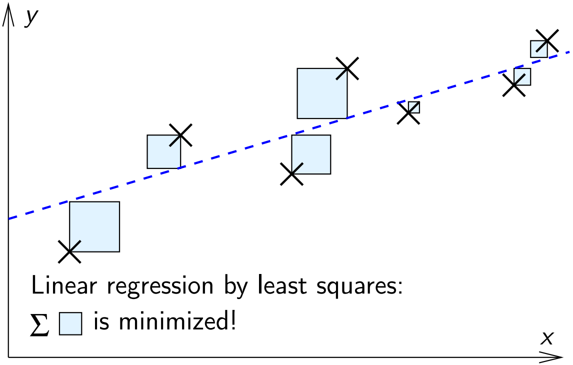
Introduction to GIS-E3010: Least-Squares Methods in Geoscience

Martin Vermeer



Aalto University
School of Engineering

Content, practicalities

- Practicalities concerning GIS-E3010:
  - Lectures, exercises, grading. *No exam*, three exercise packages, equally weighted
  - Materials are in MyCourses
- History and societal status of least-squares methods.

Gauss and Legendre

Carl Friedrich Gauss was a land surveyor in addition to being a master mathematician. He did the triangulation of Hannover using the least-squares method, but never published the method.

Adrien-Marie Legendre again was an eminent French mathematician authoring a textbook on the determination of comet orbits, containing a description of the method.

Optimisation and cost functions

- Least-squares adjustment can be considered the optimisation of a *cost function*. The cost of the imprecision of solution $\widehat{\mathrm{x}} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$ is a function of the value of the solution:

$$f\left(\widehat{\mathrm{x}}\right) = f\left(x_1, x_2, \ldots, x_n\right).$$

The simplest cost function that has an extremum, and thus can be optimised, is the quadratic function of the solution:

$$f\left(x_1, x_2, \ldots, x_n\right) = a + \sum_i b_i x_i + \sum_{i,j} c_{ij} x_i x_j.$$

Aalto University
School of Engineering

Optimisation and cost functions (2)

- The optimum is found by finding the zero of the *gradient* of the function:

$$\frac{\partial f}{\partial x_i} = b_i + \sum_j c_{ij} x_j + \sum_j c_{ji} x_j = 0,$$

or (as, without loss of generality, $c_{ij} = c_{ji}$)

$$\sum_j c_{ij} x_j = -2b_i \Rightarrow \begin{bmatrix} c_{11} & c_{21} & \cdots & c_{n1} \\ c_{12} & c_{22} & \cdots & c_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{1n} & c_{2n} & \cdots & c_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} -2b_1 \\ -2b_2 \\ \vdots \\ -2b_n \end{bmatrix}.$$

- a *linear system* of n equations in n unknowns. As long as the c_{ij} matrix is non-singular, it will have a unique solution. And as long as the matrix is positive definite, the solution will be a *cost minimum*.

Aalto University
School of Engineering

Optimisation and cost functions (3)

- Important: although the function to be optimised is *quadratic*, the solution *is linear*.
- Least-squares theory is *a linear theory*.
- This makes a lot of the math simple and elegant.

Simple practical applications

- Average
- Linear regression
- Multiple regression
- Ordinary least squares vs. weighted least squares

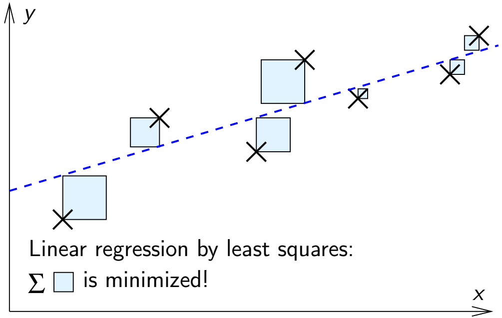
The average as a least-squares estimate

Computing the *average* probably is the simplest example of a least-squares estimate.

- Closer to the "true" value than the individual observations averaged.
- Gets better with more observations n, according to $\dfrac{1}{\sqrt{n}}$.
- Treats all observations equally.
- "Outliers" show up and can, being gross errors, be eliminated.

Aalto University
School of Engineering

Linear regression



- Assumed model: linear relationship $y = ax + b$.
- Fitting a line through a point cloud of measurements.
- Two parameters a and b estimated: intercept and slope.
- Gets better with more observation points.
- Treats all observations equally.
- "Outlier" detection.

