Content

Problems in Mathematics
&
Experiments with Mathematica

9. Functions of several variables

9.4 Curve-fitting by the least square method

Theory

Let the data set $\{(x_i, y_i) : i = 1, 2, \ldots n\}$ be given. We look for a function $h_{\alpha_0}(x)$ fitting the data set from the given family of functions $\boldsymbol{H} = \{\boldsymbol{h_\alpha(x)}\}$ for which the sum of the square of the differences is a minimum:

$$S(\alpha) = \sum_{i=1}^{n} (y_i - h_\alpha(x_i))^2$$

The method is called curve fitting by least squares.

Mathematica summary

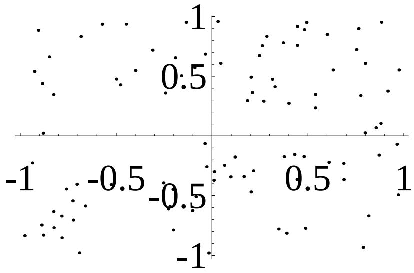
ReadList[file, RecordSeparators –> True]	Reads data from file into list
ListPlot[data]	Plots the data list
LogListPlot[data]	Logarithmic plot (Graphics`Graphics` package)
LogLogListPlot[data]	Double logarithmic plot (Graphics`Graphics` package)
Fit[data, {f_1, f_2, ...}, var]	Curve fitting with the linear combination of f_1, f_2 ...

Graphical exercises to recognize the shape of data sets

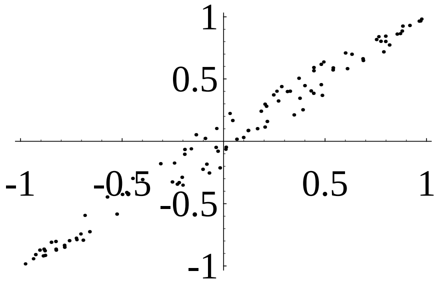
PROBLEM 9.4.1

Examine the following point sets. Is there any relation between the coordinates? Try to characterize the fitting functions.

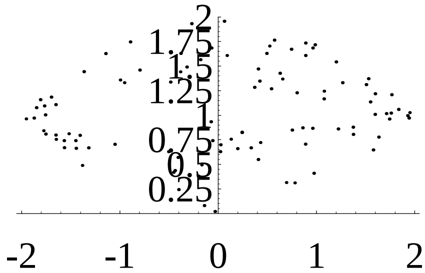
(1)


1
0.5
-1
-0.5
0.5
1
-0.5
-1

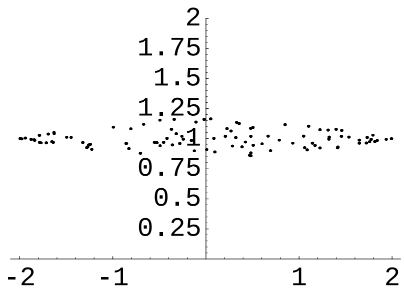
(2)


1
0.5
-1
-0.5
0.5
1
-0.5
-1

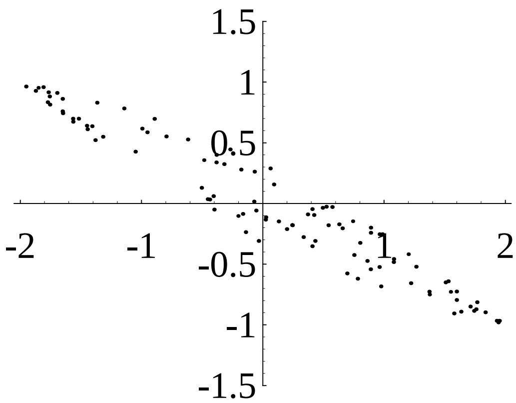
(3)


2
1.75
1.5
1.25
1
0.75
0.5
0.25
-2
-1
0
1
2

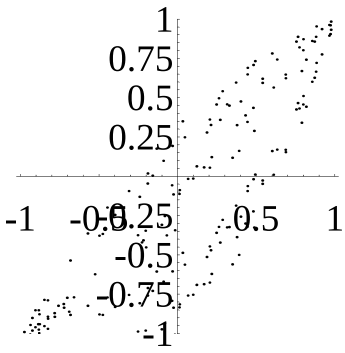
(4)


2
1.75
1.5
1.25
1
0.75
0.5
0.25
-2
-1
1
2

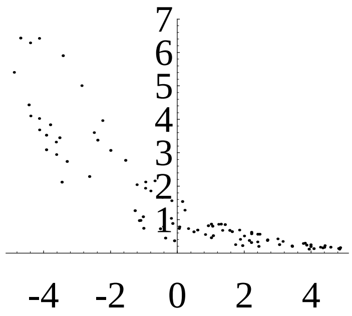
(5)


1.5
1
0.5
-2
-1
1
2
-0.5
-1
-1.5

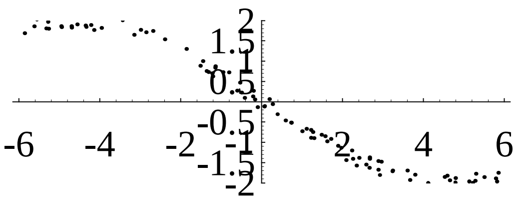
(6)



(7)



(8)



EXPERIMENT 9.4.2

Generate new data sets using the following statements:

```
data = Table[{i / 100, Random[Real, {-1, 1}]}, {i, 0, 150}];
newdata1 = data /. {x_, y_} → {12 x, 2 7^(x - 0.5) + y/5};
newdata2 = data /. {x_, y_} → {12 x, 2 5^(x - 0.5) + y/3};
ListPlot[Join[newdata1, newdata2],
  PlotRange -> {0, 10}, AspectRatio -> Automatic,
  PlotStyle -> {PointSize[0.012]}];
```

Simple fitting

The measured data are given in the list variable `data`. The family $\boldsymbol{H = \{h_\alpha(x)\}}$ is given as a linear combination:

$$h_\alpha(x) = \sum_{i=0}^{n} a_i\, f_i(x)$$

■ *Mathematica initialization*

The function that calculates the sum of squares.

```
S[ff_, data_List] :=
 Plus @@ ((((ff /. x → #1[[1]]) - #1[[2]])^2 &) /@ #1 &)[data]
```

SOLVED PROBLEM 9.4.3

Find good fitting functions to the following given data set. Check the goodness of the fitting by calculating the sum of the squares (the function `S[ff,data]`). For more precise statistical evaluation, see the *Mathematica* package `Statistics`.

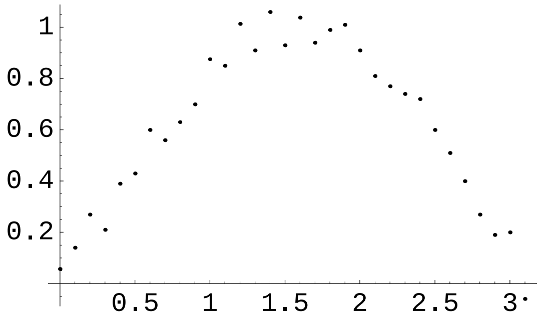
- *The data set*

```
data=
{{0, 0.057}, {0.1, 0.14}, {0.2, 0.27},{0.3, 0.21},
{0.4, 0.39},{0.5, 0.43},{0.6, 0.60}, {0.7, 0.56},
{0.8, 0.63},{0.9, 0.70}, {1., 0.875}, {1.1,0.85},
{1.2, 1.014}, {1.3, 0.91}, {1.4, 1.06},{1.5,0.93},
{1.6, 1.04}, {1.7, 0.94}, {1.8, 0.99},{1.9, 1.01},
{2., 0.91},{2.1, 0.81}, {2.2, 0.77}, {2.3, 0.74},
{2.4, 0.72}, {2.5, 0.60}, {2.6, 0.51},{2.7, 0.40},
{2.8, 0.27}, {2.9, 0.19},{3.0, 0.2}, {3.1, -0.06}};
```

○ SOLUTION

- *Plot the data*

```
ListPlot[data];
```



- *Find the fitting family of functions and perform the fitting*

```
fitf = Fit[data, {1, x, x^2, x^3}, x]
```

$0.00417981 + 1.1004\,x - 0.255144\,x^2 - 0.0344448\,x^3$

- *Check the goodness and try to improve*

```
S[fitf,data]
```

0.107834

○

PROBLEM 9.4.4 Further data sets (in the electronic version)

Generalizations: fitting with transformations

Now, we consider fitting problems in which the data set cannot be fit by a linear combination of the functions in ***H***, but a simple transformation reduces the problem to this case.
Next, we summarize the most important transformations.

• Logarithmic transformations

Let (x,y) be a point in the plane, where $y>0$. The transformation

$$(x,y)\to(x,z):=(x,\log_a y)$$

is called the logarithmic transformation. The point (x,z) can be plotted in the *xz*-system. This gives the same result, as if the point is plotted in a logarithmic coordinate system (see the chapter about the logarithmic plots). In particular, if $y_i \approx A\, a^{b\, x_i}$, then using the logarithmic transformation, the transformed data set is $z_i = b\, x_i + \log_a A$, i.e., the graph is a straight line in the *xz*-system .

• Double logarithmic transformation

Let (x,y) be a point in the plane, where y>0. The transformation

$$(x,y)\to(u,z):=(\log_a x, \log_a y)$$

is called the double logarithmic transformation. The point (u,z) can be plotted in the *uz*-plane. This gives the same result, as if the point is plotted in the double logarithmic coordinate system. In particular, if $y_i \approx A\, {x_i}^b$, then using this transformation, the transformed data set is $z_i = b \log_a x_i + \log_a A = b\, u_i + \log_a A$, i.e., it is a straight line in the *uz*-plane.

• Exponential transformation

Let (x,y) be a point in the plane. The transformation

$$(x,y)\to(x,z):=(x, a^y)$$

is called the exponential transformation. The point (u,z) can be plotted in the *uz*-system. In particular, if $y_i \approx \log_a x_i + A$, then using the transformation, the transformed data set is $z_i = a^{y_i} = a^A\, x_i$, that is a straight line in the *uz*-plane.

■ *Mathematica initialization*

```
<< Graphics`Graphics`
```

```
LogData[data_List] := data /. {x_, y_} → {x, Log[10, y]};
LogLogData[data_List] :=
  data /. {x_, y_} → {Log[10, (x)], Log[10, (y)]};
ExpData[data_List] := data /. {x_, y_} → {x, 10^y};
```

Solved Problem 9.4.5

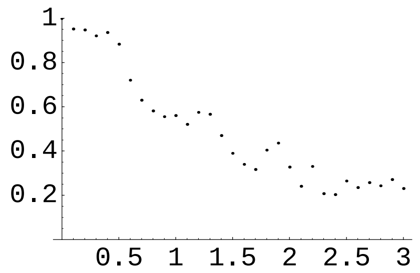
Fit an apppropriate function to the following data set.

```
data = {{0, 1.}, {0.1, 0.951}, {0.2, 0.947}, {0.3, 0.920},
   {0.4, 0.935}, {0.5, 0.883}, {0.6, 0.719}, {0.7, 0.630},
   {0.8, 0.581}, {0.9, 0.555}, {1., 0.560}, {1.1, 0.521},
   {1.199, 0.574}, {1.3, 0.566}, {1.399, 0.470},
   {1.5, 0.389}, {1.6, 0.339}, {1.7, 0.316}, {1.8, 0.403},
   {1.90, 0.435}, {2., 0.327}, {2.1, 0.240}, {2.2, 0.330},
   {2.3, 0.208}, {2.4, 0.203}, {2.5, 0.264}, {2.6, 0.257},
   {2.7, 0.257}, {2.8, 0.242}, {2.9, 0.271}, {3., 0.23}};
```

