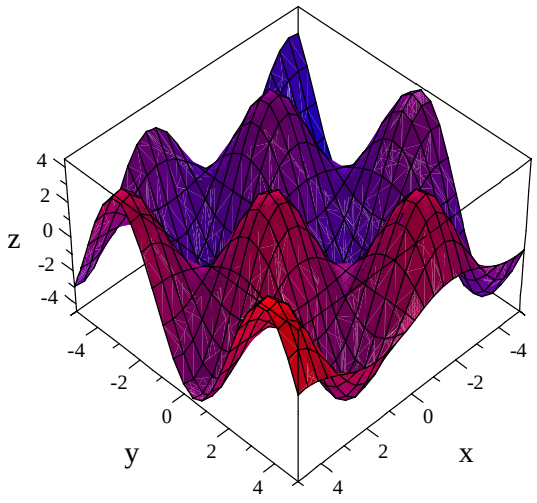


Figure 3.

On Figure 4 the graph of $A(5)$ is ilustrated,

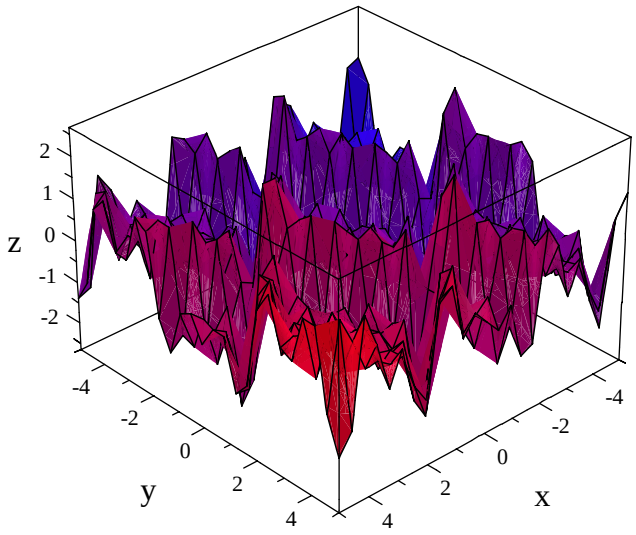


Figure 4.

and on Figures 5 and 6 the partial sums $A(5)$ is ilustated by using programme packages Scientific Workpalce 3.

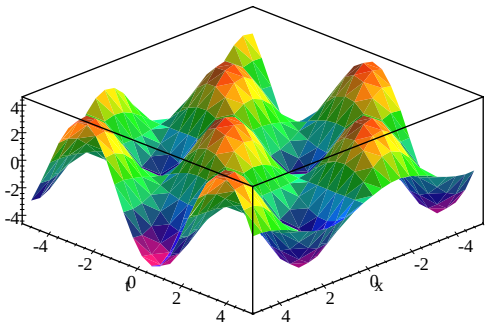


Figure 5.

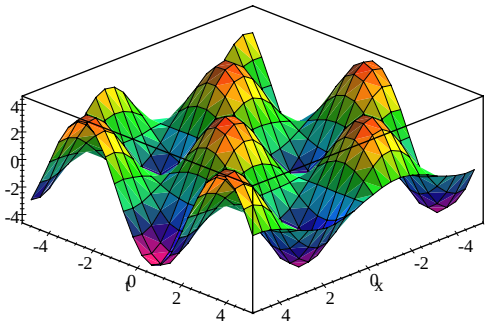


Figure 6.

Uticaj programskih paketa na usvajanje pojmova više matematike

Djurdjica Takači, Arpad Takači

UVOD

U našim školama, u Srbiji, obrada pojmova više matematike počinje u trećem razredu srednje škole, odnosno kada učenici imaju sedamnaest godina. Prema Pijažeovoj teoriji učenici su u tom periodu sposobni za formalno logičko mišljenje, što znači da su u stanju da prihvate pojmove više matematike. Usvajane pojmova više matematike se bazira na usvajanju definicija i teorema za koje nisu dovoljna samo odlična znanja elementarne matematike.

Na primer, prilikom pokazivanja konvergencije datog niza ka određenoj vrednosti, najvažnije je da učenici shvate potrebu

"*postojanja prirodnog broja* n_0, *takvog da za svako* $n > n_0$, *važi odgovarajuća nejednakost…,*."

što se ne može postići samo na osnovu znanja iz elementarne matematike. Prilikom obrade ovakvih zadataka uvek je potrebno izvršiti i određene elementarne transformacije, majoracije, minoracije, i drugo. Međutim, da bi se zadatak uradio korektno, potrebno je u određenom momentu izvesti zaključak koji karakteriše viši oblik mišljenja. U literaturi se mišljenje zasnovano na:

- aksiomama, jednostavnim teoremama koje slede iz aksioma i komplikovanijim teoremama koje slede iz dokazanih teorema;
- formalno logičkom, deduktivnom načinu zaključavanja,

naziva "*Napredno matematičko mišljenje*" ili na engleskom "*Advanced mathematical thinking"* (Tall, D., Vinner, A., *Concept Image and Concept Definitionin Mathematics with particular reference to Limits and Continuity*).

Nastavnik treba da omogući učeniku što jednostavniji prelaz sa elementarnog matematematičkog mišljenja na napredno matematičko mišljenje. U literaturi, posebno u radovima Dejvida Tola sa Univerziteta Varvik u Engleskoj i njegovih saradnika, ulozi nastavnika, u tom procesu se pridaje velika važnost. Najvažnije je da se postojeće slike koje učenici imaju o određenim pojmovima povežu i nadograde i tako prihvate pojmovi više matematike. Znači, prilikom obrade pojmova više matematike potrebno je kombinovati

- metod slike,
- metod definicije

i tako upotpuniti znanje o poznatim pojmovima sa novim pojmovima.

Na primer, učenici se sa pojmom niza sreću dosta rano u svom matematičkom obrazovanju. Međutim granična vrednost niza se obrađuje u trećem razredu srednje škole i poznato je da učenici imaju mnogo problema prilikom prihvatanja pojma granične vrednosti niza, a kasnije i granične vrednosti funkcije. Zato je potrebno povezati one mentalne slike koje učenici imaju o pojmu niza sa matematičkom definicijom niza,

funkcije koja preslikava skup prirodnih brojeva u skup realnih brojeva. Matematička definicija pojma niza podrazume da je skup vrednosti niza beskonačan, što opravdava potrebu posmatranja granične vrednosti niza.
U četvrtom razredu srednje škole učenici, pored granične vrednosti, obrađuju izvod, njegovu primenu, kao i integral (određeni i neodređeni).

U današnje vreme računari su prisutni svuda oko nas i svakako je potrebno ispitivati uticaj računara u našem obrazovanju, posebno u nastavi matematike. U ovoj knjizi je posebno naglašena upotreba računara prilikom usvajanja pojmova više matematike.
Računar se koristi kao pomoć u nastavi matematike i koriste se programski paketi *Scientific Workplace* i *GeoGebra,* sa kojima rade nastavnici a znaju da koriste i učenici.
U radovima [5], [6], [7], [8], [9], [10], [11], koristi se programski paket *Scientific Workplace* za objašnjnje pomova više matematike.
Scientific Workplace ima numeričke i grafičke mogućnosti kao i obradu teksta, dok *GeoGebra* sadrži odlične grafičke mogućnosti. Programski paket *Scientific Workplace* je veoma pogodan za rad sa učenicima i studentima jer se lako prihvaća i studeni su uglavnom usredsređeni na rešavanje odgovarajućih problema, a ne na sam programski paket.
Numeričke mogućnosti programskih paketa, kao što su određivanje izvoda, rešavanje algebarskih i diferencijalnih jednačina, nejednačina, integrala (određenih i neodređenih) i drugo, su veoma korisni za proveru rezultata određenih zaadtaka, odnosno kao pomoć pri klasičnom radu.
U radu se analizira upotreba pomenutih programskih paketa *Scientific Workplace* i *GeoGebra-e,* kao pomoć u nastavi matematike, a prilikom obrade pojmova više matematike. Iznećemo prednosti i mane primene računara u nastavi matematike.

Naprekidnost funkcija

Programski paketi *Scientific Workplace* i *GeoGebra* su veoma korisni baš kod obrade neprekidnosti funkcija u srednjoj školi i na fakultetu. Formalna definicija neprekidne funkcije se teško prihvata čak i na fakultetu pa se predlaže uvodjenje neprekidne funkcije "nežno bez podizanja olovke sa papira," što se pomoću računara veoma lako prihvata. Nastavna tema "Obrada neprekidne funkcije u srednjoj školi" detaljno je obrađena u doktorskoj disertaciji Duške Pešić i radovima [6], [7], [8].
Grafike neprekidnih funkcija možemo nacrtati pomoću pomenutih programskih paketa i ispitati osobine neprekidnih funkcija.
Međutim, poznato je da vizualizacija "crtanja neprekidne funkcije bez podizanja olovke sa hartije" izaziva kognitivne konflikte kod posmatranja funkcija

$$f(x)=\frac{1}{x},\ \ x\neq 0,\ \ f(x)=tgx,\quad x\neq\frac{2k+1}{2}\pi,\ \ k\in Z,$$

čiji grafici imaju pekide.
Nastavnici povezivanjem sa primerima iz svakodnevnog života treba da otklone nastale probleme.

Na primer, rečenica:
“Sneg je neprekidno padao ceo dan”
ukazuje da se padanje snega vezuje za vreme, što bi odgovaralo posmatranju neprekidnih funkcija samo u tačkama koje pripadaju domenu.

U daljem rada prikazaćemo rezultate testiranja učenika gimnazije „Jovan Jovanović – Zmaj“ u novom Sadu, koje je izvršeno u aprilu 2007 godine. Testirane su dve grupe učenika:
Prva grupa je imala 57 učenika kod kojih je neprekidnost obrađivana pomoću računara.
Druga grupa je imala 59 učenika i oni nisu koristili računar.
Test je preuzet iz rada [2] i imao je samo jedno pitanje:
Da li su sledeće funkcije neprekidne?
Sve funkcije su date sa grafikom osim poslednje.
U sledećoj tabeli prikazaćemo rezultate naše dve grupe, i rezultati su uporedjeni sa rezultatima Dejvida Tola označene sa T.

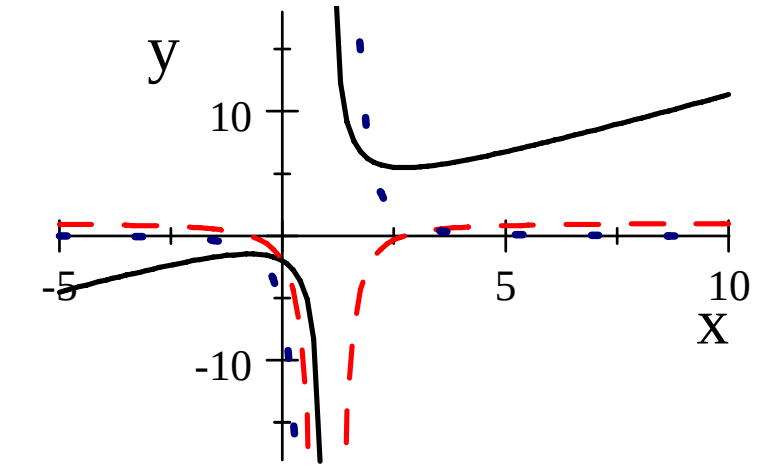
function	The first group	The second group	T
$f(x) = x^2$	100%	84,7%	100%
$g(x) = \frac{1}{x}, \quad x \neq 0$	52,6%	18,6%	14.6%
$h(x) = \begin{cases} 0 & if \quad x < 0 \\ x & if \quad x \geq 0 \end{cases}$	94,6%	52,5%	65.8%
$r(x) = \begin{cases} 0 & if \quad x < 0 \\ 1 & if \quad x \geq 0 \end{cases}$	84,2%	44,0%	68.3%
$p(x) = \begin{cases} 0 & if \quad x \in Q \\ x & if \quad x \in I \end{cases}$	69,4%	36,5%	63.4%

Primetimo da su rezultati koje su dobili učenici prve grupe koja je radila sa računarom najbolji u prepoznavanju neprekidne funkcije.

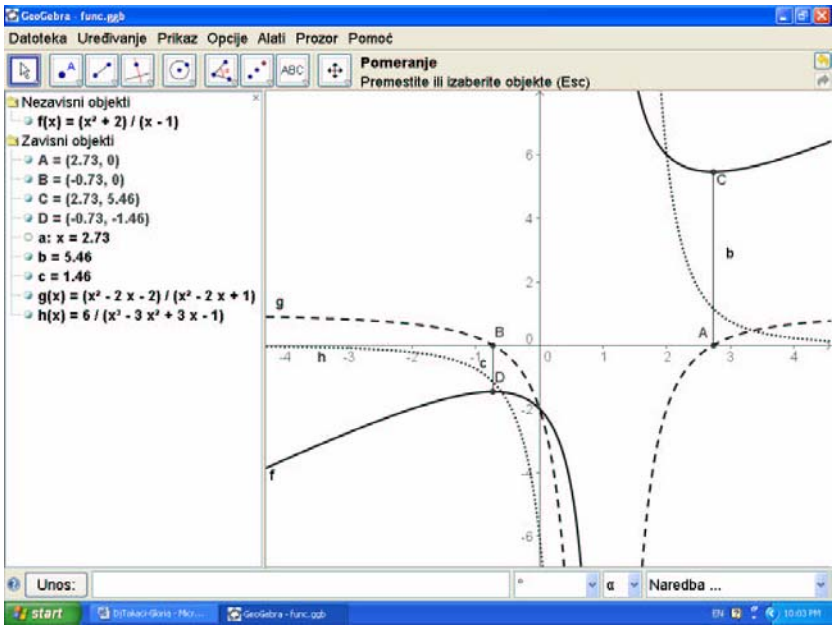
Grafičko predstavljanje funkcija

Grafičko predstvljanje funkcija pomoću računara je jedna od najznačijih primena računara u nastavi matematike. Poznato je da učenici i studenti imaju velikih problema kod isptivanja toka funkcije, a posebno kod povezivanja rezultata dobijenih ispitivanjem funkcija sa grafikom funkcija. U literaturi se u danjašnje vreme sve više koristi računar baš u tu svrhu. Navešćemo zanimljive primere.

Primer: Na slici 1 prikazan je grafik funkcije $f(x)=\frac{x^2-2}{x-1}$, njen prvi i drugi izvod pomoću programskog paketa *Scientific WorkPlace 5.5, a na slikama* 2 i 3 su prikazani isti grafici u programskom paketu *Geogebra.*

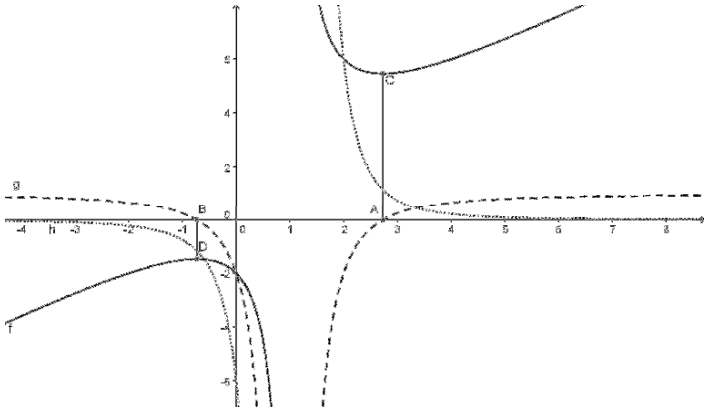


Slika 1.



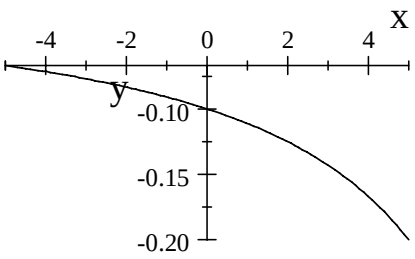
Slika 2.

Iz izloženog se vidi da je računar zaista od velike koristi kod grafičkog predstavljanja funkcije, međutim potrebno je upozoriti učenike i studente da prilikom rada sa računarom moraju biti veoma oprezni, jer se svakako javljaju novi problemi različiti od postojećih, a na koje svakako treba upozoriti studente i učenike.

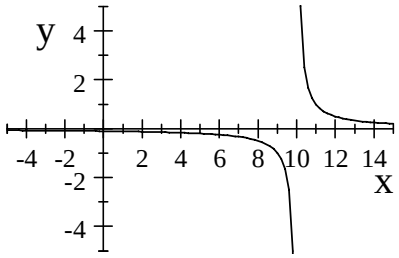


Slika 3.

Primer. Grafik funkcije $f(x)=\dfrac{1}{x-10}$ je prikazan na slikama 4 i 5.



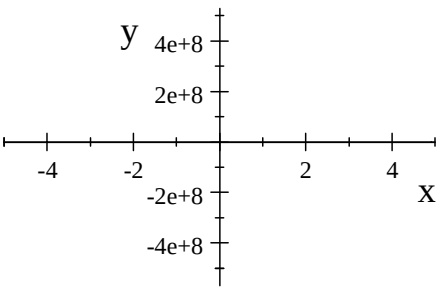
Slika 4.



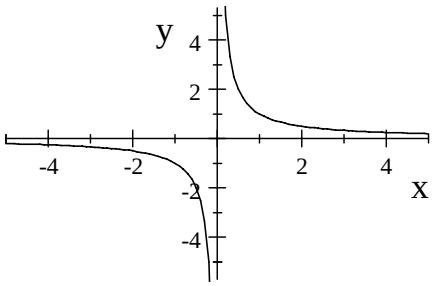
Slika 5.

Učenici će svakako primetiti da nešto nije u redu na slici 4, a da je na slici 5 zaista prikazan grafik funkcije. Znači, potrebno je ukazati da je u ovom slučaju bio problem u intervalu na x osi, jer su većina programskih paketa tako nameštena da je interval na x osi $(-5,5)$.

Grafik funkcije $f(x)=\frac{1}{x}$, $x\neq 0$, se na slici 6 ne može videti jer je prikazano $y\in(-5\cdot 10^{-8},5\cdot 10^{-8})$, ali na slici 7 je $y\in(-5,5)$.



Slika 6.



Slika 7.

Znači, učenici i studenti moraju voditi računa o intervalima i na x i na y osi, posebno ako treba da koriste grafik kao pomoć za ispitivanje toka funkcije.
Ispitivanje trigonometrijskih funkcija, kao što su na primer

$$\cos x + \tfrac{1}{2}\cos 2x, \quad \sin x \sin 3x, \quad \sin^3 x + \cos^3 x,$$

i drugih i crtanje njihovih grafika je veoma teško i za učenike i studente, pa se zato često i ne zadaje za pismene zadatke i provere znanja, medjutim sa računarom je ispitivanje toka I crtanje grafika ovakvih funkcija znatno jednostavnije. Treba napomenuti da su ovako odabrane funkcije pogodne za i klasičan rad bez računara, što znači da se računar koristi samo kao pomoć. Svakako, sada se pomoću računara mogu raditi i takve funkcije kod kojih se ne dobijaju tako lepi rezultati. Prilikom rada sa trigonometriskim funkcijama pomoću programskih paketa javljaju se novi problemi koje treba otkloniti.

Primer. Pomoću programskog paketa *Scientific Workplace* 5.5 (najnovija verzija) dobijaju se rešenja jednačina $\sin x = 0$, $\sin x + \cos x = 0$, na sledeći način:

$\sin x = 0$ **, Solution is:** $\{\pi k \mid k \in \mathrm{Z}\}$,

$\sin x + \cos x = 0$ **, Solution is:** $\{\tfrac{3}{4}\pi + \pi k \mid k \in \mathrm{Z}\}$.

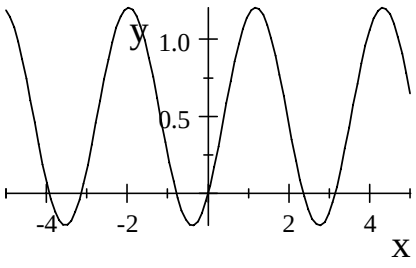
Ali, rešenje jednačine $\sin x(\sin x + \cos x) = 0$ se ne može dobiti:

$\sin x(\sin x + \cos x) = 0$ **, No solution found**.

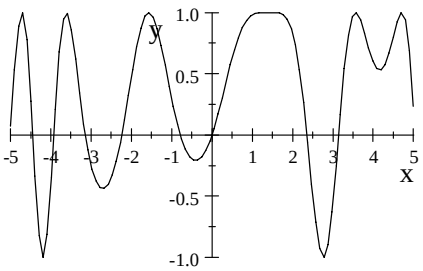
Ako se prethodna jednačina zapiše u obliku: $\sin^2 x + \sin x \cos x = 0$; tada se dobijaju tačna reršenja:

$\sin^2 x + \sin x \cos x = 0$ **, Solution is:** $\{\tfrac{3}{4}\pi + \pi k \mid k \in \mathrm{Z}\} \cup \{\pi k \mid k \in \mathrm{Z}\}$.

Grafik funkcije $f(x) = \sin^2 x + \sin x \cos x$ je dat na slici 8, međutim ako se funkcija napiše kao $f(x) = \sin x(\sin x + \cos x)$ tada se pomoću programskog paketa *Scientific Workplace* 5.5. dobija grafik na slici 9, što je pogrešno.

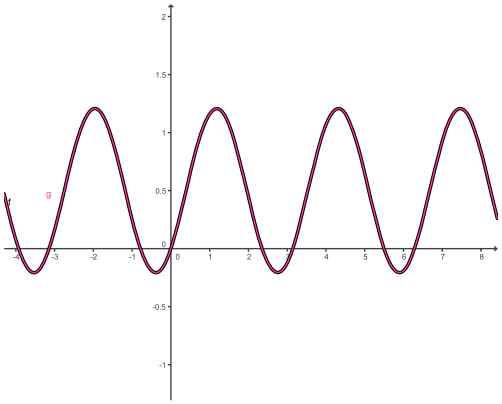


Slika 8.



Slika 9.

U programskom paketu *GeoGebra* grafik funkcije je nacrtan korektno, bez obzira u kojoj formi je zapisan analitički izraz funkcije, što je prikazano na slici 10.



Slika 10.

Integrali

Progarmski paket *Scientific Workplace* je veoma pogodan za numeričko izaračunavanje integrala, dok *GeoGebra* odmah prikaže i grafik integralne funkcije (sa konstantom $C = 0$), odnosno odgovarajuću površinu, ako je u pitanju određeni integral. Medjutim i ovde se javljaju problemi o kojima se mora voditi računa:

Primer: Integral $\int \frac{dx}{x^2 - 1}$ se pomoću *Scientific Workplace*, izračunava kao:

$$\int \frac{dx}{x^2 - 1} = -\operatorname{arctanh} x$$

Naši učenici i istudenti ne rade sa funkcijom $\operatorname{arctanh}$, i kada dobiju ovakav rezultat prilično se oneraspolože. Ali, ako se prvo data racionalna funkcija rastavi na parcijalne razlomke:

$$\frac{1}{x^2 - 1} = \frac{1}{2(x - 1)} - \frac{1}{2(x + 1)},$$

tada se pomoću *Scientific Workplace* dobija:

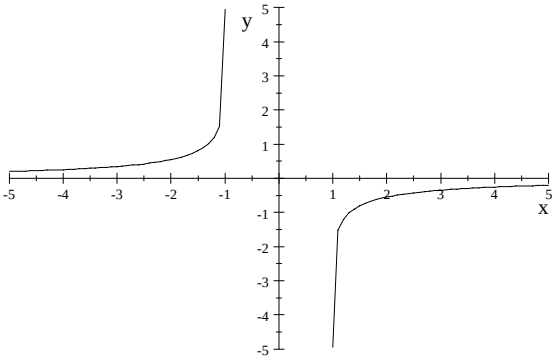
$$\int \left(\frac{1}{2(x-1)} - \frac{1}{2(x+1)} \right) dx = \frac{1}{2}\ln(x-1) - \frac{1}{2}\ln(x+1).$$

Učenicima treba skrenuti pažnju da je u prethodnom radu izostavljena i apsolutna vrednost i konstanta,
Grafik funkcije

$$\int\left(\frac{1}{2(x-1)}-\frac{1}{2(x+1)}\right)dx\ ,$$

je prikazan u programskom paketu Scientific WorkPlace, na slici 11, pa se vidi se da je u stvari prikazan ispravno grafik funkcije

$$\frac{1}{2}\ln|x-1|-\frac{1}{2}\ln|x-1|.$$



Slika 11.

Vizualni pristup definiciji izvoda funkcije

Djurdjica Takači, Marjetica Samardžijevic

Uvod

U ovom radu ćemo prikazati uvodjenja prvog izvoda realne funkcije jedne realne promenljive. Delovi rada su objavljeni u radu [8].
Koristiće se programski paketi *Scientific WorkPlace* 3 i 5.5 i *GeoGebra*.
Izvod realne funkcije jedne realne promenljive uvodi se preko granične vrednosti količnika priraštaja

$$f'(x_0) = \lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}, \qquad (1)$$

u tački x_0.
Izračunavanja graničnih vrednosti oblika (1) rade se uglavnom posle časova utvdjivanja gradiva granične vrednosti funkcija, odnosno kao primena granične vrednosti. Zanimljivo je da se u svetu, na primer u Engleskoj, učenici prvi put sreću sa graničnom vrednosti baš u momentu kada definišu prvi izvod funkcije.
U zavisnosti od nivoa znanja učenika možemo povezati izvod i neprekidnost, odnosno možemo reći da je **neprekidnost** *potreban* uslov za postojane pvog izvoda, ali nije i *dovoljan,* što se lepo ilustruje na odgovarajućim primerima.
Na osnovu iskustva je poznato da u ovom uzrasnom dobu i naši učenici, često nemaju jasnu sliku graničnog procesa, posebno kada se posmatra granična vrednost količnika priraštaja i zato se veliki značaj pridaje vizualnom pristupu definicije izvoda funkcije.

Vizualizacija prvog izvoda funkcije

U cilju uvodjenja prvog izvoda realne funkcije $f : D \to R$ (jedne realne promenljive) označićemo sa $k(h)$ količnik priraštaja:

$$k(h, x) = \frac{f(x + h) - f(x)}{h}.$$

Funkcija k jeste funkcija dve promenljive x i h, i u daljem radu ćemo menjati parametar h i pratiti grafike funkcije k. Programski paketi *Scientific WorkPlace* 3 i 5.5 i *GeoGebra* to dozvoljava.
Posebno se naglašava da, ako se vrednost h smanjuje i teži nuli, tada se grafici odgovarajućih funkcija k "približavaju" grafiku jedne funkcije, što se na slici vidi "poduplavanjem krivih". Na taj način se vizualno dolazi do grafika izvodne funkcije, koja je rezultat graničnog procesa i koja se označava sa f'.

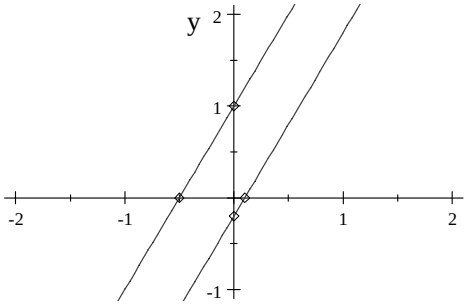
PRIMERI

.

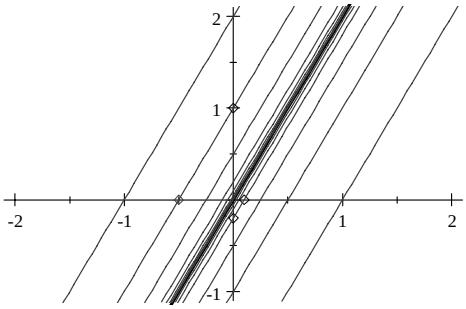
Primer1. Za funkciju $f(x) = x^2$ količnik priraštaja je

$$k(x,h) = \frac{(x+h)^2 - x^2}{h}.$$

Grafici funkcije k, za različite vrednosti h, su prikazani na slici 1 i na slici 2.



Slika 1.



Slika 2.

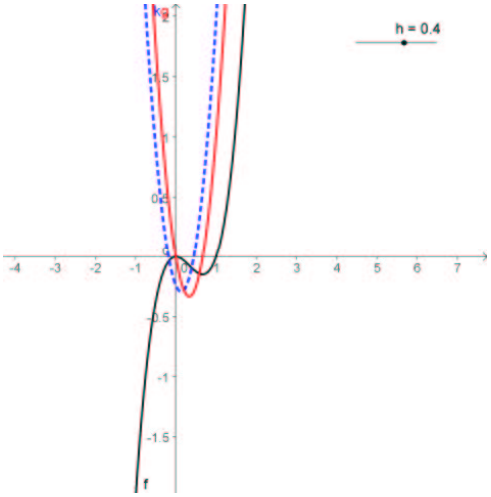
Ako izaberemo $h = 1$ i posmatramo tačke $A(-.5, 0)$ i $B(0,1)$, koje pripadaju grafiku funkcije $k(x,1)$ (slika 1) , tada je jednačina prave odredjene tačkama A i B ima oblik $y = 2x + 1$.
Analogno se za $h = -0.2$ i tačke $C(.1, 0)$ i $D(0,-0.2)$, koje pripadaju grafiku funkcije $k(x,-0.2)$ (slika 1), dobija jednačina prave odredjene tačkama C i D $y = 2x - 0.2$.
Količnik priraštaja k se može napisati kao

$$k(x,h) = \frac{(x+h)^2 - x^2}{h} = 2x + h,$$

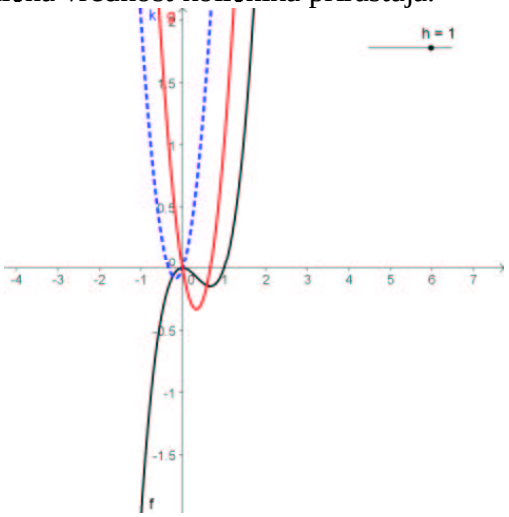
pa se za $h = 1$ i $h = -0.2$ dobija $k(x,1) = 2x + 1$, $k(x,0.2) = 2x - 0.2$, respektivno. Znači prava $y = 2x + 1$ jeste grafik funkcije $k(x,1)$, odnosno prava $y = 2x + 0.2$ jeste grafik funkcije $k(x,0.2)$.

Vizualizacija u programskom paketu Geogebra

Svi navedeni primeri u prethodnom delu rađeni su u programskom paketu *Scientific WorkPlace* 3 i 5.5. U daljem radu ćemo prikazati kako se radi u programskom paketu *GeoGebra*. U ovom primeru se posmatra funkija $f(x) = x^3 - x^2$, njen izvod i količnik priračtaja. Pored toga uveden je i klizač h, čijim pomeranjem možemo prikazati vizualizaciju granične vrednosti količnika priaštaja i prikazati kako ona teži ka izvodnoj funkciji, kada $h \to 0$.
Funkcija se u ovom programu može vrlo jednostvno promeniti naredbom redefinisanje, a time se menja i odgovarajući izvod kao i granična vrednost količnika priraštaja.



Slika 12



Slika 12

U programskom paketu *Geogebra* možemo pomorati klizače i pratiti dobijanje prvog izvoda ..\Nastava1.ggb ili E:\KURS--APRIL\ zaKNJ\Nastava1.ggb.

O vizualizaciji diferencijala funkcije

Stevan Pilipović, Đurđica Takači, Arpad Takači

Uvod

U radu se prikazuju istraživanja o uticaju računara, odnosno Computer Algebra System (CAS) u Analizi. Koristi se programski paket *GeoGebra*, koji je uveo Markus Hohenwarten, i prikazuje vizualizacija diferencijala realne funkcije jedne i dve realne promenljive. Posebna pažnja se posvećuje prikazivanju dobrih i loših strana primene programskog paketa *GeoGebra*, odnosno računara uopšte.

U prethodnom radu posmatrana je vizualizacija količnika priraštaja i tako je dobijen prvi izvod. U ovom radu se vizualno iz nagiba sečice grafika date funkcije, dobija nagib tangente grafika funkcije. Prikazuje se dobija geometrijsko značenje prvog izvoda sa posebnim osvtom na diferencijal funkcije.

Geometrijska interpretacija izvoda funkcije $f:\mathbb{R}\to\mathbb{R}$ u tački x_0

Ako je funkcija f definisana na otvorenom intervalu koji sadrži tačku x_0, tada je izvod dat sa:

$$f'(x_0)=\lim_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h},$$

pod uslovom da data granična vrednost postoji.
Ako označimo sa $A(x_0, f(x_0))$ i $B(x_0+h, f(x_0+h))$ tačke na grafiku funkcije f, tada je nagib sečice koja prolazi kroz tačke A i B dat sa:

$$k_s = \frac{f(x_0+h)-f(x_0)}{h},$$

a nagib tangente na funkciju f u tački A je dat sa

$$k_t = \lim_{h\to 0}\frac{f(x_0+h)-f(x_0)}{h},$$

odosno jednak je prvom izvodu funkcije f u tački x_0.
Broj h je priraštaj nezavisno promenljive x, a Δf je priraštaj zavisno promenljive f odnosno,

$$\Delta f = f(x_0+h)-f(x_0).$$

Funkcija f je diferecncijabilna u tački $x_0 \in [a,b]$ ako postoji linearno prislikavanje $\omega : \mathbb{R}\to\mathbb{R}$ takvo da važi

$$\Delta f = f(x_0+h)-f(x_0) = f'(x_0)h + \omega(h)h,$$

gde ω ima osobinu da $\omega(h)\to 0$ kada $h\to 0$.
Izraz $f'(x_0)h$ je diferencijal funkcije f i označava se sa df.
Diferencijal df funkcije $f : \mathbb{R}\to\mathbb{R}$ je dat sa

$$df = f'(x)dx.$$

Vizualizacija diferencijala funkcije $f:\mathbb{R}\to\mathbb{R}$

Neka je $A(x_0, f(x_0)$ tačka na grafiku funkcije $f:\mathbb{R}\to\mathbb{R}$, i neka je $B(x_0+h, f(x_0+h)$ tačka na grafiku funkcije f, gde je h razlika abscisa tačaka A i B. Označimo sa $C(x_0,0)$ i $D(x_0+h,0)$, projekcije tačaka A and B, na x-osu. Na sledećim slikama koristićemo tačke $E(x_0+h, f(x_0))$, odnosno projekciju tačke A vertikalnu pravu koja prolazi kroz tačku D. Kako je

$$\tan\beta = \frac{BE}{AE} = \frac{\Delta f}{h},$$

to je koeficijent pravca sečice koju određuju tačke A and B data sa

$$\tan\beta = \frac{BE}{AE} = \frac{\Delta f}{h} = \frac{f(x_0+h)-f(x_0)}{h}.$$

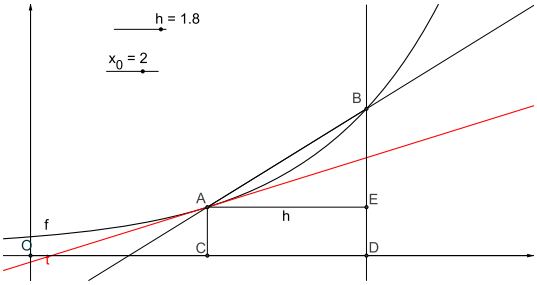
Neka je t tangenta grafika funkcije f u tački A, tada, puštajući da $h \to 0$, sečica se približava tangenti odnosno, ugao između sečice i tangente koje imaju zajedničku tačku A, teži nuli, kada $h \to 0$.

Na slikama 1-4, urađenim u *GeoGebri*, postoje dva klizača, x_0 i h. U programskom paketu *GeoGebra* koristili smo dinamičke osobine, koje dozvoljavaju promenu vrednosti x_0 i h, i time se dobija vizualizacija sledeće rečenice:

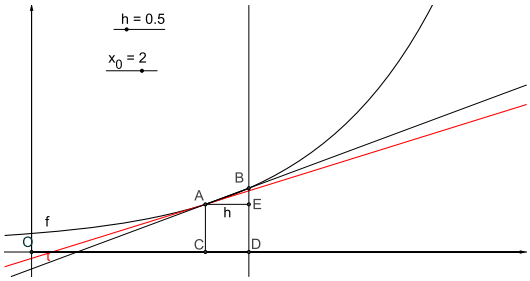
" Ako h teži nuli , tada se sečica približava tangenti ".

Promenom x_0, može se videti ista osobina.

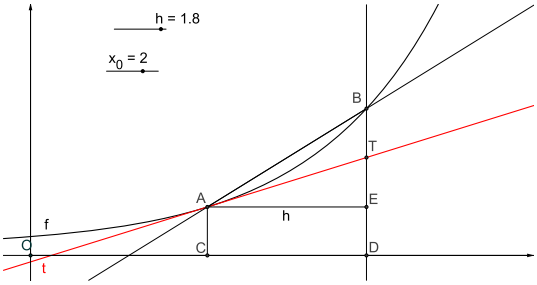
Na slikama 1 i 2, je $x_0 = 2$, a $h = 1.8$ i $h = 0.5$.



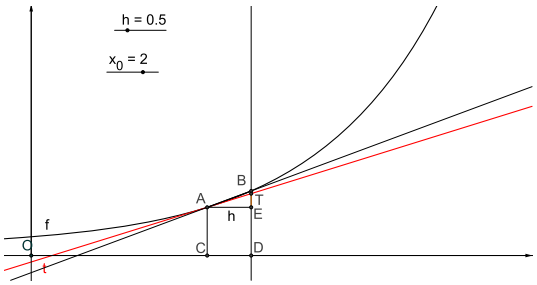
Slika 1.



Slika 2.



Slika 3.



Slika 4.

Neka je T tačka preseka tangente t i normale na x-osu u tački D (slike 3 i 4). Iz trougla ΔEAT se može se videti da je koeficijent pravca tangente na grafik funkcije f u tački x_0 $f'(x_0)$ jednak:

$$f'(x_0) = \frac{TE}{h}, \quad \text{i.e.,} \quad TE = f'(x_0)h.$$

Iz $dy = f'(x_0)dx$ sledi da je TE geometrijska interpretacija prvog izvoda diferenciala funkcije f u tački A.

Na slici 3 se posmatra $x_0 = 2$, i $h = 1.8$, a na slici 4 se posmatra $h = 0.5$ $(x_0 = 2)$.

Funkcija f je konveksna (slike 1 i 2) i zato je $BE > ET$.
Primetimo da se x_0 može pogodno menjati i pratiti šta se dešava sa promenom parametra

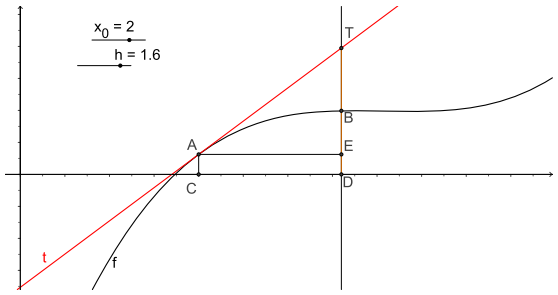
E:\KURS--APRIL\
zaKNJ\Knj-Srp\Figure

h.
Na slici 5 se posmatra $x_0 = 2,$ i $h = 1.6,$ a na slici 6 je $x_0 = 2,$ i $h = 0.3.$
Funkcija f is je konkavna (slike 3 i 4) i zato je $BE < ET$.
Na obe slike se može prikazati jednakost

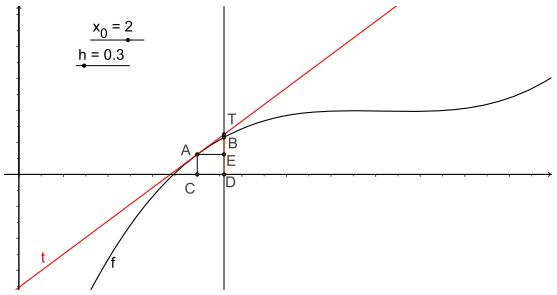
$$\Delta f = f(x_0 + h) - f(x_0) = f'(x_0)h + \tau_1(h)h,$$

a u *Geogebri* se može pratiti granični proces. Naime, kada $h \to 0$, tada izraz $\tau_1(h)h \to 0,$ teži nuli "brže" nego $f'(x_0)h.$
Takođe se može videti da je glavni deo poslednje sume ustvari diferencijal funkcije f u tački x_0, i možemo zapisati sledeću aproksimaciju

$$\Delta f \approx f'(x_0)h, \quad h \approx 0.$$



Slika 5.



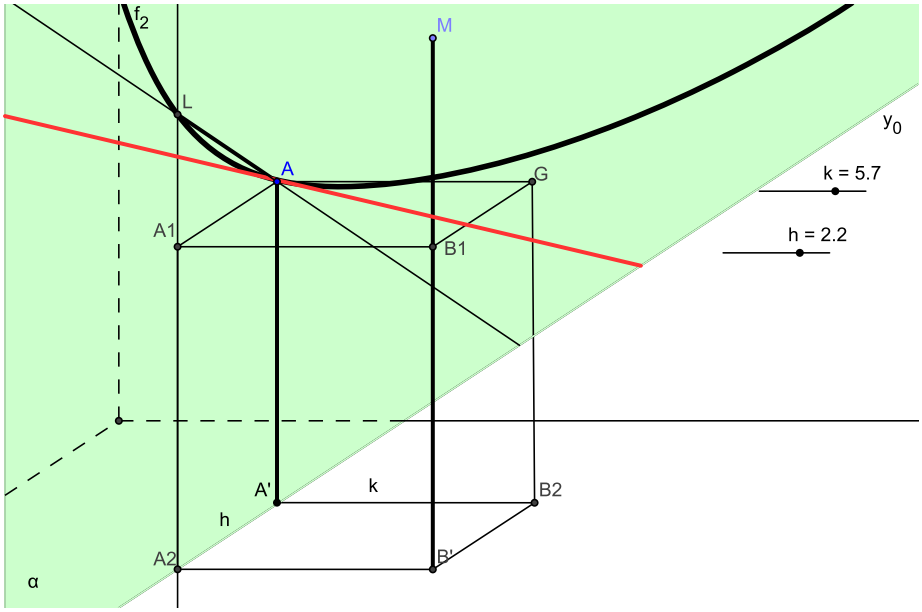
Slika 6.

Geometrijsko tumačenje parcijalnog izvoda funkcije $f : \mathbb{R}^2 \to \mathbb{R}$ u tački (x_0, y_0)

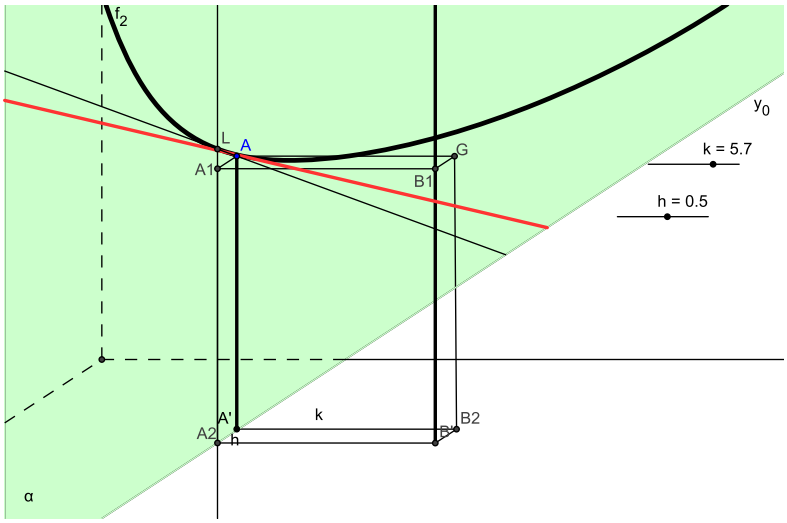
U daljem radu ćemo pikazati jednu vizualizaciju geometrijskog tumačenja parcijalnog izvoda funkcije $f : \mathbb{R}^2 \to \mathbb{R}$ u tački (x_0, y_0) pomoću programskog paketa *GeoGebra.*
Primetimo da ćemo koristiti verziju *GeoGebra-e* u dve dimienzije, i prikazati objekte u tri dimenzije.
Prvo ćemo prikazati geometrijsku interpretaciju parcijalnog izvoda $\frac{\partial f}{\partial x}(x_0, y_0)$ i izraza

$\frac{\partial f}{\partial x}(x_0, y_0)h$.

Na slici 7, posmatramo tačke $A(x_0, y_0, f(x_0, y_0))$, $A'(x_0, y_0, 0)$, $A_2(x_0 + h, y_0, 0)$, kao i tačke $B(x_0 + h, y_0 + k, f(x_0 + h, y_0 + k))$, $B'(x_0 + h, y_0 + k, 0)$. Tačke A' i A_2 pripadaju pravoj $y = y_0$, koja je paralelna sa x osom. Posmatrajmo ravan α, koja je paralelna xz -ravni i sadrži tačke A' and A_2. Ravan α seče grafik funkcije (od dve promenljive) f (površ) po krivoj f_2.



Slika 7.



Slika 8.

Ako uvedemo koordinatni sistem u ravni α, tada tačka:

- $A_1\big((x_0+h, y_0), f(x_0, y_0)\big)$,
- duž $AA' = f(x_0, y_0)$,
- duž A_2L, gde je tačka L data sa $L = (x_0+h, y_0, f(x_0+h, y_0)$,
- duž $A'A_2 = h$

pripadaju ravni α.

U ovom slučaju količnik

$$\frac{f(x_0+h, y_0) - f(x_0, y_0)}{h}$$

jeste nagib sečice koja sardži tačke A i L uravni α.
Nagib tangente t na krivu f_2 u tački A jeste granična vrednosti količnika priraštaja

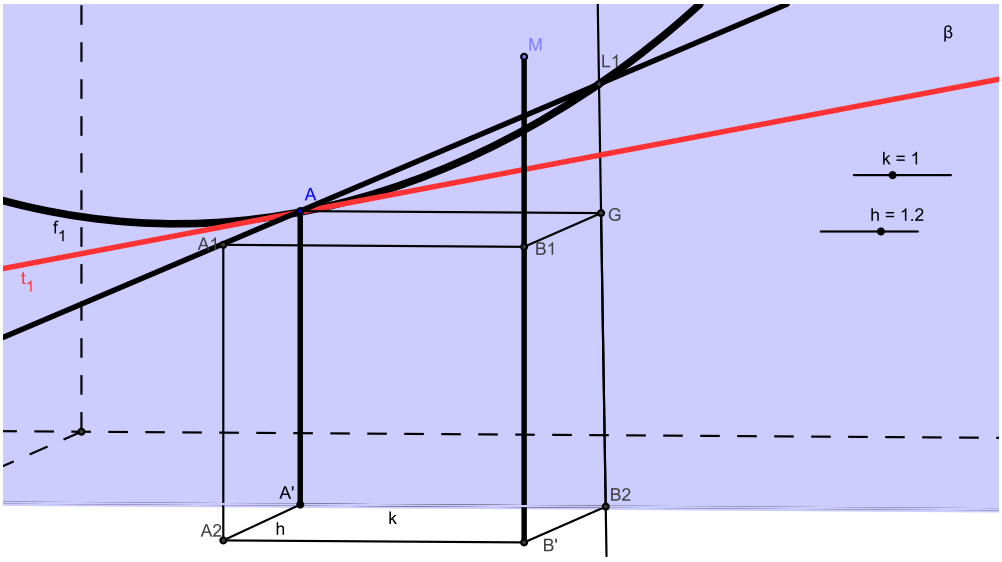
$$\frac{f(x_0+h, y_0) - f(x_0, y_0)}{h} = \frac{RT}{AT},$$

kada $h = AT$ teži 0, odnosno

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h\to 0} \frac{f(x_0+h, y_0) - f(x_0, y_0)}{h}.$$

Na slici 7 smo uzeli $h = 2.2$, a na slici 8 je $h = 0.5$, i u programskom paketu *GeoGebra* možemo menjati h, pa se sečica koja sardži tačke A i L u ravni α pribižava tangenti kada se h smanjuje.


E:\Pecuj07\
Figure7.ggb



Slika 9.

Analogno, se prikazuje geometrijsko tumačenje parcijalnog izvoda $\frac{\partial f}{\partial y}(x_0, y_0)$. Posmatramo ravan β koja je paralelna sa yz-ravni i prolazi kroy tačke A' i B_2. Presek grafika funkcije f (površ) i ravni je kriva koju smo označili sa f_1. Uvodimo koordinatni sistem u ravni β, i označimo sa $B_2L_1 = f(x_0, y_0 + k)$, i $A'B_2 = k$. Tada je količnik

$$\frac{f(x_0, y_0 + k) - f(x_0, y_0)}{k}$$

nagib sečice koja je određena tačkama A i L_1 u ravni β.

E:\Pecuj07\
Figure9.ggb

Na slici 9 je $k = 1$, koje može da se dalje smanjuje .
Analogno se kao u prethodnom se dobija geometrijsko tumačenje parcijalnog izvoda $\frac{\partial f}{\partial y}(x_0, y_0)$.

Priraštaj i differencijal funkcije $f : \mathbb{R}^2 \to \mathbb{R}$

U cilju vizualizacije diferencijala funkcije $f : \mathbb{R}^2 \to \mathbb{R}$, posmatraćemo priraštaj funkcije $f : \mathbb{R}^2 \to \mathbb{R}$ data je

$$\Delta f(x_0, y_0) = f(x_0 + h, y_0 + k) - f(x_0, y_0).$$

Za diferencijal funkcije važi sledeća teorema:

Ako funkcija $f : \mathbb{R}^2 \to \mathbb{R}$ *definisana na* $\mathbb{R}^2 = \{(x, y) : a < x < b, \quad c < y < d\}$, *i neka su prvi parcijalni izvodi* $\frac{\partial f}{\partial y}$ *i* $\frac{\partial f}{\partial y}$ *neprekidne funkcije u tački* $(x_0, y_0) \in \mathbb{R}^2$. *Ako* $(x_0 + h, y_0 + k) \in \mathbb{R}^2$, *tada važi*

$$\Delta f = f(x_0 + h, y_0 + h) - f(x_0, y_0) = \frac{\partial f}{\partial x}(x_0, y_0)h + \frac{\partial f}{\partial y}(x_0, y_0)k + \tau_1(h,k)h + \tau_2(h,k)k,$$

gde $\tau_1 \to 0, \tau_2 \to 0,$ *kada* $(h,k) \to (0,0)$.
Ako su dx i dy diferencijali x i y i $dx = h, \quad dx = k,$ tada je diferencial funkcije $f : \mathbb{R}^2 \to \mathbb{R}$ definisan kao:

$$df = \frac{\partial f}{\partial x}(x_0, y_0)dx + \frac{\partial f}{\partial y}(x_0, y_0)dy$$

Ako označimo sa T tačku preseka tangente t i vertikale u tački A_2 (Slika 10), tada iz trougla ΔATA_1 sledi da je nagib tangente t dat sa $\frac{TA_1}{h}$, odnosno:

$$TA_1 = \frac{\partial f}{\partial x}(x_0, y_0)h,$$

što znači da je TA_1 vizualizacija izraza $\frac{\partial f}{\partial x}(x_0, y_0)h.$

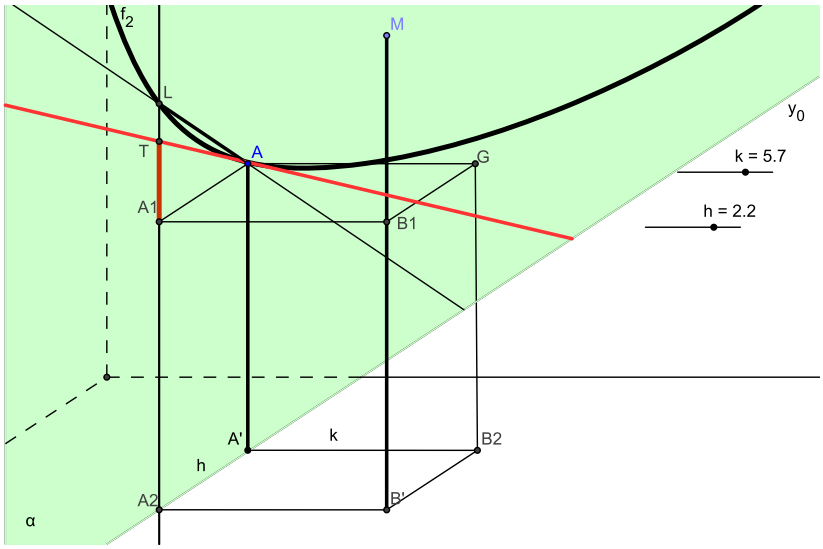
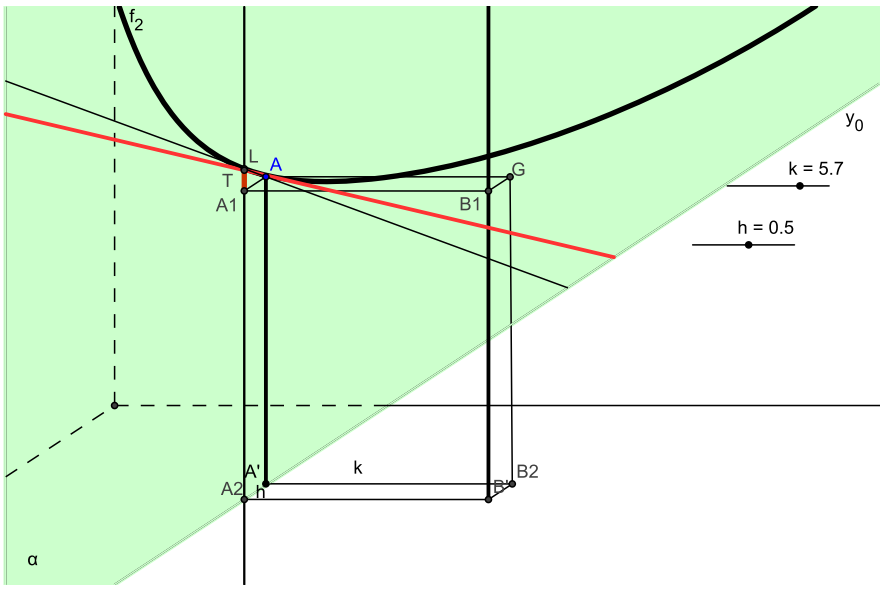


Figure 10.

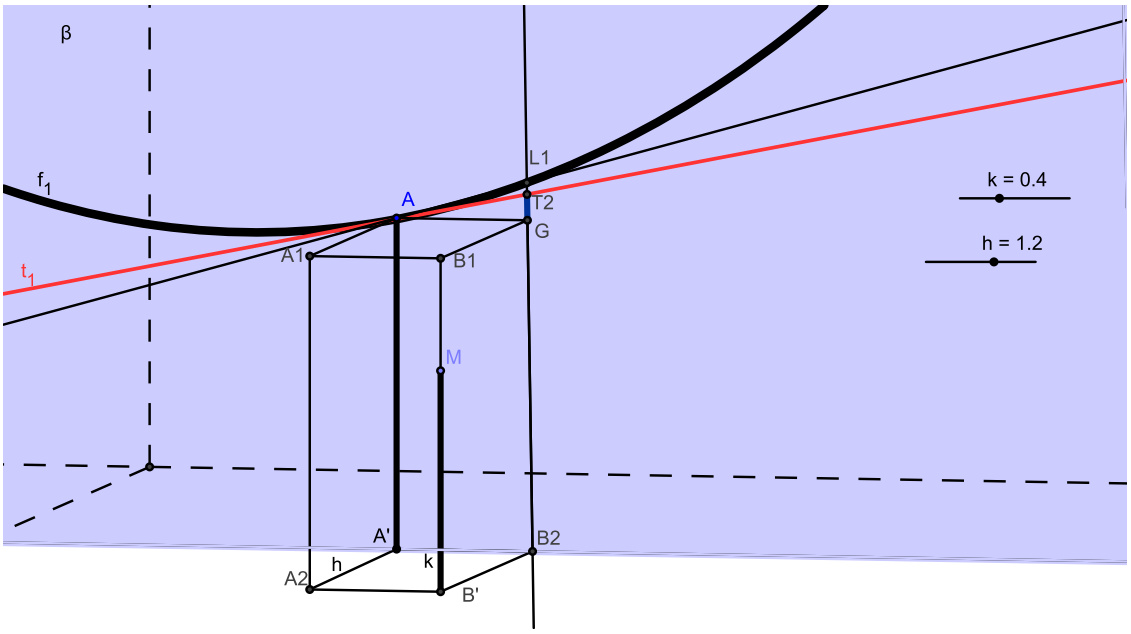
Na slici 11 je uzeto $h = .5,$ i može se videti da je dužina duži A_1T skoro ista kao i dužina duži $A_1L.$



Slika 11.

E:\KURS--APRIL\
zaKNJ\Knj-Srp\PecujF

Analogno, možemo prikazati vizualizaciju izraza $\frac{\partial f}{\partial y}(x_0, y_0)h.$



Slika 12.

Na slici 12 je $k = .4,$ i vidi se da je dužina duži GT_1 skoro ista kao i dužina duži GL_1. Analogno, se može prikazati vizualizacija izraza $\frac{\partial f}{\partial x}(x_0, y_0)h.$

Kako je

$$df = \frac{\partial f}{\partial x}(x_0, y_0)dx + \frac{\partial f}{\partial y}(x_0, y_0)dy,$$

to se

$$Df = A_1T + GT_1,$$

gde je A_1T konstruisano na Slici 11, dok je GT_1 konstruisano na Slici 12 može smatrati kao vizualization of the differential of the function $f : \mathbb{R}^2 \to \mathbb{R}$.

Furijeovi redovi

Đurđica Takači

Uvod

Furijeovi redovi su posebna vrsta funkcionalnih redova, koji su veoma značajni za rešavanja glavnih jednačina matematičke fizike.

Definicija: *Funkcija* $f : [a,b] \to \mathbf{R}$ *je* ***po delovima neprekidna*** *na intervalu* $[a,b]$ *ako se interval* $[a,b]$ *može podeliti na konačan broj podintervala* $[c_{j-1}, c_j]$, $j = 1,2,\ldots,n,$ *gde je* $a = c_0 < c_1 < \ldots < c_{n-1} < c_n = b,$ *tako da važi:*

- *na svakom podintervalu* (c_{j-1}, c_j) *funkcija* f *je neprekidna;*
- *postoje*

$$\lim_{x \to c_j^-} f(x) \quad i \quad \lim_{x \to c_j^+} f(x)$$

za sve $j = 1,2,\ldots,n.$ *na svakom podintervalu* (c_{j-1}, c_j).

Definicija: Funkcija $f : [a,b] \to \mathbf{R}$ ***ima po delovima neprekidan prvi izvod*** *na intervalu* $[a,b]$ *ako je po delovima neprekidna na* $[a,b]$ *i unutar svakog podintervala*

(c_{j-1}, c_j) *(na kome je neprekidna) ima i neprekidan prvi izvod, a u tačkama* c_{j-1} *i* c_j *ima ograničen desni odnosno levi izvod, respektivno, za sve* $j = 1, 2, \ldots, n$.

Neka je f periodična sa periodom 2π i po delovima neprekidna funkcija na intervalu $[-\pi, \pi]$. Neka je $\mathbf{N}_0 = \mathbf{N} \cup \{0\}$ i

$$A_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx \, dx, \qquad n \in \mathbf{N}_0, \tag{1}$$

$$B_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx \, dx, \qquad n \in \mathbf{N}, \tag{2}$$

Brojeve A_n, $n \in \mathbf{N}_0$ i B_n, $n \in \mathbf{N}$ zovemo **Furijeovim koeficijentima** funkcije f na $[-\pi, \pi]$. Trigonometrijski red

$$\frac{A_0}{2} + \sum_{n=1}^{\infty} (A_n \cos nx + B_n \sin nx), \tag{3}$$

je **Furijeov red** funkcije f na $[-\pi, \pi]$, ako su brojevi A_n, $n \in \mathbf{N}_0$, i B_n, $n \in \mathbf{N}$, dati relacijama (1), i (2) respektvno tj. ako su to Furijeovi koeficijenti funkcije f na $[-\pi, \pi]$.

Dovoljan uslov za jednakost $f(x)$ **sa zbirom reda u relaciji (3)**
Furijeov red funkcije f koja na intervalu $[-\pi, \pi]$ ima po delovima neprekidan prvi izvod i periodična je sa periodom 2π , konvergira u svakoj tački $x \in \mathbf{R}$. Ako je

$$f(x^{\pm}) := \lim_{h \to 0} f(x \pm h), \quad x \in \mathbf{R}$$

(tj. $f(x^+)$ označava desnu, a $f(x^-)$ levu graničnu vrednost funkcije f u tački $x \in \mathbf{R}$, tada je za svako $x \in \mathbf{R}$

$$\frac{f(x^+) + f(x^-)}{2} = \frac{A_0}{2} + \sum_{n=1}^{\infty} (A_n \cos nx + B_n \sin nx),$$

gde su A_n, $n \in \mathbf{N}_0$, i B_n, $n \in \mathbf{N}$, dati sa (1) i (2).

Ako f ima na intervalu $[-\pi, \pi]$ po delovima neprekidan prvi izvod i periodična je sa periodom 2π, tada za sve tačke $x \in \mathbf{R}$ u kojima je funkcija f neprekidna važi

$$f(x)=\frac{A_0}{2}+\sum_{n=1}^{\infty}(A_n\cos nx+B_n\sin nx), \tag{4}$$

tj. suma reda u (3) je baš jednaka $f(x)$.
Gore navedeni uslovi su sasvim odgovarajuci za funkcije sa kojima se najčešce srećemo, mada ti uslovi mogu i da se oslabe.
Na primer, funkcija $f(x)=|x|^{1/2}$ nema po delovima neprekidan prvi izvod ni na jednom intervalu koji sadrži nulu, ali ipak f ima konvergentan Furijeov red za svako $x\in[-\pi,\pi]$.
Furijeovi koeficijenti se dobijaju iz

$$\cos\alpha\cdot\cos\beta=\frac{1}{2}\cos(\alpha-\beta)+\cos(\alpha+\beta),$$

$$\sin\alpha\cdot\sin\beta=\frac{1}{2}\cos(\alpha-\beta)-\cos(\alpha+\beta),$$

$$\sin\alpha\cdot\cos\beta=\frac{1}{2}\sin(\alpha-\beta)+\sin(\alpha+\beta),$$

odakle se lako se dobijaju sledeće jednakosti za $k,n\in\mathbf{N}_0$:

$$\int_{-\ell}^{\ell}\cos\tfrac{k\pi x}{\ell}\cos\tfrac{n\pi x}{\ell}dx=\begin{cases}0, & k\neq n\\ \ell, & k=n\neq 0\\ 2\ell, & k=n=0;\end{cases}$$

$$\int_{-\ell}^{\ell}\sin\tfrac{k\pi x}{\ell}\sin\tfrac{n\pi x}{\ell}dx=\begin{cases}0, & k\neq n\\ \ell, & k=n\neq 0\\ 0, & k=n=0;\end{cases}$$

$$\int_{-\ell}^{\ell}\cos\tfrac{k\pi x}{\ell}\sin\tfrac{n\pi x}{\ell}dx=0.$$

Ako je Furijeov red na desnoj strani relacije (3) uniformno konvergentan na svakom zatvorenom i ograničenom intervalu, i ako je tačna jednakost (4), tada se taj red može integraliti član po član pa je

$$\int_{-\pi}^{\pi}f(x)\,dx=\int_{-\pi}^{\pi}\frac{A_0}{2}\,dx+\sum_{n=1}^{\infty}\left(A_n\int_{-\pi}^{\pi}\cos nx\,dx+B_n\int_{-\pi}^{\pi}\sin nx\,dx\right).$$

Ako sada pomnožimo (4) sa $\cos kx,\ k \in \mathbf{N}_0,$ i integralimo dobijeni red član po član nad $[-\pi, \pi]$, dobijamo

$$\int_{-\pi}^{\pi} f(x)\cos kx\,dx = \frac{A_0}{2}\int_{-\pi}^{\pi}\cos kx\,dx + \sum_{n=1}^{\infty}\left(A_n\int_{-\pi}^{\pi}\cos nx\cos kx\,dx + B_n\int_{-\pi}^{\pi}\sin nx\cos kx\,dx\right).$$

Ako je $k = 0,$ tada su svi integrali na desnoj strani, sem prvog, jednaki 0, tj.

$$\int_{-\pi}^{\pi} f(x)\,dx = \frac{A_0}{2}\int_{-\pi}^{\pi} dx = \pi A_0,$$

čime smo dobili "nultu" jednakost iz (3).

Ako je $k \in \mathbf{N}$ tada je

$$\int_{-\pi}^{\pi} f(x)\cos kx\,dx = A_k\int_{-\pi}^{\pi}\cos^2 kx\,dx = A_k\cdot\pi, \qquad k \in \mathbf{N}.$$

Odavde slede formule za koeficijente A_n u (1).
Analogno, posle množenja relacije (4) sa $\sin kx$ i integracije nad $[-\pi, \pi]$ dobijamo

$$\int_{-\pi}^{\pi} f(x)\sin kx\,dx = \frac{A_0}{2}\int_{-\pi}^{\pi}\sin kx\,dx + \sum_{n=1}^{\infty}\left(A_n\int_{-\pi}^{\pi}\sin nx\cos kx\,dx + B_n\int_{-\pi}^{\pi}\sin nx\sin kx\,dx\right).$$

U ovom su slučaju svi integrali, sem onog koga množi $B_k,$ jednaki nuli, pa dobijamo formule za B_n date u relaciji (2).
Dovoljne uslove za uniformnu konvergenciju Furijeovog reda na proizvoljnom zatvorenom i ograničenom intervalu daju sledeća dva tvrđenja.

Tvrđenje 1: *Ako red* $\sum_{n=1}^{\infty}(|A_n| + |B_n|)$ *konvergira, tada red u (3) konvergira uniformno na svakom zatvorenom i ograničenom intervalu.*

Tvrđenje 2: *Furijeov red neprekidne funkcije* $f,$ *koja ima po delovima neprekidan prvi izvod na intervalu* $[-\pi, \pi]$ *i periodična je sa periodom* $2\pi,$ *konvergira uniformno ka funkciji* f *na svakom zatvorenom i ograničenom intervalu.*

Furijeov red na proizvoljnom intervalu

Smenom promenljivih $x = \frac{t\pi}{\ell}$ $x \in [-\pi, \pi]$, $t \in (-\ell, \ell)$, dobijamo

$$F(t) = f(\frac{t\pi}{\ell}) = f(x).$$

Tada je Furijeov red funkcije F na $(-\ell, \ell)$ dat sa

$$A_0 + \sum_{n=1}^{\infty} (A_n \cos \frac{n\pi t}{\ell} + B_n \sin \frac{n\pi t}{\ell}),$$

a Furijeovi koeficijenti funkcije F su dati sa

$$A_n = \frac{1}{\ell} \int_{-\ell}^{\ell} F(t) \cos \frac{n\pi t}{\ell} dt, \qquad n \in \mathbf{N}_0,$$

$$B_n = \frac{1}{\ell} \int_{-\ell}^{\ell} F(t) \sin \frac{n\pi t}{\ell} dt, \qquad n \in \mathbf{N}.$$

Furijeov red parne i neparne funkcije

Neka je funkcija F neprekidna i ima po delovima neprekidan prvi izvod na intervalu $(-\ell, \ell)$.

- Ako je F **parna funkcija**, tada je za sve $n \in \mathbf{N}_0$,

$$F(x) \cos \frac{n\pi x}{\ell} \quad \text{parna funkcija,}$$

$$F(x) \sin \frac{n\pi x}{\ell} \quad \text{neparna funkcija}$$

na $(-\ell, \ell)$.

Furijeovi koeficijenti parne funkcije F su

$$A_n = \frac{1}{\ell} \int_{-\ell}^{\ell} F(x) \cos \frac{n\pi x}{\ell} dx = \frac{2}{\ell} \int_{0}^{\ell} F(x) \cos \frac{n\pi x}{\ell} dx, \qquad n \in \mathbf{N}_0,$$

$$B_n = \frac{1}{\ell} \int_{-\ell}^{\ell} F(x) \sin \frac{n\pi x}{\ell} dx = 0, \qquad n \in \mathbf{N}.$$

Ako je funkcija F neprekidna, ima po delovima neprekidan prvi izvod i **parna** je na $(-\ell, \ell)$, tada je njen Furijeov red oblika

$$F(x)=\frac{A_0}{2}+\sum_{n=1}^{\infty}A_n\cos\frac{n\pi x}{\ell},\quad x\in(-\ell,\ell)$$

i zove se **kosinusni red**.

- **Ako je** F **neparna funkcija**, tada je za sve $n\in\mathbf{N}_0$,

proizvod $F(x)\cos nx$ neparna funkcija,

proizvod $F(x)\sin nx$ parna funkcija

na $(-\ell,\ell)$.

Furijeovi koeficijenti neparne funkcije F su

$$A_n=\frac{1}{\ell}\int_{-\ell}^{\ell}F(x)\cos\frac{n\pi x}{\ell}dx=0,\ \ n\in\mathbf{N}_0,$$

$$B_n=\frac{1}{\ell}\int_{-\ell}^{\ell}F(x)\sin\frac{n\pi x}{\ell}dx=\frac{2}{\ell}\int_{0}^{\ell}F(x)\sin\frac{n\pi x}{\ell}dx,\qquad n\in\mathbf{N}.$$

Ako je funkcija F neprekidna, ima po delovima neprekidan prvi izvod i **neparna** je na $(-\ell,\ell)$, tada je njen Furijeov red

$$F(x)=\sum_{n=1}^{\infty}B_n\sin\frac{n\pi x}{\ell},\qquad x\in(-\ell,\ell)$$

i zove se **sinusni red**.