Aalto University
School of Engineering

Weighted least squares

- Linear regression works well if the observations are equally precise (*homoskedastic* observations) and uncorrelated.
- It gives the correct estimates $\widehat{a}$ and $\widehat{b}$, but *too optimistic* (too small) *uncertainties*, if the observations are correlated. This is because correlated observations contain less information than independent ones.
- For this, there is *weighted least squares*, giving also the correct statistics.

Aalto University
School of Engineering

Slings and arrows of correlation

When things correlate, all kinds of things can go wrong.

- Correlation between $\underline{x}$ and $\underline{y}$ does not prove causation $\underline{x} \to \underline{y}$; however
- Statistically significant correlation shows *some* common cause. Like $\underline{x} \to \underline{y}$, $\underline{y} \to \underline{x}$, or $\underline{z} \to \underline{x}, \underline{y}$.
- A regression model has to be *complete*: it has to contain all the unknown parameters explaining the observables.
  - *Confounders*. A confounder is a parameter not taken into account in regression.

> Example (fictional): it was found that the incidence of lung cancer is inversely correlated with radon exposure. However, when *smoking* was taken along in the regression, a positive correlation with radon was found. Explanation: radon exposure is highest in rocky areas, whereas city dwellers, living in cities on sedimentary plains, smoke most.

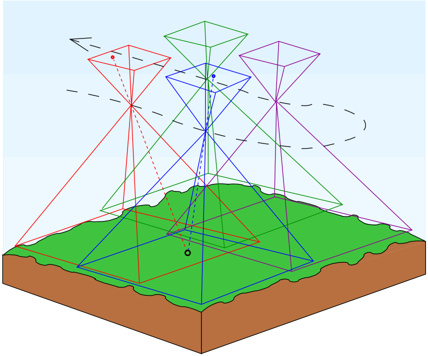
Geodetic network adjustment

Triangulation networks adjusted over large areas was a standard application of the least-squares method during the 19th and 20th centuries. Especially worth mentioning are the large triangulations of Western Europe, leading to the establishment of European Datum 1950, and the North American Datums of 1927 and 1983.

The sheer size of the networks necessitated the use of special mathematical approaches such as Helmert-Wolf blocking.

Photogrammetry, aerotriangulation

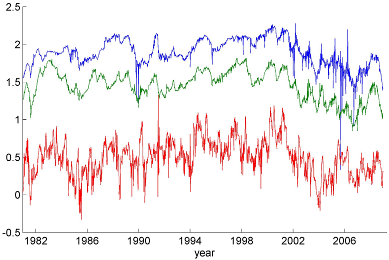
The advent of (aerial) photogrammetry meant a new field of application for least-squares adjustment method. Both in the interior and exterior orientation of images, and in the construction of multi-image models in three dimensions through *aerotriangulation*, the least-squares method finds application.



The systems of observation equations in photogrammetry can be huge; also here, special mathematical approaches are needed. E.g., iterative solution using the conjugate-gradient method.

Aalto University
School of Engineering

Time series, correlation, and all that



- Time series play a role everywhere in society: in economics, demographics, development, geophysics, everywhere things change over time.
- E.g., also in the study of climate change and its impacts.

Proper analysis of time series requires that *serial correlation* – between successive time-series elements – is properly taken into account. The method of choice for this is weighted least squares.

Statistics in society; econometry

In economics, a notation slightly different from that used in geodesy:

$$Y = \alpha + \beta X + \epsilon,$$

compare to the geodetic observation equation

$$\underline{\ell} + \underline{\mathrm{v}} = A\widehat{\mathrm{x}}$$

with $\begin{pmatrix} Y \\ \underline{\ell} \end{pmatrix}$ observation vector, $\begin{pmatrix} \beta \\ \widehat{\mathrm{x}} \end{pmatrix}$ vector of unknowns, $\begin{pmatrix} X \\ A \end{pmatrix}$ coefficient matrix / design matrix, $\begin{pmatrix} \epsilon \\ \underline{\mathrm{v}} \end{pmatrix}$ vector of errors / residuals.

- Note also the *linearization* of geodetic observation equations. In economics, linearity is usually assumed.

Aalto University
School of Engineering

Observations of gravity on the Earth's surface are, like of so many geophysical quantities, *sparse*: meteo temperatures, tide-gauge readings, ... Because of this sparsity, special techniques have been developed for proper statistical inference from this sparse data. The technique is called *least-squares collocation*.

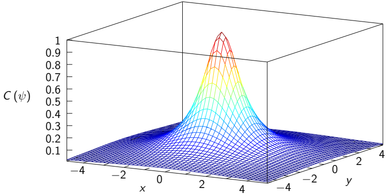
Famous researchers in least-squares collocation

- Torben Krarup (1919–2005), