The Method of Least Squares

The method of least squares

We study the following problem:
Given n points (t_i, y_i) for $i = 1, \ldots, n$ in the (t, y)-plane. How can we determine a function $p(t)$ such that

$$p(t_i) \approx y_i, \quad \text{for} \quad i = 1, \ldots, n? \tag{1}$$

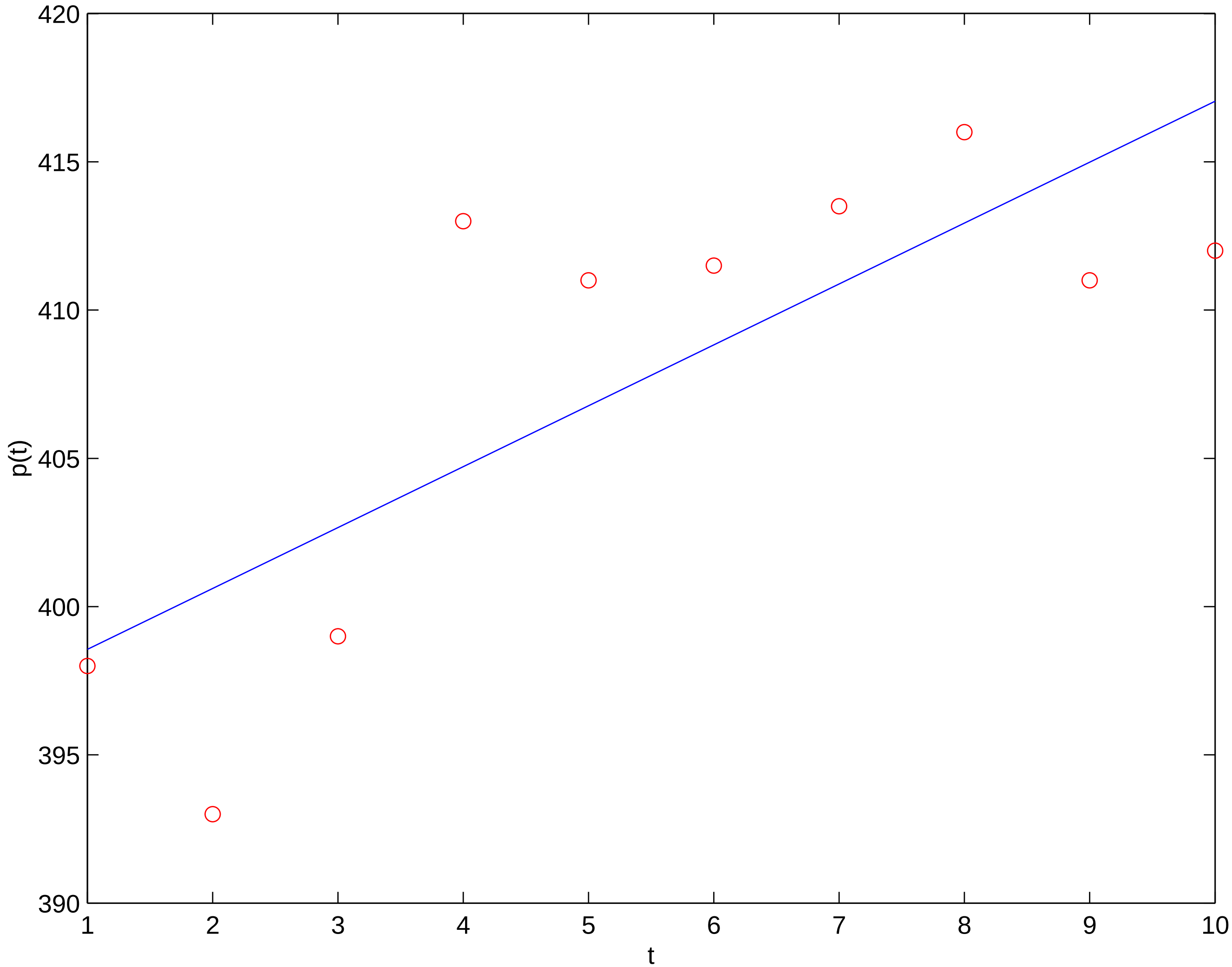

Figure 1: A set of discrete data marked by small circles is approximated with a linear function $p = p(t)$ represented by the solid line.

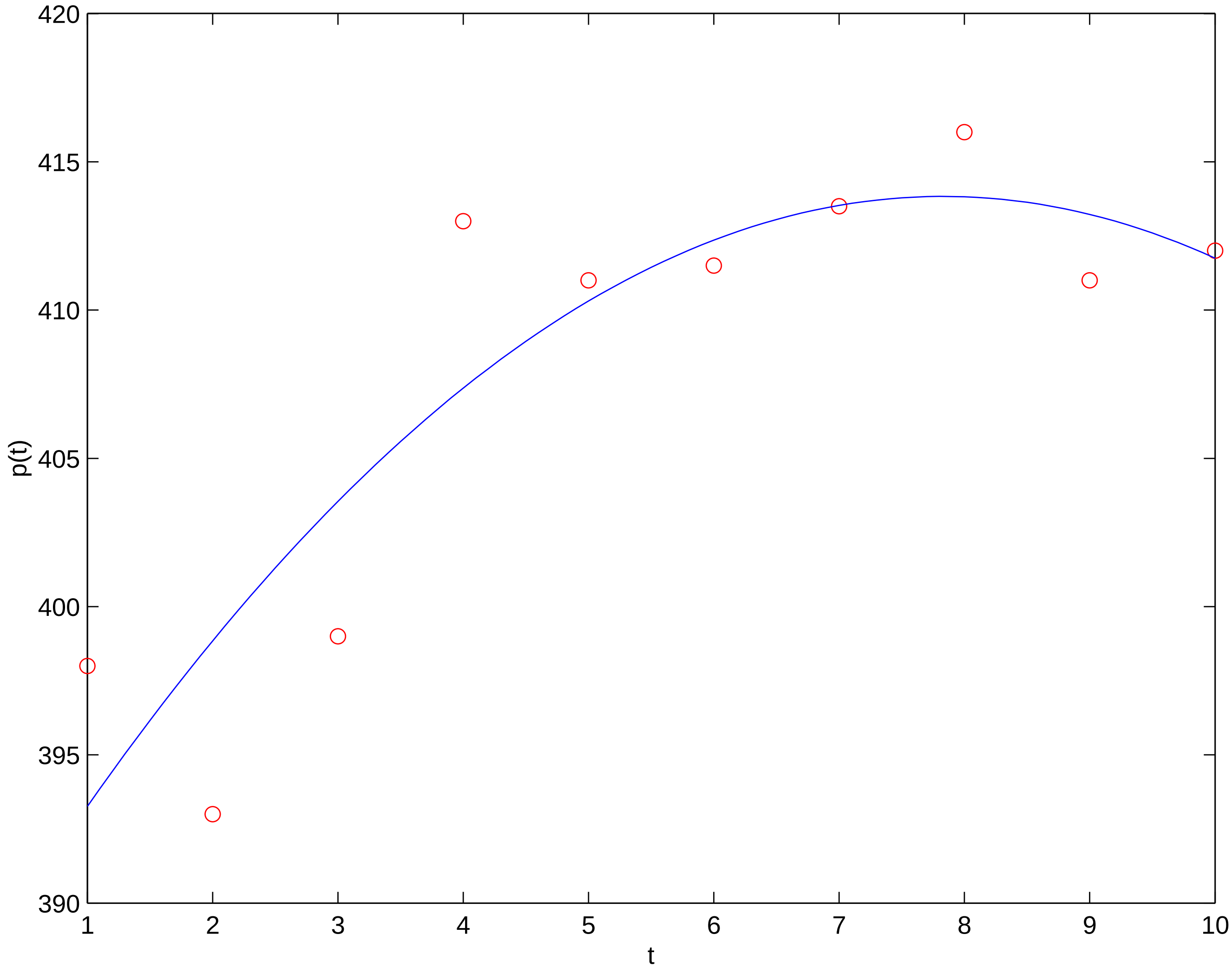

Figure 2: A set of discrete data marked by small circles is approximated with a quadratic function $p = p(t)$ represented by the solid curve.

The method of least square

- Above we saw a discrete data set being approximated by a continuous function
- We can also approximate continuous functions by simpler functions, see Figure 3 and Figure 4

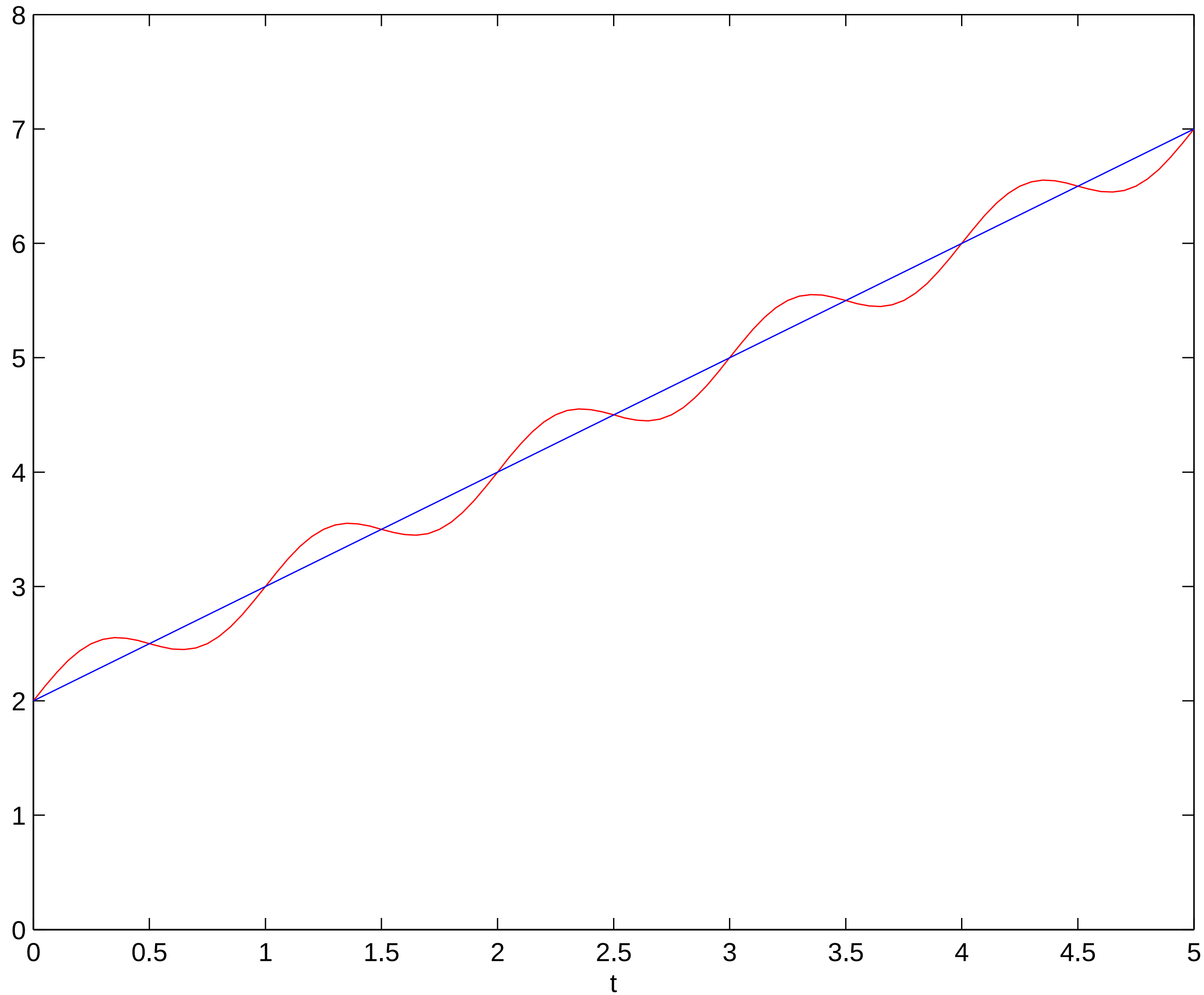

Figure 3: A function $y = y(t)$ and a linear approximation $p = p(t)$.

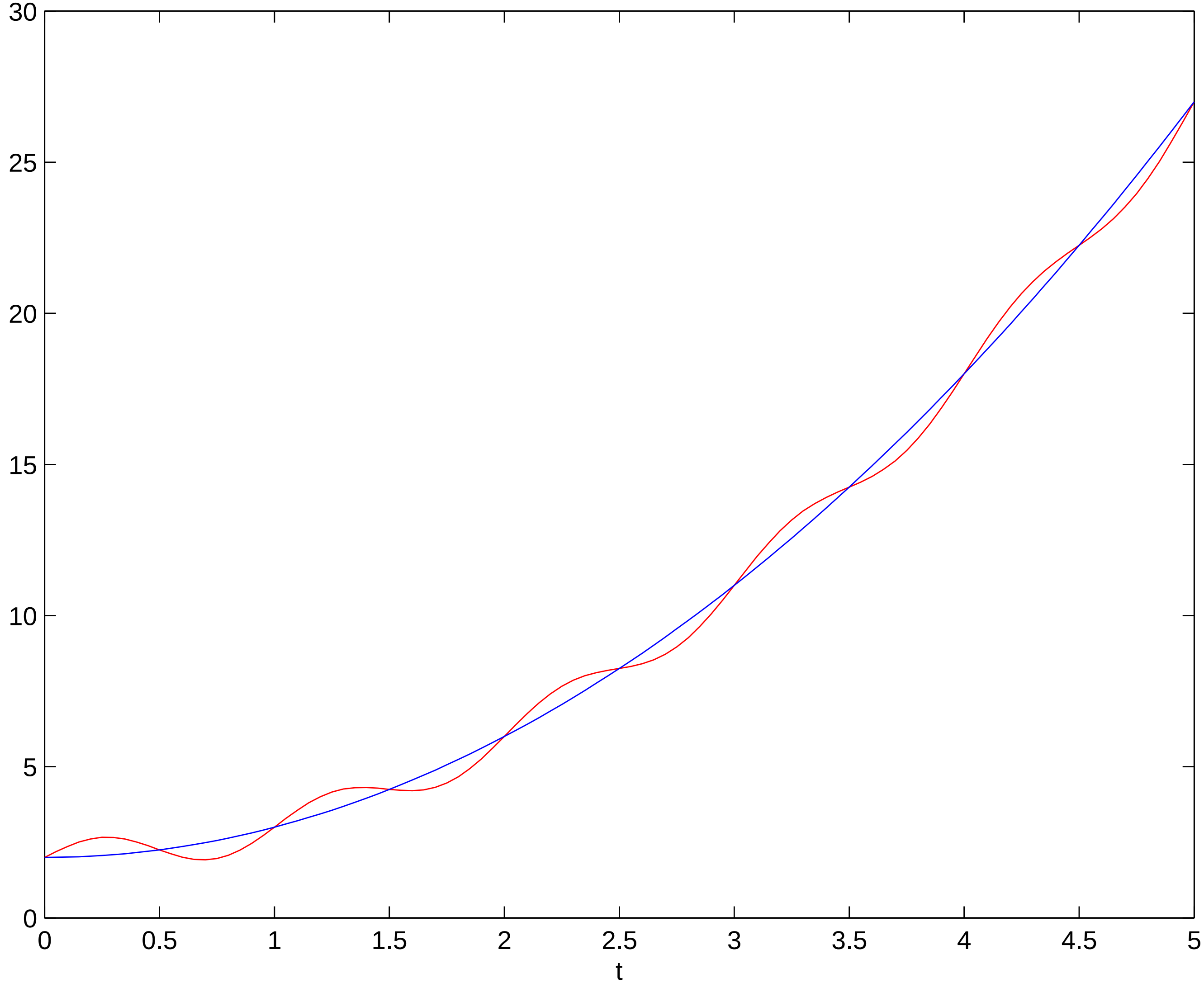

Figure 4: A function $y = y(t)$ and a quadratic approximation $p = p(t)$.

World mean temperature deviations

Calendar year	Computational year t_i	Temperature deviation y_i
1991	1	0.29
1992	2	0.14
1993	3	0.19
1994	4	0.26
1995	5	0.28
1996	6	0.22
1997	7	0.43
1998	8	0.59
1999	9	0.33
2000	10	0.29

Table 1: The global annual mean temperature deviation measured in $^{\circ}C$ for years 1991-2000.

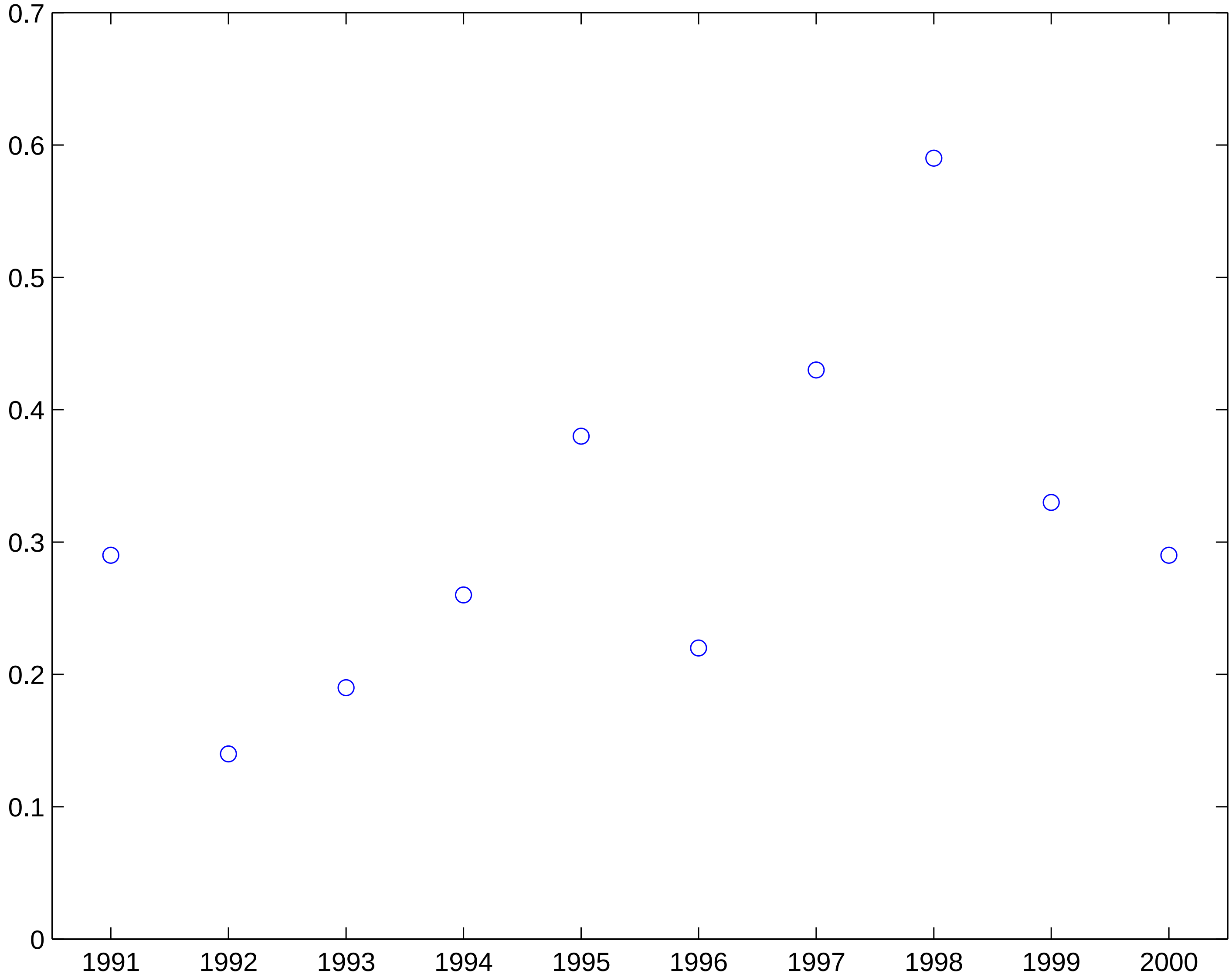

Figure 5: The global annual mean temperature deviation measurements for the period 1991-2000.

Approximating by a constant

- We will study how this set of data can be approximated by simple functions
- First, how can this data set be approximated by a constant function

$$p(t) = \alpha?$$

- The most obvious guess would be to choose α as the arithmetic average

$$\alpha = \frac{1}{10}\sum_{i=1}^{10} y_i = 0.312 \qquad (2)$$

- We will study this guess in more detail

Approximating by a constant

- Assume that we want the solution to minimize the function

$$F(\alpha) = \sum_{i=1}^{10} (\alpha - y_i)^2 \qquad (3)$$

- The function F measures a sort of deviation from α to the set of data $(t_i, y_i)_{i=1}^{10}$
- We want to find the α that minimizes $F(\alpha)$, i.e. we want to find α such that $F'(\alpha) = 0$
- We have

$$F'(\alpha) = 2\sum_{i=1}^{10} (\alpha - y_i) \qquad (4)$$

Approximating by a constant

- This leads to

$$2\sum_{i=1}^{10}\alpha^* = 2\sum_{i=1}^{10}y_i, \tag{5}$$

or

$$\alpha^* = \frac{1}{10}\sum_{i=1}^{10}y_i, \tag{6}$$

which is the arithmetic average

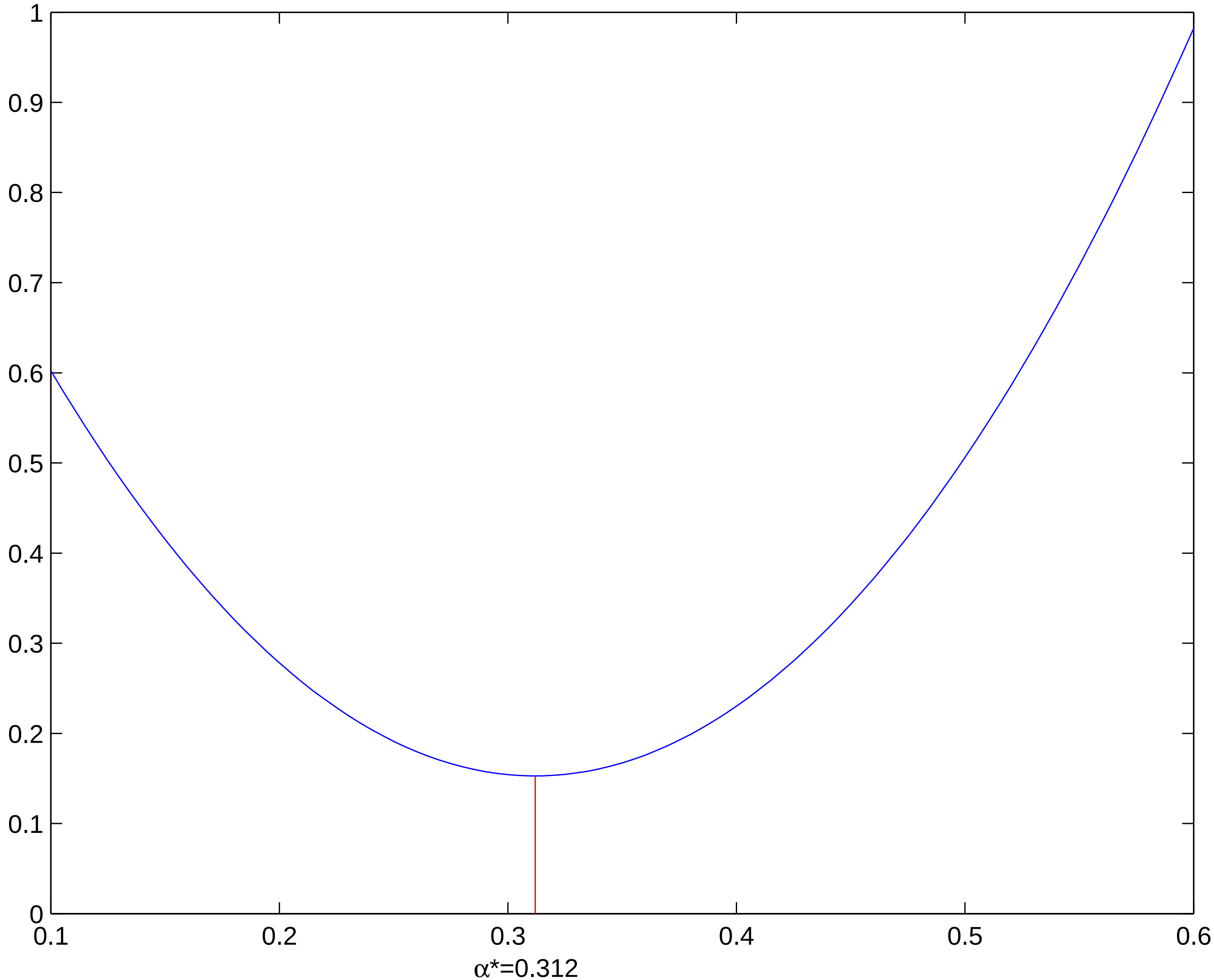

Figure 6: A graph of $F = F(\alpha)$ given by (3).