Vizualizcija Furijeovog reda

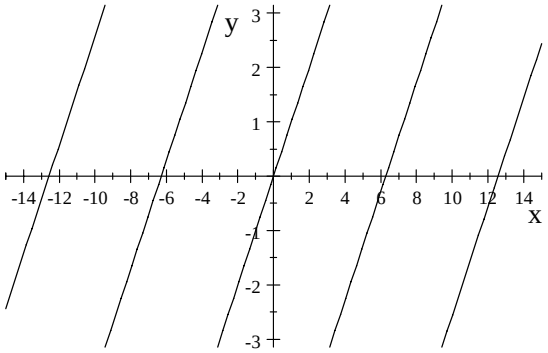
U sledećim primerima koristićemo programske pakete *Scientific WorkPlace* i *GeoGebra* za crtanje grafika po delovima neprekidne periodične funkcije, određivanje koeficijena Furievog reda, parcijalnih suma datog reda. Dinamički će se pomoću programskog paketa *GeoGebra* prikazati kako grafici niza parcijalnih suma reda se približavaju grafiku funkcije.

Primer 1: Odrediti Furijeov red periodične funkcije f sa periodom 2ℓ, za koju je $f(x)=x,\quad x\in(-\ell,\ell)$.

Funkciju f na intervalu $(-5\pi,6\pi)$. možemo zapisati kao

$$f(x)=\begin{cases} x+4\pi & \text{if} \quad -5\pi \le x < -3\pi \\ x+2\pi & \text{if} \quad -3\pi \le x < -\pi \\ x & \text{if} \quad -\pi \le x < \pi \\ x-2\pi & \text{if} \quad \pi \le x < 3\pi \\ x-4\pi & \text{if} \quad 3\pi \le x < 5\pi \\ x-6\pi & \text{if} \quad 5\pi \le x < 7\pi \end{cases}$$

a njen grafik je dat na slici 1.



Slika 1.

Funkcija f je neparna na **R**, pa su svi koeficijenti $A_n = 0, \quad n \in \mathbf{N}_0$, a za koeficijente $B_n, \quad n \in \mathbf{N}$, važi:

$$\begin{aligned} B_n &= \frac{2}{\ell}\int_0^{\ell} x \sin\frac{n\pi x}{\ell}\,dx = \frac{2}{\ell}\left(-x\frac{\ell}{n\pi}\cos\frac{n\pi x}{\ell}\Big|_0^{\ell} + \frac{\ell}{n\pi}\int_0^{\ell}\cos\frac{n\pi x}{\ell}\,dx\right) \\ &= (-1)^{n+1}\frac{2\ell}{n\pi}. \end{aligned}$$

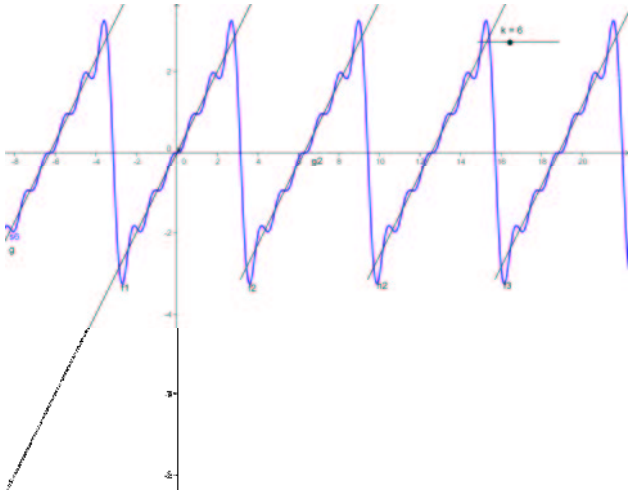
Kako je funkcija f neprekidna i ima po delovima neprekidan prvi izvod na intervalu $(-\ell,\ell)$, važi

$$f(x)=\frac{2\ell}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}\sin\frac{n\pi x}{\ell}, \qquad x \in (-\ell,\ell).$$

U programskom paketu *Geogebra* prikazana je funkcija i druga parcijalna suma Furijeovog reda na intervalu $(-2\pi,6\pi)$.

f(x),PI-x-.ggb (

) i izvezena je slika



Slika 2.

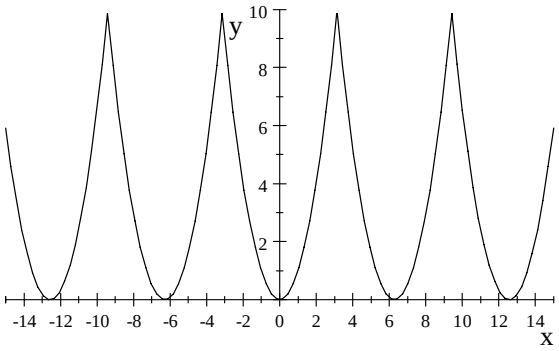
Primer 2: Razviti u Furijeov red periodičnu funkciju f sa periodom 2ℓ, za koju je $f(x)=x^2$, $-\ell \le x \le \ell$. Na osnovu toga pokazati sledeće dve jednakosti:

$$\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^2}=\frac{\pi^2}{12} \qquad \text{i} \qquad \sum_{n=1}^{\infty}\frac{1}{n^2}=\frac{\pi^2}{6}.$$

Funkciju f na intervalu $(-5\pi,6\pi)$. možemo zapisati kao

$$f(x)=\begin{cases}(x+6\pi)^2 & \text{if} \quad -7\pi \le x < -5\pi\\ (x+4\pi)^2 & \text{if} \quad -5\pi \le x < -3\pi\\ (x+2\pi)^2 & \text{if} \quad -3\pi \le x < -\pi\\ x^2 & \text{if} \quad -\pi \le x < \pi\\ (x-2\pi)^2 & \text{if} \quad \pi \le x < 3\pi\\ (x-4\pi)^2 & \text{if} \quad 3\pi \le x < 5\pi\\ (x-6\pi)^2 & \text{if} \quad 5\pi \le x < 7\pi\end{cases}$$

a njen grafik je



Slika 3.

Funkcija f je parna i za sve $n \in \mathbf{N}$ važi $B_n = 0.$ Koeficijenti $A_n,\quad n \in \mathbf{N}_0,$ se odredjuju iz

$$A_0 = \frac{1}{2\ell}\int_0^{\ell} x^2\,dx = \frac{2\ell^2}{3},$$

$$A_n = \frac{2}{\ell}\int_0^{\ell} x^2 \cos\frac{n\pi x}{\ell}\,dx = \frac{(-1)^n 4\ell^2}{n^2\pi^2},\quad n \in \mathbf{N}.$$

Prema tome je

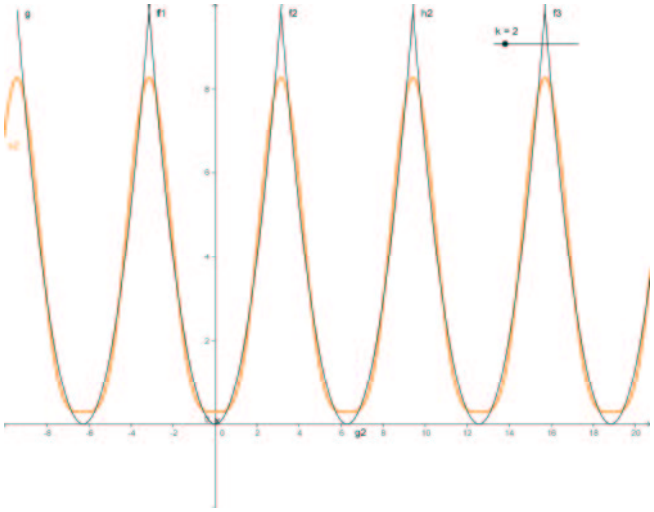
$$f(x) = \frac{\ell^2}{3} + \frac{4\ell^2}{\pi^2}\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\cos\frac{n\pi x}{\ell},\quad x \in \mathbf{R}.$$

Posebno, za $\ell = \pi$ imamo

$$x^2 = \frac{\pi^2}{3} + 4\sum_{n=1}^{\infty}\frac{(-1)^n}{n^2}\cos nx,\quad x \in [-\pi,\pi].$$

U programskom paketu *Geogebra* prikazana je funkcija i druga parcijalna suma Furijeovog reda na intervalu $(-2\pi, 6\pi)$.

f(x),PI-x^2-.ggb (E:\KURS--APRIL\f(x),PI-x^2-.ggb)i izvezena je slika



Slika 4.

Očigledno je

$$\sum_{n=1}^{\infty}\frac{(-1)^{n-1}}{n^2}+\sum_{n=1}^{\infty}\frac{1}{n^2}=2\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}.$$

Na osnovu jednakosti

$$\sum_{n=1}^{\infty}\frac{1}{n^2}=\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}+\sum_{n=1}^{\infty}\frac{1}{(2n)^2}=\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}+\frac{1}{4}\sum_{n=1}^{\infty}\frac{1}{n^2},$$

sledi

$$\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}=\frac{3}{4}\sum_{n=1}^{\infty}\frac{1}{n^2}.$$

Primer 3: Odrediti Furijeov red periodične funkcije sa periodom 2π, za koju je

$$f(x)=\begin{cases}0, & -\pi\le x<0,\\ x, & 0\le x<\pi,\end{cases}$$

i na osnovu toga pokazati jednakost

$$\sum_{n=1}^{\infty}\frac{1}{(2n-1)^2}=\frac{\pi^2}{8}.$$

Funkcija f ima po delovima neprekidan prvi izvod na intervalu $[-\pi,\pi]$, ali nije neprekidna u tačkama $(2k+1)\pi$, $k\in\mathbf{Z}$. Furijeovi koeficijenti funkcije f su redom

$$A_0=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\,dx=\frac{1}{\pi}\int_{0}^{\pi}x\,dx=\frac{\pi}{2},$$

$$B_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(x)\sin nx\,dx=\frac{1}{\pi}\int_{0}^{\pi}x\sin nx\,dx=\frac{(-1)^{n-1}}{n},\quad n\in\mathbf{N}.$$

Furijeov red date funkcije f je

$$f(x)=\frac{\pi}{4}-\sum_{n=1}^{\infty}\frac{2}{\pi}\frac{\cos(2n-1)x}{(2n-1)^2}+\frac{(-1)^n}{n}\sin nx\quad,\ x\in(\pi,\pi)$$

U tački $x=\pi$ je

$$\lim_{x\to\pi+0}f(x)=\lim_{x\to\pi+0}0=0,\ \lim_{x\to\pi-0}f(x)=\lim_{x\to\pi-0}x=\pi$$

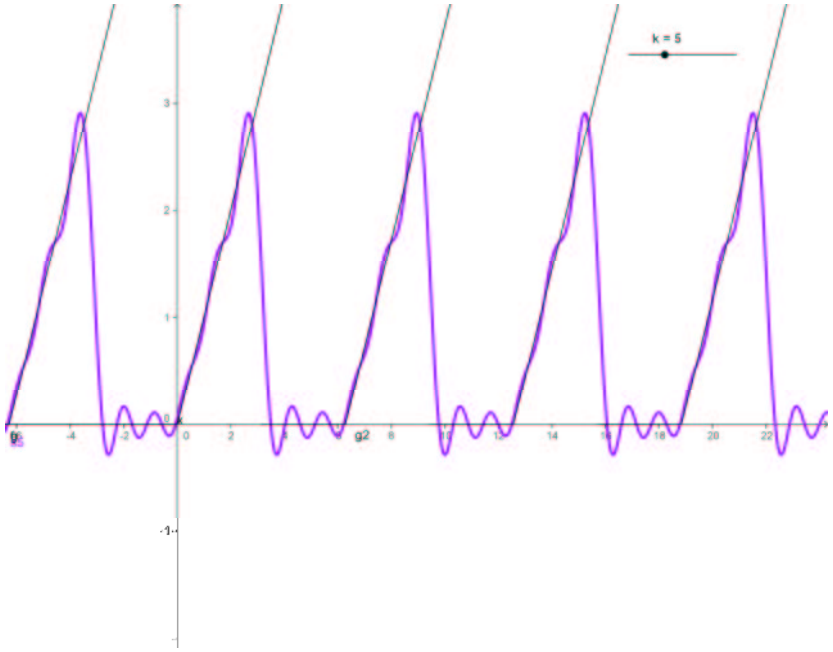
pa je

$$\frac{0+\pi}{2}=\frac{\pi}{4}-\sum_{n=1}^{\infty}\left(\frac{2}{\pi}\frac{\cos(2n-1)\pi}{(2n-1)^2}+\frac{(-1)^n}{n}\sin n\pi\right).$$

Kako je za $n\in\mathbf{N}$, $\cos(2n-1)\pi=-1$ i $\sin n\pi=0,$ tako dobijamo jednakost koju je trebalo pokazati.

U programskom paketu *Geogebra* prikazana je funkcija i druga parcijalna suma Furijeovog reda na intervalu $(-2\pi,6\pi)$.

f(x),PI-0-x.ggb ili ( )i izvezena je slika



Slika 5.

Primer 4: Odrediti Furijeov red periodične funkcije f sa periodom 2π, za koju je $f(x) = |x|, \quad -\pi \le x < \pi$

Data funkcija je neprekidna, ima po delovima neprekidan prvi izvod i parna je na **R**. Svi koeficijenti B_n, $n \in \mathbf{N}$, su jednaki nuli, a koeficijenti A_n, $n \in \mathbf{N}_0$, se određuju na sledeći način:

$$A_0 = \frac{2}{\pi}\int_0^{\pi} |x|\,dx = \frac{2}{\pi}\int_0^{\pi} x\,dx = \pi,$$

a za $n \in \mathbf{N}$ dobijamo korišćenjem parcijalne integracije

$$A_n = \frac{2}{\pi}\int_0^{\pi} |x|\cos nx\,dx = \frac{2}{\pi}\int_0^{\pi} x\cos nx\,dx = \frac{2}{\pi n^2}(\cos n\pi - 1).$$

Kako je $\cos n\pi = (-1)^n$, $n \in \mathbf{N}$, to sledi

$$A_n = \frac{2}{\pi n^2}(-1)^n - 1 = \begin{cases} -\frac{4}{n^2\pi}, & n = 1,3,5,\ldots \\ 0, & n = 2,4,6,\ldots . \end{cases}$$

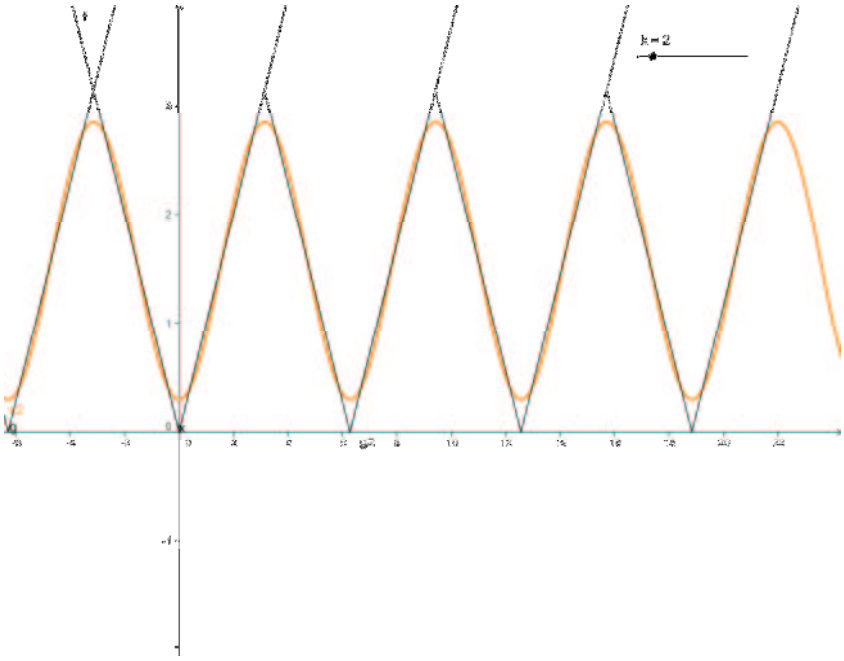
Furijeov red za datu funkciju $f(x) = |x|$ na $[-\pi,\pi]$ je

$$|x| = \frac{\pi}{2} - \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{\cos(2n-1)x}{(2n-1)^2}, \quad x \in [-\pi, \pi].$$

U programskom paketu *Geogebra* prikazana je funkcija i druga parcijalna suma Furijeovog reda na intervalu $(-2\pi, 6\pi)$.

E:\KURS--APRIL\
f(x),PI-abs(x).ggb

f(x),PI-abs(x).ggb ili ( )i izvezena je slika



Slika 6.

Napomena: U programskom paketu *Geogebra* mogu se samo menjati funkcije menjati f funkcije i pomeranjem klizača dobiti odgovarajuće parcijalne sume.

Fitovanje krvih

Đurđica Takači

U ovom delu rada prikazaćemo jednu metodu određivanja funkcije koja najbolje aproksimira podatke date odgovarajućom tabelom korišćenjem programskih paketa *Scientific Workplace* i *Exel.*

U cilju odredjivanja funkcije koje najbolje odgovaraju datim podacima koristi se i metoda najmanijh kvadrata.

Prikazaćemo metodu najmanjih kvadrata za određivanje koeficijena a i b za linearnu funkciju $f(x)=ax+b,$ koja aproksimira podatke date u tabeli $x_i, y_i,\quad i=1,2,...n.$

Ako označimo sa

$$S(a,b) = \sum_{k=1}^{n}(ax_k + b - y_k)^2,$$

funkciju S dve promenljive $a,$ i $b.$

Ekstremna vrednost funkcije dve promenljive se odredjuje iz uslova da je

$$\frac{\partial S}{\partial a} = 0, \qquad \frac{\partial S}{\partial b} = 0,$$

odnosno

$$2\sum_{k=1}^{n}(ax_k + b - y_k)(\sum_{k=1}^{n} x_k) = 0, \qquad 2\sum_{k=1}^{n}(ax_k + b - y_k) = 0$$

Na osnovu prethodnih jednačina dobijamo sistem jednačina

$$a\sum_{k=1}^{n} x_k^2 + b\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} x_k y_k$$

$$a\sum_{k=1}^{n} x_k + nb = \sum_{k=1}^{n} y_k$$

Rešenja prethodnog sistema su

$$a\sum_{k=1}^{n} x_k{}^2 + b\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} x_k y_k$$
$$a\sum_{k=1}^{n} x_k + nb = \sum_{k=1}^{n} y_k$$

odnosno koeficijenti a i b se se dobijaju iz

$$a = \frac{n\sum_{k=1}^{n} x_k y_k - \sum_{k=1}^{n} x_k \sum_{k=1}^{n} y_k}{n\sum_{k=1}^{n} x_k^2 - \left(\sum_{k=1}^{n} x_k\right)^2}$$

$$b = \frac{\sum_{k=1}^{n} x_k \sum_{k=1}^{n} y_k - \sum_{k=1}^{n} x_k \sum_{k=1}^{n} x_k y_k}{n\sum_{k=1}^{n} x_k^2 - \left(\sum_{k=1}^{n} x_k\right)^2}$$

U većini programskih paketa to se odreduje neposredno, kao na primer u Scientific Workplace. Na primer, za datu tabelu odmah dobijamo

1	3
2	6
3	8
5	14
7	21
9	27

, Polynomial fit: $y = 3.0211x - 0.42807$.

Sada ćemo pokazati jednu od metoda kako se jednostavno odredjuje funkcija koja nije linearna i kako se svodi na linearnu funkcija.

Označimo sa $a(x,t)=\frac{x+t}{2}$, $b(x,t)=\sqrt{xt}$, $h(x,t)=\frac{2xt}{x+t}$, aritmetičku, geometrijsku i harmonijsku sredinu, respektino, gde je $x = x_1$ (ili $x = y_1$) prva vrednost u tabeli, a $t = x_n$, (ili $t = y_n$) posledna vrednost u tabeli.

(Pimedba: Vrednosti x i t su uzete zato da bi se mogao konstuisati program za rad Scientific Workplace.)

Označimo sa $a = a(x_1, x_n)$, $b = b(x_1, x_n)$, $h = h(x_1, x_n)$ dobijene brojne vrednosti za x_1, i x_n iz date tabele.

Dalje, za svaku od ovih vrednosti treba odrediti iz tabele vrednosti x_i, x_{i+1}, tako da je

$x_i \le a \le x_{i+1}$, odnosno $x_i \le b \le x_{i+1}$, odnosno $x_i \le h \le x_{i+1}$.
Na osnovu toga se odrede vrednosti funkcija $c, d,$ i g , kao

$$c(x,t) = y_i + \frac{y_{i+1}-y_i}{x_{i+1}-x_i}(a - x_i)$$
$$d(x,t) = y_i + \frac{y_{i+1}-y_i}{x_{i+1}-x_i}(b - x_i)$$
$$g(x,t) = y_i + \frac{y_{i+1}-y_i}{x_{i+1}-x_i}(h - x_i).$$

Zatim se definiše tabela

$$f(x,t,y,z) = \begin{array}{|c|c|c|c|} \hline a(x,t) & a(y,z) & c(x,t) & |a(y,z)-c(x,t)| \\ \hline b(x,t) & b(y,z) & d(x,t) & |b(y,z)-d(x,t)| \\ \hline a(x,t) & b(y,z) & c(x,t) & |b(y,z)-c(x,t)| \\ \hline b(x,t) & a(y,z) & d(x,t) & |a(y,z)-d(x,t)| \\ \hline h(x,t) & a(y,z) & d(x,t) & |a(y,z)-d(x,t)| \\ \hline a(x,t) & h(x,t) & c(x,t) & |h(y,z)-c(x,t)| \\ \hline h(x,t) & h(x,t) & d(x,t) & |h(y,z)-d(x,t)| \\ \hline \end{array}$$

Znači naredbom New Definition se sačuvaju funkcije $a(x,t)$, $b(x,t)$, $h(x,t)$, zatim $c(x,t)$, $d(x,t)$, i $g(x,t)$ sa odredjenim vrednostima x_i, x_{i+1}, y_i, y_{i+1}.

Posmatra se dobijena tabela i odredi najmanja vrednost u poslednjoj koloni. Zatim se odredi kriva koja se nalazi na tom mestu iz tabele:

$x_{sa} = \frac{x_1+x_n}{2}$	$y_{sa} = \frac{y_1+y_n}{2}$	$y = ax + b$	
$x_{sg} = \sqrt{x_1 \cdot x_n}$	$y_{sg} = \sqrt{y_1 \cdot y_n}$	$y = ax^b$	$\log y = b\log x + \log a$
$x_{sa} = \frac{x_1+x_n}{2}$	$y_{sg} = \sqrt{y_1 \cdot y_n}$	$y = ab^x,\ y = ae^{bx}$	$\log y = x\log b + \log a$
$x_{sg} = \sqrt{x_1 \cdot x_n}$	$y_{sa} = \frac{y_1+y_n}{2}$	$y = a\log x + b$	$y = a\log x + b$
$x_{sh} = \frac{2x_1 \cdot x_n}{x_1+x_n}$	$y_{sa} = \frac{y_1+y_n}{2}$	$y = \frac{a}{x} + b$	$y = a\frac{1}{x} + b$
$x_{sa} = \frac{x_1+x_n}{2}$	$y_{sh} = \frac{2y_1 \cdot y_n}{y_1+y_n}$	$y = \frac{1}{ax+b}$	$\frac{1}{y} = ax + b$
$x_{sh} = \frac{2x_1 \cdot x_n}{x_1+x_n}$	$y_{sh} = \frac{2y_1 \cdot y_n}{y_1+y_n}$	$y = \frac{x}{ax+b}$	$\frac{x}{y} = ax + b$

Primer: Datu tabelu

1	2	3	4	5	6
2	17	53	123	247	430

napišemo u obliku i pomoću *Scientific Workplace*

1 2

2 17

3 53

4 123

5 247

6 430

, Polynomial fit: $y = 7.2368x - 3.0198x^2 + 2.3148x^3 - 4.3333$

Definišimo sledeće funkcije:

$$a(x,t) = \frac{x+t}{2}$$
$$b(x,t) = \sqrt{xt}$$
$$h(x,t) = \frac{2xt}{x+t}$$
$$c(x,t) = y_i + \frac{y_{i+1}-y_i}{x_{i+1}-x_i}(a(x,t) - x_i)$$
$$d(x,t) = y_i + \frac{y_{i+1}-y_i}{x_{i+1}-x_i}(b(x,t) - x_i)$$
$$g(x,t) = y_i + \frac{y_{i+1}-y_i}{x_{i+1}-x_i}(h(x,t) - x_i)$$

Za $a(1,6) = 3.5$ iz tabele se čita da je $x_i = 3,\quad x_{i+1} = 4,\quad y_i = 53,\quad y_{i+1} = 123,$

Za $b(1,6) = 2.4495$ iz tabele se čita da je $x_i = 2,\quad x_{i+1} = 3,\quad y_i = 17,\quad y_{i+1} = 53,$

Za $h(1,6) = 1.7143$ iz tabele se čita da je $x_i = 1, \quad x_{i+1} = 2, \quad y_i = 2, \quad y_{i+1} = 17,$ pa je

$$c(x,t) = 123 + 70(a(x,t) - 3)$$
$$d(x,t) = 17 + 36(b(x,t) - 2)$$
$$g(x,t) = 2 + 15(h(x,t) - 1)$$

Znači treba da se sačuva $a(x,t), \quad b(x,t), \quad h(x,t), \quad c(x,t), \quad d(x,t), \quad g(x,t)$ i $f(x,t,y,z)$ i posle naredbe Evaluate se dobija

$f(1,6,2,430) =$

$\frac{7}{2}$	216	158	58
$\sqrt{6}$	$2\sqrt{215}$	$36\sqrt{6} - 55$	$36\sqrt{6} - 2\sqrt{215} - 55$
$\frac{7}{2}$	$2\sqrt{215}$	158	$-2\sqrt{215} + 158$
$\sqrt{6}$	216	$36\sqrt{6} - 55$	$-36\sqrt{6} + 271$
$\frac{12}{7}$	216	$36\sqrt{6} - 55$	$-36\sqrt{6} + 271$
$\frac{7}{2}$	$\frac{12}{7}$	158	$\frac{8317}{54}$
$\frac{12}{7}$	$\frac{12}{7}$	$36\sqrt{6} - 55$	$36\sqrt{6} - \frac{3185}{54}$

$f(1,6,2,430) =$

3.5	216.0	158.0	58.0
2.4495	29.326	33.182	3.8559
3.5	29.326	158.0	128.67
2.4495	216.0	33.182	182.82
1.7143	216.0	33.182	182.82
3.5	1.7143	158.0	154.02
1.7143	1.7143	33.182	29.2

Iz poslednje tabele se vidi da je najmanja razlika u dugoj vrsti, što odgovara funkciji $y = ax^b$.

$$y = ax^b \quad \log y = b\log x + \log a$$

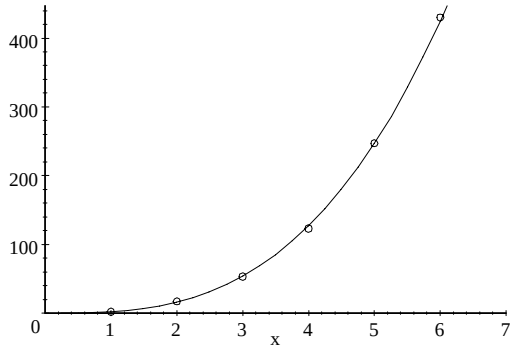
Na osnovu programskog paketa *Exel* ..\exel.xls ili 

dobili smo da je

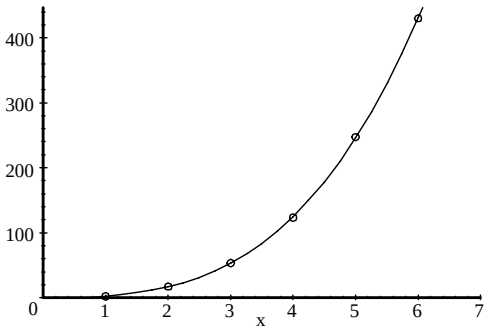
$$b = 2.979294818 \text{ i}$$
$$\log a = 0.309831568,$$

odnosno $a = 10^{0.309831568} = 2.0409.$

Znači prema prikazanoj metodi funkcija $y = 2.0409x^{2.979294818}$ najbolje aproksimira podatke date u tabeli. Na slici 1 je prikazan grafik funkcije dobijen preka programskog paketa *Scentific Workplace* prikazani grafici, a na slici 2 je prikazan grafik funkcije $y = 2.0409x^{2.979294818}$, što pokazuje da odabrana funkcija koja najbolje odgovara datoj tabeli nje jednozna;no odredjena.



Slika 1.



Slika2.

Literatura:

[1] Schmeelk, J., Takači, Dj., Takači, A., *Elementary Analysis through Examples and Exercises*, Kluwer Academic Publishers, Dordrecht/Boston/London, 1995.

[2] Kadijević, Dj., *Promoting P-C links by CAS: present state, limitations and improvements,* In J. B¨ohm (Ed.), VISIT-ME-2002, Proceedings of the Vienna International Symposium on Integrating Technology in Mathematics Education (CD), Hagenberg, Austria: bk teachware, 2002.

[3] Takači, Dj., Pešić, D., *The Continuity of Functions in Mathematical Education Visualization method*, in Serbian, Nastava matematike (Mathematics Teaching), 49, 3-4, Belgrade, 2004.

[4] Takači, Dj., Pešić, D., Tatar, J., *An introduction to the Continuity of functions using Scientific Workplace*, The Teaching of Mathematics, Vol. VI,2, Belgrade, pp. 105-112, 2003.

[5] Takači, Dj., Pešić, D., Tatar, J., *On the continuity of functions*, International Journal of Mathematical Education in Science and Technology, Taylor & Frencis, Vol. 37, No. 7, (15 October 2006), 783-791.

[6] Takači, Dj., Pešic, D., *The teaching of continuity of functions - visualization method*

(in Serbian), Nastava Matematike, Belgrade, 49, 3-4 (2003), 31-42.
[7] Takači, Đ., Herceg D., Stojković R., Possibilities and limitati-ons of Scientific Workplace in studying trigonometric functions, The Teaching of Mathematics, VIII_2 / 2006. 61-72.
[8] Takači, Đ., Samardžijević M., Vizualni pristup definiciji izvoda funkcije, Nastava matematike, LI_1-2 / 2006 str. 19-28 Beograd.
[9] Takači, Đ., Radovanović, J., The examining functions and computer, Zbornik radova na CD, Internacionalna Konferencija nastave matematematike, ICTM 3, 2006 Istambul.
[10] Takači, Đ., Herceg, D., Stojković R., *Possibilities and limitation of Scientific WorkPlace in studing trigonometric functions*, The Teaching of mathematics, Vol. VIII, 2, pp. 61—72, 2005.
[11] Tall, D., *The Transition to Advanced Mathematical Thinking: Functions, Limits, Infinity, and Proof*, in Grouws, D. A., Handbook of Research on Mathematics Teaching and Learning, Macmillan, New York, 1991, 495-511.
[12] Tall, D., *Resent Developements in the Use of Computer to Visualize and Symbolize Calculus Concepts*, The Laboratory Approach to Teaching Calculus, M.A.A. Notes, Vol. 20 (1991), 15-25.
[13] Tall, D., Vinner, A., *Concept Image and Concept Definitionin Mathematics with particular reference to Limits and Continuity*, Education Studies in Mathematics, 12 (1981), 159-169..

Project: 06SER02/02/003

Chemical Informatics

Dr Dragan Mašulović

Chapter 1

Introduction

Perhaps the best way to introduce the topic we shall be considering in this series of lectures is to quote a part of the Preface of [2]:

"One of the great benefits that mankind has from from chemistry is the possibility to synthesize substances capable of curing deseases and reducing the suffering of the ill. If we wished to synthesize a new chemical compound with more desirable properties in comparison to the compounds already known, the standard procedure would be to identify and test candidate compounds. Until recently, the standard procedure was to synthetize and test such compounds one by one and then test for the desired properties, and absence of undesirable ones. Such experimental tests are very expensive and time consuming.

"If we have an idea about the structural requirements for the compound we are looking for (and usually we do), we can produce a combinatorial library consisting of structural formulas of candidate compounds and analyze virtual compounds by means of fast algorithms. This inexpensive and non-laboratory-experiment based approach can scan a much greater range of candidate compounds and thus reduce the vast number of possibilities to a practically feasible few, which can then be tested using standard procedures."

In this series of lectures we shall first introduce the mathematical apparatus for describing and analyzing structures. We shall consider the standard numerical characteristics associated to structures such as valency and distance. We shall then talk about representations of structures with computer implementation in mind and present some simple standard algorithms for analyzing structures.

Special atention will be devoted to acyclic structures, representations of acyclic structures (again with computer implementation in mind) and Prüfer codes in particular. We shall cover the notion of a spanning tree and as an application, we discuss monocyclic structures.

We then move on to more advanced properties of structures. We shall first discuss Kekulé structures and counting Kekulé structures. After considering some particular cases we shall derive the well known general-case formula: $K(G) = K(G-e) + K(G-(e))$.

These considerations are followed by the formal treatment of symmetry in structures. We first discuss counting all vs counting nonisomorphic structures. The main tool for counting

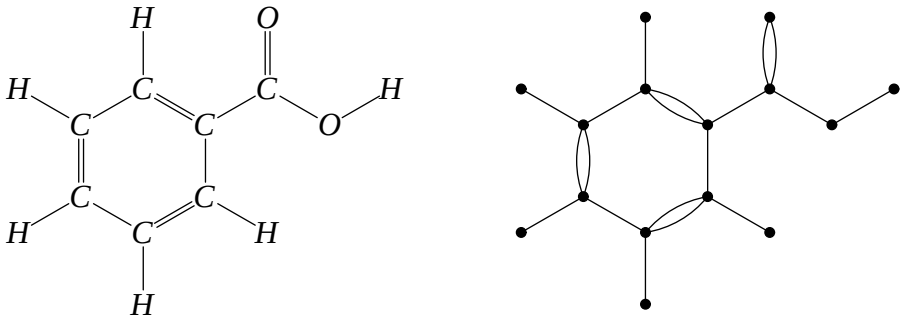
nonisomorphic symmetric structures is the Cauchy-Frobenius Lemma which we derive in the most general case, and then apply to some special cases.

As case study we discuss the probelm of generating and enumerating hexagonal systems, where all of the ideas from the previous chapters are used to demonstrate an efficient algorithm for classifying geometrically planar, simply connected polyhexes.

Chapter 2

Graphs as Models of Structures

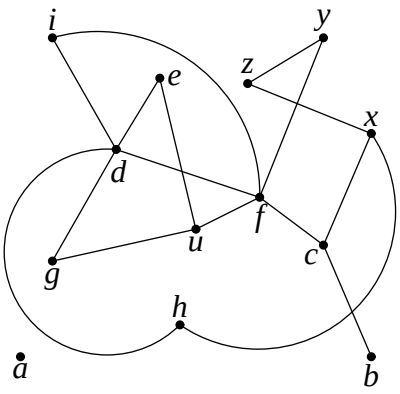
Graphs represent one of the most popular tools for modeling discrete phenomena where the abstraction of the problem involves information about certain objects being connected or not. For example, crossings in a city transportation model are joined by streets, or cities in a country are joined by roads. Or a structural formula of a chemical compound, eg. benzoic acid $C_7H_6O_2$:



2.1 Graphs

A *graph* is an ordered pair $G = (V,E)$ where V is a nonempty finite set and E is an arbitrary subset of $\binom{V}{2} = \big\{\{u,v\} \subseteq V : u \neq v\big\}$. Elements of V are called *vertices* of G, while elements of E are called *edges* of G. We shall often write $V(G)$ and $E(G)$ to denote the set of vertices and the set of edges of G, and $n(G)$ and $m(G)$ to denote the number of vertices and the number of edges of G.

The graphs are called graphs because of a very natural graphical representation they have. Vertices are usually represented as (somewhat larger) points in a plane, while edges are represented as (smooth non-selfintersecting) curves joining the respective vertices, so that adjacent vertices are joined by a curve.



$\delta(a) = 0$ (an isolated vertex)
$\delta(b) = 1$ (a leaf)
$\delta(c) = 3$
$\delta(d) = 5$
$\delta(e) = 2$
$\delta(f) = 5$
$\delta(g) = 2$
$\delta(h) = 2$
$\delta(i) = 2$
$\delta(u) = 3$
$\delta(x) = 3$
$\delta(y) = 2$
$\delta(z) = 2$

$\delta(G) = 0 \quad \Delta(G) = 5$

Figure 2.1: An example of a graph

Example 2.1 Fig. 2.1 depicts the graph $G = (V,E)$, where

$$V = \{a,b,c,d,e,f,g,h,i,u,x,y,z\}$$
$$E = \big\{\{b,c\},\{c,x\},\{c,f\},\{h,x\},\{z,x\},\{f,y\},\{y,z\},\{f,u\},\{f,d\},\{f,i\},$$
$$\{e,u\},\{g,u\},\{d,e\},\{d,g\},\{d,h\},\{d,i\}\big\}$$
$$n = 13$$
$$m = 16.$$

If $e = \{u,v\}$ is an edge of a graph, we say that u and v are *adjacent*, and that e is *incident* with u and v. We also say that u is a *neighbour* of v. The *degree of a vertex* v, denoted by $\delta_G(v)$, is the number of edges incident to v. If G is clear from the context, we simply write $\delta(v)$. By $\delta(G)$ we denote the least, and by $\Delta(G)$ the greatest degree of a vertex in G. A vertex with degree 0 is said to be an *isolated vertex*. A vertex of degree 1 is called a *leaf of* G. A vertex is said to be *even*, resp. *odd* according as $\delta(v)$ is an even or an odd integer. A graph is *regular* if $\delta(G) = \Delta(G)$. In other words, in a regular graph all vertices have the same degree. See Fig. 2.1.

Theorem 2.2 (The First Theorem of Graph Theory) *If $G = (V,E)$ is a graph with m edges, then $\sum_{v \in V} \delta(v) = 2m$.*

Proof. Since every edge is incident to two vertices, every edge is counted twice in the sum on the left. □

Corollary 2.3 *In any graph the number of odd vertices is even.*

Theorem 2.4 *If $n(G) \geqslant 2$, there exist vertices $v,w \in V(G)$ such that $v \neq w$ and $\delta(v) = \delta(w)$.*

Proof. Let $V(G) = \{v_1, \ldots, v_n\}$ and suppose that $\delta(v_i) \neq \delta(v_j)$ whenever $i \neq j$. Without loss of generality we may assume that $\delta(v_1) < \delta(v_2) < \ldots < \delta(v_n)$. Since there are only n possibilities for the degree of a vertex (0, 1, ..., $n-1$) it follows that $\delta(v_1) = 0$, $\delta(v_2) = 1$, ..., $\delta(v_n) = n-1$. But then v_n is adjacent to every other vertex of a graph, including the isolated vertex v_1. Contradiction. □

A graph $H = (W, E')$ is a *subgraph* of a graph $G = (V, E)$, in symbols $H \leqslant G$, if $W \subseteq V$ and $E' \subseteq E$. A subgraph H of G is a *spanning subgraph* if $W = V(G)$. A subgraph H is an *induced subgraph* of G if $E' = E \cap \binom{W}{2}$. Induced subgraphs are usually denoted by $G[W]$. The edges of an induced subgraph of G are all the edges of G whose both ends are in W. A set of vertices $W \subseteq V(G)$ is *independent* if $E(G[W]) = \varnothing$, i.e. no two vertices in W are adjacent in G. By $\alpha(G)$ we denote the maximum carinality of an independent set of vertices in G. If $A, B \subseteq E(G)$ are disjoint, by $E(A, B)$ we denote the set of all edges in G whose one end is in A and the other in B.

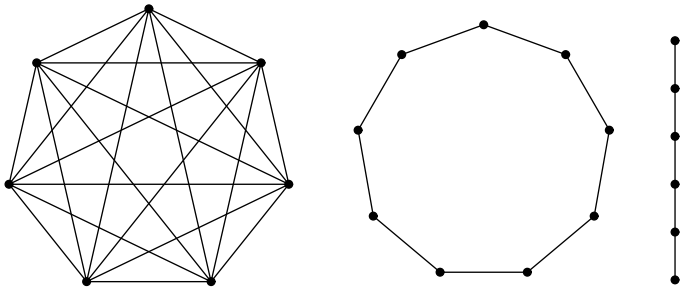


Figure 2.2: K_7, C_9 and P_6

A *complete graph on n vertices* (or an *n-clique*) is a graph with n vertices where each two distinct vertices are adjacent. A complete graph on n vertices is denoted by K_n. A *cycle* of length n, denoted by C_n, is the graph with n vertices where the first vertex is adjacent to the second one, and the second vertex to the third one, and so on, the last vertex is adjacent to the first. A *path with n vertices*, denoted by P_n, is a graph where the first vertex is adjacent to the second one, and the second vertex to the third one, and so on, and the penultimate vertex is adjacent to the last one, but the last vertex is *not* adjacent to the first. We say that the path with n vertices has length $n-1$. Fig. 2.2 depicts K_7, C_9 and P_6.

Theorem 2.5 *If $\delta(G) \geqslant 2$ then G contains a cycle.*

Proof. Let $x_1 \ldots x_{k-1}\, x_k$ be the longest path in G. Since $\delta(x_k) \geqslant \delta(G) \geqslant 2$, x_k has a neighbour v distinct from x_{k-1}. If $v \notin \{x_1, \ldots, x_{k-2}\}$ then $x_1 \ldots x_{k-1}\, x_k\, v$ is a path with more vertices than the longest path, which is impossible. Therefore, $v = x_j$ for some $j \in \{1, \ldots, k-2\}$ so $x_j \ldots x_k$ are vertices of a cycle in G. □

Graphs G_1 and G_2 are *isomorhic*, and we write $G_1 \cong G_2$, if there is a bijection $\varphi : V(G_1) \to V(G_2)$ such that $\{x, y\} \in E(G_1) \Leftrightarrow \{\varphi(x), \varphi(y)\} \in E(G_2)$. For example graphs G and G_2 in Fig. 2.3 are isomorphic, while G and G_1 are not.

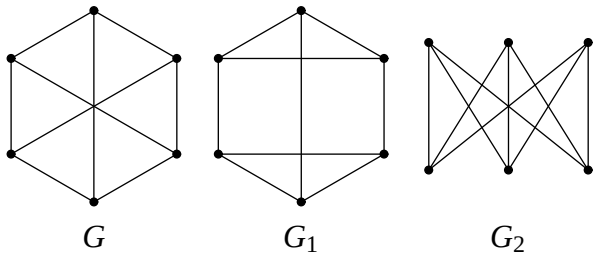


Figure 2.3: $G \cong G_2$, but $G \ncong G_1$

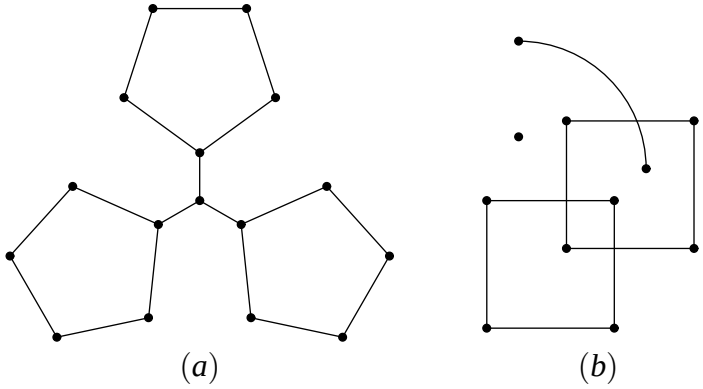


Figure 2.4: (a) A connected graph; (b) A graph with $\omega = 4$

Theorem 2.6 *Let $G_1 \cong G_2$ and let φ be an isomorphism between G_1 and G_2. Then $n(G_1) = n(G_2)$, $m(G_1) = m(G_2)$ and $\delta_{G_1}(x) = \delta_{G_2}(\varphi(x))$ for every $x \in V(G_1)$.*

2.2 Connectedness

A *walk* in a graph G is any sequence of vertices and edges $v_0\ e_1\ v_1\ e_2\ v_2 \ldots v_{k-1}\ e_k\ v_k$ such that $e_i = \{v_{i-1}, v_i\}$ for all $i \in \{1, \ldots, k\}$. Note that an edge or a vertex may appear more than once in a walk. We say that k is the *length* of the walk. If $v_0 \neq v_k$ we say that the *walk connects v_0 and v_k*. A *closed walk* is a walk $v_0\ e_1\ v_1\ \ldots\ v_{k-1}\ e_k\ v_k$ where $v_0 = v_k$. Clearly, a path is a walk where neither vertices nor edges are allowed to repeat, and a cycle is a closed walk where neither edges nor vertices are allowed to repeat, except for the first and the last vertex.

We define a binary relation θ on $V(G)$ by $x\theta y$ if $x = y$ or there is a walk that connects x and y. Clearly, θ is an equivalence relation on $V(G)$ and hence partitions $V(K)$ into blocks $S_1, \ldots, S_t$. These blocks or the corresponding induced subgraphs (depending on the context) are called *connected components* of G. The number of connected components of G is denoted by $\omega(G)$. A graph G is *connected* if $\omega(G) = 1$. An example of a connected graph and of a graph with four connected components are given in Fig. 2.4.

Lemma 2.7 *$S \subseteq V(G)$ is a connected component of G if and only if no proper superset $S' \supset S$ induces a connected subgraph of G.*

Theorem 2.8 *A graph G is connected if and only if $E(A,B) \neq \varnothing$ for every partition $\{A,B\}$ of $V(G)$.*

Proof. ($\Rightarrow$) Let G be a connected graph and $\{A,B\}$ a partition of $V(G)$. Take any $a \in A$ and $b \in B$. Now G is connected, so there is a path $x_1 \ldots x_k$ that connects a and b. Since $x_1 = a$ and $x_k = b$, there is a j such that $x_j \in A$ and $x_{j+1} \in B$ whence $E(A,B) \neq \varnothing$.

($\Leftarrow$) Suppose G is not connected and let S_1, ..., S_ω be the connected components. Then Lemma 2.7 yields $E(S_1, \bigcup_{j=2}^{\omega} S_j) = \varnothing$. □

The *distance* $d_G(x,y)$ between vertices x and y of a connected graph G is defined by $d_G(x,x) = 0$, and in case $x \neq y$,

$$d_G(x,y) = \min\{k : \text{ there is a path of length } k \text{ that connects } x \text{ and } y\}.$$

Theorem 2.9 *Let $G = (V,E)$ be a connected graph. Then (V, d_G) is a metric space, i.e. for all $x,y,z \in V$ the following holds:*

(D1) $d_G(x,y) \geqslant 0$;

(D2) $d_G(x,y) = 0$ *if and only if* $x = y$;

(D3) $d_G(x,y) = d_G(y,x)$; *and*

(D4) $d_G(x,z) \leqslant d_G(x,y) + d_G(y,z)$.

If G is obvious, instead of $d_G(x,y)$ we simply write $d(x,y)$. The *diameter* $d(G)$ of a connected graph G is the maximum distance between two of its vertices:

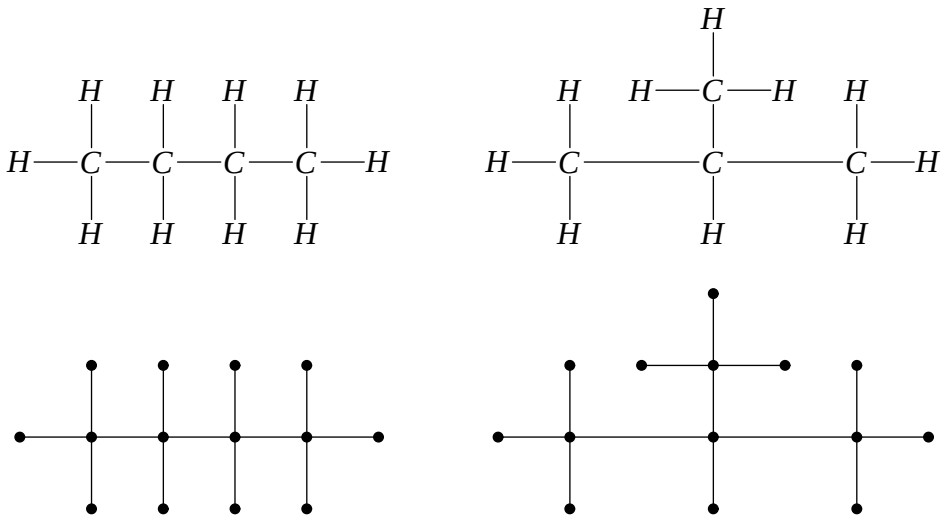
$$d(G) = \max\{d(x,y) : x,y \in V(G)\}.$$

Example 2.10 (a) $d(G) = 1$ if and only if G is a complete graph.

(b) $d(P_n) = n-1$ and $d(C_n) = \lfloor \dfrac{n-1}{2} \rfloor$.

2.3 Trees and monocyclic structures

Historically, the first treatment of graphs as models of chemical compounds appeared in 1889 when Arthur Cayley was engaged in counting isomers of hydrocarbons. A mathematical model of a hydrocarbon is called a *tree* for obvious reasons:



Formally, a *tree* is a connected graph with no cycles. It requires a little bit of thinking to see that a tree is a *minimal connected graph with the given set of vertices*. The following theorem shows that in a way trees capture the essence of the property of being connected. Recall that a spanning subgraph of a graph $G = (V,E)$ is a graph $H = (W,E')$ such that $W = V$ and $E' \subseteq E$. If H is a tree, we say that H is a *spanning tree of* G.

Theorem 2.11 *A graph is connected if and only if it has a spanning tree.*

Proof. Clearly, if a graph G contains a connected spanning subraph H then G is also connected. Therefore if a graph has a spanning tree, it is connected. For the converse, take any connected graph G and construct a sequence of graphs $G_0, G_1, G_2, \ldots$ as follows: $G_0 = G$; if G_i has a cycle, take any edge e_i that lies on a cycle and let $G_{i+1} = G_i - e_i$, otherwise put $G_{i+1} = G_i$. Each G_i is a spanning subgraph of G and each G_i is connected since by removing an edge that lies on a cycle we cannot turn a connected graph into a disconnected one. Moreover, if $G_i = G_{i+1}$ then $G_i = G_j$ for all $j > i$. Let m be the number of edges of G. Since we cannot remove more than m edges from G, we conclude that $G_{m+1} = G_{m+2}$. By construction of the sequence this means that G_{m+1} has no cycles. Therefore, G_{m+1} is a spanning tree of G. □

We will now show that each tree with n vertices has $n-1$ edges and that each two of the three properties listed below implies the remaining one:

- being connected,
- having no cycles, and
- $m = n-1$.

Lemma 2.12 *Each tree with at least two vertices has at least two leaves.*

Proof. Let G be a tree with $n \geqslant 2$ vertices and let $v_1, v_2, \ldots, v_k$ be the longest path in the tree. Then $k \geqslant 2$ since G is a connected graph with at least two vertices. If $\delta(v_1) > 1$ then v_1 has a

neighbour x distinct from v_2. If x is a new vertex, i.e. $x \notin \{v_3, \ldots, v_k\}$, then the path x, v_1, v_2, …, v_k is longer than the longest path in G, which is impossible. If, however, $x \in \{v_3, \ldots, v_k\}$ then G has a cycle, which contradicts the assumption that G is a tree. Therefore, v_1 is a leaf. The same argument shows that v_k is another leaf. □

Theorem 2.13 *Let $G = (V,E)$ be a tree with n vertices and m edges. Then $m = n-1$, and consequently $\sum_{v \in V} \delta(v) = 2(n-1)$.*

Proof. The second part of the theorem follows from the First Theorem of Graph Theory, so let us show that $m = n-1$. The proof is by induction on n. The cases $n = 1$ and $n = 2$ are trivial. Assume that the statement is true for all trees with less then n vertices and consider a tree G with n vertices. By Lemma 2.12 there is a leaf x in G and it is not a cut-vertex. Hence, $G - x$ is connected. Clearly, $G - x$ does not have cycles (removing vertices and edges cannot introduce cycles), so $G - x$ is a tree with less than n vertices. By the induction hypothesis, $m' = n' - 1$, where $m' = m(G-x)$ and $n' = n(G-x)$. But $m' = m-1$ and $n' = n-1$ since x is a leaf, whence $m = n-1$. □

Theorem 2.14 *Let G be a graph with n vertices and m edges. If $m = n-1$ and G has no cycles then G is connected (hence a tree).*

Proof. Suppose that $m = n-1$ and that G has no cycles. Let S_1, …, S_ω be the connected components of G. Each connected component is a tree, so $m_i = n_i - 1$ for all i, where $m_i = m(S_i)$ and $n_i = n(S_i)$. Therefore $\sum_{i=1}^{\omega} m_i = \sum_{i=1}^{\omega} n_i - \omega$ i.e. $m = n - \omega$ (since $m = \sum_{i=1}^{\omega} m_i$ and $n = \sum_{i=1}^{\omega} n_i$). Now, $m = n-1$ yields $\omega = 1$, i.e. G is connected. □

Theorem 2.15 *Let G be a connected graph with $n \geqslant 2$ vertices and m edges and let $m = n-1$. Then G has no cycles (and hence it is a tree).*

Proof. According to Theorem 2.11 the graph $G = (V,E)$ has a spanning tree $H = (V,E')$. Since H is a tree Theorem 2.13 yields $m(H) = n(H) - 1 = n-1$. Assumption $m = n-1$ now implies $m(H) = m$ and thus from $E' \subseteq E$ we conclude $E' = E$. Therefore, $G = H$ and so G is a tree. □

Corollary 2.16 *A connected graph with n vertices and m edges is a tree if and only if $m = n-1$.*

We shall conclude the section by a result on the number of distinct trees. Let us first note that when counting structures we can count distinct structures and non-isomorphic structures. For example, there are 16 distinct trees on a four element set, but only two nonisomorphic, see Fig. 2.5. It is not surprising that counting nonisomorphic structures is more difficult.

Theorem 2.17 (Cayley 1889) *There are n^{n-2} distinct trees with n vertices.*

Proof. Let $V = \{1, \ldots, n\}$ be a finite set that serves as a set of vertices. The proof we are going to present is due to H. Prüfer[1]. The idea is to encode each tree on V by a sequence of integers $(a_1, \ldots, a_{n-2})$ and thus provide a bijection $\varphi : \mathscr{T}_n \to \{1,2,\ldots,n\}^{n-2}$, where $\mathscr{T}_n$ denotes the set of all trees on V.

[1] H. Prüfer, *Neuer Beweis eines Satzes über Permutationen*, Archiv der Math. und Phys. (3) 27(1918), 142–144

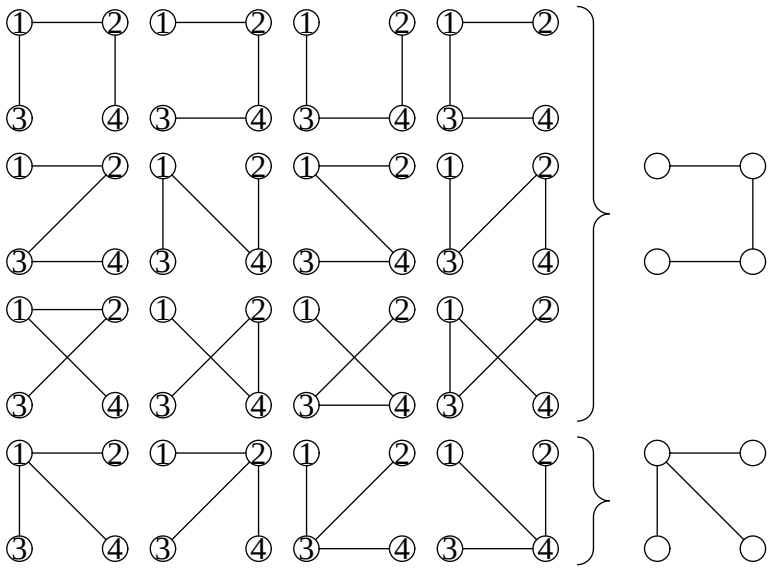


Figure 2.5: Sixteen distinct and only two nonisomorphic trees with four vertices

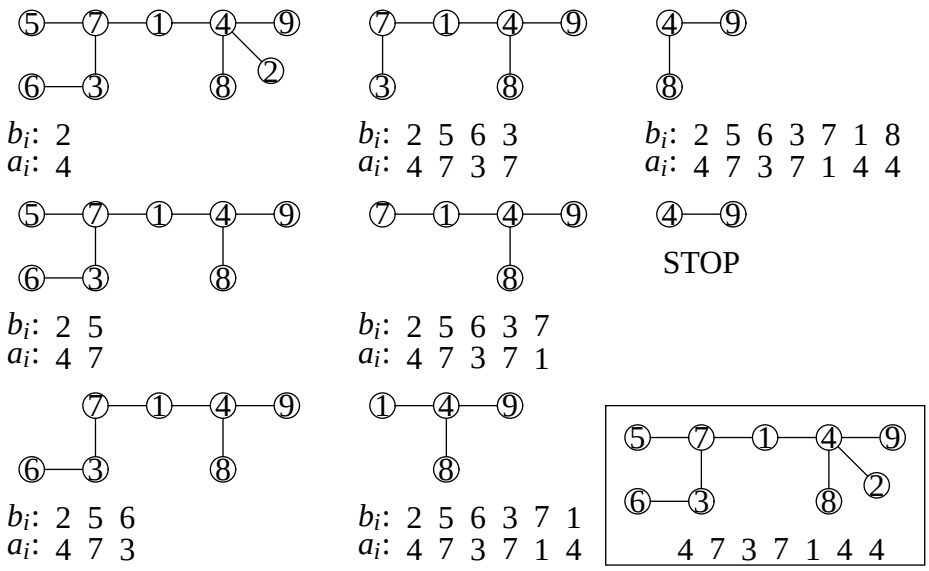


Figure 2.6: The Prüfer code of a tree

We first show how to construct the Prüfer code of a tree. Let T be a tree with the set of vertices V. We shall construct a sequence of trees (T_i) and two sequences of integers, the code (a_i) and an auxiliary sequence (b_i). Let $T_1 = T$. Given T_i, let b_i be the smallest leaf of the tree (vertices are integers, so out of all integers that appear as leaves we choose the smallest) and let a_i be its only neighbour. Now put $T_{i+1} = T_i - b_i$ and repeat until a tree with two vertices is obtained. The code of the tree is now $(a_1, a_2, \ldots, a_{n-2})$. An example is given in Fig. 2.6. Thus, we have a function $\varphi : \mathscr{T}_n \to \{1,\ldots,n\}^{n-2}$ that takes a tree onto its Prüfer code.

Conversely, given a sequence $(a_1,\ldots,a_{n-2})$ we can construct the tree as follows. For $S \subseteq \{1,\ldots,n\}$ let $\operatorname{mix} S = \min(\{1,\ldots,n\} \setminus S)$ denote the minimal number not in S (*mi*nimal e*x*cluded). Put $a_{n-1} = n$ and then construct b_1, b_2, …, b_{n-1} by

$$b_i = \operatorname{mix}\{a_i,\ldots,a_{n-1},b_1,\ldots,b_{i-1}\}$$

(for $i = 1$ there are no b_j's in the set). For example in case of $(4,7,3,4,1,4,4)$ we have $a_8 = 9$ and:

$$\begin{aligned}
b_1 &= \operatorname{mix}\{4,7,3,4,1,4,4,9\} = 2\\
b_2 &= \operatorname{mix}\{\ \ 7,3,4,1,4,4,9,2\} = 5\\
b_3 &= \operatorname{mix}\{\ \ \ \ 3,4,1,4,4,9,2,5\} = 6\\
b_4 &= \operatorname{mix}\{\ \ \ \ \ \ 4,1,4,4,9,2,5,6\} = 3\\
b_5 &= \operatorname{mix}\{\ \ \ \ \ \ \ \ 1,4,4,9,2,5,6,3\} = 7\\
b_6 &= \operatorname{mix}\{\ \ \ \ \ \ \ \ \ \ 4,4,9,2,5,6,3,7\} = 1\\
b_7 &= \operatorname{mix}\{\ \ \ \ \ \ \ \ \ \ \ \ 4,9,2,5,6,3,7,1\} = 8\\
b_8 &= \operatorname{mix}\{\ \ \ \ \ \ \ \ \ \ \ \ \ \ 9,2,5,6,3,7,1,8\} = 4
\end{aligned}$$

This process is called the *reconstruction procedure* since, as we shall see, it produces a tree whose Prüfer code is $(a_1,\ldots,a_{n-2})$.

Let us show that $\{\{b_i,a_i\} : 1 \leqslant i \leqslant n\}$ is the set of edges of a tree. If $i < j$ then, by construction, $b_j = \operatorname{mix}\{a_j,\ldots,a_{n-1},b_1,\ldots,b_i,\ldots,b_{j-1}\}$, so $b_j \neq b_i$. We see that all b_i's are distinct and smaller than $n = a_{n-1}$. Therefore, $\{b_1,\ldots,b_{n-1}\} = \{1,\ldots,n-1\}$ and hence $\{b_1,\ldots,b_{n-1}, a_{n-1}\} = \{1,\ldots,n-1,n\}$. Moreover, if $i \leqslant j$ then $a_j \notin \{b_1,\ldots,b_j\}$ since $b_i = \operatorname{mix}\{a_i, \ldots, a_j, \ldots, a_{n-1}, b_1, \ldots, b_{i-1}\}$, so from $\{b_1,\ldots,b_{n-1},a_{n-1}\} = \{1,\ldots,n-1,n\}$ it follows that $a_j \in \{b_{j+1},\ldots,b_{n-1},a_{n-1}\}$. To summarize,

$$\begin{aligned}
&a_j \in \{b_{j+1},b_{j+2},\ldots,b_{n-1},a_{n-1}\} \text{ and}\\
&b_j \notin \{a_{j+1},b_{j+1},a_{j+2},b_{j+2},\ldots,a_{n-1},b_{n-1}\},
\end{aligned} \qquad \text{for all } j. \tag{$\star$}$$

To build the graph we start from $\{b_{n-1},a_{n-1}\}$ and then add edges $\{b_{n-2},a_{n-2}\}$, $\{b_{n-3},a_{n-3}\}$, …, $\{b_1,a_1\}$ one by one. From $(\star)$ it follows that at each step we extend the graph by one new vertex b_i and one new edge $\{b_i,a_i\}$ that connects the new vertex to an existing one. Therefore, the graph we obtain at the end is connected, and a connected graph with n vertices and $n-1$ edges has to be a tree (Corollary 2.16). Thus, we have a function $\psi : \{1,\ldots,n\}^{n-2} \to \mathscr{T}_n$ that takes a code and produces a tree.

To complete the proof, we have to show that φ and ψ are inverses of one another, i.e. $\varphi \circ \psi = \mathrm{id}$ and $\psi \circ \varphi = \mathrm{id}$. We show only $\psi \circ \varphi = \mathrm{id}$ i.e. $\psi(\varphi(T)) = T$ for all $T \in \mathscr{T}_n$ (the other equality is left as an exercise). For a tree T, a vertex $v \in V(T)$ is an *internal vertex* T if $\delta_T(v) > 1$. Let $\mathrm{int}(T)$ denote the set of all internal vertices of T.

Take any $T \in \mathscr{T}_n$, let $(a_1, \ldots, a_{n-2})$ be its Prüfer code and $(b_1, \ldots, b_{n-2})$ the auxiliary sequence. At the end of the procedure of constructing the Prüfer code two vertices remain the the graph, the vertex $a_{n-1} = n$ and its neighbour whom we denote by b_{n-1}. Starting from $(a_1, \ldots, a_{n-1})$ the reconstruction procedure produces a sequence of integers b'_1, ..., b'_{n-1}. We will show that $b_i = b'_i$ for all i. Assume also that $n \geqslant 3$.

Since b_1 is adjacent to a_1 in T and $n \geqslant 3$, a_1 cannot be a leaf of T so $a_1 \in \mathrm{int}(T)$. The same argument shows that $a_2 \in \mathrm{int}(T - b_1)$, $a_3 \in \mathrm{int}(T - b_1 - b_2)$, and in general, $a_{i+1} \in \mathrm{int}(T - b_1 - \ldots - b_i)$. Since $\mathrm{int}(T - v) \subseteq \mathrm{int}(T)$ whenever v is a leaf of T and $n(T) \geqslant 2$, it follows that $\mathrm{int}(T - b_1 - \ldots - b_i) = \{a_{i+1}, \ldots, a_{n-2}\}$. In particular, $\mathrm{int}(T) = \{a_1, \ldots, a_{n-2}\}$. Since each vertex of a tree with at least two vertices is either a leaf or an internal vertex we obtain that

$$V(T - b_1 - \ldots - b_i) \setminus \mathrm{int}(T - b_1 - \ldots - b_i)$$

is the set of leaves of $T - b_1 - \ldots - b_i$. Now $V(T - b_1 - \ldots - b_i) = \{1, \ldots, n\} \setminus \{b_1, \ldots, b_i\}$ and $\mathrm{int}(T - b_1 - \ldots - b_i) = \{a_{i+1}, \ldots, a_{n-2}\}$, so the set of leaves of $T - b_1 - \ldots - b_i$ is

$$\begin{aligned}\Big(\{1, \ldots, n\} \setminus \{b_1, \ldots, b_i\}\Big) \setminus \{a_{i+1}, \ldots, a_{n-2}\} = \\ = \{1, \ldots, n\} \setminus \{a_{i+1}, \ldots, a_{n-2}, b_1, \ldots, b_i\}.\end{aligned}$$

It is now easy to show that $b_i = b'_i$ by induction on i. As we have seen, b_1 is a leaf of T, so $b_1 \in \{1, \ldots, n\} \setminus \{a_1, \ldots, a_{n-2}\}$. But b_1 is the smallest such integer, whence $b_1 = \min(\{1, \ldots, n\} \setminus \{a_1, \ldots, a_{n-2}\}) = \mathrm{mix}\{a_1, \ldots, a_{n-2}\} = b'_1$. Assume that $b_j = b'_j$ for all $j \in \{1, \ldots, i\}$ and consider b_{i+1}. It is the smallest leaf in $T - b_1 - \ldots - b_i$ so, with the help of induction hypothesis

$$\begin{aligned}b_{i+1} &= \min(\{1, \ldots, n\} \setminus \{a_{i+1}, \ldots, a_{n-2}, b_1, \ldots, b_i\}) \\ &= \mathrm{mix}\{a_{i+1}, \ldots, a_{n-2}, b_1, \ldots, b_i\} = \mathrm{mix}\{a_{i+1}, \ldots, a_{n-2}, b'_1, \ldots, b'_i\} = b'_{i+1}\end{aligned}$$

Therefore, $\{a_i, b_i\} = \{a_i, b'_i\}$ for all i and the tree produced by the reconstruction procedure is T, the tree we started with. □

Closely related to trees are *monocyclic structures*, that correspond to chemical compounds with one benzene ring, such as benzoic acid $C_7H_6O_2$ depicted on page 3.

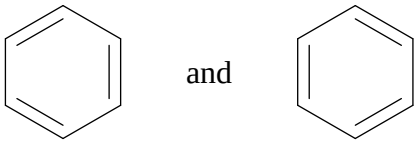
A graph G is *monocyclic* if it is connected and contains precisely one cycle. Since monocyclic structures are very close to trees (removing an arbitraty edge from the cycle yields a tree), there is a characterization of monocyclic structures that parallels that of the trees. By reducing to trees, it can easily be show that each pair of the three properties listed below implies the remaining one:

- being connected,
- having precisely one cycle, and
- $m = n$.

Theorem 2.18 (a) *For each monocyclic graph we have $n = m$.*
(b) *A conneted graph satisfying $m = n$ contains precisely one cycle.*
(c) *A graph with precisely one cycle and satisfying $m = n$ has to be connected.*

2.4 Kekulé structures

Kekulé structure is a representation of an aromatic molecule (such as benzene), with fixed alternating single and double bonds, in which interactions between multiple bonds are assumed to be absent. For benzene, for example, Kekulé structures are:

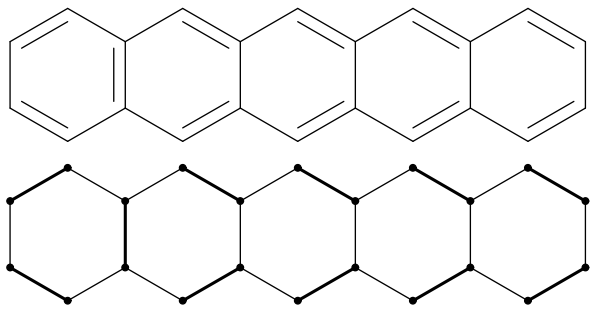

and

Mathematically, a model of a Kekulé structure consists of a set of edges of the coresponding graph such that no two edges in the set are adjacent and that every vertex is covered by exactly one of the selected edges. Such a set of edges is referred to as a perfect matching, and we now supply the definitions of the corresponding notions.

A *matching* of a graph $G = (V,E)$ is a set of edges $M \subseteq E$ such that for all $e_1, e_2 \in M$:

$$\text{if } e_1 \neq e_2 \text{ then } e_1 \cap e_2 = \varnothing.$$

The following is an example of a Kekulé structure and the corresponding set of edges of the graph of the compound:



We say that a matching M of a graph G is *maximal* if for every other matching M' the following holds: $M' \supseteq M$ implies $M' = M$. This means that no proper superset of M is a matching of G. A matching M of a graph G is *maximum* if for every other matching M' the following holds: $|M'| \leqslant |M|$. This means that no other a matching of G excedes M in the number of edges. Finally, a matching M of a graph G is *perfect* if every vertex of the graph belongs to precisely one edge of in the matching. The picture immediately above depicts a perfect matching.

We shall now discuss the existence of matchings in graphs. It is easy to see that every graph has a maximal matching and a maximum matching. However, not all graphs have perfect matchings.

Lemma 2.19 *If a graph G has a perfect matching, then $n(G)$ has to be even.*

Proof. Let $M = \{e_1, \ldots, e_k\}$ be a perfect matching of G. By the definition of the matching, distinct edges of M are disjoint, so $\left|\bigcup_{i=1}^{k} e_i\right| = 2k$, which is an even integer. For every matching we have $\bigcup_{i=1}^{k} e_i \subseteq E(G)$. However, the requirement that M be a perfect matching means that $\bigcup_{i=1}^{k} e_i \subseteq E(G)$, so $n(G) = 2k$, which is even. □

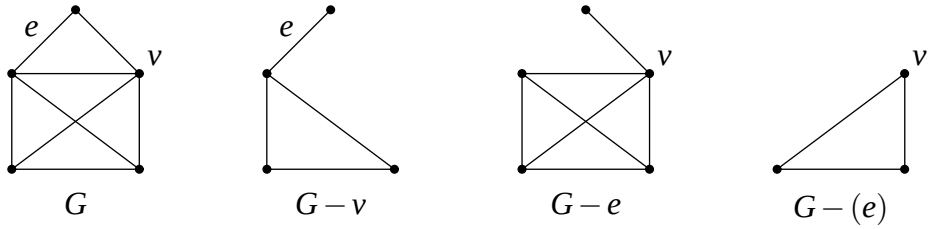
There is a very deep result tue to Tutte which provides a neccessary and sufficient condition for a graph to have a perfect matching. Let $\omega_{\text{odd}}(G)$ denote the number of odd components of G, where a component $S \subseteq V(G)$ is *odd* if $|S|$ is an odd integer.

Theorem 2.20 (Tutte, 1947) *A graph G has a perfect matching if and only if*

$$\omega_{\text{odd}}(G - W) \leqslant |W|,$$

for all subsets $W \subseteq V(G)$.

Finally, we consider the general formula for the number of perfect matchings of a graph G. Let $e = x, y$ be an edge and v a vertex of a graph G. By $G - e$ we denote the graph obtained from G by removing the edge e, while $G - v$ denotes the graph obtained from G by removing v and all the edges of G incident to v. Moreover, by $G - (e)$ we denote the graph $G - x - y$:



Theorem 2.21 *Let $K(G)$ denote the number of distinct Kekulé structures (= perfect matchings) in G. Then for every edge $e \in E(G)$ we have:*

$$K(G) = K(G - e) + K(G - (e)).$$

Proof. Let G be a graph and e and arbitrary edge of G. The class of all perfect matchings of G splits naturally into two disjoint classes: the perfect matchings of G that contan e, and those that don't. Therefore, if $\mathscr{M}(G)$ denotes the class of all perect matchings of G, $\mathscr{M}_e(G)$ denotes the class of all perfect matchings of G that contain e, and $\mathscr{M}'_e(G)$ denotes the class of all perfect matchings of G that do not contain e, then

$$|\mathscr{M}(G)| = |\mathscr{M}_e(G)| + |\mathscr{M}'_e(G)|.$$

The main idea of the proof is to show that $|\mathscr{M}_e(G)| = K(G - (e))$, while $|\mathscr{M}'_e(G)| = K(G - e)$.

Let M be a perfect matching in G and assume that $e \notin M$. Then M is a perfect matching of $G - e$. On the other hand, every perfect matching of $G - e$ is also a perfect matching of G, so $|\mathscr{M}'_e(G)| = |\mathscr{M}(G - e)| = K(G - e)$.

Assume, now, that M is a perfect matching of G such that $e \in M$ and let $e = \{x,y\}$. Since e belongs to M, the endvertices x and y of e are covered by e, so no edge incident to x or y belongs to M. Therefore, $M' = M \setminus \{e\}$ is a perfect matching of $G - (e)$. On the other hand, every perfect matching M of $G - (e)$ can be extended to a perfect matching $M \cup \{e\}$ of G. Therefore, $|\mathscr{M}_e(G)| = |\mathscr{M}(G - (e))| = K(G - (e))$. This completes the proof. $\square$

Chapter 3

Computer Implementation

There are many ways to represent graphs in the memory of a computer, and in this chapter we shall discuss two:

- adjacency matrices, and
- lists of neighbours.

3.1 Adjacency matrices

The simplest way to store a graph in the memory of a computer is to represent the graph using its adjacency matrix. Let G be a graph with $V(G) = \{1, 2, \ldots, n\}$. The *adjacency matrix of* G denoted by $A(G)$ is an $n \times n$ matrix matrica $A = [a_{ij}]$ where

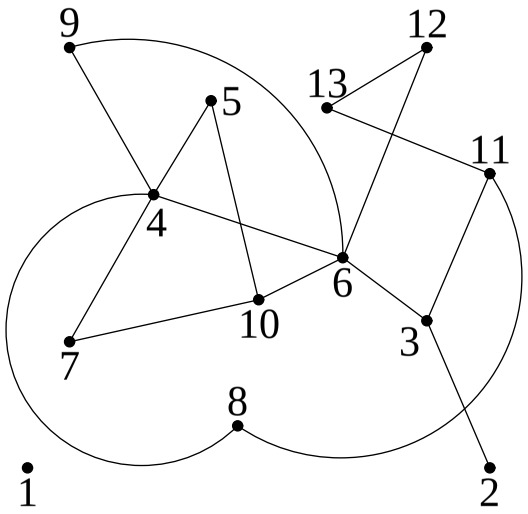
$$a_{ij} = \begin{cases} 0, & \text{vertices } i \text{ and } j \text{ are not adjacent,} \\ 1, & \text{vertices } i \text{ and } j \text{ are adjacent.} \end{cases}$$

If a vertex u is adjacent to a vertex v, then v is also adjacent to u, whence follows that $a_{ij} = a_{ji}$ for all i, j, i.e., $A(G)$ is a *symmetric matrix*. Fig. 3.1 contains an example.

We shall represent graphs as records consisting of two fields: the number of vertices of the graph and the adjacency matrix of the graph.

```
const
  MaxNoVertices = 50;
  Null = 0;

type
  Graph = record
    N : integer;
    adjacent : array [1 .. MaxNoVertices, 1 .. MaxNoVertices]
               of Boolean
  end;
```

$G =$



$A(G)$	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	1	0	0	0	0	1	0	0
4	0	0	0	0	1	1	1	1	1	0	0	0	0
5	0	0	0	1	0	0	0	0	0	1	0	0	0
6	0	0	1	1	0	0	0	0	1	1	0	1	0
7	0	0	0	1	0	0	0	0	0	1	0	0	0
8	0	0	0	1	0	0	0	0	0	0	1	0	0
9	0	0	0	1	0	1	0	0	0	0	0	0	0
10	0	0	0	0	1	1	1	0	0	0	0	0	0
11	0	0	1	0	0	0	0	1	0	0	0	0	1
12	0	0	0	0	0	1	0	0	0	0	0	0	1
13	0	0	0	0	0	0	0	0	0	0	1	1	0

Figure 3.1: A graph and its adjacency matrix

The constant value `Null` is reserved for the situations where no vertex of the graph is to be returned by a computation. The following three functions will be extremely useful in the sequel. The function `Deg(G, v)` returns the degree of `v` in `G`:

```
function Deg(var G : Graph; v : integer) : integer;
var
  sum, i : integer;
begin
  sum := 0;
  for i := 1 to G.N do
    if G.adjacent[v, i] then
      sum := sum + 1;
  Deg := sum
end;
```

The function `FirstNeighbour(G, v)` returns the (lexicographically) first neighbour of `v`, or `Null` if `v` has no neighbours:

```
function FirstNeighbour(var G : Graph; v : integer) : integer;
var
  i : integer;
  go : Boolean;
begin
  i := 1;
  go := true;
  while go and (i <= G.N) do
    if G.adjacent[v, i] then
      go := false
    else
      i := i + 1;
  if go then
    FirstNeighbour := Null
  else
    FirstNeighbour := i
end;
```

while `NextNeighbour(G, v, x)` returns the neighbour of `v` that comes (lexicographically) immediately after `x`, or `Null` if there are no more neighbours of `v`:

```
function NextNeighbour(var G : Graph; v, x : integer) : integer;
var
  i : integer;
  go : Boolean;
begin
  i := x + 1;
```

```
  go := true;
  while go and (i <= G.N) do
    if G.adjacent[v, i] then
      go := false
    else
      i := i + 1;
  if go then
    NextNeighbour := Null
  else
    NextNeighbour := i
end;
```

Example 3.1 As an example, we show the outline of a program that reads a graph from a file and then for every vertex of the graph prints its degree and the list of its neighbours.

```
program TheFirstExample;
const
  MaxNoVertices = 50;
  Null = 0;

type
  Graph = record
    N : integer;
    adjacent : array [1 .. MaxNoVertices, 1 .. MaxNoVertices]
               of Boolean
  end;

var
  G : Graph;
  v, w : integer;
  f : text;

function Deg(var G : Graph; v : integer) : integer;
  begin ... end;

function FirstNeighbour(var G : Graph; v : integer) : integer;
  begin ... end;

function NextNeighbour(var G : Graph; v, x : integer) : integer;
  begin ... end;

begin
  assign(f, 'G.txt'); reset(f);
```

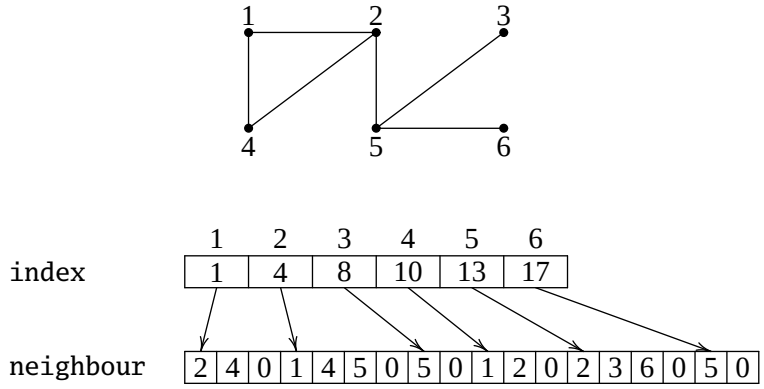


Figure 3.2: A graph and its representation by lists of neighbours

```
  ReadGraph(f, G); close(f);
  for v := 1 to G.N do
    begin
      write(v:2, Deg(G, v):3, ':');
      w := FirstNeighbour(G, v);
      while w <> Null do
        begin
          write(w:3);
          w := NextNeighbour(G, v, w)
        end;
      writeln
    end
end.
```

3.2 Lists of neighbours

We say that a graph is *sparse* if the number of edges of the graph is "relatively small" with respect to the number of its vertices. Sparse graphs can be very efficiently represented using two arrays: an arrays containing the lists of neighbours for every vertex, and an array containing, for each vertex, a pointer to the beginning of its list of neighbours. For convenience, we shall assume that every list of neighbours ends with `Null`. An example is given in Fig. 3.2.

Using this idea, we can represent sparse graphs as records with three fields: the number of vertices of the graph, the array containing the pointers, and the graph containing the lists of neighbours.

```
const
  MaxNoVertices = 50;
  MaxNeighbourListLen = 200;
  Null = 0;
```

```
type
  SGraph = record
    N : integer;
    index : array [1 .. MaxNoVertices] of integer;
    neighbour : array [1 .. MaxNeighbourListLen] of integer
  end;
```

The representation of a graph G with n vertices and m edges using adjacency matrices requires $O(n^2)$ memory units, while representation using lists of neighbours requires $O(m+n)$ memory units. The choice of the representation, therefore, depends on whether we have an additional information on the expected number of edges of the graph or not. If the problem under consideration provides no reasonable upper bound on the number of edges of the graph, it is usually recommended to represent graphs using adjacency matrices. If, however, we have an upper bound on m and if $n+m$ is sufficiently smaller than n, it pays off to represent the graph using lists of neighbours.

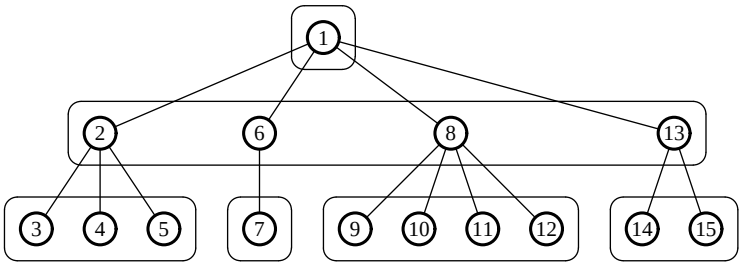
Sometimes the choice of a suitable representation depends on the operations we intend to perform on graphs. For example, it may pay off to use adjacecny matrix representation for very sparse graphs if the algorithms we use heavily depend on operations that are more easily implemented on matrices.

3.3 Depth-first search and connectedness

Many algorithms that manipulate graphs require a systematic way of traversing its vertices and in this section we describe a standard algorithm of traversing vertices of a graph known as *depth-first search*, or DFS for short.

Depth-first search (DFS) is a recursive algorithm that starting from a vertex traverses all the vertices in its *connected component* by first examining its first available neighbour, then the first available neightbour of the vertex the algoritm has just traversed, then the first available neighbour thereof, proceeding "deeper and deeper" as long as possible. Once all the neighbours of a vertex have been traversed, the algorithm back-tracks to traverse the next available neigbnour of the parent-vertex, until all the neighbours of the first vertex have been traversed. The algorithm can abstractly be described as follows:

```
procedure DFS(v)
    foreach neighbour w of v do
      if w has not been traversed then
        begin
          mark w
          DFS(w)
        end
```



and in order to obtain a working implementation we just have to fill in some details. We need an array

```
var
  Idx : array [1 .. MaxNoVertices] of integer;
```

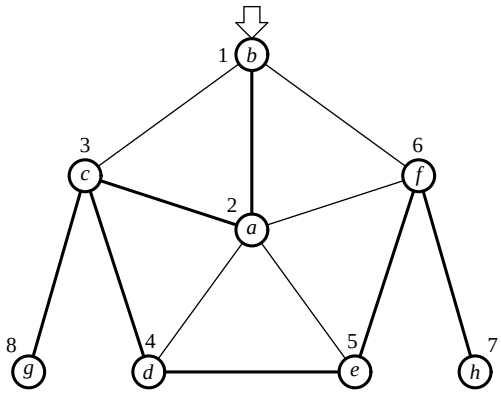
where the algorithm stores the index of the vertex: `Idx[v] = k` if and only if the algorithm traverses `v` in its *k*-th step. We use this array to record whether the vertex has been traversed or not. We shall initialize all the entries of the array to zero

```
for i := 1 to G.N do Idx[i] := 0;
```

and we now know that a vertex has been traversed if and only if `Idx[v] <> 0`. We shall also need a global variable `Lbl` which contains the next available index:

```
procedure DFS(v : integer);
{ G, Lbl, Idx are global and have to be }
{ initialized before invoking the procedure }
var
  i, w : integer;
begin
  Lbl := Lbl + 1;
  Idx[v] := Lbl;
  w := FirstNeighbour(G, v);
  while w <> Null do
    begin
      if Idx[w] = 0 then DFS(w);
      w := NextNeighbour(G, v, w)
    end
end;
```

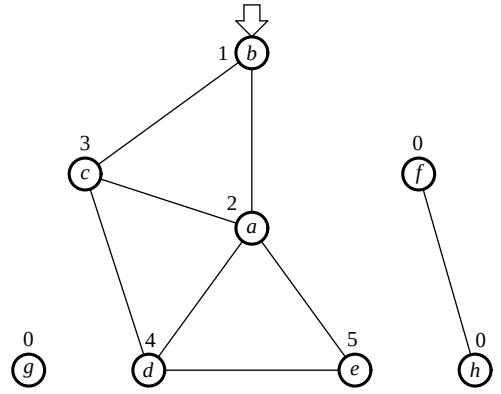
Example. In the adjacent figure we see a graph and the values of the variable `Idx` if we start DFS from vertex *b*. The algorithm first traverses *b*; then its neighbour *a*; then the first available neighbour of *a*, which turns out to be *c*; then *d*, and so on, until *h*. When the algorithm reaches the vertex *c* while getting out of recursive calls, it notes that its neighbour *g* has not been traversed, so another recursive call takes place. At the end, the algorithm backtracks to *b*, notes that all the neighbours of *b* have been traversed and terminates.



Let us stress again that DFS traverses all the vertices in the *connected component* of a vertex, as the following example shows.

Example. Consider the graph given in the adjacent figure and start DFS at vertex *b*. The order of traversal (values of `Idx`) is also indicated, and we see that for certain vertices the value of `Idx` is 0. Starting from *b* the algorithm has no way of reaching vertices *g*, *f* and *h* because these three vertices do not belong to the connected component of *b*.

An easy modification of the original DFS procedure leads to the algorithm that traverses vertices of both connected and disconnected graphs:




```
procedure FullDFS(v : integer);
{ G, Lbl, Idx are global variables }
begin
  DFS(v);
  for v := 1 to G.N do
    if Idx[v] = 0 then DFS(v)
end;
```

which leads to an easy and efficient algorithm that tests whether a graph is connected:

```
function Connected(var G : Graph) : boolean;
var
  v : integer;
  go : boolean;
  Idx : array [1 .. MaxNoVertices] of integer;

  procedure DFS(v : integer);
  var
```

```
    i : integer;
    w : integer;
  begin
    Idx[v] := 1;
    w := FirstNeighbour(G, v);
    while w <> Null do
      begin
        if Idx[w] = 0 then DFS(w);
        w := NextNeighbour(G, v, w)
      end
  end;

begin
  for v := 1 to G.N do Idx[v] := 0;
  DFS(1);

  v := 2;
  go := true;
  while go and (v <= G.N) do
    if Idx[v] = 0 then
      go := false
    else
      v := v + 1;
  Connected := go
end;
```

This algorithm also demonstrates another possible way to mark vertecies of the graphs we have traversed. Instead of storing the index of a vertex during traversal, this uses `Idx` to store only 0 or 1 according as the vertex *v* belongs to the connected component of the vertex 1 or not. This algorithm can easily be modified to actually count connected components of a graph.

3.4 The search tree the DFS algorithm

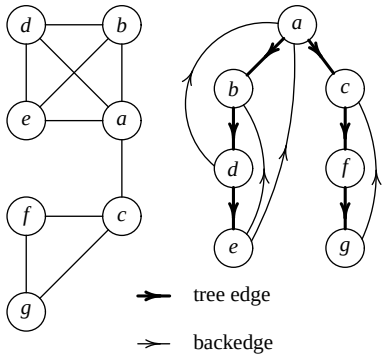
A careful look at DFS reveals that during the execution of the algorithm upon a connected graph, a special spanning tree of the graph is constructed. This tree is referred to as the *search tree of the DFS algorithm*. Therefore, the correctness of the algorithm `Connected` presented in the previous section follows from Theorem 2.11 which states that a graph is connected if and only if it has a spanning tree.

If *G* is not a connected graph, then DFS constructs a spanning tree of one of its connected components. The `FullDFS` procedure presented above constructs a spanning tree of each of its connected components.

A *rooted tree* is a tree where one of its vertices is distinguished. That distinguished vertex is

called the *root* of a tree. Clearly, the root of the search tree of the DFS algorithm is the vertex which appears as the input to the algorithm.

A spanning tree of a connected graph contains only some of the edges of the graph. The other edges, that is, those that do not appear in the spanning tree, are referred to as the *backedges*. The name comes from a detailed analysis of the DFS algorithm. Assume that we are traversing a graph and that we started the traversal in vertex *x*. Then each edge we examine in the main loop of the algorithm either leads to a new edge that has not been traversed until now, or to a vertex that has already been included into the search tree. In the latter case, the edge leads "back" to a vertex the algorithm has already visited, so it is a "backedge".



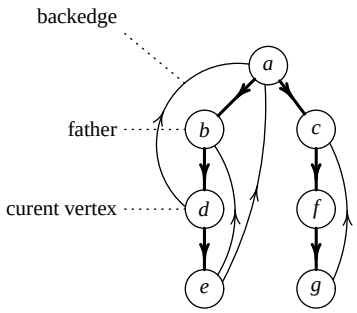
The presence of backedges indicates the existence of cycles in the graph, as the following theorem shows:

Theorem 3.2 *A connected graph has a cycle if and only if the search tree of the DFS algorithm has a backedge.*

This observation leads to an efficient algorithm which checks whether a graph is a tree.

```
function IsTree(var G : Graph) : Boolean;
{ We assume that G is connected }
var
  Idx : array [1 .. MaxNoVertices] of integer;
  i : integer;
  go : Boolean;

  procedure DFS(v, father : integer);
  var
    w : integer;
  begin
    Idx[v] := 1;
    w := FirstNeighbour(G, v);
    while go and (w <> Null) do
      begin
        if Idx[w] = 0 then
          DFS(w, v)
        else if w <> father then
          go := false;
        w := NextNeighbour(G, v, w)
      end
  end;
begin
```

```
    for i := 1 to G.N do Idx[i] := 0;
    go := true;
    DFS(1, Null);
    IsTree := go
  end;
```



We see that the algorithm is a straightforward modification of the DFS algorithm. The DFS procedure receives not only the vertex *v* from which the traversal is to continue, but also the *father* of that vertex, that is, the vertex from which we discovered *v*. This helps detect backedges: an edge is a backedge if it goes from *v* into a vertex *w* the algorithm has already visited, but which is *not* the father of *v*.

Chapter 4

Structures and Symmetry

One might say that humand kind was designed to comprehend and appreciate symmetry. The notion of symmetry in geometry transposes to the more abstract level of structures as the notion of automorphism.

4.1 A few words on groups

In this chapter we use some basics of group theory. To start with, a *group* is a set G together with a binary operation $\cdot$ (multiplication) such that the following requiremets (the axioms of group theory) are met:

(G1) multiplication is associative, i.e., $x(yz) = (xy)z$ for all $x, y, z \in G$;

(G2) there exists an $e \in G$ which acts as a *neutral element* for the multiplication, i.e., $xe = ex = x$ for all $x \in G$, and

(G3) for every $x \in G$ there is a $y \in G$ such that $xy = yx = e$; y is the called the *inverse of* x and, being unique for the given x, is usually denoted by x^{-1}.

The prototypical example of a group is the set of all bijections of a given set X onto itself. This group is called the *symmetric group* and denoted by $\mathrm{Sym}(X)$. If X is an n-element set, instead of $\mathrm{Sym}(X)$ we also write $\mathrm{Sym}(n)$.

A subset H of G is a *subgroup* of G if H is a group in itself with respect to the restriction of the multiplication of G, and shares the neutral element with G.

Example 4.1 The set of integers $\mathbb{Z}$ together with addition $+$ forms a group. The set $2\mathbb{Z}$ of even integers together with addition of integers forms a group, and it is a subgroup of $\mathbb{Z}$. Note that the two groups share the the neutral element 0.

If H is a subgroup of G, then $\{H \cdot g : g \in G\}$ is a partition of G. In other words, either $H \cdot g_1$ and $H \cdot g_2$ coincide, or are disjoint, for all $g_1, g_2 \in G$. We, therefore, immediately get

Theorem 4.2 (Lagrange Theorem) *If G is a finite group with n elements, and H a subgroup of G with k elements, then $k \mid n$.*

Sets of the form $H \cdot g$ are called *right cosets* of H in G and the set $G \setminus H = \{H \cdot g : g \in G\}$ is called *the set of right cosets* of H in G.

A *group homomorphism* from the groups G with multiplication $\cdot$ into the group H with multiplication $*$ is any mapping $f : G \to H$ satisfying

$$f(xy) = f(x) * f(y), \quad \text{for all } x, y \in G.$$

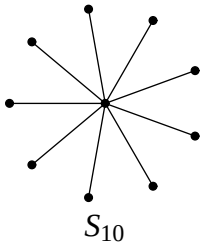
An *isomorphism* between groups G and H is a homomorphism from G into H which is bijective. In that case we say that G and H are *isomorphic* and write $G \cong H$.

Example 4.3 (a) The set of integers $\mathbb{Z}$ together with addition $+$ forms a group. The set $\mathbb{Z}_6 = \{0,1,2,3,4,5\}$ together with addition $+_6$ modulo 6 forms another group. The mapping $f : \mathbb{Z} \to \mathbb{Z}_6$ defined by $f(x) = x \bmod 6$ which takes an integer into its reminder modulo 6 is a group homomorphism.

(b) The set of reals $\mathbb{R}$ together with addition forms a group. The set of positive reals $\mathbb{R}^+$ together with multiplication forms another groups. The two groups are isomorphic, as we can easily see by considering the bijective mapping $f : \mathbb{R}^+ \to \mathbb{R}$ given by $f(x) = \log x$.

Automorphism groups of structures constitute an important class of groups. An *automorphsm* of a graph $G = (V,E)$ is every isomorhism $\varphi : V \to V$ from the graph onto itself. By $\mathrm{Aut}(G)$ we denote the set of all the automorphisms of G. The set $\mathrm{Aut}(G)$ of automorphisms of a graph together with the composition of mappings is a group.

Example 4.4 Let us now describe $\mathrm{Aut}(K_n)$, $\mathrm{Aut}(S_n)$ and $\mathrm{Aut}(P_n)$ for $n \geqslant 3$, where S_n denotes the *star* with n vertices, i.e., the graph where one vertex is adjacent to all the remaining vertices, and there are no other edges in the graph, see Fig. 4.1.



S_{10}

Figure 4.1: A star S_{10}

(a) Every bijection $\varphi : V(K_n) \to V(K_n)$ is an automorphism of K_n, so the group of automorphisms of K_n is isomorphic (as a group) to the symmetric group $\mathrm{Sym}(n)$.

(b) Every automorphism of a star on n vertices has to keep the central vertex fixed, while freely permuting the remaining $n-1$ vertices of the graph. Therefore, $\mathrm{Aut}(S_n)$ is isomorphic (as a group) to the symmetric group $\mathrm{Sym}(n-1)$.

(c) A path has only two automorphisms: the identity mapping which keeps all the vertices fixed, and the "flip" which "mirrors" the path by mapping the first vertex of the path onto the last one (and vice versa), the second vertex of the path onto the penultimate vertex of the path (and vice versa), and so on. Therefore, $|\mathrm{Aut}(P_n)| = 2$.

Every homomorphism $f : G \to H$ between two groups uniquely determines an equivalence relation θ by $(x,y) \in \theta$ if $f(x) = f(y)$. The relation θ is usually referred to as the *kernel* of f and denoted by $\ker f$.

A binary relation $\rho \subseteq G^2$ on a group G is called *congruence* if it is an equivalence relation such that

$$(x,y) \in \rho \text{ and } (u,v) \in \rho \text{ implies } (xu,yv) \in \rho, \qquad \text{for all } x,y,u,v \in G.$$

The kernel of every homomorphism is a congruence on the domain of the homomorphism. Conversely, every congruence is a kernel of some homomorphism.

For every congruence θ of a group G, the set of equivalence classes $\{g/\theta : g \in G\}$ can be turned into a group by defining the multiplication as follows:

$$(g_1/\theta) \cdot (g_2/\theta) = (g_1 g_2)/\theta.$$

This group is called the *factor group* of G and is denoted by G/θ.

Theorem 4.5 (The First Isomorphism Theorem) *Let $f : G \to H$ be a surjective homomorphism from a group G into a group H. Then $G/(\ker f) \cong H$.*

4.2 Group actions

Let X be a set and G a group. *The action of G on X* is a mapping $\mu : X \times G \to G$ such that

- $\mu(x,e) = x$, and
- $\mu(\mu(x,g),h) = \mu(x,gh)$.

An action of a group G on a set X will be denoted by (G,X). Instead of $\mu(x,g)$ we shall write x^g, so that the above two laws take the rather familiar form $x^e = x$ and $(x^g)^h = x^{gh}$.

Every $g \in G$ determines a mapping $\tau_g : X \to X : x \mapsto x^g$. Since G is a group, τ_g is a permutation of X for every $g \in G$ Therefore, every group action (G,X) determines a homomorphism $\lambda : G \to \mathrm{Sym}(X) : g \mapsto \tau_g$. Coversely, every homomorphism $\lambda : G \to \mathrm{Sym}(X)$ determines an action of the group G on X by $x^g := \Big(\lambda(g)\Big)^{-1}(x)$. This correspondence establishes a Cayely-type representation theorem for group actions.

The homomorphism $\lambda : G \to \mathrm{Sym}(X)$ that corresponds to the group action (G,X) is not necessarily injective, since it might happen that distinct elements of G act in the same fashion on X. We say that the actions (G,X) is *faithful* if the corresponding homomorphism is injective. If the group action (G,X) is not faithful, then the kernel $\theta = \ker(\lambda)$ of λ identifies the elements of G that act in the same fashion, so instead of (G,X) one can consider the faithful group action

$(G/\theta, X)$ given by $x^{g/\theta} = x^g$. Therefore, we can safely assume that the group actions we work with are faithful.

A group action (G,X) induces a binary relation $\sim$ on X as follows: we let $x \sim y$ if there exists a $g \in G$ such that $x^g = y$. Clearly, $\sim$ is an equivalence relation, and the equivalence classes of $\sim$ are referred to sa *orbits* of the group action (G,X). The orbit of an $x \in X$ has the form $\{x^g : g \in G\}$. This is why we shall denote the orbit of x by x^G. Let X/G denote the set of all orbits of the group action (G,X).

For $x \in X$, let G_x denote the *stabilizer of x*, i.e. the set of all group elements that leave x fixed:

$$G_x = \{g \in G : x^g = x\}.$$

Clearly, G_x is a subgroup of G. For $g \in G$, let $\mathrm{fix}(g)$ denote the *set of all fixpoints* of g:

$$\mathrm{fix}(g) = \{x \in X : x^g = x\}.$$

Theorem 4.6 (Lagrange Theorem) *Let (G,X) be a faithful group actions. Then for every $x \in X$ we have $|G| = |G_x| \cdot |x^G|$.*

Proof. Let $G_x \setminus G = \{G_x \cdot g : g \in G\}$ be the set of right cosets of G_x in G. From the Lagrange Theorem (for groups) we know that

$$|G_x \setminus G| = \frac{|G|}{|G_x|}.$$

Therefore, in order to show the claim of the theorem, it suffices to show that $|x^G| = |G_x \setminus G|$. Consider the mapping $\varphi : x^G \to G_x \setminus G : x^g \mapsto G_x \cdot g$.

- φ is well defined. Let $x^g = x^h$. Then $x^{gh^{-1}} = x^{hh^{-1}} = x^1 = x$, whence gh^{-1} stabilizes x. This implies $gh^{-1} \in G_x$, so $g \in G_x \cdot h$. This shows that $G_x \cdot g = G_x \cdot h$.
- φ is clearly surjective.
- φ injective. Let $G_x \cdot g = G_x \cdot h$. Then $g = kh$ for some $k \in G_x$. Now, $x^g = x^{kh} = (x^k)^h$. Since k stabilizes x we have that $x^k = x$, and thus $x^g = x^h$.

Therefore, φ is a bijection, which proves the theorem. □

Theorem 4.7 (Cauchy-Frobenius (Burnside) Lemma) *Assume that the group G acts faithfully on X. Then*

$$|X/G| = \frac{1}{|G|} \sum_{g \in G} |\mathrm{fix}(g)|.$$

Proof. For an arbitrary formula Φ let

$$\chi(\Phi) = \begin{cases} 1, & \Phi \\ 0, & \neg\Phi. \end{cases}$$

It is now easy to see that $|\text{fix}(g)| = \sum_{x\in X} \chi(x^g = x)$ and $|G_x| = \sum_{g\in G} \chi(x^g = x)$. Therefore,

$$\sum_{g\in G} |\text{fix}(g)| = \sum_{g\in G}\sum_{x\in X} \chi(x^g = x) = \sum_{x\in X}\sum_{g\in G} \chi(x^g = x) =$$
$$= \sum_{x\in X} |G_x| = \sum_{x\in X} \frac{|G|}{|x^G|} = |G| \cdot \sum_{x\in X} \frac{1}{|x^G|}.$$

Let $X/G = \{\Omega_1, \ldots, \Omega_s\}$. Then

$$\sum_{x\in X} \frac{1}{|x^G|} = \sum_{x\in\Omega_1\cup\ldots\cup\Omega_s} \frac{1}{|x^G|} = \sum_{i=1}^{s}\sum_{x\in\Omega_i} \frac{1}{|x^G|}.$$

We are now going to show that $\sum_{x\in\Omega_i} \frac{1}{|x^G|} = 1$. For $x \in \Omega_i$ we have $x^G = \Omega_i$, so that

$$\sum_{x\in\Omega_i} \frac{1}{|x^G|} = \sum_{x\in\Omega_i} \frac{1}{|\Omega_i|} = |\Omega_i| \cdot \frac{1}{|\Omega_i|} = 1.$$

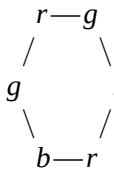
This implies

$$\sum_{g\in G} |\text{fix}(g)| = |G| \cdot \sum_{x\in X} \frac{1}{|x^G|} = |G| \cdot \sum_{i=1}^{s} 1 = |G| \cdot s = |G| \cdot |X/G|,$$

which concludes the proof. □

4.3 Pólya action

Let us start with an apparently simple problem. Let us try to count the number of necklaces with 6 beads, where each bead can be in one of the three colors r, g and b.

The formalization of the problem is as follows. Let $V = \{0,1,2,3,4,5\}$, $C = \{r,g,b\}$ and $X = C^V$. An element of X is a mapping $f : V \to C$, hence, a sequence of colors. Two such sequences represent the same necklace if there is a cyclic permutation of V that takes one of the sequences to the other. For example, the sequences $f = (r,g,r,r,b,g)$ and $g = (r,r,b,g,r,g)$ represent the same necklace 

, since $f = g \circ \sigma^{-1}$ for the cyclic permutation $\sigma = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 0 & 1 \end{pmatrix}$.

Therefore, in the problem of counting necklaces we have the group $\mathbb{Z}_6$ acting on V as a group of cyclic permutations. If we now "extend" the group action $(\mathbb{Z}_6, V)$ from V to C^V in such a way that $f^\sigma = f \circ \sigma^{-1}$, then the orbits of the extended action consist of configurations that represent the same necklace. We see, then, that the problem of counting distinct necklaces in its abstract form reduces to the problem of counting orbits of a group action, for which we have just developed tools – the Cauchy-Frobenius Lemma.

Above considerations motivate the following definition.

Definition 4.8 Let C and V be nonempty sets and let G act on V as a group of permutations. Then $f^\sigma = f \circ \sigma^{-1}$ defines an action of G on C^V called the *Pólya action*.

Let us now go back to the problem of counting necklaces. In the model we have just described, the group $\mathbb{Z}_6$ acts as a permutation group on C^V through the Pólya action. The orbit of an element $f \in C^V$ consists of all strings of colors of length 6 which represent the same necklace. Thus, the number of distinct necklaces equals the number of orbits in the Pólya action of the group $\mathbb{Z}_6$ on the set C^V. According to the Cauchy-Frobenius Lemma,

$$|C^V/\mathbb{Z}_6| = \frac{1}{6} \sum_{\sigma \in \mathbb{Z}_6} |\text{fix}(\sigma)|,$$

where the fixpoints of σ's are to be counted in C^V.

Let us now take a look at the sets $\text{fix}(\sigma)$: $f \in \text{fix}(\sigma)$ means that $f = f^\sigma$, i.e., that $f(k) = f(\sigma^{-1}(k))$ for all $k \in V$. Note that this means that f has to be constant on the cycles of σ. This is such an important conclusion, that we shall frame it:

$f \in \text{fix}(\sigma)$ in the Pólya action if and only if f is constant on the cycles of σ.

The group $\mathbb{Z}_6$ as a permutation group consists of the following permutations given in their cyclic representation:

$$\text{id},\ (012345),\ (543210),\ (024)(135),\ (420)(531),\ (03)(14)(25).$$

- The permutation id has 6 cycles. On each of the cycles the mapping f can take each of the three values, so $|\text{fix}(\text{id})| = 3^6$.
- The permutation $\sigma = (012345)$ has only one cycle, so $|\text{fix}(\sigma)| = 3$.
- Similarly, $|\text{fix}(\sigma)| = 3$ for $\sigma = (543210)$.
- The permutation $\sigma = (024)(135)$ has 2 cycles. On each of the cycles f can take any of the three values, so $|\text{fix}(\sigma)| = 3^2$.
- Similarly, $|\text{fix}(\sigma)| = 3^2$ for $\sigma = (420)(531)$.
- Finally, the permutation $\sigma = (03)(14)(25)$ has three cycles, whence $|\text{fix}(\sigma)| = 3^3$.

Therefore, the number of distinct necklaces equals $\frac{1}{6}(3^6 + 2 \cdot 3 + 2 \cdot 3^2 + 3^3) = 130$.

We have just seen in this example that permutations with the same "cyclic structure" have the same number of fixed elements in the Pólya action. *This is a general phenomenon, rather than an isolated case!* We shall, therefore, introduce two new notions: the cyclic type, and the cyclic number of a permutation.

Definition 4.9 Let X be a finite set with $|X| = n$ and let G act on X as a permutation group. Suppose that a permutation $\sigma \in G$ has a_1 cycles of length 1, a_2 cycles of length 2, etc., a_n cycles of length n. Then the *cyclic type* of sigma is the sequence $\mathrm{ct}(\sigma) = (a_1, a_2, \ldots, a_n)$, and the *cyclic number* of σ is the number $\mathrm{cn}(\sigma) = a_1 + a_2 + \ldots + a_n$.

It is clear that $\mathrm{cn}(\sigma)$ is the number of cycles of σ, and that $1 \cdot a_1 + 2 \cdot a_2 + \ldots + n \cdot a_n = n$ if $(a_1, a_2, \ldots, a_n)$ is the cyclic type of a permutation from G.

Note that cn and ct depend on the group action under consideration rather then on the abstract properties of the group. Therefore, the same element of a group can have distinct cyclic types in distinct group actions.

Theorem 4.10 (Cauchy-Frobenius Lemma for the Pólya Action) *For any faithful Pólya action of G on C^V the following holds:*

$$|C^V/G| = \frac{1}{|G|} \sum_{\sigma \in G} |C|^{\mathrm{cn}(\sigma)}.$$

Proof. According to the Cauchy-Frobenius Lemma,

$$|C^V/G| = \frac{1}{|G|} \sum_{\sigma \in G} |\mathrm{fix}(\sigma)|$$

so that it suffices to show that $|\mathrm{fix}(\sigma)| = |C|^{\mathrm{cn}(\sigma)}$ for every $\sigma \in G$. Take any $\sigma \in G$ and let $\mathrm{cn}(\sigma) = k$. As we have already noted, $f \in \mathrm{fix}(\sigma)$ if and only if f is constant on the cycles of σ. Therefore $|\mathrm{fix}(\sigma)|$ equals the number of mappings $f : V \to C$ that are constant on the cycles of σ. On each of the k cycles of σ the mapping f can take any of the $|C|$ values, whence $|\mathrm{fix}(\sigma)| = |C|^k = |C|^{\mathrm{cn}(\sigma)}$. □

4.4 Counting nonisomorphic graphs

Each graph $G = (V, E)$ can be understood as a mapping $f_E : \binom{V}{2} \to \{0,1\}$, where f_E is the characteristic function of the set E: $f_E(e) = \chi(e \in E)$.

The symmetric group $\mathrm{Sym}(V)$ acts on V in a straightforward way: $x^\sigma = \sigma^{-1}(x)$. Assume now that $\mathrm{Sym}(V)$ acts on $\binom{V}{2}$ "coordinate-wise": $\{x,y\}^\sigma = \{x^\sigma, y^\sigma\}$. This action then "extends" to the Pólya action of $\mathrm{Sym}(V)$ on $2^{\binom{V}{2}}$ where

$$f^\sigma\Big(\{x,y\}\Big) = f\Big(\{\sigma^{-1}(x), \sigma^{-1}(y)\}\Big)$$

for all $\sigma \in \mathrm{Sym}(V)$ and $f \in 2^{\binom{V}{2}}$.

The following obvious lemma is the key to applications of Pólya theory on the problem of counting nonisomorphic graphs:

Lemma 4.11 *Graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ are isomorphic if and only if there is a $\sigma \in \mathrm{Sym}(V)$ such that $f_{E_1}^{\sigma} = f_{E_2}$.*

In other words, two graphs are isomorphic if and only if they belong to the same orbit of the Pólya action we have just described, whence follows that the number of nonisomorphic graphs on a finite set V equals the number of orbits of the Pólya action of $\mathrm{Sym}(V)$ on $2^{\binom{V}{2}}$. The Cauchy-Frobenius Lemma for Pólya actions now straightforwardly implies:

Theorem 4.12 *The number of nonisomorphic graphs on a set with n vertices is*

$$\frac{1}{n!} \sum_{\sigma \in \mathrm{Sym}(n)} 2^{\mathrm{cn}(\sigma)}.$$

In spite of the nicely looking theorem above, it is still terribly complicated (impossible in the general case!) to compute the number of nonisomorphic graphs on n vertices. What makes it so complicated is the fact that $\mathrm{cn}(\sigma)$ is the number of cycles in the cyclic representation of the permutation σ in the action of the group $\mathrm{Sym}(n)$ on the set $\binom{n}{2}$. *This number is not necessarily equal to the number of cycles of the permutation σ in the "ordinary" action of* $\mathrm{Sym}(n)$ *on* $\{1, \ldots, n\}$. The action of $\mathrm{Sym}(n)$ on $\binom{n}{2}$ interpreted as a subgroup G_n of $\mathrm{Sym}(\binom{n}{2})$ shows that, although $\mathrm{Sym}(n)$ and G_n are abstractly isomorphic, the cyclic structures of the corresponding elements are *essentially different*.

Chapter 5

Counting Hexagonal Systems

A polyhex system is a connected system of congruent regular hexagons such that any two hexagons either share exactly one vertex, one edge or they are disjoint. Presently, we shall be interested only in the class of geometrically planar, simply connected polyhexes. A polyhex is geometrically planar when it does not contain any overlapping edges, and a simply connected polyhex has no holes (see Fig. 5.1). In this way the helicenes and coronoids (molecules with "holes") are excluded.

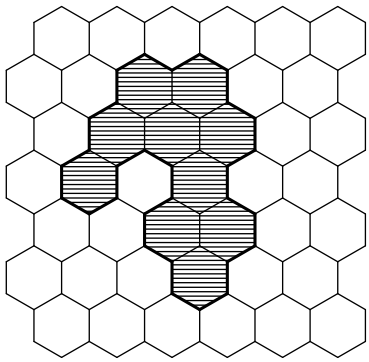


Figure 5.1: A geometrically planar, simply connected polyhex with $h = 10$ hexagons

The geometrically planar, simply connected polyhexes have often been referred to as "benzenoids". These systems may conveniently be defined in terms of a cycle on a hexagonal lattice; the system is found in the interior of this cycle, which represents the boundary (usually called the *perimeter*) of the system. In order to avoid confusion, we use the term *hexagonal system* (HS) for a geometrically planar, simply connected polyhex.

A classical paper on counting polyhex hydrocarbons dates back to 1968 [1], but it was not before 1983 that the Düsseldorf-Zagreb group (Knop and Trinajstić with collaborators) published their results from computerized enumerations hexagonal systems to $h = 10$, where h is the number of hexagons within the perimeter of the system [3].

The sad part of the story is that there is no general formula for the number of hexagonal systems with the given number of hexagons. Therefore, all we can do to enumerate and calssify them

is to use brute force: fast computers and algorithms. This chapter presents one such algorithm which has been used to enumerate nonisomorphic hexagonal systems with $h \leqslant 17$ hexagons and to classify them according to their perimeter.

5.1 The basics

Two different hexagonal systems are considered *isomorphic* if they are congruent in the sense of euclidean geometry. For example, Fig. 5.2 depicts three congruent hexagonal systems.

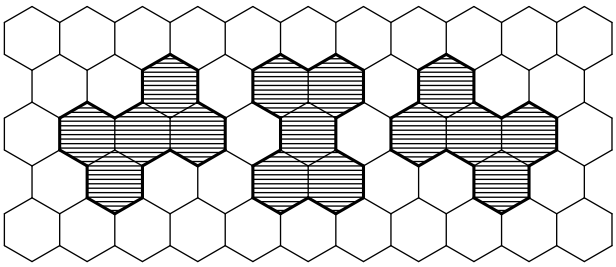
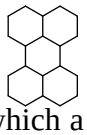


Figure 5.2: Three congruent hexagonal systems

Another point of view is that the hexagonal system  appears in three different *orientations* in the hexagonal grid. The number of ways in which a hexagonal system appears in the hexagonal grid is of ultimate importance when counting nonisomorphic, or *essentially distinct*, hexagonal systems with the given number of hexagons. The crucial notion in determining that number is the number of *symmetries* of the system.

The geometric notion of symmetry corresponds to the algebraic notion of an *automorphism* of the system. In that respect, symmetric structures are rich in automorphisms, and the we take the cardinality of the automorphism group as a measure of the symmetry of the system.

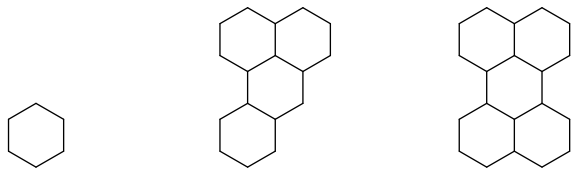
In order to be more specific, we now introduce the formal notions. Let S be an arbitrary set of points in the euclidean plane. An *automorphism* of S is every bijective distance-preserving mapping f of the plane which maps S onto itself, i.e., such that $f(S) = S$. The set of all such mappings

$$\mathrm{Aut}(S) = \{f : f(S) = S \text{ and } f \text{ is a bijective distance-preserving mapping}\}$$

carries the structure of a group under the function composition. This is why $\mathrm{Aut}(S)$ is referred to as the *automorphism group* of S.

For example, let S_1 be the hexagonal system consisting of precisely one hexagon, Fig. 5.3 (a). Then $|\mathrm{Aut}(S)| = 12$ since there are 6 rotations and 6 axial symmetries that map S_1 onto itself. It is easy to see that:

Lemma 5.1 *Let S be a hexagonal system. Then $|\mathrm{Aut}(S)| \leqslant 12$.*



(a) (b) (c)

Figure 5.3: Three hexagonal systems

As another example, let S_4 be the hexagonal system consisting of four hexagons depicted in Fig. 5.3 (b) and let S_5 be the hexagonal system consisting of five hexagons depicted in Fig. 5.3 (c). Then $|\mathrm{Aut}(S_4)| = 1$ since the trivial map which maps every point of the plane onto itself is the only map which maps S_4 onto itself. On the other hand, $|\mathrm{Aut}(S_5)| = 4$: there are two rotations and two axial symmetries that take S_5 onto itself.

As a consequence of the Cauchy-Frobenius Lemma (Lemma 4.7) we have the following:

Theorem 5.2 *Let S be a hexagonal system and let $k = |\mathrm{Aut}(S)|$. Then S appears in $12/k$ different orientations in the hexagonal grid.*

For example, for the hexagonal system S_5 depicted in Fig. 5.3 (c) we know that $|\mathrm{Aut}(S_5)| = 4$, so S_5 appears in $12/4 = 3$ different orientations in the hexagonal grid. The three different orientations are depicted in Fig. 5.2.

A careful inspection shows that the automorphism group of an arbitrary hexagonal system is one of the following groups: D_{6h}, C_{6h}, D_{3h}, C_{3h}, D_{2h}, C_{2h}, C_{2v} and C_s. These are well-known groups and their numbers of elements are:

Group	D_{6h}	C_{6h}	D_{3h}	C_{3h}	D_{2h}	C_{2h}	C_{2v}	C_s
No. of elem's	12	6	6	3	4	2	2	1

Let $H(h)$ be the number of all hexagonal systems with h hexagons including the isomorphic copies, and let $N(h)$ denote the number of nonisomorphic hexagonal systems with h hexagons. Furthermore, for a group $G \in \{D_{6h}, C_{6h}, D_{3h}, C_{3h}, D_{2h}, C_{2h}, C_{2v}, C_s\}$ let $N(G,h)$ denote the number of nonisomorphic hexagonal systems with h hexagons whose automorphism group is G. Then

$$\begin{aligned} H(h) = {} & N(D_{6h},h) + 2N(C_{6h},h) + 2N(D_{3h},h) + 4N(C_{3h},h) \\ & + 3N(D_{2h},h) + 6N(C_{2h},h) + 6N(C_{2v},h) + 12N(C_s,h) \end{aligned} \tag{5.1}$$

since a hexagonal system S with the automorphism group G appears in $12/|G|$ different orientations in the hexagonal grid. On the other hand,

$$\begin{aligned} N(h) = {} & N(D_{6h},h) + N(C_{6h},h) + N(D_{3h},h) + N(C_{3h},h) \\ & + N(D_{2h},h) + N(C_{2h},h) + N(C_{2v},h) + N(C_s,h). \end{aligned} \tag{5.2}$$

The number we are interested in is $N(h)$, the number of nonisomorphic hexagonal systems with h hexagons, and this number is not easy to compute. Instead of computing $N(h)$ directly, we shall first compute the numbers $H(h)$, $N(D_{6h},h)$, $N(C_{6h},h)$, $N(D_{3h},h)$, $N(C_{3h},h)$, $N(D_{2h},h)$, $N(C_{2h},h)$ and $N(C_{2v},h)$, then use formula 5.1 to compute $N(C_s,h)$, and finally use formula 5.2 to compute $N(h)$.

5.2 The algorithm

In this section we are going to outline the algorithm used to compute the number $H(h)$ of all *distinct* (but not necessarily nonisomorphic) hexagonal systems with h hexagons. In order to eliminate translationally equivalent systems (see Fig. 5.4) we introduce the notion of a cage and enumerate all hexagonal systems that are properly placed in the cage.

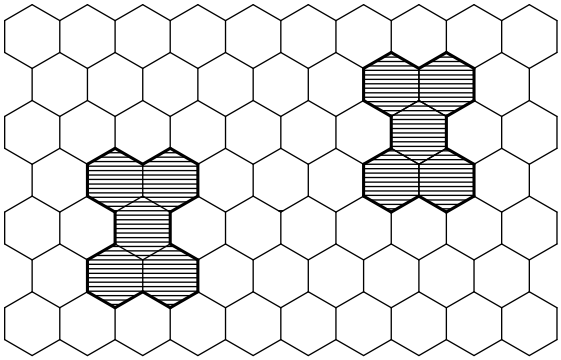


Figure 5.4: Two translationally equivalent hexagonal systems

A *cage* is a rather regular region of hexagonal grid in which we try to catch all relevant hexagonal systems. This algorithm uses a triangular cage, where the triangle is equliateral. Let Cage(h) denote the triangular cage with h hexagons along each side. Fig. 5.5 shows Cage(9) and demonstrates how a coordinate system can be introduced into the Cage(h).

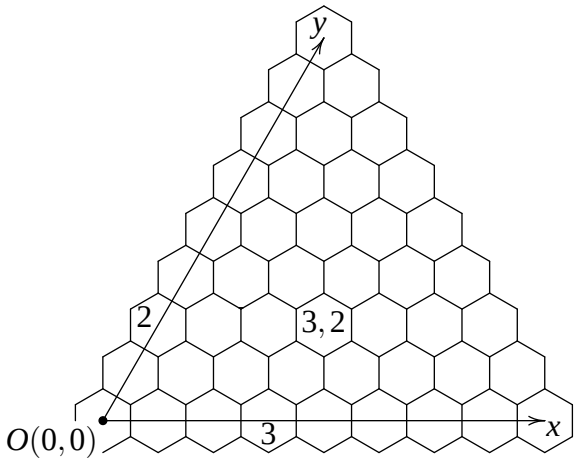


Figure 5.5: Cage(9)

Each hexagonal system with at most h hexagons can be placed in the cage in such a way that at least one of its hexagons is on the x-axis of the cage, and at least one of its hexagons is on the y-axis of the cage. We shall say that such hexagonal systems are *properly placed* in the cage. Fig. 5.6 depicts a hexagonal system with $h = 9$ hexagons properly placed in the Cage(9).

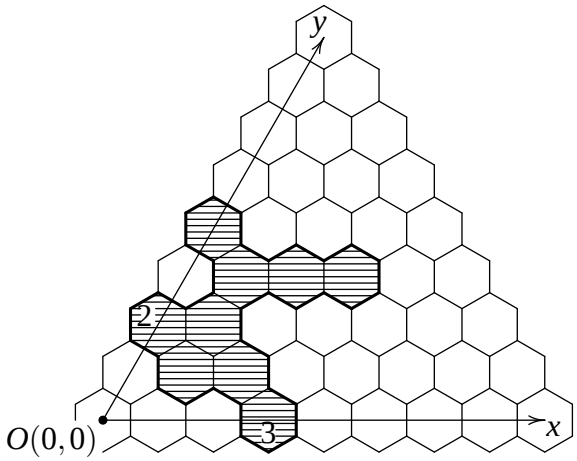


Figure 5.6: A propely placed hexagonal system with $h = 9$ hexagons

Reducing our search to the cage does not interfere with the general strategy outlined in the introduction to this chapter. Namely,

Theorem 5.3 *Let S be a hexagonal system with h hexagons and let $k = |\mathrm{Aut}(S)|$. Then there are $12/k$ different hexagonal systems which are isomorphic to S and properly placed in* $\mathrm{Cage}(h)$.

Thus, we generate and enumerate all hexagonal systems that are properly placed in the cage. Since we know how many times symmetric hexagonal systems have appeared, we can easily determine the number of all hexagonal systems with the trivial symmetry group C_s. This suffices to determine the number of all nonisomorphic hexagonal systems with the given number of hexagons, provided we have enumerated all symmetric hexagonal systems. By doing this, *we avoid isomorphism tests* which are considered to be the most time consuming parts of similar algorithms.

Consider a hexagonal system which is properly placed in the $\mathrm{Cage}(h)$, let p be the smallest coordinate of all of its hexagons on the x-axis and q be the smallest coordinate of all of its hexagons on the y-axis. Hexagons with coordinates $(p,0)$ and $(0,q)$ (with respect to the coordinate system of the cage) shall be of great importance to us and we refer to them as *key hexagons*.

Let $H(p,q)$ denote the set of all hexagonal systems with $\leqslant h$ hexagons satisfying the following two conditions:

- the hexagonal system is properly placed in Cage(h)
- its key hexagon on x-axis has coordinate p and its key hexagon on y-axis has coordinate q.

(Fig. 5.6 depicts one element of $H(3,2)$).

The family $\{H(p,q) : 0 \leqslant p \leqslant h-1, 0 \leqslant q \leqslant h-1\}$ is a partition of the set of all hexagonal systems with $\leqslant h$ hexagons that are properly placed in Cage(h). Because of the overall beauty of symmetry, it can be easily verified that $|H(p,q)| = |H(q,p)|$, for all $p,q \in \{0,1,\ldots,h-1\}$. Thus, the enumeration of all properly placed hexagonal systems reduces to determining $|H(p,q)|$ for all $p \geqslant q$. This gives the algorithm outlined in Fig. 5.7.

```
initialize Cage(h);
total := 0;
for q := 0 to h − 1 do
    for p := q to h − 1 do
        determine H(p,q);
        n := |H(p,q)|;
        if p = q then
            total := total + n
        else
            total := total + 2n
        fi
    od
od
```

Figure 5.7: The first iteration of the algorithm

Given the numbers $0 \leqslant q \leqslant p \leqslant h-1$ and Cage(h), determining $|H(p,q)|$ reduces to generating all the hexagonal systems from $H(p,q)$. We do that by generating their boundary line. A quick glance at Fig. 5.6 reveals that the boundary line of a properly placed hexagonal system can be divided in two parts: the left part of the boundary (from the readers point of view) which starts on the y-axis below the key hexagon and finishes at the first junction with x-axis, and the rest of the boundary which we call the right part of the boundary.

We recursively generate the left part of the boundary line. As soon as it reaches the x-axis, we start generating the right part. All the time we take care of the length of the boundary line as well as the area of the hexagonal system. The trick which gives the area of the hexagnal system is simple: each time the boundary goes up, we subtract the corresponding x coordinate. Each time the boundary goes down, we add the corresponding x coordinate. The "zig—zag" movements do not interfere the area. Once the generating is over, the area of the hexagonal system gives the number of hexagons circumscribed in this manner. We have to do this because of useless

hexagonal systems. Namely, during the generation of systems with h hexagons that belong to $H(p,q)$, systems with $> h$ hexagons also appear, so we have to eliminate them.

It would be a great waste of time (and computing power) to insist on generating elements of $H(p,q)$ strictly. This would require additional tests to decide whether the left part of the boundary has reached x-axis precisely at hexagon p or not. On the other hand, once we find out we have reached a hexagon, say, $p+2$, why should we ignore what we have achieved and discard this path in order to generate it again within attempts to find $H(p+2,q)$? This is why we are going to introduce another partition of the set of all properly placed hexagonal systems.

Given h and Cage(h), put $H^*(q) = \cup_{j=q}^{h-1} H(j,q)$, for all $q = 0,1,\ldots,h-1$. It is obvious that $\{H^*(q) : 0 \leqslant q \leqslant h-1\}$ is a partition of the set of all hexagonal systems with h hexagons that are properly placed in Cage(h). Instead of having two separate phases (generating $H(p,q)$ and adding appropriate numbers to *total*), we now have one phase in which generating and counting are put together. All we have to take care of is to prevent appearances of hexagonal systems from $H(p,q)$ with $p < q$. But this causes no overhead because it can be achieved by forbidding some left and some down turns in the matrix representing the cage. On the contrary, introducing the forbidden turns accelerates the process of generating the boundary line. Having all this in mind, the second iteration of the algorithm can be formulated as in Fig. 5.8.

As we can see, the algorithm is a straightforward example of backtracking, thus facing all classical problems of the technique: even for small values of h the search tree misbehaves so it is essential to cut it as much as possible. One idea that cuts some edges of the tree is based on the fact that for larger values of q there are some parts of the cage that cannot be reached by hexagonal system with $\leqslant h$ hexagons, but can easily be reached by useless hexagonal systems that emerge as a side-effect. That is why we can, knowing q, forbid some regions of the cage.

The other idea that reduces the search tree is counting the boundary hexagons. A *boundary hexagon* is a hexagon which has at least one side in common with the boundary line and which is in the interior of the hexagonal system we are generating. It is obvious that boundary hexagons shall be part of the hexagonal system, so we keep track on their number. We use that number as a *very good* criterion for cutting off useless edges in the search tree. The idea is simple: further expansion of left/right part of the boundary line is possible if there are less than h boundary hexagons the line has passed by.

The third idea that speeds up the algorithm is living on credit. When we start generating the left part of the boundary, we do not know where exactly is it going to finish on x-axis, but we know that *it is going to finish on x-axis*. In other words, knowing that there is one hexagon on the x-axis that is going to become a part of the hexagonal system, we can count that hexagon as a bundary hexagon in advance. Thus, many useless hexagonal systems are discarded before the left part of the boundary lands on the x-axis.

All these ideas collected in one place represent the core of the algorithm (Fig. 5.9).

5.3 The implementation

What has been shown in the previous section is just the main idea of the algorithm presented in [7]. Many purely technical things had to be added in order to make a working computer program out of it.

```
procedure ExpandRightPart(ActualPos);
begin
    if EndOfRightPart then
        n := NoOfHexagons();
        if n ⩽ h then
            determine p;
            if p = q then
                total[n] := total[n] + 1
            else
                total[n] := total[n] + 2
            fi
        fi
    else
        FindPossible(ActualPos, FuturePos);
        while RightPartCanBeExpanded(ActualPos, FuturePos) do
            ExpandRightPart(FuturePos);
            CalcNewFuturePos(ActualPos, FuturePos)
        od
    fi
end;
procedure ExpandLeftPart(ActualPos);
begin
    if EndOfLeftPart then
        ExpandRightPart(RightInitPos(q))
    else
        FindPossible(ActualPos, FuturePos);
        while LeftPartCanBeExpanded(ActualPos, FuturePos) do
            ExpandLeftPart(FuturePos);
            CalcNewFuturePos(ActualPos, FuturePos)
        od
    fi
end;
begin
    initialize Cage(h);
    set total[1...h] to 0;
    for q := 0 to h − 1 do
        initialize y-axis key hexagon(q);
        ExpandLeftPart(LeftInitPos(q))
    od
end
```

Figure 5.8: The second iteration of the algorithm

```
procedure ExpandRightPart(ActualPos, BdrHexgns);
begin
    if EndOfRightPart then
        n := NoOfHexagons();
        if n ⩽ h then
            determine p;
            if p = q then total[n] := total[n] + 1
            else total[n] := total[n] + 2
            fi
        fi
    else
        FindPossible(ActualPos, FuturePos);
        while RightPartCanBeExpanded(ActualPos, FuturePos)
        and BdrHexgns ⩽ h do
            ExpandRightPart(FuturePos, update(BdrHexgns));
            CalcNewFuturePos(ActualPos, FuturePos)
        od
    fi
end;
procedure ExpandLeftPart(ActualPos, BdrHexgns);
begin
    if EndOfLeftPart then
        ExpandRightPart(RightInitPos(q), updCredit(BdrHexgns))
    else
        FindPossible(ActualPos, FuturePos);
        while LeftPartCanBeExpanded(ActualPos, FuturePos)
        and BdrHexgns ⩽ h do
            ExpandLeftPart(FuturePos, update(BdrHexgns));
            CalcNewFuturePos(ActualPos, FuturePos)
        od
    fi
end;
begin
    initialize Cage(h);
    set total[1...h] to 0;
    for q := 0 to h − 1 do
        initialize y-axis key hexagon(q);
        ExpandLeftPart(LeftInitPos(q), InitBdrHexgns(q))
    od
end
```

Figure 5.9: The core of the algorithm

This algorithm has been implemented for IBM PC compatible computers in Modula–2. The program consists of five modules and more than 1900 bruto program lines. It has been used to determine the number of all properly placed hexagonal systems with $h \leqslant 17$.

Enumerating all properly placed hexagonal systems with 17 hexagons is a very lengthy process. This is why we had to divide the task into several smaller tasks (as the matter of fact, we had divided the enumeration process for $h = 17$ into 197 smaller tasks), which made it possible to run the program on several sites: Novi Sad, Ottawa and Trondheim. The parallelization was performed by hand, and according to two natural criteria: the coordinate of the y-axis key hexagon and, since this was too coarse, according to the initial piece of the boundary.

Bibliography

[1] Balaban A. T., Harary F., *Chemical Graphs, Enumeration and Proposed Nomenclature of Benzenoid Cata-Condensed Polycyclic Aromatic Hydrocarbons*, Tetrahedron 24 (1968), 2505–2516

[2] Gutman I. (Ed.), *Mathematical Methods in Chemistry*, Prijepolje, 2006

[3] Knop J. V., Szymanski K., Jeričević O., Trinajstić N., *Computer Enumeration and Generation of Benzenoid Hydrocarbons and Identification of Bay Regions*, J. Comput. Chem. 4(1983), 23–32

[4] van Lint J. H., Wilson R. M., *A Course in Combinatorics*, 2nd Ed., Cambridge University Press, 2001

[5] Pemmaraju S., Skiena S., *Computational Discrete Mathematics*, Cambridge University Press, 2003

[6] Roberts F. S., Tesman B., *Applied Combinatorics*, 2nd Ed., Pearson Education Inc., 2005

[7] Tošić R., Mašulović D., Stojmenović I., Brunvoll J., Cyvin B. N., Cyvin S. J., *Enumeration of Polyhex Hydrocarbons to h = 17*, J. Chem. Inf. Comput. Sci., 35(2) (1995) 181–187

[8] West D. B., *Introduction to Graph Theory*, 2nd Ed., Prentice Hall, 2001

Project: 06SER02/02/003

Informatika u hemiji

Dr Dragan Mašulović

Glava 1

Uvod

Najbolji način da se napravi uvod u probleme kojima ćemo se baviti u ovom nizu predavanja je da se citira deo Predgovora iz [2]:

"Jedna od najvećih dobrobiti koju imamo od hemije je mogućnost da se sintetišu supstance koje su sposobne da leče bolesti ili bar olakšaju patnje onih koji trpe bol. Kada želimo da sintetišemo novo hemijsko jedinjenje koje ima bolje osobine od onih koje već poznajemo, standardna procedura se sastoji u tome da se identifikuju i testira čitav niz kandidata. Sve do nedavno, standardna procedura je podrazumevala da se kandidati sintetišu jedan po jedan i potom da se proveri da li poseduju željena svojstva, i da li moža imaju neka od neželjenih svojstava. Takvi eksperimenti su veoma skupi i traju veoma dugo.

"Ako imamo neku predstavu o strukturnik zahtevima koje jedinjenje za kojim tragamo treba da ispunjava (a takvu predstavu obično imamo), sada smo u mogućnosti da generišemo kombinatornu biblioteku koju čine strukturne formule jedinjenja koja se javljaju kao kandidati, i da potom analiziramo virtuelna jedinjenja upotrebom brzih algoritama. Ovim jeftinim pristupom koji se ne zasniva na laboratorijskim testovima moguće je pretražiti mnogo veći raspon kandidata i tako svesti veliki broj mogućnosti na svega nekoliko razumnih, koji će potom biti testirani u laboratoriji, upotrebom standardnih procedura."

U ovom nizu predavanja prvo uvodimo matematički aparat koji je pogodan za opisivanje i rezonovanje o strukturama. Razmatraćemo standardne numeričke karakteristike koje se mogu dodeliti strukturama, kao što su valentnost elementa strukture i rastojanje elemenata strukture. Potom ćemo razmatrati reprezentacije struktura, pri čemo ćemo na umu stalno imati implementaciju na računaru, a na kraju ćemo demonstrirati nekoliko jednostavnih algoritama za analiziranje struktura.

Posebnu pažnju ćemo posvetiti aciklčnim strukturama, reprezentacijama acikličnih struktura (i ovaj put sa računarskom implementacijom na umu), a posebnu pažnju ćemo posvetiti Prüferovom kodu povezane aciklične strukture. Razmatraćemo i pokrivajuća stabla strukture, a kao primenu svega rečenog, prodiskutovaćemo monocikličnе strukture.

Potom prelazimo na napredne osobine struktura. Razmatraćemo Kekuléove strukture i brojanje Kekuléovih struktura date strukture. Nakon razmatranja nekih posebnih slučajeva dokazujemo dobro poznatu opštu formulu za broj Kekuléovih struktura: $K(G) = K(G-e) + K(G-(e))$.

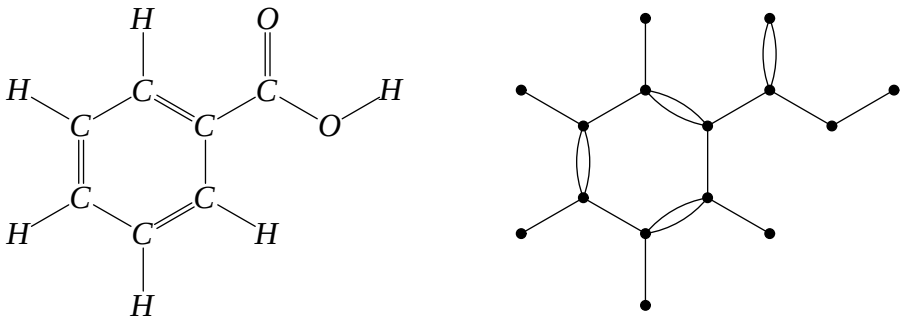
Nakon ovih razmatranja, prikazujemo osnove matematičkog aparata koji je namenjen formalizovanju pojam simetrije u strukturama. Prvo razmatramo brojanje svih struktura prema brojanju neizomorfnih struktura. Osnovni alat za brojanje meizomorfnih simetričnih struktura je Cauchy-Frobeniusova Lema koju izvodimo u opštem obliku, a potom je primenjujemo na neke posebne slučajeve.

Kao studiju slučaja razmatramo problem renerisanja i prebrajanja heksagonalnih sistema, gde se sve ranije navedene ideje koriste kako bi se razvio efikasan algoritam za klasifikovanje planarnih, prosto povezanih heksagonalnih sistema.

Glava 2

Grafovi kao modeli struktura

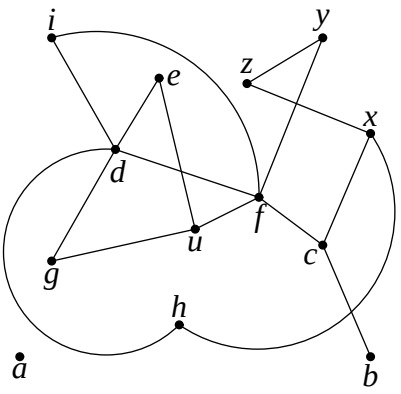
Grafovi predstavljaju jedno od najpopularnijih sredstava za modelovanje diskretnih fenomena kod kojih je potrebno predstaviti informaciju o tome da li su izvesni objekti u vezi ili ne, na primer, raskrsnice i ulice u nekom gradu, ili autoputevi i gradovi u nekoj državi. Ili, na primer, strukturna formula hemijskog jedinjenja, kao što je $C_7H_6O_2$:



2.1 Grafovi

Graf je uređeni par $G = (V,E)$, gde je V neprazan konačan skup, a E proizvoljan podskup skupa $V^{(2)} = \big\{\{u,v\} \subseteq V : u \neq v\big\}$. Elemente skupa V zovemo *čvorovi* grafa G, dok elemente skupa E zovemo *grane* grafa G. Često ćemo skup čvorova i skup grana grafa G označavati sa $V(G)$ i $E(G)$, a broj čvorova i broj grana grafa G sa $n(G)$ i $m(G)$.

