AESB2440: Geostatistics & Remote Sensing

Least Squares

Lecture 6, May 1, 2015,

Roderik Lindenbergh

Dept. of Geoscience & Remote Sensing

Delft University of Technology

Least Squares

Least Squares

- Measurement redundancy
- Optimal fit
- Minimizing fit error
- Least Squares Solution

Fit examples

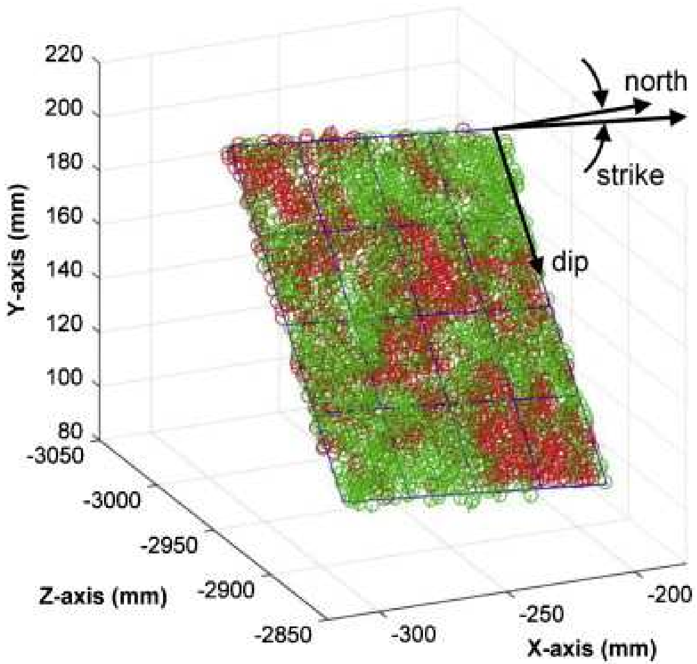
- Velocity example
- Plane examples
- Curve fitting
- Changes in the Texel dunes

Weighted least squares

- incorporating measurement quality

Digital Terrain Analysis

First start: look at planes!



Planes give you:

- Surface Approximation
- Surface Orientation
- Surface Slope
- Surface Roughness

Later: look at derivatives.

Image source: http://www.sciencedirect.com/science/article/pii/S1365160912001724

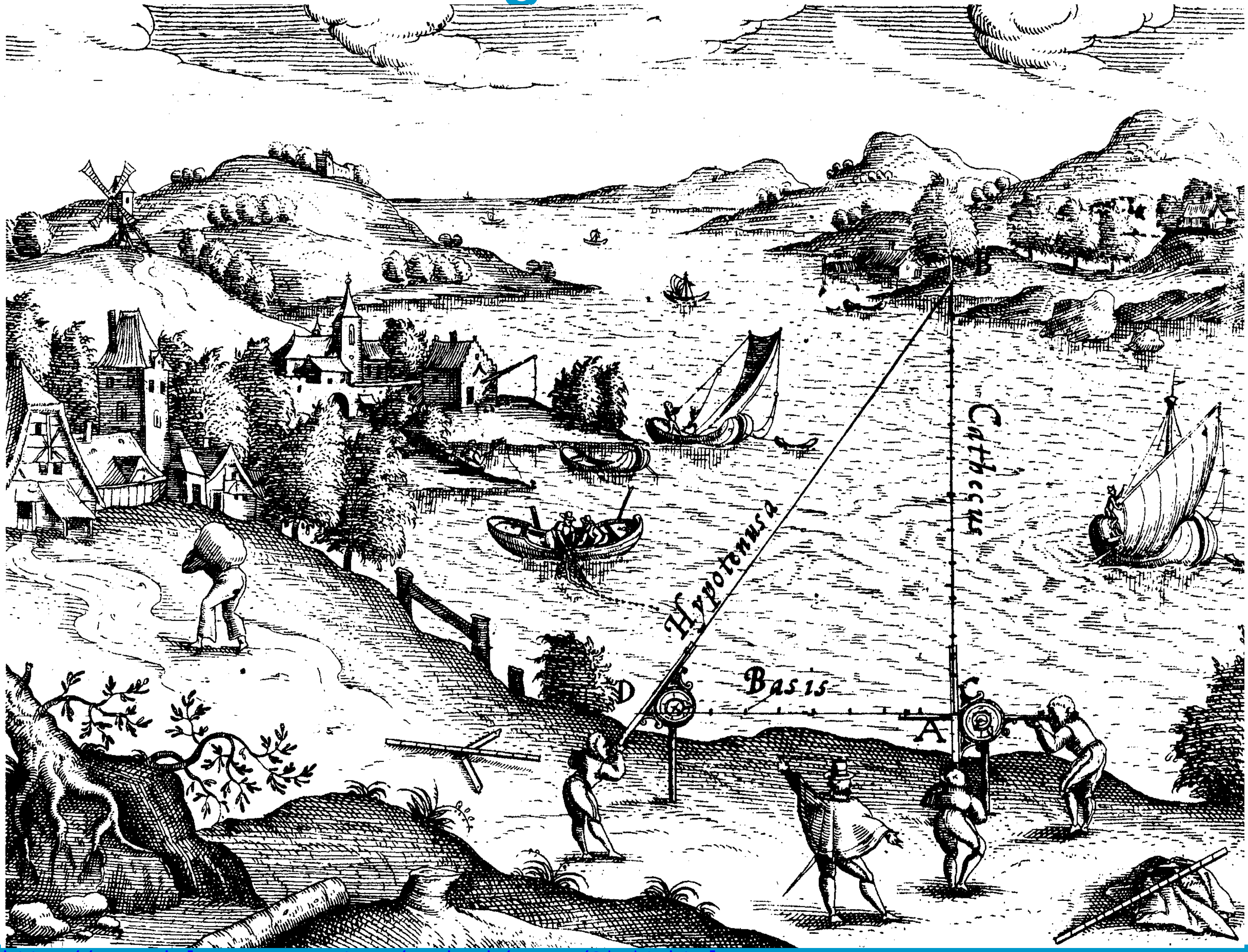
A. Least Squares

Least Squares References

Linear Algebra and its applications, Lay, D.C.,
3rd edition, Addison Wesley (2003).
Notably Chapter 6.5: Least-Squares Problems

Primer on Mathematical Geodesy, C.C.J.M. Tiberius,
TU Delft, Faculty of Civil Engineering and Geosciences, (2014).
Available as pdf via Blackboard
(Notably Chapter 4)

Gauss: measuring the Earth



Source http://www.ajaloomuuseum.ut.ee/vvebook/pages/8$_$1.html

Measurement Redundancy

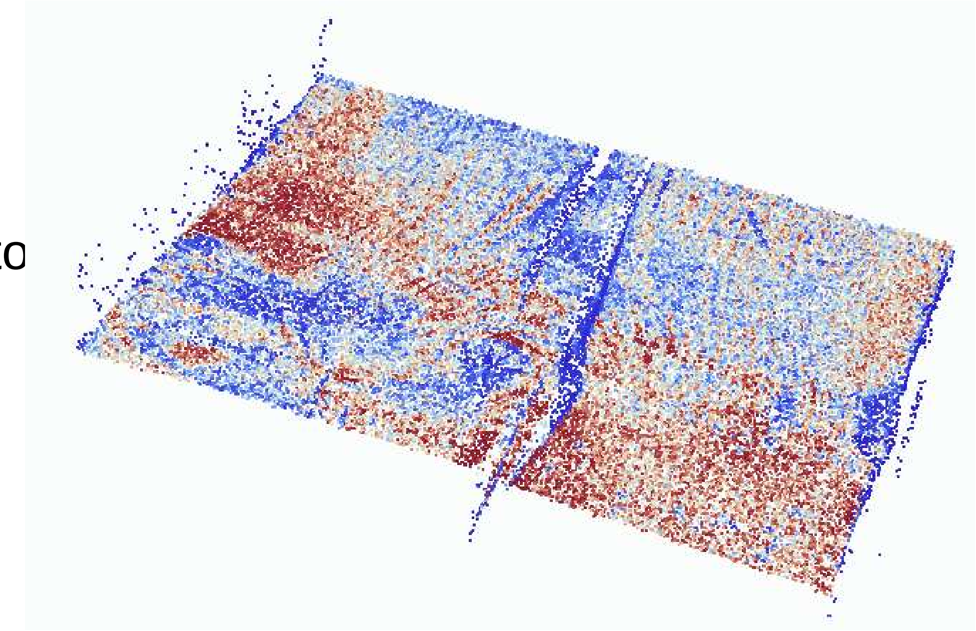
Question.
How many points in 3D are needed to estimate a plane?

Question.
How many parameters are needed to fix an arbitrary plane in 3D?

The (measurement) redundancy is:

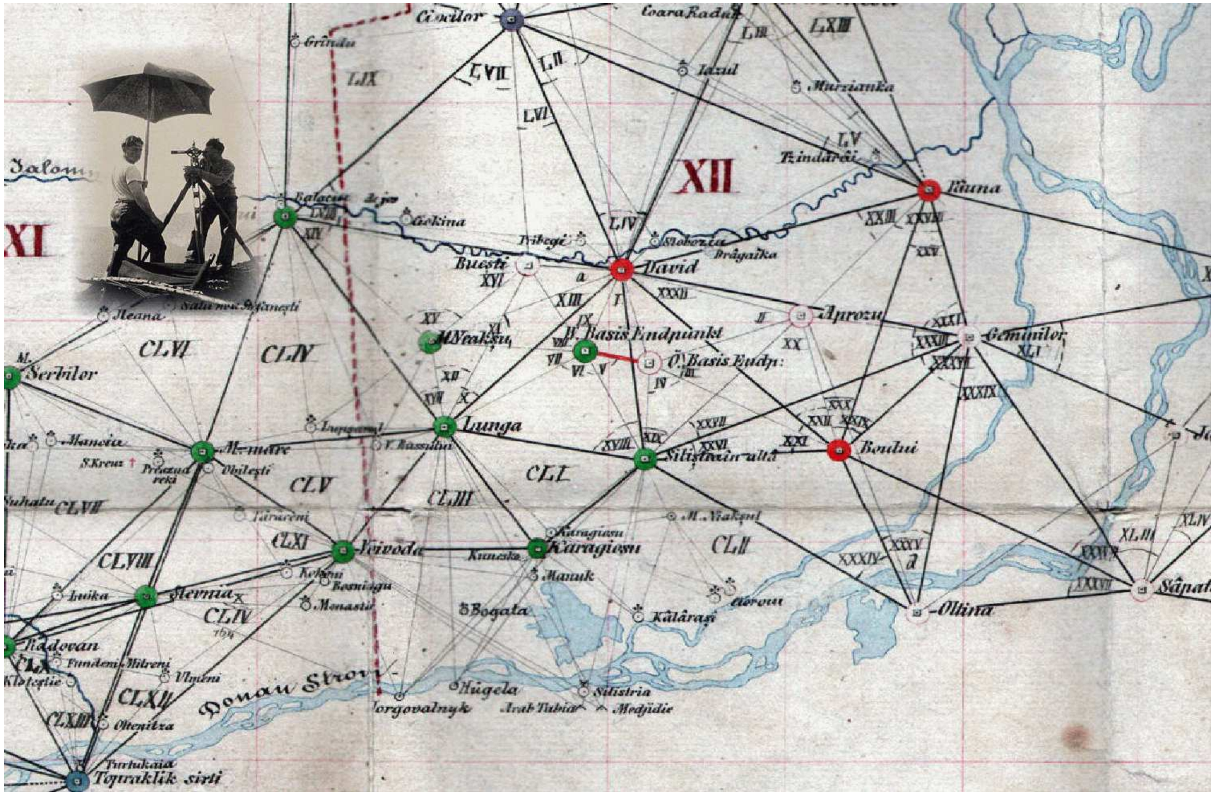
- the number of observations, m, used for an estimation
- minus the number of parameters, n, used to desribe the geometric model

Often, $m >> n$.



Example.
Wooden (planar) beam face sampled by 50 000 laser scan points.

Redundancy in triangulation networks



Measuring more then strictly necessary enables error identification

Source `https://confluence.qps.nl/display/KBE/Howto+Computation+Setup`

Linear system

Input: m measurements $y_1, y_2, \ldots, y_m$
Output: (Wished) n parameters $x_1, \ldots, x_n$

Assumption: Measurements have a linear relationship with the parameters.

That is, for each y_i, $i = 1, \ldots, m$ we may write

$$y_i \quad = \quad a_{i1}x_1 + a_{i2}x_2 + \ldots a_{in}x_n$$

To establish the relation: determine the unknown values of the coefficients

$$a_{ij}, \qquad \text{with} \qquad i = 1, \ldots, m \qquad \text{and} \qquad j = 1, \ldots, n$$

In matrix form: with $A = a_{ij}$ the matrix of coefficients, $\mathbf{y} = \{y_1, \ldots, y_m\}$ the vector of observations, and $\mathbf{x} = \{x_1, \ldots, x_n\}$ the parameter vector we get:

$$\mathbf{y} \quad = \quad A\mathbf{x}$$

Least Squares

$\mathbf{y}$	-	Vector of observations
A	-	Model matrix
$\mathbf{x}$	-	Vector of model parameters
$\mathbf{e}$	-	Vector of residuals, or, mismatches between observations and model

Given: the observations

Assumed: a model, describing the physical or mathematical reality or geometry, so

$$\mathbf{y} = A \cdot \mathbf{x} + \mathbf{e}$$

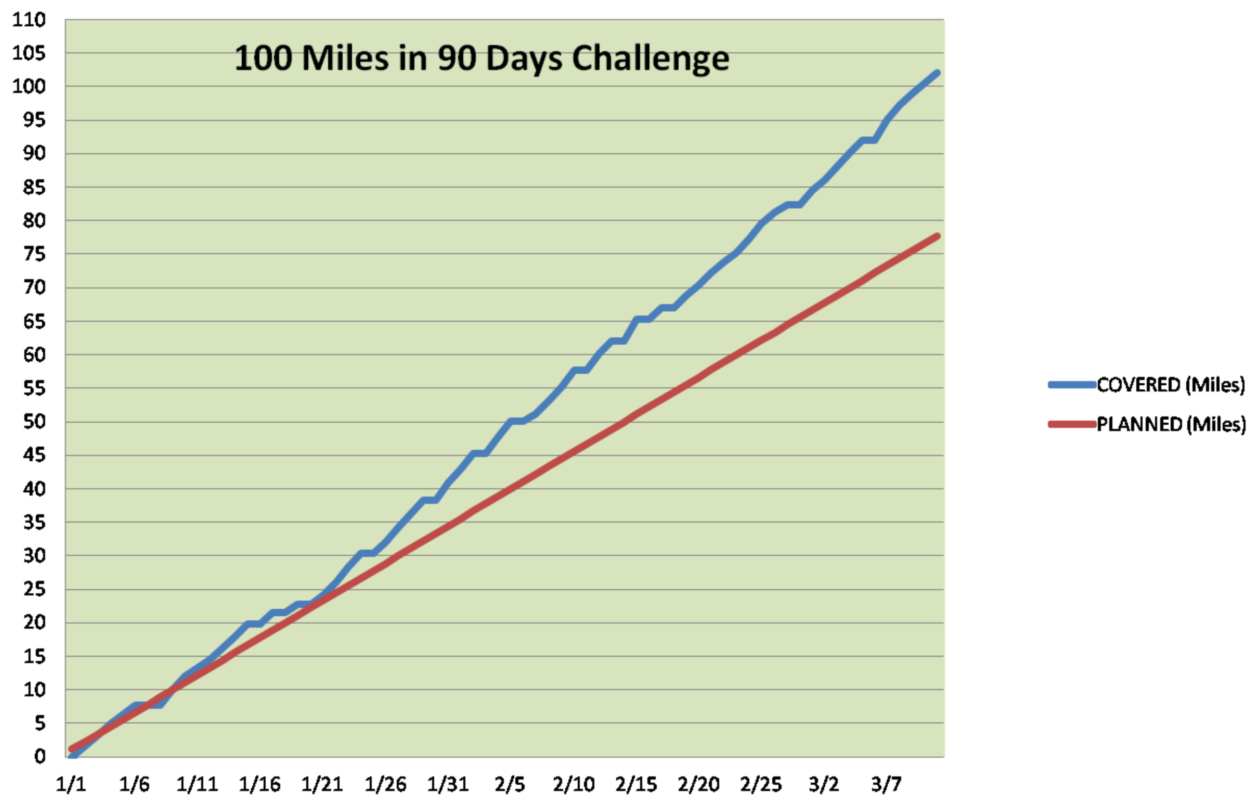
Wanted: best choice for the parameters of the model

$\Rightarrow$ Minimize the errors in some sense

Definition. Best: Minimize the squares of the errors: $\mathbf{e}^T \cdot \mathbf{e}$

Dept. of Geoscience & Remote Sensing

Covered distance example



Source http://personapaper.com/article/26547-100-miles-in-90-days---final-progress-report

Velocity Example

The covered distance $d(t)$ is assumed to depend linearly on the constant velocity:

$$d(t) \quad = \quad v \cdot t$$

Suppose covered distances y_1, y_2 and y_3 are observed at moments $t = 2$, $t = 3$ and $t = 4$. What was the velocity v?

[Proof.]

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \quad \text{and} \quad A = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}, \quad \text{and} \quad \mathbf{x} = x = v$$

$\Rightarrow$

$$\mathbf{e} = \mathbf{y} - A \cdot x = \begin{pmatrix} y_1 - 2v \\ y_2 - 3v \\ y_3 - 4v \end{pmatrix}$$

We try to minimize the mean error in the least squares sense:

Solution velocity example

[Proof.] So find v such that

$$\mathbf{e}^T\mathbf{e}(v) = (y_1 - 2v)^2 + (y_2 - 3v)^2 + (y_3 - 4v)^2$$

is minimal. The expression $\mathbf{e}^T\mathbf{e}(v)$ has a minimum $\Rightarrow$

$$\begin{aligned} \frac{d}{dv}\mathbf{e}^T\mathbf{e}(v) &= 0 \\ -2[(y_1 - 2v)2 + (y_2 - 3v)3 + (y_3 - 4v)4] &= 0 \end{aligned}$$

Or, after reordering:

$$2y_1 + 3y_2 + 4y_3 = 2 \cdot 2v + 3 \cdot 3v + 4 \cdot 4v$$

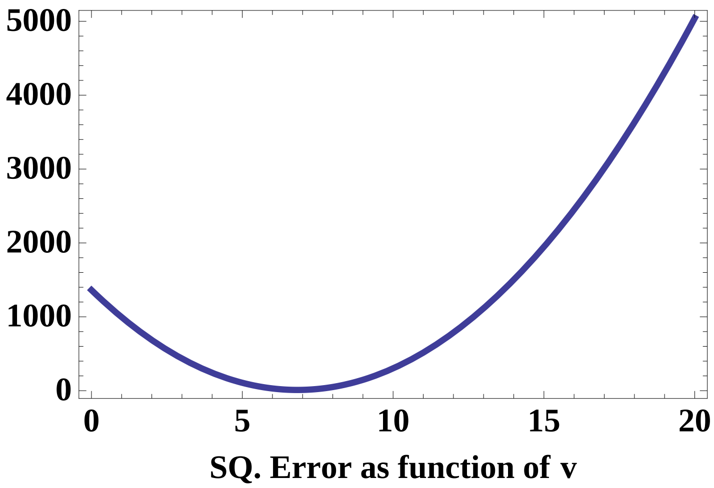
So,

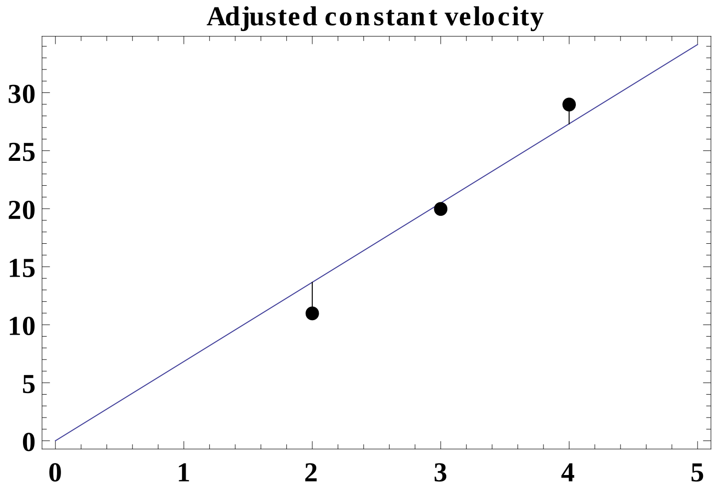
$$v = \frac{2y_1 + 3y_2 + 4y_3}{2 \cdot 2 + 3 \cdot 3 + 4 \cdot 4} = \frac{A^T\mathbf{y}}{A^TA}$$

Illustration, velocity example

[Proof.] Take $y_1 = 11$, $y_2 = 20$ and $y_3 = 29$. Then

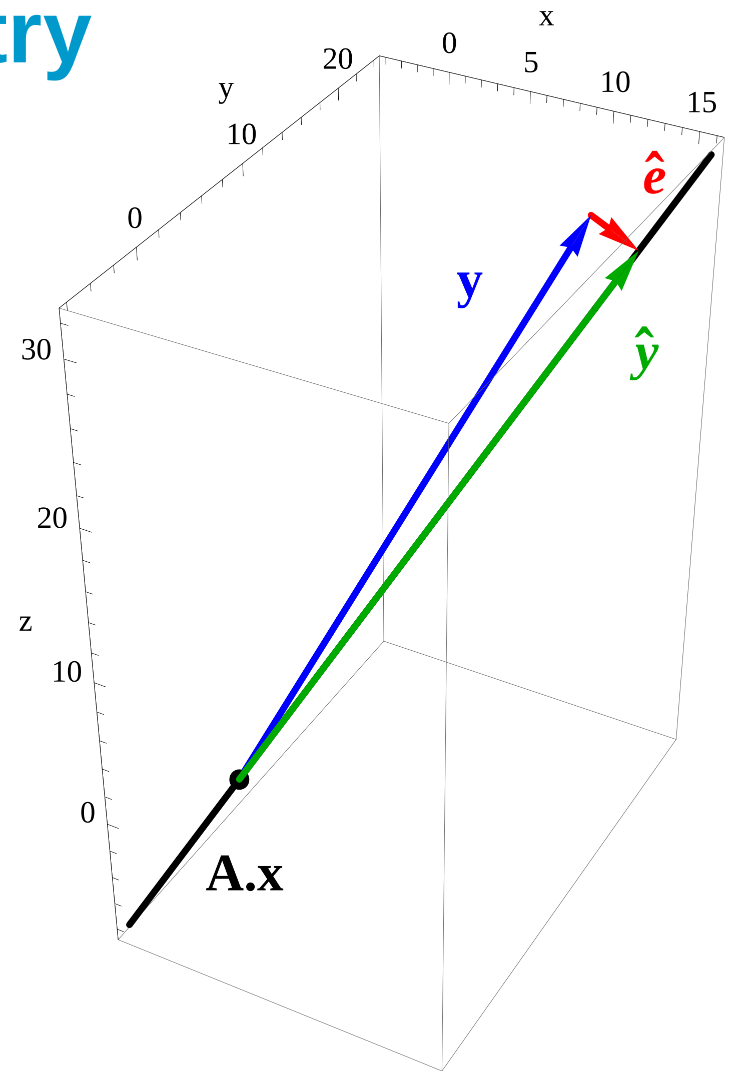
$$\begin{aligned} \mathbf{e}^T\mathbf{e}(v) &= (y_1 - 2v)^2 + (y_2 - 3v)^2 + (y_3 - 4v)^2 \\ &= 1362 - 396v + 29v^2 \end{aligned}$$





[Proof.] $\frac{d}{dv}\mathbf{e}^T\mathbf{e} = 58v - 396 = 0 \Rightarrow v \approx 6.83.$

Least Squares Geometry



Model is spanned by A. Every choice for parameter vector $\mathbf{x}$ corresponds to another location in the model space of A.