Approximating by a constant

- Since

$$F''(\alpha) \;=\; 2\sum_{i=1}^{10} 1 = 20 > 0, \qquad (7)$$

  it follows that the arithmetic average is the minimizer for F

- We can say that the average value is the optimal constant approximating the global temperature
- This way of defining an optimal constant, where we minimize the sum of the square of the distances between the approximation and the data, is referred to as *the method of least squares*
- There are other ways to define an optimal constant

Approximating by a constant

- Define

$$G(\alpha) \;=\; \sum_{i=1}^{10} (\alpha - y_i)^4 \tag{8}$$

- $G(\alpha)$ also measures a sort of deviation from α to the data
- We have that

$$G'(\alpha) \;=\; 4\sum_{i=1}^{10} (\alpha - y_i)^3 \tag{9}$$

- And in order to minimize G we need to solve $G'(\alpha) = 0$, (and check that $G''(\alpha) > 0$)

Approximating by a constant

- Solving $G'(\alpha) = 0$ leads to a nonlinear equation that can be solved with the Newton iteration from the previous lecture
- We use Newton's method with
  - initial approximation: $\alpha_0 = 0.312$
  - tolerance specified by: $\varepsilon = 10^{-8}$

  This gives $\alpha^* \approx 0.345$, in three iterations
- α^* is a minimum of G since

$$G''(\alpha^*) \;=\; 12\sum_{i=1}^{10}(\alpha^* - y_i)^2 > 0$$

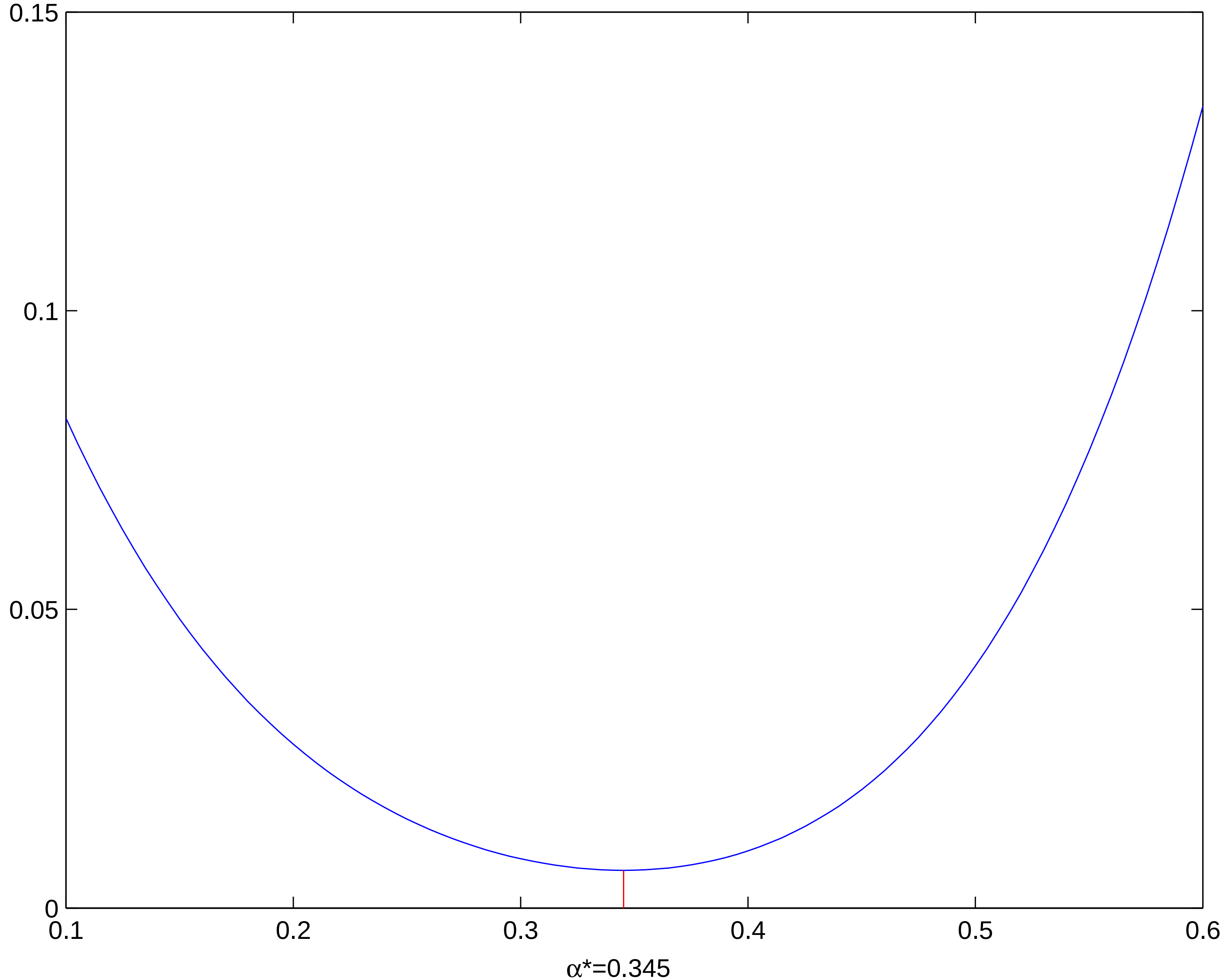

Figure 7: A graph of $G = G(\alpha)$ given by (8).

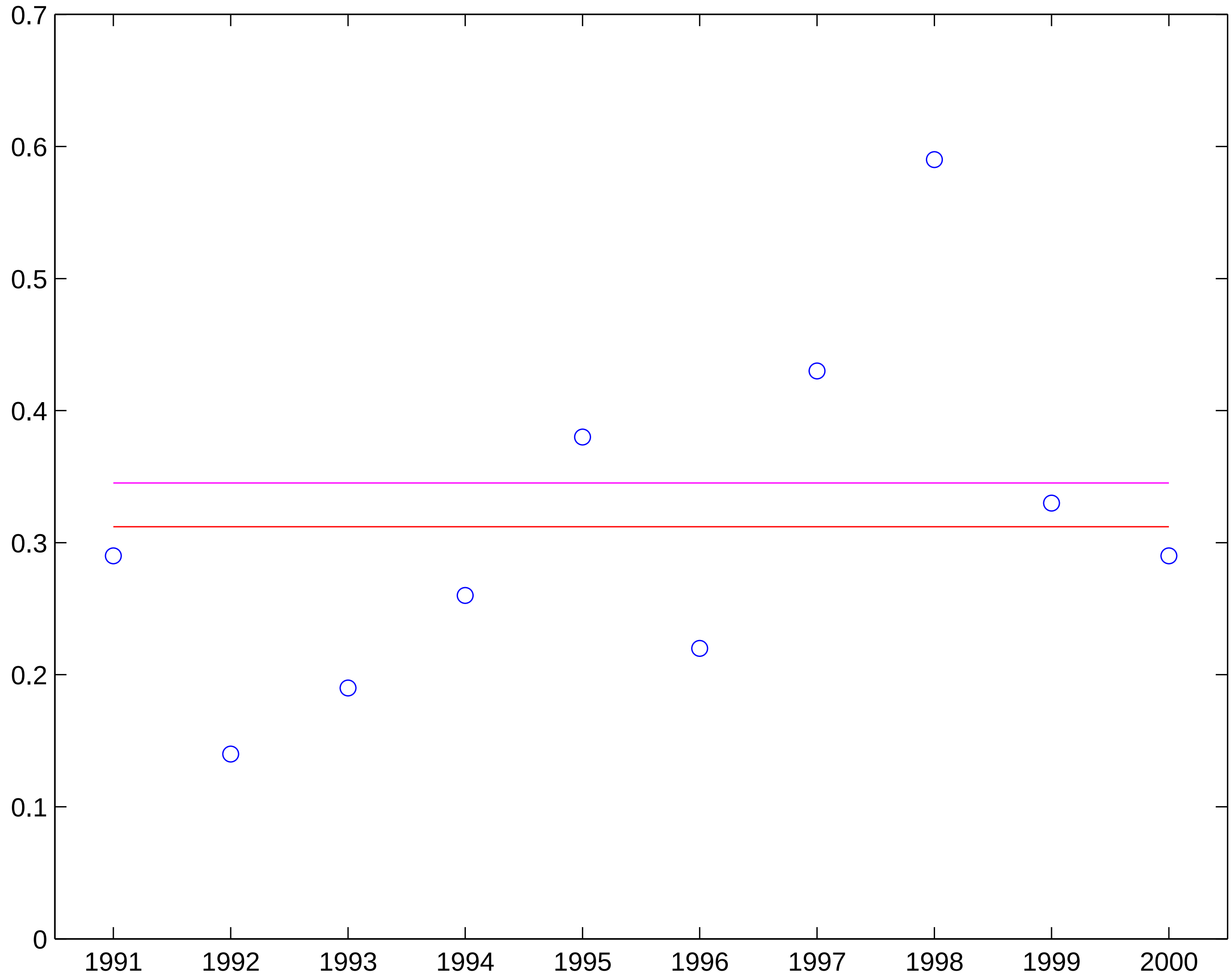

Figure 8: Two constant approximations of the global annual mean temperature deviation measurements from year 1991 to 2000.

Approximating by a linear function

- Now we will study how we can approximate the world mean temperature deviation with a linear function
- We want to determine two constants α and β such that

$$p(t) = \alpha + \beta t \tag{10}$$

  fits the data as good as possible in the sense of least squares

Approximating by a linear function

- Define

$$F(\alpha,\beta) \;=\; \sum_{i=1}^{10}(\alpha+\beta t_i - y_i)^2 \qquad (11)$$

- In order to minimize F with respect to α and β, we can solve

$$\frac{\partial F}{\partial \alpha} \;=\; \frac{\partial F}{\partial \beta} = 0 \qquad (12)$$

Approximating by a linear function

We have that

$$\frac{\partial F}{\partial \alpha} = 2\sum_{i=1}^{10}(\alpha + \beta t_i - y_i), \tag{13}$$

and therefore the condition $\frac{\partial F}{\partial \alpha} = 0$ leads to

$$10\alpha + \left(\sum_{i=1}^{10} t_i\right)\beta = \sum_{i=1}^{10} y_i. \tag{14}$$

Approximating by a linear function

Here

$$\sum_{i=1}^{10} t_i \;=\; 1+2+3+\cdots+10 \;=\; 55,$$

and

$$\sum_{i=1}^{10} y_i \;=\; 0.29+0.14+0.19+\cdots+0.29 \;=\; 3.12,$$

so we have

$$10\alpha+55\beta \;=3.12. \tag{15}$$

Approximating by a linear function

Further, we have that

$$\frac{\partial F}{\partial \beta} \;=\; 2\sum_{i=1}^{10}(\alpha+\beta t_i - y_i)t_i,$$

and therefore the condition $\frac{\partial F}{\partial \beta} = 0$ gives

$$\left(\sum_{i=1}^{10} t_i\right)\alpha + \left(\sum_{i=1}^{10} t_i^2\right)\beta \;=\; \sum_{i=1}^{10} y_i t_i.$$

Approximating by a linear function

We can calculate

$$\sum_{i=1}^{10} t_i^2 \;=\; 1+2^2+3^2+\cdots+10^2 = 385,$$

and

$$\sum_{i=1}^{10} t_i y_i \;=\; 1\cdot 0.29+2\cdot 0.14+3\cdot 0.19+\cdots+10\cdot 0.29 = 20,$$

so we arrive at the equation

$$55\alpha+385\beta \;=20. \tag{16}$$

Approximating by a linear function

We now have a 2×2 system of linear equations which determines α and β:

$$\begin{pmatrix} 10 & 55 \\ 55 & 385 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 3.12 \\ 20 \end{pmatrix}.$$

With our knowledge of linear algebra, we see that

$$\begin{aligned} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} &= \begin{pmatrix} 10 & 55 \\ 55 & 385 \end{pmatrix}^{-1} \begin{pmatrix} 3.12 \\ 20 \end{pmatrix} \\ &= \frac{1}{825} \begin{pmatrix} 385 & -55 \\ -55 & 10 \end{pmatrix} \begin{pmatrix} 3.12 \\ 20 \end{pmatrix} \approx \begin{pmatrix} 0.123 \\ 0.034 \end{pmatrix}. \end{aligned}$$

Approximating by a linear function

We conclude that the linear model

$$p(t) = 0.123 + 0.034t \tag{17}$$

approximates the data optimally in the sense of least squares.