Grafovi se zovu grafovi zbog veoma prirodne grafičke reprezentacije koja je postala uobičajen način komunikacije među ljudima koji se bave grafovima. Čvorove predstavljamo kružićima u ravni, a grane glatkim krivim bez samopreseka koje spajaju odgovarajuće čvorove.



$\delta(a) = 0$ (izolovani čvor)
$\delta(b) = 1$ (viseći čvor)
$\delta(c) = 3$
$\delta(d) = 5$
$\delta(e) = 2$
$\delta(f) = 5$
$\delta(g) = 2$
$\delta(h) = 2$
$\delta(i) = 2$
$\delta(u) = 3$
$\delta(x) = 3$
$\delta(y) = 2$
$\delta(z) = 2$

$\delta(G) = 0$ $\Delta(G) = 5$

Slika 2.1: Primer grafa

Primer 2.1 Na Sl. 2.1 je prikazan graf $G = (V,E)$, gde je

$$\begin{aligned} V &= \{a,b,c,d,e,f,g,h,i,u,x,y,z\} \\ E &= \big\{\{b,c\},\{c,x\},\{c,f\},\{h,x\},\{z,x\},\{f,y\},\{y,z\},\{f,u\},\{f,d\},\{f,i\}, \\ &\qquad \{e,u\},\{g,u\},\{d,e\},\{d,g\},\{d,h\},\{d,i\}\big\} \\ n &= 13 \\ m &= 16. \end{aligned}$$

Ako je $e = \{u,v\}$ grana grafa, kažemo da su u i v *susedni*, i da je grana e *incidentna* sa u i v. Takođe kažemo da je čvor u *sused* čvora v. *Stepen čvora* v, u oznaci $\delta_G(v)$ jednak je broju grana koje su incidentne sa v. Ako je jasno o kom grafu se radi, pišemo prosto $\delta(v)$. Sa $\delta(G)$ označavamo najmanji, a sa $\Delta(G)$ najveći stepen čvora u grafu G. Za čvor stepena 0 kažemo da je *izolovan*, a za čvor stepena 1 da je *viseći čvor u* G. Za čvor kažemo da je *paran*, odnosno *neparan* u zavisnosti od toga da li je $\delta(v)$ paran ili neparan broj. Graf je *regularan* ako je $\delta(G) = \Delta(G)$. Drugim rečima, graf je regularan ako svi njegovi čvorovi imaju isti stepen. Videti Sl. 2.1.

Teorema 2.2 (Prva teorema teorije grafova) *Ako je $G = (V,E)$ graf sa m grana, onda je* $\sum_{v \in V} \delta(v) = 2m$.

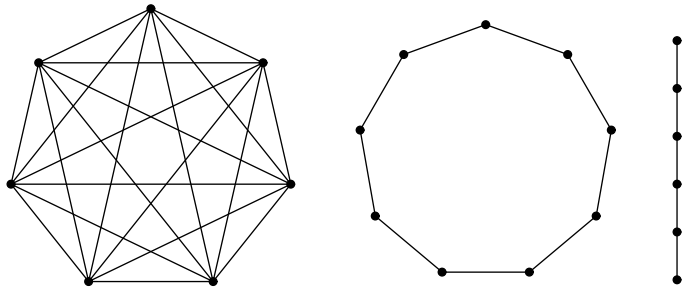
Dokaz. Zato što je svaka grana incidentna sa dva čvora, svaka grana je prebrojana dva puta u sumi na levoj strani jednakosti. □

Posledica 2.3 *U svakom grafu broj neparnih čvorova je paran.*

Teorema 2.4 *Ako je $n(G) \geqslant 2$, onda postoje čvorovi $v, w \in V(G)$ takvi da je $v \neq w$ i $\delta(v) = \delta(w)$.*

Dokaz. Neka je $V(G) = \{v_1, \ldots, v_n\}$ i neka je $\delta(v_i) \neq \delta(v_j)$ za $i \neq j$. Bez umanjenja opštosti možemo pretpostaviti da je $\delta(v_1) < \delta(v_2) < \ldots < \delta(v_n)$. Pošto imamo samo n mogućnosti za stepen čvora (0, 1, …, $n-1$) sledi da je $\delta(v_1) = 0$, $\delta(v_2) = 1$, …, $\delta(v_n) = n-1$. No, tada je čvor v_n susedan sa svim ostalim čvorovima u grafu, uključujući izolovani čvor v_1. Kontradikcija. □

Graf $H = (W, E')$ je *podgraf* grafa $G = (V, E)$, pišemo $H \leqslant G$, ako je $W \subseteq V$ i $E' \subseteq E$. Podgraf H grafa G je *pokrivajući podgraf grafa* G ako je $W = V(G)$. Graf H je *indukovani podgraf* grafa G ako je $E' = E \cap W^{(2)}$. Indukovani podgraf ćemo označavati sa $G[W]$. Grane indukovanog podgrafa grafa G su sve grane grafa G čija oba kraja pripadaju skupu W. Skup čvorova $W \subseteq V(G)$ je *nezavisan* ako je $E(G[W]) = \varnothing$, tj. među čvorovima iz W ne postoje dva koji su susedni u G. Sa $\alpha(G)$ označavamo maksimalnu kardinalnost nezavisnog skupa čvorova u G. Ako su $A, B \subseteq V(G)$ disjunktni, sa $E(A, B)$ označavamo skup svih grana čiji jedan kraj je u A, a drugi u B.



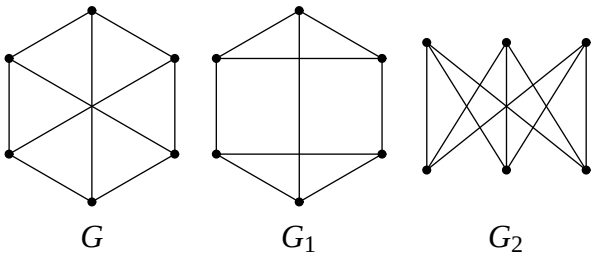
Slika 2.2: K_7, C_9 i P_6

Kompletan graf sa n čvorova (ili *n-klika*) je graf sa n čvorova kod koga su svaka dva različita čvora susedni. Kompletan graf sa n čvorova označavamo sa K_n. *Kontura sa n čvorova*, u oznaci C_n, je graf sa n čvorova kod koga je prvi čvor susedan sa drugim, drugi sa trećim, i tako dalje, poslednji čvor je susedan sa prvim. *Put sa n čvorova*, u oznaci P_n, je graf kod koga je prvi čvor susedan sa drugim, drugi sa trećim, i tako dalje, pretposlednji susedan sa poslednjim, a poslednji *nije* susedan sa prvim. Kažemo da put sa n čvorova ima *dužinu* $n-1$. Na Sl. 2.2 su prikazani K_7, C_9 i P_6.

Teorema 2.5 *Ako je $\delta(G) \geqslant 2$, graf G sadrži konturu.*

Dokaz. Neka je $x_1 \ldots x_{k-1}\, x_k$ najduži put u G. Pošto je $\delta(x_k) \geqslant \delta(G) \geqslant 2$, čvor x_k ima suseda v koji nije x_{k-1}. Ako $v \notin \{x_1, \ldots, x_{k-2}\}$ onda je $x_1 \ldots x_{k-1}\, x_k\, v$ put koji je duži od najdužeg puta u G, što je nemoguće. Dakle, $v = x_j$ za neko $j \in \{1, \ldots, k-2\}$, pa su $x_j \ldots x_k$ čvorovi konture u G. □

Grafovi G_1 i G_2 su *izomorfni* ako postoji bijekcija $\varphi : V(G_1) \to V(G_2)$ takva da je $\{x, y\} \in E(G_1) \Leftrightarrow \{\varphi(x), \varphi(y)\} \in E(G_2)$. Za bijekciju φ kažemo da je *izomorfizam*. Pišemo $G_1 \cong G_2$. Na primer, grafovi G i G_2 na Sl. 2.3 su izomorfni, a G i G_1 nisu.



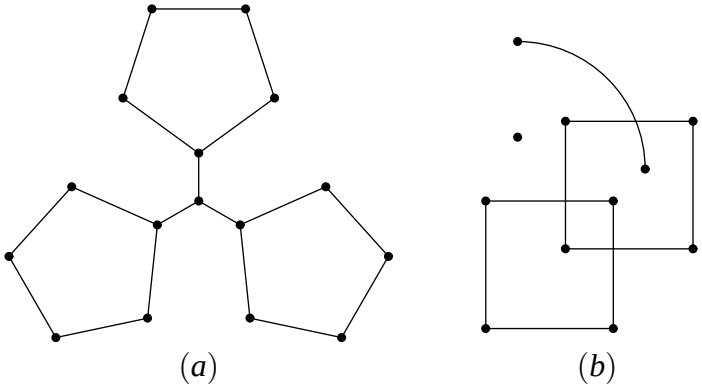
Slika 2.3: $G \cong G_2$, ali $G \ncong G_1$

Teorema 2.6 *Neka je $G_1 \cong G_2$ i neka je φ proizvoljan izomorfizam ta dva grafa. Tada je $n(G_1) = n(G_2)$, $m(G_1) = m(G_2)$ i $\delta_{G_1}(x) = \delta_{G_2}(\varphi(x))$ za svako $x \in V(G_1)$.*

2.2 Povezanost

Šetnja u grafu G je svaki niz čvorova i grana $v_0\ e_1\ v_1\ e_2\ v_2 \ldots v_{k-1}\ e_k\ v_k$ kod koga je $e_i = \{v_{i-1}, v_i\}$ za sve $i \in \{1, \ldots, k\}$. Primetimo da se jedan čvor ili grana može više puta pojaviti u šetnji. Kažemo da je *k dužina* šetnje. Ako je $v_0 \neq v_k$ kažemo da *šetnja povezuje v_0 sa v_k*. *Zatvorena šetnja* je šetnja $v_0\ e_1\ v_1\ \ldots\ v_{k-1}\ e_k\ v_k$ kod koje je $v_0 = v_k$. Primetimo da je put šetnja kod koje se ne ponavljaju ni čvorovi ni grane, a kontura je zatvorena šetnja kod koje se ne ponavljaju ni čvorovi ni grane, osim prvog čvora koji se pojavljuje na kraju šetnje.

Na skupu $V(G)$ definišemo binarnu relaciju θ ovako: $x\theta y$ ako $x = y$ ili postoji šetnja koja spaja x sa y. Jasno je da je θ relacija ekvivalencije na $V(G)$ i stoga razbija skup $V(G)$ na blokove $S_1, \ldots, S_t$. Ove blokove, ili odgovarajuće indukovane podgrafove (kako nam kad više odgovara) zovemo *komponente povezanosti* grafa G. Broj komponenti povezanosti grafa G označavamo sa $\omega(G)$. Graf G je *povezan* ako je $\omega(G) = 1$. Primer povezanog grafa i primer grafa sa četiri komponente povezanosti dati su na Sl. 2.4.



Slika 2.4: (a) Povezan graf; (b) Graf kod koga je $\omega = 4$

Lema 2.7 *Komponente povezanosti grafa su maksimalni povezani podgrafovi tog grafa. Preciznije, $S \subseteq V(G)$ je komponenta povezanosti grafa G ako i samo ako ne postoji pravi nadskup $S' \supset S$ koji indukuje povezan podgraf grafa G.*

Teorema 2.8 *Graf G je povezan ako i samo ako $E(A,B) \neq \varnothing$ za svaku particiju $\{A,B\}$ skupa $V(G)$.*

Dokaz. ($\Leftarrow$) Pretpostavimo da G nije povezan graf i neka su $S_1, \ldots, S_\omega$, $\omega \geqslant 2$, njegove komponente povezanosti. Tada Lema 2.7 implicira $E(S_1, \bigcup_{j=2}^{\omega} S_j) = \varnothing$.

($\Rightarrow$) Neka je G povezan graf i $\{A,B\}$ proizvoljna particija skupa $V(G)$. Uzmimo proizvoljne $a \in A$ i $b \in B$. Obzirom da je G povezan, postoji šetnja $x_1\ e_2\ x_2\ \ldots\ e_k\ x_k$ koja povezuje a sa b. Zbog $x_1 = a$ i $x_k = b$, mora postojati j sa osobinom $x_j \in A$ i $x_{j+1} \in B$, odakle $e_{j+1} \in E(A,B)$, pa $E(A,B) \neq \varnothing$. □

Rastojanje $d_G(x,y)$ čvorova x i y povezanog grafa G definišemo ovako:

$$d_G(x,x) = 0,$$
$$d_G(x,y) = \min\{k : \text{ postoji put dužine } k \text{ koji spaja } x \text{ sa } y\}, \text{ za } x \neq y.$$

Teorema 2.9 *Neka je $G = (V,E)$ povezan graf. Tada je (V,d_G) metrički prostor, tj. za sve $x,y,z \in V$ zadovoljeno je sledeće:*

(D1) $d_G(x,y) \geqslant 0$;

(D2) $d_G(x,y) = 0$ *ako i samo ako* $x = y$;

(D3) $d_G(x,y) = d_G(y,x)$; *i*

(D4) $d_G(x,z) \leqslant d_G(x,y) + d_G(y,z)$.

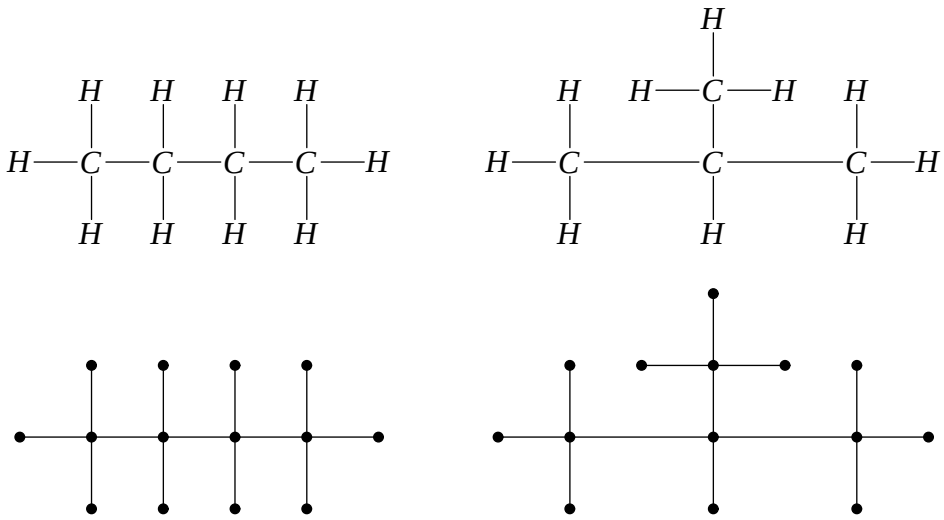
Ako se graf G podrazumeva, umesto $d_G(x,y)$ pišemo samo $d(x,y)$. *Dijametar* povezanog grafa G, u oznaci $d(G)$, je najveće rastojanje dva čvora tog grafa:

$$d(G) = \max\{d(x,y) : x,y \in V(G)\}.$$

Primer 2.10 (a) $d(G) = 1$ ako i samo ako je G kompletan graf.
(b) $d(P_n) = n-1$ i $d(C_n) = \left\lfloor \dfrac{n-1}{2} \right\rfloor$.

2.3 Stabla i monociklične strukture

Istorijski posmatrano, prva upotreba grafova kao modela hemijskih jedinjenja se pojavila 1889. godine kada je Arthur Cayley rešavao problem broja izomera ugljovodonika. Matematički model ugljovodonika zovemo *stablo* iz očiglednih razloga:



Formlano posmatrano, *stablo* je povezan graf koji ne sadrži konture. Nije teško videti da je stablo *minimalan povezan graf na datom skupu čvorova*. Naredna teorema pokazuje da stabla na izvestan način opisuju esenciju pojma povezanosti. Podsetimo se da je graf $H = (W, E')$ pokrivajući podgraf grafa $G = (V, E)$ ako je $W = V$ i $E' \subseteq E$. Ako je uz to H još i stablo, kažemo da je *H pokrivajuće stablo grafa G*.

Teorema 2.11 *Graf je povezan ako i samo ako ima pokrivajuće stablo.*

Dokaz. Ako graf ima pokrivajući podgraf koji je povezan, onda i polazni graf mora biti povezan. Dakle, ako graf ima pokrivajuće stablo, on je povezan. Da bismo pokazali obrnuto tvrđenje, posmatrajmo proizvoljan povezan graf G i konstruišimo niz grafova G_0, G_1, G_2, ... kako sledi: $G_0 = G$; ako G_i ima konturu, uzmimo proizvoljnu granu e_i koja leži na nekoj konturi i stavimo $G_{i+1} = G_i - e_i$; u suprotnom stavimo $G_{i+1} = G_i$. Svaki od grafova G_i je pokrivajući podgraf grafa G i svaki od grafova G_i je povezan zato što grana koja pripada konturi ne može biti most. Štaviše, ako je $G_i = G_{i+1}$ onda je $G_i = G_j$ za sve $j > i$. Neka je m broj grana grafa G. Zato što iz grafa ne možemo ukloniti više od m grana, zaključujemo da je $G_{m+1} = G_{m+2}$. Po konstrukciji to znači da graf G_{m+1} nema konture. Dakle, graf G_{m+1} je pokrivajuće stablo grafa G. □

Sada ćemo pokazati da svako stablo sa n čvorova ima $n-1$ grana i da svake dve od sledeće tri osobine impliciraju onu treću:

- graf je povezan,
- graf nema konture, i
- $m = n-1$.

Lema 2.12 *Stablo sa bar dva čvora ima bar dva viseća čvora.*

Dokaz. Neka je G stablo sa $n \geqslant 2$ čvorova i neka je $v_1\, v_2\, \ldots\, v_k$ najduži put u stablu. Tada je $k \geqslant 2$ jer je G povezan graf sa bar dva čvora. Ako je $\delta(v_1) > 1$ onda v_1 ima suseda x koji nije v_2. Ako

bi x bio novi čvor, tj. $x \notin \{v_3, \ldots, v_k\}$, tada bi put $x\, v_1\, v_2\, \ldots\, v_k$ bio duži od najdužeg puta u G, što nije moguće. S druge strane, ako je $x \in \{v_3, \ldots, v_k\}$ onda G ima konturu, što je u suprotnosti sa pretpostavkom da je G stablo. Dakle, v_1 je viseći čvor. Na isti način pokazujemo da je i v_k viseći čvor. □

Teorema 2.13 *Neka je $G = (V,E)$ stablo sa n čvorova i m grana. Tada je $m = n-1$, i shodno tome $\sum_{v \in V} \delta(v) = 2(n-1)$.*

Dokaz. Drugi deo tvrđenja je direktna posledica Prve teoreme teorije grafova. Dokažimo sada da je $m = n-1$. Dokaz provodimo indukcijom po n. Slučajevi $n = 1$ i $n = 2$ su trivijalni. Pretpostavimo da je tvrđenje tačno za sva stabla sa manje od n čvorova i posmatrajmo stablo G sa n čvorova. Prema Lemi 2.12 stablo G ima viseći čvor x. Lako se vidi da je stepen artikulacionog čvora je najmanje 2. Zato x nije artikulacioni čvor grafa G, pa je $G-x$ povezan. Jasno je i to da $G-x$ nema konture (uklanjanje čvorova ne može da dovede do toga da se u grafu pojave konture), odakle sledi da je $G-x$ stablo sa manje od n čvorova. Prema induktivnoj hipotezi, $m' = n'-1$, gde je $m' = m(G-x)$ i $n' = n(G-x)$. Kako je $m' = m-1$ i $n' = n-1$ (x je viseći čvor), zaključujemo da je $m = n-1$. □

Teorema 2.14 *Neka je G graf sa n čvorova i m grana. Ako je $m = n-1$ i ako G nema konture onda je G povezan (i stoga stablo).*

Dokaz. Neka je $m = n-1$, neka G nema konture i neka su $S_1, \ldots, S_\omega$ komponente povezanosti grafa G. Svaka komponenta je stablo, pa je $m_i = n_i - 1$ za sve i, gde je $m_i = m(S_i)$ i $n_i = n(S_i)$. Odatle je $\sum_{i=1}^{\omega} m_i = \sum_{i=1}^{\omega} n_i - \omega$ tj. $m = n - \omega$ (jer je $m = \sum_{i=1}^{\omega} m_i$ i $n = \sum_{i=1}^{\omega} n_i$). Iz pretpostavke $m = n-1$ sada lako zaključujemo da je $\omega = 1$, tj. G je povezan graf. □

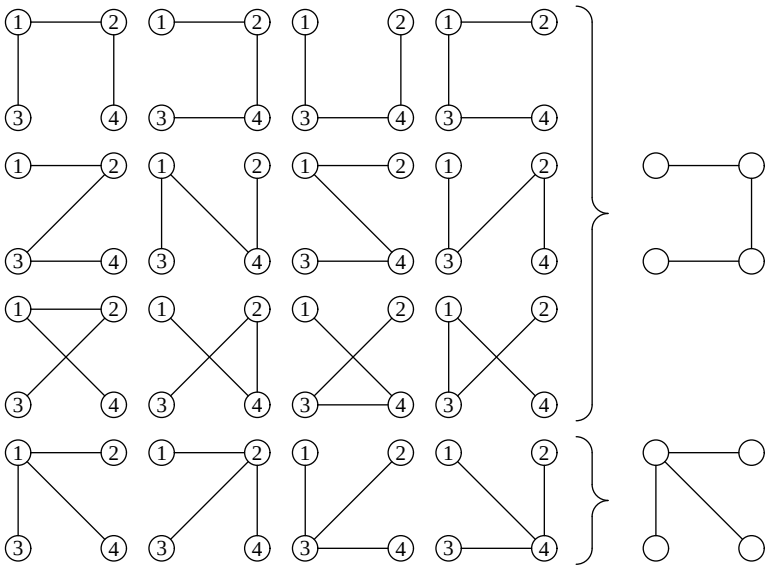
Teorema 2.15 *Neka je G povezan graf sa $n \geqslant 2$ čvorova i m grana i neka je $m = n-1$. Tada G nema konture (i stoga je G stablo).*

Dokaz. Prema Teoremi 2.11 graf $G = (V,E)$ ima pokrivajuće stablo $H = (V,E')$. Zato što je H stablo, Teorema 2.13 nam daje $m(H) = n(H) - 1 = n-1$. Pretpostavka $m = n-1$ sada implicira $m(H) = m$ pa iz $E' \subseteq E$ zaključujemo da je $E' = E$. Dakle, $G = H$ i tako je G stablo. □

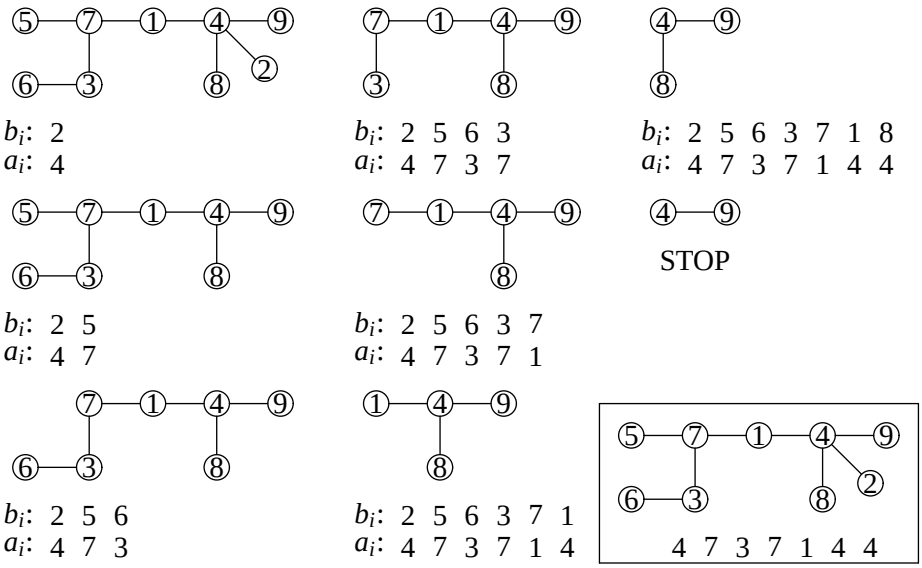
Posledica 2.16 *Povezan graf sa n čvorova i m grana je stablo ako i samo ako je $m = n-1$.*

Odeljak o stablima ćemo zaključiti značajnim rezultatom o broju različitih stabala na datom skupu čvorova. Važno je napomenuti da kada brojimo strukture možemo brojati *različite* strukture, ili *neizomorfne* strukture. Na primer, postoji 16 *različitih* stabala na skupu od četiri elementa, a samo dva *neizomorfna*, kako je to pokazano na Sl. 2.5. Ne treba da nas iznenadi to što je brojanje neizomorfnih struktura daleko komplikovanije od brojanja različitih struktura.

Teorema 2.17 (Cayley 1889) *Broj različitih stabala sa n čvorova je n^{n-2}.*



Slika 2.5: Šesnaest različitih i samo dva neizomorfna stabla na skupu od četiri čvora



Slika 2.6: Prüferov kod stabla

Dokaz. Neka je $V = \{1, \ldots, n\}$ konačan skup koji ćemo koristiti kao skup čvorova. Ideja dokaza je da se svako stablo sa skupom čvorova V kodira nizom $(a_1, \ldots, a_{n-2})$ elemenata skupa V i da se pokaže da je na taj način dobijena bijekcija $\varphi : \mathscr{T}_n \to \{1, 2, \ldots, n\}^{n-2}$, gde je $\mathscr{T}_n$ skup svih stabala sa skupom čvorova V.

Prvo ćemo pokazati kako se konstruiše Prüferov kod stabla. Neka je T stablo sa skupom čvorova V. Konstruisaćemo niz stabala $(T_i)_{1 \leqslant i \leqslant n-2}$ i dva niza celih brojeva, Prüferov kod $(a_i)_{1 \leqslant i \leqslant n-2}$ i pomoćni niz $(b_i)_{1 \leqslant i \leqslant n-2}$. Neka je $T_1 = T$. Ako nam je dato stablo T_i, neka je b_i najmanji viseći čvor stabla T_i (čvorovi stabla su celi brojevi, pa od svih brojeva koji se javljaju kao viseći čvorovi stabla odaberemo najmanji) i neka je a_i njegov jedini sused. Sada stavimo $T_{i+1} = T_i - b_i$ i ponovimo postupak dok ne dobijemo stablo sa samo dva čvora. Prüferov kod polaznog stabla je $(a_1, a_2, \ldots, a_{n-2})$. Primer konstrukcije koda dat je na Sl. 2.6. Tako smo dobili funkciju $\varphi : \mathscr{T}_n \to \{1, \ldots, n\}^{n-2}$ koja svakom stablu dodeljuje njegov Prüferov kod.

Obrnuto, ako nam je dat neki niz $(a_1, \ldots, a_{n-2})$ celih brojeva iz skupa V odgovarajuće stablo konstruišemo na sledeći način. Za $S \subseteq \{1, \ldots, n\}$ neka je sa

$$\operatorname{mix} S = \min(\{1, \ldots, n\} \setminus S)$$

označen najmanji broj koji nije u S (*mi*nimal e*x*cluded). Stavimo $a_{n-1} = n$ i konstruišemo b_1, b_2, …, b_{n-1} ovako:

$$b_i = \operatorname{mix}\{a_i, \ldots, a_{n-1}, b_1, \ldots, b_{i-1}\}$$

(za $i = 1$ u skupu nema nijednog elementa b_j, naravno). Na primer, u slučaju niza $(4,7,3,4,1,4,4)$ imamo da je $a_8 = 9$ i

$$\begin{aligned}
b_1 &= \operatorname{mix}\{4,7,3,4,1,4,4,9\} = 2\\
b_2 &= \operatorname{mix}\{\ \ 7,3,4,1,4,4,9,2\} = 5\\
b_3 &= \operatorname{mix}\{\ \ \ \ 3,4,1,4,4,9,2,5\} = 6\\
b_4 &= \operatorname{mix}\{\ \ \ \ \ \ 4,1,4,4,9,2,5,6\} = 3\\
b_5 &= \operatorname{mix}\{\ \ \ \ \ \ \ \ 1,4,4,9,2,5,6,3\} = 7\\
b_6 &= \operatorname{mix}\{\ \ \ \ \ \ \ \ \ \ 4,4,9,2,5,6,3,7\} = 1\\
b_7 &= \operatorname{mix}\{\ \ \ \ \ \ \ \ \ \ \ \ 4,9,2,5,6,3,7,1\} = 8\\
b_8 &= \operatorname{mix}\{\ \ \ \ \ \ \ \ \ \ \ \ \ \ 9,2,5,6,3,7,1,8\} = 4
\end{aligned}$$

Ovaj proces se zove *procedura rekonstrukcije* zato što, kao što ćemo pokazati, ona određuje stablo čiji Prüferov kod je $(a_1, \ldots, a_{n-2})$.

Pokažimo, prvo, da je $\{\{b_i, a_i\} : 1 \leqslant i \leqslant n\}$ skup grana nekog stabla. Ako je $i < j$ onda je, po konstrukciji, $b_j = \operatorname{mix}\{a_j, \ldots, a_{n-1}, b_1, \ldots, b_i, \ldots, b_{j-1}\}$, pa $b_j \neq b_i$. Vidimo da su svi b_i različiti i manji od $n = a_{n-1}$. Dakle, $\{b_1, \ldots, b_{n-1}\} = \{1, \ldots, n-1\}$ i zato je $\{b_1, \ldots, b_{n-1}, a_{n-1}\} = \{1, \ldots, n-1, n\}$. Štaviše, ako je $i \leqslant j$ onda $a_j \notin \{b_1, \ldots, b_j\}$ zbog $b_i = \operatorname{mix}\{a_i, \ldots, a_j, \ldots, a_{n-1}, b_1, \ldots, b_{i-1}\}$. Zato $\{b_1, \ldots, b_{n-1}, a_{n-1}\} = \{1, \ldots, n-1, n\}$ implicira da je $a_j \in \{b_{j+1}, \ldots, b_{n-1}, a_{n-1}\}$. Da rezimiramo,

$$\begin{aligned}
&a_j \in \{b_{j+1}, b_{j+2}, \ldots, b_{n-1}, a_{n-1}\} \text{ i}\\
&b_j \notin \{a_{j+1}, b_{j+1}, a_{j+2}, b_{j+2}, \ldots, a_{n-1}, b_{n-1}\},
\end{aligned} \qquad \text{za sve } j. \qquad (\star)$$

Da bismo konstruisali graf, poći ćemo od grane $\{b_{n-1}, a_{n-1}\}$ i potom dodavati grane $\{b_{n-2}, a_{n-2}\}$, $\{b_{n-3}, a_{n-3}\}$, ..., $\{b_1, a_1\}$ jednu po jednu. Zbog $(\star)$ lako se vidi da u svakom koraku grafu dodajemo novi čvor b_i i novu granu $\{b_i, a_i\}$ koja spaja b_i sa već postojećim čvorom. Zato je graf koji se na kraju dobija povezan, a povezan graf sa n čvorova i $n-1$ grana je stablo (Posledica 2.16). Dakle, sada imamo i funkciju $\psi : \{1, \ldots, n\}^{n-2} \to \mathscr{T}_n$ koja uzima proizvoljan kod i dodeljuje mu stablo.

Da bismo završili dokaz dovoljno je da pokažemo da je ψ inverzna funkcija funkcije φ, tj. da je $\varphi \circ \psi = \mathrm{id}$ i $\psi \circ \varphi = \mathrm{id}$. Ovde ćemo dokazati samo $\psi \circ \varphi = \mathrm{id}$ tj. $\psi(\varphi(T)) = T$ za sve $T \in \mathscr{T}_n$, dok drugu jednakost ostavljamo za vežbu. Za čvor v stabla T kažemo da je *unutrašnji čvor stabla* ako je $\delta_T(v) > 1$. Neka je $\mathrm{int}(T)$ skup svih unutrašnjih čvorova stabla T.

Uzmimo proizvoljno stablo $T \in \mathscr{T}_n$, neka je $(a_1, \ldots, a_{n-2})$ njegov Prüferov kod i neka je $(b_1, \ldots, b_{n-2})$ odgovarajući pomoćni niz. Na kraju procedure konstrukcije Prüferovog koda u stablu ostaju samo dva čvora, čvor $a_{n-1} = n$ i njegov sused koga ćemo označiti sa b_{n-1}. Polazeći od niza $(a_1, \ldots, a_{n-1})$ procedura rekonstrukcije određuje niz celih brojeva $b'_1, \ldots, b'_{n-1}$. Pokazaćemo da je $b_i = b'_i$ za sve i. Jasno je da možemo pretpostaviti da je $n \geqslant 3$.

Zato što je b_1 susedan sa a_1 u T i zbog $n \geqslant 3$, čvor a_1 ne može biti viseći čvor stabla T, tako da je $a_1 \in \mathrm{int}(T)$. Na isti način pokazujemo da je $a_2 \in \mathrm{int}(T - b_1)$, $a_3 \in \mathrm{int}(T - b_1 - b_2)$, i uopšte, $a_{i+1} \in \mathrm{int}(T - b_1 - \ldots - b_i)$. Zbog $\mathrm{int}(T - v) \subseteq \mathrm{int}(T)$ kadgod je v viseći čvor stabla T i $n(T) \geqslant 2$, zaključujemo da je

$$\mathrm{int}(T - b_1 - \ldots - b_i) = \{a_{i+1}, \ldots, a_{n-2}\}.$$

Posebno, $\mathrm{int}(T) = \{a_1, \ldots, a_{n-2}\}$. Ako stablo ima bar dva čvora, svaki čvor tog stabla je ili viseći čvor ili unutraršnji čvor, odakle sledi da je

$$V(T - b_1 - \ldots - b_i) \setminus \mathrm{int}(T - b_1 - \ldots - b_i)$$

skup svih visećih čvorova stabla $T - b_1 - \ldots - b_i$. Kako je $V(T - b_1 - \ldots - b_i) = \{1, \ldots, n\} \setminus \{b_1, \ldots, b_i\}$ i $\mathrm{int}(T - b_1 - \ldots - b_i) = \{a_{i+1}, \ldots, a_{n-2}\}$, lako se vidi da je

$$\begin{aligned}\Big(\{1, \ldots, n\} \setminus \{b_1, \ldots, b_i\}\Big) \setminus \{a_{i+1}, \ldots, a_{n-2}\} = \\ = \{1, \ldots, n\} \setminus \{a_{i+1}, \ldots, a_{n-2}, b_1, \ldots, b_i\}\end{aligned}$$

skup visećih čvorova stabla $T - b_1 - \ldots - b_i$. Sada se indukcijom po i pokazuje da je $b_i = b'_i$. Kao što smo videli, b_1 je viseći čvor stabla T, pa je $b_1 \in \{1, \ldots, n\} \setminus \{a_1, \ldots, a_{n-2}\}$. Međutim, b_1 je najmanji takav broj, odakle je $b_1 = \min(\{1, \ldots, n\} \setminus \{a_1, \ldots, a_{n-2}\}) = \mathrm{mix}\{a_1, \ldots, a_{n-2}\} = b'_1$. Pretpostavimo sada da je $b_j = b'_j$ za sve $j \in \{1, \ldots, i\}$. Čvor b_{i+1} je najmanji viseći čvor u $T - b_1 - \ldots - b_i$, pa koristeći induktivnu hipotezu dobijamo

$$\begin{aligned}b_{i+1} &= \min(\{1, \ldots, n\} \setminus \{a_{i+1}, \ldots, a_{n-2}, b_1, \ldots, b_i\}) \\ &= \mathrm{mix}\{a_{i+1}, \ldots, a_{n-2}, b_1, \ldots, b_i\} = \mathrm{mix}\{a_{i+1}, \ldots, a_{n-2}, b'_1, \ldots, b'_i\} = b'_{i+1}.\end{aligned}$$

Dakle, $\{a_i, b_i\} = \{a_i, b'_i\}$ za sve i, pa se procedurom rekonstrukcije dobija baš stablo T. $\square$

Po strukturi veoma bliske stablima su *monociklične strukture* koje odgovaraju hemijskim jedinjenjima sa jednim benzenovim prstenom, kao što je $C_7H_6O_2$ (videti ilustraciju na str. 3).

Graf G je *monocikličan* ako je povezan i sadrži tačno jednu konturu. Obzirom na to da su monociklične strukture veoma bliske stablima (uklanjanjem bilo koje grane sa konture dobija se stablo), lako dolazimo do karakterizavije monocikličnih struktura koja odgovara karakterizaciji stabala. Svođenjem na stabla lako se može pokazati da svaka dva od naredna tri svojstva impliciraju ono treće:

- graf je povezan,
- graf ima ta čno jednu konturu, i
- $m = n$.

Teorema 2.18 (a) *Za svaki monociklični graf je* $n = m$.
(b) *Povezan graf koji zadovoljava* $m = n$ *sadrži tačno jednu konturu.*
(c) *Graf sa tačno jednom konturom koji zadovoljava* $m = n$ *je povezan.*

2.4 Kekuléove strukture

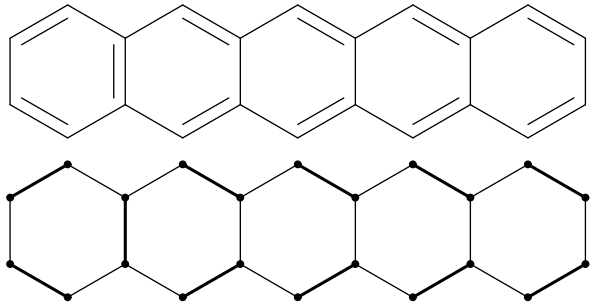
Kekuléova struktura predstavlja reprezentaciju aromatičnog molekula (kao što je benzen), sa alternirajućim jednostrukim i dvostrukim vezama, kod kojih se pretpostavlja da ne postoji interakcija među višestrukim vezama. Kekuléove strukture benzena, na primer, su:

Formalno posmatrano, model Kekuléove strukture čini skup grana odgovarajućeg grafa takav da ne postoje dve od tih grana koje su susedne, i sa dodatnom osobinom da je svaki čvor grafa incidentan sa tačno jednom granom is uočenog skupa grana. Skup grana grafa sa navedenim osobinama se naziva još i perfektan mečing. Dajemo sada precizne definicije.

Mečing grafa $G = (V, E)$ je skup grana $M \subseteq E$ tog grafa takav da za svake dve grane $e_1, e_2 \in M$ važi sledeće:

$$\text{ako je } e_1 \neq e_2 \text{ onda je } e_1 \cap e_2 = \varnothing.$$

Evo primera Kekuléove strukture i odgovarajućeg skupa grana u grafu jedinjenja:



Kažemo da je mečing M grafa G *maksimalan* ako za svaki drugi mečing M' u grafu imamo da iz $M' \supseteq M$ sledi $M' = M$. To znači da nijedan pravi nadskup skupa M nije mečing grafa G. Mečing M grafa G *maksimum* ako za svaki drugi mečing M' u grafu imamo $|M'| \leqslant |M|$. To znači nijedan drugi mečing grafa G nema veći broj grana. Konačno, mečing M grafa G je *perfektan* ako je svaki čvor grafa incidentan sa tačno jednom granom u M. Slika iznad ovog pasusa ilustruje jedan perfektan mečing.

Sada ćemo preći na diskusiju problema egzistencije mečinga u grafu. Lako se vidi da svaki graf ima maksimalan mečing i da ima maksimum mečing. Međutim, nemaju svi grafovi savršeni mečing.

Lema 2.19 *Ako graf G ima perfektan mečing onda je $n(G)$ paran broj.*

Dokaz. Neka je $M = \{e_1, \ldots, e_k\}$ perfektan mečing grafa G. Prema definiciji mečinga, različite grane u M su disjunktne, tako da je $\left|\bigcup_{i=1}^{k} e_i\right| = 2k$, što je paran broj. Za svaki mečing važi sledeće: $\bigcup_{i=1}^{k} e_i \subseteq E(G)$. Međutime, M je perfektan mečing, tako da mora biti $\bigcup_{i=1}^{k} e_i \subseteq E(G)$. Zato je $n(G) = 2k$, što je paran broj. □

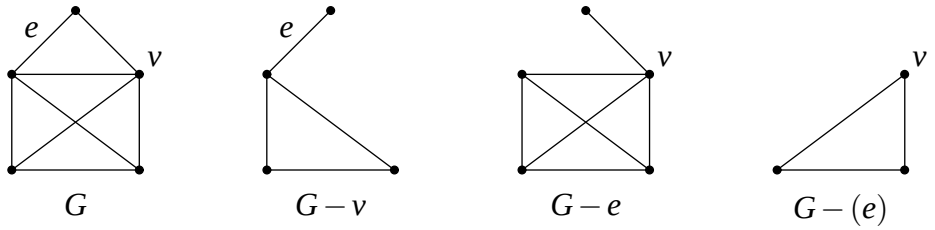
Tutte je 1947. godine dokazao veoma dubok rezultat koji daje potreban i dovoljan uslov da bi graf imao perfektan mečinkg. Neka je sa $\omega_{\text{odd}}(G)$ označen broj neparnih komponenti grafa G, pri čemu za komponentu $S \subseteq V(G)$ kažemo da je *neparna* ako je $|S|$ neparan ceo broj.

Teorema 2.20 (Tutte, 1947) *Graf G ima perfektan mečing ako i samo ako je*

$$\omega_{\text{odd}}(G - W) \leqslant |W|,$$

za sve podskupove $W \subseteq V(G)$.

Na kraju dajemo opštu formulu za broj perfektnih mečinga grafa G. Neka je $e = x,y$ grana grafa G, a v čvor u tom grafu. Sa $G - e$ označavamo graf koji se dobija od G uklanjanjem grane e, a sa $G - v$ označavamo graf koji se dobija od grafa G uklanjanjem čvora v i svih grana grafa G koje su incidentne sa v. Osim toga, sa $G - (e)$ označavamo graf $G - x - y$:



Teorema 2.21 *Neka je sa $K(G)$ označen broj različitih Kekuléovih struktura (= perfektnih mečinga) grafa G. Tada za svaku granu $e \in E(G)$ važi sledeće:*

$$K(G) = K(G-e) + K(G-(e)).$$

Dokaz. Neka je G graf i e njegova proizvoljna grana. Klasa svih perfektnih mečinga grafa G se na prirodan način može podeliti na dve disjunktne klase: jednu čine svi perfektni mečinzi grafa G koji sadrže granu e, a drugu perfektni mečinzi grafa G koji je ne sadrže. Prema tome, ako je sa $\mathscr{M}(G)$ označena klasa svih perfektnih mečinga grafa G, sa $\mathscr{M}_e(G)$ klasa svih perfektnih mečinga grafa G koji sadrže e, a sa $\mathscr{M}'_e(G)$ klasa svih perfektnih mečinga grafa G koji ne sadrže e, tada je

$$|\mathscr{M}(G)| = |\mathscr{M}_e(G)| + |\mathscr{M}'_e(G)|.$$

Osnovna ideja dokaza je da se pokaže da je $|\mathscr{M}_e(G)| = K(G-(e))$, dok je $|\mathscr{M}'_e(G)| = K(G-e)$.

Neka je M perfektan mečing grafa G takav da $e \notin M$. Tada je M perfektan mečing grafa $G-e$. S druge strane, svaki perfektan mečing grafa $G-e$ je i perfektan mečing grafa G, i zato je $|\mathscr{M}'_e(G)| = |\mathscr{M}(G-e)| = K(G-e)$.

Neka je sada M perfektan mečing grafa G takav da je $e \in M$ i neka je $e = \{x,y\}$. Kako e pripada mečingu M, čvorovi x i y su pokriveni granom e, pa nijedna druga grana koja je incidentna sa x ili y ne može pripadati mečingu M. Zato je, $M' = M \setminus \{e\}$ perfektan mečing grafa $G-(e)$. s druge strane, svaki perfektan mečing M grafa $G-(e)$ se može proširiti do perfektnog mečinga $M \cup \{e\}$ grafa G. Zato je $|\mathscr{M}_e(G)| = |\mathscr{M}(G-(e))| = K(G-(e))$, čime je dokaz završen. □

Glava 3

Implementacija na računaru

Postoji više načina da se graf reprezentuje u memoriji računara. U ovoj glavi ćemo pokazati i prodiskutovati dva:

- reprezentaciju preko matrice susedstva, i
- reprezentaciju preko liste suseda.

3.1 Matrica susedstva

Prvi način predstavljanja grafova koga ćemo pomenuti i koga ćemo najčešće koristiti se zasniva na tome da se graf predstavi preko svoje matrice susedstva. Neka je G graf sa skupom čvorova $\{1, 2, \ldots, n\}$. Tada je *matrica susedstva grafa G* (engl. *adjacency matrix*), u oznaci $A(G)$, matrica $[a_{ij}]$ formata $n \times n$ takva da je

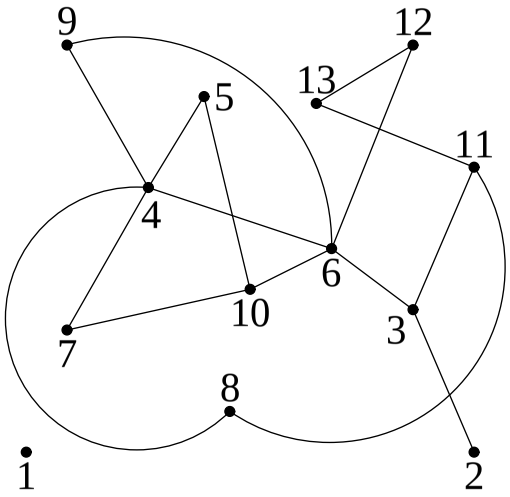
$$a_{ij} = \begin{cases} 0, & \text{čvorovi } i \text{ i } j \text{ nisu susedni,} \\ 1, & \text{čvorovi } i \text{ i } j \text{ su susedni.} \end{cases}$$

Ako je čvor u susedan sa čvorom v, onda je i čvor v susedan sa čvorom u, tako da je matrica $A(G)$ uvek simetrična. Primer je dat na Slici 3.1.

Grafove ćemo predstavljati kao slogove koji imaju dva polja: broj čvorova i matricu susedstva.

```
const
  MaxNoVertices = 50;
  Null = 0;

type
  Graph = record
    N : integer;
    adjacent : array [1 .. MaxNoVertices, 1 .. MaxNoVertices]
               of Boolean
  end;
```

$G =$



$A(G)$	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	1	0	0	0	0	1	0	0
4	0	0	0	0	1	1	1	1	1	0	0	0	0
5	0	0	0	1	0	0	0	0	0	1	0	0	0
6	0	0	1	1	0	0	0	0	1	1	0	1	0
7	0	0	0	1	0	0	0	0	0	1	0	0	0
8	0	0	0	1	0	0	0	0	0	0	1	0	0
9	0	0	0	1	0	1	0	0	0	0	0	0	0
10	0	0	0	0	1	1	1	0	0	0	0	0	0
11	0	0	1	0	0	0	0	1	0	0	0	0	1
12	0	0	0	0	0	1	0	0	0	0	0	0	1
13	0	0	0	0	0	0	0	0	0	0	1	1	0

Slika 3.1: Graf i njegova matrica susedstva

Konstanta `Null` je rezervisana za situacije kada ne postoji čvor grafa koji ispunjava zahteve algoritma. Tada algoritam vrati `Null` kao rezultat rada. Sledeće tri funkcije ćemo veoma često koristiti. Funkcija `Deg(G, v)` određuje stepen čvora `v` u grafu `G`:

```
function Deg(var G : Graph; v : integer) : integer;
var
  sum, i : integer;
begin
  sum := 0;
  for i := 1 to G.N do
    if G.adjacent[v, i] then
      sum := sum + 1;
  Deg := sum
end;
```

Funkcija `FirstNeighbour(G, v)` vraća redni broj prvog suseda čvora *v*, ili `Null` ako `v` nema suseda:

```
function FirstNeighbour(var G : Graph; v : integer) : integer;
var
  i : integer;
  go : Boolean;
begin
  i := 1;
  go := true;
  while go and (i <= G.N) do
    if G.adjacent[v, i] then
      go := false
    else
      i := i + 1;
  if go then
    FirstNeighbour := Null
  else
    FirstNeighbour := i
end;
```

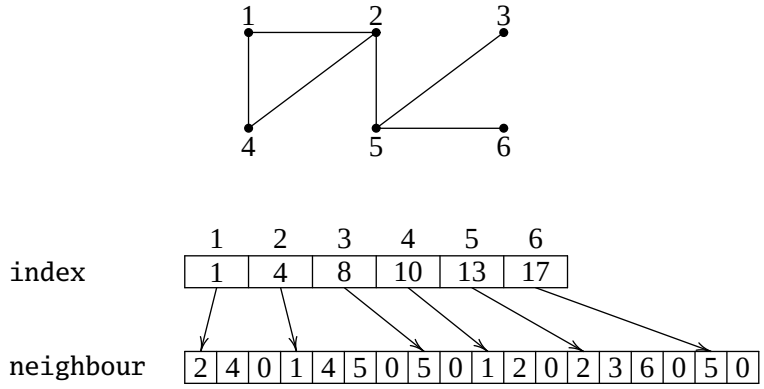
dok funkcija `NextNeighbour(G, v, x)` vraća redni broj sledećeg suseda čvora `v` koji se nalazi posle čvora `x`, ili `Null` ako `v` nema više suseda:

```
function NextNeighbour(var G : Graph; v, x : integer) : integer;
var
  i : integer;
  go : Boolean;
begin
  i := x + 1;
```

```
  go := true;
  while go and (i <= G.N) do
    if G.adjacent[v, i] then
      go := false
    else
      i := i + 1;
  if go then
    NextNeighbour := Null
  else
    NextNeighbour := i
end;
```

Primer 3.1 Kao primer navodimo kostur programa koji učitava graf iz datoteke i potom za svaki čvor grafa ispisuje njegov stepen i listu njegovih suseda.

```
program TheFirstExample;
const
  MaxNoVertices = 50;
  Null = 0;

type
  Graph = record
    N : integer;
    adjacent : array [1 .. MaxNoVertices, 1 .. MaxNoVertices]
               of Boolean
  end;

var
  G : Graph;
  v, w : integer;
  f : text;

function Deg(var G : Graph; v : integer) : integer;
  begin ... end;

function FirstNeighbour(var G : Graph; v : integer) : integer;
  begin ... end;

function NextNeighbour(var G : Graph; v, x : integer) : integer;
  begin ... end;

begin
  assign(f, 'G.txt'); reset(f);
```



Slika 3.2: Graf predstavljen listama suseda

```
  ReadGraph(f, G); close(f);
  for v := 1 to G.N do
    begin
      write(v:2, Deg(G, v):3, ':');
      w := FirstNeighbour(G, v);
      while w <> Null do
        begin
          write(w:3);
          w := NextNeighbour(G, v, w)
        end;
      writeln
    end
end.
```

3.2 Lista suseda

Za graf kažemo da je *redak* (engl. *sparse*) ako ima relativno mali broj grana (štagod to značilo u ovom trenutku). Retke grafove možemo efikasnije predstaviti koristeći dva niza: drugi niz sadrži spiskove suseda svih čvorova, dok prvi niz u kućici j sadrži početak spiska suseda čvora j. Svaki spisak suseda se završava oznakom `Null`. Primer je dat na Sl. 3.2.

Retke grafove ćemo predstavljati kao slogove koji imaju tri polja: broj čvorova, niz koji sadrži početak svake liste suseda, i niz koji sadrži liste suseda svih čvorova.

```
const
  MaxNoVertices = 50;
  MaxNeighbourListLen = 200;
  Null = 0;
```

```
type
  SGraph = record
    N : integer;
    index : array [1 .. MaxNoVertices] of integer;
    neighbour : array [1 .. MaxNeighbourListLen] of integer
  end;
```

Primetimo da nam je za reprezentaciju grafa preko matrice susedstva potreban memorijski prostor veličine $O(n^2)$, gde je n broj čvorova grafa, dok nam je za reprezentaciju retkog grafa dovoljan memorijski prostor veličine $O(m+n)$, gde je m broj grana grafa. Dakle, izbor reprezentacije grafa zavisi od toga da li imamo dodatnu informaciju o broju grana ili ne. Ako nemamo gornje ograničenje na broj grana, onda ćemo uvek birati reprezentaciju preko matrice susedstva. Ako, međutim, znamo da će broj grana imati neku unapred zadatu gornju granicu i ako je tada $m+n$ dovoljno manje od n^2, opredelićemo se za upravo opisanu reprezentaciju grafa.

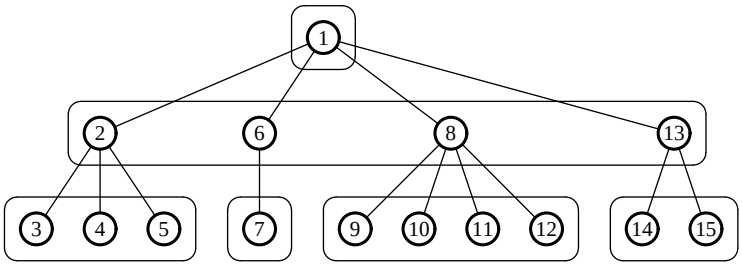
Osim toga, odluka o izboru reprezentacije zavisi i od operacija koje treba da obavljamo nad grafom. Ponekad ćemo se i za retke grafove opredeliti za reprezentaciju preko matrice susedstva ako se operacije koje treba da obavimo nad grafom ne mogu efikasno implementirati u ovoj reprezentaciji. Naravno, važi i obrnuto.

3.3 Pretraživanje prvo u dubinu (DFS)

Mnogi algoritmi na grafovima zahtevaju da se ustanovi sistematski način obilaska obilaska čvorova grafa. U ovom odeljku opisujemo jedan standardan algoritam za obilazak čvorova grafa koji je poznat pod imenom *pretraživanje prvo u dubinu* (engl. *depth-first search*), i koga kratko označavamo sa DFS.

Algoritam obilaska grafa pretraživanjem prvo u dubinu predstavlja klasičan rekurzivni algoritam koji obilazi sve čvorove jedne komponente povezanosti grafa. Krenuvši od nekog čvora u grafu, posetimo njegovog prvog suseda, potom prvog suseda prvog suseda, i tako u dubinu dokle god je to moguće. Kada algoritam obiđe sve susede nekog čvora, vrati se u prethodni čvor da bi proverio da li je neki njegov sused ostao neoznačen. Ako jeste, algoritam se "spusti" na tu stranu sve dok može, itd. Slikovito, to bi izgledalo ovako:

```
procedure DFS(v)
    foreach sused w čvora v do
      if w nije označen then
        begin
          označi w
          DFS(w)
        end
```



a da bismo implementirali algoritam trebaće nam još nekoliko detalja. Pre svega, trebaće nam niz

```
var
  Idx : array [1 .. MaxNoVertices] of integer;
```

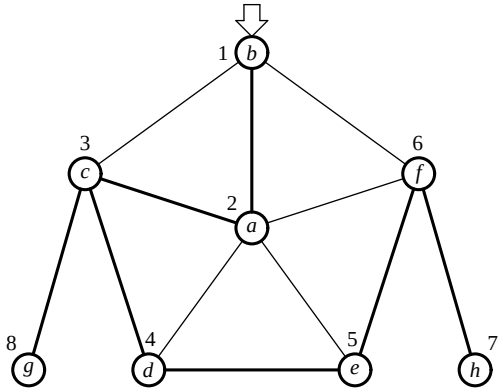
u kome pamtimo kojim redom smo posetili koji čvor: `Idx[v] = k` ako i samo ako smo u *k*-tom koraku posetili čvor v. Takođe, ovaj niz nam treba i da bismo mogli da utvrdimo da li smo neki čvor već posetili. Zato ga inicijalizujemo tako da je pre početka obilaženja `Idx[v] = 0` za sve v:

```
for i := 1 to G.N do Idx[i] := 0;
```

Sada znamo da smo čvor v posetili ako i samo ako je `Idx[v] <> 0`. Koristićemo još i celobrojnu promenljivu `Lbl` koja sadrži narednu raspoloživu labelu. Procedura `DFS` koja obilazi čvorove grafa `G` počev od čvora v izgleda ovako, pri čemu su zbog efikasnosti promenljive `G`, `Lbl` i `Idx` ovaj put globalne:

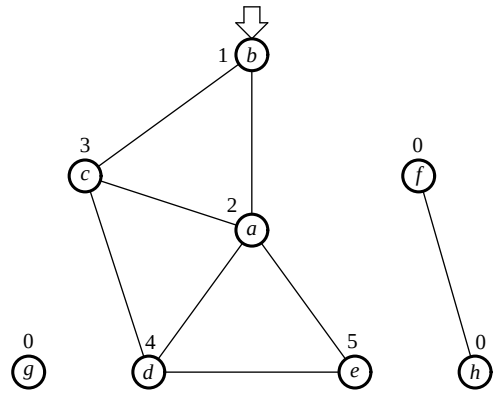
```
procedure DFS(v : integer);
{ G, Lbl, Idx su globalne promenljive i treba ih }
{ inicijalizovati pre poziva procedure }
var
  i, w : integer;
begin
  Lbl := Lbl + 1;
  Idx[v] := Lbl;
  w := FirstNeighbour(G, v);
  while w <> Null do
    begin
      if Idx[w] = 0 then DFS(w);
      w := NextNeighbour(G, v, w)
    end
end;
```

Primer. Na slici pored dat je graf i dat je redosled obilaska njegovih čvorova ukoliko se DFS algoritam startuje od čvora *b*. Prvo posetimo čvor *b*; potom njegovog prvog suseda, *a*; potom prvog suseda čvora *a* što je *c*; onda prvog suseda čvora *c* što je *d* itd. sve do *h*. U povratku iz rekurzije, kada dođemo do *c* primetićemo da sused *g* čvora *c* nije posećen, tako da ćemo sada posetiti i njega, i na kraju se vratiti sasvim iz rekurzije završivši time obilazak.



Napomenimo još jednom da DFS obilazi sve čvorove *komponente povezanosti* čvora koji se prosledi proceduri kao ulazni argument. To je pokazano i u sledećem primeru:

Primer. Ako pustimo algoritam DFS na graf na slici pored počev od čvora *b*, dobićemo navedeni redosled obilazaka čvorova koji je smešten u elemente niza `Idx`. Vidimo da za neke čvorove odgovarajući element niza `Idx` ima vrednost 0. Kada se DFS pusti od čvora *b*, algoritam nema načina da dosegne čvorove *g*, *f* i *h* zato što ne postoji put od *b* do ova tri čvora (oni ne pripadaju komponenti povezanosti kojoj pripada *b*).



Jednostavnom modifikacijom originalne DFS procedure dobijamo algoritam koji obliazi sve čvorove kako povezanih tako i nepovezanih grafova:

```
procedure FullDFS(v : integer);
{ G, Lbl, Idx su globalne promenljive }
begin
  DFS(v);
  for v := 1 to G.N do
    if Idx[v] = 0 then DFS(v)
end;
```

a ovaj pristup vodi do efikasnog algoritma kojim se može proveriti da li je graf povezan:

```
function Connected(var G : Graph) : boolean;
var
  v : integer;
  go : boolean;
  Idx : array [1 .. MaxNoVertices] of integer;

  procedure DFS(v : integer);
  var
    i : integer;
```

```
    w : integer;
  begin
    Idx[v] := 1;
    w := FirstNeighbour(G, v);
    while w <> Null do
      begin
        if Idx[w] = 0 then DFS(w);
        w := NextNeighbour(G, v, w)
      end
  end;

begin
  for v := 1 to G.N do Idx[v] := 0;
  DFS(1);

  v := 2;
  go := true;
  while go and (v <= G.N) do
    if Idx[v] = 0 then
      go := false
    else
      v := v + 1;
  Connected := go
end;
```

Ovaj algoritam takođe pokazuje drugi način da se označe čvorovi grafa koje smo do sada obišli. Umesto da tokom obilaska u odgovarajući element niza `Idx` smestimo indeks čvora prilikom obilaska grafa, ovaj algoritam koristi niz `Idx` kako bi smestio informaciju o tome da li je algoritam već posetio čvor. Ovaj pristup se lako može modifikovati kako bi se dobio algoritam koji zapravo broji komponente povezanosti grafa.

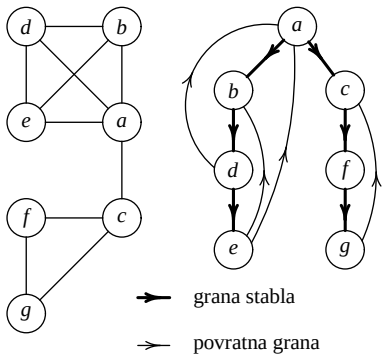
3.4 Stablo pretraživanja DFS algoritma

Pažljivom analizom DFS algoritma može se ustanoviti da ovaj algoritam tokom izvršavanja nad povezanim grafom konstruiše jedno posebno pokrivajuće stablo tog grafa. To stablo zovemo *stablo pretraživanja DFS algoritma*. Prema tome, korektnost procedure `Connected` iz prethodnog odeljka sledi na osnovu Teoreme 2.11 koja tvrdi da je graf povezan ako i samo ako ima pokrivajuće stablo.

Ako graf nije povezan, DFS određuje pokrivajuće stablo samo jedne njegove komponente povezanosti, dok algoritam `FullDFS` određuje pokrivajuće stablo svake njegove komponente povezanosti.

Korensko stablo je stablo kome je posebno istaknut jedan čvor. Taj čvor zovemo *koren stabla*. Kod DFS algoritma koren stabla pretraživanja je čvor od koga počinjemo pretraživanje.

Pokrivajuće stablo povezanog grafa sadrži samo neke grane tog grafa. Ostale grane grafa zovemo *povratne grane u grafu* (engl. *backedges*). Ime je nastalo detaljnim posmatranjem DFS algoritma. Pretpostavimo da smo na neki graf pustili DFS počev od nekog čvora. Tada neka grana ili vodi ka čvoru koga još nismo obišli, u kom slučaju će to biti grana stabla pretraživanja, ili vodi ka nekom čvoru koga smo već obišli. U tom slučaju, kada se stablo pretraživanja crta od gore ka dole, takve grane vode "u natrag" i zato se zovu povratne grane.



Povratne grane sada lako možemo da iskoristimo za dobijanje efikasnog algoritma koji proverava da li je dati povezan graf stablo, kako pokazuje sledeća teorema:

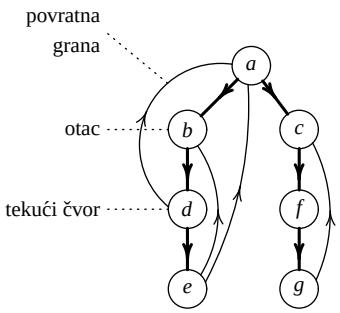
Teorema 3.2 *Povezan graf ima konturu ako i samo ako stablo pretraživanja dobijeno prilikom DFS ima povratnu granu.*

Algoritam koji proverava da li je dati povezan graf stablo se sada lako dobija modifikacijom DFS algoritma:

```
function IsTree(var G : Graph) : Boolean;
{ We assume that G is connected }
var
  Idx : array [1 .. MaxNoVertices] of integer;
  i : integer;
  go : Boolean;

  procedure DFS(v, father : integer);
  var
    w : integer;
  begin
    Idx[v] := 1;
    w := FirstNeighbour(G, v);
    while go and (w <> Null) do
      begin
        if Idx[w] = 0 then
          DFS(w, v)
        else if w <> father then
          go := false;
        w := NextNeighbour(G, v, w)
      end
  end;
begin
  for i := 1 to G.N do Idx[i] := 0;
  go := true;
```

```
    DFS(1, Null);
    IsTree := go
  end;
```

DFS proceduri se ovaj put mora proslediti ne samo čvor od koga počinjemo DFS već i redni broj njegovog prethodnika (ili oca) u stablu pretraživanja. Ovo je neophodno da bismo mogli da identifikujemo povratne grane: grana je povratna ako ide od tekućeg čvora ka nekom označenom čvoru koji nije otac tekućeg čvora.



Glava 4

Strukture i simetrija

Ljudi su stvoreni da uočavaju i poštuju simetriju. Pojam geometrijske simetrije na nivou apstraktnih struktura odgovara pojmu automorfizma.

4.1 Nekoliko reči o grupama

U ovoj glavi ćemo koristiti osnove teorije grupa, i zato ćemo sada napraviti mali uvod. *Grupu* čine skup G i binarna operacija $\cdot$ (množenje) takvi da su zadovoljeni sledeći uslovi (aksiome grupe):

(G1) množenje je asocijativno: $x(yz) = (xy)z$ za sve $x, y, z \in G$;

(G2) postoji $e \in G$ koji se ponaša kao *neutralni element* za množenje: $xe = ex = x$ za sve $x \in G$, i

(G3) za svako $x \in G$ postoji $y \in G$ takav da je $xy = yx = e$; y se zove *inverzni element za* x i, budući da je jedinstven za dato x, označava se sa x^{-1}.

Prototip pojma grupe čini skup svih bijekcija skupa X u sebe. Ova grupa se zove *simetrična* i označava se sa $\mathrm{Sym}(X)$. Ako X ima n elemenata umesto $\mathrm{Sym}(X)$ često pišemo $\mathrm{Sym}(n)$.

Podskup H skupa G je *podgrupa* grupe G ako H obrazuje grupu u odnosu na restrikciju operacije množenja koja je definisana na G, i ima isti neutralni element kao G.

Primer 4.1 Skup celih brojeva $\mathbb{Z}$ zajedno sa operacijom sabiranja $+$ čini grupu. Skup $2\mathbb{Z}$ parnih celih brojeva zajedno sa operacijom sabiranja celih brojeva čini grupu, i to je podgrupa grupe $\mathbb{Z}$. Obe grupe imaju isti neutralni element 0.

Ako je H podgrupa grupe G, tada je $\{H \cdot g : g \in G\}$ particija skupa G. Drugim rečima, ili se $H \cdot g_1$ i $H \cdot g_2$ poklapaju, ili su disjunktni, za sve $g_1, g_2 \in G$. Tako dobijamo sledeći važan rezultata:

Teorema 4.2 (Lagrangeova teorema) *Ako je G konačna grupa sa n elemenata, a H podgrupa grupe G sa k elemenata, onda $k \mid n$.*

Skupovi oblika $H \cdot g$ se zovu *desni koseti* podgrupe H u G, pa se skup $G \setminus H = \{H \cdot g : g \in G\}$ zove *skup desnih koseta* podgrupe H u G.

Homomorfizam grupe G sa operacijom množenja $\cdot$ u grupu H sa operacijom množenja $*$ je svako preslikavanje $f : G \to H$ koje zadovoljava:

$$f(xy) = f(x) * f(y), \quad \text{za sve } x,y \in G.$$

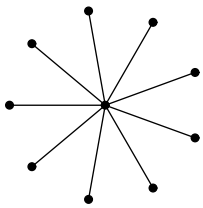
Izomorfizam grupa G i H homomorfizam grupe G u grupu H koji je bijektivan. kažemo još i da su frupe G i H *izomorfne* i pišemo $G \cong H$.

Primer 4.3 (a) Skup celih brojeva $\mathbb{Z}$ sa sabiranjem $+$ obrazuje grupu. Skup $\mathbb{Z}_6 = \{0,1,2,3,4,5\}$ sa sabiranjem $+_6$ po modulu 6 takođe obrazuje grupu. Preslikavanje $f : \mathbb{Z} \to \mathbb{Z}_6$ definisano sa $f(x) = x \bmod 6$ koje ceo broj preslikava na njegov ostatak po modulu 6 je homomorfizam.

(b) Skup realnih brojeva $\mathbb{R}$ sa operacijom sabiranja čini grupu. Skup strogo pozitivnih realnih brojeva $\mathbb{R}^+$ sa operacijom množenja takođe čini grupu. Ove dve grupe su izomorfne, zato što je preslikavanje $f : \mathbb{R}^+ \to \mathbb{R}$ dato sa $f(x) = \log x$ bijektivni homomorfizam.

Važnu klasu grupa čine grupe automorfizama struktura. *Automorfizam* grafa $G = (V,E)$ je svaki izomorfizam $\varphi : V \to V$ grafa u sebe. Sa $\mathrm{Aut}(G)$ označavamo skup svih automorfizama grafa G. Skup $\mathrm{Aut}(G)$ sa operacijom kompozicije funkcija $\circ$ obrazuje grupu.

Primer 4.4 Opišimo $\mathrm{Aut}(K_n)$, $\mathrm{Aut}(S_n)$ i $\mathrm{Aut}(P_n)$ za $n \geqslant 3$, gde S_n označava *zvezdu* sa n čvorova, tj., čiji jedan čvor je susedan sa svim ostalim čvorovima, i ne postoje drugi čvorovi grafa koji su susedni, Sl. 4.1.



S_{10}

Slika 4.1: Zvezda S_{10}

(a) Svaka bijekcija $\varphi : V(K_n) \to V(K_n)$ je automorfizam grafa K_n, tako da je grupa automorfizama grafa K_n izomorfna (kao grupa) simetričnoj grupi $\mathrm{Sym}(n)$.

(b) Svaki automorfizamzvezde sa n čvorova mora da fiksira centralni čvor, dok ostalih $n-1$ čvorova može slobodno da permutuje. Zato je $\mathrm{Aut}(S_n)$ izomorfno (kao grupa) simetričnoj grupi $\mathrm{Sym}(n-1)$.

(c) Put ima samo dva automorfizma: identičko preslikavanje koje fiksira svaki čvor, i "flip" koji "obrće" put tako što preslikava prvi čvor puta u poslednji, drugi u pretposlednji, i tako dalje. Dakle, $|\mathrm{Aut}(P_n)| = 2$.

Svaki homomorfizam $f : G \to H$ grupe G u grupu H jednoznačno određuje relaciju ekvivalencije θ na sledeći način: $(x,y) \in \theta$ ako je $f(x) = f(y)$. Relacija θ se zove *jezgro* homomorfizma f i označava se sa $\ker f$.