Observations $\mathbf{y}$ are generally not in the model space of A.

Goal: Find point $\hat{\mathbf{x}}$ in model space that minimizes distance to observations $\mathbf{y}$.

Solution: the perpendicular projection of $\mathbf{y}$ in the model space, the adjusted vector of observations $\hat{\mathbf{y}} = A \cdot \hat{\mathbf{x}}$

Shortest vector between observations $\mathbf{y}$ and adjusted observations $\hat{\mathbf{y}}$ is the error vector $\hat{\mathbf{e}}$: $\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$

Example: Errors in observed distances

[Proof.] We got $v \approx 6.83$. Therefore the corrected or adjusted distances are:

$$\hat{\mathbf{y}} = \begin{pmatrix} 13.66 \\ 20.48 \\ 27.31 \end{pmatrix}$$

And the errors, or the differences between observed and corrected distances:

$$\mathbf{e} = \mathbf{y} - A \cdot x = \begin{pmatrix} y_1 - 2v \\ y_2 - 3v \\ y_3 - 4v \end{pmatrix} = \begin{pmatrix} 11 - 2 \cdot 6.83 \\ 20 - 3 \cdot 6.83 \\ 29 - 4 \cdot 6.83 \end{pmatrix} = \begin{pmatrix} -2.66 \\ -.48 \\ 1.69 \end{pmatrix}$$

Least Squares Principle

Project vector of observations into model, that is, error vector $\hat{e}$ should be perpendicular to model space:

$$A^T \cdot \hat{\mathbf{e}} \quad = \quad 0$$

or

$$\begin{aligned} A^T \cdot (\mathbf{y} - A\hat{\mathbf{x}}) &= 0 \\ A^T\mathbf{y} - A^T A\hat{\mathbf{x}} &= 0 \\ A^T\mathbf{y} &= A^T A\hat{\mathbf{x}} \end{aligned}$$

Which gives the general solution

$$\hat{\mathbf{x}} \quad = \quad (A^T A)^{-1} A^T \mathbf{y}$$

Dept. of Geoscience & Remote Sensing

Least Squares - Calculus

Let $\mathbf{y} \approx A\mathbf{x}$ with A an $m \times n$ matrix of rank$(A) = n$. The least squares solution of the system $\mathbf{y} \approx A\mathbf{x}$ is defined as

$$\hat{\mathbf{x}} = \text{arg min}_{\mathbf{x} \in \mathbb{R}^n} (\mathbf{y} - A \cdot \mathbf{x})^T (\mathbf{y} - A \cdot \mathbf{x})$$

The difference between the vector of observations $\mathbf{y}$ and the adjusted vector of observations $\hat{\mathbf{y}} = A\hat{\mathbf{x}}$ is the least squares residual vector

$$\hat{\mathbf{e}} = \mathbf{y} - A \cdot \hat{\mathbf{x}}$$

The size of the squared norm

$$\|\hat{\mathbf{e}}\|^2 = \hat{\mathbf{e}}^T \cdot \hat{\mathbf{e}}$$

is a scalar measure for the inconsistency of the linear system.

Question What is $\hat{\mathrm{e}}$ in case the system is consistent?

Least Squares - Solution

The least squares solution of the linear system on the previous slide is given by

$$\hat{\mathbf{x}} \quad = \quad (A^T A)^{-1} A^T \mathbf{y}$$

[Proof.] Consider the function

$$\begin{aligned} f(\mathbf{x}) &= (\mathbf{y} - A\mathbf{x})^T (\mathbf{y} - A\mathbf{x}) \\ &= \mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T A\mathbf{x} + \mathbf{x}^T A^T A\mathbf{x} \end{aligned}$$

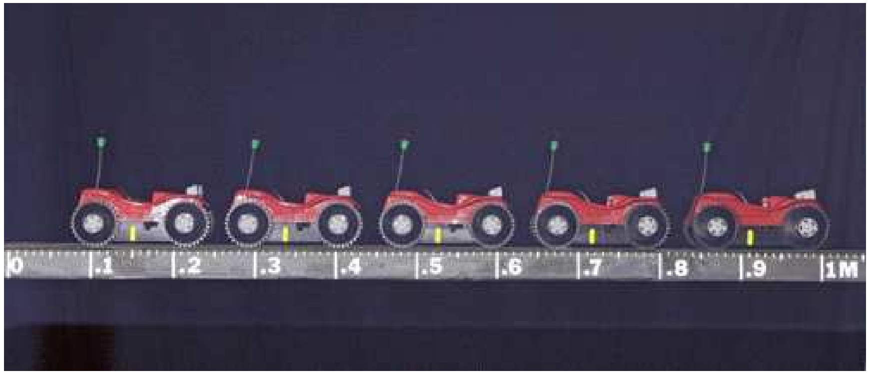
$\hat{\mathbf{x}}$ minimizes $f(\mathbf{x}) \Rightarrow$

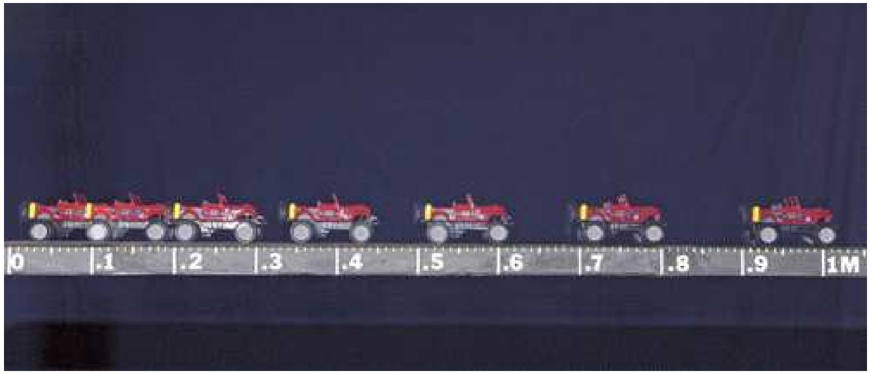
1. $\nabla f = 0$ ($\hat{\mathbf{x}}$ is a singular point).
2. The Hessian of f is positive definite. (The singular point is a minimum).

The gradient of $f(\mathbf{x})$ is given by: $\nabla f = -2A^T \mathbf{y} + 2A^T A\mathbf{x}$ and the Hessian of $f(\mathbf{x})$ by $H(f) = 2A^T A$. The $H(f)$ is a quadratic form and therefore positive definite for all x, so the solution of $\nabla f = 0$ is indeed the minimum.

$\nabla f = 0$ implies $A^T A\mathbf{x} = A^T \mathbf{y}$. Because the matrix $A^T A$ is positive definite, it is also invertible, so the solution follows as $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{y}$.

A more complicated A model





Source www.adformatie.nl

Source http://www.be.wednet.edu/Page/2720

Second Velocity Example

The position of a car is assumed to depend linearly on the unknown (constant) velocity v:

$$x_i \quad = \quad x_0 + v \cdot t_i \qquad\qquad i = 1, \ldots, m$$

with x_i - the distance measurement at time t_i, and
x_0 - the unknown initial position of the car.

How do we formulate the estimation of velocity and initial position as a least squares problem?

General Recipe

- What is your vector of observations?
- Identify the parameters that you want to estimate
- Examine how these parameters relate to the observations
- Use this relation to fix the model matrix A
- Check: are all dimensions OK?

Solution

[Proof.]

$$\mathbf{y} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \approx \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ v \end{pmatrix} = A \cdot \mathbf{x}$$

and the solution is given by $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{y}$

Numerical second velocity example

Assume for example: $x_1 = x_2 = 1$ at $t = -1, 1$ and,
$x_3 = 3$ at $t = 2$.

[Proof.]

$$\mathbf{y} = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} \approx \begin{pmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} x_0 \\ v \end{pmatrix} = A \cdot \mathbf{x}$$

Therefore

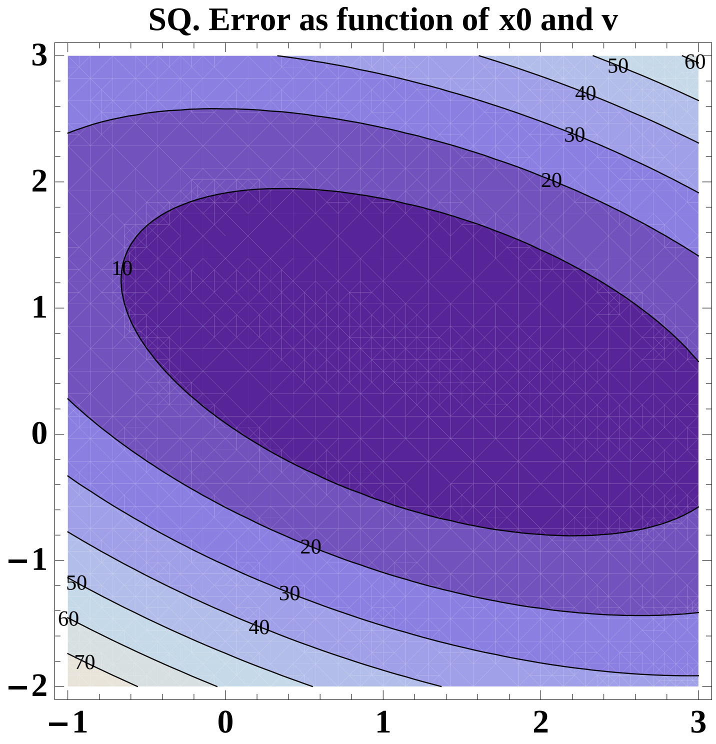
$$A^T A = \begin{pmatrix} 3 & 2 \\ 2 & 6 \end{pmatrix}, \quad \text{and} \quad A^T \mathbf{y} = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$$

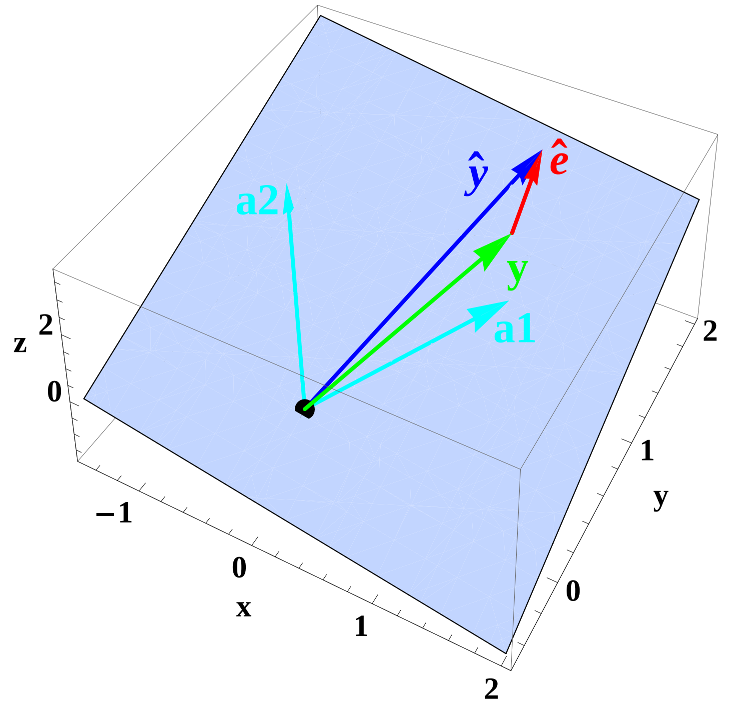
So,

$$\hat{\mathbf{x}} = \begin{pmatrix} \hat{x}_0 \\ \hat{v} \end{pmatrix} = \frac{1}{7} \begin{pmatrix} 9 \\ 4 \end{pmatrix}, \quad \hat{\mathbf{y}} = \frac{1}{7} \begin{pmatrix} 5 \\ 13 \\ 17 \end{pmatrix} \quad \text{and} \quad \hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}} = \frac{1}{7} \begin{pmatrix} 2 \\ -6 \\ 4 \end{pmatrix}$$

Note that indeed $\hat{\mathbf{e}}$ is orthogonal to the columns of matrix A!

Sum of squared errors





a_1: 1st column of matrix A

a_2: 2nd column of matrix A

Everything blue: in $\mathcal{R}(A)$

B. Plane Fitting

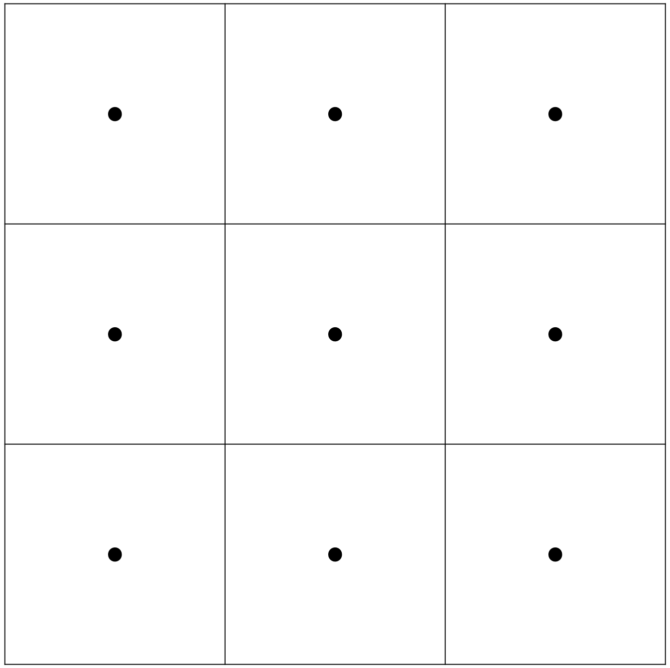
Local 3 x 3 window

Input:
raster file/regular grid of elevations

Output:

- Local normal
- Local roughness

Method:
Local planar fit



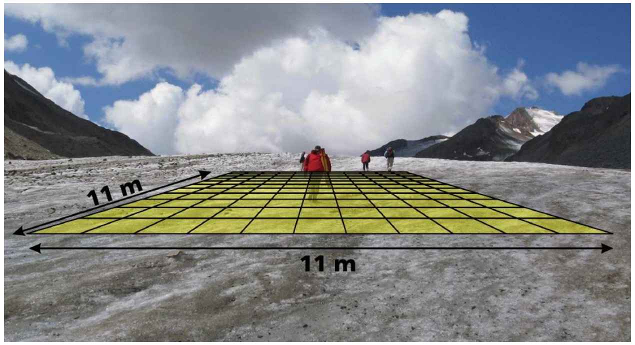
Question. What is the effect of choosing a larger (e.g. 6×6) window size?

Question. How do you get the local normal?

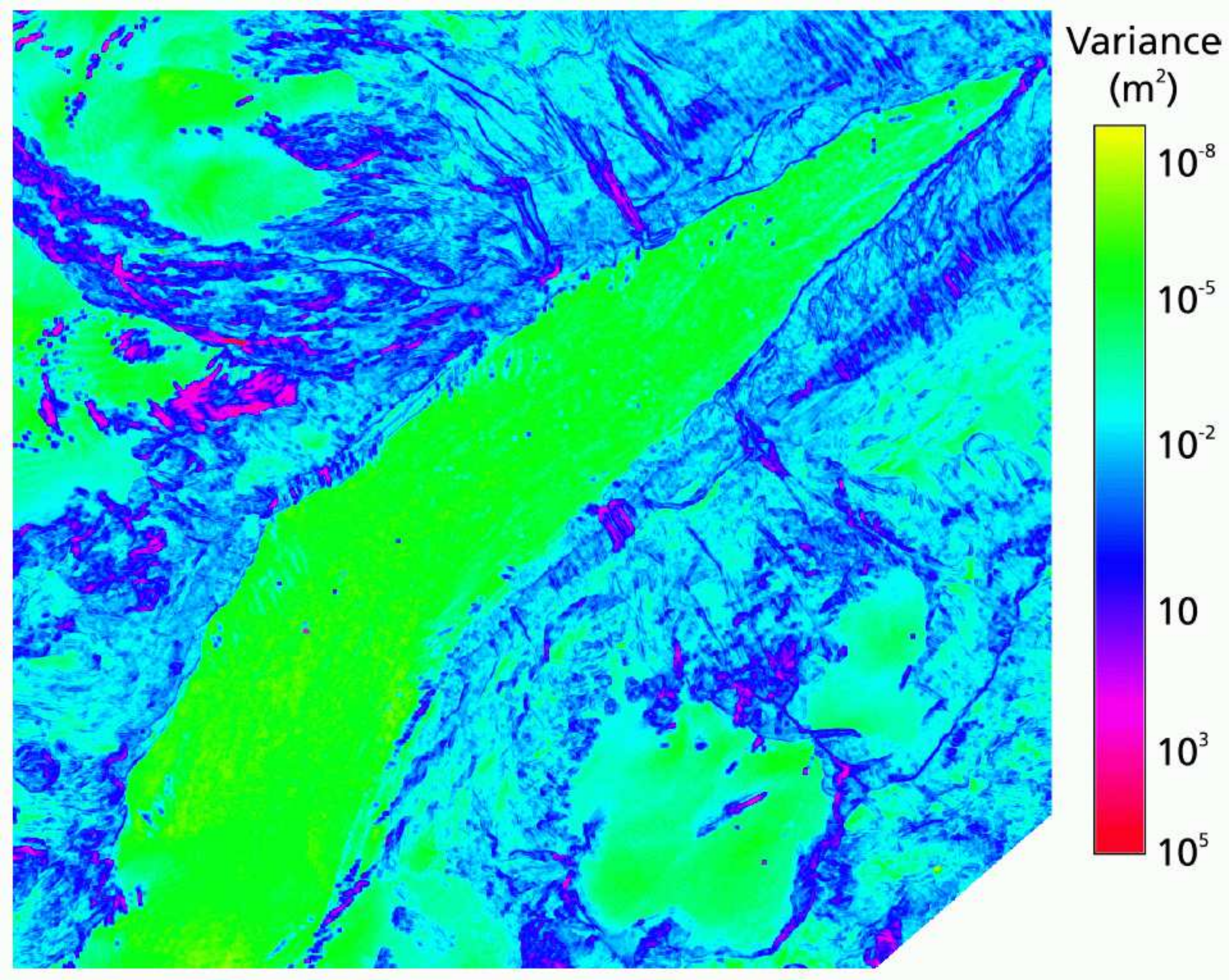
Question. How do you get the/a local roughnes

Variance or Glacier Roughness

Hintereisferner - Tirol

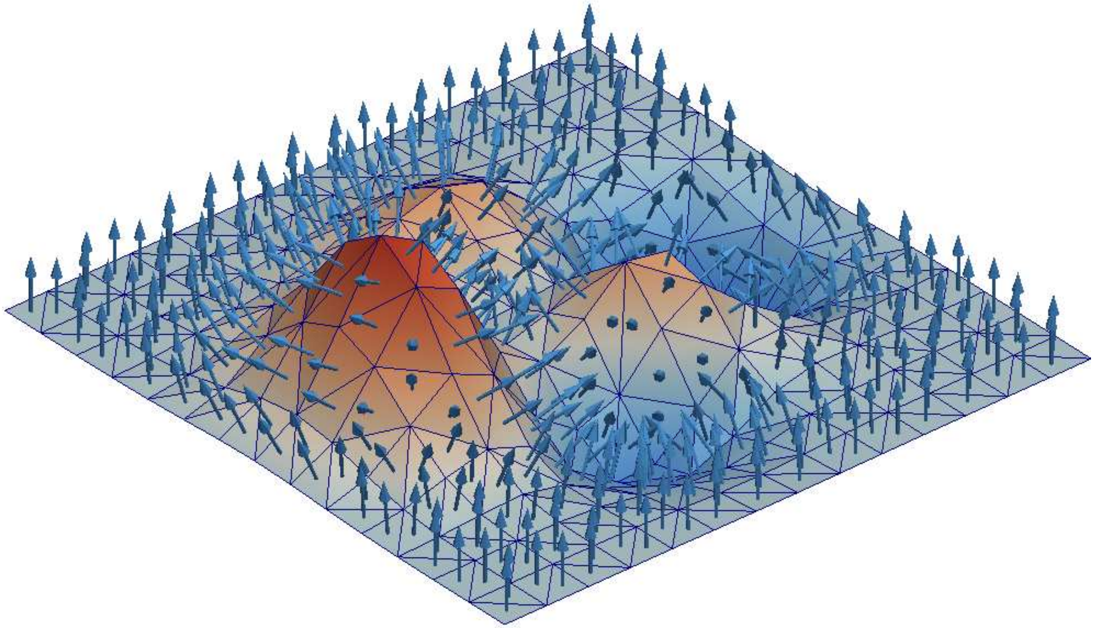


Variance window



(Images: Martin Kodde)

Local normals



Here computed for a simulated TIN (Triangulated Irregular Network)

Source http://www.mathworks.com/matlabcentral/fileexchange/authors/37194

Example, 3x3 window

Given: Elevation data $\Rightarrow$

	1	2	3
3	7.7	7.9	8.3
2	7.6	7.4	7.5
1	7.2	7.2	6.7

Obtain an equation of the best fitting plane

$$z = a \cdot x + b \cdot y + c$$

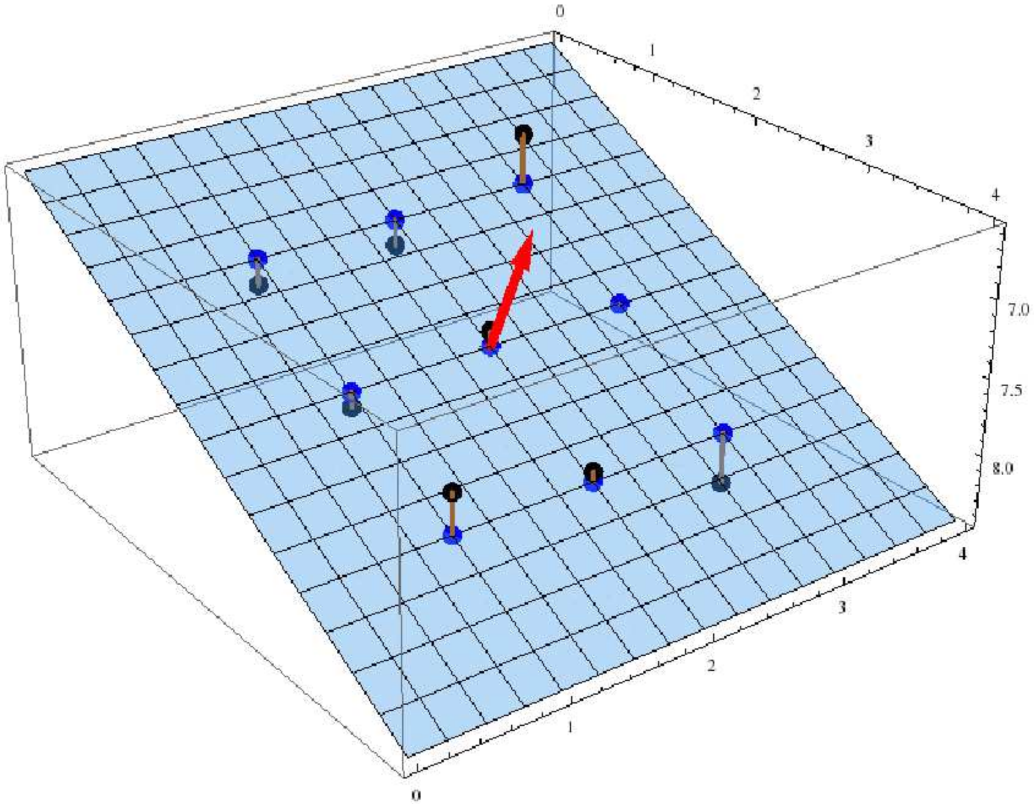
From that, derive:

1. Normal: vector $\mathbf{n} = (a, b, -1)^T$
2. Rougness: deviations from the local plane

Exercise

Obtain the plane equations, the normal and the roughness using the least squares framework

Resulting plane



Black points: input elevations

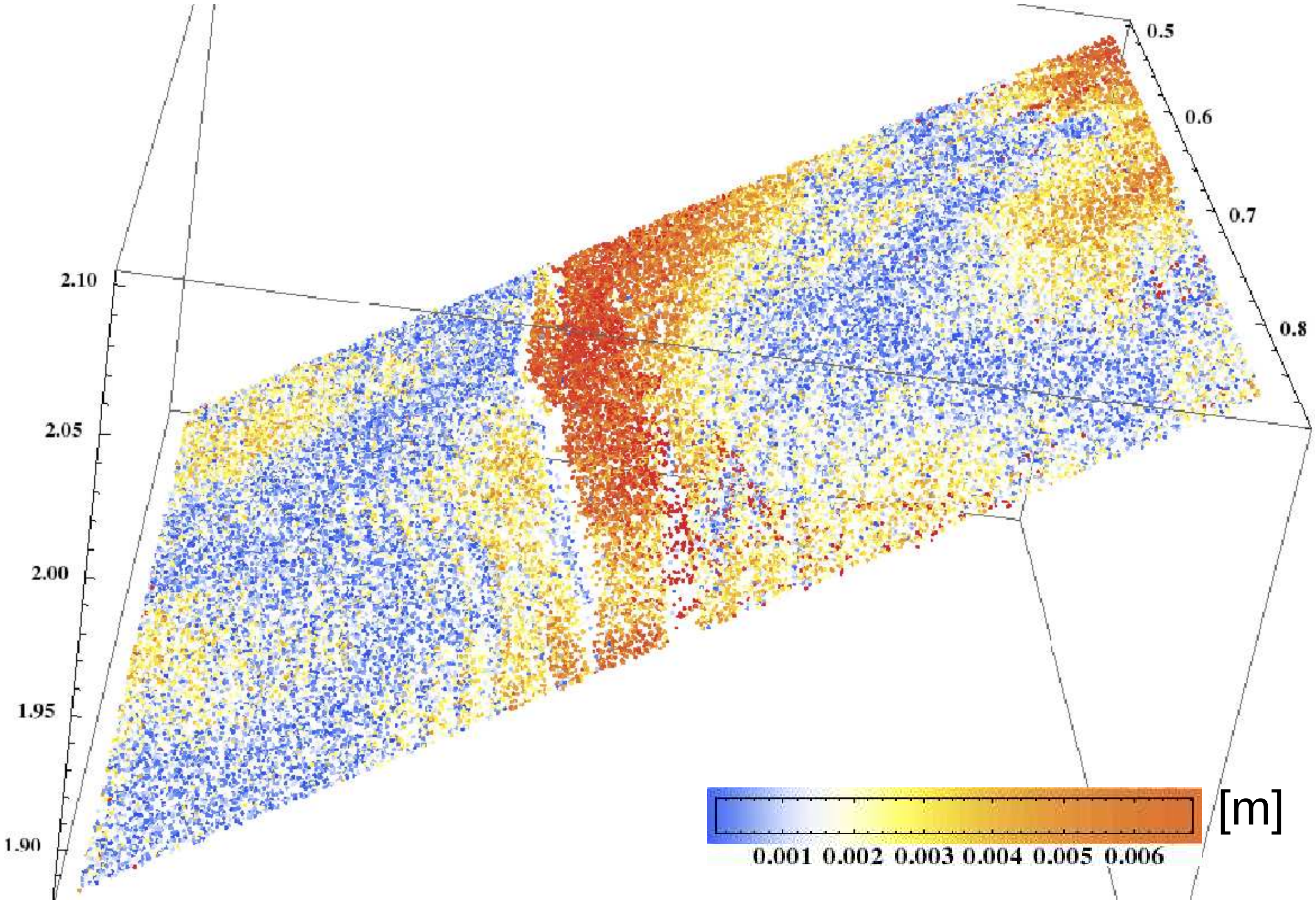
Blue points: adjusted elevations

Blue plane: plane fitting the elevations best in the least squares sense

Brown lines: elevation residuals

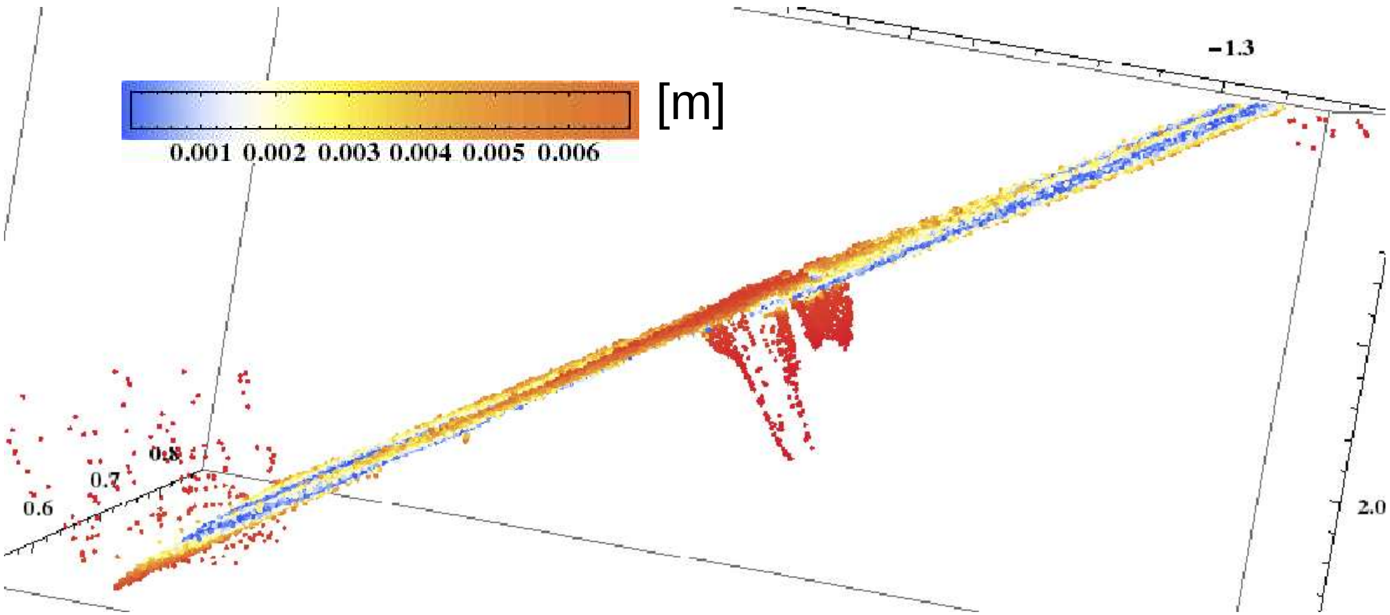
Red arrow: surface normal

Results, planar fit, wooden beam



Plot shows beam points, adjusted to plane by least squares
The coloring indicates the absolute size of the error e_i

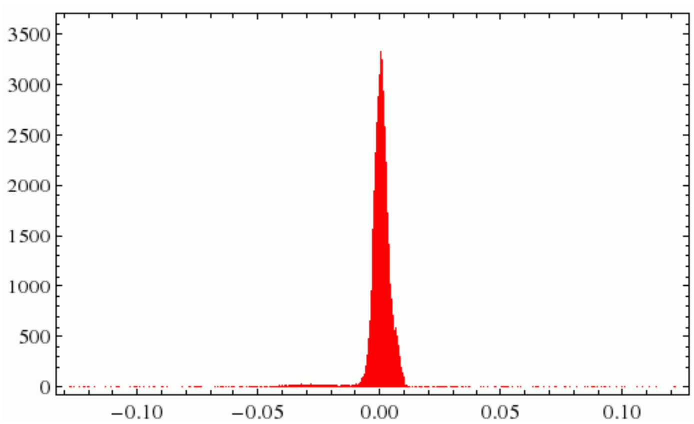
3D Results, planar fit, wooden beam

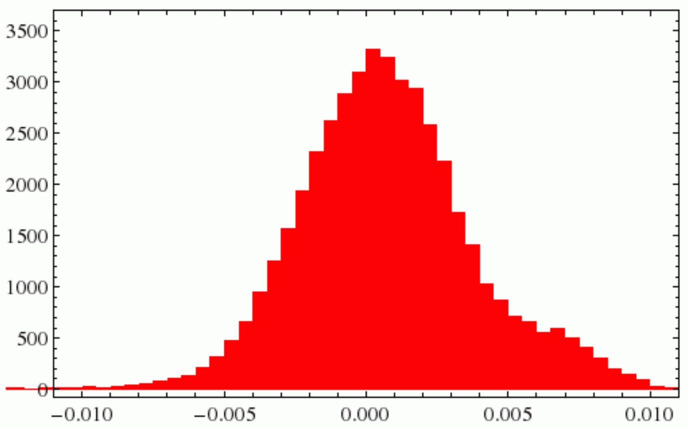


Plot shows original 3D beam points, with coloring indicating the absolute size of the error e_i.

Question. What 3 choices did the fitter had to choice as vector of observations?

Distribution of LSQ adjustment errors





Question. Why are the errors distributed around zero?

Question. Why is the (cropped) histogram not symmetric? Are the residuals normally distributed?

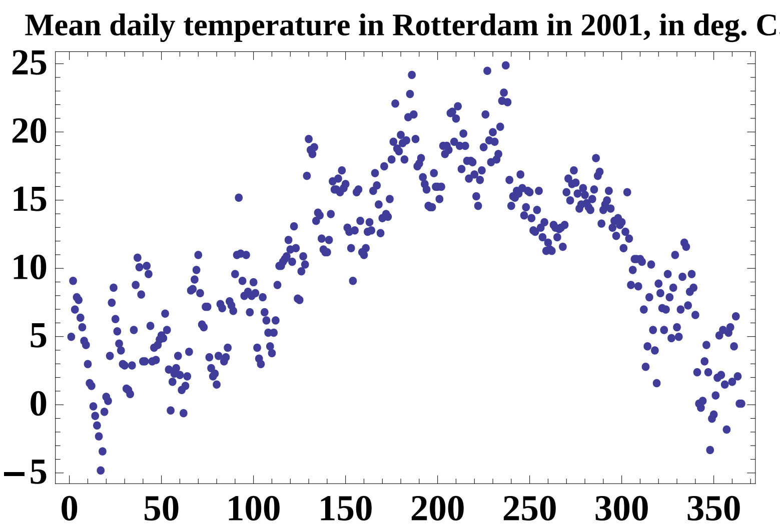
Question. What does the histogram tell us on the quality of the measurements?

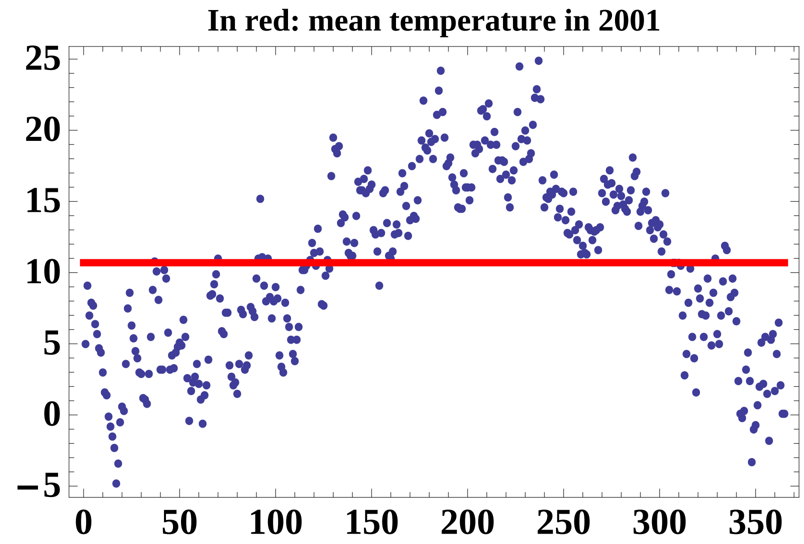
Question. How to determine the sample standard deviation of the residuals?

C. Least Squares Curve fitting

Dept. of Geoscience & Remote Sensing

Introduction Weather data





Question: How do we use Least Squares to estimate the mean temperature in 2001?

[Proof.] For $i = 1, \ldots, 365$ write

$$y_i \quad = \quad 1 \cdot x,$$

so, $\mathbf{y} = \{y_1, \ldots, y_{365}\}^T$, $\mathbf{x} = \{x\}$, and $A = \{1, 1, \ldots, 1\}^T$. Now $\bar{x} = (A^T A)^{-1} A^T \mathbf{y} = 10.6^\circ$.

Fitting polynomials

How can we fit a general (not necessarily linear) polynomial to the weather data?

Write, with $T_i = y_i$, the temperature at day $t_i = 1, \ldots, 365$,

$$T_i \quad = \quad a_0 + a_1 t_i + a_2 t_i^2 + a_3 t_i^3 + \ldots$$

Question: what is the vector of observations, the parameter vector and the A-matrix?

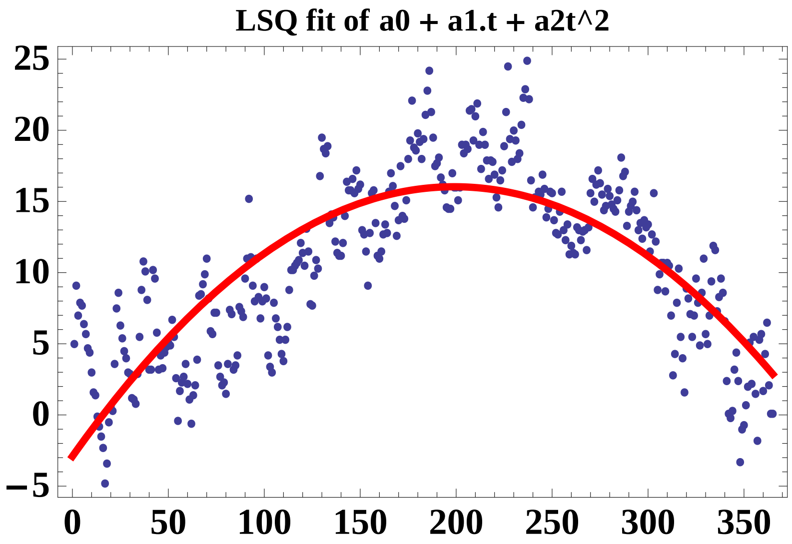
[Proof.] so, $\mathbf{y} = \{T_1, T_2, \ldots, T_{365}\}^T$, $\mathbf{x} = \{a_0, a_1, a_2, \ldots, \}^T$ and

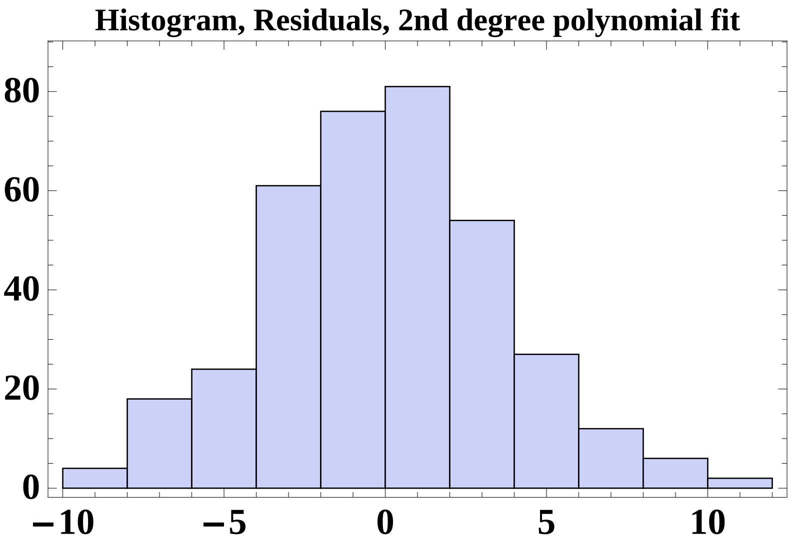
$$\begin{aligned} A_i &= \{1, t_i, t_i^2, t_i^3, \ldots\}, \qquad \text{for } i = 1, \ldots, 365 \\ &= \{1, i, i^2, i^3, \ldots\} \end{aligned}$$

Question: what is the redundancy of the resulting least squares system?

[Proof.] Nr. observations - Nr. parameters = 365 - (degree of polynomial + 1)

Result: polynomial of degree 2





$\sqrt{(\mathbf{e}^T\mathbf{e})/365}$ = 3.7

$\hat{\mathbf{x}} \approx (-2.90, 0.19, -0.00048)^T$

Question: Other method to check whether this A-model fits well?

General curve fitting

Suppose $f(t)$ is an unknown function of which we measure m function values

$$y_i \approx f(t_i) \qquad\qquad i = 1, 2, \ldots, m$$

It is considered known that $f(t)$ can be written as a linear combination of n base functions:

$$f(t) \quad = \quad \sum_{j=1}^{n} c_j \phi_j(t),$$

with c_j the coefficient that expresses the contribution of base function $\phi_j(t)$ to $f(t)$.

The vector of coefficients

$$\mathbf{x} \quad = \quad \{c_1, c_2, \ldots, c_n\}$$

is estimated by solving the following Least Squares system:

Estimating base coefficients

Solve $\mathbf{y} = A\mathbf{x}$, given by

$$\begin{pmatrix} f(t_1) \\ f(t_2) \\ \vdots \\ f(t_m) \end{pmatrix} = \begin{pmatrix} \phi_1(t_1) & \phi_2(t_1) & \dots & \phi_n(t_1) \\ \phi_1(t_2) & \phi_2(t_2) & \dots & \phi_n(t_2) \\ \vdots & \vdots & & \vdots \\ \phi_1(t_i) & \phi_2(t_i) & \dots & \phi_n(t_i) \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

Fitting trigonometric function

How to fit

$$h(t) = a_0 + a_1 \cos(\omega t) + a_2 \sin(\omega t) :$$

to the annual weather data:

1. What is ω?
2. What value of ω should we take?
3. What is the A-model?

[Proof.]

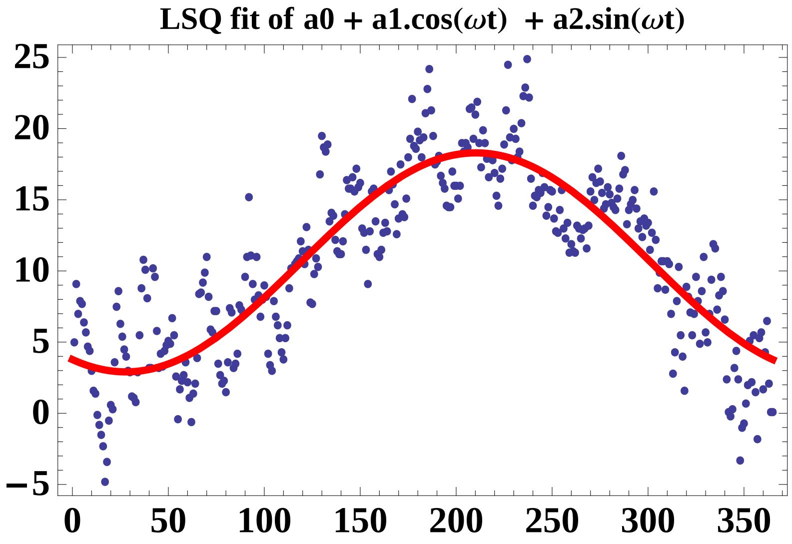
1. ω is the ‘angular velocity‘ and determines the period of the trigonometric function. If $\omega = 1$, the period T is 2π. So, in general $T = (2\pi)/\omega$.
2. For the weather data, we have $T = 365$. So we should take

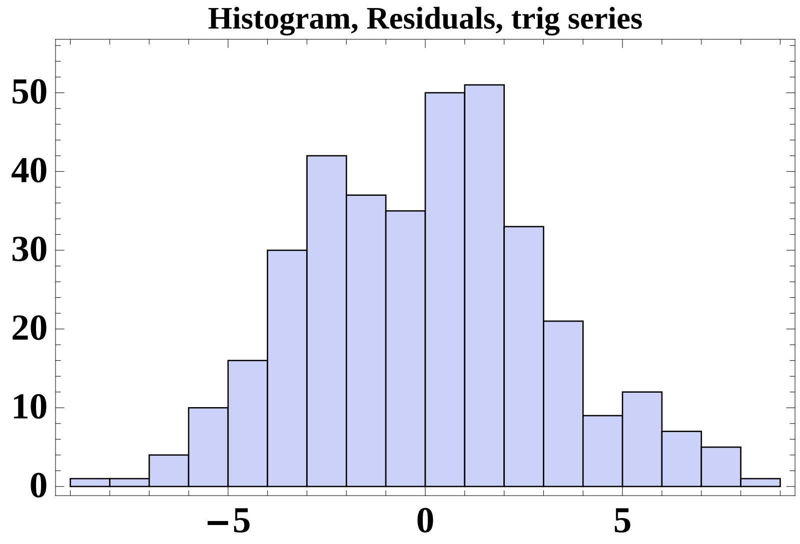
$$\omega = (2\pi)/T = (2\pi)/365 = \pi/182.5.$$

3. The i-th column of the A-matrix reads:

$$A_i \;=\; (1, \cos(\omega i), \sin(\omega i))$$

Trigonometric fit results





$$\sqrt{(\mathbf{e}^T\mathbf{e})/365} = 3.04$$

$$\hat{\mathbf{x}} \approx (10.6, -6.83, -3.54)^T$$

Trigo questions

Question: Why is a_0 similar to the mean temperature in 2001?

[Proof.] Model is the mean + seasonal deviation from the mean

Question: Why can we conclude in this special case that our trigonometric model fits better than the polynomial of degree 2?

[Proof.] Redundancy is the same

Question: can we also obtain ω using least squares fitting?

[Proof.] No, we used an ad-hoc method to obtain a value for ω

D. Texel case study

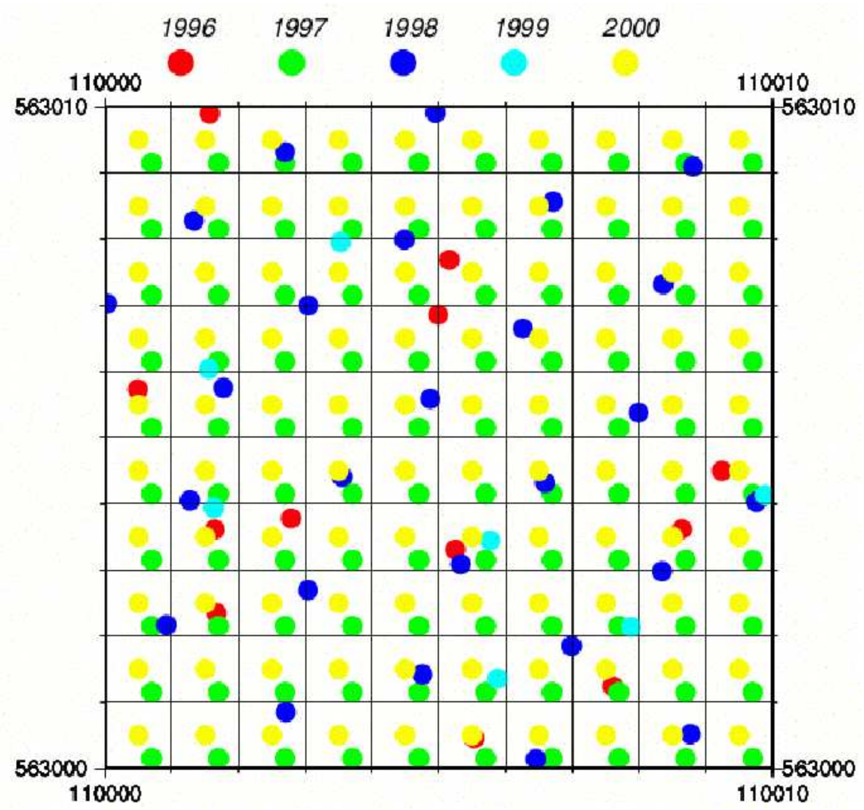
Case study: South-West Texel

Airborne laser data from 1996 to 2001 covering beach and dunes.

year	# heights
1996	224 521
1997	3 577 200
1998	1 054 817
1999	607 860
2000	2 914 654
2001	2 810 642

Source: www.kustfoto.nl

Different ground-points in distinct years.



Grid-point wise analysis after interpolation to regular grid.

Single position modeling.

Assume that for every grid point position (x, y) a full observation vector $\mathbf{h} \in I\!R^m$ is given

We look for a linear model A such that

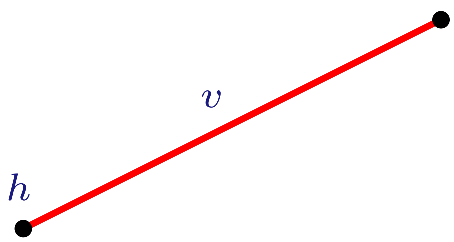
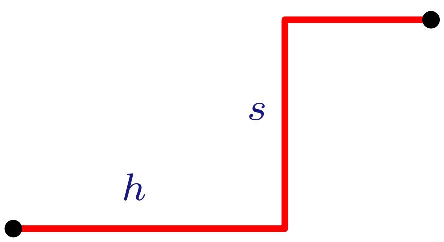
$$\mathbf{h} = \begin{pmatrix} h_{96} \\ h_{97} \\ \vdots \\ h_{01} \end{pmatrix}, \qquad E\{\mathbf{h}\} = A \cdot \mathbf{x},$$

where $\mathbf{x} \in I\!R^n$ denotes the vector of model parameters.

Model expressing no deformation

Question: what A model expresses stability?

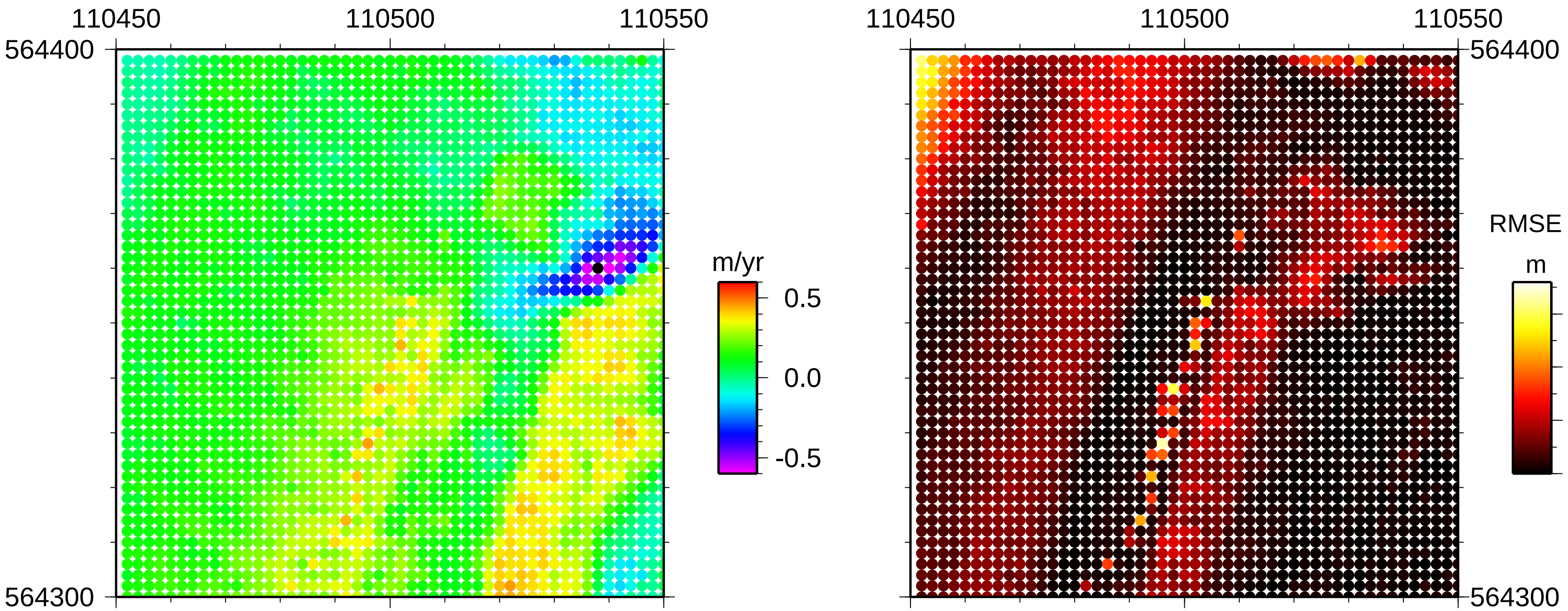
Kinematic model alternatives.

Constant velocity model	Instantaneous suppletion model
	
$A = \begin{pmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_m & 1 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} v \\ h \end{pmatrix};$	$A = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix}, \mathbf{x} = \begin{pmatrix} h \\ s \end{pmatrix}.$

h = height; v = velocity; s = suppletion.

Question: Other 2-parameter models??

Example: 2500 velocities and residuals.



Number of 'observations' per grid-point: $m = 7$

Number of model parameters: $n = 2$

Redundancy: $q := m - n = 5$.

Conclusions

Plane fitting

- Easy starting point for e.g. terrain analysis
- Gives surface approximation, orientation and a measure of roughness
- Use a lot or less points for a fit

Least Squares

- Method to incorporate redundant observations in geometric fitting
- Minimizes error in the least squares sense
- Model used to fit may be more or less appropriate
- So far, didn’t incorporate quality of observations

Exercises

Exercise 6.1 Find the least-squares solution of the following two linear systems of equations:

$$\begin{aligned} x_1 + 2x_2 &= 5 \\ 2x_1 - x_2 &= 0 \\ 5x_1 + x_2 &= 6 \end{aligned} \qquad \begin{aligned} 2x_1 + 4x_2 - 3x_3 &= 8 \\ x_1 + x_2 - x_3 &= 3 \\ 2x_1 - 2x_2 + 3x_3 &= -1 \end{aligned}$$

Exercise 6.2 Consider the linear system of equations $y \approx Ax$ with

matrix $A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{pmatrix}$, and vector $y = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$

Find the least-squares solution $\hat{x}$ and the projection $\hat{y}$ of vector y onto the column space of A.

Exercises

Exercise 6.3 Consider the inconsistent linear system of equations $y \approx Ax$ with

$$\text{matrix } A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad \text{and vector } y = \begin{pmatrix} 4 \\ 5 \\ 9 \end{pmatrix}$$

Let $\hat{y}$ be the projection of vector y onto the column space of A, i.e., the least-squares solution of the 'measurements' y. Find the least-squares residual vector $\hat{e} = y - \hat{y}$.

Exercise 6.4 An object is moving along a straight line. The following measurements y_i of the object's position have been made at corresponding times t_i.

Time t_i, in [s]	-1	0	1	2
Position y_i, in [m]	2	0	-3	-6

To the data a parabolic model $y = x_0 + vt + \frac{1}{2}at^2$ is fitted using least squares. What is the (unweighted) least squares solution for the acceleration a?

Exercises

Exercise 6.5 An object is moving along a straight line. The following measurements y_i of the object's position have been made at corresponding times t_i.

Time t_i, in [s]	-1	0	1	2
Position y_i, in [m]	-2	0	3	5

To the data a linear model $y = x_0 + vt$ is fitted using least squares. Find the (unweighted) least squares solution for the speed v?

Exercises

Exercise 6.6 The monthly sales of a certain product are subject to seasonal fluctuations. The sales data might be modeled by a function of the form

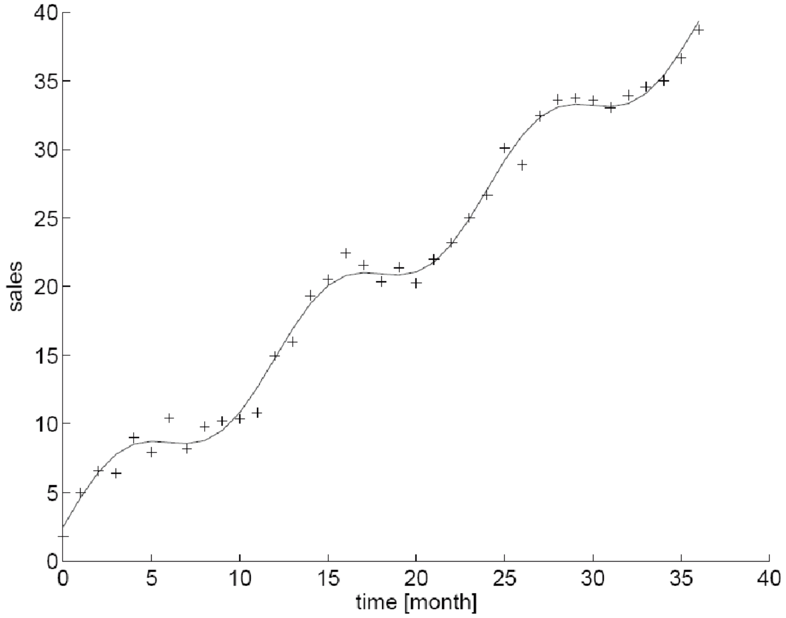
$$y = x_1 + x_2(t) + x_3 \sin(2\pi t/12)$$

with t the time in months. The term $x_1 + x_2 t$ gives the basic sales trend, whereas the term $x_3 \sin(2\pi t/12)$ is representing the seasonal changes, see figure below.

Assume that sales data y_i are available at times $t_1, t_2, \ldots, t_m$. Find the design matrix of the linear system of equations that leads to a least-squares fit of

$$y = x_1 + x_2(t) + x_3 \sin(2\pi t/12)$$

to the sales data.



Dept. of Geoscience & Remote Sensing

Exercises

Exercise 6.7 Consider an airplane taking laser height measurements. In part of the surveyed area gas is extracted from the subsurface. The airplane flies along a straight line. It takes measurements at positions $x = 0, 2, 4, 6$ and 8 km along this line. The measurements taken are distances h_i from the airplane to the ground and are listed below:

x_i (m)	0	2000	4000	6000	8000
h_i (m)	50.334	50.595	51.144	55.226	58.648

From independent measurements it is known that the area is flat from 0 m to $x_{\text{start}} = 3000m$. From x_{start} on the area is subsiding due to the gas extraction. This subsidence shows a linear behavior,

$$h = c_0 + c_1(x - x_{\text{start}}).$$

Apply least-squares curve fitting to the data to determine the slope c_1 in the area $x > x_{\text{start}}$.

Exercise, plane fitting

Exercise 6.8 In this exercise we will fit a plane using least squares to the following 3D points:

$$p_1 = (1, 1, 7.2),\, p_2 = (2, 1, 7.2),\, p_3 = (3, 1, 6.7),\, p_4 = (1, 2, 7.6),\, p_5 = (2, 2, 7.4),$$
$$p_6 = (3, 2, 7.5),\, p_7 = (1, 3, 7.7),\, p_8 = (2, 3, 7.9),\, p_9 = (3, 3, 8.3)$$

Compare also Slides 29 and 30.

a). What is the number of observations, and how many parameters need to be estimated?

b). What is the vector of observations $\mathbf{y}$ in this case?

c). What is the vector of (yet unknown) plane parameters $\mathbf{x}$?

d). Rewrite the plane equation $z = a \cdot x + b \cdot y + c$ as an inner product $z = r_A \cdot (a, b, c)$. So notably, how does the vector r_A looks like?

The only ingredient still needed for a least squares fit is the model matrix A. The i-th row of this 9×3 matrix is a copy of the vector r_A corresponding to the i-th observation. For example, the sixth row of the A matrix looks like $A_6 = (3, 2, 1)$.

e). Write down the full model matrix A.

f). Now get the plane parameters $\hat{\mathbf{x}}$ by solving

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{y}$$

Hint: this can be easily done in Matlab: http://www.mathworks.nl/help/matlab/ref/mldivide.html

Exercise, Plane fitting, continued

Exercise 6.9 This exercise is a continuation of Exercise 6.8. In the following we will use the plane parameters to evaluate the quality of fit and to obtain a normal of the plane.

a). Determine the vector of adjusted observations

$$\hat{\mathbf{y}} = A \cdot \hat{\mathbf{x}}$$

b). Where are the adjusted observations situated?

c). Determine the vector of residuals $\hat{\mathbf{e}}$, i.e. the distances between observations and adjusted observations:

$$\hat{\mathbf{e}} = \mathbf{y} - \hat{\mathbf{y}}$$

d). Determine the standard deviation of your plane fit result by determininig

$$\sigma = \frac{\sqrt{\hat{\mathbf{e}}^T \cdot \hat{\mathbf{e}}}}{n},$$

where n denotes the number of observations.

e). Determine a normal $\mathbf{n}_P$ of the plane from the plane parameter vector $\hat{\mathbf{x}}$.

f). Verify that $\mathbf{n}_P$ is indeed a normal by evaluating the inner product with two independent vectors in the plane P. Take for example two suitable difference vectors between adjusted observations.

Answers, Exercise 6.1

Find the least-squares solution of the following two linear systems of equations:

$$
\begin{aligned}
x_1 + 2x_2 &= 5 \\
2x_1 - x_2 &= 0 \\
5x_1 + x_2 &= 6
\end{aligned}
\qquad\qquad
\begin{aligned}
2x_1 + 4x_2 - 3x_3 &= 8 \\
x_1 + x_2 - x_3 &= 3 \\
2x_1 - 2x_2 + 3x_3 &= -1
\end{aligned}
$$

[Proof.] 1st system

$$
A = \begin{pmatrix} 1 & 2 \\ 2 & -1 \\ 5 & 1 \end{pmatrix}, \mathbf{y} = \begin{pmatrix} 5 \\ 0 \\ 6 \end{pmatrix}, A^T A = \begin{pmatrix} 30 & 5 \\ 5 & 6 \end{pmatrix}, \quad A^T \mathbf{y} = \begin{pmatrix} 35 \\ 16 \end{pmatrix}
$$

$$
\begin{pmatrix} \hat{x}_1 \\ \hat{x}_1 \end{pmatrix} = (A^T A)^{-1} A^T \mathbf{y} = \begin{pmatrix} 0.84 \\ 1.97 \end{pmatrix}
$$

Answer, 2nd system

[Proof.]

$$A = \begin{pmatrix} 2 & 4 & -3 \\ 1 & 1 & -1 \\ 2 & -2 & 3 \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} 8 \\ 3 \\ -1 \end{pmatrix}$$

So

$$A^T A = \begin{pmatrix} 9 & 5 & -1 \\ 5 & 21 & -19 \\ -1 & -19 & 19 \end{pmatrix}, \qquad A^T \mathbf{y} = \begin{pmatrix} 17 \\ 37 \\ -30 \end{pmatrix}$$

Therefore

$$\begin{pmatrix} \hat{x}_1 \\ \hat{x}_1 \end{pmatrix} = (A^T A)^{-1} A^T \mathbf{y} = \begin{pmatrix} 1.5 \\ 0.5 \\ -1 \end{pmatrix}$$

Answers, Exercise 6.2

Consider the linear system of equations $y \approx Ax$ with

$$\text{matrix } A = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ 0 & 0 \end{pmatrix}, \qquad \text{and vector } y = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}$$

Find the least-squares solution $\hat{x}$ and the projection $\hat{y}$ of vector y onto the column space of A.

[Proof.]

$$A^T A = \begin{pmatrix} 2 & 5 \\ 5 & 13 \end{pmatrix}, \qquad A^T = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 3 & 0 \end{pmatrix}, \qquad A^T \mathbf{y} = \begin{pmatrix} 9 \\ 23 \end{pmatrix}$$

Therefore

$$\begin{pmatrix} \hat{x}_1 \\ \hat{x}_1 \end{pmatrix} = (A^T A)^{-1} A^T \mathbf{y} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \qquad \text{and} \qquad \hat{\mathbf{y}} = \mathbf{y} - A\hat{\mathbf{x}} = \begin{pmatrix} 4 \\ 5 \\ 0 \end{pmatrix}$$

Answers, Exercise 6.3

Consider the inconsistent linear system of equations $y \approx Ax$ with

$$\text{matrix } A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad \text{and vector } y = \begin{pmatrix} 4 \\ 5 \\ 9 \end{pmatrix}$$

Let $\hat{y}$ be the projection of vector y onto the column space of A, i.e., the least-squares solution of the 'measurements' y. Find the least-squares residual vector $\hat{e} = y - \hat{y}$.

[Proof.]

$$\begin{pmatrix} \hat{x}_1 \\ \hat{x}_1 \end{pmatrix} = \begin{pmatrix} 6 \\ 2.5 \end{pmatrix}, \qquad \hat{\mathbf{y}} = \begin{pmatrix} 3.5 \\ 6 \\ 8.5 \end{pmatrix}, \qquad \hat{\mathbf{e}} = \begin{pmatrix} 0.5 \\ -1 \\ 0.5 \end{pmatrix}$$

Dept. of Geoscience & Remote Sensing

Answers, Exercise 6.4

An object is moving along a straight line. The following measurements y_i of the object's position have been made at corresponding times t_i.

Time t_i, in [s]	-1	0	1	2
Position y_i, in [m]	2	0	-3	-6

To the data a parabolic model $y = x_0 + vt + \frac{1}{2}at^2$ is fitted using least squares. What is the (unweighted) least squares solution for the acceleration a?

[Proof.]

$$\begin{cases} 2 &= x_0 + (-1)v + \frac{1}{2}(-1)^2 a \\ 0 &= x_0 + 0v + \frac{1}{2}(0)^2 a \\ -3 &= x_0 + 1v + \frac{1}{2}(1)^2 a \\ -6 &= x_0 + 2v + \frac{1}{2}(2)^2 a \end{cases}$$

PTO

Answers, Exercise 6.4, continued

[Proof.] So,

$$A = \begin{pmatrix} 1 & -1 & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 1 & \frac{1}{2} \\ 1 & 2 & 2 \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} 2 \\ 0 \\ -3 \\ -6 \end{pmatrix} \qquad \hat{\mathbf{x}} = \begin{pmatrix} \hat{x}_0 \\ \hat{v} \\ \hat{a} \end{pmatrix} = \begin{pmatrix} -.15 \\ -2.45 \\ -.5 \end{pmatrix}$$

Answers, Exercise 6.5

An object is moving along a straight line. The following measurements y_i of the object's position have been made at corresponding times t_i.

Time t_i, in [s]	-1	0	1	2
Position y_i, in [m]	-2	0	3	5

To the data a linear model $y = x_0 + vt$ is fitted using least squares. Find the (unweighted) least squares solution for the speed v?

[Proof.]

$$\begin{cases} -2 & = & x_0 + (-1)v \\ 0 & = & x_0 + 0v \\ 3 & = & x_0 + 1v \\ 5 & = & x_0 + 2v \end{cases}$$

So $\mathbf{y} = (-2, 0, 3, 5)^T$ and $A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$. Therefore $\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{y} = (0.3, 2.4)^T$.

We are looking for the speed, which is the 2nd parameter, so the answer is $v = 2.4$.

Answers, Exercise 6.6

The monthly sales... Find the design matrix ... to the sales data

$$A = \begin{pmatrix} 1 & t_1 & \sin(2\pi t_1/12) \\ 1 & t_2 & \sin(2\pi t_2/12) \\ \vdots & & \\ 1 & t_m & \sin(2\pi t_m/12) \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_m \end{pmatrix}, \text{ and } \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

Answers, Exercise 6.7

Consider an airplane taking laser height measurements. In part of the surveyed area gas is extracted from the subsurface. The airplane flies along a straight line. It takes measurements at positions $x = 0, 2, 4, 6$ and 8 km along this line. The measurements taken are distances h_i from the airplane to the ground and are listed below:

x_i (m)	0	2000	4000	6000	8000
h_i (m)	50.334	50.595	51.144	55.226	58.648

From independent measurements it is known that the area is flat from 0 m to $x_{\text{start}} = 3000m$. From x_{start} on the area is subsiding due to the gas extraction. This subsidence shows a linear behavior, $h = c_0 + c_1(x - x_{\text{start}})$. Apply least-squares curve fitting to the data to determine the slope c_1 in the area $x > x_{\text{start}}$.

Matlab answer, Exercise 6.7:

```
close all
clear all

A = [1 0; 1 0; 1 1000; 1 3000; 1 5000]
y = [50.334; 50.595; 51.144; 55.226; 58.648]

xhat = inv((A'*A))*(A'*y);
yhat = A*xhat;
ehat = y-yhat;

c0 = xhat(1,:)
c1 = xhat(2,:)
```

Answers, Exercise 6.8

a) Observations: 9; Nr. of parameters: 3, as you can write a plane as
$z = ax + by + c$.
Parameters to be estimated: a, b and c.

b) Vector of observations:
vecy = (7.2, 7.2 ,6.7, 7.6, 7.4, 7.5 ,7.7 ,7.9 ,8.3).

c) Vector of paramaters
vecx = (a, b, c).

d) r_A = (x, y, 1).

e) ->

```
vecy = {7.2, 7.2, 6.7, 7.6, 7.4, 7.5, 7.7, 7.9, 8.3};
matA = {{1, 1, 1}, {2, 1, 1}, {3, 1, 1}, {1, 2, 1},
  {2, 2, 1}, {3, 2, 1}, {1, 3, 1}, {2, 3, 1}, {3, 3, 1}};
MatrixForm[
 %]
```

trixForm=

$$\begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \\ 3 & 1 & 1 \\ 1 & 2 & 1 \\ 2 & 2 & 1 \\ 3 & 2 & 1 \\ 1 & 3 & 1 \\ 2 & 3 & 1 \\ 3 & 3 & 1 \end{pmatrix}$$

f) vecxhat =
Inverse(Transpose(matA).matA)).Transpose(matA).vecy
is
(0, 0.466667, 6.56667)

Answers, Exercise 6.9

a) vecyhat = matA.vecxhat gives
vecyhat = (7.03, 7.03, 7.03, 7.5, 7.5, 7.5, 7.97, 7.97, 7.97)

b) The adjusted observations are inside the determined plane

c) ehat = vecy - vecyhat, gives
ehat = (0.17, 0.17, -0.33, 0.1, -0.1, 0, -0.27, -0.07, 0.33).

d) There is a mistake in this exercise, the sqrt should be taken of everything, so

$$\sigma = \sqrt{\frac{\mathbf{e} \cdot \mathbf{e}}{n}}$$

The resulting st.dev is therefore $\sigma = 0.20$

e) A normal of the plane is given by (a, b, -1):
Coefficient c is just the offset from the origin. If c=0, it would follow that $ax + by - z = 0$. So, $(a, b, -1)$ is perpendicular to each point (x, y, z) fullfilling the plane equation.
So a normal is: $(0, 0.466667, -1)$

f) Take for example q1, q2, q4, the adjusted versions of p1, p2 and p4.
$q1 = (1, 1, 7.033)$; $q2 = (2, 1, 7.033)$ and $q3 = (1, 2, 7.5)$
normal.(q2-q1) = 0 and normal.(q4-q1) = 0