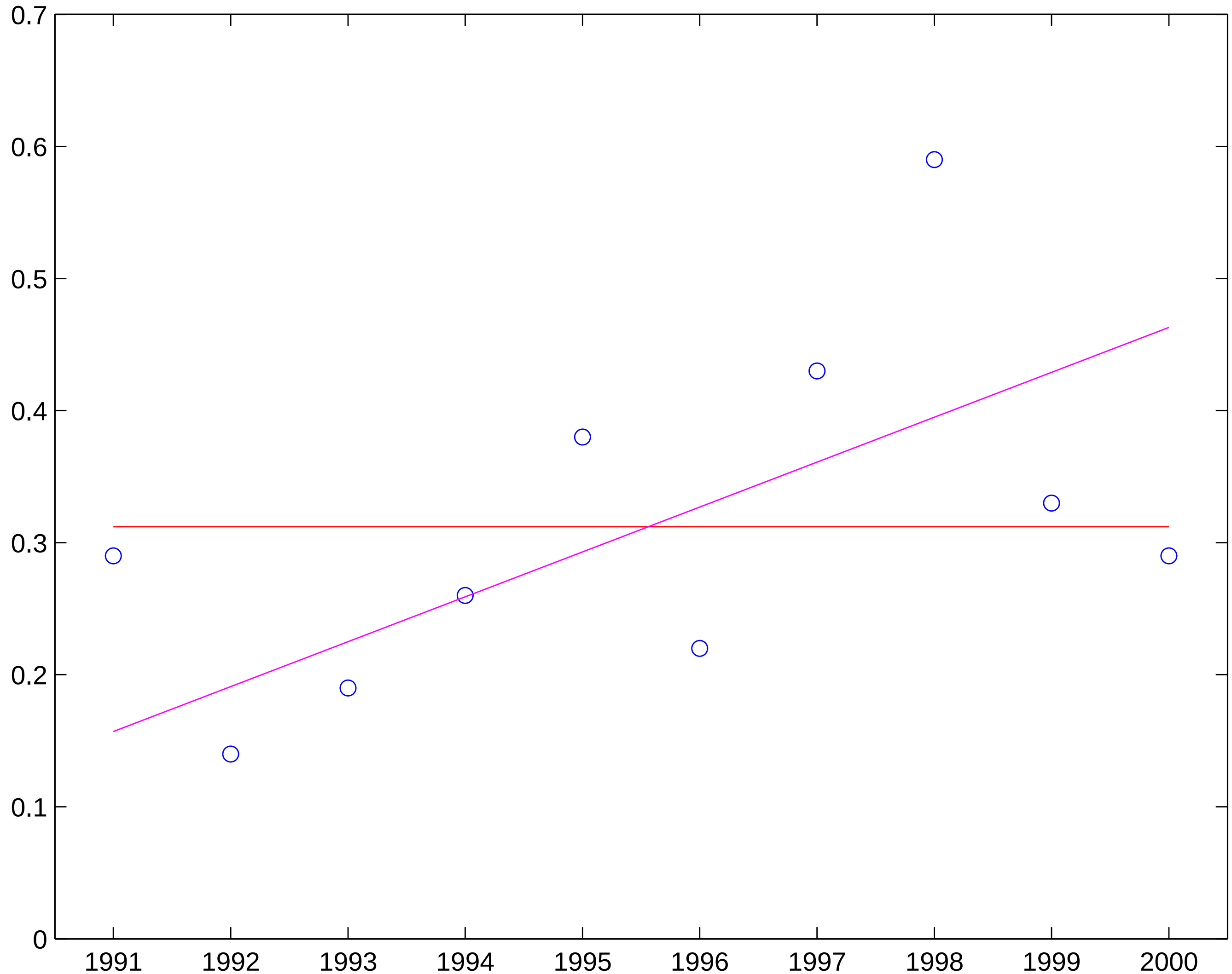

Figure 9: Constant and linear least squares approximations of the global annual mean temperature deviation measurements from year 1991 to 2000.

Approx. by a quadratic function

- We now want to determine constants α, β and γ, such that the quadratic polynomial

$$p(t) = \alpha + \beta t + \gamma t^2 \tag{18}$$

fits the data optimally in the sense of least squares

- Minimizing

$$F(\alpha, \beta, \gamma) = \sum_{i=1}^{10} (\alpha + \beta t_i + \gamma t_i^2 - y_i)^2 \tag{19}$$

requires

$$\frac{\partial F}{\partial \alpha} = \frac{\partial F}{\partial \beta} = \frac{\partial F}{\partial \gamma} = 0 \tag{20}$$

Approx. by a quadratic function

- $\frac{\partial F}{\partial \alpha} = 2\sum_{i=1}^{10}\left(\alpha + \beta t_i + \gamma t_i^2 - y_i\right) \;=\; 0$ leads to

$$10\alpha + \left(\sum_{i=1}^{10} t_i\right)\beta + \left(\sum_{i=1}^{10} t_i^2\right)\gamma \;=\; \sum_{i=1}^{10} y_i$$

- $\frac{\partial F}{\partial \beta} = 2\sum_{i=1}^{10}\left(\alpha + \beta t_i + \gamma t_i^2 - y_i\right) t_i \;=\; 0$ leads to

$$\left(\sum_{i=1}^{10} t_i\right)\alpha + \left(\sum_{i=1}^{10} t_i^2\right)\beta + \left(\sum_{i=1}^{10} t_i^3\right)\gamma \;=\; \sum_{i=1}^{10} y_i t_i$$

- $\frac{\partial F}{\partial \gamma} = 2\sum_{i=1}^{10}\left(\alpha + \beta t_i + \gamma t_i^2 - y_i\right) t_i^2 \;=\; 0$ leads to

$$\left(\sum_{i=1}^{10} t_i^2\right)\alpha + \left(\sum_{i=1}^{10} t_i^3\right)\beta + \left(\sum_{i=1}^{10} t_i^4\right)\gamma \;=\; \sum_{i=1}^{10} y_i t_i^2$$

Approx. by a quadratic function

Here

$$\sum_{i=1}^{10} t_i = 55, \quad \sum_{i=1}^{10} t_i^2 = 385, \quad \sum_{i=1}^{10} t_i^3 = 3025,$$
$$\sum_{i=1}^{10} t_i^4 = 25330, \quad \sum_{i=1}^{10} y_i = 3.12, \quad \sum_{i=1}^{10} t_i y_i = 20,$$
$$\sum_{i=1}^{10} t_i^2 y_i = 138.7,$$

which leads to the linear system

$$\begin{pmatrix} 10 & 55 & 385 \\ 55 & 385 & 3025 \\ 385 & 3025 & 25330 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 3.12 \\ 20 \\ 138.7 \end{pmatrix}. \quad (21)$$

Solving the linear system (21) with, e.g., matlab we get

$$\begin{aligned} \alpha &\approx -0.4078, \\ \beta &\approx 0.2997, \\ \gamma &\approx -0.0241. \end{aligned} \tag{22}$$

We have now obtained three approximations of the data

- The constant
$$p_0(t) = 0.312$$
- The linear
$$p_1(t) = 0.123 + 0.034t$$
- The quadratic
$$p_2(t) = -0.4078 + 0.2997t - 0.0241t^2$$

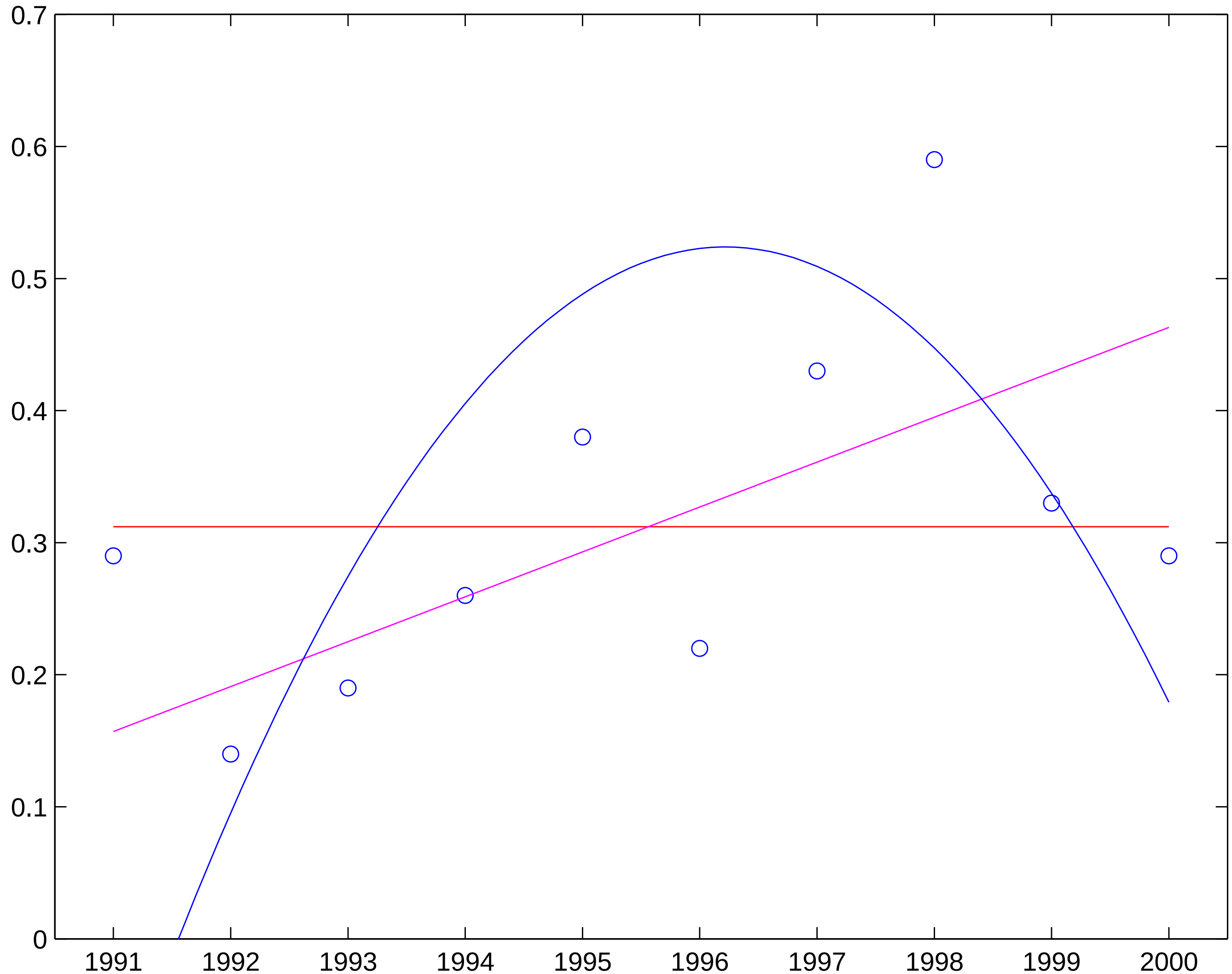

Figure 10: Constant, linear and quadratic approximations of the global annual mean temperature deviation measurements from the year 1991 to 2000.

Summary

Approximating a data set

$$(t_i, y_i) \quad i = 1, \ldots, n,$$

with a constant function

$$p_0(t) \;=\; \alpha.$$

Using the method of least squares gives

$$\alpha \;=\; \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{23}$$

which is recognized as the arithmetic average.

Summary

Approximating the data set with a linear function

$$p_1(t) \quad = \quad \alpha + \beta t$$

can be done by minimizing

$$\min_{\alpha,\beta} F(\alpha,\beta) = \min_{\alpha,\beta} \sum_{i=1}^{n} (p_1(t_i) - y_i)^2,$$

which leads to the following 2×2 linear system

$$\begin{pmatrix} n & \sum_{i=1}^{n} t_i \\ \sum_{i=1}^{n} t_i & \sum_{i=1}^{n} t_i^2 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} t_i y_i \end{pmatrix}. \qquad (24)$$

Summary

A quadratic approximation on the form

$$p_2(t) \quad = \quad \alpha + \beta t + \gamma t^2$$

can be done by minimizing $\min_{\alpha,\beta,\gamma} F(\alpha,\beta,\gamma) = \min_{\alpha,\beta,\gamma} \sum_{i=1}^{n} (p_2(t_i) - y_i)^2$, which leads to the following 3×3 linear system

$$\begin{pmatrix} n & \sum_{i=1}^{n} t_i & \sum_{i=1}^{n} t_i^2 \\ \sum_{i=1}^{n} t_i & \sum_{i=1}^{n} t_i^2 & \sum_{i=1}^{n} t_i^3 \\ \sum_{i=1}^{n} t_i^2 & \sum_{i=1}^{n} t_i^3 & \sum_{i=1}^{n} t_i^4 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i t_i \\ \sum_{i=1}^{n} y_i t_i^2 \end{pmatrix}. \tag{25}$$

Large Data Sets

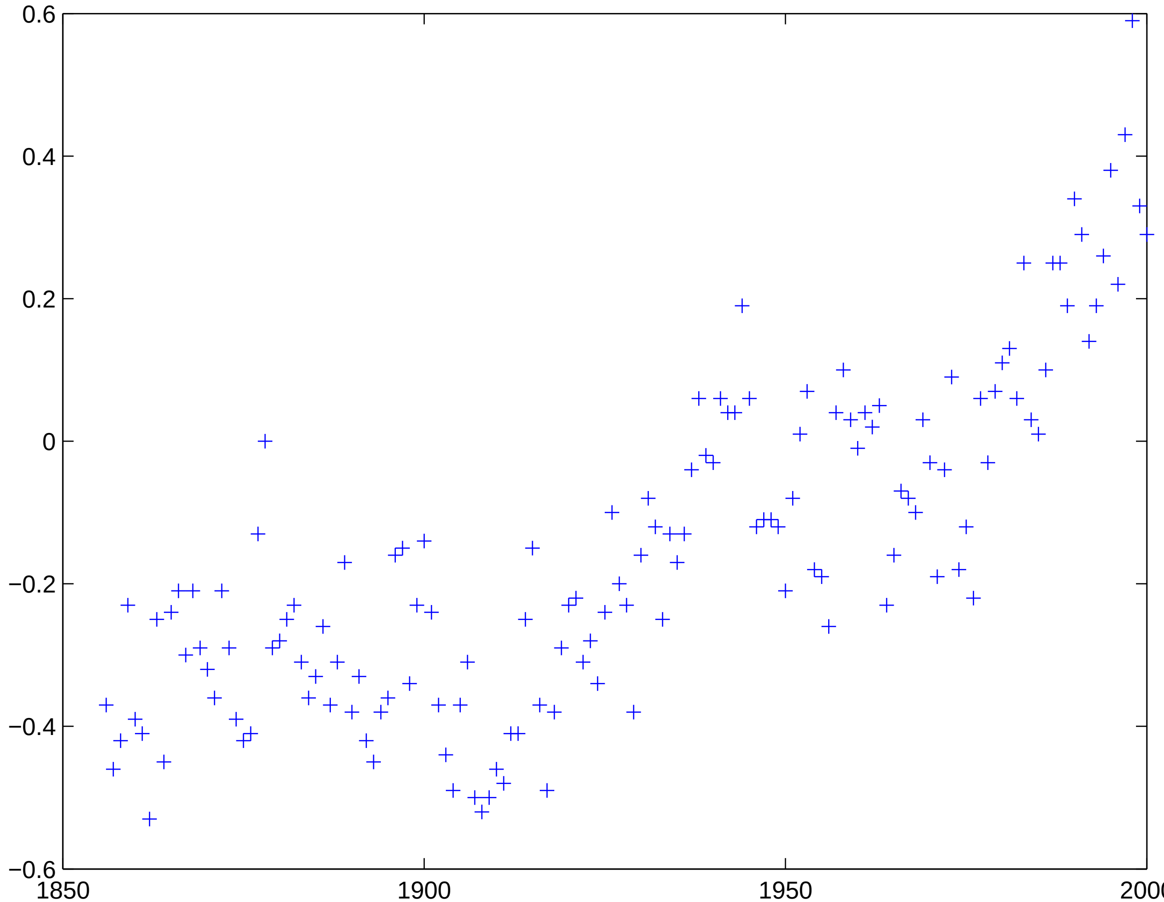

Figure 11: The global annual mean temperature deviation measurements from the year 1856 to 2000.

Application to the temperature data

We use the derived methods to model the global temperature deviation from 1856 to 2000. Here

$$
\begin{aligned}
n = 145, \qquad & \sum_{i=1}^{145} t_i = 10585, \\
\sum_{i=1}^{145} t_i^2 = 1026745, \qquad & \sum_{i=1}^{145} t_i^3 = 1.12042 \cdot 10^8, \\
\sum_{i=1}^{145} t_i^4 = 1.30415 \cdot 10^{10}, \qquad & \sum_{i=1}^{145} y_i = -21.82, \\
\sum_{i=1}^{145} t_i y_i = -502.43, \qquad & \sum_{i=1}^{145} t_i^2 y_i = 19649.8,
\end{aligned}
\tag{26}
$$

where we have used $t_i = i$, i.e., $t_1 = 1$ corresponds to the year of 1856, $t_2 = 2$ corresponds to the year of 1857 etc.

Application to the temperature data

First we get the constant model

$$p_0(t) \approx -0.1505. \tag{27}$$

The coefficients α and β of the linear model are obtained by solving the linear system (24), i.e.

$$\begin{pmatrix} 145 & 10585 \\ 10585 & 1026745 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} -21.82 \\ -502.43 \end{pmatrix}.$$

Consequently,

$$\alpha \approx -0.4638 \quad \text{and} \quad \beta \approx 0.0043,$$

so the linear model is given by

$$p_1(t) \approx -0.4638 + 0.0043t.$$

Application to the temperature data

Similarly, the coefficients α, β and γ of the quadratic model are obtained by solving the linear system (25), i.e.

$$\begin{pmatrix} 145 & 10585 & 1026745 \cdot 10^6 \\ 10585 & 1026745 \cdot 10^6 & 1.12042 \cdot 10^8 \\ 1026745 \cdot 10^6 & 1.12042 \cdot 10^8 & 1.30415 \cdot 10^{10} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} -21.82 \\ -502.43 \\ 19649.8 \end{pmatrix}$$

The solution of this system is given by

$\alpha \approx -0.3136$, $\beta \approx -1.8404 \cdot 10^{-3}$ and $\gamma \approx 4.2005 \cdot 10^{-5}$,

so the quadratic model is given by

$$p_2(t) \approx -0.3136 - 1.8404 \cdot 10^{-3} t + 4.2005 \cdot 10^{-5} t^2. \quad (28)$$

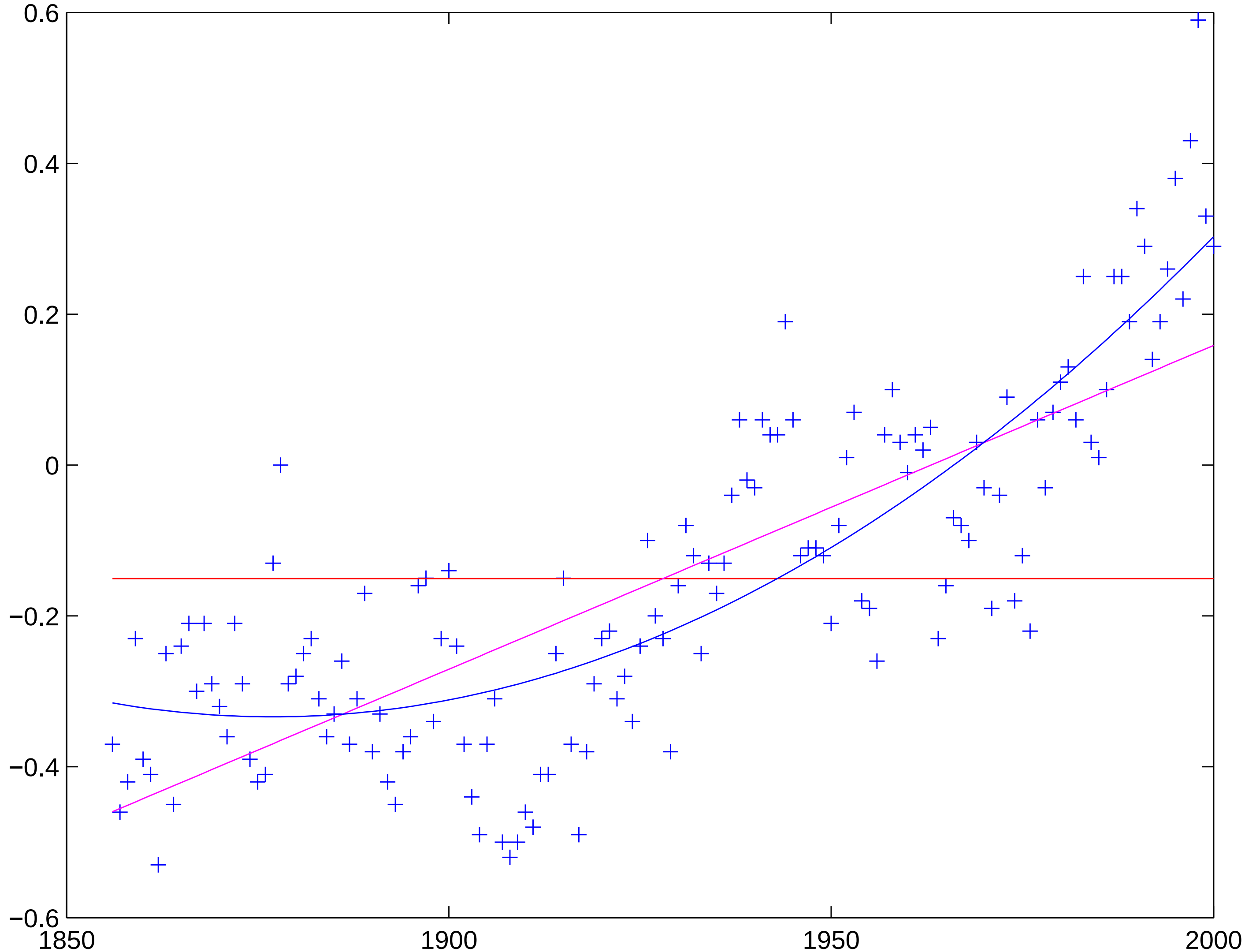

Figure 12: Constant, linear and quadratic least squares approximations of the global annual mean temperature deviation measurements; from 1856 to 2000.

Application to population models

We now consider the growth of the world population.

Year	Population (billions)
1950	2.555
1951	2.593
1952	2.635
1953	2.680
1954	2.728
1955	2.780

Table 2: The total world population from 1950 to 1955.

Exponential growth

First we model the data in Table 2 using the exponential growth model

$$p'(t) = \alpha p(t), \qquad p(0) = p_0, \tag{29}$$

with solution $p(t) = p_0 e^{\alpha t}$.

- We have earlier mentioned that for this model α has to be estimated
- We shall now estimate α relative to this data, in the sense of least squares

Exponential growth

- Put $t = 0$ at 1950 and measure t in years
- $p_0 = 2.555$
- We want to determine the parameter α for data from Table 2 and

$$\frac{p'(t)}{p(t)} = \alpha \tag{30}$$

- Since only p is available, we have to approximate $p'(t)$ using the standard formula

$$p'(t) \approx \frac{p(t+\Delta t) - p(t)}{\Delta t} \tag{31}$$

Exponential growth

- By choosing $\Delta t = 1$, we estimate α to be

$$\alpha_n = \frac{p(n+1) - p(n)}{p(n)} \tag{32}$$

- Let b_n be the relative annual growth in percentage, i.e.

$$b_n = 100\,\frac{p(n+1) - p(n)}{p(n)} \tag{33}$$

Exponential growth

Year	n	$p(n)$	$b_n = 100\,\dfrac{p(n+1)-p(n)}{p(n)}$
1950	0	2.555	1.49
1951	1	2.593	1.62
1952	2	2.635	1.71
1953	3	2.680	1.79
1954	4	2.728	1.91
1955	5	2.780	

Table 3: Estimated values of $\frac{p'(t)}{p(t)}$ using (33) based on the numbers of the world's population from 1950 to 1955.

Exponential growth

- We can to compute a constant approximation b, to this data set in the sense of least squares, i.e. the average

$$b = \frac{1}{5}\sum_{n=0}^{4} b_n = \frac{1}{5}\big(1.49 + 1.62 + 1.71 + 1.79 + 1.91\big) = 1.704$$

- Since $b_n = 100\alpha_n$ we get

$$\alpha = \frac{1}{100} b = 0.01704 \tag{34}$$

- This gives us the model

$$p(t) = 2.555\, e^{0.01704\, t} \tag{35}$$

Exponential growth

- In the year 2000 (t=50) this model gives

$$p(50) \;=\; 2.555\,e^{0.01704\times 50} \approx 5.990 \tag{36}$$

- The actual population in the 2000 was 6.080 billions
- The relative error is

$$\frac{6.080-5.990}{6.080}\cdot 100\% \;=\; 1.48\%, \tag{37}$$

which is remarkably small

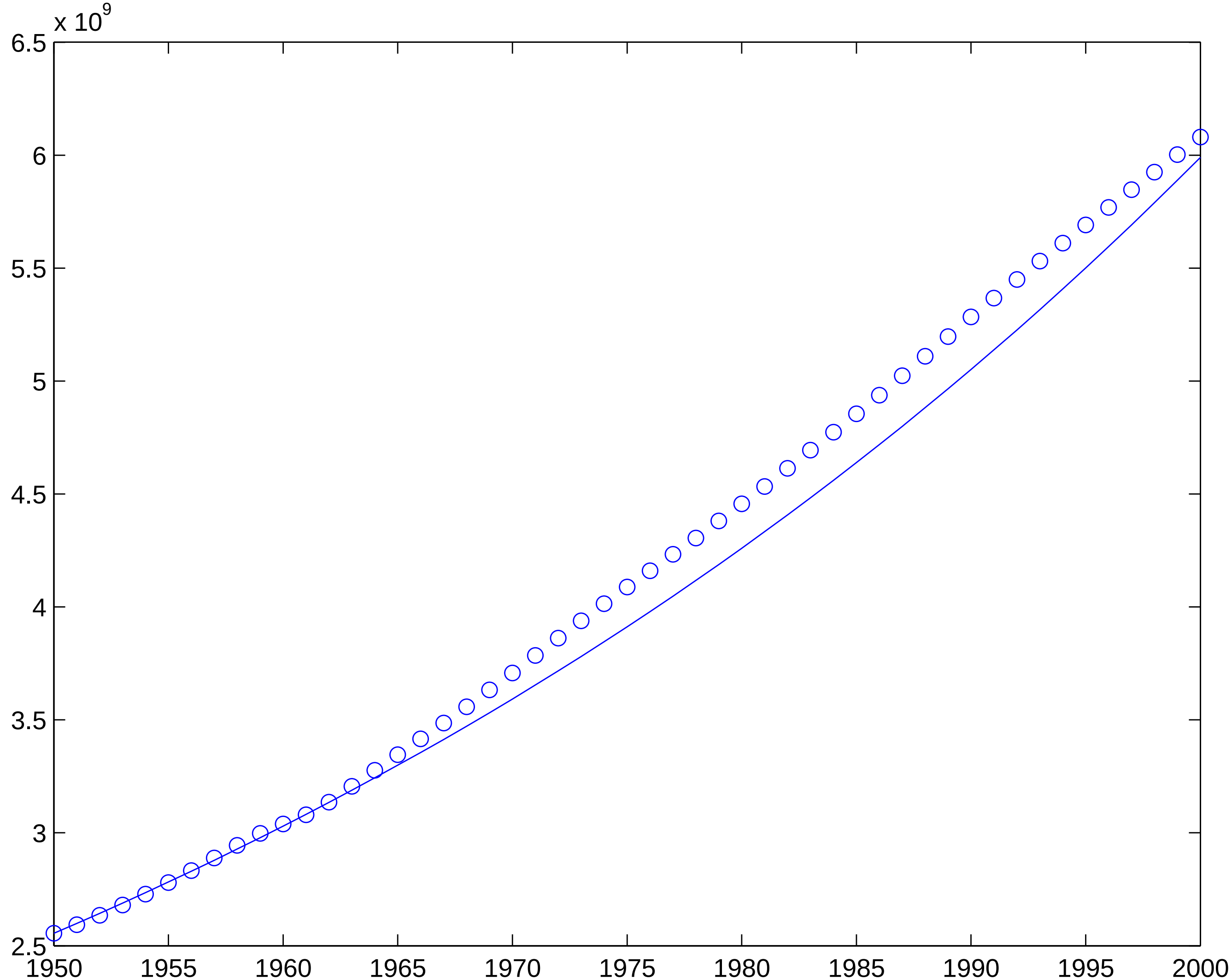

Figure 13: The figure shows the graph of an exponential population model $p(t) = 2.555e^{0.01704t}$ together with the actual measurements marked by 'o'.

Exponential growth

- Will this model fit in the future as well?
- We try to use the development from 1990 to 2000 (see Figure 13) to predict the world population in 2100

Year	n	$p(n)$	$b_n = 100\,\frac{p(n+1)-p(n)}{p(n)}$
1990	0	5.284	1.57
1991	1	5.367	1.55
1992	2	5.450	1.49
1993	3	5.531	1.46
1994	4	5.611	1.43
1995	5	5.691	1.37
1996	6	5.769	1.35
1997	7	5.847	1.33
1998	8	5.925	1.32
1999	9	6.003	1.28
2000	10	6.080	

Table 4: The calculated b_n values associated with an exponential population model for the world between 1990 and 2000.

Exponential growth

- The average of the b_n values is

$$b = \frac{1}{10}\sum_{n=0}^{9} b_n = 1.42$$

- Therefore $\alpha = \frac{b}{100} = 0.0142$
- Starting from t=0 in the year 2000, the model reads

$$p(t) = 6.080\, e^{0.0142\, t} \tag{38}$$

- This model predicts that there will be

$$p(100) = 6.080\, e^{1.42} \approx 25.2 \tag{39}$$

billion people living on the earth in the year of 2100

Logistic growth of the world pop.

- We study how the world population can be modeled by the logistic growth model

$$p'(t) = \alpha p(t)(1 - p(t)/\beta) \tag{40}$$

- By defining $\gamma = -\alpha/\beta$, we can rewrite (40)

$$\frac{p'(t)}{p(t)} = \alpha + \gamma p(t) \tag{41}$$

- Hence, we can determine constants α and γ by fitting a linear function to the observations of

$$\frac{p'(t)}{p(t)} \tag{42}$$

Logistic growth

- As above we define

$$b_n = 100\,\frac{p(n+1)-p(n)}{p(n)} \tag{43}$$

- We now want to determine two constants A and B, such that the data are modeled as accurately as possible by a linear function

$$b_n \approx A + Bp(n), \tag{44}$$

in the sense of least squares

Logistic growth

- A and B can be determined by the following 2×2 linear system,

$$\begin{pmatrix} 10 & \sum_{n=0}^{9} p(n) \\ \sum_{n=0}^{9} p(n) & \sum_{n=0}^{9} (p(n))^2 \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} \sum_{n=0}^{9} b_n \\ \sum_{n=0}^{9} p(n) b_n \end{pmatrix} \quad (45)$$

- We have

$$\sum_{n=0}^{9} p(n) = 56.5, \qquad \sum_{n=0}^{9} (p(n))^2 = 319.5,$$

$$\sum_{n=0}^{9} b_n \approx 14.1, \qquad \sum_{n=0}^{9} p(n) b_n \approx 79.6$$

Logistic growth

- By solving the (2×2) system

$$\begin{pmatrix} 10 & 56.5 \\ 56.5 & 319.5 \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} 14.1 \\ 79.6 \end{pmatrix}$$

  we get

$$A = 2.7455 \quad \text{and} \quad B = -0.2364$$

- Inserting this in the logistic model we get

$$p'(t) \approx 0.027\, p(t)\,(1 - p(t)/11.44) \tag{46}$$

- This indicates that the carrying capacity of the earth is about 11.44 billions

Logistic growth

- Let $t = 0$ correspond to the year 2000, which gives the initial condition

$$p(0) \approx 6.08$$

- The analytical solution to the differential equation is

$$p(t) \approx \frac{69.5}{6.08 + 5.36\,e^{-0.027t}} \tag{47}$$

- This model predicts that there will be

$$p(100) \approx 10.79$$

billion people on the earth in the year of 2100

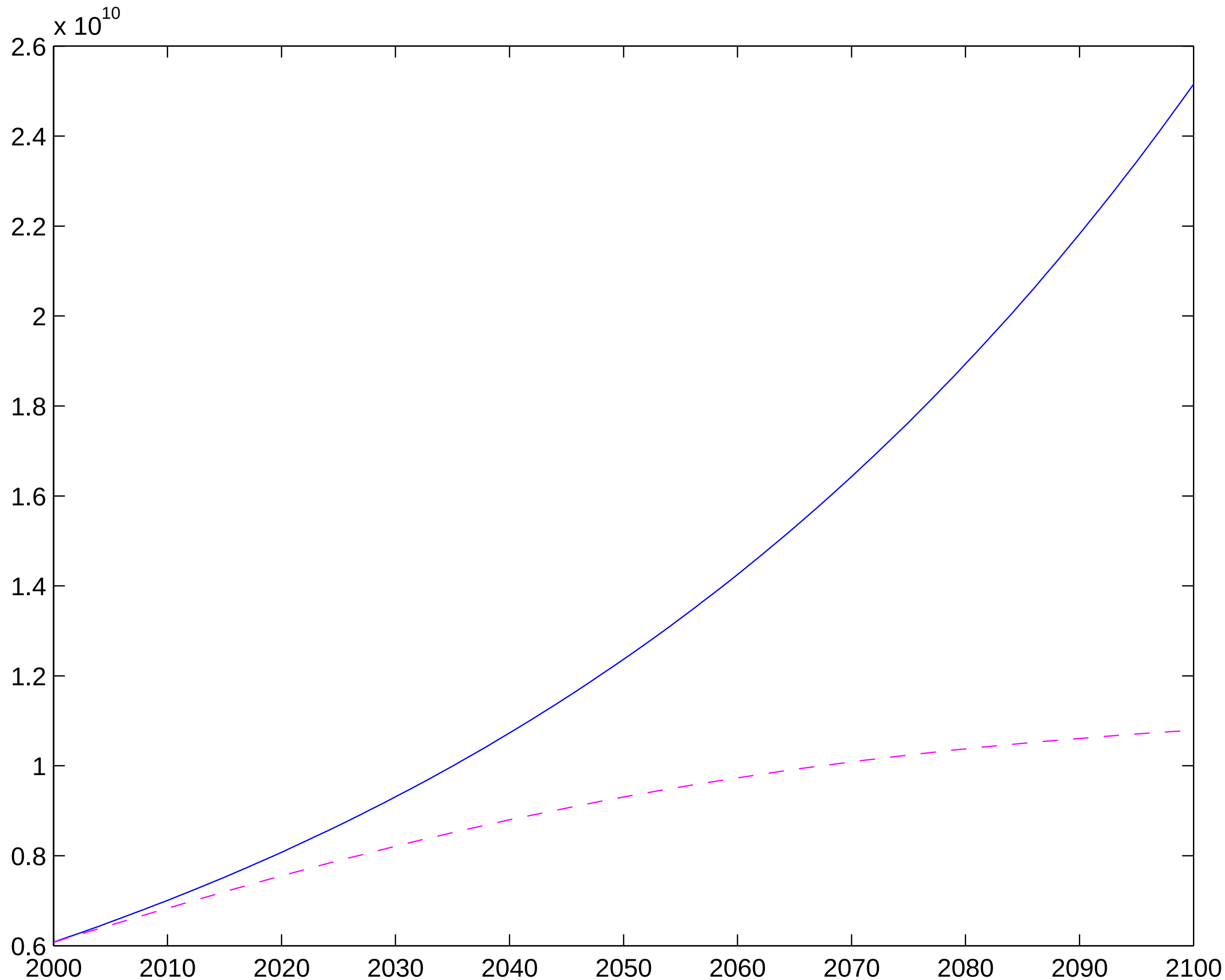

Figure 14: Predictions of the population growth on the earth based on an exponential model (solid curve) and a logistic model (dashed curve).

Approximations of Functions

- Above we have studied continuous representation of discrete data
- Next we will consider continuous approximation of continuous functions
- Consider the function

$$y(t) = \ln\left(\frac{1}{10}\sin(t) + e^t\right) \quad (48)$$

- In Figure 15 we see that $y(x)$ seems to be close to the linear function $p(t) = t$ on the interval $[0,1]$
- In Figure 16 we see that $y(x)$ seems to be even closer to the linear function plotted on $t \in [0,10]$

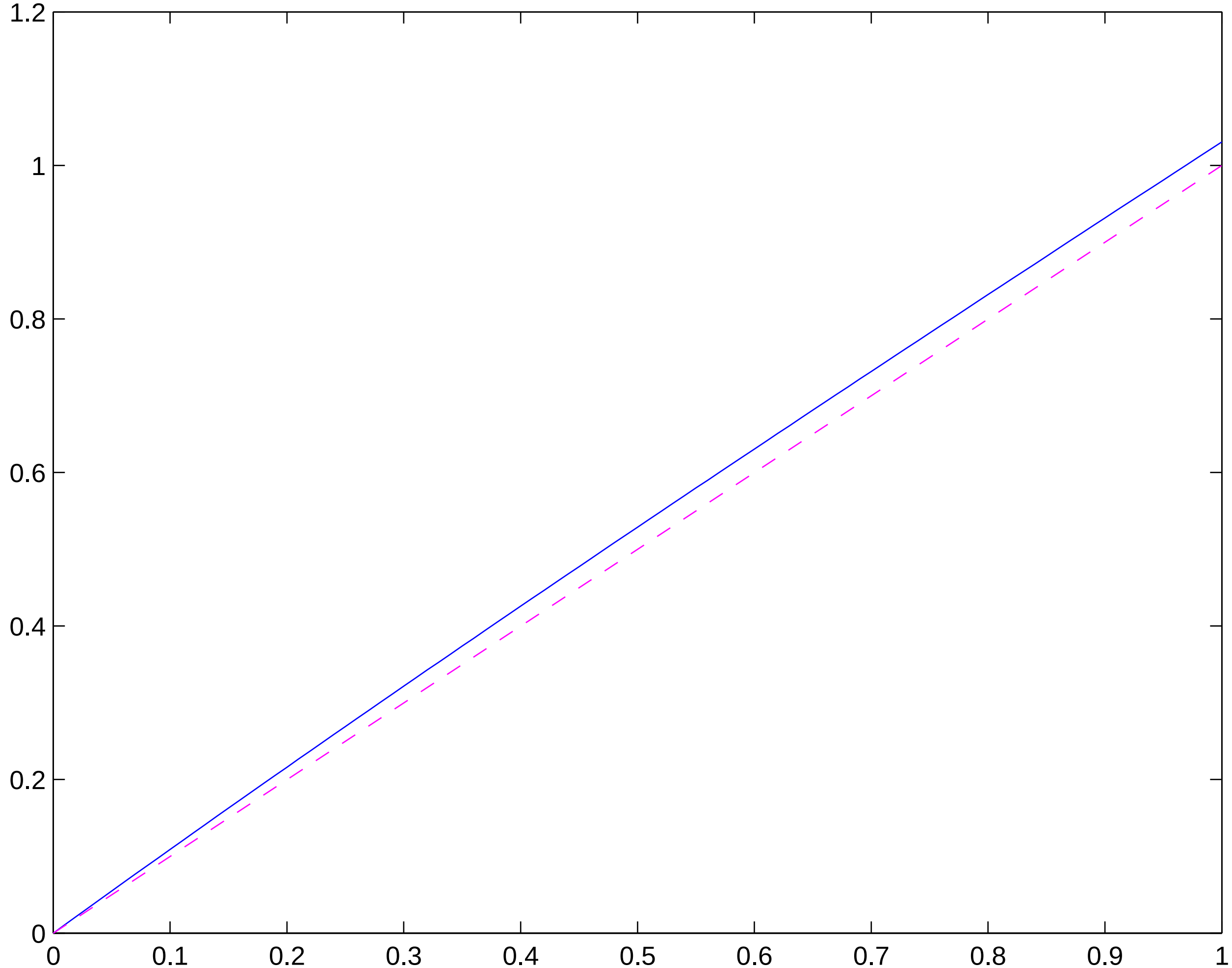

Figure 15: The function $y(t) = \ln\left(\frac{1}{10}\sin(t) + e^t\right)$ (solid curve) and a linear approximation (dashed line) on the interval $t \in [0,1]$.

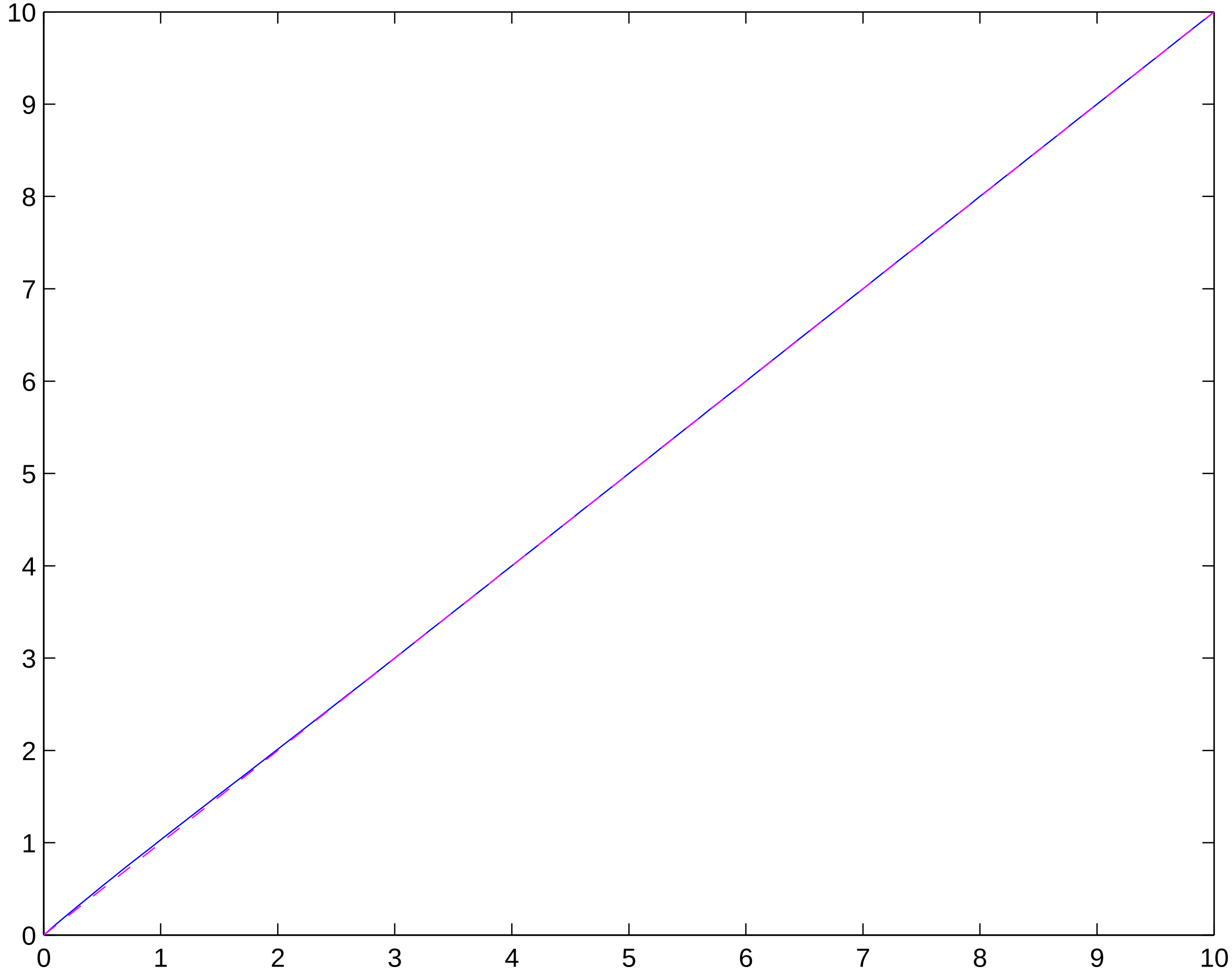

Figure 16: The function $y(t) = \ln\left(\frac{1}{10}\sin(t) + e^t\right)$ (solid curve) and a linear approximation (dashed line) on the interval $t \in [0, 10]$.

Approximations by constants

- For a given function $y(t)$, $t \in [a,b]$, we want to compute a constant approximation of it

$$p(t) = \alpha \tag{49}$$

for $t \in [a,b]$, in the sense of least squares

- That means that we want to minimize the integral

$$\int_a^b (p(t) - y(t))^2 \, dt = \int_a^b (\alpha - y(t))^2 \, dt$$

Approximations by constants

- Define the function

$$F(\alpha) \;=\; \int_a^b (\alpha - y(t))^2\, dt \qquad (50)$$

- The derivative with respect to α is

$$F'(\alpha) \;=\; 2\int_a^b (\alpha - y(t))\, dt$$

- And solving $F'(\alpha) = 0$ gives

$$\alpha \;=\; \frac{1}{b-a}\int_a^b y(t)dt \qquad (51)$$

Note that

- The formula for α is the integral version of the average of y on $[a,b]$. In the discrete case we would have written

$$\alpha = \frac{1}{n}\sum_{i=1}^{n} y_i, \qquad (52)$$

If y_i in (52) is $y(t_i)$, where $t_i = a + i\Delta t$ and $\Delta t = \frac{b-a}{n}$, then

$$\frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{b-a}\Delta t \sum_{i=1}^{n} y(t_i) \approx \frac{1}{b-a}\int_a^b y(t)\,dt.$$

We therefore conclude that (51) is a natural continuous version of (52).

Note that

- We used

$$\frac{d}{d\alpha}\int_a^b(\alpha - y(t))^2 dt = \int_a^b \frac{\partial}{\partial\alpha}(\alpha - y(t))^2 dt$$

  Is that a legal operation? This is discussed in Exercise 5.

- The α given by (51) is a minimum, since

$$F''(\alpha) = 2(b-a) > 0$$

Example 15; const. approx.

Consider

$$y(t) \;=\; \sin(t)$$

defined on $0 \le t \le \pi/2$. A constant approximation of y is given by

$$\begin{aligned} p(t) \;=\; \alpha \quad &\overset{(51)}{=} \quad \frac{2}{\pi}\int_0^{\pi/2} \sin(t)\,dt \;=\; \frac{-2}{\pi}\left[\cos(t)\right]_0^{\pi/2} \\ &= \quad \frac{-2}{\pi}(0-1) \;=\; \frac{2}{\pi}. \end{aligned}$$

Example 16; const. approx.

Consider

$$y(t) \;=\; t^2 + \frac{1}{10}\cos(t)$$

defined on $0 \leq t \leq 1$. A constant approximation of y is given by

$$\begin{aligned} p(t) = \alpha \overset{(51)}{=} & \int_0^1 \left(t^2 + \frac{1}{10}\cos(t)\right) dt = \left[\frac{1}{3}t^3 + \frac{1}{10}\sin(t)\right]_0^1 \\ = & \frac{1}{3} + \frac{1}{10}\sin(1) \approx 0.417. \end{aligned}$$

Approximations by Linear Functions

- Now, we search for a linear approximation of a function $y(t)$, $t \in [a,b]$, i.e.

$$p(t) = \alpha + \beta t \tag{53}$$

in the sense of least squares

- Define

$$F(\alpha,\beta) = \int_a^b (\alpha + \beta t - y(t))^2 dt \tag{54}$$

- A minimum of F is obtained by finding α and β such that

$$\frac{\partial F}{\partial \alpha} = \frac{\partial F}{\partial \beta} = 0$$

Approximations by Linear Functions

- We have

$$
\begin{aligned}
\frac{\partial F}{\partial \alpha} &= 2\int_a^b (\alpha + \beta t - y(t))\, dt \\
\frac{\partial F}{\partial \beta} &= 2\int_a^b (\alpha + \beta t - y(t)) t\, dt
\end{aligned}
$$

- Therefore α and β can be determined by solving the following linear system

$$
\begin{aligned}
(b-a)\alpha + \frac{1}{2}(b^2 - a^2)\beta &= \int_a^b y(t)\, dt \\
\frac{1}{2}(b^2 - a^2)\alpha + \frac{1}{3}(b^3 - a^3)\beta &= \int_a^b t\, y(t)\, dt
\end{aligned}
\tag{55}
$$

Example 15; linear approx.

Consider

$$y(t) \;=\; \sin(t)$$

defined on $0 \le t \le \pi/2$.
We have

$$\int_0^{\pi/2} \sin(t)\,dt \;=\; 1$$

and

$$\int_0^{\pi/2} t\sin(t)\,dt \;=\; 1.$$

Example 15; linear approx.

The linear system now reads

$$\begin{pmatrix} \pi/2 & \pi^2/8 \\ \pi^2/8 & \pi^3/24 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The solution is

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \frac{1}{\pi^2} \begin{pmatrix} 8\pi - 24 \\ \frac{96}{\pi} - 24 \end{pmatrix} \approx \begin{pmatrix} 0.115 \\ 0.664 \end{pmatrix}.$$

Therefore the linear approximation is given by

$$p(t) \approx 0.115 + 0.664t.$$

Example 16; linear approx.

Consider

$$y(t) \;\; = \;\; t^2 + \frac{1}{10}\cos(t)$$

defined on $0 \leq t \leq 1$. The linear system (55) then reads

$$\begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \;\; = \;\; \begin{pmatrix} \frac{1}{3} + \frac{1}{10}\sin(1) \\ \frac{3}{20} + \frac{1}{10}\cos(1) + \frac{1}{10}\sin(1) \end{pmatrix},$$

with solution $\alpha \approx -0.059$ and $\beta \approx 0.953$.
We conclude that the linear least squares approximation is given by

$$p(t) \;\; \approx \;\; -0.059 + 0.953t.$$

Approx. by Quadratic Functions

- We seek a quadratic function

$$p(t) = \alpha + \beta t + \gamma t^2 \tag{56}$$

  that approximates a given function $y = y(t)$, $a \leq t \leq b$, in the sense of least squares

- Let

$$F(\alpha, \beta, \gamma) = \int_a^b (\alpha + \beta t + \gamma t^2 - y(t))^2 dt \tag{57}$$

- Define α, β and γ to be the solution of the three equations:

$$\frac{\partial F}{\partial \alpha} = \frac{\partial F}{\partial \beta} = \frac{\partial F}{\partial \gamma} = 0$$

Approx. by Quadratic Functions

- By taking the derivatives, we have
  - $$\frac{\partial F}{\partial \alpha} = 2\int_a^b (\alpha + \beta t + \gamma t^2 - y(t))\,dt$$
  - $$\frac{\partial F}{\partial \beta} = 2\int_a^b (\alpha + \beta t + \gamma t^2 - y(t))\,t\,dt$$
  - $$\frac{\partial F}{\partial \gamma} = 2\int_a^b (\alpha + \beta t + \gamma t^2 - y(t))\,t^2\,dt$$

- The coefficients α, β and γ can now be determined from the linear system

$$
\begin{aligned}
(b-a)\alpha + \frac{1}{2}(b^2-a^2)\beta + \frac{1}{3}(b^3-a^3)\gamma &= \int_a^b y(t)\,dt \\
\frac{1}{2}(b^2-a^2)\alpha + \frac{1}{3}(b^3-a^3)\beta + \frac{1}{4}(b^4-a^4)\gamma &= \int_a^b t\,y(t)\,dt \\
\frac{1}{3}(b^3-a^3)\alpha + \frac{1}{4}(b^4-a^4)\beta + \frac{1}{5}(b^5-a^5)\gamma &= \int_a^b t^2 y(t)\,dt
\end{aligned}
$$

Example 15; quad. approx.

For the function

$$y(t) \;=\; \sin(t), \quad 0 \le t \le \pi/2,$$

the linear system reads

$$\begin{pmatrix} \pi/2 & \pi^2/8 & \pi^3/24 \\ \pi^2/8 & \pi^3/24 & \pi^4/64 \\ \pi^3/24 & \pi^4/64 & \pi^5/160 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} \;=\; \begin{pmatrix} 1 \\ 1 \\ \pi - 2 \end{pmatrix},$$

and the solution is given by $\alpha \approx -0.024$, $\beta \approx 1.196$ and $\gamma \approx -0.338$, which gives the quadratic approximation

$$p(t) \;=\; -0.024 + 1.196\,t - 0.338\,t^2.$$

Example 16; quad. approx.

Let us consider

$$y(t) = t^2 + \frac{1}{10}\cos(t)$$

for $0 \leq t \leq 1$. The linear system takes the form

$$\begin{pmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} \frac{1}{3} + \frac{1}{10}\sin(1) \\ \frac{3}{20} + \frac{1}{10}\cos(1) + \frac{1}{10}\sin(1) \\ \frac{1}{5} + \frac{1}{5}\cos(1) - \frac{1}{10}\sin(1) \end{pmatrix}$$

and the solution is given by $\alpha \approx 0.100$, $\beta \approx -0.004$ and $\gamma \approx 0.957$, and the quadratic approximation is

$$p(t) = 0.100 - 0.004t + 0.957t^2.$$

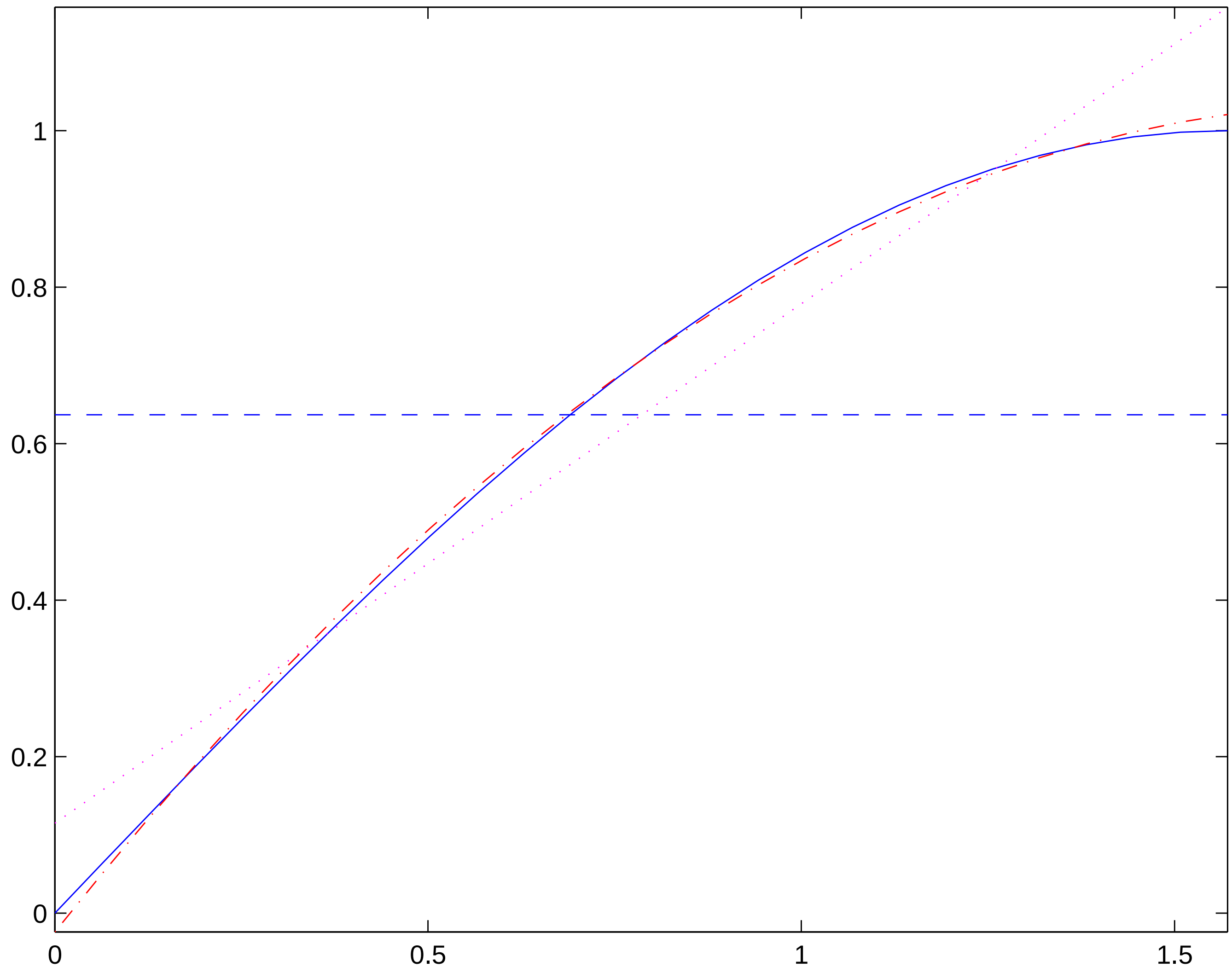

Figure 17: The function $y(t) = \sin(t)$ (solid curve) and its least squares approximations: constant (dashed line), linear (dotted line) and quadratic (dashed-dotted curve).

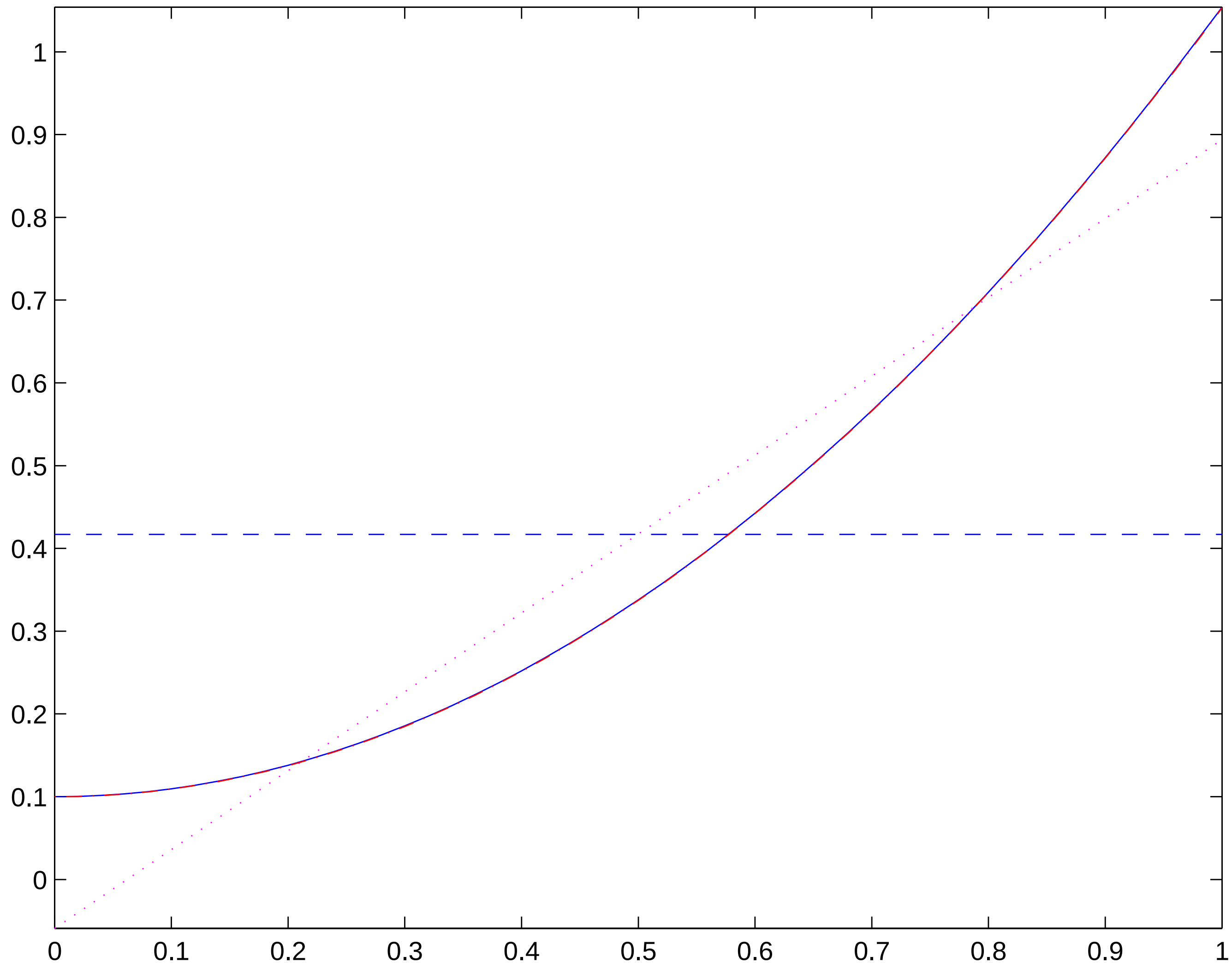

Figure 18: The function $y(t) = t^2 + \frac{1}{10}\cos(t)$ (solid curve) and its least squares approximations: constant (dashed line), linear (dotted line) and quadratic (dashed-dotted curve).