Binarna relacija $\rho \subseteq G^2$ na skupu G je *kongruencija* grupe G ako je to relacija ekvivalencije koja ima sledeću osobinu:

$$\text{ako je } (x,y) \in \rho \text{ i } (u,v) \in \rho \text{ onda je } (xu,yv) \in \rho, \quad \text{za sve } x,y,u,v \in G.$$

Jezgro svako g homomorfizma je kongruencija na domenu tog homomorfizma. Obrnuto, svaka kongruencija je jezgro nekog homomorfizma.

Za svaku kongruenciju θ grupe G, skup klasa ekvivalencije $\{g/\theta : g \in G\}$ može postati grupa ako se množenje definiše na sledeći način:

$$(g_1/\theta) \cdot (g_2/\theta) = (g_1 g_2)/\theta.$$

Ovako definisana frupa se zove *faktor grupa* grupe G i označava se sa G/θ.

Teorema 4.5 (Prva teorema o izomorfizmu) *Neka je $f : G \to H$ sirjektivni homomorfizam iz grupe G u grupu H. Tada je $G/(\ker f) \cong H$.*

4.2 Dejstvo grupe

Neka je X skup i G grupa. *Dejstvo grupe G na skup X* je preslikavanje $\mu : X \times G \to G$ takvo da

- $\mu(x,1) = x$, i
- $\mu(\mu(x,g),h) = \mu(x,gh)$.

Dejstvo grupe G na skup X označavamo sa (G,X). Umesto $\mu(x,g)$ pišemo x^g, pa prethodna dva zakona dobijaju sledeći oblik: $x^1 = x$ i $(x^g)^h = x^{gh}$.

Svako $g \in G$ određuje preslikavanje $\tau_g : X \to X : x \mapsto x^g$. Zato što je G grupa, τ_g je permutacija skupa X za svako g (τ_g je sirjektivno zato što je $\tau_g(x^{g^{-1}}) = x$, a injektivno zato što iz $x^g = y^g$ sledi $(x^g)^{g^{-1}} = (y^g)^{g^{-1}}$ tj. $x = y$). Odatle, svako dejstvo (G,X) određuje homomorfizam $\lambda : G \to \mathrm{Sym}(X) : g \mapsto \tau_g$. Obrnuto, svaki homomorfizam $\lambda : G \to \mathrm{Sym}(X)$ određuje jedno dejstvo grupe X na skupu X: $x^g := \Big(\lambda(g)\Big)^{-1}(x)$. Tako dobijamo teoremu reprezentacije Cayelyjevog tipa za dejstvo grupe.

Homomorfizam $\lambda : G \to \mathrm{Sym}(X)$ pridružen dejstvu (G,X) ne mora biti injektivan zato što može da se desi da različiti elementi grupe G deluju na isti način. Za dejstvo (G,X) kažemo da je *verno* ako je odgovarajući homomorfizam λ injektivan. Ukoliko dejstvo (G,X) nije verno, $\ker(\lambda)$ ujednačava elemente grupe G koji deluju na isti način, pa se umesto dejstva (G,X) može posmatrati dejstvo $(G/\theta,X)$ za $\theta := \ker(\lambda)$, dato sa $x^{g/\theta} = x^g$ koje je verno.

Na skupu X uvodimo relaciju $\sim$ ovako: $x \sim y$ ako postoji $g \in G$ takvo da je $x^g = y$. Lako se vidi da je $\sim$ relacija ekvivalencije. Klase ekvivalencije ove relacije zovemo *orbite* dejstva (G,X). Za $x \in X$ odgovarajuća orbita ima oblik $\{x^g : g \in G\}$ i označavamo je kratko sa x^G. Skup svih orbita označavamo sa X/G (umesto sa $X/\sim$).

Za $x \in X$ sa G_x označavamo *stabilizator elementa x,* tj. skup svih elemenata grupe koji ne deluju na njega:

$$G_x := \{g \in G : x^g = x\}.$$

Jasno je da je G_x podgrupa grupe G. Za $g \in G$, sa $\mathrm{fix}(g)$ označavamo *skup svih fiksnih tačaka* od g:

$$\mathrm{fix}(g) := \{x \in X : x^g = x\}.$$

Teorema 4.6 (Lagrangeova teorema) *Neka je (G,X) verno dejstvo grupe G. Tada za svako $x \in X$ imamo da je $|G| = |G_x| \cdot |x^G|$.*

Dokaz. Neka je $G_x \setminus G = \{G_x \cdot g : g \in G\}$ skup desnih koseta podgrupe G_x u grupi G. Znamo da je

$$|G_x \setminus G| = \frac{|G|}{|G_x|}.$$

Zato, da bismo pokazali tvrđenje, dovoljno je da pokažemo da je $|x^G| = |G_x \setminus G|$.

Posmatrajmo preslikavanje $\varphi : x^G \to G_x \setminus G : x^g \mapsto G_x \cdot g$.

- φ je dobro definisano. Neka je $x^g = x^h$. Onda je $x^{gh^{-1}} = x^{hh^{-1}} = x^1 = x$, što znači da gh^{-1} stabilizuje x. Odatle je $gh^{-1} \in G_x$, pa je $g \in G_x \cdot h$. Zato je $G_x \cdot g = G_x \cdot h$.
- φ je sirjektivno. Jasno.
- φ je injektivno. Neka je $G_x \cdot g = G_x \cdot h$. Onda je $g = kh$ za neko $k \in G_x$. Sada je $x^g = x^{kh} = (x^k)^h$. Kako k stabilizuje x imamo da je $x^k = x$, i tako je $x^g = x^h$.

Dakle, φ je bijekcija, i tvrđenje je pokazano. □

Teorema 4.7 (Cauchy-Frobeniusova (Burnsideova) lemma) *Neka grupa G verno deluje na X. Tada je*

$$|X/G| = \frac{1}{|G|} \sum_{g \in G} |\mathrm{fix}(g)|.$$

Dokaz. Za neku formulu Φ uvedimo sledeću oznaku:

$$\chi(\Phi) = \begin{cases} 1, & \Phi \\ 0, & \neg\Phi. \end{cases}$$

Sada se lako vidi da je $|\mathrm{fix}(g)| = \sum_{x \in X} \chi(x^g = x)$ i $|G_x| = \sum_{g \in G} \chi(x^g = x)$. Zato je

$$\sum_{g \in G} |\mathrm{fix}(g)| = \sum_{g \in G} \sum_{x \in X} \chi(x^g = x) = \sum_{x \in X} \sum_{g \in G} \chi(x^g = x) =$$

$$= \sum_{x \in X} |G_x| = \sum_{x \in X} \frac{|G|}{|x^G|} = |G| \cdot \sum_{x \in X} \frac{1}{|x^G|}.$$

Neka je $X/G = \{\Omega_1, \ldots, \Omega_s\}$. Onda je

$$\sum_{x \in X} \frac{1}{|x^G|} = \sum_{x \in \Omega_1 \cup \ldots \cup \Omega_s} \frac{1}{|x^G|} = \sum_{i=1}^{s} \sum_{x \in \Omega_i} \frac{1}{|x^G|}.$$

Pokazaćemo da je $\sum_{x \in \Omega_i} \frac{1}{|x^G|} = 1$. Za $x \in \Omega_i$ je $x^G = \Omega_i$, tako da je

$$\sum_{x \in \Omega_i} \frac{1}{|x^G|} = \sum_{x \in \Omega_i} \frac{1}{|\Omega_i|} = |\Omega_i| \cdot \frac{1}{|\Omega_i|} = 1.$$

Zato je

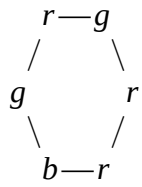
$$\sum_{g \in G} |\mathrm{fix}(g)| = |G| \cdot \sum_{x \in X} \frac{1}{|x^G|} = |G| \cdot \sum_{i=1}^{s} 1 = |G| \cdot s = |G| \cdot |X/G|.$$

Ovim je tvrđenje pokazano. □

4.3 Pólyino dejstvo

Razmotrimo jedan jednostavan problem za početak: odrediti broj ogrlica sa 6 perli, gde svaka perla može biti u jednoj od tri boje r, g, b. Dve ogrlice ne razlikujemo ukoliko se jedna može dobiti od druge primenom neke rotacije.

Evo kako se problem može formalizovati. Neka je $V = \{0,1,2,3,4,5\}$, $C = \{r,g,b\}$ i $X = C^V$. Svaki element skupa X je neko preslikavanje $f : V \to C$, dakle, neki niz boja. Dva niza predstavljaju istu ogrlicu ukoliko postoji ciklička permutacija skupa V koja jedan niz prevodi u drugi. Na primer, nizovi $f := (r,g,r,r,b,g)$ i $g := (r,r,b,g,r,g)$ predstavljaju istu ogrlicu



, i pri tome je $f = g \circ \sigma^{-1}$ za $\sigma = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 0 & 1 \end{pmatrix}$ (ciklički pomeraj za dva). Dakle, imamo grupu $\mathbb{Z}_6$ koja deluje na V kao grupa permutacija. Ako dejstvo ove grupe "proširimo" sa V na C^V tako da je $f^\sigma = f \circ \sigma^{-1}$, onda se orbite ovako proširenog dejstva sastoje od nizova koji predstavljaju istu ogrlicu. Tako se problem brojanja ogrlica svodi na problem određivanja broja orbita nekog dejstva, a za to imamo razvijenu tehniku – Cauchy-Frobenius lemu (Teorema 4.7). Ovim razmatranjima je motivisana sledeća definicija.

Definicija 4.8 Neka su C i V neprazni skupovi i neka grupa G deluje na skup V kao grupa permutacija. Dejstvo grupe G na skup C^V dato sa $f^\sigma := f \circ \sigma^{-1}$ zovemo *Pólyino dejstvo*.

Vratimo se problemu brojanja ogrlica. U modelu koga smo upravo opisali grupa $\mathbb{Z}_6$ deluje na skup C^V na Pólyin način. Orbita elementa $f \in C^V$ se sastoji od svih nizova koji predstavljaju

istu ogrlicu kao i niz f. Tako dobijamo da je broj orglica upravo jednak broju orbita u Pólyinom dejstvu grupe $\mathbb{Z}_6$ na skup C^V. Prema CF-lemi je

$$|C^V/\mathbb{Z}_6| = \frac{1}{6}\sum_{\sigma\in\mathbb{Z}_6} |\text{fix}(\sigma)|,$$

gde se fiksni elementi traže u skupu C^V. Pogledajmo sada kako izgledaju skupovi $\text{fix}(\sigma)$: $f \in \text{fix}(\sigma)$ znači da je $f = f^\sigma$, tj. da je $f(k) = f(\sigma^{-1}(k))$ za sve $k \in V$. Primetimo da ovo znači da je f konstantan na ciklusima od σ! Ovo je toliko važan zaključak da ćemo ga uokviriti:

$f \in \text{fix}(\sigma)$ u Pólyinom dejstvu ako i samo ako je f konstantan na ciklusima permutacije σ.

Elementi grupe $\mathbb{Z}_6$ su sledeće permutacije:

$$\text{id},\ (012345),\ (543210),\ (024)(135),\ (420)(531),\ (03)(14)(25)$$

- Permutacija id ima 6 ciklusa. Na svakom ciklusu f može imati bilo koju od tri vrednosti, pa je $|\text{fix}(\text{id})| = 3^6$.
- Permutacija $\sigma = (012345)$ ima jedan ciklus, pa je $|\text{fix}(\sigma)| = 3$.
- Za $\sigma = (543210)$ je kao i gore $|\text{fix}(\sigma)| = 3$.
- Permutacija $\sigma = (024)(135)$ ima dva ciklusa. Na svakom ciklusu f može imati bilo koju od tri vrednosti. Zato je $|\text{fix}(\sigma)| = 3^2$.
- Za $\sigma = (420)(531)$ je kao i gore $|\text{fix}(\sigma)| = 3^2$.
- Permutacija $\sigma = (03)(14)(25)$ ima tri ciklusa. Na svakom ciklusu f može imati bilo koju od tri vrednosti. Zato je $|\text{fix}(\sigma)| = 3^3$.

Tako dobijamo da je broj ogrlica jednak $\frac{1}{6}(3^6 + 2\cdot 3 + 2\cdot 3^2 + 3^3) = 130$.

U prethodnim razmatranjima smo videli da permutacije sa istom cikličkom strukturom imaju isti broj fiksnih elemenata. *Ovo nije izolovan primer, već opšti fenomen.* Zato ćemo uvesti dva pojma: ciklički tip permutacije i ciklički broj permutacije.

Definicija 4.9 Neka je $|X| = n$ i neka grupa G deluje na skup X kao grupa permutacija. Neka permutacija $\sigma \in G$ ima a_1 ciklusa dužine 1, a_2 ciklusa dužine 2, itd, a_n ciklusa dužine n. Tada se niz $\text{ct}(\sigma) = (a_1, a_2, \ldots, a_n)$ zove *ciklički tip* permutacije σ, a broj $\text{cn}(\sigma) = a_1 + a_2 + \ldots + a_n$ *ciklički broj* permutacije σ.

Jasno je da je $\text{cn}(\sigma)$ broj ciklusa permutacije σ, kao i da je $1\cdot a_1 + 2\cdot a_2 + \ldots + n\cdot a_n = n$ za ciklički tip $(a_1, a_2, \ldots, a_n)$ svake permutacije iz G. Primetimo da cn i ct ne zavise od grupe već od uočenog dejstva, kao i da elementi iste grupe mogu da imaju različite cikličke parametre kada deluju na različite načine.

Teorema 4.10 (Cauchy-Frobenius lema za Pólyino dejstvo) *Neka grupa G verno deluje na skup C^V na Pólyin način. Tada je*

$$|C^V/G| = \frac{1}{|G|}\sum_{\sigma\in G}|C|^{\mathrm{cn}(\sigma)}.$$

Dokaz. Prema Cauchy-Frobenius lemi je

$$|C^V/G| = \frac{1}{|G|}\sum_{\sigma\in G}|\mathrm{fix}(\sigma)|$$

tako da je dovoljno pokazati da je $|\mathrm{fix}(\sigma)| = |C|^{\mathrm{cn}(\sigma)}$ za svako $\sigma \in G$. Neka je $\sigma \in G$ prozivoljno i neka je $\mathrm{cn}(\sigma) = k$. Kao što smo ranije primetili (i uokvirili), $f \in \mathrm{fix}(\sigma)$ ako i samo ako je konstantan na ciklusima od σ. Zato je $|\mathrm{fix}(\sigma)|$ jednak broju preslikavanja $f : V \to C$ koja su konstantna na ciklusima od σ. Na svakom od k ciklusa f može da uzme bilo koju od $|C|$ vrednosti. Pri tome je izbor vrednosti na jednom ciklusu nezavisan od izbora vrednosti na ostalim ciklusima. Zato preslikavanja sa ovom osobinom ima $|C|^k = |C|^{\mathrm{cn}(\sigma)}$. □

4.4 Brojanje grafova

Za neprazan skup V neka je

$$\binom{V}{2} = \Big\{\{x,y\} \subseteq V : x \neq y\Big\}.$$

Tada se graf (V,E) može predstaviti kao preslikavanje $f_E : \binom{V}{2} \to \{0,1\}$, gde je f_E karakteristična funkcija skupa E: $f_E(e) = \chi(e \in E)$.

Na skup V deluje simetrična grupa $\mathrm{Sym}(V)$ na uobičajen način: $x^\sigma = \sigma^{-1}(x)$. Uzmimo da grupa $\mathrm{Sym}(V)$ deluje na skup $\binom{V}{2}$ "po koordinatama": $\{x,y\}^\sigma = \{x^\sigma, y^\sigma\}$. Tada se to dejstvo "proširuje" do Pólyinog dejstva na skup $2^{\binom{V}{2}}$. To znači da za $\sigma \in \mathrm{Sym}(V)$ i $f \in 2^{\binom{V}{2}}$ imamo

$$f^\sigma\Big(\{x,y\}\Big) = f\Big(\{\sigma^{-1}(x), \sigma^{-1}(y)\}\Big).$$

Sledeća veoma jednostavna lema sadrži ključ za primenu Pólyine teorije na problem brojanja neizomorfnih grafova:

Lema 4.11 *Grafovi (V,E_1) i (V,E_2) su izomorfni ako i samo ako postoji $\sigma \in \mathrm{Sym}(V)$ takva da je $f_{E_1}^\sigma = f_{E_2}$.*

Tako dobijamo da su grafovi izomorfni ako i samo ako pripadaju istoj orbiti, pa je broj neizomorfnih grafova na skupu V jednak broju orbita u dejstvu grupe $\mathrm{Sym}(V)$ na skup $2^{\binom{V}{2}}$. Cauchy-Frobenius lema za Pólyino dejstvo sada daje sledeći rezultat:

Teorema 4.12 *Broj neizomorfnih grafova na skupu od n čvorova je dat sa*

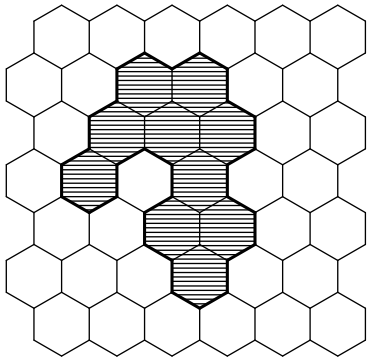
$$\frac{1}{n!}\sum_{\sigma\in \mathrm{Sym}(n)} 2^{\mathrm{cn}(\sigma)}.$$

Broj $\mathrm{cn}(\sigma)$ je broj ciklusa permutacije σ u dejstvu grupe $\mathrm{Sym}(n)$ na skup $\binom{n}{2}$. *Ovaj broj nije jednak broju ciklusa permutacije* σ *pri uobičajenom dejstvu grupe* $\mathrm{Sym}(n)$ *na skup n.* Odatle potiče glavni problem pri utvrđivanju broja neizomorfnih grafova na skupu od *n* elemenata. Kada se dejstvo grupe $\mathrm{Sym}(n)$ na skup $\binom{n}{2}$ interpretira kao podgrupa G_n grupe $\mathrm{Sym}(\binom{n}{2})$, dobija se da su grupe $\mathrm{Sym}(n)$ i G_n apstraktno izomorfne. Ciklička struktura njihovih elemenata je, međutim, *potpuno različita*. Važno je znati da apstraktno izomorfne algebarske strukture mogu imati potpuno različita kombinatorna svojstva.

Glava 5

Brojanje heksagonalnih sistema

Poliheks sistem je povezan sistem podudarnih pravilnih heksagona koji ima osobinu da su svaka dva heksagona u tom sistemu ili disjunktni, ili imaju zajedničko teme, ili imaju zajedničku ivicu. U ovoj glavi nas interesuju geometrijski planarni, prosto povezani poliheksi. Poliheks je geometrijski planaran kada može da se smesti u ravan, dok prosto povezan poliheks nema "rupu". (Sl. 5.1). Na taj način su isključeni heliceni i koronoidi (primeri molekula sa "rupana").



Slika 5.1: Geometrijski planaran, prosto povezan poliheks sa $h = 10$ heksagona

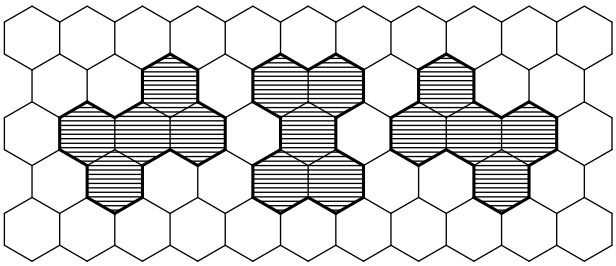
Geometrijski planarni, prosto povezani poliheksi se zovu još i "benzenoidi". Ovi sistemi su jednoznačno određeni svojim rubom, što je kontura u šestougaonoj mreži. Da bismo izbegli konfuziju za geometrijski planaran, prosto povezan poliheks koristićemo termin *heksagonalni sistem* (HS).

Ključni rad na temu brojanja heksagonalnih sistema se pojavio još 1968. godine [1], ali je tek 1983. godine Düsseldorf-Zagreb groupa (Knop i Trinajstić sa saradnicima) objavila svoje rezultate računarskog prebrajanja heksagonalnih sistema za $h = 10$, gde je h broj heksagona obuhvaćenih perimetrom sistema [3].

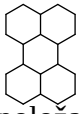
Tužna je istina da ne postoji opšta formula za određivanje broja neizomorfnih heksagonalnih sistema sa datim brojem heksagona. Jedino što nam preostaje je da se ovi sistemi generišu, prebroje i klasifikuju upotrebom brutalne sile: potrebom brzih računara i algoritama. U ovoj glavi prikazaćemo jedan efikasan algoritam koji je upotrebljen za prebrajanje neizomorfnih heksagonalnih sistema sa $h \leqslant 17$ heksagona i njihovu klasifikaciju prema dužini perimetra.

5.1 Osnove

Za dva različita heksagonalna sistema kažemo da su *izomorfni* ako su podudarni u smislu euklidske geometrije. Na primer, na Sl. 5.2 su prikazana tri izomorfna heksagonalna sistema.



Slika 5.2: Tri izomorfna heksagonalna sistema

Drugi način da se sagleda lva situacija je da se heksagonalni sistem  pojavljuje u tri različita *položaja* u heksagonalnoj mreži. Određivanje broja različitih položaja datog heksagonalnog sistema u heksagonalnoj mreži je presudno za određivanje broja neizomorfnih, ili *esencijalno različitih*, heksagonalnih sistema sa datim brojem heksagona. Da bismo mogli da odredimo ovaj broj, potrebno je odrediti broj *simetrija* sistema.

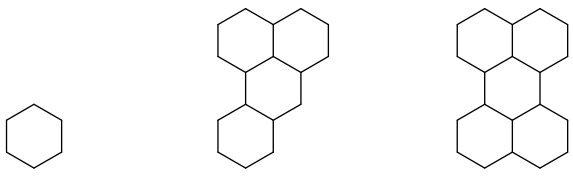
Geometrijski pojam simetrije odgovara apstraktnom pojmu *automorfizma* strukture. na taj način, simetrične strukture su bogate automorfizmima, pa broj elemenata grupe automorfizama uzimamo za meru simetričnosti sistema.

Sada ćemo preciznije odrediti pojmove o kojima smo govorili u prethodnim razmatranjima i daćemo formalne definicije odgovarajućih pojmova. Neka je S neki skup tačaka u euklidskoj ravni. *Automorfizam* skupa S je svako bijektivno preslikavanje F ravni u sebe koje čuva rastojanja, a koje S preslikava na sebe, dakle, koje ima osobinu da je $f(S) = S$. Skup svih takvih preslikavanja

$$\mathrm{Aut}(S) = \{f : f(S) = S \text{ i } f \text{ bijektivno preslikavanje koje čuva rastojanja}\}$$

ima strukturu grupe u odnosu na kompoziciju funkcija. Zato se $\mathrm{Aut}(S)$ zove još i *grupa automorfizama* sistema S.

Na primer, neka je S_1 heksagonalni sistem sa tačno jednim heksagonom, Sl. 5.3 (a). Tada je $|\mathrm{Aut}(S)| = 12$ zato što postoji 6 rotacija i 6 osnih simetrija koje preslikavaju S_1 na sebe. Lako se vidi da važi sledeće:



(a) (b) (c)

Slika 5.3: Tri heksagonalna sistema

Lema 5.1 *Neka je S heksagonalni sistem. Tada je* $|\mathrm{Aut}(S)| \leqslant 12$.

Kao sledeći promer posmatrajmo heksagonalni sistem S_4 sa četiri heksagona koji je prikazan na Sl. 5.3 (b), a neka je S_5 heksagonalni sistem sa pet heksagona koji je prikazan na Sl. 5.3 (c). Tada je $|\mathrm{Aut}(S_4)| = 1$ zato što je trivijalno preslikavanje koje svaku tačku ravni slika na sebe jedino preslikavanje koje preslikava S_4 na sebe. Sa druge strane, $|\mathrm{Aut}(S_5)| = 4$: postoje dve rotacije i dve osne simetrije koje preslikavaju S_5 na sebe.

Kao posledicu Cauchy-Frobeniusove leme (Lema 4.7) dobijamo:

Teorema 5.2 *Neka je S heksagonalni sistem i neka je* $k = |\mathrm{Aut}(S)|$. *Tada S može da se pojavi u* $12/k$ *različitih položaja u heksagonalnoj mreži.*

Na primer, za heksagonalni sistem S_5 prikazan na Sl.. 5.3 (c) znamo da je $|\mathrm{Aut}(S_5)| = 4$, odakle sledi da S_5 može da se pojavi u $12/4 = 3$ različita položaja u heksagonalnoj mreži. Ta tri položaja su prikazana na Sl. 5.2.

Pažljivom analizom se može ustanoviti da grupa automorfizama proizvoljnog heksagonalnog sistema mora biti jedna od sledećih grupa: D_{6h}, C_{6h}, D_{3h}, C_{3h}, D_{2h}, C_{2h}, C_{2v} ili C_s. To su dobro poznate male grupe i broj njihovih elemenata je:

Grupa	D_{6h}	C_{6h}	D_{3h}	C_{3h}	D_{2h}	C_{2h}	C_{2v}	C_s
Broj elemenata	12	6	6	3	4	2	2	1

Neka je sa $H(h)$ označen broj svih heksagonalnih sistema sa h heksagona, uključujući i izomorfne kopije, a neka je sa $N(h)$ označen broj neizomorfnih heksagonalnih sistema sa h heksagona. Dalje, neka je za grupu $G \in \{D_{6h}, C_{6h}, D_{3h}, C_{3h}, D_{2h}, C_{2h}, C_{2v}, C_s\}$ sa $N(G,h)$ označen broj neizomorfnih heksagonalnih sistema sa h heksagona čija grupa automorfizama je G. Tada je

$$\begin{aligned} H(h) = {} & N(D_{6h},h) + 2N(C_{6h},h) + 2N(D_{3h},h) + 4N(C_{3h},h) \\ & + 3N(D_{2h},h) + 6N(C_{2h},h) + 6N(C_{2v},h) + 12N(C_s,h) \end{aligned} \tag{5.1}$$

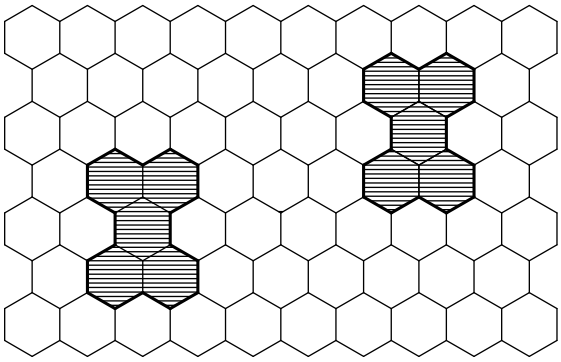
zato što se heksagonalni sistem S čija grupa automorfizama je G pojavljuje u $12/|G|$ različitih položaja u heksagonalnoj mreži. Sa druge strane,

$$\begin{aligned} N(h) = {} & N(D_{6h},h) + N(C_{6h},h) + N(D_{3h},h) + N(C_{3h},h) \\ & + N(D_{2h},h) + N(C_{2h},h) + N(C_{2v},h) + N(C_s,h). \end{aligned} \tag{5.2}$$

Nas interesuje broj $N(h)$, broj neizomorfnih heksagonalnih sistema sa h heksagona, a taj broj nije nimalo lako izračunati. Umesto da direktno računamo $N(h)$, izračunaćemo prvo brojeve $H(h)$, $N(D_{6h},h)$, $N(C_{6h},h)$, $N(D_{3h},h)$, $N(C_{3h},h)$, $N(D_{2h},h)$, $N(C_{2h},h)$ i $N(C_{2v},h)$, potom ćemo na osnovu formule 5.1 izračunati $N(C_s,h)$, i na kraju ćemo na osnovu formule 5.2 izračunati $N(h)$.

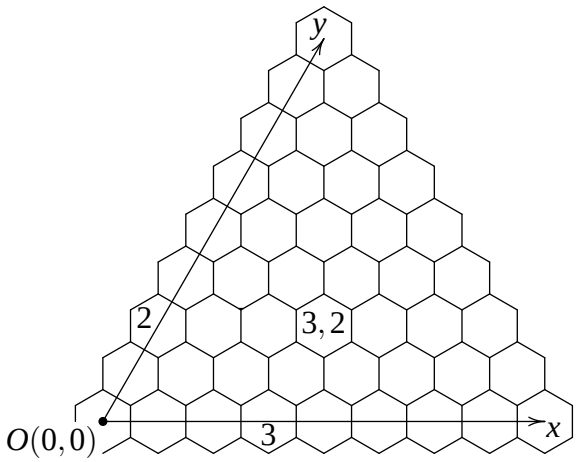
5.2 Algoritam

U ovom odljeku ćemo dati skicu algoritma kojim se efikasno može odrediti broj $H(h)$ svih *različitih* (ne nužno i neizomorfnih) heksagonalnih sistema sa h heksagona. Kako bismo eliminisali translatorno ekvivalentne sisteme (videti Sl. 5.4) uvešćemo pojam kaveza i pokazati da je dovoljno prebrojati samo one heksagonalne sisteme koji su pravilno smešteni u kavezu.



Slika 5.4: Dva translatorno ekvivalentna heksagonalna sistema

Kavez je relativno pravilan deo heksagonalne mreže u kome pokušavamo da uhvatimo sve relevantne heksagonalne sisteme. Algoritam koga ćemo demonstrirati koristi trougaoni kavez, gde je trougao jednakostraničan. Neka je sa Cage(h) označen trougaoni kavez sa h heksagona duž svake strane. Na Sl. 5.5 je prikazan Cage(9), kao i način da se u kavez uvede koordinatni sistem.

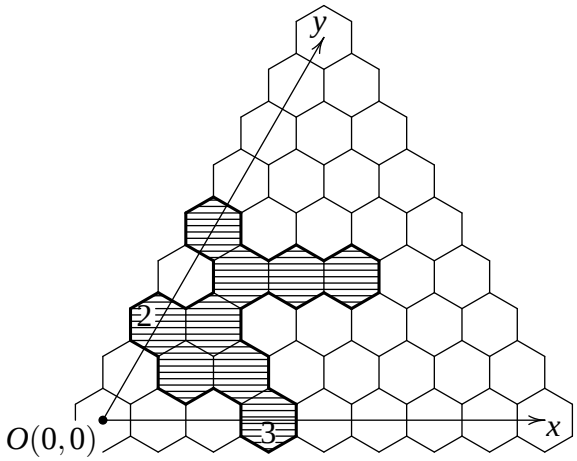


Slika 5.5: Cage(9)

Svaki heksagonalni sistem sa najviše h heksagona se može postaviti u kavez tako da se bar jedan od njegovih heksagona nalazi na x-osi kaveza i da se bar jedan od njeovih heksagona nalazi na y-osi kaveza. Tada kažemo da je heksagonalni sistem *pravilno smešten* u kavez. Na Sl. 5.6 je prikazan heksagonalni sistem sa $h = 9$ heksagona koji je pravilno smešten u kavez Cage(9).

Sužavanjem pažnje sa cele heksagonalne mreže na kavez ni na koji način ne utiče na opštu strategiju koju smo naveli u uvodu ove glave. Važi, naime, sledeće:

Teorema 5.3 *Neka je S heksagonalni sistem sa h heksagona i neka je $k = |\mathrm{Aut}(S)|$. Tada postoji $12/k$ različitih heksagonalnih sistema koji su izomorfni sa S i koji su pravilno smešteni*



Slika 5.6: Heksagonalni sistem sa $h = 9$ heksagona koji je pravilno smešten u kavez

u Cage(h).

odatle sledi da je dovoljno probrojati sve heksaonalne sisteme koji su pravilno smešteni u kavez. Pošto znamo koliko se puta simetrični heksagonalni sistemi pojavljuju u kavezu, možemo lako da odredimo broj svih heksagonalnih sistema čija grupa automorfizama je trivijalna grupa C_s. Taj podatak, nam je dovoljan da se odredi broj svih heksagonalnih sistema sa datim brojem heksagona, ako znamo broj svih simetričnih heksagonalnih sistema. Ovim pristupom se *izbegavaju testovi izomorfnosti* koji su vremenski najskuplji deo sličnih algoritama.

Posmatrajmo jedan heksagonalni sistem koji je pravilno smešten u kavez Cage(h), neka je p najmanja od svih koordinata njegovih heksagona koji se nalaze na x-osi, a q najmanja od svih koordinata njegovih heksagona koji se nalaze na y-osi. Heksagone tog sistema čije koordinate su $(p,0)$ i $(0,q)$ (u odnosu na koordinatni sistem kaveza) će biti veoma značajni za algoritam koji predstavljamo, i zvaćemo ih *ključni heksagoni*.

Neka $H(p,q)$ označava skup svih heksagonalnih sistema sa $\leqslant h$ heksagona koji zadovoljavaju sledeća dva uslova:

- heksagonalni sistem je pravilno smešten u Cage(h), i
- njegov ključni heksagon na x-osi koordinatu p, dok njegov ključni heksagona na y-osi ima koordinatu q.

(Na Sl. 5.6 je prikazan jedan element skupa $H(3,2)$).

Familija $\{H(p,q) : 0 \leqslant p \leqslant h-1, 0 \leqslant q \leqslant h-1\}$ je particija skupa svih heksagonalnih sistema sa $\leqslant h$ heksagona koji su pravilno smešteni u Cage(h). Lako se može pokazati da je $|H(p,q)| = |H(q,p)|$, za sve $p,q \in \{0,1,\ldots,h-1\}$. Tako se proces prebrajanja svih pravilno smeštenih heksagonalnih sistema svodi na određivanje broja $|H(p,q)|$ za sve $p \geqslant q$. Prva verzija algoritma koja je zasnovana na ovoj ideji skicirana je na Sl. 5.7.

Kada su nam dati pbrojevi $0 \leqslant q \leqslant p \leqslant h-1$ i kavez Cage(h), određivanje broja $|H(p,q)|$ se svodi na generisanje i prebrajanje svih heksagonalnih sistema iz skupa $H(p,q)$, što se može

```
inicijalizuj Cage(h);
total := 0;
for q := 0 to h − 1 do
    for p := q to h − 1 do
        odredi H(p,q);
        n := |H(p,q)|;
        if p = q then
            total := total + n
        else
            total := total + 2n
        fi
    od
od
```

Slika 5.7: Prva verzija algoritma

realizovati generisanjem rubne linije sistema. Pogled na Sl. 5.6 otkriva da se rubna linija heksagonalnog sistema koji je pravilno smešten u kavez može podeliti na dva dela: na levi deo linije (gledano iz perspektive čitaoca) koji počinje na y-osi ispod ključnog heksagona i završava na x-osi, i na ostatak rubne linije koji zovemo desni deo linije.

Algoritam rekurzivno generiše levi deo linije. Čim levi deo linije dodirne x-osu počinje se sa generisanjem desnog dela linije. Sve vreme vodimo računa o dužini rubne linije kao i o površini dela heksagonalnog sistema koji je linijom obuhvaćen. Kada se proces generisanja završi površina obuhvaćena linijom daje zapravo informaciju o broju heksagona u tom sistemu. Na ovaj način se lako otklanjaju beskorisni (tj. oni sa $> h$ heksagona koje ne brojimo) heksagonalni sistemi koji se prirodno javljaju prilikom generisanja svih heksagonalnih sistema iz skupa $H(p,q)$.

Insistirati na tome da se generišu isključivo heksagonalni sistemi koji pripadaju skupu $H(p,q)$ bi predstavljalo veliki gubotak vremena i računskih resursa, zato što bi to zahtevalo da se izvršavaju dodatni testovi koji bi proveravali da li je levi deo ruba dodirnuo x-osu tačno kod heksagona p. S druge strane, kada levi deo ruba dodirne x-osu kod heksagona, recimo, $p+2$, zašto da ignorišemo takav heksagonalni sistem koga ćemo ionako morati ponovo da generišemo kasnije kada na red dođe $H(p+2,q)$? Zato ćemo sada uvesti drugačiju particiju skupa svih heksagonalnih sistema koji su pravilno smešteni u kavez.

Neka nam je dat h i kavez Cage(h). Stavimo $H^*(q) = \cup_{j=q}^{h-1} H(j,q)$, za sve $q = 0,1,\ldots,h-1$. Jasno je da je $\{H^*(q) : 0 \leqslant q \leqslant h-1\}$ articija skupa svih heksagonalnih sistema sa h heksagona koji su pravilno smešteni u Cage(h). Umesto sve posebne faze u radu algoritma (generisanje skupa $H(p,q)$ i dodavanje odgovarajućih vrednosti na *total*), nova verzija algoritma koju ćem opokazati ima samo jednu fazu tokom koje se generisanje i brojanje obavljaju istovremeno. Jedino treba voditi računa o tome da se spreči pojava heksagonalnih sistema iz skupa $H(p,q)$ gde je $p < q$. No, to može postići tako što je u heksagonalnoj mreži zabrane neka "leva skretanja" i

“skretanja nadole”. Štaviše, uvođenjem ovakvih zabrana se čak i ubrzava rad algoritma. Sve ove ideje na jednom mestu mogu se naći u verziji algoritma koji je prikazan na Sl. 5.8.

kao što vidimo, radi se o klasičnom primeru bektrek algoritma, koji stoga pati od svih problema od kojih pate bektrek algoritmi: stablo pretraživanja postaje preveliko čak i za relativno male vrednosti h, tako da je važno otsecati grane u ovom stablu kadgod je to moguće. Jedna ideja koja se koristi u procesu “orezivanja” stabla se sastoji u tome da za velike vrednosti parametra q postoje delovi kaveza u koje heksagonalni sistem $\leqslant h$ heksagona ne može da dospe, dok se u tom delu kaveza veoma lagodno baškare beskorisni heksagonalni sistemi za koje nismo zainteresovani. Zato algoritam tokom inicijalizacije kaveza za zadato q zabranjuje rubnoj liniji da skrene u te delove kaveza.

Druga ideja koja se koristi da se “oreže” stablo pretraživanja se sastoji u tome da se tokom generisanja rubne linije broje rubni heksagoni. *Rubni heksagon* je heksagon koji ima zajedničku ivicu sa rubnom linijom i nalazi se sa one strane rubne linije sa koje će jednog dana biti unutrašnjost heksagonalnog sistema. Jasno je da će rubni heksagoni svakako biti deo heksagonalnog sistema i da stoga njihov broj ne sme biti veći od h. Ovo se pokazalo kao veoma dobar kriterijum za otsecanje beskorisnih grana u stablu pretraživanja. Ideja je krajnje jednostavna: produžavanje levog/desnog dela rubne linije je moguće samo ako smo do sada registravali ne više od h rubnih heksagona.

Treća ideja koja ubrzava algoritam je život na kredit. Kada algoritam otpočne sa generisanjem levog dela rubne linije, mi ne znamo tačno gde će ta linija prvi put dodirnuti x-osu, ali znamo da će se to sigurno desiti. Drugim rečima, znajući da jedan heksagon sa x-ose mora da bude deo heksagonalnog sistema, taj heksagom možemo uračunati unapred, i time otkloniti još više beskorisnih heksagonalnih sistema čak i pre nego što levi deo rubne linije dodirne x-osu.

Sve ideje o kojima je do sada bilo reči sakupljene su u algoritmu koji je prikazan na Sl. 5.9.

5.3 Implementacija

U prethodnom odeljku smo, naravno, dali samo pregled osnovnih ideja algoritma koji je detaljno opisan u [7]. Još mnogo tehničkih detalja je moralo biti realizovano kako bi se od osnovne ideje dobio efektivan računarski program.

Program je napisan za IBM PC kompatibilne računare u programskom jeziku Modula–2. Program se sastoji iz pet modula i ima preko 1900 bruto linija programskog koda. On je upotrebljen za određivanje broja svih neizomorfnih heksagonalnih sistema sa $h \leqslant 17$ heksagona.

Enumeracija svih heksagonalnih sistema sa 17 heksagona je veoma dugotrajan proces. Zato je ceo posao enumeracije podeljen na više manjih zadataka (konkretno, za $h = 17$ smo podelili posao na 197 manjih zadataka), čime je omogućeno da se program izvršava paralelno na više geografskih lokacija: u Novom Sadu, Ottawi i Trondheimu. Paralelizacija je obavljena ručno, a prema dva prirodna kriterijuma: u obzir je uzeta koordinata ključnog heksagona na y-osi i, obzirom da se ispostavilo da je ova podela previše gruba, na osnovu početnog segmenta rubne linije.

```
procedure ExpandRightPart(ActualPos);
begin
    if EndOfRightPart then
        n := NoOfHexagons();
        if n ⩽ h then
            odredi p;
            if p = q then
                total[n] := total[n] + 1
            else
                total[n] := total[n] + 2
            fi
        fi
    else
        FindPossible(ActualPos, FuturePos);
        while RightPartCanBeExpanded(ActualPos, FuturePos) do
            ExpandRightPart(FuturePos);
            CalcNewFuturePos(ActualPos, FuturePos)
        od
    fi
end;
procedure ExpandLeftPart(ActualPos);
begin
    if EndOfLeftPart then
        ExpandRightPart(RightInitPos(q))
    else
        FindPossible(ActualPos, FuturePos);
        while LeftPartCanBeExpanded(ActualPos, FuturePos) do
            ExpandLeftPart(FuturePos);
            CalcNewFuturePos(ActualPos, FuturePos)
        od
    fi
end;
begin
    inicijalizuj Cage(h);
    postavi total[1...h] na 0;
    for q := 0 to h − 1 do
        inicijalizuj ključni heksagon na y-osi(q);
        ExpandLeftPart(LeftInitPos(q))
    od
end
```

Slika 5.8: Druga verzija algoritma

```
procedure ExpandRightPart(ActualPos, BdrHexgns);
begin
    if EndOfRightPart then
        n := NoOfHexagons();
        if n ⩽ h then
            odredi p;
            if p = q then total[n] := total[n] + 1
            else total[n] := total[n] + 2
            fi
        fi
    else
        FindPossible(ActualPos, FuturePos);
        while RightPartCanBeExpanded(ActualPos, FuturePos)
        and BdrHexgns ⩽ h do
            ExpandRightPart(FuturePos, update(BdrHexgns));
            CalcNewFuturePos(ActualPos, FuturePos)
        od
    fi
end;
procedure ExpandLeftPart(ActualPos, BdrHexgns);
begin
    if EndOfLeftPart then
        ExpandRightPart(RightInitPos(q), updCredit(BdrHexgns))
    else
        FindPossible(ActualPos, FuturePos);
        while LeftPartCanBeExpanded(ActualPos, FuturePos)
        and BdrHexgns ⩽ h do
            ExpandLeftPart(FuturePos, update(BdrHexgns));
            CalcNewFuturePos(ActualPos, FuturePos)
        od
    fi
end;
begin
    inicijalizuj Cage(h);
    postavi total[1...h] na 0;
    for q := 0 to h − 1 do
        inicijalizuj ključni heksagon na y-osi(q);
        ExpandLeftPart(LeftInitPos(q), InitBdrHexgns(q))
    od
end
```

Slika 5.9: Treća verzija algoritma

Bibliografija

[1] Balaban A. T., Harary F., *Chemical Graphs, Enumeration and Proposed Nomenclature of Benzenoid Cata-Condensed Polycyclic Aromatic Hydrocarbons*, Tetrahedron 24 (1968), 2505–2516

[2] Gutman I. (Ed.), *Mathematical Methods in Chemistry*, Prijepolje, 2006

[3] Knop J. V., Szymanski K., Jeričević O., Trinajstić N., *Computer Enumeration and Generation of Benzenoid Hydrocarbons and Identification of Bay Regions*, J. Comput. Chem. 4(1983), 23–32

[4] van Lint J. H., Wilson R. M., *A Course in Combinatorics*, 2nd Ed., Cambridge University Press, 2001

[5] Pemmaraju S., Skiena S., *Computational Discrete Mathematics*, Cambridge University Press, 2003

[6] Roberts F. S., Tesman B., *Applied Combinatorics*, 2nd Ed., Pearson Education Inc., 2005

[7] Tošić R., Mašulović D., Stojmenović I., Brunvoll J., Cyvin B. N., Cyvin S. J., *Enumeration of Polyhex Hydrocarbons to h = 17*, J. Chem. Inf. Comput. Sci., 35(2) (1995) 181–187

[8] West D. B., *Introduction to Graph Theory*, 2nd Ed., Prentice Hall, 2001

Project: 06SER02/02/003

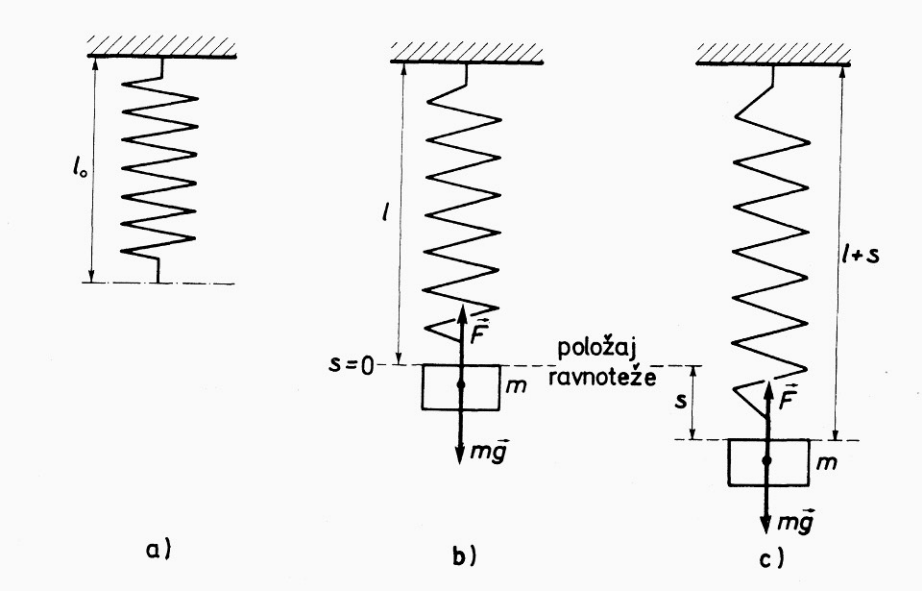
Oscillation and waves, signals

Agneš Kapor

OSCILACIJE

Kretanje koje se ponavlja u jednakim vremenskim razmacima se naziva *periodičnim kretanjem.* Posebna vrsta periodičnog kretanja je *oscilovanje.* Pri oscilovanju se materijalna tačka kreće oko ravnotežnog položaja tj. prelazi istu trajektoriju najpre u jednom a zatim u suprotnom smeru. Posle određenog vremenskog intervala koji se zove *period (T)*, materijalna tačka ponavlja celo kretanje. Trajanje jednog potpunog oscilovanja je T i za to vreme telo dva puta prođe kroz ravnotežni položaj.

Uzmimo primer sistema opruga+masa (Slika 1.) Kada sistem izvedemo iz ravnotežnog položaja i pustimo, sistem počinje da osciluje.



Slika 1.

Rezultujuća sila koja prouzrokuje oscilovanje je vektorska suma sile elastičnosti opruge i težine tela mase m.

$$F = mg - k\left(s + l - l_0\right) = k\left(l - l_0\right) - k\left(l - l_0\right) - ks$$

$$\vec{F} = -k\vec{s}$$

gde je s pomak iz ravnotežnog položaja (elongacija) a k pozitivna elastična konstanta opruge.

Ovako dobijen sila je *elastična* ili *harmonijska* sila koja je proporcionalna pomeranju s iz položaja ravnoteže i suprotnog je smera i ima opšti naziv *restituciona sila*. Sistem koji osciluje pod dejstvom restitucione sile se zove **harmonijski oscilator**.

Matematički je najjednostavnije opisati harmonijsko oscilovanje pomoću harmonijske funkcije. Mnoga oscilatorna kretanja u prirodi se mogu aproksimativno opisati preko kretanja harmonijskog oscilatora.

Kao primer opisaćemo model

- oscilovanja elastične opruge,
- matematičko klatno i
- električno LC kolo.

Ukoliko se kretanje odvija duž jednog pravca, govorimo o **linearnom harmonijskom oscilatoru.** Da bi se utvrdilo kako se kreće harmonijski oscilator moramo rešiti jednačinu kretanja:

$\vec{F} = m\vec{a}$ i $\vec{F} = -kx\vec{i}$ su iste sile tako da dobijamo izraz:

$-kx = m\dfrac{d^2x}{dt^2}$ odnosno

$$\frac{d^2x}{dt^2} + \frac{k}{m}x = 0$$

koja se naziva diferencijalna jednačina harmonijskog oscilovanja, a $\dfrac{k}{m} = \omega_0^2$ *sopstvena kružna frekvencija harmonijskog oscilatora.*

Prema teoriji diferencijalnih jednačina postoje dva linearno nezavisna rešenja jednačine ($\sin\omega_0 t$) i ($\cos\omega_0 t$) a opšte rešenje je linerana kombinacija ta dva nezavisna rešenja.

$$x(t) = a\sin\omega_0 t + b\cos\omega_0 t$$

Zamenimo li da je $a = A\cos\varphi$ i $b = A\sin\varphi$ izraz dobija oblik:

$$x(t) = A\cos\varphi\sin\omega_0 t + A\sin\varphi\cos\omega_0 t$$

$$x(t) = A\sin(\omega_0 t + \varphi)$$

$x(t)$ - elongacija; A- amplituda; ω_0 - kružna frekvencija ($T = \frac{2\pi}{\omega_0}$); $(\omega_0 t + \varphi)$- faza oscilovanja; φ - početna faza u trenutku t=0.

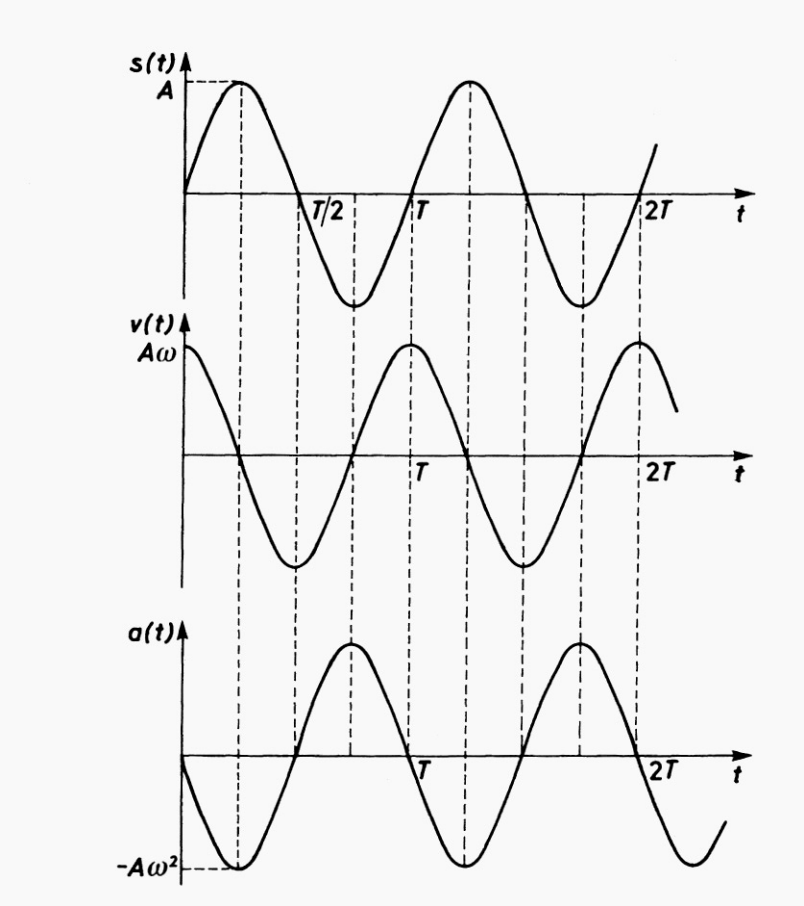
Prvi izvod elongacije po vremenu daje brzinu oscilatora:

$$\mathrm{v}(t) = \frac{dx}{dt} = A\omega_0 \cos(\omega_0 t + \varphi)$$

Drugi izvod elongacije po vremenu daje ubrzanje oscilatora.

$$\mathrm{a}(t) = \frac{d^2x}{dt^2} = -A\omega_0^2 \sin(\omega_0 t + \varphi)$$

Promena ovih funkcija prikazana je na Slici 2.



Slika 2.

Rešenje diferencijalne jednačine kretanja harmonijskog oscilatora i sličnih diferencijalnih jednačina još se jednostavnije može odrediti predstavljajući rešenje u obliku kompleksnog brojeva.

Svaki kompleksan broj $z = x + i\,y$ se može prikazati u polarnim koordinatama:

$z = |z| \cdot e^{i\varphi}$ gde je $|z| = \sqrt{x^2 + y^2}$ i $\varphi = arctg\,\frac{y}{x}$ Uzimajući u obzir Ojlerovu relaciju $e^{i\varphi} = \cos\varphi + i\sin\varphi$ možemo dobiti sopstvenu kružnu frekvenciju oscilatora iz diferencijalne jednačine kretanja za proizvoljno pomeranje s :

$$\frac{d^2 s}{dt^2} + \frac{k}{m} s = 0$$

zamenom pređenog puta $s(t) = A\,e^{i(\omega_0 t + \varphi)}$ i njegovog drugog izvoda po vremenu u tu jednačinu što nam daje rezultata $\omega_0 = \sqrt{\frac{k}{m}}$.

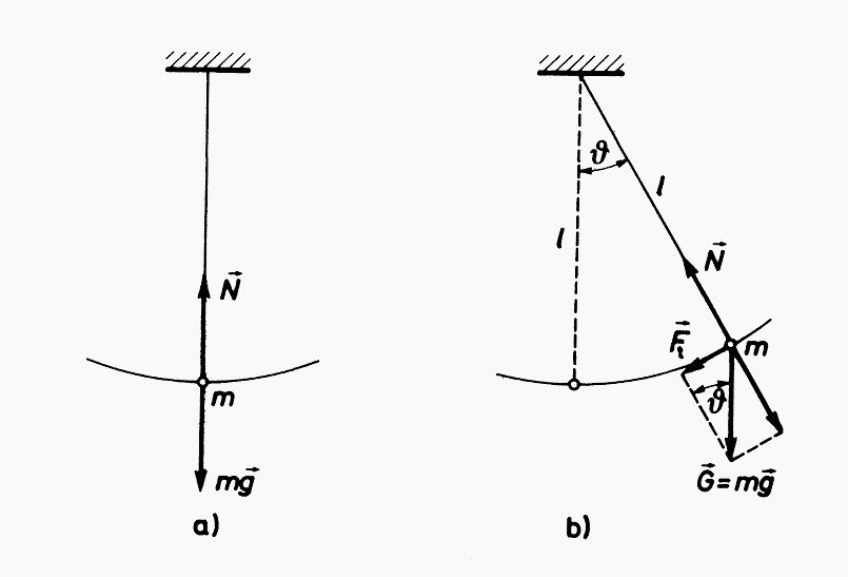
$s(t) = A\cos(\omega_0 t + \varphi)$ - je realni deo rešenja diferencijalne jednačine

$s(t) = A\sin(\omega_0 t + \varphi)$ - je imaginarni deo rešenja diferencijalne jednačine.

Budući da su rešenja matematički ekvivalentna u fizici se bira jedno od njih.

Matematičko klatno

Malo telo mase m koje osciluje obešeno o nerastegljivu laganu nit dužine l čiju masu zanemarujemo se zove *matematičko klatno* (Slika 3)



Slika 3.

Matematičko klatno harmonijski osciluje samo za male amplitude. Za veće amplitude oscilovanje nije harmonijsko. Jednačina kretanja klatna je:

$$ma_i = -mg\sin\vartheta \qquad \text{gde je} \qquad a_i = l\frac{d^2\vartheta}{dt^2}$$

$$\frac{d^2\vartheta}{dt_2} + \frac{g}{l}\sin\vartheta = 0$$

Razvoj funkcije u red:

$$\sin\vartheta = \vartheta - \frac{\vartheta^3}{3!} + \frac{\vartheta^5}{5!} - \frac{\vartheta^7}{7!} + ...(-1)^n \frac{\vartheta^{2n+1}}{(2n+1)!} \pm .....$$

daje u prvoj aproksimaciji rešenje kao za linearni harmonijski oscilator:

$$\vartheta = \vartheta_0 \sin(\omega_0 t + \varphi) \qquad T = \frac{2\pi}{\omega_0} = 2\pi\sqrt{\frac{l}{g}}$$

Za veće amplitude $\sin\vartheta$ se ne može aproksimirati uglom pa se jednačina kretanja ne može tako jednostavno rešiti. U tom slučaju je izraz za period oscilovanja dat u obliku beskonačnog reda:

$$T = 2\pi\sqrt{\frac{l}{g}}\left(1+\frac{1}{2^2}\sin^2\frac{\vartheta}{2}+\frac{1^2}{2^2}\frac{3^2}{4^2}\sin^4\frac{\vartheta}{2}+\frac{1^2}{2^2}\frac{3^2}{4^2}\frac{5^2}{6^2}\sin^6\frac{\vartheta}{2}+....\right)$$

Pošto se članovi reda brzo smanjuju, često je pri izračunavanju perioda dovoljno uvesti prva dva ili tri člana reda a ostale zanemariti. Dobijeni rezultat je tada u granici greške merenja perioda oscilovanja. Često primenjivana formula je i preko perioda oscilovanja za male oscilacije koji označavamo sa $T_0 = 2\pi\sqrt{\frac{l}{g}}$. Tada je:

$$T = T_0\left(1+\frac{1}{4}\sin^2\frac{\vartheta}{2}+\frac{9}{64}\sin^4\frac{\vartheta}{2}+.....\right)$$

Primer:

Koliki je period oscilovanja matematičkog klatna dužine 1 m za amplitude:

a) 6° , b) 15° c) 60°

Kolika je relativna greška ako u poslednja dva slučaja pretpostavimo da period ne zavisi od amplitude? ($g = 9.80665\frac{m}{s^2}$ $\pi = 3.141593$)

Rešenje:

a) $\vartheta = 6^0$ $(0.1047 rad)$ $\sin\vartheta = 0.1045$ Pošto se radi o malom uglu umesto $\sin\vartheta$ možemo uzeti ugao i izračunati period oscilovanja po formuli: $T = 2\pi\sqrt{\frac{l}{g}} = 2.00641 s$

b) $\vartheta = 15^0$ $(0.2618\ rad)$ $\sin\vartheta = 0.2588$

$$T = 2\pi\sqrt{\frac{l}{g}}\left(1+\frac{1}{4}\sin^2\frac{15^0}{2}+.....\right) = 2.0064(1+0.0041) = 2.015\ s$$

Kada bi se period oscilovanja računao izrazom za male oscilacije greška bi iznosila:

$$\frac{\Delta T}{T} = \frac{2.015-2.0064}{2.0064} = 0.4\ \%$$

c) $\vartheta = 60^0$ $(1.047\ rad)$ $\sin\vartheta = 0.866$

$$T = 2\pi\sqrt{\frac{l}{g}}\left(1+\frac{1}{4}\sin^2\frac{60^0}{2}+\frac{1}{4}\frac{9}{16}\sin^4\frac{60^0}{2}+....\right)$$

$T = 2.0064(1+0.0625+0.0088+....) = 2.15\ s$ Relativna greška ako bi se period računao izrazom aza male oscilacije bi iznosila:

$$\frac{\Delta T}{T} = \frac{2.15-2.0064}{2.0064} = 6.7\ \%$$

ELEKTRIČNO OSCILATORNO KOLO (LC)

Jedan od najvažnijih primera oscilujućeg sistema imamom kod elektriciteta. Pojam *naizmenične struje* nije ništa drugo nego oscilujuća električna struja. Ukoliko formiramo strujno kolo koje se satoji od solenoida induktivnosti L i kondenzatora kapaciteta C (slika 4.) možemo napisati analognu diferencijalnu jednačinu kao kod mehaničkog harmonijskog oscilatora.

Napon na kondenzatoru kapaciteta C je

$V_C = \frac{Q}{C}$ gde je Q naelektrisanje kondenzatora. Struja kroz kolo koje sadrži taj kondenzator je $I = \frac{dQ}{dt}$ ili $Q = -\int I\,dt$ gde znak minus označava da struja teče u takvom smeru da smanjuje naelektrisanje kondenzatora. Indukovani napon na zavojnici (solenoidu) iznosi:

$$V_L = -L\frac{dI}{dt}$$

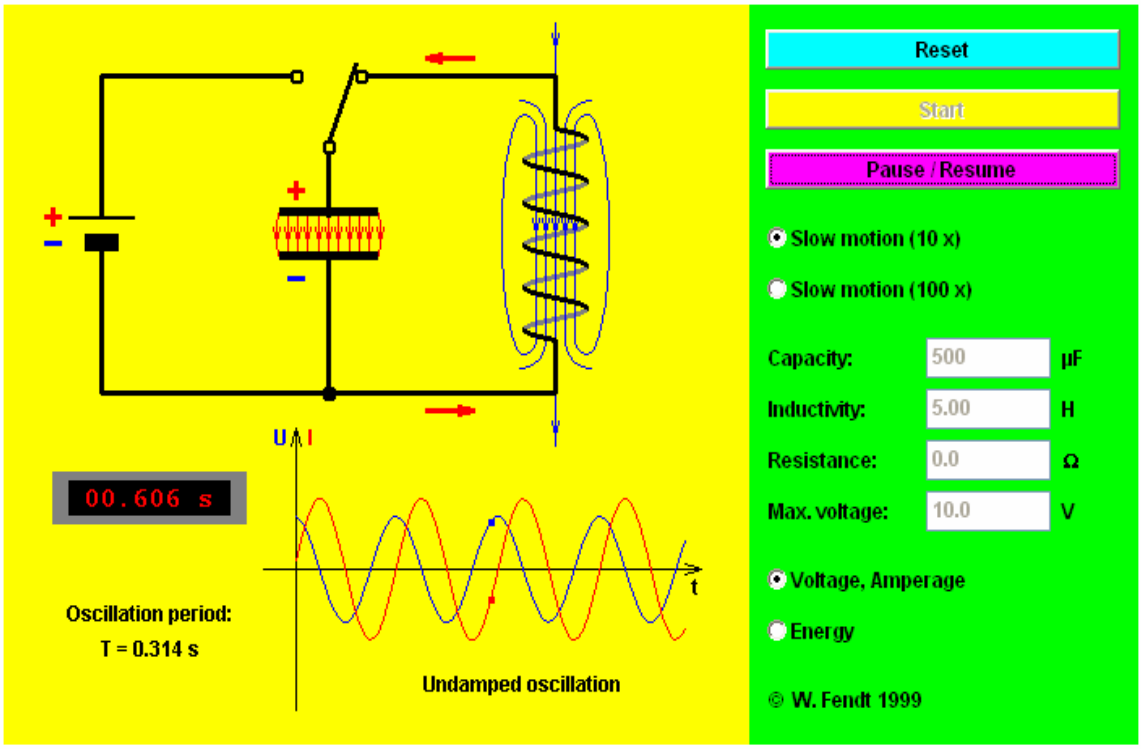
Pošto je zbir padova napona u zatvorenom strujnom krugu jednak nuli dobijamo relaciju:

$$-L\frac{dI}{dt}+\frac{Q}{C}=O \qquad -L\frac{d^2Q}{dt^2}+\frac{Q}{C}=0$$

To je upravo i diferencijalna jednačina oscilatornog kretanja spirane opruge tako da rešenje dobijamo u istom obliku:

$$Q = Q_0 \sin(\omega_0 t + \varphi) \qquad \omega_0 = \left(\frac{1}{LC}\right)^{\frac{1}{2}}$$

Promena naelektrisanja Q može se povezati sa promenom struje i dobiti izraz za promenu



Slika 4.

jačine naizmenične struje. Naravno ovaj primer je samo za idealno LC kolo. U svim realnim strujnim kolima postoji i termogeni otpor R na kojem je pad napona $RI = R\frac{dQ}{dt}$ i koji u diferencijalnoj jednačini dodaje linearni član i menja rešenje slično kao i u jednačinama kretanja mehaničkih oscilatora ukoliko se uzme u obzir sila trenja.

ENERGIJA HARMONIJSKOG OSCILATORA

Pri oscilovanju materijalne tačke dolazi do stalnog prelaza kinetičke energije u potencijalnu i obrnuto.

Izraz za kinetičku energiju harmonijskog oscilatora može se dobiti iz poznatog oblika promene puta u toku vremena:

$$s(t) = A\sin(\omega_0 t + \varphi)$$

a korišćenjem izraza za kinetičku energiju materijalne tačke mase m

$$E_k = \frac{mv^2}{2} = \frac{kA^2}{2}\cos^2(\omega_0 t + \varphi)$$

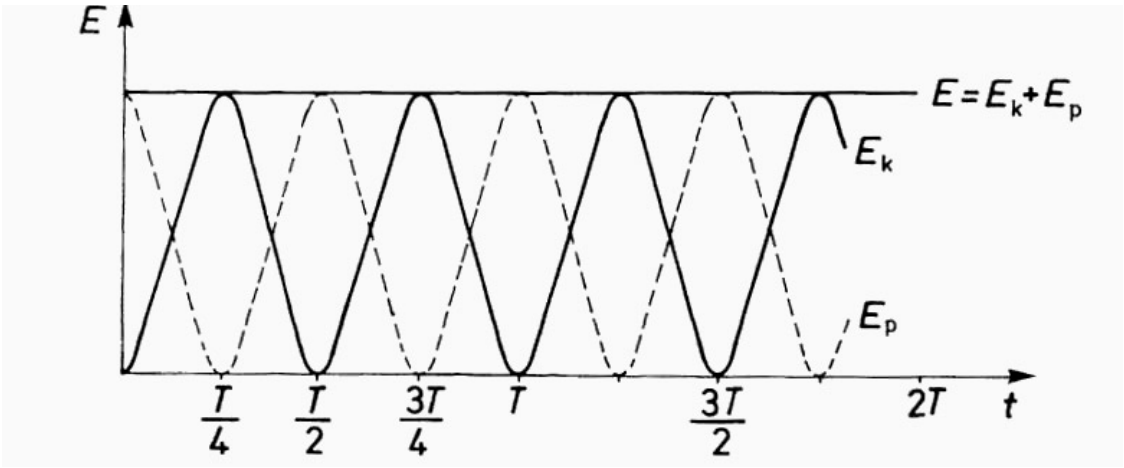
Kada na materijalnu tačku mase m deluje elstična sila $F = -ks$ njena potencijalna energija je jednaka radu te sile pri pomaku tačke za elongaciju s iz ravnotežnog položaja.

$$E_p = \frac{ks^2}{2} = \frac{kA^2}{2}\sin^2(\omega_0 t + \varphi)$$

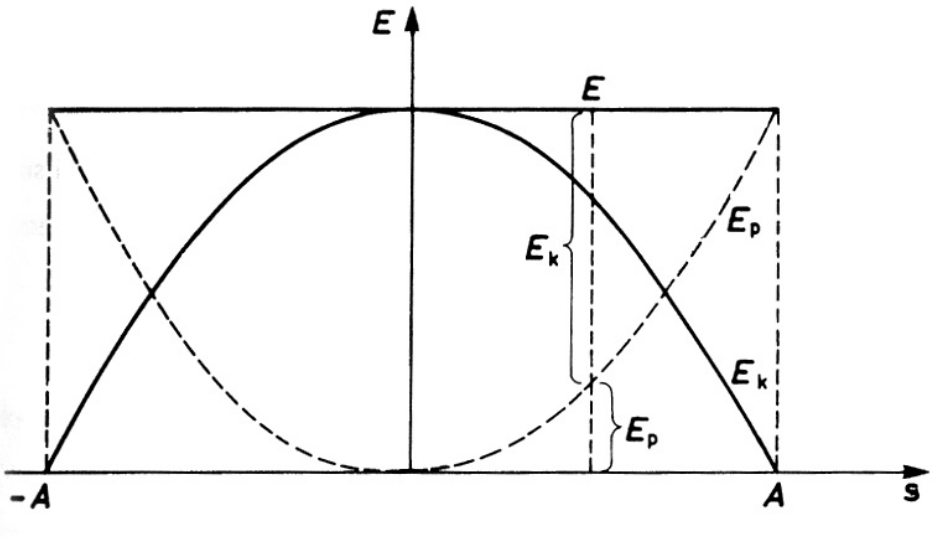
Ukupna energija harmonijskog oscilatora je :

$$E = E_k + E_p = \frac{kA^2}{2}\left[\cos^2(\omega_0 t + \varphi) + \sin^2(\omega_0 t + \varphi)\right] = \frac{kA^2}{2}$$

Grafički prikaz promene kinetičke i potencijalne energija dat je na slici 5.i 6. U svakom trenutku (za svaku elongaciju oscilatora) ukupna energija je konstantna $E = E_k + E_p = const.$



Slika 5.

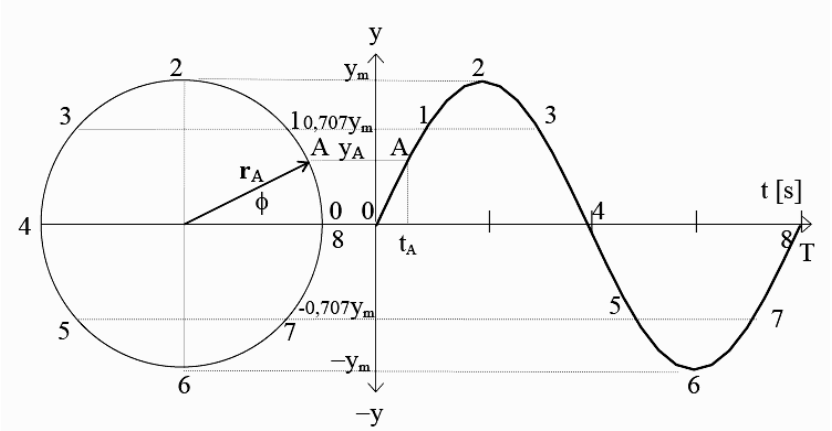


Slika 6.

SLAGANJE HARMONIJSKIH OSCILATORA

Prikaz harmonijskog oscilatora pomoću rotirajućeg vektora

Harmonijsko oscilovanje možemo povezati sa jednolikim kretanjem po kružnici (vidi sliku 7).



Slika 7.

Projekcija rotirajućeg vektora dužine A na x i y osu pravouglog koordinatnog sistema u ravni je:

$$x = A\cos(\omega_0 t + \varphi) \qquad y = A\sin(\omega_0 t + \varphi)$$

Neka tačka koja se jednoliko kreće po kružnici ima projekciju, na bilo koji prečnik kružnice, koja harmonijski osciluje. Vektor $\overrightarrow{OA}$ koji spaja koordinatni početak i tačku A se zove **rotirajući vektor** (fazor).

Pri istovremenom dejstvu više različitih restitucionih sila na oscilator, on će vršiti složeno kretanje, koje je po *principu nezavisnosti dejstva sila,* geometrijski zbir pojedinačnih oscilacija.

$$x_1 = A_1\cos(\omega_1 t + \varphi_1) \qquad x_2 = A_2\cos(\omega_2 t + \varphi)$$

Rezultujuće kretanje u opštem slučaju može da se izrazi kao zbir harmonijskih oscilacija:

$$x = x_1 + x_2 = A_1\cos(\omega_1 t + \varphi_1) + A_2\cos(\omega_2 t + \varphi_2)$$

Matematički je rešenje ovoga izraza složeno jer sadrži šest nezavisnih veličina ($A_1, A_2, \omega_1, \omega_2, \varphi_1, \varphi_2$)

SLAGANJE DVA HARMONIJSKA OSCILATORA ISTOG PRAVCA, SMERA I JEDNAKIH PERIODA

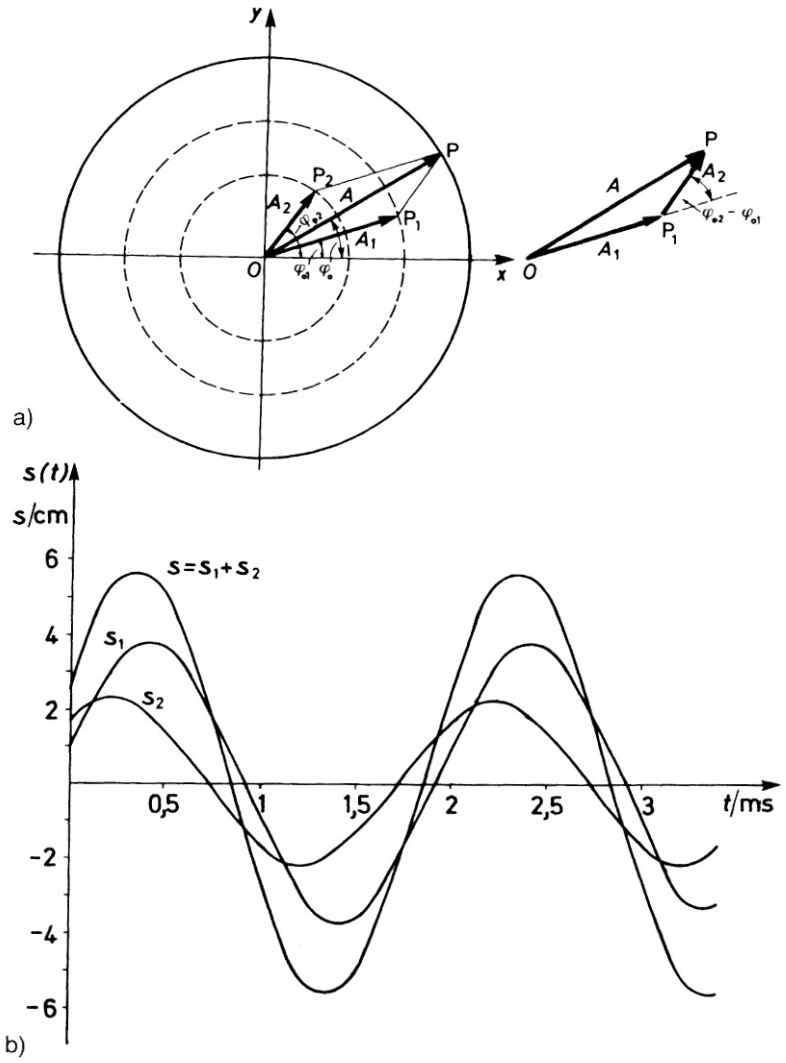
Matematička forma ovih oscilatora i njihovog zbira je sledeća:

$$s_1 = A_1\cos(\omega t + \varphi_1) \qquad s_2 = A_2\cos(\omega t + \varphi_2)$$

$$s = s_1 + s_2 = A_1\cos(\omega t + \varphi_1) + A_2\cos(\omega t + \varphi_2)$$

Analiza ovog kretanja se pojednostavljuje pomoću vektorskog dijagrama odnosno amplitudnog rotirajućeg vektora (Slika 8.).

$$\vec{A} = \vec{A}_1 + \vec{A}_2 \quad \text{i} \quad \varphi_2 - \varphi_1 = const$$



Slika 8.

Analogno se može slagati i veći broj koherentnih harmonijskih oscilatora iste kružne frekvencije:

$$s = A_1 \cos(\omega t + \varphi_1) + A_2 \cos(\omega t + \varphi_2) + A_3 \cos(\omega t + \varphi_3) + ...A_i \cos(\omega t + \varphi_i)$$

i rešenje se dobija u obliku:

$$s = A\cos(\omega t + \varphi)$$

gde se rezultujuća amplituda dobija kao intenzitet rezultante vektora dobijenog sabiranjem vektora amplituda, a početna faza φ odgovara uglu koji taj rezultujući vektor zaklapa sa x osom.

SLAGANJE HARMONIJSKIH OSCILOVANJA ISTOG PRAVCA I SMERA ALI RAZLIČITIH PERIODA. OSCILATORNA KOLEBANJA

Kretanje oscilatora koji se sabiraju se može predstaviti pomoću harmonijskih funkcija oblika:

$$s_1 = A\cos(\omega_1 t + \varphi_0) \qquad s_2 = A\cos(\omega_2 t + \varphi_0) \qquad \omega_2 - \omega_1 \langle\langle\ \omega_1 + \omega_2$$

Rezultujuće oscilovanje je zbir dve kosinusne funkcije koji se može tranformisati pomoću poznae relacije za zbir kosinusa dva ugla:

$$\cos\alpha + \cos\beta = 2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$$

Rezultujuće oscilovanje se tada može predstaviti funkcijom (Slika 9):

$$s = 2A\cos\left(\frac{\omega_2-\omega_1}{2}t\right)\cos\left(\frac{\omega_2+\omega_1}{2}t+\varphi_0\right)$$

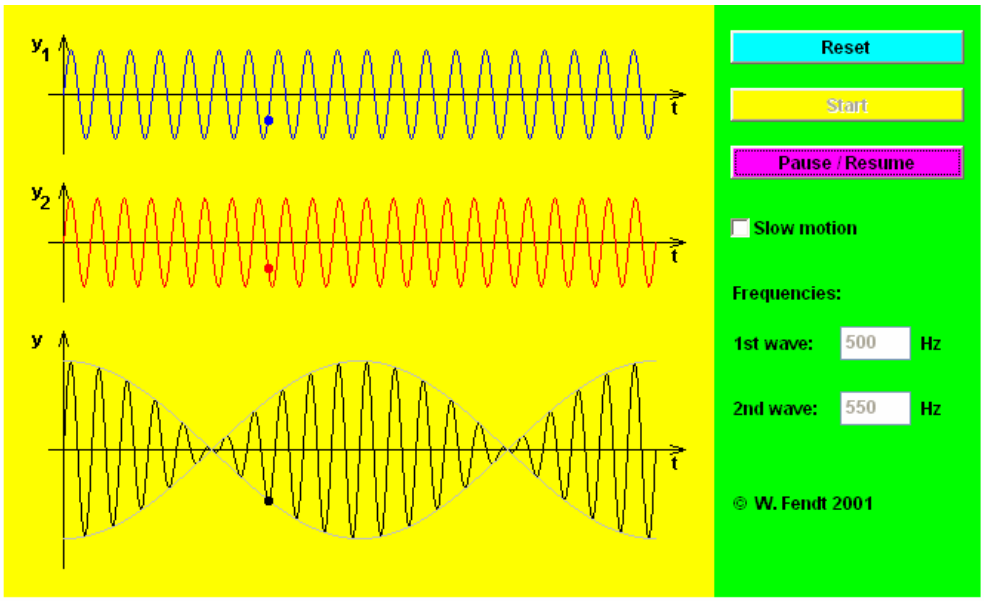
Zbog početnog uslova za razliku kružnih frekvencija, možemo uvesti novu funkciju amplitude koja zavisi od vremena i ima oblik:

$A(t) = 2A\cos\frac{\omega_2-\omega_1}{2}t$ i koja predtavlja amplitudu rezultujućeg oscilovanja. Amplituda se menja sa vremenom sa frekvencijom $\omega_A = \frac{\omega_2-\omega_1}{2}$ odnosno periodom

$$T_A = \frac{2\pi}{\omega_A} = \frac{4\pi}{\omega_2-\omega_1} \qquad \omega_2 \to \omega_1 \quad T_A \to \infty$$

što znači da amplituda takvih oscilacija teži konstantnoj vrednosti $2A$ koja ne zavisi od vremena (Slika 9.).

Drugi jednostavan slučaj dabiranja dva ili više harmonijskih oscilovanja istog pravca, dešava se kad se periodi odnose kao celi brojevi. U ovom slučaju rezultujuće oscilovanje će imati period kao i komponentno oscilovanje sa najvećom periodom, samo će njegov oblik biti vrlo složen.



Slika 9.

PREDSTAVLJANJE NEHARMONIJSKIH OSCILATORNIH PROCESA POMOĆU HARMONIJSKIH OSCILACIJA

Ukoliko saberemo više harmonijskih oscilovanja čije su frekvencije celobrojni umnošci neke osnovne frekvencije , dobijamo rezultujuće oscilovanje koje više nije jednostavna harmonijska funkcija već složena periodična funkcija $f(t)$, koja se ponavlja nakon perioda T .

Ukoliko uzmemo oscilatorna kretanja prikazana relacijama:

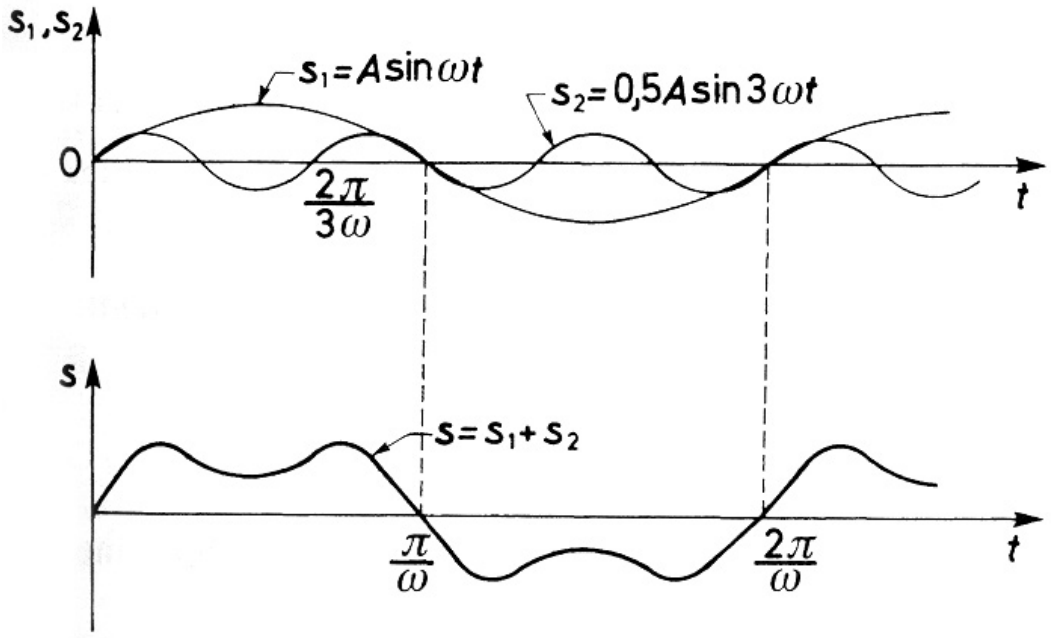
$$s_1 = 2\sin\omega t \qquad s_2 = \sin 3\omega t$$

tada je njihov zbir jednak:

$$s = s_1 + s_2 = 2\sin\omega t + \sin 3\omega t$$

Rezultujuće oscilovanje više nije harmonijsko ali je periodična funkcija perioda $T_1 = \dfrac{2\pi}{\omega}$

(Slika 10.).



Slika 10.

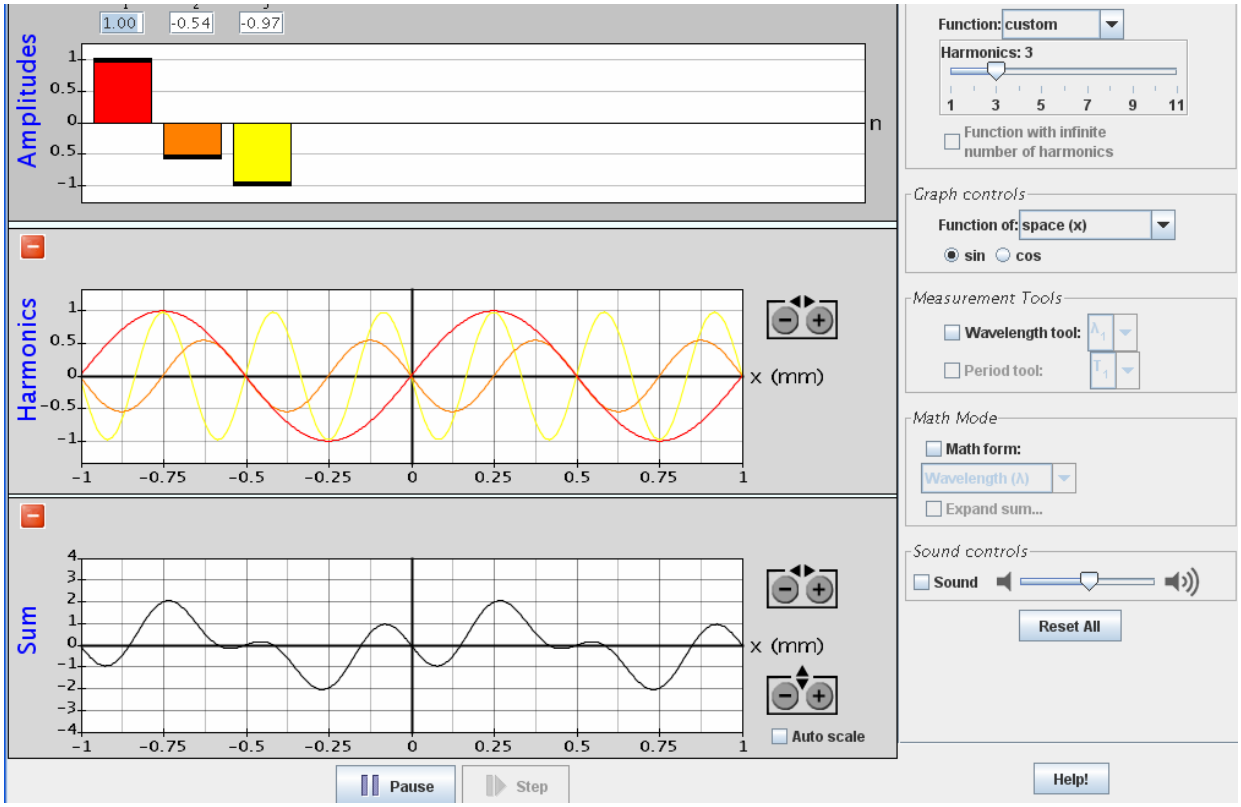
Frekvencija ω se zove **osnovna frekvencija** a celobrojni umnošci ove frekvencije $2\omega, 3\omega, 4\omega$.....n$\omega$,.... zovu se **viši harmonici.** Sabirajući jednostavne harmonijske oscilacije čije su frekvencije celobrojni umnošci osnovne frekvencije, možemo uzimajući odgovarajući

broj viših harmonika sa odabranim amplitudama, dobiti bilo koju periodičnu funkciju. Primer je dat na slici 10.a.

Isto tako se svaka periodična funkcija $\mathrm{f}(t)$ perioda T može izraziti beskonačnim redom harmonijskih članova:

$$\mathrm{f}(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos n\omega t + b_n \sin n\omega t \right)$$

To je poznati Furijeov red za periodičnu funkciju $\mathrm{f}(t)$ s periodom $T = \frac{2\pi}{\omega}$. U praksi se amplitude članova višeg reda sve više smanjuju tako da se Furijeov red često sastoji od samo nekoliko članova.



Slika 10.a.

Furijeovi koeficijnti dati su formulama:

$$a_n = \frac{2}{T}\int_0^T \mathrm{f}(\mathrm{t})\cos n\omega t\, dt \qquad n = 0,1,\ 2,....$$

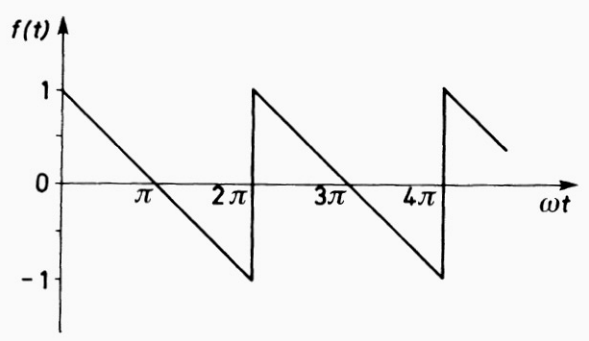
$$b_n = \frac{2}{T}\int_0^T \mathrm{f}(t)\sin n\omega t\, dt \qquad n = 1,2,3.....$$

Rastavljanje nesinusne periodične funkcije na njene harmonijske komponente se naziva *spektralna analiza*.

Grafički prikaz amplitude kao funkcije frekvencije zove se spektar periodičnog nesinusnoh oscilovanja. Takav spektar je diskretan i sastoji se od frekvencija $\omega, 2\omega, 3\omega,....$

Primer 1:

Na slici 11. je prikazana testerasta kriva. Tako se prikazuje jačina struje u vertikalnom otklonu snopa elektrona koji pada na ekran televizora.



Slika 11.

To je periodična funkcija čija se jednačina u intervalu $\left[0,\ \frac{2\pi}{\omega}\right]$ može napisati kao:

$$\mathrm{f}(t) = -\frac{\omega}{\pi}t + 1$$

Ova funkcija se može prikazati kao suma harmonijskih oscilovanja odnosno može se razviti u Furijeov red:

$$\mathrm{f}(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty}\left(a_n \cos n\omega t + b_n \sin n\omega t\right)$$

Pomoću date formule možemo izračunati Furijeove koeficijente:

$$a_n = \frac{2}{T}\int_0^T\left(-\frac{\omega}{\pi}t + 1\right)\cos n\omega t\, dt = 0$$

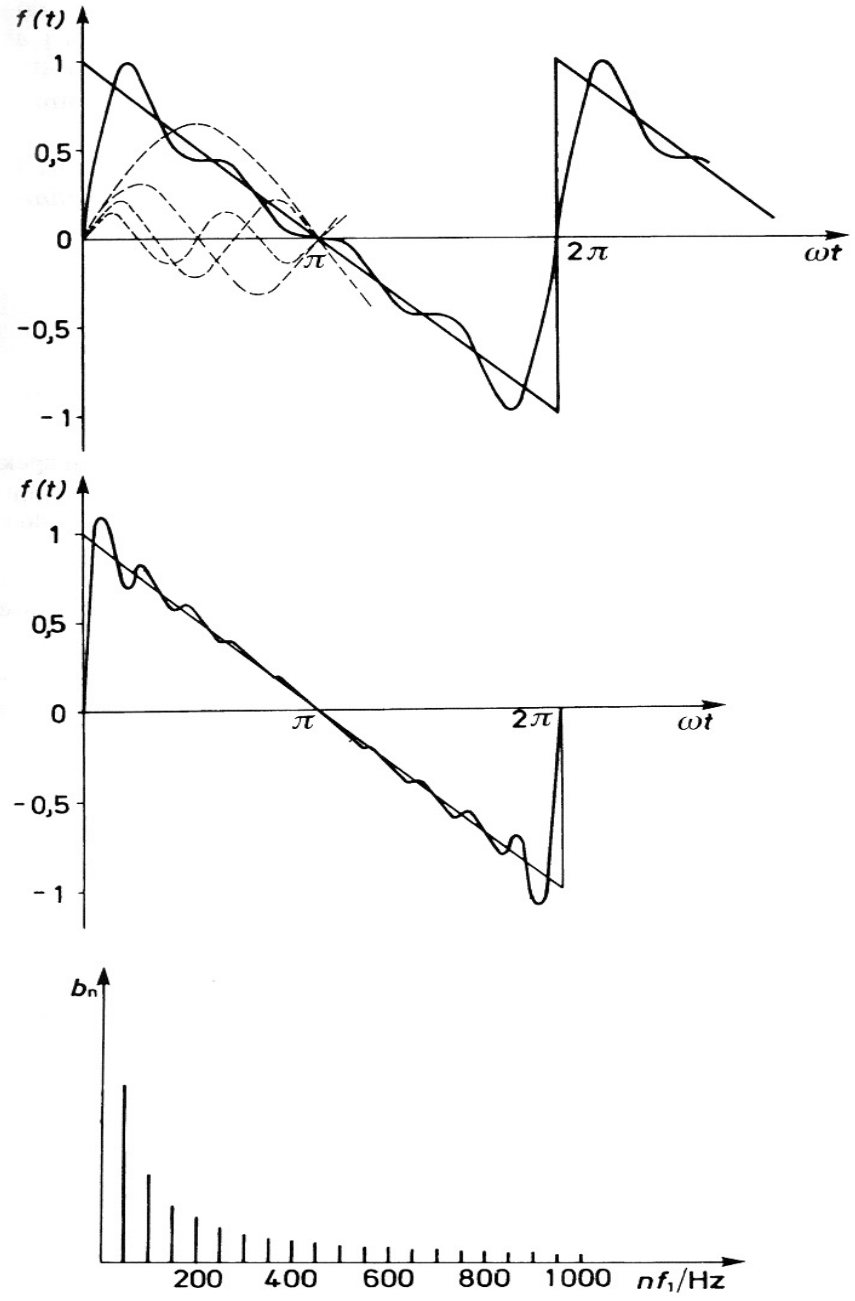
$$b_n = \frac{2}{T}\int_0^T\left(-\frac{\omega}{\pi}t + 1\right)\sin n\omega t\, dt = \frac{2}{n\pi}$$

Furijeov red za testerastu funkciju prikazanu na slici 11. glasi:

$$\mathrm{f}(t) = \frac{2}{\pi}\left(\sin\omega t + \frac{\sin 2\omega t}{2} + \frac{\sin 3\omega t}{3} + \frac{\sin 4\omega t}{4} + ...\right)$$

Što više članova reda uzmemo u obzir dobićemo bolju aproksimaciju. Na Slici 12.a. prikazan je zbir prva četiri člana reda a na Slici 12.b zbir prvih deset članova reda. Diskretan spektar sa

vrednostima koeficijenata b_n koji predstavljaju amplitude pojedinih harmonika prikazan je na Slici 12.c.

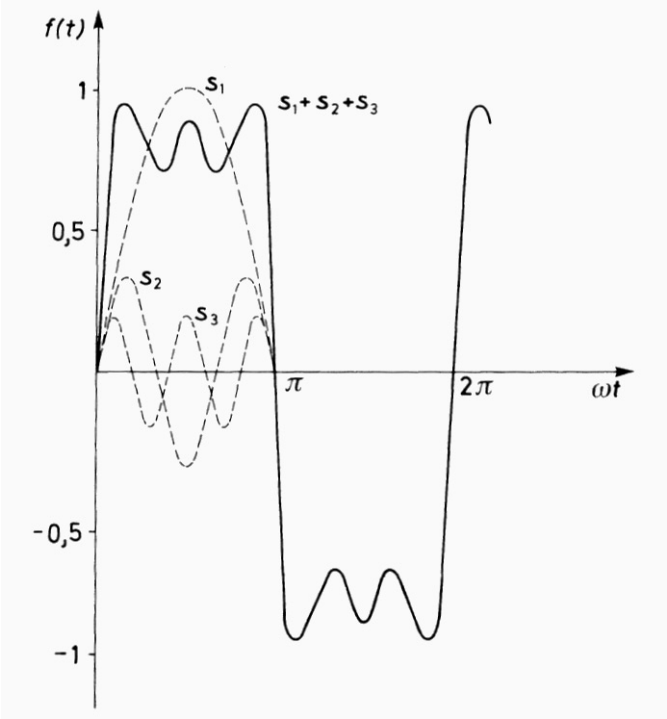


Slika 12.

Primer 2.

Odrediti funkciju $f(t)=\left(\sin\omega t-\frac{1}{3}\sin 3\omega t+\frac{1}{5}\sin 5\omega t\right)$ u intervalu $0\langle\omega t\langle 2\pi$ sabirajuci pojedine harmonijske oscilacije. Zadata funkcija je zbir tri harmonijske oscilacije različitih

amplituda i frekvencija $\omega,\ 3\omega,\ 5\omega$. Ako nacrtamo oscilacije i saberemo ih Slika 13.a. Dobijamo periodočnu nesinusoidalnu funkciju perioda $\frac{2\pi}{\omega}$.



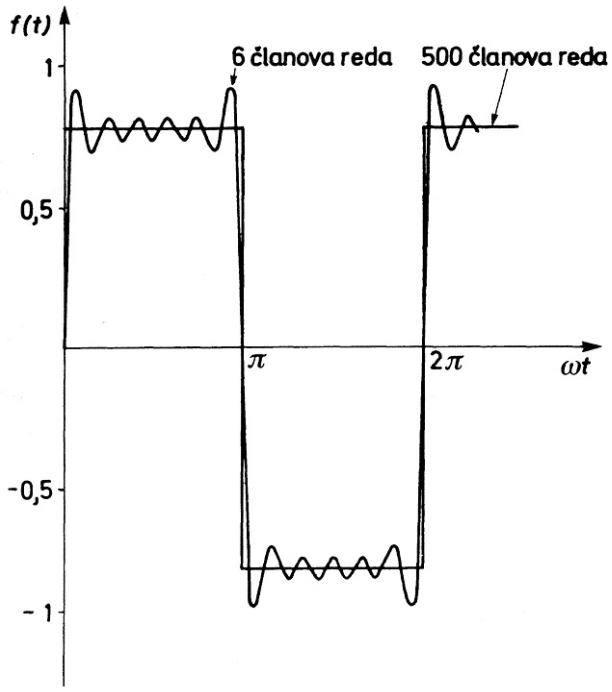
Slika 13.a

Ako umesto prva tri članaza funkciju $f(t)$ saberemo veliki broj članova reda kao što se vidi na slici 13.b. Tada ćemo dobiti periodičnu funkciji oblika

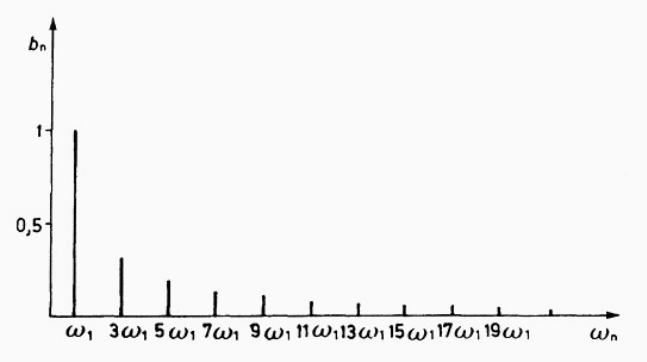
$$f(t)=\frac{\pi}{4} \quad \text{za } 0 \langle \omega t \langle 2\pi \qquad \text{i}$$

$$f(t)=-\frac{\pi}{4} \quad \text{za } \pi \langle \omega t \langle 2\pi$$

Diskretni spektar ovog oscilovanja prikazan je na Slici 13.c.



Slika 13.b



Slika 13.c.

FURIJE ANALIZA NEPERIODIČNIH FUNKCIJA

Ukoliko se neperiodične funkcije analiziraju Furijeovom metodom tada umesto diskretnog spektra imamo *kontinuirani spektar frekvencija*.

Kretanje može biti oscilatorno ali da nema periodični karakter. Takvo aperiodično kretanje (neprimer prigušeno oscilovanje) ne može da se razvije u Furijeov red sa neprekidnim nizom frekvencija koje su celobrojni umnožak osnovne frekvencije ω. Ovakvo kretanje može da se razloži na beskonačan niz harmonijskih oscilatornih kretanja, pri čemu se frekvencija «susednih» oscilacila beskonačno malo razlikuje jedno od druge a amplituda ΔA_i pojedinačnih elementarnih oscilacija, beskonačno su male.

$$F(\omega t)=\frac{1}{\pi}\int_0^{\infty} d\alpha \int_{-\infty}^{+\infty} F(\theta)\cos\alpha\,(\theta-\omega t)\,d\theta$$

Takvom oscilovanju grafički ne odgovara «linijski spektar» već «neprekidni spektar», što znači da postoje oscilacije sa svim frekvencijama.

REZONANCIJA

Šta bi bio fizički smisao razlaganja neharmonijskog procesa na harmonijske? Pretpostavimo da je zavisnost nekog procesa od vremena predtavljena funkcijom $f(t)$ koja se može razložiti u Furijeov red.

$$f(t)=\sum_{n=0}^{\infty}\left(A_n\cos n\omega t+B_n\sin n\omega t\right)$$

Za registrovanje samo jednog člana reda zasebno (jednog harmonika) treba podesiti takve uslove eksperimenta pri kojima bi se ispoljio samo taj harmonik. Pretpostavimo da je $f(t)$ prinudna sila , koja deluje na sistem (rezonator), koji može da vrši prinudno oscilovanje. Neka taj rezonator ima sopstvenu frekvenciju oscilovanja koja se podudara sa frekvencijom $k\omega$ jednog od harmonika reda. Ako je rezonantna kriva rezonatora tako oštra da frekvencija susednih harmonika $(k+1)\omega$ leže u oblasti malih amplituda prinudnog oscilovanja, rezonator će praktično vršiti oscilovanje samo sa frekvencijom $k\omega$ i amplitudom koja je proporcionalna amplitudi harmonika. Menjajući rezonantnu frekvenciju rezonatora mogu se *sukcesivno* stvarati uslovi za registrovanje ostalih harmonika iz reda.

Ovakve pojave se sreću kod uređaja za analizu spektralnog sastava nekog fizičkog procesa, na primer, svetlosti ili zvuka ili električnih oscilacija.

SLAGANJE UZAJAMNO NORMALNIH HARMONIJSKIH OSCILACIJA

Ukoliko sabiramo uzajamno normalne oscilacije, rezultujuće kretanje se može opisati na sledeći način:

$$x = a_1 \cos(\omega t + \varphi_1)$$

$$y = a_2 \cos(\omega t + \varphi_2)$$

$$\frac{x^2}{a_1^2} + \frac{y^2}{a_2^2} - \frac{2xy}{a_1 a_2}\cos(\varphi_2 - \varphi_1) = \sin^2(\varphi_2 - \varphi_1)$$

Jednačina predstavlja jednačinu elipse čije su karakteristike zavisne od fazne razlike $(\varphi_2 - \varphi_1)$ ukoliko amplitude oscilovanja a_1 *i* a_2 imaju određene vrednosti.

a) Neka je $\varphi_2 - \varphi_1 = 0$ *ili* $2k\pi$ $k = 1,2,3,....$

Jednačina kretanja tada ima oblik:

$\left(\frac{x}{a_1} - \frac{y}{a_2}\right)^2 = 0$ $y = \frac{a_2}{a_1}x$ odnosno jednačina prave.

Ovakvo oscilovanje se naziva *linearno polarizovano.* Položaj tačke u odnosu na koordinatni početak se izražava relacijom:

$s = \sqrt{x^2 + y^2} = \sqrt{a_1^2 + a_2^2}\cos(\omega t + \varphi)$

b) Neka je fazna razlika komponentnih oscilacija $\varphi_2 - \varphi_1 = \pi$ *ili* $(2k+1)\pi$

Jednačina kretanja tada ima oblik:

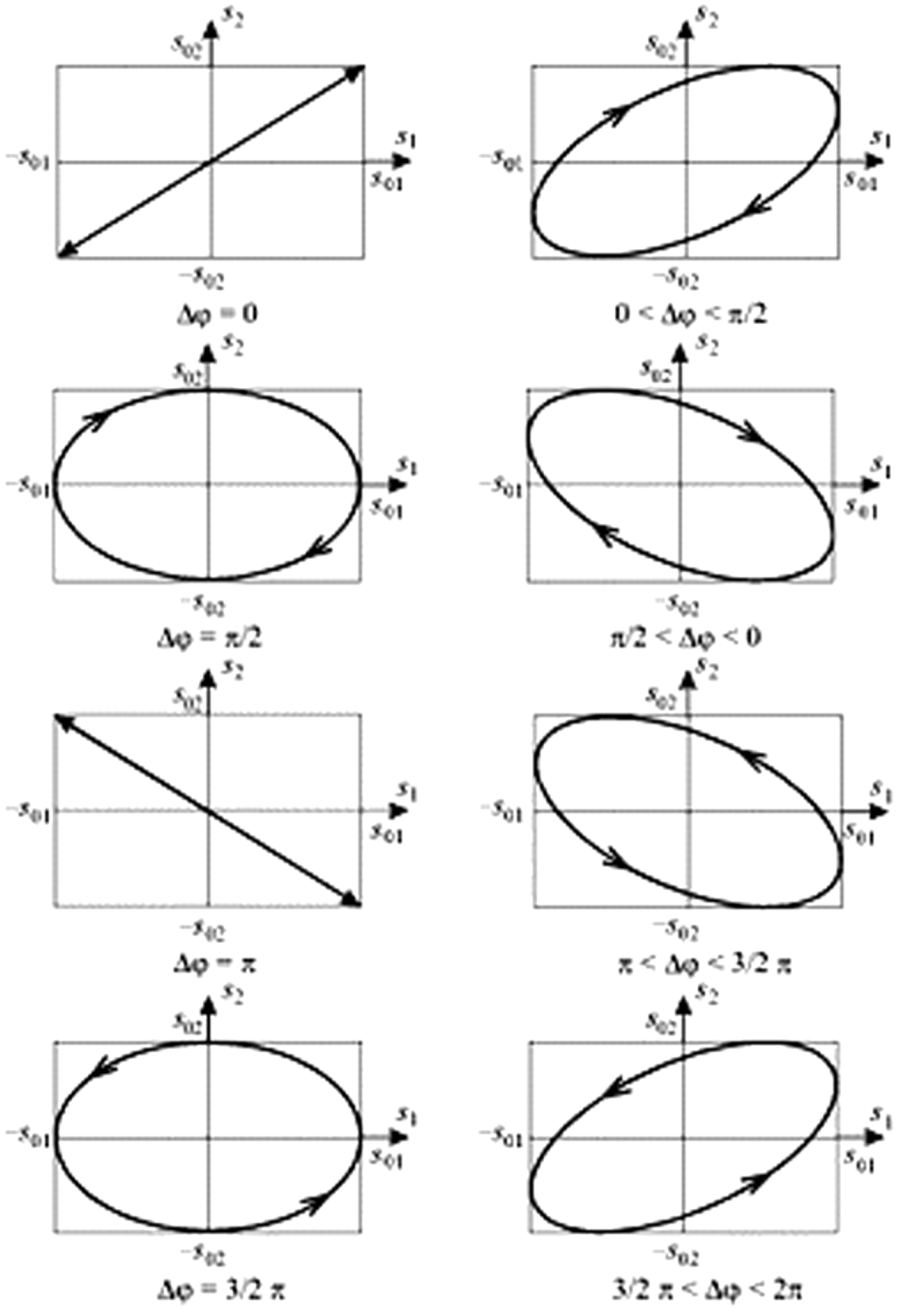
$\frac{x^2}{a_1^2} + \frac{y^2}{a_2^2} + \frac{2xy}{a_1 a_2} = 0$ ili $\left(\frac{x}{a_1} + \frac{y}{a_2}\right)^2 = 0$ $y = -\frac{a_2}{a_1}x$ koje reprezentuje linerno kretanje ili jednačinu prave koja prolazi kroz II i IV kvadrant koordinatnog sistema, odnosno razlikuje se od prehodnog kretanja po koeficijentu pravca prave.

c) Neka je fazna razlika komponentnih oscilacija $\varphi_2 - \varphi_1 = \frac{\pi}{2}$

Tada je rešenje jednačine oblika:

$\frac{x^2}{a_1^2}+\frac{y^2}{a_2^2}=1$ To je jednačina elipse čije su poluose jednake amplitudama komponentnih oscilacija. Ovakvo oscilovanje se naziva *eliptično-polarizovano* (Slika 14.)

U slučaju da su amplitude komponentnih oscilacija jednake $a_1 = a_2 = a$ jednačina prelazi u: $x^2 + y^2 = 1$ što predstavlja jednačinu kruga a oscilovanje se naziva *cirkularno-polarizovano*.



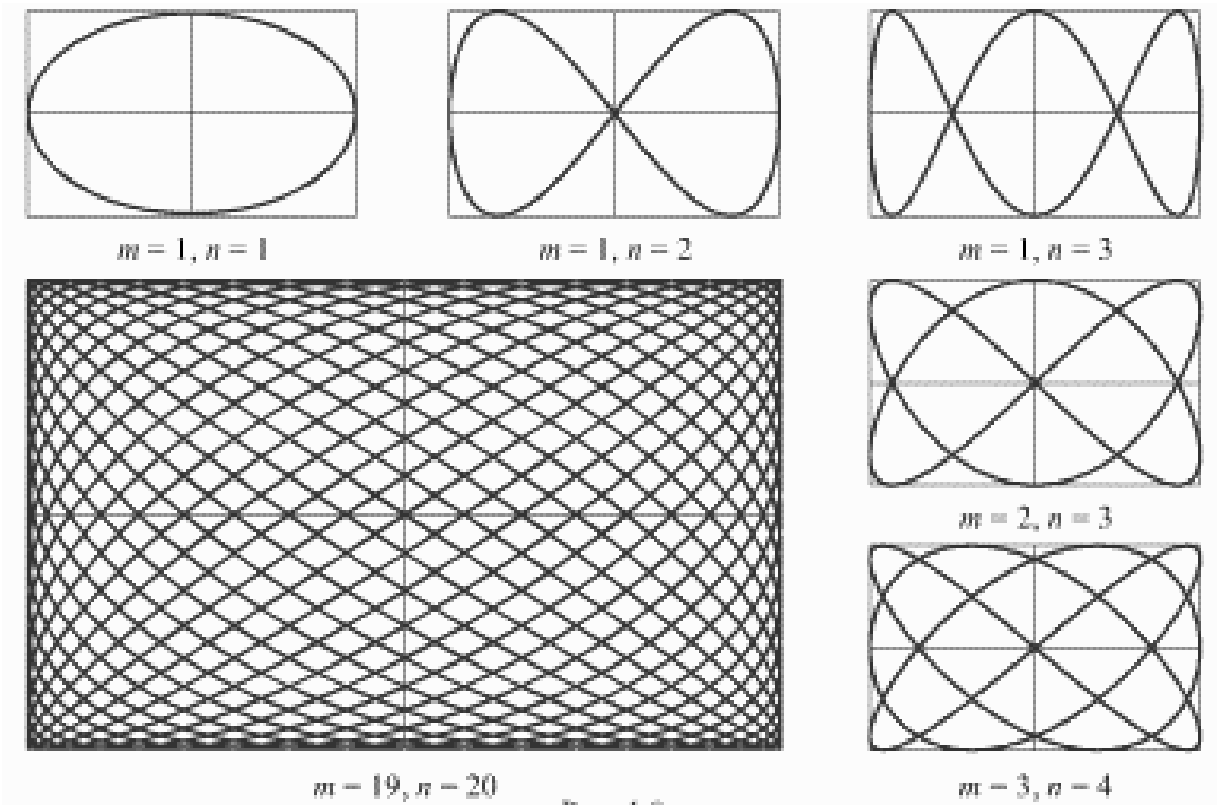
Slika 14.

Lisažuove figure su krive linije koje predstavljaju putanje materijalne tačke koja istovremeno osciluje u dva međusobno normalna pravca. U opštem slučaju: **amplitude, periode i fazna razlika** komponentnih oscilacije mogu biti različite.

Promena odnosa **amplituda** izaziva promenu **oblika** elipse od kruga do prave.

Promena **fazne razlike** $\Delta\varphi = \varphi_2 - \varphi_1$ izaziva promenu elipse rezultujuće putanje i po **obliku** i po **orijentaciji**.

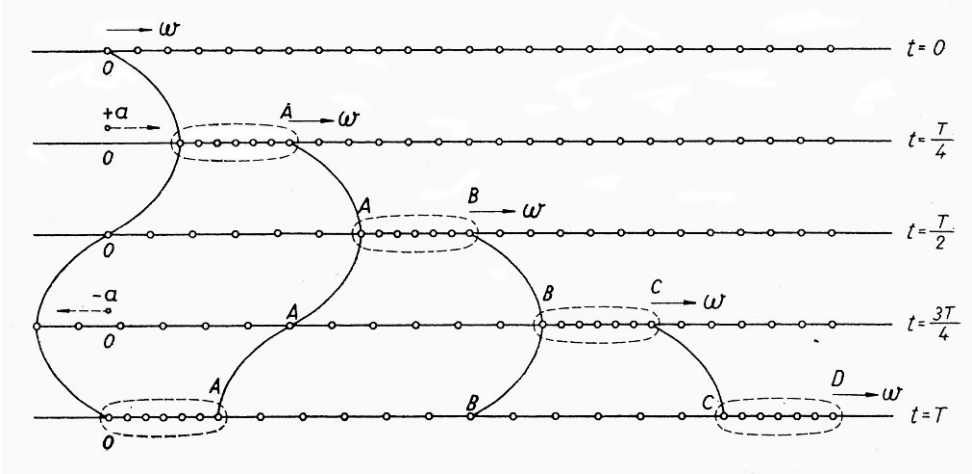
Razlika perioda izaziva neprekidnu promenu fazne razlike koja utiče na deformaciju elipse dajući vrlo komplikovane figure (Slika 15.)



Slika 15.

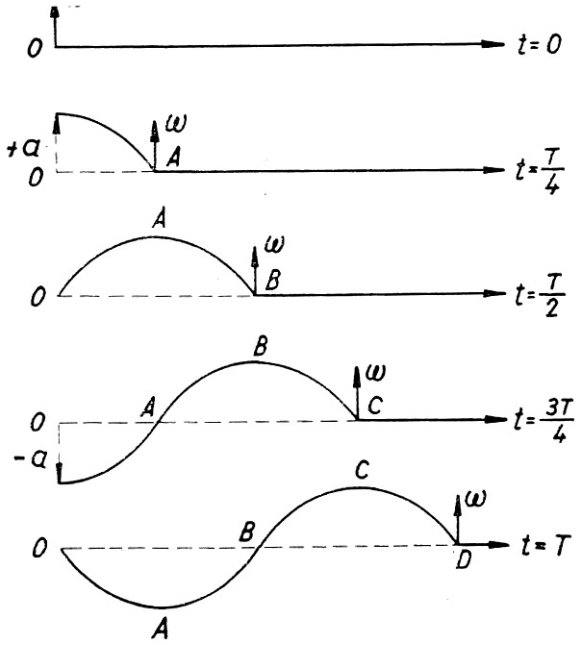
TALASI

Pojava širenja oscilovanja u nekoj sredini naziva se *talas*. Ako se tačka koja osciluje nalazi u materijalnoj sredini čiji su delići povezani među sobom tada se energija oscilovanja tačke predaje okolnim tačkama izazivajući njihovo oscilovanje. Pri širenju oscilovanja, oscilujući delići se ne premeštaju zajedno sa prostiranjem oscilovanja, nego osciluju oko svojih ravnotežnih položaja. Ako delić osciluje po pravoj duž koje se širi talas, takav talas se



zove l o n g i t u d i n a l a n.

Ako je oscilovanje delića normalno na pravac prostiranja talasa, takav talas se zove t r a n s v e r z a l a n.

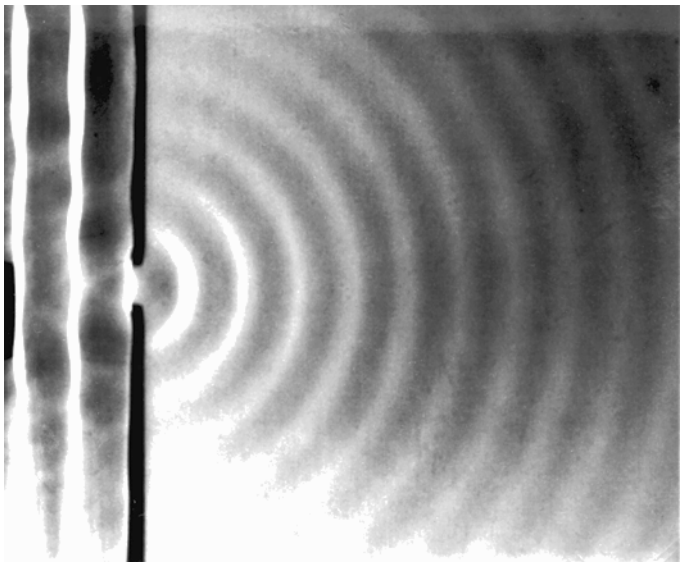


Ako u materijanoj sredini ne postoje sile elastičnosti pri međusobnom pomeranju paralelnih slojeva, tada se transverzalni talasi ne mogu obrazovati. Tečnosti i gasovi predstavljaju sredine u kojima se transverzalni talasi ne prostiru. Ako u sredini nastaju sile elastičnosti pri deformaciji sabijanja i istezanja, u takvoj sredini se mogu širiti longitudinalni talasi (tečnosti, gasovi i čvrsta tela).

Rastojanje koje određena faza oscilovanja pređe u jednom periodu oscilovanja, naziva se t a l a s n a d u ž i n a λ. Talasna dužina predstavlja najmanje rastojanje između tačaka sredine koje osciluju u istoj fazi. Pod brzinom prostiranja talasa podrazumeva se f a z n a b r z i n a V. Početna faza se za vreme jednako p e r i o d u T pomerilo za rastojanje jednako talasnoj dužini λ.

$$V = \frac{\lambda}{T}$$

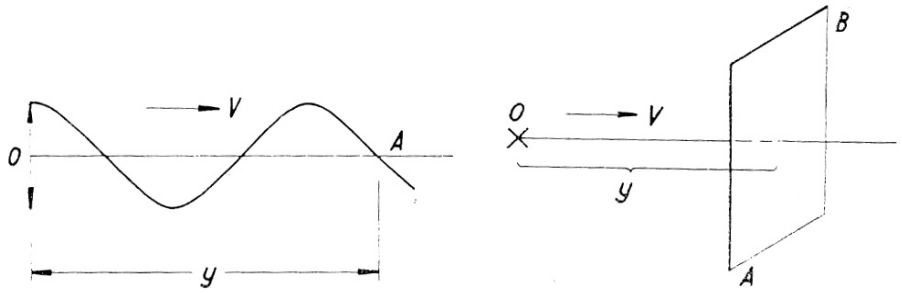
Geometrijsko mesto tačaka do kojih je u nekom vremenu došlo oscilovanje nazivamo t a l a s n i m f r o n t o m. U materijalnoj sredini možemo izdvojiti geometrijsko mesto tačaka koje osciluje sa istim fazama. Ovaj skup tačaka obrazuje površinu istih faza ili tzv, t a l a s n u p o v r š i n u. Talasni front je poseban slučaj talasne površine. Na slici 16. dat je primer širenja talasa na površini vode u talasnoj kadi (koja služi za demonstriranje prostiranja talasa u materijalnoj sredini). Vidi se nailazak ravnog talasa (tamne i svetle uspravne pruge pre otvora) na otvor koji po Hajgensovom principu postaje izvor novog sfernog talasa čiji se kružni oblik talasnog fronta (površine) formira posle otvora na koji je naišao poremećaj.



Slika 16.

TALASNA JEDNAČINA

Talasni proces će biti poznat ako se zna vrednost x u svakom trenutku za svaku tačku prave duž koje se talas prostire. Treba znati pomeranje x kao funkciju vremena i koordinata



ravnotežnog položaja tačke.Ako sa O označimo koordinatni početak i centar oscilovanja

Slika 17.

(Slika17.) tada se oscilacije u toj tački izvode po zakonu:

$$x = a\cos\omega t$$

gde je a-amplituda oscilovanja, ω-kružna frekvencija, t-vreme računato od početka oscilovanja. Oscilacije koje se šire od tačke O doći će do tačke A posle vremena τ:

$$\tau = \frac{y}{V}$$

gde je V brzina prostiranja poremećaja (talasa) u datoj materijalnoj sredini.

Tačka A poćinje da osciluje sa vremenom τ kasnije od tačke O. Smatrajući da se talasi koji se šire duž posmatrane prave ne prigušuju, dobijamo da tačka A kada talas dođe do nje počinje da osciluje sa amplitudom a i kružnom frekvencijom ω po relaciji:

$$x = a\cos\omega t'$$

t' - vreme računato od trenutka kada je tačka A počela da osciluje $t' = t - \tau$

$$x = a\cos\omega\left(t - \tau\right)$$

$$x = a\cos\omega\left(t - \frac{y}{V}\right)$$

koja predtavlja jednačinu ravnog talasa koji se širi u pravcu y. Ako zamislimo ravan talas koji se prostire u smeru suprotnom od smera u kome raste rastojanje y tada je:

$$x = a\cos\omega\left(t + \frac{y}{V}\right)$$

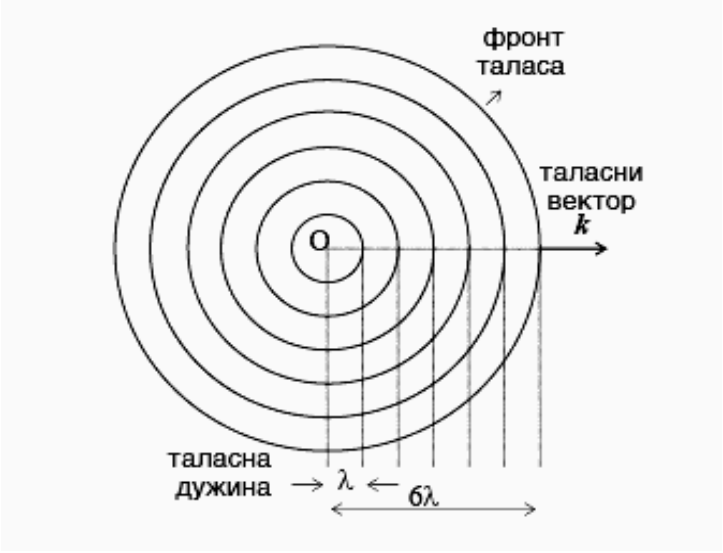
Talas ima i prostornu i vremensku periodičnost. Dati delić sredine koji se odlikuje određenom vrednošću y, vrši u toku vremena harmonisko oscilatorno kretanje:

$$x = a\cos\omega\left(t - \frac{y}{V}\right) = a\cos\left(\omega t - \alpha\right) \qquad \alpha = \frac{\omega y}{V} = 2\pi\frac{y}{\lambda}$$

Veličina α je konstanta za datu tačku i predstavlja početnu fazu oscilovanja u toj tački. Dve tačke kojima odgovaraju odstojanja y_1 i y_2 od koordinatnog početka imaju razliku faza:

$$\alpha_2 - \alpha_1 = 2\pi\frac{y_2 - y_1}{\lambda}$$

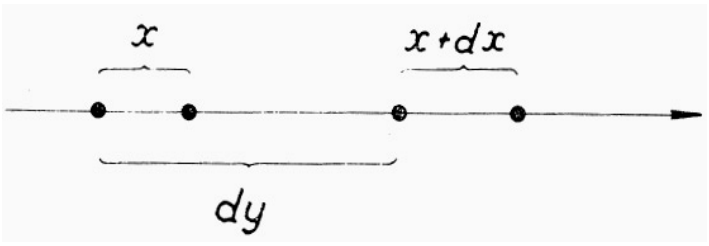
Dve tačke koje se nalaze na međusobnom rastojanju koje je jednako talasnoj dužini λ: $y_2 - y_1 = \lambda$ imaju faznu razliku $\alpha_2 - \alpha_1 = 2\pi$. Takve tačke osciluju u **istoj fazi.** Za tačke čije je međusobno rastojanje $y_2 - y_1 = \frac{\lambda}{2}$ fazna razlika iznosi $\alpha_2 - \alpha_1 = \pi$ i za takve tačke se kaže da osciluju u **suprotnim fazama**. Kod sfernih talasa (Slika 18.) amplituda se smanjuje obrnuto proporcionalno rastojanju r od izvora oscilovanja: $x = \frac{a}{r}\cos\omega\left(t - \frac{r}{V}\right)$



Slika 18.

DINAMIKA PROSTIRANJA OSCILACIJA U ELASTIČNOJ SREDINI

U materijalnoj sredini će se širiti one oscilacije koje su uslovljene pomeranjem izazvanim elastičnom deformacijom te sredine. Izaberimo niz tačaka koje pripadaju neprekidnoj sredini i leže na jednoj pravaoj duž koje se širi longitudinalni talas. Pomerenje neke tačke, koja leži na toj pravoj, od ravnotežnog položaja označimo sa x. Rastojanje između tačaka duž te prave označimo sa y. Pomeranje tačaka na rastojanju dy menja se za



Slika 19.

veličinu dx (Slika 19.): $\frac{dx}{dy} = s$ gde je s relativna deformacija sredine. Kada je s > 0 rastojanje između tačaka se povećava što označava istezanje sredine, dok je s < 0 rastojanje između tačaka se smanjuje tj. sredina se sabija. Imajući u vidu talasnu jednačinu može se naći veza između relativne deformacije i brzine poremećaja:

$$v = \frac{dx}{dt} = -a\omega \sin \omega\left(t - \frac{y}{V}\right)$$

$$s = \frac{dx}{dy} = \frac{a\omega}{V} \sin \omega\left(t - \frac{y}{V}\right)$$

$$\frac{dx}{dt} = -V\frac{dx}{dy}$$

Deformacija sredine po apsolutnoj vrednosti je najveća u onim tačkama u kojima je brzina oscilujućih čestica najveća u oblasti u kojoj tačka prolazi kroz ravnotežni položaj.

Iz talasne jednačine:

$$x = a\cos\omega\left(t - \frac{y}{V}\right)$$

dvostrukim diferenciranjem po vremenu i po položaju mogu se dobiti relacije:

$\frac{d^2x}{dt^2} = -a\omega^2 \cos\omega\left(t - \frac{y}{V}\right)$ i $\frac{d^2x}{dy^2} = -\frac{a\omega^2}{V^2}\cos\omega\left(t - \frac{y}{V}\right)$ koje su povezane diferencijalnom jednačinom talasnog kretanja:

$$\frac{d^2x}{dt^2} = V^2 \frac{d^2x}{dy^2}$$

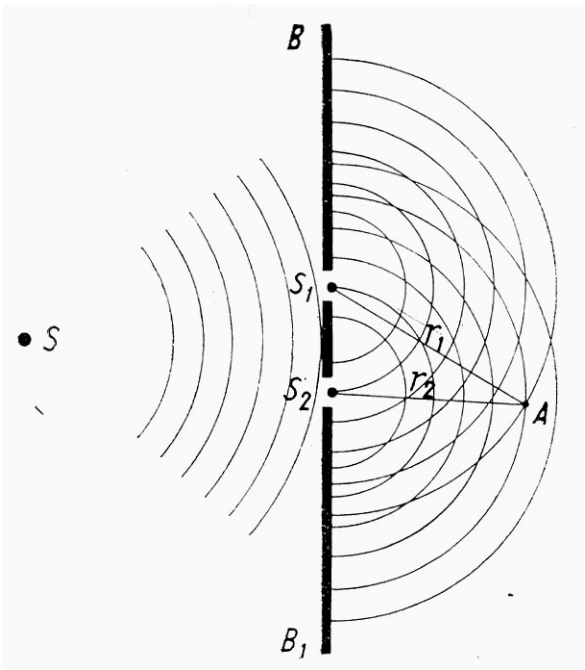
Rešenje ove jednačine opisuje širenje talasnog kretanja proizvoljnog oblika, brzinom V u nekoj materijalnoj sredini.

INTERFERENCIJA TALASA

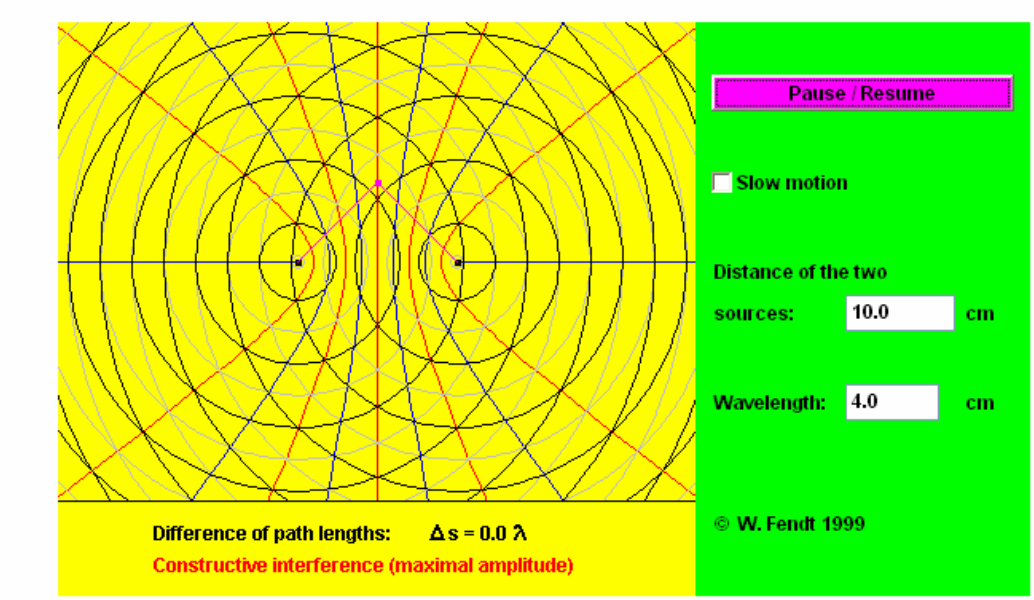
U nekoj materijalnoj sredini mogu da se istovremeno šire oscilacije koje polaze iz raznih centara oscilovanja (Slika 20.). Ako se dva različita sistema talasa, koji polaze iz različitih izvora poklaapaju u nekoj oblasti, a zatim ponovo razilaze, tada će se svaki taalas dalje kretati kao da se nisu ni sreli. To je *princip superpozicije*.

U oblasti poklapanja talasa, oscilacije se superponiraju jedna na drugu i nastaje *interferencija talasa*.

Kada izvori talasa osciluje istom frekvencijom, imaju iste pravce oscilovanja i iste faze ili stalnu faznu razliku, takvi izvori se nazivaju **koherentnim.** Ovakvo slaganje oscilovanja se naziva interferencijom koja potiče od koherentnih izvora.



a)



b)

Slika 20.

ZVUČNI TALASI

Zvučni izvori emituju zvučne talase u okolni prostor. Ukoliko je sredina u kojoj se zvučni izvor nalazi izotropna, zvučni talasi se prostiru istom brzinom u svim pravcima. Kada se zvučni talas širi, prenosi se mehanički poremećaj sredine koji je nastao u izvoru zvuka. Kako se na ovaj način vrši prenošenje energije od zvučnog izvora kroz okolnu sredinu, to se kaže da izvor 'zrači' zvučnu energiju u okolnu sredinu.

Ljudsko uho čuje oscilacije koje se prenose kroz materijalnu sredinu u rasponu frekvencija približno 20 do 20000 Hz i za mehaničke talase tih frekvencija kažemo da su *zvučni talasi*.

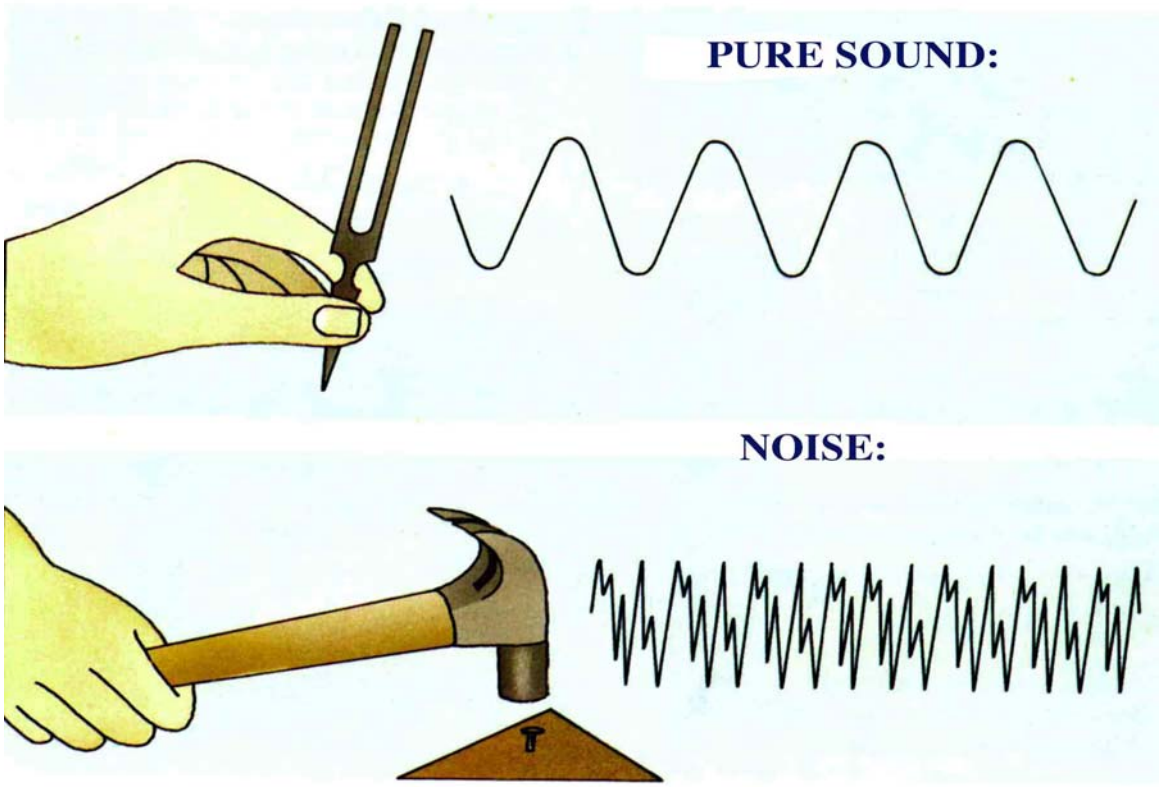
Mehanizam prostiranja zvučnih talasa u svin agregatnim stanjima je da su to *longitudinalni talasi*, koji delujući na čovekovo uho izazivaju u svesti osećaj čuvenja. Razlikujemo tri vrste zvuka:ton, šum i prasak.

Ton nastaje periodičnim oscilovanjem izvora. Ako je oscilovanje harmonijsko nastaje prost ili čist ton čije se prostiranje opisuje jednačinom ravnog talasa. Složenom tonu odgovara anharmonijsko oscilovanje koje može biti razloženo na harmonijsko. Skup svih prostih tonova određenih frekvencija i amplituda daje *akustični spektar*, koji je važna fizička osobina složenog tona. Prost ton nastaje oscilovanjem zvučne viljuške (Slika 21.), zategnute žice,

metalnog štapa, vazdušnog stuba itd., dok složene tonove mogu proizvesti muzički instrumenti, glasovni aparati čoveka i životinja, itd.

Šum je zvuk koji se sastoji od oscilacija složene neperiodične prirode. Možemo ga posmatrati kao neperiodično promenljivi složeni ton. Nastaje glasnim govorom više ljudi, vibracijama (Slika 19) i pri radu mašina, kao i u saobraćaju.

Prasak (zvučni udar) predtavlja složeni zvučni talas koji naglo nastaje (pucanj, eksplozija) i brzo dostiže maksimum intanziteta da bi isto tako nestao bez ponavljanja.



Slika 21.

Odgovor uva na zvučne talase zavisi od promene viška pritiska vazduha ili sredinu kroz koju se poremećaj prostire, a ne od promene pomeranja oscilatora. Neki tipovi mikrofona tđ. Deluju na principu promene pritiska. Upravo stoga je korisno da se talas posmatra kao *talas pritiska* a ne kao talas pomeranja.

$$p = -B\frac{dx}{dy}$$

Ako pomeranje u materijalnoj sredini prikažemo relacijom:

$$x = a\sin(ky - \omega t)$$

gde je $k = \frac{\omega}{V} = \frac{2\pi}{\lambda}$ intenzitet talasnog vektora, dobijamo izraz za lokalnu promenu pritiska

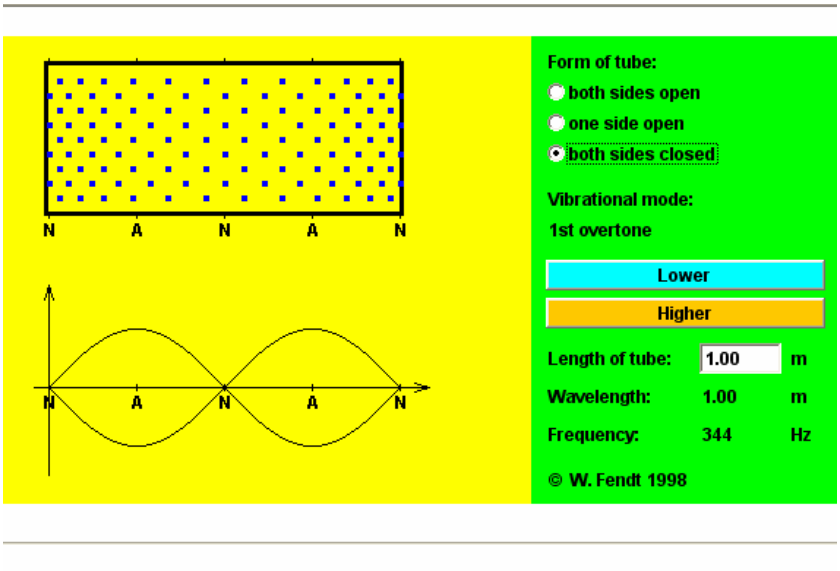
$p = -kB\,a\cos(ky - \omega t)$

gde $kB\,a$ predstavlja maksimalni višak pritiska ili *amplitudu pritiska* p_m

$$p = -p_m\cos(ky - \omega t)$$

ZVUČNI UDARI

Stacionarni talasi u vazdušnom stubu mogu se navesti kao primer interferencije talasa. Oni nastaju kada se dva talasna niza iste amplitude i frekvencije kreću kroz istu oblast prostora u suprotnim smerovima (Slika 22.).



Slika 22.

Može se posmatrati i druga vrsta interferencije koja nastaje kada se kroz istu oblast prostora kreću dva talasna niza jednakih amplituda ali različitih frekvencija. Amplituda rezltujućeg oscilovanja se menja sa vremenom. Ove promene prouzrokuju promene u subjektivnoj jačini zvuka koji se naziva *zvučnim udarima*.

Rezultujuće oscilovanje ima frekvenciju $\frac{\omega_1 + \omega_2}{2}$. Amplituda se menja sa vremenom sa frekvencijom $\frac{\omega_1 - \omega_2}{2}$. Ako su frekvencije ω_1 i ω_2 bliske tada je njihova razlika mala i

amplituda fluktuira vrlo sporo. Kada je amplituda velika, zvuk je jak a kada je mala zvuk je jako slab. Zvučni udari ili maksimumi amplitude se pojavljuju kada je $\cos\frac{\omega_1-\omega_2}{2}t=\pm 1$. Pošto se svaka od ovih vrednosti pojavljuje u jednom periodu, broj udara u sekundi jednak je razlici frekvencija talasa.

FURIJE ANALIZA TALASNOG KRETANJA

Kod oscilatornog kretanja je pokazano da se svako periodično kretanje (periodična vremenska funkcija) može prikazati kao zbir sinusnih (harmonijskih) funkcija frekvencije ω, 2ω, 3ω,....nω. To se može proširiti i na neperiodične funkcije i umesto diskretnih, prikazati ih sa kontinuiranim spektrom frekvencija.

Talasno kretanje je takođe periodićna funkcija ali dve promenljive: vremena i prostora. Na periodičnu funkciju $f=f\left(t\pm\frac{y}{V}\right)$ koja označava talasno kretanje, takođe se može primeniti Furijeova analiza:

$$f\left(t\pm\frac{y}{V}\right)=\frac{a_0}{2}+\sum a_n\cos n\,\omega\left(t\pm\frac{y}{V}\right)+\sum b_n\sin n\,\omega\left(t\pm\frac{y}{V}\right)$$

gde su Furije koeficijenti a_0, $a_n, i\, b_n$ dati sličnim formulama kao i kod oscilacija. Odatle sledi da se svako talasno kretanje može prikazati kao zbir sinusnih(harmonijskih) talasa.

Project: 06SER02/02/003

Chemistry models in Environment Protection

Dr Ivana Ivančev Tumbas

1. Environmental chemistry and modeling: what do we need and why?

1.1. Why do we need modeling in Environmental Chemistry?

Environmental chemistry is used to solve environmental problems. Typical examples are:

- diffusion of accidental pollution in air, water or soil which threats to the ecosystem or human health and which requires efficient and prompt reaction of authorities
- clean technology choice for the company which has to deal with its waste in most efficient way (e.g. to fulfill legislation regarding emission limit values with best available technologies and to effectively operate and control it)
- management in the field of wastewater, sewers, water supply or control of the polluters in certain area related to emission to air or river basin which is required by authorities

To solve these and similar problems one has to have the knowledge in ecology, environmental science and technology. It is not a rare case that different profiles of experts are engaged to solve one environmental problem due to its complexity. Very often teams consist of chemists, civil engineers, mechanical engineers, hydrogeologists, biologists, physicians, doctors, law experts, geologists, meteorologists, etc.

The scope of environmental chemistry consists of chemistry of biological treatment processes, chemistry of physico-chemical treatment methods, chemistry of the pathways and effects of both biochemical components and toxic substances in the environment including pathways and effects of toxic substances in organisms.

Chemistry and biology are most widely applied in environmental sciences and they provide us knowledge necessary to do the environmental management. They teach us how different chemicals influence biological systems and, based on the knowledge, we are able to predict those reactions as well as to use them for the benefit of our society. On the other side, models are synthesis of the knowledge on the system and beside the fact that they are very effective as a tool in environmental management, they provide us with knew knowledge about the system which help us to understand it and to do the better prediction of its behavior. Thus, using the models, sometimes we can find new relationships between variables and further use this new knowledge to improve the efficiency of the cleaning technologies or simply to do the better management of our environment.

By doing environmental management we:

- Protect ecosystems
- Protect human beings and resources

- Achieve sustainable development
- Reach optimal results with our technology related to its influence on environment

Doing environmental management by using models we are more successful. Since the best protection can be done by prevention, one can conclude that the effective prediction of the system behavior in certain circumstances can be of utmost importance to take preventive actions (e.g. disconnection of endangered wells from water supply system if there is a threat of spreading of pollution) or to simply manage the system with the highest efficiency (e.g. usage of membranes for water purification with optimal influent quality in order to prolong their usage and by doing that save maintenance costs for the company).

Beside the fact that we want to know how environment functions and to predict how it reacts on different kind of pressures caused by activities of humankind, we also use environmental chemistry to manage the processes in different treatment techniques used to prevent environmental pollution (e.g. pollution of water, air and soil). All physico-chemical reactions related to processes used for treatment of waste gases, water and soil can be modeled with more or less efficiency by mathematical equations and the models are improving all the time. The impact of man and his technology on environment must be quantified where there is a possibility of relating the application of environmental technology to corresponding effects in the ecosystem. Furthermore, the goal of applied ecotechnologies is enabling ecosystems to cope with pollutions.

Pollutants enter environment by three different ways: wastewater, air pollution and solid waste. Pollution prevention and control is possible by using best available technologies which minimize emission of pollutants, save resources and energy.

Some of them related to prevention of air pollution are related to:

- particle and toxic gases removal by usage of filters where classification is based on particle size,
- wash towers where separation is based on particle density and
- cyclones where centrifugal force is used to separate particles from air.

Beside particle removal, chemical reactions occur to remove sulphur dioxide, NOx, or adsorption takes place in removal of volatile organic compounds.

In wastewater treatment, different kind of physico-chemical treatments are applied: sedimentation, flotation, filtration, neutralization, precipitation, flocculation, ion-exchange and adsorption, electrochemical methods, thermal methods, etc. Each of those processes is well described by mathematical equations and possible to model. One of the most important treatments in wastewater treatment is conventional activated sludge process. Example of biological treatment processes model used in waste water treatment plant is a model of man-made ecological system.

Related to solid waste, we have the possibility of either dispose it to the landfill or to incinerate it. When we are talking about disposal it is of utmost importance to have sanitary landfill by excluding the possibility to further contaminate environment. Therefore it might be interesting and useful to know and

model groundwater risk and mobility of pollutants, to know about decomposition processes in landfills and to manage removal and use of landfill gas.

Treatment of polluted soil can be done *in situ* or *ex situ* by using chemicals, but also bioremediation process can be applied if biodegradable pollutants has to be removed. Biodegradation is chemical transformation of molecules by microorganisms (mostly bacteria) via different intermediers to natural products (e.g.CO_2, H_2O and inorganic salts). It occurs in stepwise fashion and is usually not the result of the activity of a single specific organism. Usually, several strains of microorganisms, often existing synergistically, are involved. So, we have to deal with different metabolic pathways and enzyme systems.

Relating to thermal decomposition of waste, very important is to minimize the formation of toxic products like dioxins or furans by control of the combustion process. One of very important factors is a composition of the waste which relates to the quality of incineration process and it should be managed in most efficient way: to get as much as possible energy with the lowest possible pollution.

1.2. What do we need to model?

1.2.1. Cycles of elements

Our environment is divided into spheres. Biosphere is all the layers of the Earth inhabited by living organisms. It consists of ecosystems. The atmosphere is the mass of air surrounding the Earth with the fluid upper limit bordering on outer space. Ecosphere includes only inhabited space. Meteorologically the atmosphere is divided into troposphere and tropopause, stratosphere and stratopause, mesosphere and mesopause and thermosphere. They differ in physical properties.

The highest region of the soil inhabited by organisms is pedosphere. It borders on the lithosphere, the outermost rock layer down to the depth of 100 km. The pedosphere is penetrated by atmosphere and hydrosphere.

All geological processes on Earth are described as a cycle of materials. Humans interfere with natural cycles in particular via the anthropogenic exploitation of natural resources (mining, water use) and by waste emission. The waste than enter the cycles.

There are four main cycles of nutrient elements: carbon cycle, the nitrogen cycle, the sulphur cycle, the phosphorous cycle. They all shown transformations of four nutrient elements in the nature between their several forms.

Thus, in carbon cycle, carbon takes various forms which are in relations: carbon dioxide dissolves in water giving HCO_3^- which can be either precipitated in forms of salts (Ca and Mg precipitates) or used in process of photosynthesis. Thus carbon enters in organic forms related to biogeochemical processes and further can be used (e.g. as fossil fuel) in production of xenobiotics. Organic carbon may be biodegraded yielding CO_2 again. Inorganically bounded carbon in precipitates can be dissolved again depending on the conditions in water matrix.

Regarding nitrogen, it is important to know that it can be present in the nature in various forms: organic (e.g. NH_2 groups of proteins) and inorganic (ammonia, nitrites, nitrates, various oxides). Fixation of atmospheric nitrogen by microorganisms and further microbial decay of the organic matter that contain nitrogen in protein form yield ammonia which is by *Nitrosomonas spp.* and *Nitrobacter spp.* Further it is

transformed into nitrate and nitrite form. Those forms are by process of denitrification further transformed again into molecular nitrogen.

Beside, we can talk about metal cycles and special cycles of environmental chemicals that enter the ecosphere as a result of human activity. Fore example, very important anthropogenic and toxic chemicals are heavy metals. They can be present in the environment in forms of ions or complexes which both interact with particulate matter in water. Thus metals can be bounded to sediment or adsorbed in soils in different forms.

Organic pollutants, depending on their solubility, might be attached on particles of sediment or might be dissolved in water traveling long distance. Their fate might include biodegradation, photodecomposition, oxidation, chemical speciation etc. Depending on solubility and volatility contaminants are spread within the air, water and soil.

1.2.2. The most common types of pollution

Some of very important environmental chemicals are: pesticides, industrial chemicals (e.g. solvents, reagents), polychlorinated biphenyles and dioxins, aromatic and polyaromatic hydrocarbons, organochlorine compounds, detergents, etc.

Most important agricultural pollution is caused by usage of pesticides. We use herbicides most frequently (40%), insecticides (30%) and fungicides (20%). The other groups of pesticides are not so frequently used (rodenticides, growth regulators, etc). Besides pesticides, significant agricultural pollution is application of fertilizers (phosphates, K, nitrates). Beside the fact that these compounds enter the food chain via plants, the most important influence on the environment is water pollution since they are very often transported through soil layers into the shallow aquifers of groundwater or surface water. There they can seriously influence water quality of the water supply source, or accumulate in aquatic organisms.

Most important pollution from the energy production is air pollution caused by emission of sulphur oxides, nitrogen oxides, carbon monoxide, hydrocarbons and particles.

Last, but not least important, is pollution released from households: detergents, whitening agents, personal care products, medicines, colors, solvents, glue etc.

The behavior of substances is determined by molecular and physico-chemical characteristics: molecular mass, functional groups, volatility, solubility, Nernst's coefficient of distribution, adsorption behavior, etc. Their behavior in environment is influenced by environmental factors as well: humidity, soil type and characteristics (e.g. humic content or particle size), temperature, pH, redox potential, presence of microorganisms, etc. Effects of toxicants in environment can be localized in space and time, but the level of pollution depends on toxicant and ecosystem characteristics. For example, rainy weather and soil treatment can enhance pollution transport through the soil or running water has mixing influence in different directions on distribution of pollution and particles in water body. Mixing causes dilution. Also, photolytic or hydrolytic degradation or precipitation can happen. Higher humidity of the atmosphere prevents volatilization of chemicals from the soil or water surface and enhances penetration into dry soil.

Natural waters are endangered with lot of inorganic, organic and biological pollutants. Large number of them is a consequence of unacceptable waste disposal. Some of pollutants are highly toxic (e.g. cadmium) while sometimes pollutants are not toxic but they cause other fatal changes (oxygen depletion due to high content of biodegradable organic matter). Some constituents are normal in small concentrations, but very toxic in high concentrations (NaCl). Various chemical processes are happening in water bodies: redox processes, chelation, photosynthesis, precipitation, acid.base reactions, microbial reactions, gas exchange, leaching and uptake from sediment.

1.2.3. Model types

Model types used in environmental chemistry are:

- biogeochemical models,
- ecotoxicological and toxicological models,
- chemical speciation models,
- biological treatment processes models and
- physical-chemical treatment processes models

Biogeochemical models focus on the processes and transformations of various compounds in ecosystem present both as pollutants or naturally occurring substances. Ecosystems modeled by use of biogeochemical models are rivers, lakes, reservoirs, ponds, estuaries, coastal zones, open sea, wetlands, grassland, desert, forests, agriculture land etc. Conceptual diagram of simple biogeochemical model can be described as in Fig. 1.1.

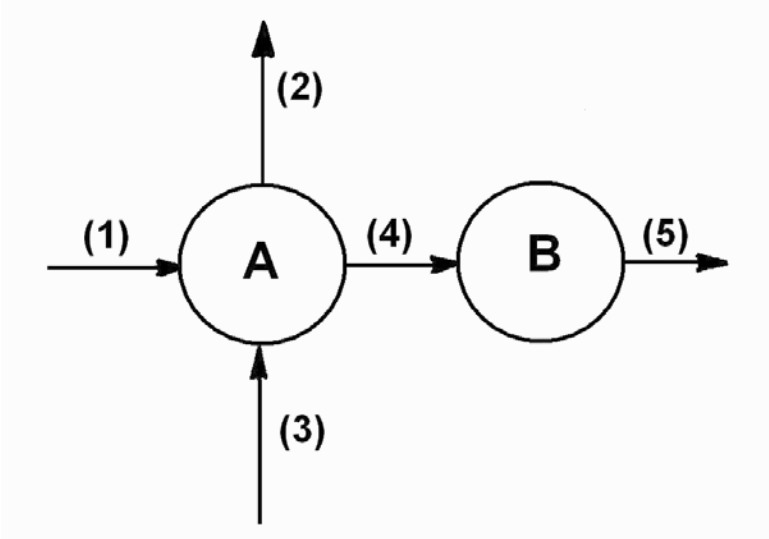


Figure 1.1. Example of conceptual diagram

A and B are state variables (e.g. concentration units or population density). Arrows indicate processes. Accumulation can be described as difference between

input and output. In dynamic models one describes change of variables A and B with time:

$$\frac{dA}{dt} = process(1) - process(2) + process(3) - process(4) \quad (1.1)$$

$$\frac{dB}{dt} = process(4) - proces(5) \quad (1.2)$$

It is crucial to have good formulation of processes which can be described by mathematical equations.

Contrary to mentioned dynamic model, static model is constructed by setting

$$\frac{dA}{dt} = \frac{dB}{dt} = 0 \quad (1.3)$$

These equations can be used to find A and B in the steady state situation. In general, dynamic models require more comprehensive database.

Models of ecological processes are used to form model of ecosystems. Both of them can be described with more or less details. Biogeochemical models deal with biochemical and geochemical compounds and elements in ecosystems. When they are used for the control of pollution, they must include fate and distribution of both pollutants and natural compounds.

Typical chemical processes in biogeochemical models are oxidation, hydrolysis, photolysis, reduction, acid-base reactions. Biological processes are growth, production, mortality, immigration, emigration. Microbiological processes are growth of microorganisms, nitrification, reduction of sulfate, microbiological oxidation etc. Typical physical processes are transport of compounds between air and water, advection, diffusion etc. In general, the more complex the formulation, more parameters are included and better data base is needed. Equations which can be used for submodel development are described in literature.

Ecotoxicological models

Ecotoxicological models deal with fate and effects of toxic substance in ecosystem and organisms. Thus they can describe fate and transport of toxic substance in ecosystem, but also population dynamics which include effects of toxic substance in ecosystem (e.g. mortality or abundance of any kind of toxic effects). Some examples of ecotoxicological models are: food chain models, static models of the mass flows of toxic substances, a dynamic model of a toxic substance in a trophic level, ecotoxicological models with effect components and ecotoxicological models in population dynamics. An example of such a model is a distribution of pollutants such as PCBs in water, sediment, fish and other organisms. When modeled, this distribution has to take in account every input and output. For example PCBs can come into fish by water (very low fraction due to negligible solubility), but also by

suspended particles where PCBs can be adsorbed, or through food chain from other organism. Excretion should also be taken into account.

Chemical speciation models

Chemical speciation models deal with concentrations of the various forms of different components in a given ecosystem. They are often based on differential equations presenting reactants or product concentration changes in time. Those models can be included in ecotoxicological and biogeochemical models.

Chemical speciation is most often described as a steady state allocation of a chemical component in various forms. Model consists of several equilibrium equations (redox processes, complex formation, hydrolysis, photolysis, acid-base reactions, mass transfer from one phase to the other and adsorption).

Chemical reaction is in equilibrium when the amounts of reactants and products stop changing with the time. At the equilibrium the rates of forward and reverse reactions are equal and for the general reaction:

$$aA + bB \leftrightarrow cC + dD$$

We use equilibrium expression where K is equilibrium constant and it is calculated as follows:

$$K=[C]^c[D]^d/[A]^a[B]^b \qquad (1.4)$$

It depends only on temperature. According to the LeChatelier's Principle, for a given change in conditions, an equilibrium shifts into direction that opposes the change.

Complex formation

Complex formation is very important issue related to environmental chemistry. Ability of metal ions to form complexes influences their toxicity since structure and mobility of the metal is changed within the complex. Complexation increase solubility, alter distribution between reduced and oxidized forms, influence toxicity, change properties related to adsorption or ion exchange and change the stability of the colloids which contain metal. Metals which can form complexes are: copper, silver, mercury, lithium, aluminium, iron, nickel, manganese, etc. There are numerous ligands present in environment: natural humic and fulvic compounds, synthetic chemicals like EDTA from anthropogenic pollution, anions like carbonate, silicate, nitrate, o-phosphate, iodide, acetate, amino-acids, etc. Depending on ligand and metal concentrations and stability constant (K) of complex formation at given pH, redox and temperature conditions, complex formation can be predicted from the following equilibrium:

$$Me^{n+} + L^{m-} \longrightarrow MeL^{(n-m)+} \qquad (I\text{-}1)$$

$$\frac{[MeL^{(n-m)+}]}{[Me^{n+}][L^{m-}]} = K \qquad (1.5)$$

Hydrolysis processes influence metal mobility since they might increase solubility of metal ions. It covers processes which proceed with water, hydroxide and hydrogen ions. Here are examples of iron and aluminium:

$$Al(OH)_3 + OH^- \rightarrow AL(OH)_4^- \qquad (I\text{-}2)$$

$$Al(OH)_3 + 3H^+ + nH_2O \rightarrow Al(H_2O)_{n+3}^{3+} \qquad (I\text{-}3)$$

$$Fe(OH)_3(H_2O)_3(s) + 3H^+ = Fe(OH)_4(H_2O)_2^- + 4H \qquad (I\text{-}4)$$

Decreasing pH value toxic effects of metal increase due to formation of metal aqua ions.

Hydrolysis

Hydrolysis of organic compounds is interesting for environmental chemistry because of different toxic effects of different organic compounds. For some chemicals hydrolysis rates are independent of natural pH values (4-9), while for the others they depend on pH.

For all those reactions we have rate constants which can be expressed by change in product or reactant concentration in time. Relatively simple test can be used to determine kinetic rates from which one can estimate concentration of reactant or product under given condition. When intervals are long enough (half a day or more) equilibrium parameters are used for modeling while for short intervals of few hours or minutes, kinetic parameters are used for modeling.

The rate law is a mathematical function, specifically a differential equation, describing the turnover rate of the compound of interest as a function of the concentrations of the various species participating in the reaction. In general way, we can write the macroscopic rate law for the transformation rate:

$$D[org]/dt = k[i]^i[B]^b[C]^c... \qquad (1.6)$$

Where the exponents i,b,c,... indicate the order of the reaction with respect to the corresponding species: org, B, C,…This empirical law does not reveal the mechanism of the reaction considered. Even simple reaction may proceed by several distinct reaction steps (elementary molecular changes) in which chemical bonds are broken and new bonds are formed to convert the compound into the observed product. Each of these steps, including back reactions, may be important in determining the overall reaction rate. Therefore, the reaction rate constant, k, may be a composite of reaction rate constants of several elementary reaction steps.

The simplest example is transformation rate of compound A to compound B. It is expressed mathematically by a first-order rate law:

$$d[A]/dt = -k[A] \qquad (1.7)$$

where k is refered to as the first-order rate constant. If the water molecules were involved in the slowest step of the reaction, the rate-determining step, the reaction should be described by a second-order rate law:

$$d[A]/dt=-k'[A][H_2O] \quad (1.8)$$

where k' is second-order rate constant. Since water is usually present in large excess and change of its concentration is negligible, by setting $k=k'[H_2O]$, we again get first order constant, but now, since we know that the mechanism involves more species on molecular level we call it pseudo-first order rate constant.

Although the most reactions with which we are concerned are not truly first order, it is common for modeling purposes to make assumptions that allow us to reduce the order of the reaction law, ideally to pseudo-first order.

Redox processes

Redox processes are of great importance in environmental chemistry since lot of metals can exchange electrons with different constituents in environment. Thus arsenic might be very strongly bounded to ferromanganese nodules under aerobic conditions, while under anaerobic conditions toxic arsenic will release due to reduction metal ions. The another example is FeS_2 exposure to the air by reduced water in mines where following processes occur and damage environment with huge amounts of sulfuric acid formed:

$$2FeS_2 + 2H_2O + 7O_2 = 2FeSO_4 + 2H_2SO_4$$

$$4FeSO_4 + O_2 + 2H_2SO_4 = 2Fe_2(SO_4)_3 + 2H_2O \quad (I\text{-}5)$$

$$Fe_2(SO_4)_3 + 6H_2O = 2Fe(OH)_3(s) + 3H_2SO_4$$

Reaction of oxidation is very important in aquatic systems where several oxidants can be present: radicals, singlet oxygen, peroxides, ozone, etc. They can have half life very short (few milliseconds) but also to be more stable.

Rate of oxidation reaction is given:

$$v = k_{ox} \cdot [C][ox] \quad (1.9)$$

Where k_{ox} is specific second order constant for oxidation at specific temperature and [C] and [ox] are molar concentrations of the chemical compounds and the oxidant respectively. The total rate of oxidation if most oxidants work simultaneously is the sum of the rates for each reaction of each kind of oxidant:

$$R_{ox}= (k_{Ox1}[Ox_1]+k_{Ox2}[Ox_2]\ldots\ldots)\cdot[C]=\left(\sum_{n=1}^{n=n} k_{Ox_n} \cdot [Ox_n]\right)\cdot[C] \quad (1.10)$$

Integration between the time limits 0 and t gives

$$\ln[C_0]/[C_t] = \sum_{n=1}^{n=n} k_{Ox_n} \cdot [Ox_n] \cdot t \qquad (1.11)$$

or if the half time, $t_{1/2}$ is used

$$t_{1/2} = \ln 2 / (\sum k_{Ox_n} \cdot [Ox_n]) \qquad (1.12)$$

Most redox processes are very fast relative to the time steps applied in most environmental models. Good example is kinetic of oxidation of iron (II) and manganese (II). Both oxidation processes take minutes. When intervals of half a day or one day are used in models it is clear that equilibrium description will be enough, but when intervals of few hours or even minutes are used, a description of oxygenation kinetics will be required. When the amount of oxygen is limited, the transfer of oxygen may be the rate-determining process. If it is so, the oxygen is consumed almost instantly and we have to model transfer of oxygen as accurately as possible and to take into account possible anaerobic decomposition. Oxidation of iron and manganese will proceed by a rate determined by the transfer of oxygen.

Acid-base reactions

Almost all the processes in nature are pH dependant. Thus content of ammonia and carbon dioxide in water is pH dependant, biological processes have their optimum pH value (usually 6-8), mobility of heavy metals is pH dependent. Furthermore efficiency of chlorine used as disinfecting agent is also dependant on pH (Figure 1.2) since different forms of hypochlorous acid have different disinfecting ability.

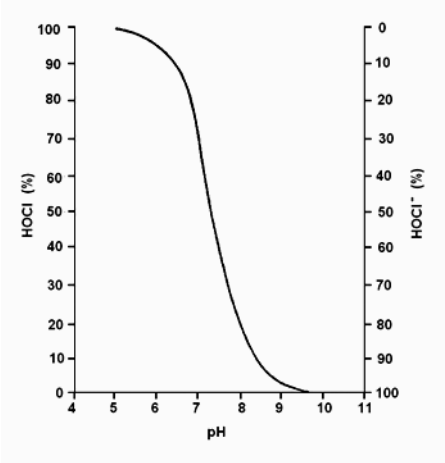


Figure 1.2. Influence of pH on speciation of HOCl in water.

When the composition of aquatic system is known it is possible to calculate both alkalinity and the buffering capacity and find out which species will be present in solution.

Furthermore, processes of adsorption, ion exchange and photochemical reactions also have great importance in environmental modeling. Adsorption is in detail explained in chapter 4.

Literature:

- S. E.Jørgensen (1991) Modelling in Environmental Chemistry, Developments in Environmental Modelling, 17, Elsevier, Amsterdam-London-New York-tokyo
- S.E. Manahan (1992) Toxicological Chemistry, Second Edition, Lewis Publishers, Chelsea Michigan , USA

2. Eyring equation-example of basic chemistry model

Example of basics on chemistry modeling that will be presented here is Eyring rate equation. The bimolecular reaction is considered by “transition state theory”:

$$A + B \xrightarrow{k} C \qquad v = \frac{dC}{dt} = k[A][B] \tag{2.1}$$

The reactants react by forming unsteady intermediate on the reaction pathway:

$$A + B \xleftrightarrow{k_a} AB \xrightarrow{k_2} C \tag{2.2}$$

There is an “energy barrier” on the pathway between the reactants (A, B) and the product (C). The barrier determines a minimum of energy necessary for the reaction to occur. It is called activation energy.

While approaching each other, reactants possess kinetic energy and their potential energy is constant. The approaching reactant molecules had sufficient kinetic energy to overcome the mutual repulsive forces between the electron clouds of their constituent atoms and thus come very close to each other. An 'activated complex' AB or 'transition state' is formed at the potential energy maximum. The high-energy complex is unstable and it breaks and generates the product C or degrades back to the reactants A and B. Principles of the transition state theory are: there is a thermodynamic equilibrium between the transition state and the state of reactants at the top of the energy barrier and the rate of chemical reaction is proportional to the concentration of the reactants.

The change in the concentration of the complex AB over time can be described by the following equation:

$$\frac{d[AB]}{dt} = k_1[A][B] - k_{-1}[AB] - k_2[AB] \tag{2. 3}$$

Due to the equilibrium between the 'activated complex' AB and the reactants A and B, the components $k_1 \cdot [A] \cdot [B]$ and $k_{-1} \cdot [AB]$ cancel out. Thus the rate of the direct reaction is proportional to the concentration of AB :

$$\frac{dC}{dt} = -\frac{d[AB]}{dt} = k_2[AB] \tag{2.4}$$

k_2 is given by statistical mechanics:

$$k_2 = \frac{k_B T}{h} \tag{2.5}$$

k_B = Boltzmann's constant [$1.381 \cdot 10^{-23}$ J · K^{-1}]
T = absolute temperature in degrees Kelvin (K)
h = Plank constant [$6.626 \cdot 10^{-34}$ J · s]
k_2 is called 'universal constant for a transition state' (~ $6 \cdot 10^{-12}$ sec^{-1} at room temperature).

Additionally, [AB] can be derived from the quasi stationary equilibrium between AB and A, B by applying the mass action law:

$$[AB] = K^*[A][B] \tag{2.6}$$

K* = thermodynamic equilibrium constant

Due to the equilibrium that will be reached rapidly, the reactants and the activated complex decrease at the same rate. Therefore, considering both equation (2.5) and (2.6), equation (2.4) becomes:

$$-\frac{d[AB]}{dt} = \frac{k_B T}{h} K^*[A][B] \tag{2.7}$$

Comparing the derived rate law (2.1) and the expression (2.7) yields for the rate constant of the overall reaction

$$k = \frac{k_B T}{h} K^* \tag{2.8}$$

Additionally, thermodynamics gives a further description of the equilibrium constant:

$$\Delta G = -RT \ln K^* \tag{2.9}$$

Furthermore $\Delta G^{\ddagger}$ is given by

$$\Delta G = \Delta H - T\Delta S \tag{2.10}$$

R = Universal Gas Constant = 8.3145 J/mol K
ΔG = free activation enthalpy [kJ · mol^{-1}]
ΔS = activation entropy [J · mol^{-1} · K^{-1}]
ΔH = activation enthalpy [kJ · mol^{-1}]

ΔG is the free activation enthalpy (Gibb's free energy) and it represents the determining driving power for a reaction. The sign of ΔG determines if a reaction is spontaneous or not. If it is lower than zero reaction is spontaneous, if it is equal to zero, reaction has achieved equilibrium and if it is higher than zero reaction is not

spontaneous. Combining Equation (2.9) and the expression (2.10) and solving for lnk one can get:

$$\ln K^{*} = -\frac{\Delta H}{RT} + \frac{\Delta S}{R} \tag{2.11}$$

The *Eyring equation*: is found by substituting equation (2. 11) into equation (2. 8):

$$k = \frac{k_B T}{h} e^{-\frac{\Delta H}{RT}} e^{\frac{\Delta S}{R}} \tag{2.12}$$

$$\ln k = \ln \frac{k_B}{h} T - \frac{\Delta H}{R} \frac{1}{T} + \frac{\Delta S}{R} \tag{2.13}$$

$$\ln \frac{k}{T} = -\frac{\Delta H}{R} \frac{1}{T} + \ln \frac{k_B}{h} + \frac{\Delta S}{R} \tag{2.14}$$

A plot of ln(k/T) versus 1/T produces a straight line with the familiar form

y = -mx + b, where

x = 1/T
y = ln(k/T)
m = - $\Delta H^{\ddagger}$/ R
b = y (x = 0)

ΔH can be calculated from the slope m of this line: ΔH = -m · R .

From the y-intercept $\Delta S^{\ddagger}$ can be determined and thus the calculation of $\Delta G^{\ddagger}$ for the appropriate reaction temperatures according to equation (10) is allowed. The activation energy E_a is related to the activation enthalpy $\Delta H^{\ddagger}$ as follows

$$y(x = 0) = \ln \frac{k_B}{h} + \frac{\Delta S}{R} \tag{2.15}$$

$$E_a = \Delta H + RT \tag{2.16}$$

Low values of E_a and ΔH mean fast rate reactions and high values of Ea and ΔH mean slow rate. The typical values of E_a and ΔH lie between 20 and 150 [kJ / mol].

Example

Kuempel and co-workers (Inorg. Chem., 12, 1036 (1973)) obtained the following pseudo first order rate data for the hydrolysis reaction:

$$Ru(NH_3)_5Br^+ + H_2O \xrightarrow{k} Ru(NH_3)_5H_2O^{2+} + Br^-$$

k (sec^{-1})	T (ºC)
1.2	15
3.8	20
5.4	25
8.3	30
12.2	35

Calculate the free energy of activation for the reaction.

Literature

- S. E.Jørgensen (1991) Modelling in Environmental Chemistry, Developments in Environmental Modelling, 17, Elsevier, Amsterdam-London-New York-Tokyo
- Varma, A., Morbidelli, M. (1997) Mathematical Methods in Chemical Engineering, Oxford University Press, New York, Oxford.

3. Modeling tools: transport and reactions

There are two types of transport in natural systems:

- Transport by random motion (e.g. molecular diffusion, dispersion) and
- Transport by directed motion (e.g. advection in water currents, settling suspended particles due to gravitation)

Transport processes caused by random motion are diffusive, while those resulting from directed motions are advective.

3.1. Random motion

Random motion is ubiquitous. At the molecular level the thermal motions of atoms and molecules are random. Randomness means that the movement of an individual portion of the medium (i.e. molecule, a water parcel, etc) can not be described deterministically. Basic description of transport by random motion is Fick's law (gradient-flux law). Flux is defined as a quantity of physical unit (e.g. concentration) which is transported per unit area (perpendicular to flow direction) and per time. It is assumed that the subsystems A and B and the distance between them, $\Delta x_{A/B}$, become infinitely small. The difference in concentration tends toward zero. The ratio of the two differences, Δ concentration: $\Delta x_{A/B,}$ is equal to the spatial gradient of the concentration and usually different from zero:

$$F_x = -b \frac{d}{dx}(concentration) \tag{3.1}$$

Where b is a constant

The minus sign indicates that the flux points are against the gradient. Instead the subscript A/B, we use the subscript x to design the coordinate axes along which the flux occurs. It is assumed that the flux is determined by the variation of a local property. This variation (such as the gradient of the concentration, temperature, preasure, etc) is "driving force" for transport. Mathematically, the gradient is a local property of a function. There are lots of physical processes obeying the gradient flux law: molecular diffusion, conduction of heat, flow through porous medium, etc.

Second Fick's low states that the local concentration range with time due to diffusive transport process is proportional to the second spatial derivative of the concentration:

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} \tag{3.2}$$

3.2. Boundaries in the environment

Many important processes in the environment occur at boundaries. Boundaries are surfaces at which properties of a system change (e.g. air-water interface, sediment-water interface). They are characterized by physical and chemical processes and very often it is enough to use equilibrium concept to describe the boundary (e.g. Henry's low). There are three different types of boundaries (according to the shape of the generalized diffusivity profile across the boundary):

- bottleneck boundary with sharp change in diffusivity since the transfer coefficient is different in one single zone than in the bulk (water surface of the river, between two turbulent zones)
- wall boundary (e.g. sediment-water interface in lakes or oceans) where we have boundary between two different compartments where only one (water) is turbulent.
- diffusive boundary-dispersion at the edge of a pollutant front. Here there is no big change in diffusivity.

Different mathematical equations are used to explain those boundaries in modeling of fate and transport of pollutants in ecosystems.

3.3. Box models

The simplest and often most suitable modeling tool is the one-box model. One box models describe the system as a single spatially homogenous entity. Homogenous means that no further spatial variation is considered. Those models can have one or several state variables, for instance, the mean concentration of one or several compounds which are influenced both by external "forces" and by internal processes.

In a system which has constant volume, the mass balance of compound i is described by equation:

$$\frac{dC_i}{dt} = \frac{1}{V}\left(I_i - Q_i - \Sigma R_i + \Sigma P_i\right) \tag{3.3}$$

Where ΣR_i and ΣP_i are internal consumption and production rates of processes related to chemical *i* respectively. The concentrations are influenced by inputs I_i and by outputs Q_i.

In **linear one box model with one variable** we use linear function to describe external forces. For example, external input is not dependent on C_i.

$$f_p(C_i) = a_p(t) + b_p(t)C_i \tag{3.4}$$

This model can be used for example to calculate total mass and mean concentration of pollutant in the lake from repeated measurements of pollutant concentrations. Calculation includes apriori knowledge related the spatial distribution of pollutant, horizontal concentration gradients in lake must be so small that total

mass can be calculated as a volume-weighted average of the concentrations measured along a vertical profile at the deepest location of the lake. Furthermore, input and output must be known. If the pollutant is volatile it appears that the only significant removal mechanism (other than loss at the outlet) is air-water exchange. In situ reactions and sorption on sediments are supposed to be not relevant.

Example of **one box model with two variables** is the case with two chemicals, A and B, where A is transformed into B by a chemical process and vice versa. The system is described by two concentrations C_A and $C_{B,}$ by two zero order input functions, by two first order output functions, $k_A C_A$ and $k_B C_{B,}$ and by the first order transformations from A to B and vice versa. If there are no transformation between A and B, we could describe each chemical separately by linear one box model.

Two-box models are useful for describing systems consisting of two spatial subsystems which are connected by one or several transport processes. Furthermore, we have linear multidimensional models.

Finally, model variables (i.e. concentration) depend on time and on space. During the last several decades, the field of three dimensional space-time modeling has undergone a rapid development. The main mechanisms are transformations, directed transport (advection, flow, settling of particles) and random transport (diffusion and dispersion).

Literature

- Schwarcenbah R. P., Gschwend P. M., Imboden D.M. (2003) Environmental organic chemistry, Second Edition, Wiley-Interscience, John Wiley&Sons, Hoboken, New Jersey

4. Adsorption modeling

Adsorption is attachment of molecules from a liquid or gaseous solution at another phase (two-dimensional surface). It is very important in various aspects of environmental chemistry since it is common process, both in nature and in different treatment techniques of waste streams. It influences fate and transport of variety of pollutants and thus it is of great importance in the risk assessment.

Compounds can be adsorbed both from water and from the air onto the soil, clays, various surfaces, onto natural organic matter, etc. Use of granular activated carbon is very common: for the treatment of the waste gases and removal of organic compounds from them (e.g. volatile organic compounds in fuel gasses), for removal of organic compounds from drinking water, etc. For example, understanding of adsorption process in water technology consists of knowledge related to special characteristics of activated carbon, knowledge about adsorption equilibrium and kinetics, and about process design. Related to the process design there are possibilities to apply activated carbon as powdered activated carbon or in fixed bed columns filled with granular activated carbon. Furthermore, several other modern processes are developing (e.g. hybrid processes where powdered activated carbon and membrane filtration are combined, or combined adsorbers with permeable synthetic collectors). Modeling of those processes enable us to predict behavior of pollutants and to manage the treatment plant more successfully.

4.1. Adsorption equilibrium

From adsorption equilibrium data one can calculate adsorption capacity of the activated carbon. It is determined by an adsorption isotherm which describes equilibrium in closed system, which consists of solution of the substance(s) which we want to remove and amount of carbon brought into the contact with the solution.

For the evaluation of an adsorption isotherm, defined quantities of activated carbon are added to several bottles containing the same, predefined volume of the solution with single substance initial concentration. Solutions should be shaked until adsorption equilibrium is reached and than equilibrium concentration in each solution is determined for each carbon dose applied. Time needed to reach equilibrium can vary from hours to days and even years. From these data adsorption parameters can be calculated (eq. 4.1). The extent to which the full surface area of an activated carbon can be used for adsorption depends on the concentration of solute in the solution with which the carbon is mixed. The specific relationship which defines adsorption isotherm is constant for constant temperature. Adsorption equilibrium is most frequently described by empirical Freundlich equation:

$$q=K_F c^n \qquad (4.1)$$

where

q – solid phase concentration of the solute which describes the quantity of the adsorbed substance
K_F-Freundlich constant

c-equilibrium concentration in the solution
n-Freundlich exponent

Freundlich exponent and Freundlich constant can be easily determined by nonlinear regression transforming the equation into:

$$\lg q = \lg K_F + n \lg c \qquad (4.2)$$

The slope of log c versus log q is equal to the exponent, n, and the solid phase concentration at $C_{iw}=1$, equals Freundlich constant K_F. Adsorption capacity of carbon increases when K_F value increases and when n decreases. Value of n represents change in free energy of sorption of solute on sorbent in certain concentration range. When *n=1*, izotherm is linear and free energy of sorption is same for all concentrations, when n<1, isotherm is concave and with increase of sorbate concentration free energy of sorption decrease. When ni>1, isotherm is convex and with increase of sorbate concentration free energy of sorption increase.

Furthermore, Langmuir equation can be used for the explanation of the adsorption equilibrium:

$$q = q_m \cdot \frac{K_L \cdot c}{1 + K_L \cdot c} \qquad (4.3)$$

where,

K_L - Langmuir constant,
q_{m-} maximum solid phase concentration at monomolecular cover of the adsorbtion surface,
c - equilibrium concentration and
q - solid phase concentration at adsorption equilibrium.

The Langmuir equation assumes that the free energy, enthalpy, and enthropy change due to adsorption are constant for all solid phase concentrations. This is often referred to as homogenous surface. Frequently this condition is not fulfilled and the enthalpy of adsorption changes with increasing of solid phase concentration.

The two constants of Langmuir isotherm equation can be determined from isotherm data, either by applying a nonlinear regression method or, if the test results are plotted in a suitable manner by linear regression.

Multisolute adsorption

In the case when not only one substance is present in solution, we have multisolute system where substances are competing for adsorption sites. In this case solid phase concentrations of any substance will be reduced in comparison to solid-phase concentration in a single solute system. Well known model for the description of the adsorption in multisolute systems is based on the Ideal Adsorbed Solution Theory (IAST). It is based on assumption that adsorption equilibrium occurs between

two dimensional surfaces and the solution. The adsorption medium is taken as thermodynamically inert, the adsorption sites at the activated carbon surface are accessible in the same degree to all absorbable organic molecules in the solution and the adsorption equilibrium is reversible. For the calculation of the adsorption equilibrium in a mixture only adsorption data for single substances are needed. Freundlich equation is used and it is assumed that the Freundlich exponent is constant in the range $c_i=0$ to $c_i=c_{0.}$

$$c_i = \frac{q_i}{\sum_{j=1}^{N} q_j} \left(\frac{\sum_{j=1}^{N} \frac{q_j}{n_j}}{\frac{K_{F,i}}{n_i}} \right)^{1/n_i} \quad \text{N= number of components} \tag{4.4}$$

The IAST is often used as a basis for the description of the adsorption of an unknown mixture of adsorbable organic substances. Adsorption analysis with IAST also can be used to monitor removal of the substances with different adsorbability during water treatment process.

4.2. Adsorption kinetic

Kinetic of adsorption is the rate of reaching adsorption equilibrium by two diffusion steps: from the solution to the external surface of the adsorbent (external mass transfer or film diffusion) and than to diffuse into the pore system of the particles where they adsorb at the sites onto the inner surface (internal mass transfer). Mathematical description of the film diffusion is given by Fick's Law:

$$n_{L,i} = D_{L,i} \frac{dc_i}{d\delta} \tag{4.5}$$

Where

$n_{L,i}$ is the mass transfer rate per unit of surface area
$D_{L,i}$ is the aqueous phase diffusion coefficient of adsorbate i
δ is the location within the boundary layer of a thickness δ

By integrating the equation we obtain:

$$n_{L,i} = \beta_{L,i}(c_i - c_i^*) \tag{4.6}$$

where

$\beta_{L,i}$ is film diffusion coefficient, c_i^* is the equilibrium concentration of the component *i* at the outer surface of the activated carbon particle and c_i is the concentration of the component in the bulk solution. The determination of the surface diffusion coefficient can be accomplished by experiment or by use of well known empirical correlations.

Internal mass transfer may occur by the diffusion of the molecules in the liquid filled pores (pore diffusion) or by the diffusion of the adsorbed molecules onto the walls of the pores (surface diffusion). In the later case the driving force is the solid phase gradient and instead of aqueous phase diffusion coefficient we deal with surface phase diffusion coefficient and particle density:

$$n_{S,i} = \rho_p \cdot D_S \cdot \frac{\partial q_i}{\partial r} \tag{4.7}$$

The forces that govern the uptake kinetics of organic solutes from dilute solution by porous carbon are frequently quite different from those which control the ultimate capacity of the carbon for adsorption. The rate–limiting mechanism generally is either film diffusion or intraparticle transport, depending largely on the hydrodynamic character of the system in which the porous carbon is used. In contrast, the final position of adsorptive equilibrium is governed by the forces of adsorption, either chemical or physical in nature. As the result of possible differences in the nature of kinetic and equilibrium forces, factors that enhance rates of uptake may well decrease the capacity of porous carbon for certain adsorbates; the converse may also be true.

4.3. Process applications

Today we use mainly two types of water treatment processes based on activated carbon adsorption: powdered activated carbon (PAC) in sequential adsorption operation and fixed bed adsorber process with granular activated carbon (GAC).

Powdered activated carbon

Powdered activated carbon adsorption is frequently called contact filtration, as the typical application includes treatment in a mixing tank followed by filtration or settling. If the water to be treated contains C_0 as initial concentration of the substance in the volume V, where activated carbon is added in quantity m and after reaching equilibrium we measure equilibrium concentration C mass balance can be written:

$$VC_0+mq_0=VC+mq \tag{4.8}$$

Where q is solid phase concentration. Initial solid phase concentration is zero, so equation 4.8 can be rearranged as follows:

$$q = \frac{V}{m}(C_0 - C) \tag{4.9}$$

When Freudlich parameters are known for chemical it is possible to calculate the amount of activated carbon needed for certain removal efficiency based on equation 4.9. The calculation is based on calculation of q from equation 4.1, and afterwards required amount of carbon (m) based on equation of 4.9.

Granular activated carbon (GAC) adsorption

In continuous operation the water and the adsorbent are in contact throughout the entire process without periodic separation of the two phases (example of GAC fixed bed adsorbers).

The design of a fixed bed adsorber and the prediction of the length of the adsorption cycle require knowledge about saturation at the break point. Usually process is conducted in series of carbon columns whose contact times with water which is pumped through the column range from 15-60 min.

The point at which the impurity in a column effluent exceeds the treatment objective is called the breakpoint (Cb in Figure 4.1). The part of the curve between the initial leakage and the point where the column effluent concentration is the same as the influent is called the breakthrough curve. During the adsorption, the upper section of a column saturates with impurities, whereas the lower section remains virgin. Between this two extremes lies the adsorption zone, where removal of the impurity actually takes place. As the column becomes saturated, the adsorption zone moves downward through the bed. Typical breakthrough curve is presented in Figure 4.1

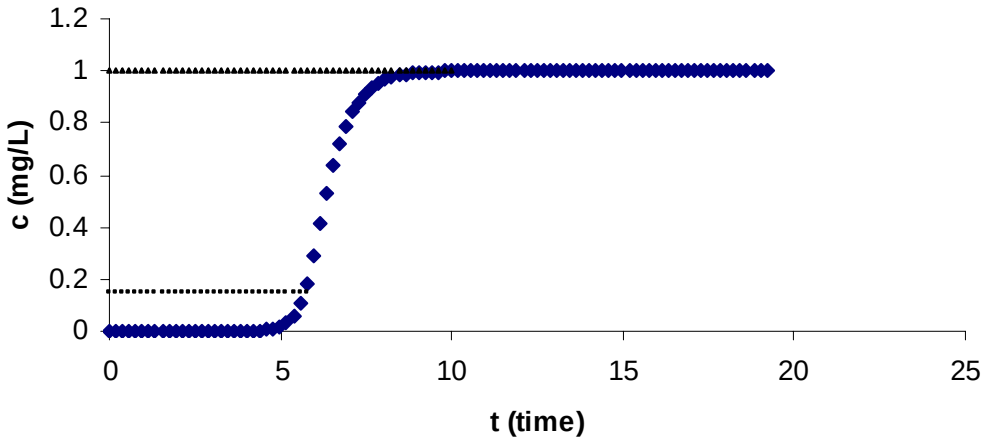


Figure 4.1. Breakthrough curve

The breakthrough curve should be steep and the solute concentration in the effluent rises rapidly from close to zero to that of incoming water. Some low arbitrary value of C_B is chosen as the break-point concentration and the column is considered exhausted when the effluent concentration has raised to some other arbitrarily chosen concentration of near value $C_{0.}$

The adsorption zone, the part of the adsorbent in which the concentration changes from C_B to near the C_0 has constant height of Z_A (m). The adsorbent behind the zone has been completely saturated with the solute. Within the zone degree of saturation with adsorbate varies from 0-100%. The height of the zone depends on the rate of adsorption and the solution flow rate. Smaller carbon particle size, larger diffusion coefficient of adsorbate and greater the strength of adsorption of adsorbate (e.g. larger freundlich K value) decrease the height of the zone.

For the situation where the breakthrough concentration is defined as the minimum detectable concentration, the critical height of the activated carbon column is equal to the height of the mass transfer zone. This height leads to immediate appearance of an effluent concentration equal to breakthrough concentration when the column is started up. The critical height, the flow rate and the area of the column are used for the calculation of the minimum tank volume occupied by the activated carbon divided by the volumetric flow rate (empty bed contact time):

$$\frac{Z_{critical}}{Q/A} = EBCT_{\min} \tag{4.10}$$

The EBCT has significant effect on the performance of the carbon. For a given situation, a critical depth of GAC and corresponding minimum EBCT exist that must be exceeded if the adsorber is to produce water of acceptable quality.

When breakthrough concentration is greater than minimum detectable concentration the critical height is less than height of mass transfer zone. EBCT is calculated as folows

$$\text{EBCT=V/Q=}\frac{L_{Bed}}{Q/A} \tag{4.11}$$

Where

V-bulk volume of carbon in contactor
Q- volumetric flow rate to contactor
L_{Bed}-bed depth
A-bed area

The mass of solute adsorbed per unit mass of adsorbent increases as percent exhaustion increases and correspondingly the number of bed volumes of water before breakthrough will also increase to a maximum value. Increasing EBCT, or bed depth impacts treatment costs. If adsorber is larger, fixed costs are increased. Operating costs decrease because of decreasing carbon usage rate and replacement frequency. Thus it is very important to work with optimum depth or contact time.

4.4. Homogenous surface diffusion model

Several useful mathematical models of the adsorption process are available. The homogenous surface diffusion model (HSDM) and modifications of it have been widely used to predict performance of adsorption systems. For single substance dispersed flow HSDM consists of the several elements: liquid phase storage, transport by advection, dispersion and adsorption:

$$\varepsilon \cdot \frac{\partial c(t,z)}{\partial t} + v_F \cdot \frac{\partial c(t,z)}{\partial z} - D_Z \cdot \varepsilon \frac{\partial^2 c(t,z)}{\partial z^2} + \frac{6 \cdot \beta_{L\cdot} \cdot (1-\varepsilon)}{d_P} \cdot \left(c(t,z) - c^*(t,z)\right) = 0 \quad (4.12)$$

where

ε- bed porosity
v_F-linear filter velocity
D_z- diffusion coefficient
β_L- film transfer coefficient
d_p - particle density

Linear Driving Force Model (LDF) (Kummel, 1990; Worch 1991) is based on the Dispersed-Flow Homogenous Surface Diffusion Model (HSDM). It is simpler because a linear gradient for homogenous surface diffusion in a GAC-grain is assumed. Model is based on above mentioned equation without consideration of dispersion. Both external film diffusion and intraparticle diffusion are considered within the model:

$$\frac{dq}{dt} = \frac{k_F a_V}{\rho_b}(c - c_S) \quad (4.13)$$

where

$k_F a_V$- volumetric mass transfer coefficient for film diffusion (mass transfer coefficient, k_F, multiplied with the area available for mass transfer normalized to the filter volume, a_V),

c_S-concentration at the external particle surface, ρ_b- bed density

$$\frac{dq}{dt} = k_S a_V (q_S - q\ ) \quad (4.14)$$

Where

$k_S a_V$- volumetric mass transfer coefficient for intraparticle diffusion (mass transfer coefficient, k_S, multiplied with with the area available for mass transfer normalized to the filter volume, a_V,

q is loading and q_S is loading at the external particle surface.

In a case of film diffusion, the concentration difference between the bulk phase and external surface of grain acts as a driving force. In case of intraparticle diffusion the driving force is difference between the loading at the external surface and mean loading of the particle. The mass transfer coefficient $k_S a_V$ can be related to the pore diffusion coefficient, $D_{P,}$ as well to the surface diffusion coefficient, D_S:

$$k_S a_V = \frac{15 D_S}{r_p^2} \tag{4.15}$$

$$k_S a_V = \frac{15 D_p}{r_p^2} \frac{c_0}{q_0 \rho_p} \tag{4.16}$$

where

r_p- particle radius,
c_0-initial concentration,
q_0-equilibrium loading related to $c_{0,}$
ρ_p- particle density.

In case of single solute adsorption concentration and loading at the external particle surface area are related by the Freundlich isotherm, while in case of mixture adsorption, an appropriate mixture adsorption model has to be used.

The program LDF (version 2.3) developed by Worch (2007) can be used to calculate adsorber breakthrough curves for single solute adsorbates or multisolute adsorbate mixtures using IAST model and LDF model. The input data required for breakthrough calculation are: concentrations, Freundlich coefficients and exponents of the adsorbate or adsorbate mixture components, mass transfer coefficients for the transport process film diffusion and intraparticle diffusion, adsorbent mass, volumetric flow rate and bed density.

Literature

- Schwarcenbah R. P., Gschwend P. M., Imboden D.M. (2003) Environmental organic chemistry, Second Edition, Wiley-Interscience, John Wiley&Sons, Hoboken, New Jersey
- S. E.Jørgensen (1991) Modelling in Environmental Chemistry, Developments in Environmental Modelling, 17, Elsevier, Amsterdam-London-New York-tokyo
- Worch E. (2007) Program LDF Version 2.3, Program Documentation.
- Sontheimer H., Crittenden J., Summers S. (1988) Activated carbon for Water Treatment, DVGW-Forschungstelle, Karlsruhe

- Snoeyink (1990) Adsorption of organic compounds in Water Quality and Treatment, AWWA, Fourth Edition, McGraw-Hill, Inc.
- Kümmel.R.; Worch,E. (1990) Adsorption aus wässrigen Lösungen. 1. Aufl. Leipzig: Deutscher Verlag für Grundstoffindustrie**.**
- Worch,E. (1991) *Zur Vorausberechnung der Gemischadsorption in Festbettadsorbern*, Teil I: Mathematisches Modell Chem. Tech., 43 , 3 111-114

5. Supplement- Biomass growth and kinetics in water treatment

Wastewater treatment utilizes biodegradation for removal of organic pollution responsible for high biological and chemical oxygen demand in municipal and some of industrial wastewaters. Bacterial comunity degrades organic mater by aerobic respiration and produce carbon dioxide, water, energy and new biomass. Organic nitrogen is transformed into ammonia or nitrate and organic phosphorous into ortophosphate. If the wastewater is not treated, this process is going on in the recipient that leads to oxygen consumption in rivers, lakes and consequently to death of fish.

There are several types of the treatment: suspended biomass (e.g. conventional activated sludge treatment) or fixed biofilm (trickling filter, rotating biodiscs, biologicaly activated carbon filters).

Conventional activated sludge process uses microorganisms for removal of pollution from water in their log phase of growth. After that phase they start to floculate and form settlable solids which are removed by settling. Part of the sludge is recirculated in the process and part of it goes to further utilization (e.g. methane and soil conditioner production).

In trickling filter wastewater slowly flows over the stones or some other crude material which surface is coated by microorganisms.

Rotating biodiscs are plastic discs immersed into the water and microorganisms are attached on their surface. Air is supplied by their rotation while aerobic respiration is going on.

Biologicaly activated carbon filters are used for both waste and drinking water treatment. They combine the processes of activated carbon adsorption and biological degradation of organic load due to attached biofilm on activated carbon.

To achieve biomass growth within a system, it is necessary that the retention time of biomass is long enough to reach the phase of reproduction. Time period necessary to reach the reproduction phase depends on growth rate of cells which is the function of metabolism and pollution utilization.

Rate of bacterial cells is defined by:

$$r_g = \mu X \tag{5.1}$$

where

r_g- rate of cell growth, mass/unit volume x time
μ - specific growth rate, time^{-1}
X – microorganism concentration, mass/unit volume

Cell growth is limited with supstrate concentration. It is defined by Monod equation:

$$\mu = \mu_m \frac{S}{K_S + S} \tag{5.2}$$

where

μ - specific growth rate, time^{-1}
μ_m - maximum specific growth rate, time^{-1}
S- concentration of the growth- limiting supstrate in the solution
K_S- half velocity constant, concentration of the supstrate at one half the maximum growth rate, mass/unit volume

Combination of equations (5.1) and (5.2) gives the equation for growth rate:

$$r_g = \frac{\mu_m XS}{K_S + S} \tag{5.3}$$

This equation of growth rate should be corrected for the energy spent on cell maintenance, cell death and predation. That is endogenous decay. So, the equation for total growth rate (mass/volume unit x time) is:

$$r_g{}' = \frac{\mu_m XS}{K_S + S} - k_d X \tag{5.4}$$

where

k_d- coefficient of endogenous decay, time^{-1}

Suspended growth treatment processess

For the continuous/flow stirred tank reactor with cellular recycle mass balans is that the biomass acumulation is equal to the difference of rate of flow of microorganism into the system boundary and rate of flow of microorganism out of the system boundary plus net growth of microorganism within the system boundary. Mathematicaly that means:

$$\frac{dX}{dt} V_r = QX_0 - [Q_w X + Q_e X_e] + V_r (r_g{}') \tag{5.5}$$

where
Qw- flow rate of liquid containing the biological cells to be wasted from the systemr
Qe- efluent flow at the outlet of the settler
Xe-microorganism conentration in settler efluent
r_g'- total microorganism growth rate, mass VSS/unit volume x time
V_r- reactor volume

If in equation (5.5) is included equation for the growth rate with assumption that in inflow water cell concentration is zero and steady state conditions prevail (dX/dt=0), we get the new equation:

$$\frac{Q_W X + Q_e X_e}{V_r X} = -Y \frac{r_{SU}}{X} - k_d \tag{5.6}$$

or

$$\frac{1}{\theta_c} = -Y \frac{r_{SU}}{X} - k_d \tag{5.7}$$

where θc is average retention time.

This parameter θc can be used as a process control parameter. Certain procent of biomass has to be removed from the reactor in order to be able to control the microorganism growth. So, if it is known that parameter has the value of 10 days for achievment of stabilization, it is necessery that 1/10 cell mass is every day taken out from the system.

Attached growth treatment processes

Biofilm is the solid phase which includes extracellular polymer gel, microbial cells and other different particles incorporated into the biofilm structure of biotic or abiotic origin. 1-D models usually deal with mass transport and biochemical reactions. Later models are able to deal with multisupstrate biofilm dynamics with several microorganism species. 2-D and 3-D modes have morphological element. Great attention is given to biofilm heterogenicity (geometry, chemistry, biology and physics). For biofilm modeling it is necessary to define several submodels:

1. biomass growth and decay based on nutrient consumption
2. model of biomass division and spreading to describe the increase of volume due to increased number of bacteria and possibly extracellular polymer production and spreading could be considered here.
3. supstrate transport and kinetics and equilibrium of reactions
4. biofilm detachment
5. fluid flow
6. biofilm attachment

One of the most pronounced problems is to accommodate to the same model all relevant, both slow (growth, decay, detachment) and fast processes (diffusion, reactions).

In general, there are two groups of models: based on activity and characteristics of individual cells (IbM, eng. individual based modeling) and models based on biomass as multiphase system (BbM, eng. biomass-based models).

Relating to the biomass growth, cell growth by absorbing nutrients and when it achieves critical mass it divides into two cells. General equation for the rate of mass

change m of bacteria i, in the time t and in the place x *[x y z]* can be written as follows:

$$\frac{dm_i}{dt} = r_{Xj}(m_i(t), C_S(x,t),... \tag{5.8}$$

Cs is parameter consisted of all concentrations of substrates and products which influence bacterial growth. Simple Monod's kinetic is acceptable in most of the cases. Sometimes model has to be complicated by inhibition relations, requirements for biomass decay, etc. The simplest models for biofilm spreading are similar to the crystal growth models.

Basic processes which contribute to the enlargement of the biofilm volume are determined by nutrient availability. Decay processes are also determined by the concentration of certain substances. Transport and reactions of substrate are defined by the physical lows. Dissolved substances are transported by molecular diffusion and convection. Concentration gradients are formed due to substrate utilization and product formation in biofilm. Rate of accumulation of the substance *i* in volume element must be in equilibrium wity transport rate within the borders of the volume element and with total rate of transformation (e.g. chemical reaction R_i). So, material balance for chemical species *i* am as follows:

$$\frac{\partial C_{Si}}{\partial t} = D_i \nabla^2 C_{Si} - u\nabla_{Si} + R_i(Cs, Cx) \tag{5.9}$$

Biofilm in trickling waste water filter

Growth of biofilm on trickling biologicaly active filter is given by the equation

$$r_S = f_0 h k_0 S^2 / (K_m + S) \tag{5.10}$$

where

r_s is the growth rate in thin layer,
h-depth of the layer,
k_0-maximum rate of substrate removal,
S-average supstrate concnetration,
K_m-half rate constant and
f_0-factor

Mass bilans for substrate removal in treatment process is given by the equation:

$$(\partial S / \partial t) dV = QS - Q(S + (\partial S / \partial z) dz) + dzW(-f_0 h k_0 S^2 / K_m + S) \tag{5.11}$$

Q-water flow,

W-width of the area where biofilm grows and
Z-adsorber height.

Since for steady state we have $\partial S/\partial t=0$ above mentioned mass balance can be simplified

$$Q(dS/dZ)= -f_0k_0hWS^2/(K_m+S) \quad (5.12)$$

Since $K_m \langle$ S we can write

$$dS/dZ=-f_0k_0hWS/Q \quad (5.13)$$

Integration from $S_e$to S_0 and from 0 to Z we get:

$$S_e / S_0 = e^{-(f_0hk_0)WZ/Q} \quad (5.14)$$

Where

S_e is concentration of substrate in effluent, and
S_0 concentration of susbtrate in influent

Further

$$WZ/Q=A_S/Q=ZA/QxA_S/V=S_aZA/Q \quad (5.15)$$

where

A_S surface of the filter filling,
V-volume of the biofilter,
S_a- specific surface of the filling,
A –biofilter crosssectional area

Product of f_0hk_0 can be replaced with one constant of substrate removal, k_s, so we can write

$$S_e / S_0 = e^{-k_S S_a ZA/Q} \quad (5.16)$$

Literature:

- Picioreanu C. And M.C.M. van Loosdrecht (2003) Use of mathematical modelling to study biofilm development and morphology in Biofilms in Medicine, Industry and Environmental Biotechnology-Characteristics,

Analysis and Control (Ed. Lens, P., Moran, A.P. Mahony T., Stoodley, P., O'Flaherry, V.) IWA Publishing

- Metcalf & Eddy, Inc (1991) Biological Unit Processes in Wastewater Engineering: Treatment, Disposal and Reuse (Eds. Tchobanoglous G. and Burton F. L) McGraw-Hill.

Project: 06SER02/02/003

Hemijski modeli u zaštiti životne sredine

Dr Ivana Ivančev Tumbas

1.Hemija životne sredine i modeliranje: šta i zbog čega nam je potrebno?

1.1. Zašto nam je potrebno modeliranje u hemiji životne sredine?

Poznavanje hemije životne sredine nam pomaže u rešavanju problema vezanih za upravljanje njenom zaštitom. Tipični primeri takvih problema su:

- Difuzija akcidentnog zgađenja u vazduh, vodu ili zamljište koja preti ekosistemu ili ljudskom zdravlju i koja zahteva efikasnu i brzu reakciju vlasti
- Izbor čiste tehnologije za kompaniju koja treba da reši problem sopstvenog otpada na najefikasniji način (ispunjavajući zahteve zakona u vidu graničnih vrednosti emisije sa najboljim dostupnim tehnologijama kojima treba upravljati)
- Prostorno planiranje u upravljanju otpadnim vodama, kanalizacijom, vodosnabdevanjem ili kontroli zagađivača u određenom području u okviru plana upravljanja rečnim slivom, emisijom i kontrolu aerozagađenja

Za rešavanje ovih i sličinih problema neophodno je raspolagati znanjem iz oblasti ekologije, zaštite životne sredine i tehnologije. Nije redak slučaj da različiti profili eksperata zajedno učestvuju u rešavanju problema koji su izuzetno kompleksni. Vrlo često se timovi sastoje od hemičara, građevinskih inženjera, mašinskih inžinjera, hidrogeologa, biologa, fizičara, lekara, pravnika, geologa, meteorologa i dr.

Područje hemije životne sredine obuhvata hemiju bioloških procesa prečišćavanja, hemiju fizičko-hemijskih procesa prečišćavanja, hemiju sudbine i transporta hemikalija u okolini i efekte biohemijskih i toksičnih supstanci, posebno na organizmima.

Hemija i biologija su možda najviše primenjivane nauke u oblasti nauka o životnoj sredini. One nam pružaju znanje neophodno za adekvatno upravljanje životnom sredinom. Uče nas kako različite hemikalije utiču na biološke sisteme i, zasnovano na tom znanju, mi smo u mogućnosti da predviđamo reakcije, i to znanje koristimo na dobrobit šire društvene zajednice. Sa druge strane, modeli su sinteza znanja o nekom sistemu i pored činjenice da su efikasno sredstvo upravljanja, oni nam omogućuju i sticanje novog znanja koje nam pomaže da bolje razumemo sistem i predvidimo njegovo ponašanje. Tako, koristeći modele ponekad možemo uočiti nove odnose među promenljivim i dalje koristiti to znanje zarad poboljšanja efikasnosti pojedinih tehnologija koje služe u prečišćavanju otpadnih tokova ili generalno za poboljšanje upravljanja kvalitetom životne sredine.

Upravljanje životnom sredinom omogućava

- Zaštitu ekosistema
- Zaštitu ljudskih bića i resursa

- Održivi razvoj
- Postizanje optimalnih rezultata u tehnologiji u odnosu na njen uticaj na životnu sredinu

Upravljanje životnom sredinom uz upotrebu matematičkih modela je efikasnije. Prevencija je najbolja zaštita i u tom smislu predviđanje ponašanja sistema ima najznačajniju ulogu prilikom preduzimanja preventivnih mera u određenim situacijama (npr. isključivanje serije bunara iz sistema vodosnabdevanja u slučaju pretnje akcidentnim zagađenjem ili npr. upravljanje membranskim filtracijom u sistemu proizvodnje vode za piće održavanjem optimalnog kvaliteta ulazne vode i time produženje veka trajanja membrane i smanjivanja operativnih troškova u proizvodnji).

Pored činjenice da želimo da znamo kako naša životna sredina funkcioniše i da predvidimo kako reaguje na različite vrste pritisaka uzrokovane aktivnostima čoveka, poznavanje hemije životne sredine je korisno u upravljanju tehnikama koje koriste različite procese u sprečavanju zagađenja (npr. vode, vazduha i zemljišta). Sve fizičko-hemijske reakcije na kojima se zasnivaju procesi za tretman otpadnih voda, gasova i zemljišta mogu biti modelirane matematičkim jednačinama sa manje ili više efikasnosti. Modeli se stalno dalje razvijaju i usavršavaju. Uticaj čoveka na životnu sredinu mora biti kvantifikovan u slučajevima kada postoji mogućnost povezivanja primene tehnologija koje čuvaju ili restauriraju životnu sredinu vezano za određene efekte u ekosistemu. Cilj ekotehnologije je da ojača ekosistemime u borbi sa zagađenjem.

Zagađenje u životnu sredinu dospeva na tri različita načina: otpadnom vodom, otpadnim gasovima i kao čvrst otpad. Prevencija zagađenja i kontrola su moguće upotrebom najboljih dostupnih tehnika koje minimiziraju emisiju polutanata, štede resurse i energiju.

Tako, u sprečavanju zagađenja vazduha koriste se:

- Filtri za uklanjanje čestica i toksičnih gasova u kojima je klasifikacija zasnovana na veličini čestica,
- Tornjevi za pranje u kojima je separacija zasnovana na gustini čestica
- Cikloni u kojima se koristi centrifugalna sila da bi se čestice razdvojile od vazduha

Sem uklanjanja čestica, dodatno se koriste hemijske reakcije za uklanjanje sumpornih i azotovih oksida ili se koristi adsorpcija za uklanjanje isparljivih organskih jedinjenja.

U preradi otpadnih voda koriste se različiti fizičko-hemijski tretmani: sedimentacija, flotacija, različite vrste filtracije, neutralizacije, precipitacije, flokulacije, jonske izmene i adsorpcije, elektrohemijske metode, termalne metode i dr. Svaki od ovih procesa je u potpunosti opisan mamtematičkim jednačinama i moguće ga je modelirati. Jedan od najvažnijih tretmana otpadnih voda je konvencionalni tretman aktivnim muljem. Taj primer modela biološkog prečišćavanja, o kome će biti više reči u poglavlju 5, je model ekosistema koji je napravio čovek. Stanište predstavlja reaktor, a biocenozu zajednica mikroorganizama u reaktoru koja vrši mineralizaciju organskog zagađenja otpadne vode.

U vezi sa čvrstim otpadom, imamo mogućnost ili da ga odložimo na zemljište, ili da vršimo sagorevanje, odn. incineraciju. Kada govorimo o odlaganju na zemljište, veoma je važno ispoštovati sva pravila modernog projektovanja i izgradnje sanitarnih deponija koje onemogućavaju dalju kontaminaciju životne sredine. U tom smislu je interesantno znati i modelirati rizik od zagađenja podzemnih voda i pokretljjivost zagađenja, znati procese razgradnje na deponijama i upravljati gasom koji se formira na deponiji njegovim korišćenjem.

Tretman zagađenog zemljišta se može vršiti *in situ* ili *ex situ* upotrebom hemikalija, ali i procesima bioremedijacije koji se sve više primenjuju u slučajevima kada treba rešiti problem biorazgradljivog zagađenja. Biodegradacija podrazumeva hemijske promene, od malih modifikacija molekula uz pomoć mikroorganizama (u najvećoj meri bakterija), do njegovog kompletnog razaranja u neškodljive oblike koji se nalaze u prirodi. Ona se dešava korak po korak i nije rezultat samo aktivnosti jednog specifičnog organizma. Obično nekoliko rodova organizma koji žive u sinergizmu su umešani u takve procese. To podrazumeva niz metabiličkih šema i velik broj enzimskih sistema.

U slučaju termalne dekompozicije otpada (incineracije) mora se takođe voditi računa o formiranju nekih neželjenih jedinjenja. To su toksični dioksini i furani čija se emisija u ovom procesu može umanjiti kontrolom procesa sagorevanja. Jedan od značajnih faktora je sastav samog otpada koji se sagoreva. Cilj je dobiti što više energije, uz što je manje moguće zagađenje.

1.2. Šta modeliramo?

1.2.1. Ciklusi elemenata

Naša životna sredina je podeljena u sfere. Biosferu čine svi nivoi Zemlje nastanjeni živim organizmima. Ona se sastoji iz ekosistema. Atmosfera predstavlja masu vazduha koja okružuje zemlju sa gornjom granicom fluida koja predstavlja razgraničenje prema spoljašnjem prostoru. Ekosfera obuhvata samo nastanjem prostor. Meteorološki, atmosfera je podeljena na troposferu i tropopauzu, stratosferu i stratopauzu, mezosferu i mezopauzu i termosferu. Fizičke osobine ovih delova zemlje su različite.

Gornji sloj zemljišta, nastanjen organizmima, naziva se pedosfera. Graniči se sa delom sačinjenim od stena koji se zove litosfera i koji se proteže do dubine od 100 km. Atmosfera i hidrosfera penetriraju u pedosferu.

Svi geološki procesi na Zemlji su opisani ciklusima elemenata, odnosno materijala. Ljudi interaguju sa prirodnim ciklusima kroz eksploataciju prirodnih resursa (korišćenje ruda, voda) i proizvodnjom otpada koji ulazi u cikluse.

Postoje četiri osnovna ciklusa nutrijentnih elemenata: ciklus ugljenika, azota, sumpora i fosfora, u kojima se dešavaju transformacije ovih elemenata kroz nekoliko različitih formi.

Tako, u ciklusu ugljenika, on ima različite forme koju su međusobno povezane reakcijama: ugljen-dioksid se rastvara u vodi dajući bikarbonate i karbonate koji se, ili talože u obliku soli kalcijuma i magnezijuma, ili bivaju upotrebljeni u procesu fotosinteze. Kroz fotosintezu ugljenik dobija organsku formu povezanu sa biogeohemijskim procesima, a kasnije može naći primenu u vidu fosilnih goriva u proizvodnji ksenobiotika. Organski ugljenik se može biološki razgraditi dajući

ponovo CO_2. neorganski vezan ugljenik u talozima može se ponovo rastvrati u zavisnoti od uslova u vodenoj sredini.

Što se tiče azota, on može biti prisutan u prirodi u različitim formama: organski (npr. NH_2 grupe proteina) i neogranske (amonijak, nitriti, nitrati, različiti oksidi). Fiksiranje atmosferskog azota mikroorganizmima i dalji mikrobiološki raspad organske materije koja ga sadrži produkuje amonijak koji pomoću mikroorganizama *Nitrosomonas spp.* i *Nitrobacter spp.* Dalje biva transformisan u nitrat i nitrit. Dalje, ovi se oblici u procesu denitrifikacije transformišu u molekulski azot.

Pored toga, možemo govoriti o ciklusu metala i posebnim ciklusima hemikalija koje dospevaju u ekosferu kao razultat ljudske aktivnosti. Vrlo važan antropogeni uticaj jesu teksični teški metali. Oni mogu biti prisutni u životnoj sredini u obliku jona i u obliku kompleksa koji interaguju npr. sa česticama u vodi. To može dovesti do njihovog inkorporiranja u sedimente i zemljište u različitoj formi.

Organske hemikalije, u zavisnosti od rastvorljivosti, mogu biti adsorbovane na čestice sedimenta ili rastvorene u vodi, putujući na taj način daleko od mesta zagađenja. Njihova dalja sudbina u životnoj sredini može biti biološka razgradnja, fotodekompozicija, oksidacija, pojavljivanje u obliku različitih formi i dr. Rastvorljivost i isparljivost određuju raspodelu hemikalija između tri dela životne sredine: vode, vazduha i zemljišta.

1.2.2. Najčešći tipovi zagađenja

Neke vrlo važne hemikalije za životnu sredinu su: pesticidi, industrijske hemikalije (rastvarači, reagensi), polihlorovani bifenili i dioksini, aromatični i poliaromatični ugljovodonici, organohlorne komponente i dr.

Najznačajnije poljoprivredno zagađenje prouzrokovano je pesticidima. Herbicide koristimo u najvećoj meri (40%), a insekticide i fungicide nešto manje (30%, odn. 20% respektivno). Druge grupe pesticida nisu u tako širokoj upotrebi (rodenticidi, regulatori rasta). Pored pesticida, poljoprivredno zagađenje su i đubriva. I pesticidi i đubriva pored toga što ulaze u lanac ishrane preko biljaka, vrše značajan uticaj na kvalitet voda pošto se zemljištem transportuju u plitke vodonosne slojeve podzemnih voda ili spiranjem dospevaju u površinske vode. Uticaj može biti do te mere negativan, da izvorišta vode za piće postaju neupotrebljiva. Takođe, može doći do akumualcije štetnih materija u vodenim organizmima (npr. u ribama).

Zagađenje sektora proizvodnje energije je uglavnom zagađenje vazduha uzrokovano emisijom oksida sumpora, azota, ugljen-monoksida, ugljovodonika i čestica.

Poslednje, ali ne i manje važno, je zagađenje koje emituju pojedinačna domaćinstva: deterdženti, sredstva za beljenje, sredstva za ličnu higijenu, lekovi, boje, rastvarači, lepak i dr.

Ponašanje supstanci je određeno njihovim molekulskim i fizičko-hemijskim karakteristikama: molekulska masa, funkcionalne grupe, isparljivost, rastvorljivost, Nernstov koeficijent raspodele, adsorpciono ponašanje i dr. Pored ovih karakteristika, ponašanje u životnoj sredini uslovljeno je i faktorima okoline: vlažnost, tip zemljišta (sadržaj organske materije, veličina čestica), temperature, pH, redoks potencijal, prisustvo mikroorganizama i dr. Efekti toksičnih materija u životnoj sredini mogu biti lokalizovani u prostoru i vremenu, ali nivo zagađenja zavisi od vrste zagađujuće materije i karakteristika samog ekosistema. Na primer, kišno vreme i obrada zemljišta

mogu uzrokovati poboljšanje transporta kroz zemljište, ili u drugom slučaju otpadna voda trpi uticaj mešanja u različitim pravcima pri distribuciji zagađenja u nekom recipijentu. Mešanje uzrokuje razblaženje hemikalija. Takođe, fotolitička i hidrolitička razgradnja ili precipitacija mogu da se dese. Visoka vlažnost atmosfere sprečava isparavanje hemikalija sa zemljišta ili površine vode i pospešuje prodor u suvo zemljište.

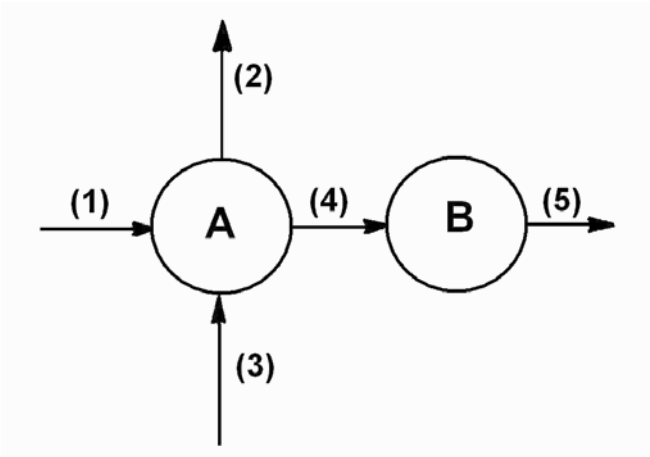
Prirodne vode su ugrožene prisustvom velikog broja neorganskih, organskih i bioloških zagađenja. Većina njih je posledica neprihvatljivog odlaganja otpada. Neki polutanti su visoko toksični (npr. kadmijum), dok drugi to nisu. Međutim, čak i netoksični polutanti kao razgradljiva organska materija, mogu do te mere potrošiti kiseonik u vodotoku da izazovu pomor riba, što predstavlja indirektno toksično elovanje. Neki konstituenti su normalno prisutni u malim količinama, dok u visokim koncentracijama mogu biti izuzetno štetni (npr. obična so). Velik broj procesa se dešava u vodenim ekosistemima: redoks reakcije, kompleksiranje, fotosinteza, precipitacija, kiselo-bazne reakcije, mikrobiološko delovanje, izmena gasova, taloženje i rastvaranje materija iz sedimenta.

1.2.3. Vrste modela

Vrste modela koje se koriste u hemiji životne sredinu su:

- Biogeohemijski modeli,
- Ekotoksikološki i toksikološki modeli,
- Modeli hemijskih vrsta,
- Modeli procesa biološkog tretmana
- Modeli procesa fizičko-hemijskog tretmana

Biogeohemijski modeli se fokusiraju na procese i transformacije različitih komponenti u ekosistemu, bilo da se radi o zagađenju ili prirodno prisutnim supstancama. Pomoću ovih modela predstavljaju se reke, jezera, akumulacije, lagune, rukavci, obalne zone, otvoreno more, močvare, pustinje, šume, poljoprivredno zemljište i dr. Na slici 1-1. predstavljen je konceptualni dijagram jednostavnog biogeohemijskog modela.



Slika 1.1. Primer konceptualnog dijagrama

A i B su promenljive stanja (npr. jedinice koncentracije ili gustina naseljenosti karkateristična za određenu vrstu). Strelice ukazuju na procese. Akumulacija se može opisati kao razlika između ulaza i izlaza. U dinamičkim modelima opisuju se promene promenljivih A i B u vremenu:

$$\frac{dA}{dt} = proces(1) - proces(2) + proces(3) - proces(4) \tag{1.1}$$

$$\frac{dB}{dt} = proces(4) - proces(5) \tag{1.2}$$

Najznačajnije je imati dobru formulaciju procesa koja se može opisati matematičkim jednačinama.

Suprotno pomenutom dinamičkom modelu, statički model je opisan iskazom

$$\frac{dA}{dt} = \frac{dB}{dt} = 0 \tag{1.3}$$

Ove jednačine se mogu koristiti za određivanje vrednosti A i B u uslovima dinamičke ravnoteže. Uopšteno govoreći, dinamički modeli zahtevaju obimnu bazu podataka.

Modeli ekoloških procesa se koriste za izradu modela ekosistema. I jedni i drugi modu biti opisani sa više ili manje detalja. Biogeohemijski modeli se odnose na biohemijske i geohemijske jedinjenja i elemente ekosistema. Kada se koriste u kontroli zagađenja moraju da obuhvate i sudbinu i raspodelu zagađenja i prirodnih komponenti.

Uobičajeni hemijski procesi koji su obuhvaćeni biogeohemijskim modelima su oksidacija, hidroliza, fotoliza, redukcija i kiselo bazne ravnoteže. Biološki procesi su rast, produkcija, mortalitet, imigracija, emigracija. Mikrobiološpki procesi su rast mikroorganizama, nitrifikacija, redukcija sulfata, mikrobiološka oksidacija i sl. Tipični fizički procesi su transport jedinjenja između vode i vazduha, advekcija, difuzija i dr. Upšteno govoreći, što je kompleksnija formulacija, više parametara je uključeno i neophodna je bolja i kompletnija baza podataka. ednačine koje se koriste za pomenute podmodele različitih procesa opisani su u literaturi

Ekotoksikološki modeli

Ekotoksikološki modeli se odnose na sudbinu i efekte toksičnih supstanci u ekosistemima i organizmima. Oni opisuju sudbinu i transport toksične supstance u ekosistemu, ali i populacionu dinamiku koja obuhvata efekte toksičnih supstanci (npr. mortalitet ili zastupljenost bilo kakvog toksičnog efekta). Neki primeri ekotoksikoloških modela su: model lanca ishrane, statički modeli protoka supstanci, dinamički modeli toksikanta u trofičnom nivou, ekotoksikološki modeli sa efektima toksikanata i ekotoksikološki modeli populacione dinamike. Primer je distribucija zagađenja PCB u vodi, sedimentu, ribama i drugim organizmima. Prilikom

modeliranja neophodno je uzeti u obzir svaki ulaz i izlaz. Npr. molekul PCB može dsopeti u ribu vodom (vrlo niska frakcija usled zanemarljive rastvorljivosti), al ii preko suspendovanih čestica na kojima je adsorbovan ili preko lanca ishrane, odnosno drugih kontaminiranih organizama. Izlučivanje je takoše potrebno uzeti u obzir.

Hemijski modeli

Ovi modeli se odnose na koncentracije različitih formi raznih jedinjenja u određenom ekosistemu. Često su zasnovani na diferencijalnim jednačinama koje predstavljaju promenu koncentracije reaktanta ili proizvoda u vremenu. Oni mogu biti podmodeli ekotoksikoloških i biogeohemijskih modela. Zasnovani su isključivo na hemijskim informacijama. Najčešće se radi o različitim formama hemijskih jedinjenja koje se nalaze u stanju ravnoteže. Modeli se sastoje od nekoliko jednačina koje se odnose na ravnotežu, a najzastupljeniji procesi su redoks procesi, formiranje kompleksa, hidroliza, fotoliza, kiselo-bazne ravnoteže, transfer mase iz jedne u drugu fazu, adsorpcija.

Hemijska reakcija je u stanju ravnoteže u slučaju da se količina reaktanata i proizvoda ne menja u vremenu. Brzine direktne i povratne reakcije su iste za uopštenu reakciju:

$$aA + bB \leftrightarrow cC + dD$$

Izraz za ravnotežu u kojem je K konstanta ravnoteže izgleda ovako:

$$K=[C]^c[D]^d/[A]^a[B]^b \qquad (1.4)$$

Konstanta ravnoteže zavisi jedino od temperature. U skladu sa Le Šatelje-ovim principom, za određenu promenu uslova ravnoteža se pomera u pravcu koji se suprostavlja promeni.

Formiranje kompleksa

Formiranje kompleksnih jedinjenja je veoma značajno u hemiji okoline. Sposobnost metalnih jona da formiraju komplekse utiče na njihovu toksičnost pošto se strutura i mobilnost menjaju u slučaju stvaranja kompleksa. Kompleksiranje povećava rastvorljivost, utiče na raspodelu između redukovanih i oksidovanih formi, utiče na toksičnost, menja osobine vezane za adsorpciju ili jonsku izmenu i menja stabilnost koloida koji sadrže metal. Metali koji stvaraju komplekse su: bakar, srebro, živa, litijum, aluminijum, gvožđe, nikal, mangan i dr. U prirodi postoji velik broj liganada: prirodne huminske i fulvinske komponente, sintetske hemikalije poput EDTA, anjone poput karbonata, silikata, nitrata, o-fosfata, jodida, acetata, amino-kiselina i dr. U zavisnosti od koncentracije metala i liganda i konstante stabilnosti kompleksa (K) pri određenoj pH vrednosti , redoks potencijalu i temperaturi, nastajanje kompleksa moguće je predvideti na osnovu ravnoteže:

$$Me^{n+} + L^{m-} \longrightarrow MeL^{(n-m)+} \qquad (I\text{-}1)$$

$$\frac{[MeL^{(n-m)+}]}{[Me^{n+}][L^{m-}]} = K \qquad (1.5)$$

Proces hidrolize utiče na mobilnost metala pošto može doći do povećanja rastvorljivosti metalnih jona. Ona se dešava u vodi, sa hidroksidnim i hidronijum jonima, a primeri aluminijuma i gvožđa opisani su jednačinama:

$$Al(OH)_3 + OH^- \rightarrow AL(OH)_4^- \qquad (I\text{-}2)$$

$$Al(OH)_3 + 3H^+ + nH_2O \rightarrow Al(H_2O)_{n+3}^{3+} \qquad (I\text{-}3)$$

$$Fe(OH)_3(H_2O)_3(s) + 3H^+ = Fe(OH)_4(H_2O)_2^- + 4H \qquad (I\text{-}4)$$

Smanjenje pH vrednosti često puta povećava toksičnost metala usled formiranja metalnih akva jona.

Hidroliza

Hidroliza organskih komponenti je takođe interesantna sa aspekta hemije okoline zbog različitih toksičnih efekata koje ispoljavaju različite organske supstance. Brzina reakcije hidrolize može da zavisi od pH vrednosti.

Konstante brzine se mogu predstaviti promenom koncentracije reaktanta ili proizvoda u vremenu. Jednostavnim testom moguće je odrediti kinetičke brzine iz kojih je onda moguće procenjivati koncentraciju reaktanta ili proizvoda pod određenim uslovima. Kada su u pitanju dugi intervali (pola dana i više) ravnotežni parametri se koriste za modeliranje, dok u slučaju kratkih intervala, od nekoliko sati ili čak minuta, za modeliranje koristimo kinetičke parametre.

Zakon brzine hemijske reakcije je matematička funkcija, specifična diferencijalna jednačina koja opisuje promenu (utrošak ili nastanak) jedne komponente kao funkciju koncentracija različitih hemijskih vrsta koje učestvuju u hemijskoj reakciji. Uopšteno, možemo pisati za brzinu transformacije komponente org:

$$D[org]/dt = k[i]^i[B]^b[C]^c... \qquad (1.6)$$

Gde eksponenti i,b,c,... govore koji je red reakcije u odnosu na odgovarajuće hemijske vrste: org, B, C,…Ovaj empirijski zakon ne govori ništa o mehanizmu reakcije. Čak i jednostavna reakcija može da se dešava u nekoliko razdvojenih stepena (elementarne promene na molekulima) u kojima se hemijske veze kidaju u nove veze formiraju pri konverziji jedne hemijske vrste u drugu. Svaki od ovih koraka, uključujući i povratne reakcije, može biti važan u određivanju ukupne brzine reakcije. Tako, konstanta brzine reakcije, k, predstavlja kombinaciju konstanti brzina reakcije nekoliko elementarnih reakcionih koraka.

Najjednostavniji primer je transformacija komponente A do komponente B. Njena brzina je proporcionalna koncentraciji i matematički se izražava zakonom brzine prvog reda:

$$d[A]/dt = -k[A] \qquad (1.7)$$

gde je k konstanta brzine reakcije prvog reda. Ako su molekuli vode uključeni u najsporiji korak reakcije, onaj koji određuje brzinu, reakciju treba opisati zakonom brzine reakcije drugog reda:

$$d[A]/dt=-k'[A][H_2O] \quad (1.8)$$

gde je k' konstanta brzine reakcije drugog reda. Pošto je voda obično prisutna u velikom višku i promena njene koncentracije zanemarljiva, zamenom $k=k'[H_2O]$, ponovo se dobija konstanta prvog reda, ali je zovemo konsantom brzine reakcije pseudo-prvog reda jer mehanizam obuhvata više vrsta na molekulskom nivou.

Mada većina reakcija koje srećemo nisu uistinu prvog reda, vrlo je pogodno da se za svrhe modeliranja prave pretpostavke koje nam dozvoljavaju redukciju reda reakcije, idealno ako je moguće, pseudo-prvog reda.

Redoks procesi

Vrlo su važni u životnoj sredini jer većina metala izmenjuje svoje elektrone sa drugim jedinjenjima u životnoj sredini. Tako, npr. arsen može biti snažno vezan za feromanganske agglomerate u aerobnim uslovima, dok pod anaerobnim uslovima dolazi do oslobađanja toksičnog arsena usled redukcije metala. Drugi primer je gvožđe sulfid koji u otpadnim vodama rudnika pri kontaktu sa vazduhom uzrokuje formiranje velikih količina sumporne kiseline koja nanosi ogromnu štetu okolini:

$$2FeS_2 + 2H_2O + 7O_2 = 2FeSO_4 + 2H_2SO_4$$

$$4FeSO_4 + O_2 + 2H_2SO_4 = 2Fe_2(SO_4)_3 + 2H_2O \quad (I\text{-}5)$$

$$Fe_2(SO_4)_3 + 6H_2O = 2Fe(OH)_3(s) + 3H_2SO_4$$

Reakcija oksidacije je veoma važna u vodenim ekosistemima gde je prisutno nekoliko oksidacionih vrsta: radikali, singletni kiseonik, peroksidi, ozon, i dr. Neki od njih su kratkozivuci, dok su drugi stabilniji. Brzina oksidacione reakcije predstavljena je jednačinom:

$$v = k_{ox} \cdot [C][ox] \quad (1.9)$$

Gde je k_{ox} specifična konstanta drugog reda za reakciju oksidacije na specifičnoj temperaturi, a [C] i [ox] su molarne koncentracije hemijske vrste oka se oksiduje i oksidanta respektivno. Ukupna brzina oksidacije kada većina oksidacionih vrsta simultano vrši oksidaciju je suma brzina svake reakcije ponaosob:

$$R_{ox}= (k_{Ox1}[Ox_1]+k_{Ox2}[Ox_2]\ldots\ldots)\cdot[C]=\left(\sum_{n=1}^{n=n} k_{Ox_n} \cdot [Ox_n]\right)\cdot[C] \quad (1.10)$$

Integracija u vremenskim granicama od 0 do nekog vremena t daje

$$\ln[C_0]/[C_t] = \sum_{n=1}^{n=n} k_{Ox_n} \cdot [Ox_n] \cdot t \quad (1.11)$$

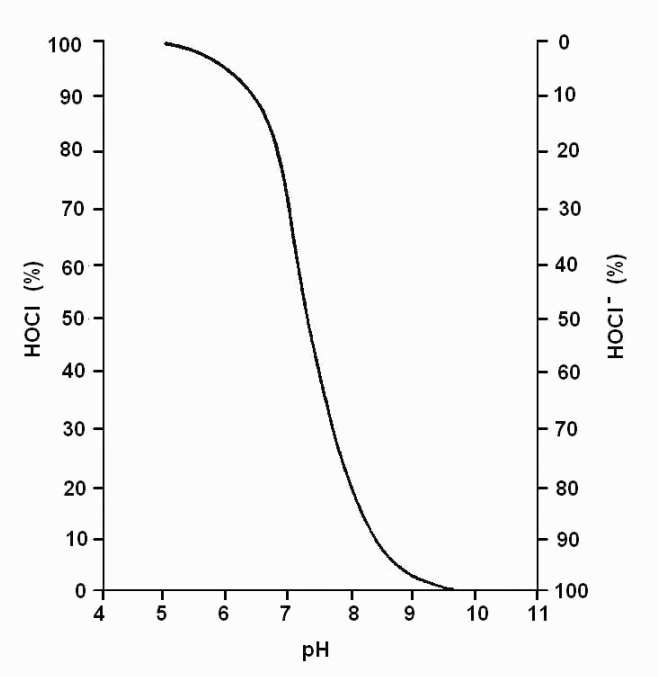
Ili, ako se koristi vreme poluraspada, $t_{1/2}$

$$t_{1/2} = \ln 2 / (\sum k_{Ox_n} \cdot [Ox_n]) \quad (1.12)$$

Mnogi redoks procesi su veoma brzi relativno u odnosu na vremenske razmake koji se primenjuju u većini modela. Dobar primer za to je kinetika oksidacije gvožđa (II) i mangana (II). Oba oksidaciona procesa se dešavaju u okviru nekoliko minuta. Kada se u modelima koriste vremenski intervali od pola dana ili dan, jasno je da je dovoljan podatak i ravnoteža, ali kada se u modelima koriste intervali od nekoliko sati ili minuta, mora se uzeti u obzir i kinetika oksidacije. Kada je ograničena količina kiseonika, njegov prenos može biti limitirajući faktor za brzinu reakcije. U koliko je to tako, a kiseonik se potroši skoro trenutno, podmodel pre svega mora prvo opisati prenos kiseonika što je tačnije moguće i u isto vreme anaerobni raspad organske materije. Oksidacija gvožđa i mangana će se dešavati brzinom određenom upravo prenosom kiseonika.

Kiselo-bazne reakcije

Skoro svi procesi u životnoj sredini zavise od pH vrednosti. Tako, sadržaj amonijaka i ugljen-dioksida u vodi zavisi od pH vrednosti, biološki procesi imaju svoju optimalnu pH vrednost (obično 6-8), pokretljivost teških metala je zavisna od pH, efikasnost hlora kao dezinficijensa je zavisna od pH (slika 1.2) pošto različite forme hipohlorne kiseline imaju različitu sposobnost dezinfekcije.



Slika 1. 2. Uticaj pH na prisutnost različitih formi HOCl u rastvoru

Kada je poznat sastav akvatičnog ekosistema moguće je izračunati alkalitet i puferski kapacitet in a tak način saznati koje vrste će biti prisutne u rastvoru.

Slično prethodno navedenim procesima, veliki značaj u modeliranju procesa u životnoj sredini imaju i procesi adsorpcije (vidi poglavlje 4), jonske izmene i fotohemijskih reakcija.

Literatura:

- S. E.Jørgensen (1991) Modelling in Environmental Chemistry, Developments in Environmental Modelling, 17, Elsevier, Amsterdam-London-New York-tokyo
- S.E. Manahan (1992) Toxicological Chemistry, Second Edition, Lewis Publishers, Chelsea Michigan , USA

2. Eyring-ova jednačina – primer jednostavnog hemijskog modela

Primer koji će biti predstavljen ovde jedan je od najjednostavnijih fizičko-hemijskih modela. On opisuje zavisnost brzine reakcije od temperature. Zasnovan je na teoriji prelaznog stanja.

$$A + B \xrightarrow{k} C \qquad v = \frac{dC}{dt} = k[A][B] \tag{2.1}$$

Reaktanti stupaju u reakciju formirajući nestabilni intermedijer

$$A + B \xleftrightarrow{k_a} AB \xrightarrow{k_2} C \tag{2.2}$$

Postoji "energetska barijera" na putu između reaktanata (A, B) i proizvoda (C). Barijera određuje minimum energije neophodne da se desi reakcija, i ta energija se naziva aktivacionom energijom. Čestice se približavaju jedna drugoj. One poseduju kinetičku energiju i konstantnu potencijalnu energiju. Približavajući molekuli reaktanta imaju dovoljno kinetičke energije da prevaziđu međusobno odbijanje elektronskih oblaka svojih atoma. Nastaje aktivirani intermedijer AB ili "prelazno stanje" sa maksimumom potencijalne energije. Radi se o nestabilnom kompleksu koji se raspada i formira proizvod C ili ponovo reaktante A i B.

Principi teorije prelaznog stanja su: postoji termodinamička ravnoteža između prelaznog stanja i stanja reaktanata na vrhu energetske barijere i brzina hemijske reakcije je proporcionalna koncentraciji reaktanata koji se nalaze u prelaznom stanju visoke energije.

Promena koncentracije kompleksa AB tokom vremena može se opisati sledećom jednačinom:

$$\frac{d[AB]}{dt} = k_1[A][B] - k_{-1}[AB] - k_2[AB] \tag{2. 3}$$

Zbog postojanja ravnoteže između aktiviranog kompleksa AB i reaktanata A i B, komponente $k_1 \cdot [A] \cdot [B]$ i $k_{-1} \cdot [AB]$ se poništavaju. Tako, brzina direktne reakcije je proporcionalna koncentraciji AB:

$$\frac{dC}{dt} = -\frac{d[AB]}{dt} = k_2[AB] \tag{2.4}$$

k_2 se računa:

$$k_2 = \frac{k_B T}{h} \tag{2.5}$$

k_B = Boltzmanova konstanta [$1.381 \cdot 10^{-23}$ J · K^{-1}]
T = apsolutna temperature u stepenima Kelvina (K)
h = Plankova konstanta [$6.626 \cdot 10^{-34}$ J · s]

k_2 je univerzalna konstanta prelaznog stanja (~ $6 \cdot 10^{-12}$ sec^{-1} na sobnoj temperaturi).

Dodatno, [AB] može biti izvedena iz kvazistacionarnog stanja ravnoteže između AB i A, B primenjujući zakon o dejstvu masa:

$$[AB] = K^*[A][B] \tag{2.6}$$

K^* = termodinamička konstanta ravnoteže

Zbog ravnoteže koja se brzo postiže, koncentracija reaktanata i aktiviranog kompleksa opada istom brzinom. Zbog toga, uzimajući u obzir obe jednačine 2.5 i 2.6 jednačina 2.4 postaje:

$$-\frac{d[AB]}{dt} = \frac{k_B T}{h} K^*[A][B] \tag{2.7}$$

Poređenje izvednog zakona brzine (2.1) i izraza 2.7 daje za konstantu brzine celokupne reakcije

$$k = \frac{k_B T}{h} K^* \tag{2.8}$$

Dodatno, termodinamički konstanta ravnoteže može se opisati jednačinom:

$$\Delta G = -RT \ln K^* \tag{2.9}$$

ΔG je određeno

$$\Delta G = \Delta H - T\Delta S \tag{2.10}$$

R = univerzalna gasna konstanta = 8.3145 J/mol K
ΔG = slobodna entalpija aktivacije[$kJ \cdot mol^{-1}$]
ΔS = entropija aktivacije [$J \cdot mol^{-1} \cdot K^{-1}$]
ΔH = activation enthalpy [$kJ \cdot mol^{-1}$]

ΔG je sloboddna aktivaciona entalpija (Gibsova slobodna energija) . U skladu sa jednačinom (2.9) ΔG predstavlja pokretačku silu reakcije, a znak određuje njenu spontanost. Ako je manja od nule, reakcija je spontana, ako je jednaka nuli, reakcija je

postigla stanje ravnoteže, a ako je veća od nule, reakcija nije spontana. Kombinacija jednačina (2-9) i izraza (2-10) i rešavanje za *lnk* daje:

$$\ln K^* = -\frac{\Delta H}{RT} + \frac{\Delta S}{R} \tag{2.11}$$

Eyring-ova jednačina se dobija supstitucijom jednačine (2.11) u jednačinu (2.8):

$$k = \frac{k_B T}{h} e^{-\frac{\Delta H}{RT}} e^{\frac{\Delta S}{R}} \tag{2.12}$$

$$\ln k = \ln \frac{k_B}{h} T - \frac{\Delta H}{R} \frac{1}{T} + \frac{\Delta S}{R} \tag{2.13}$$

$$\ln \frac{k}{T} = -\frac{\Delta H}{R} \frac{1}{T} + \ln \frac{k_B}{h} + \frac{\Delta S}{R} \tag{2.14}$$

Predstavljanje ln(k/T) u zavisnoti od 1/T daje pravu liniju (y = -mx + b), gde je

x = 1/T
y = ln(k/T)
m = - ΔH / R
b = y (x = 0)

ΔH se može računati iz nagiba *m*:

$\Delta H = -m \cdot R$.

na osnovu odsečka

$$y(x = 0) = \ln \frac{k_B}{h} + \frac{\Delta S}{R} \tag{2.15}$$

$$E_a = \Delta H + RT \tag{2.16}$$

Niske vrednosti E_a i ΔH znače da se reakcija odvija velikom brzinom a visoke vrednosti E_a i $\Delta H^{\ddagger}$, da je brzina reakcije mala. Tipične vrednosti E_a i $\Delta H^{\ddagger}$ su između 20 i 150 [kJ / mol].

Primer

Kuempel i sar. (Inorg. Chem., 12, 1036 (1973) su dobili sledeći set podataka za pseudo prvi red reakcije hidrolize:

$$Ru(NH_3)_5Br^+ + H_2O \xrightarrow{k} Ru(NH_3)_5H_2O^{2+} + Br^-$$

k (sec^{-1})	T (ºC)
1.2	15
3.8	20
5.4	25
8.3	30
12.2	35

Izračunaj slobodnu energiju aktivacije reakcije.

Literatura

- S. E.Jørgensen (1991) Modelling in Environmental Chemistry, Developments in Environmental Modelling, 17, Elsevier, Amsterdam-London-New York-Tokyo
- Varma, A., Morbidelli, M. (1997) Mathematical Methods in Chemical Engineering, Oxford University Press, New York, Oxford.

3. Alati u modeliranju: transport i reakcije

Postoje dva tipa transporta u prirodnim sistemima:

- Usled slučajnog, odn. nasumičnog kretanja (npr. molekulska difuzija, disperzija) i
- Usled usmerenog kretanja (npr. advekcija u struji vode, taloženje suspendovanih čestica usled gravitacije)

Transportni procesi uzrokovani slučajnim kretanjem su difuzivni, dok oni koji potiču od usmerenog kretanja su advektivni.

3.1. Slučajno kretanje

Slučajno, odn. nasumično kretanje, je prisutno svuda oko nas. Na molekulskom nivou termički pokreti atoma i molekula su nasumični. Slučajnost znači da pokret individualnog dela nekog medijuma (npr. molekul, deo vode i dr) ne može biti opisan tačno. Osnovni opis transporta nasumičnim kretanjem je Fikov zakon (zakon gradijenta fluksa). Fluks je definisan kao količina neke fizičke veličine (npr. koncentracije) koja se transportuje u jedinici vremena kroz jedinicu površine upravnu na pravac toka. Pretpostavljeno je da su podsistemi A i B i rastojanje među njima $\Delta x_{A/B}$, beskonačno mali. Razlika u koncentraciji teži nuli. Odnos ove dve razlike, Δkoncentracije: $\Delta x_{A/B,}$ jednak je prostornom gradijentu koncentracije i obično je različit od nule:

$$F_x = -b \frac{d}{dx}(koncentracija) \tag{3.1}$$

gde je b konstanta

Znak minus indicira da su tačke nasuprot gradijentu. Umesto subskripta A/B koristi se indeks x da označi osu duž koje se fluks dešava. Pretpostavlja se da je fluks određen promenama lokalnih svojstava. Ove promene (kao gradijent koncentracije, temperature, pritiska i sl.) predstavljaju pokretačku silu transporta. Matematički, gradijent je lokalno svojstvo funkcije. Puno fizičkih procesa se ponaša po ovom zakonu: molekulska difuzija, provođenje toplote, protok fluida kroz porozni medijum

Drugi Fikov zakon kaže da je lokalni opseg koncentracija u vremenu zbog difuzivnog transporta proporcionalnom drugom prostornom izvodu koncentracije:

$$\frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} \tag{3.2}$$

3.2. Granice u životnoj sredini

Mnogi važni procesi u životnoj sredini odvijaju se na granicama. To su površine na kojima se značajno ili diskontinualno menjaju osobine sistema (npr. granična poršina voda–vazduh i sediment-voda). Granice su okarakterisane fizičkim i hemijskim procesima i često puta je dovoljno upotrebiti koncept ravnoteže da bi se opisala neka granica (npr. Henrijev zakon za granicu voda-vazduh). Postoje tri tipa granica (prema obliku opšteg profila difuzije preko granice):

- Granice "uskog grla" sa jednom zonom u kojoj se koeficijent transfera značajno smanjuje u odnosu na okolinu (npr. površina vode reke između dve turbulentne zone, vode i vazduha)
- "Zidna" granica (npr. sediment-voda u jezerima i okeanima) gde je samo jedna zona turbulentna
- Difuzna granica – disperzija na ivici fronta polutanta, kada nema velike promene difuziviteta

Različite matematičke jednačine se koriste da bi se objasnile ovako definisane granice pri modeliranju sudbine i transporta polutanata u ekosistemima.

3.3. Modeli kutije

Najjednostavnija i najčešće najpodesnija alatka za modeliranje je model jedne kutije (eng. one-box model). Ovi modeli opisuju system kao jedinstven prostorno homogen entitet. Homogen znači da nema daljih prostornih varijacija uzetih u razmatranje. Ovi modeli mogu imati jednu ili više promenljivih stanja, npr. srednja koncentracija jedne ili više komponenti na koje utiču spoljne promenljive i unutrašnji procesi.

U sistemu koji ima konstantnu zapreminu maseni bilans komponente i je opisan jednačinom:

$$\frac{dC_i}{dt} = \frac{1}{V}\left(I_i - Q_i - \Sigma R_i + \Sigma P_i\right) \tag{3.3}$$

Gde su ΣR_i i ΣP_i izrazi za ukupnu internu brzinu potrošnje i proizvodnje procesa koji obuhvataju komponentu i. Koncentracije su uslovljene ulazom I_i i izlazom Q_i. U linearnom modelu jedne kutije (eng. linear one box model) sa jednom promenljivom koristimo linearnu funkciju da opišemo spoljašnje promenljive. Na primer, spoljni input nije zavistan od C_i.

$$f_p(C_i) = a_p(t) + b_p(t)C_i \tag{3.4}$$

Ovaj model se može koristiti za proračun ukupne mase i srednje koncentracije polutanta u jezeru na osnovu ponovljenih merenja koncentracije polutanta na različitim dubinama. Račun podrazumeva prvobitno znanje vezano za prostornu distribuciju polutanta, horizontalni gradijenti koncentracije moraju biti tako mali da se

ukupna masa može računati kao srednja vrednost koncentracija merenih duž vertikalnog profila na najdubljoj lokaciji jezera. Moraju se znati ulazna i izlazna koncentracija. Ako je pollutant isparljiv, jedini značajni mehanizam uklanjanja (sem gubitka na ulazu) je razmena vazduh-voda. Reakcije *in situ* i sorpcija na sedimentima se ne smatraju relevantnima.

Primer modela **jedne kutije sa dve promenljive** je slučaj dve hemikalije, A i B, gde se A transformiše u B hemijskom reakcijom i obratno. Sistem je opisan pomoću dve koncentracije C_A i C_B, pomoću dve ulazne funkcije nultog reda, J_A i J_B, pomoću dve izčazme funkcije prvog reda, k_AC_A i k_BC_B, i transformacijom prvod reda od A u B i obratno. Ako ne bi bilo transformacije među hemijskim vrstama A i B, mogli bismo opisati svaku hemikaliju zasebnim linearnim modelom jedne kutije.

Model dve kutije (eng. two box model) se koristi uspešno u opisivanju sistema koji imaju dva prostorna podsistema povezana jednim ili sa nekoliko transportnih procesa. Sledeći korak su linearni multidimensionalni modeli.

Konačno, promenljive modela (npr. koncentracija) se menjaju u vremenu i prostoru. Tokom poslednjih nekoliko decenija, polje trodimenzionalnoh prostorno vremenskog modeliranja je prošlo kroz veoma brz razvoj. Glavni mehanizmi su transformacije, usmeren transport (advekcija, tok, taloženje čestica) ili nasumični transport (difuzija i disperzija). Važno je zapamtiti da razlikujemo laminarni i turbulentni tok. Disperzija može biti prikazana istim zakonom kojim prikazujemo i difuziju ali je ona po prirodi drugačija. Ona je rezultat brzine smicanja koja predstavlja razliku susednih strujnica u advektivnom toku. Na primer, usled turbulentne razmene upravne na pravac toka, vodene čestice kontinualno menjaju strujnice duž kojih se kreću. Kako se strujnice kreću različitim brzinama, svaki delić vode ima svoju individualnu brzinu.

Literature

- Schwarcenbah R. P., Gschwend P. M., Imboden D.M. (2003) Environmental organic chemistry, Second Edition, Wiley-Interscience, John Wiley&Sons, Hoboken, New Jersey

4. Modeliranje adsorpcije

Adsorpcija je pričvršćivanje molekula iz tečnog ili gasovitog rastvora na drugu fazu (dvodimenzionalnu površinu). Vrlo je čest process u prirodi, ali i u tretmanima otpadnih tokova, bilo da se radi o vodi ili vazduhu. Proces adsorpcije utiče na transport i sudbinu hemikalija u okolini i ima veliku važnost u proceni rizika.

Supstance mogu biti adsorbovane i iz vode i iz vazduha na zemljište, glinu, različite površine, prirodnu organsku materiju i dr. Upotreba granulovanog aktivnog uglja je vrlo česta u tretmanu otpadnih gasova i uklanjanju organskih komponenti iz njih (npr. isparljive organske komponente prisutne u dimnim gasovima), za uklanjanje organskih materija iz vode za piće, i dr. Na primer, razumevanje procesa adsorpcije u tehnologiji vode se sastoji od znanja vezanog za specijalne karakteristike aktivnog uglja, znanja o adsorpcionoj ravnoteži i kinetic ii dizajnu samog procesa. Vezano za dizajn procesa, postoji mogućnost primene aktivnog uglja u prahu (PAC) ili granulovanog aktivnog uglja u adsorberima sa fiksiranim slojem. Nekoliko novih procesa se razvija (npr. hibridni procesi u kojima se kombinuje membranska filtracija sa aktivnim ugljen ili adsorberi sa permeabilnim sintetičkim kolektorima). Modeliranje nam omogućava da predvidimo ponašanje polutanta i bolje upravljamo tretmanom vode ili vazduha.

4.1. Adsorpciona ravnoteža

Na osnovu podataka o adsorpcionoj ravnoteži može se računati adsorpcioni kapacitet aktivnog uglja. On se određuje na osnovu adsorpcione izoterme koja opisuje ravnotežu u zatvorenom sistemu koji se sastoji od rastvora supstanci koje želimo da uklonimo i količine uglja koja je u kontaktu sa rastvorom.

Za evaluaciju adsorpcione izoterme dodaju se definisane količine aktivnog uglja u nekoliko boca koje sadrže istu, unapred definisanu, zapreminu rastvora sa početnom koncentracijom jedne rastvorene supstance. Rastvori se mućkaju dok se ne postigne adsorpciona ravnoteža i na kraju se odredi ravnotežna koncentracija supstance za svaku primenjenu dozu uglja. Vreme neophodno za postizanje ravnoteže može varirati od nekoliko sati do nekoliko dana, ali i godina. Iz ovih podataka računaju se adsorpcioni parametri (eq. 4.1). Stepen do kog se može koristiti slobodna površina uglja za adsorpciju zavisi od koncentracije rastvorka u rastvoru koji se meša sa ugljem. Specifična relacija definiše adsorpcionu izotermu na određenoj temperaturi. Najčešće je adsorpciona ravnoteža opisana empirijskom Freundlich-ovom jednačinom:

$$q=K_F c^n \qquad (4.1)$$

gde su

q – koncentracija na čvrstj fazi koja opisuje količinu adsorbovane supstance
K_F-Freundlich-ova konstanta
c- ravnotežna koncentracija u rastvoru
n-Freundlich-ov eksponent

Freundlich-ov eksponent i konstanta mogu se lako odrediti nelinearnom regresijom transformacijom jednačine u:

$$\lg q = \lg K_F + n \lg c \qquad (4.2)$$

Nagib log c u zavisnosti od log q jednak je eksponentu, n, a koncentracija na čvrstoj fazi pri $C_{iw}=1$, jednaka je Freundlich-ovoj konstanti, K_F. Adsorpcioni kapacitet uglja raste sa porastom vrednosti K_F i kada opada n.

Vrednost n predstavlja promenu u slobodnoj energiji sorpcije rastvorka na sorbent u odgovarajućem opsegu koncentracija. Kada je *n = 1*, izoterma je linearna i slobodna energija sorpcije je jednaka za sve koncentracije, kada je n < 1, izoterma je konkavna i sa porastom koncentracije sorbata slobodna energija sorpcije opada. Kada je ni > 1, izoterma je konveksna i sa porastom koncentracije sorbata slobodna energija sorpcije raste.

Pored Freundlich-ove jednačine, za objašnjenje adsorpcione ravnoteže može se koristiti i Langmuir-ova jednačina:

$$q = q_m \cdot \frac{K_L \cdot c}{1 + K_L \cdot c} \qquad (4.3)$$

Gde su

K_L-Langmuir-ova konstanta,
q_m- maksimalna koncentracija na čvrstoj fazi u monomolekulskom sluju koji pokriva adsorpcionu površinu,
c- ravnotežna koncentracija,
q - koncentracija na čvrstoj fazi u stanju ravnoteže.

Langmuir-ova jednačina ima za pretpostavku da je promena slobodne energije, entalpije i entropije usled adsorpcije konstantna za sve koncentracije na čvrstoj fazi. To se obično povezuje sa homogenom površinom. Međutim, često ovaj uslov nije ispunjen i entalpija adsorpcije se menja sa porastom koncentracije na čvrstoj fazi.

Dve konstante Langmuir-ove izoterme se mogu odrediti na osnovu podataka primenom nelinearne regresije, ili ako se rezultati adsorpcionog testa crtaju na odgovarajući način, linearnom regresijom.

Multikomponentna adsorpcija

U slučaju kada je u rastvoru prisutno više supstanci koje se adsorbuju imamo pojavu kompeticije za adsorpciona mesta među njima. To je tzv. višekomponentni system. U ovom slučaju koncentracije supstanci na čvrstoj fazi bilo koje od supstanci će biti redukovane u poređenju sa jednokomponentnim sistemom.

Dobro poznat model za opisivanje višekomponentne adsorpcije je zasnovan na teoriji idealne adsorpcije u rastvoru (eng. Ideal Adsorbed Solution Theory (IAST)). Ona je zasnovana na pretpostavci da se adsorpciona ravnoteža ravnoteža dešava

između dvodimenzionalne površine i rastvora. Adsorpcioni medijum je termodinamički inertan, a adsorpciona mesta npr. površine aktivnog uglja su dostupna svim organskim molekulima prisutnim u rastvoru na isti način i adsorpciona ravnoteža je reverzibilna.

Za račun adsorpcione ravnoteže u smeši neophodno je poznavati samo podatke za jednokomponentne sisteme. Freundlich-ova jednačina je upotrebljena i pretpostavljeno je da je Freundlich-ov eksponent konstantan u opsegu $c_i=0$ to $c_i=c_{0.}$

$$c_i = \frac{q_i}{\sum_{j=1}^{N} q_j \left(\frac{\sum_{j=1}^{N} \frac{q_j}{n_j}}{\frac{K_{F,i}}{n_i}} \right)^{1/n_i}} \quad \text{N= broj komponenti} \qquad (4.4)$$

IAST se često koristi kao osnova za opisivanje adsorpcije u nepoznatim smešama adsorbujućih organskih supstanci. Adsorpciona analiza pomoću IAST takođe se može koristiti za posmatranje uklanjanja supstanci sa različitim adsorpcionim afinitetima tokom tretmana voda.

4.2. Adsorpciona kinetika

Kinetika adsorpcije je brzina postizanja adsorpcione ravnoteže pomoću dva difuziona procesa: iz rastvora na spoljašnju površinu adsorbenta (eksterni transfer mase ili difuzija kroz film) i difuzija kroz pore sistema čestica gde se dešava adsorpcija na unutrašnjoj površini (interni transfer mase). Matematički opis difuzije kroz film je dat Fick-ovim zakonom:

$$n_{L,i} = D_{L,i} \frac{dc_i}{d\delta} \qquad (4.5)$$

Gde su

$n_{L,i}$ brzina transfera mase po jedinici površine
$D_{L,i}$ vodeni koeficijent difuzije adsorbata i
δ lokacija u okviru graničnog sloja debljine δ

Intergraljenjem se dobija:

$$n_{L,i} = \beta_{L,i}(c_i - c_i^*) \qquad (4.6)$$

Gde je

$\beta_{L,i}$ koefficijent difuzije kroz film,
c_i^* ranotežna koncentracija komponente i na spoljnoj površini čestice aktivnog uglja i
c_i koncentracija komponente u rastvoru.

Određivanje površinskog koeficijenta difuzije može se izvršiti eksperimentalno ili upotrebom dobro poznatih empirijskih korelacija.

Unutrašnji transfer mase dešava se usled difuzije molekula kroz tečnošću ispunjene pore ili difuzijom adsorbovanih molekula po zidovima pora (površinska difuzija). U poslednjem slučaju pokretačka sila je gradijent na čvrstoj fazi i umesto vodenog difuzionog koeficijenta, u tom slučaju govorimo o difuzionom koeficijentu površinske faze i gustini čestica:

$$n_{S,i} = \rho_p \cdot D_S \cdot \frac{\partial q_i}{\partial r} \tag{4.7}$$

Sile koje utiču na kinetiku adsorpcije organskih supstanci iz razblaženih rastvora na poroznom uglju su često sasvim različite od onih koje kontrolišu konačni kapacitet uglja za adsorpciju. Ograničavajući mehanizam je obično ili difuzija kroz film ili međučestični transport, što u velikoj meri zavisi od hidrodinamičkog karaktera sistema u kome se aktivni ugalj koristi. Suprotno, krajnja pozicija ravnoteže je određena silama adsorpcije, koje su po prirodi hemijske ili fizičke. Kao rezultat mogućih razlika u prirodi kinetičkih i ravnotežnih sila, faktori koji poboljšavaju brzinu adsorpcije mogu znatno umanjiti kapacitet poroznog uglja za neke adsorbate i obratno.

4.3. Primena u procesima

Danas se u tretmanu voda koriste dva tipa procesa zasnovana na adsorpciji na aktivnom uglju: process sa aktivnim ugljem u prahu (PAC) u sekvencijalnom operaciji adsorpcije i proces sa granulovanim aktivnim ugljem (GAC) u adsorberu sa fiksiranim slojem.

Aktivni ugalj u prahu

Adsorpcija na aktivnom uglju u prahu se često puta naziva i kontaktna filtracija pošto tipična primena podrazumeva tretman u tanku u kome se ugalj meša sa vodom i nakon toga se vrši filtracija ili taloženje. Ako voda koja treba da se obradi sadrži početnu koncentraciju supstance koju je neophodno ukloniti, C_0, u zapremini V, u koju se dodaje aktivni ugalj u količini m i nakon dostizanja ravnoteže merimo koncentraciju C, može se pisati jednačina masenog bilansa:

$$VC_0+mq_0=VC+mq \tag{4.8}$$

Gde je q koncentracija supstance na čvrstoj fazi.

Pošto je početna koncentracija na čvrstoj fazi nula, jednačinu 4.8 možemo preurediti:

$$q = \frac{V}{m}(C_0 - C) \tag{4.9}$$

Kada su poznati Freudlich-ovi parametri za neku hemikaliju moguće je izračunati količinu aktivnog uglja neophodnu za uklanjanje supstance do određenog stepena na osnovu jednačine 4.9. Prvo se računa koncentracija na čvrstoj fazi iz jednačine 4.1, i nakon toga neophodna količina uglja (m) zasnovana na jednačini 4.9.

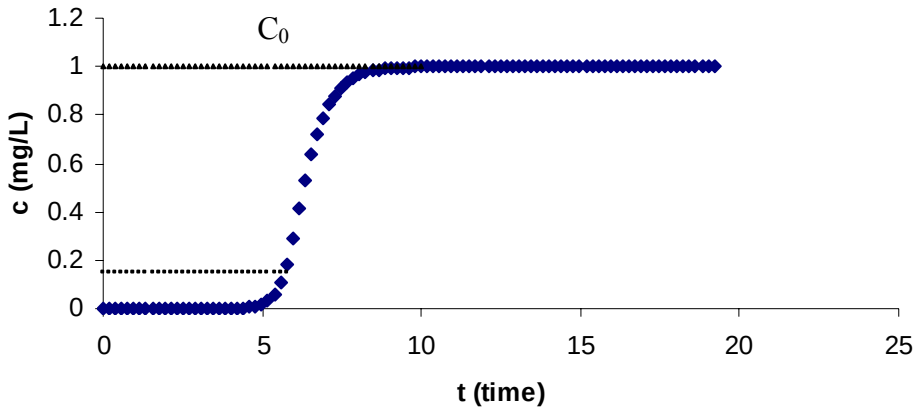
Adsorpcija na granulovanom aktivnom uglju

Tokom kontinualne operacije voda i adsorbent su u kontaktu tokom celog procesa bez periodičnog razdvajanja faza (primer GAC adsorbera sa fiksiranim slojem).

Projektovanje adsorbera sa fiksiranim slojem i predviđanje dužine adsorpcionog ciklusa zahteva znanje o stepenu zasićenosti uglja u tački proboja. Obično se proces vodi u seriji kolona sa aktivnim ugljem kroz koje se pumpa voda i vreme kontakta je u opsegu 15-60 min.

Tačka u kojoj nečistoća, odnosno koncentracija hemikalije prevazilazi kriterijum za kvalitet efluenta naziva se tačka proboja (Cb na slici 4.1). Deo krive između početnog pojavljivanja u efluentu i tačke u kojoj je koncentracija u efluentu jednaka ulaznoj naziva se kriva proboja). Tokom adsorpcije gornji deo kolone se zasiti nečistoćama, dok donji deo uglja ostaje relativno svež. Između ova dva stanja nalazi se adsorpciona zona u kojoj se dešava uklanjanje. Tokom vremena kolona se polako zasićava, a adsorpciona zona se polako pomera na dole. Tipična kriva proboja predstavljena je na slici 4.1.

Kriva proboja treba da je strma, odnosno koncentracija rastvorka od interesa u efluentu brzo raste od vrednosti bliskih nuli do one koje se nalaze u ulaznoj vodi. Proizvoljno uzeta niska vrednost C_B je izabrana kao probojna koncentracija, a kolona se smatra potrošenom kada se koncentracija u efluentu popune na drugu proizvoljnu vrednost blisku vrednosti C_0.



Slika 4.1. Kriva proboja

Adsorpciona zona, deo adsorbenta u kome se koncentracija menja od C_B do blizu C_0 ima konstantnu visinu Z_A (m). Adsorbent iza zone je kompletno zasićen rastvorkom. U okviru zone, stepen zasićenja varira od 0-100%. Visina zone zavisi od brzine adsorpcije i brzine protoka rastvora. Smanjenje veličine čestica ugljenika, povećanje koeficijenta difuzije adsorbata i jača adsorpcija (veća vrednost Freudlich-ovog parametra K) smanjuju visinu adsorpcione zone.

U situaciji u kojoj je probojna koncentracija definisana kao minimalna koncentracija koja se može detektovati, kritična visina kolone sa aktivnim ugljem je jednaka visini zone transfera mase. Ukoliko je adsorber visok tačno koliko i zona transfera mase, to dovodi do trenutne pojave supstance u efluentu u koncentraciji jednakoj probojnoj koncentraciji čim adsorber počne sa radom. Kritična visina, brzina protoka i površinski presek kolone se koriste za račun vremena zadržavanja (minimalna zapremina adsorbera ispunjena ugljem i podeljena sa zapreminskim protokom, eng. empty bed contact time):

$$\frac{Z_{critical}}{Q/A} = EBCT_{\min} \tag{4.10}$$

EBCT ima značajan efekat na učinak uglja. Za datu situaciju kritična dubina GAC i odgovarajuće minimalno EBCT postoje koje more biti premašeno ako želimo da nam adsorber proizvodi vodu prihvatljivog kvaliteta.

Kada je probojna koncentracija veća od minimalne koncentracije koju možemo detektovati , kritična visina je manja nego visina zone transfera mase.

$$\text{EBCT=V/Q=}\frac{L_{Bed}}{Q/A} \tag{4.11}$$

Gde je

V- zapremina uglja u kontaktoru
Q- zapreminska brzina protoka
L_{Bed}-dubina sloja
A-površina sloja

Masa supstance adsorbovana po jediničnoj masi adsorbenta povećava se kako raste procenat iskorišćenja adsorbenta i shodno tome broj bed volumena vode koji je obrađen pre proboja takođe raste do maksimalne vrednosti. Povećanje EBCT, ili dubine sloja utiče na troškove procesa. Ako je adsorber veći, fiksni troškovi rastu. Operativni troškovi opadaju jer opada brzina korišćenja uglja i frekvencija njegove zamene. Zbog toga je izuzetno važno raditi sa optimalnom dubinom sloja i vremenom kontakta.

4.4. Model difuzije po homogenoj površini (HSDM)

Nekoliko korisnih matematičkih modela adsorpcionih procesa su dostupni. Jedna od njih je model difuzije po homogenoj površini (eng. homogenous surface

diffusion model (HSDM)) koji se zajedno sa svojim modifikacijama široko koristi u predviđanjima rada adsorpcionih sistema.

U slučaju jednokomponentnog sistema HSDM model za disperzni tok sastoji se od nekoliko elemenata: kocentrisanje u tečnosti, transport advekcijom, disperzija i adsorpcija:

$$\varepsilon \cdot \frac{\partial c(t,z)}{\partial t} + v_F \cdot \frac{\partial c(t,z)}{\partial z} - D_Z \cdot \varepsilon \frac{\partial^2 c(t,z)}{\partial z^2} + \frac{6 \cdot \beta_{L\cdot} \cdot (1-\varepsilon)}{d_P} \cdot \left(c(t,z) - c^*(t,z)\right) = 0 \quad (4.12)$$

Gde su

ε- poroznost sloja
ν_F- linearna brzina filtera
D_z- difuzioni koeficijent
β_L- koeficijent transfera u filmu
d_p – gustina čestica

Model linearne pokretačke sile (eng. Linear Driving Force Model (LDF)) (Kummel, 1990; Worch 1991) je zasnovan na HSDM modelu za disperzni tok. U njemu je pretpostavljen linearni gradijent za difuziju po homogenoj površini zrna GAC. Model je zasnovan na gore pomenutoj jednačini, bez izraza za disperziju i njime su obuhvaćene spoljašnja difuzija kroz film i unutrašnji transfer mase kroz česticu:

$$\frac{dq}{dt} = \frac{k_F a_V}{\rho_b}(c - c_S) \quad (4.13)$$

Gde je

$k_F a_V$- koeficijent zapreminskog tranfera mase za difuziju kroz film (koeficijent transfera mase, k_F, pomnožen sa površinom dostupnom za transfer mase normalizovanom na zapreminu filtra, a_V),

c_S-koncentracija na spoljašnoj površini čestice, ρ_b- gustina sloja

$$\frac{dq}{dt} = k_S a_V (q_S - q\ ) \quad (4.14)$$

gde je

$k_S a_V$- koeficijent zapreminskog transfera mase za unutrašnji transport kroz čestice (koeficijent transfera mase), k_S, pomnožen sa površinom dostupnom za transfer normalizovanom na zapreminu filtra, a_V,

q je opterećenje i q_S je opterećenje na spoljašnjoj površini čestice.

U slučaju difuzije kroz film, razlika u koncentraciji između rastvora i spoljašnje površine zrna je pokretačka sila procesa. Kod unutrašnjeg transporta unutar čestice pokretačka sila je razlika opterećenja na spoljašnjoj površini čestice i srednjeg opterećenja čestice. Koeficijent prenosa mase $k_S a_V$ povezan je sa koeficijentom difuzije kroz pore, D_P, kao is a površinskim difuzionim koeficijentom, D_S:

$$k_S a_V = \frac{15 D_S}{r_p^2} \tag{4.15}$$

$$k_S a_V = \frac{15 D_p}{r_p^2} \frac{c_0}{q_0 \rho_p} \tag{4.16}$$

Gde je

r_p- prečnik zrna,
c_0-početna koncentracija,
q_0-ravnotežno opterećenje povezano sa c_0,
ρ_p- gustina zrna.

U slučaju jednokomponentne adsorpcije koncentracija i opterećenje na spoljašnoj površini zrna su povezani sa Freundlich-ovom izotermom, dok je u slučaju multikomponentnih sistema neophodno koristiti druge modele.

Program LDF (verzija 2.3) (Worch (2007)) može se koristiti za proračun adsorbera i krive proboja za jednokomponentne i višekomponentne sisteme upotrebom pomenutih IAST i LDF modela. Ulazni podaci zahtevaju za krivu proboja koncentracije, Freundlich-ove koeficijente i eksponente za adsorbate ili smeše, koeficijente transfera mase za difuziju u filmu i unutar zrna, masu adsorbenta, volumetrijsku brzinu protoka i gustinu sloja.

Literatura

- Schwarcenbah R. P., Gschwend P. M., Imboden D.M. (2003) Environmental organic chemistry, Second Edition, Wiley-Interscience, John Wiley&Sons, Hoboken, New Jersey
- S. E.Jørgensen (1991) Modelling in Environmental Chemistry, Developments in Environmental Modelling, 17, Elsevier, Amsterdam-London-New York-tokyo
- Worch E. (2007) Program LDF Version 2.3, Program Documentation.
- Sontheimer H., Crittenden J., Summers S. (1988) Activated carbon for Water Treatment, DVGW-Forschungstelle, Karlsruhe
- Snoeyink (1990) Adsorption of organic compounds in Water Quality and Treatment, AWWA, Fourth Edition, McGraw-Hill, Inc.
- Kümmel.R.; Worch,E. (1990) Adsorption aus wässrigen Lösungen. 1. Aufl. Leipzig: Deutscher Verlag für Grundstoffindustrie**.**

- Worch,E. (1991) *Zur Vorausberechnung der Gemischadsorption in Festbettadsorbern*, Teil I: Mathematisches Modell Chem. Tech., 43 , 3 111-114

5. Dodatak- Rast biomase i kinetika u tretmanu voda

Tretman optadnih voda koristi biorazgradnju za uklanjanje organskog zagađenja odgovornog za visoku biološku i hemijsku potrošnju kiseonika u komunalnim i nekim industrijskim otpadnim vodama. Bakterijska zajednica razgrađuje organsku materiju aerobnom respiracijom i proizvodi ugljen-dioksid, vodu, energiju i novu biomasu. Organski azot se pretvara u amonijum jon ili nitrat, organski fosfor u ortofosfat. U slučaju da se voda ne prerađuje, procesi se dešavaju u recipijentu, što vodi do potrošnje kiseonika u vodotoku i mogućeg pomora ribe.

Postoje dva tipa tretmana: u suspendovanoj biomasi (primer konvencionalnog tretmana sa aktivnom muljem) ili sa fiksiranim biofilmom (kapajući filtar, rotirajući biodisk ili filtri sa biološki aktivnim ugljem).

Konvencionalni tretman sa aktivnim muljem koristi mikroorganizme za razgradnju organskog zagađenja u njihovoj log fazi rasta. Nakon toga ćelije počinju da flokulišu i formiraju lako taloživu čvrstu materiju. Deo mulja se recirkuliše nazad u proces, a deo se dalje prerađuje stabilizacijom u cilju dobijanja kondicionera za zemljište i proizvodnju metana.

U kapajućim filtrima otpadna voda kaplje preko stenja ili nekog čvrstog materijala prekrivenog mikroorganizmima. Vazduhom se obogaćuje raspršivanjem.

Rotirajući biološki reaktori su plastični diskovi do pola uronjeni u vodu i na površini nose miikroorganizme koji vrše razgradnju zagađenja. Rotiranjem se snabdevaju kiseonikom neophodnim za šroces aerobne respiracije.

Biološki aktivni ugljevi mogu da se koriste i u tretmanu čistih i u tretmanu otpadnih voda. Oni kombinuju procese adsorpcije na aktivnom uglju i mikrobiološke razgradnje organskih materija pomoću biofilma pričvršćenog na površinu aktivnog uglja.

Da bi biomasa rasla u nekom sistemu njeno vreme zadražavanja mora biti dovoljno dugo da se dostigne faza reprodukcije. Dužina neophodnog vremenskog perioda za dostizanje ove faze zavisi od brzine rasta ćelije koja je funkcija metabolizma i iskorišćenja zagađenja.

Brzina rasta bakterijskih ćelija definiše se:

$$r_g = \mu X \tag{5.1}$$

gde su

r_g- brzina rasta bakterija, masa/jedinica zapremine x vreme
μ - specifična brzina rasta, vreme^{-1}
X – komcentracija mikroorganizama, masa/jedinica zapremine

U protočnoj kulturi rast mikroorganizama ograničen je koncentracijom supstrata. On je definisan Monod-ovom jednačinom:

$$\mu = \mu_m \frac{S}{K_S + S} \tag{5.2}$$

gde su

μ - specifična brzina rasta, vreme^{-1}
μ_m - maksimalna specifična brzina rasta, vreme^{-1}
S- koncentracija ograničavajućeg supstrata u rastvoru
K_S-polusaturaciona konstanta, koncentracija supstrata pri kojoj mikroorganizmi dostižu polovinu maksimalne brzine rasta, masa/jedinica zapremine

Kombinacijom jednačina (5.1) i (5.2) dobija se jednačina za brzinu rasta:

$$r_g = \frac{\mu_m XS}{K_S + S} \tag{5.3}$$

Da bi bio precizniji, izraz za brzinu rasta se koriguje za onu količinu energije koja se troši na održavanje ćelije, njenu smrt i predaciju. Smatra se da je smanjenje biomase usled ovih procesa proporcionalno količini biomase, odnosno koncentraciji mikroorganizama. To je endogeni raspad koji se definiše na sledeći način:

$$r_d = -k_d X \tag{5.4}$$

gde je

k_d- koeficijent endogenog raspada, vreme^{-1}

Kombinovanjem sa prethodno navedenim jednačinama dobija se izraz za ukupnu brzinu rasta (masa/jedinici zapremine x vreme):

$$r_g' = \frac{\mu_m XS}{K_S + S} - k_d X \tag{5.5}$$

Uticaj temperature na brzinu bioloških reakcija je vrlo važan jer ne samo metabolička aktivnost nego i transfer gasa, kao i taložne karakteristike se sa temperaturom menjaju:

$$r_T = r_{20}\theta^{(T-20)} \tag{5.6}$$

gde je

r_T-reakciona brzina na T°C
r_{20}- reakciona brzina na 20°C
θ- temperaturni koeficijent
T-temperatura, °C

Procesi sa suspendovanom mikroflorom

Za reaktor sa potpunim mešanjem i odvođenjem viška biomase iz reaktora, važi maseni bilans da je akumulacija biomase jednaka razlici ulazne i izlazne biomase uvećanoj za rast mikroorganizama u elementu zapremine koji se posmatra. Matematički, to glasi:

$$\frac{dX}{dt}V_r = QX_0 - [Q_w X + Q_e X_e] + V_r(r_g') \tag{5.7}$$

gde je

Qw- protok tečnosti kojim se iznosi višak biomase iz reaktora
Qe- protok efluenta na izlazu iz taložnika
Xe-koncentracija mikroorganizama u efluenta iz taložnika

Ako u jednačinu (5.7) uvrstimo jednačinu za brzinu rasta (5.9) uz pretpostavku da je koncentracija ćelija u ulaznoj vodi nula i da važe ravnotežni uslovi (dX/dt=0) dobija se:

$$\frac{Q_W X + Q_e X_e}{V_r X} = -Y\frac{r_{SU}}{X} - k_d \tag{5.8}$$

ili

$$\frac{1}{\theta_c} = -Y\frac{r_{SU}}{X} - k_d \tag{5.9}$$

gde je θc srednje vreme zadržavanja ćelija definisano kao masa mikroorganizama u reaktoru, podeljena sa masom mikroorganizama koja se uklanja iz sistema svakoga dana.

Parametar θc je moguće koristiti kao kontrolni parametar procesa, bez potrebe da se određuje aktivna biomasa ili količina supstrata koja je iskorišćen. To je zasnovano na činjenici da u cilju kontrole rasta mikroorganizama, a time i stabilizacije otpada, određeni procenat ćelijske mase mora biti izveden iz sistema svaki dan. Tako, ako je utvrđeno da parametar ima vrednost 14 dana za postizanje željene stabilizacije, onda svakog dana 1/14 ćelijske mase treba izbaciti iz sistema.

Procesi sa imobilisanom mikroflorom

Biofilm predstavlja čvrstu fazu sa ekstracelularnim polimernim gelom, ćelijama i drugim različitim česticama inkorporiranim u njegovu strukturu. 1-D modeli biofilma odnose se uglavnom na transport mase i biohemijske reakcije, a kasnije su razvijeni stratifikovani modeli sposobni da predstave dinamiku multisupstratnih biofilmova sa više vrsta mikroorganizama. 2-D i 3-D modeli biofilma imaju i morfološki, odn. strukturni elemenat. Velika pažnja poklanja se heterogenosti

biofilma (geometrijska, hemijska, biološka i fizička). Pri modeliranju biofilma neophodno je definisati nekoliko podmodela:

1. model rasta biomase odnosno razgradnje biomase zasnovan na potrošnji nutrijenata
2. model raspodele i širenja biomase koji mogu da obuhvate i proizvodnju i širenje ekstracelularnih polimera
3. model za transport supstrata i kinetiku odn. ravnotežu reakcija
4. model za odvajanje biofilma
5. model protoka tečnosti
6. model pričvršćivanja biofilma

Jedan od osnovnih problema pri modeliranju jeste prilagoditi istom modelu sve relevantne, kako spore (rast biofilma, raspad, odvajanje), tako i brze procese (difuzija, reakcije).

Postoje dve grupe modela: modeli zasnovani na aktivnostima i osobinama individualnih ćelija (IbM, eng. individual based modeling) i modeli zasnovani na biomasi kao multifaznom sistemu (BbM, eng. biomass-based models).

Što se tiče rasta biomase, ćelija se uvećava apsorpcijom nutrijenata i kada dostigne kritičnu masu deli se u dve ćelije. Uopštena jednačina za brzinu koja opisuje promenu mase *m* bakterije *i* smeštene u momentu *t* na mestu *x [x y z]* može se napisati:

$$\frac{dm_i}{dt} = r_{Xj}(m_i(t), C_S(x,t), \ldots \tag{5-10}$$

Gde je Cs veličina koja sadrži sve koncentracije substrata i proizvoda koji mogu uticati na rast bakterija.

Jednostavna Monodova kinetika je u najvećem broju slučajeva sasvim prihvatljiva. Ponekad je model neophodno komplikovati inhibicijom supstrat-proizvod, zahtevima za održavanjem i raspadom mase.

Najjednostavniji modeli za širenje novonastalog biofilma slični su modelima rasta kristala.

Osnovni procesi koji doprinose povećanju zapremine biofilma su određeni dostupnošću nutrijenata. Procesi raspadanja su takođe određeni koncentracionim nivoima pojedinih supstanci. Transport i reakcije supstrata su opisani dobro poznatim zakonima fizike. Rastvorene supstance mogu biti transportovane molekulskom difuzijom i konvekcijom (ponekad se koristi i termin advekcija). Difuzija je određena koncentracionim gradijentom i opisana Fikovim zakonom. Konstanta proporcionalnosti fluksa mase koncentracionom gradijentu zove se difuzioni koeficijent Di (m^2/h). Gradijenti koncentracije nastaju usled potrošnje supstrata i formiranja proizvoda u biofilmu. Konvektivni fluks rastvorka u bilo kojoj tački u prostoru proporcionalan je brzini tečnosti (m^2/s), koncentraciji transportovane supstance i $C_{Si.}$ Prostorna raspodela koncentracije svake relevantne hemijske vrste računa se na osnovu sistema dinamičke materijalne ravnoteže. To znači da brzina akumulacije supstance *i* u elementu zapremine mora biti u ravnoteži sa brzinom transporta u okviru granica date zapremine i ukupnom brzinom transformacije u datoj

zapremini (odn. hemijske reakcije R_i). Tako, materijalni bilans za hemijsku vrstu *i* u elementu zapremine biofilma matematički glasi:

$$\frac{\partial C_{Si}}{\partial t} = D_i \nabla^2 C_{Si} - u\nabla_{Si} + R_i(Cs, Cx) \tag{5.11}$$

Rast biofilma na biološki aktivnom kapajućem filtru dat je jednačinom:

$$r_S = f_0 h k_0 S^2 / (K_m + S) \tag{5-12}$$

Gde je

r_s brzina rasta u tankom sloju,
h- dubina sloja,
k_0- maksimalna brzina uklanjanja supstrata,
S-srednja koncentracija supstrata,
K_m- konstanta polubrzine i
f_0-faktor

maseni bilans za uklanjenje supstrata definisan je jednačinom :

$$(\partial S / \partial t) dV = QS - Q(S + (\partial S / \partial z) dz) + dzW(-f_0 h k_0 S^2 / K_m + S) \tag{5-13}$$

Q-protok vode,
W-širina površine na kojoj raste biofilm i
Z-visina adsorbera.

Kako u ravnotežnom stanju vredi $\partial S/\partial t=0$ gore pomenuti maseni bilans se pojednostavljuje

$$Q(dS/dZ) = -f_0 k_0 h W S^2/(K_m+S) \tag{5-14}$$

A kako je $K_m \langle$ S možemo pisati

$$dS/dZ = -f_0 k_0 h W S/Q \tag{5-15}$$

Intergaljenjem od S_e do S_0 i od 0 do Z dobijamo:

$$S_e / S_0 = e^{-(f_0 h k_0) WZ / Q} \tag{5-16}$$

Gde je S_e koncentracija supstrata u efluentu,
S_0 koncentracija supstrata u ulaznoj vodi

Dalje vredi

$$WZ/Q=A_S/Q=ZA/QxA_S/V=S_aZA/Q \quad (5\text{-}17)$$

gde je

A_S površina punjenja filtra,
V-zapremina biofiltra,
S_a- specifična površina punjenja,
A –poprečni presek

Proizvod f_0hk_0 možemo zameniti jednom konstantom uklanjanja supstrata, k_s, pa se može pisati

$$S_e / S_0 = e^{-k_S S_a ZA/Q} \quad (5\text{-}18)$$

Literatura:

- Picioreanu C. And M.C.M. van Loosdrecht (2003) Use of mathematical modelling to study biofilm development and morphology in Biofilms in Medicine, Industry and Environmental Biotechnology-Characteristics, Analysis and Control (Ed. Lens, P., Moran, A.P. Mahony T., Stoodley, P., O'Flaherry, V.) IWA Publishing
- Metcalf & Eddy, Inc (1991) Biological Unit Processes in Wastewater Engineering: Treatment, Disposal and Reuse (Eds. Tchobanoglous G. and Burton F. L) McGraw-Hill.