○ Solution

- *Plot the data*

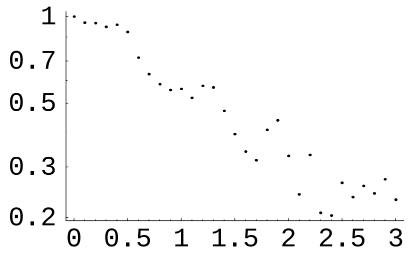
```
pldata=ListPlot[data, PlotRange->{0,1}];
```



- *Graphical investigation of the data*

The dataset is possibly exponential. Check it by using `LogListPlot`. See the section Logarithmic transformations.

```
LogListPlot[data];
```



Really, the logarithmic figure looks linear, the original one is exponential.

- *The transformation*

Perform logarithmic transformation on the original data set

```
logdata=LogData[data];
ListPlot[logdata];
```

- *The fitting*

Fit a straight line to the transformed data set.

```
logfitf=Fit[logdata,{1,x},x]
```

```
-0.00209232 - 0.236231 x
```

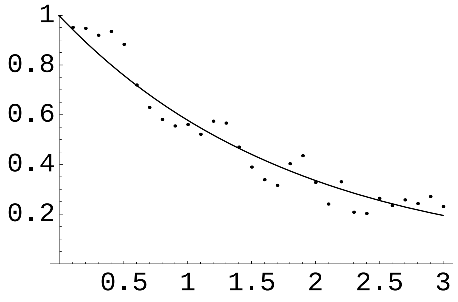
- *The fitting function to the original data set*

```
fitf = 10^logfitf
```

$10^{-0.00209232-0.236231\,x}$

- *The graph of the fitting function*

```
plfit=Plot[fitf,{x,0,3}, DisplayFunction->Identity];
Show[plfit,pldata, DisplayFunction->$DisplayFunction,
PlotRange->{0,1}];
```



PROBLEM 9.4.6 Further data sets (in the electronic version)

EXPERIMENT 9.4.7

Generate new data sets using the following statements. Try to apply new transformations.

```
data = Table[{x, N[log (5 x) + Random[Real, {0, 0.15}], 2]},
  {x, 1, 4, 0.1}]
```

Generalization: nonlinear fitting

Here, we consider problems that cannot be reduced to the previous cases. For example, such problems arise for logistic flows (see the chapter on [logistic differential equations]). In some cases, linearization causes some statistical problems, that also explains the using of the more complicated nonlinear methods.

■ *Mathematica initialization*

```
<< "Statistics`NonlinearFit`"
```

SOLVED PROBLEM 9.4.8 A logistic data set

Fit an apppropriate function to the following data set.

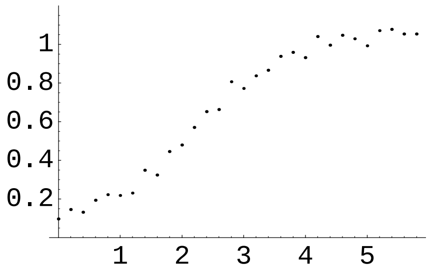
```
data = {{0, 0.096}, {0.2, 0.145}, {0.4, 0.132}, {0.6, 0.193},
    {0.8, 0.223}, {1., 0.218}, {1.2, 0.230}, {1.4, 0.349},
    {1.6, 0.324}, {1.8, 0.445}, {2., 0.479}, {2.2, 0.57},
    {2.4, 0.652}, {2.6, 0.663}, {2.8, 0.806}, {3., 0.772},
    {3.2, 0.837}, {3.4, 0.866}, {3.6, 0.938}, {3.8, 0.958},
    {4., 0.931}, {4.2, 1.04}, {4.4, 0.995}, {4.6, 1.047},
    {4.8, 1.029}, {5., 0.993}, {5.2, 1.071}, {5.4, 1.078}};
```

○ SOLUTION

- *The data were generated by*

- *Plot the data*

```
pldata=ListPlot[data, PlotRange->{0,1.2}];
```



- *Find the model*

The data likely describe a logistic growth. Such curves are solutions of a differential equation of form

$$x' = k\, x^{\alpha}\,(A - x^{\beta})$$

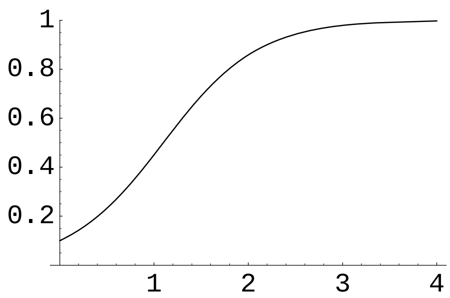
(see the chapter on logistic differential equations). The solutions of the simplest logistic equation $x' = k\,x\,(A - x)$ are of form

$$\text{model} = \frac{A\, e^{k A t}}{\frac{A - x_0}{x_0} + e^{k A t}};$$

where x_0 is the initial value of the process at t=0.

- *Check the shape of the model:*

```
Plot[
  Evaluate[model /. {A -> 1, k -> 2, x0 -> 0.1}], {t, 0, 4},
    PlotRange -> {0, 1}];
```



- *The fitting*

```
fitf = NonlinearFit[data, model,
  {t}, {{k, 0.5}, {A, 0.5}, {x0, 0.5}},
  AccuracyGoal → 15, ShowProgress → True]
```

Iteration:1 ChiSquared:4.81576 Parameters:{0.5, 0.5, 0.5}

Iteration:2 ChiSquared:1.36286 Parameters:{0.5, 0.981934, 0.2356}

Iteration:3 ChiSquared:0.265892 Parameters:{0.996184, 0.989087, 0.104634}

Iteration:4 ChiSquared:0.0260442 Parameters:{1.08874, 1.07489, 0.085055}

Iteration:5 ChiSquared:0.0258225 Parameters:{1.07148, 1.08077, 0.0857481}

Iteration:6 ChiSquared:0.0258219 Parameters:{1.07206, 1.08071, 0.0857186}

Iteration:7 ChiSquared:0.0258219 Parameters:{1.07205, 1.08072, 0.0857186}

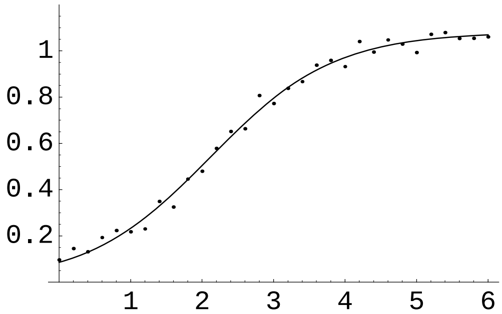
Iteration:8 ChiSquared:0.0258219 Parameters:{1.07205, 1.08072, 0.0857186}

$$\frac{1.08072\, e^{1.15858\, t}}{11.6077 + e^{1.15858\, t}}$$

Compare the results with the parameters used in generating the data. Explain the difference. Pay attention to the random part of the function.

- *The graph of the fitting function*

```
plfit = Plot[fitf, {t, 0, 6}, DisplayFunction → Identity];
Show[plfit, pldata, DisplayFunction → $DisplayFunction,
  PlotRange → {0, 1.2}];
```



PROBLEM 9.4.9 Further data sets (in the electronic version)

EXPERIMENT 9.4.10

Generate new data sets using the following statements. Try to apply new models.

```
data =
    Table[{x, N[Log[5 x] + Random[Real, {0, 0.15}], 2]},
  {x, 1, 4, 0.1}]
```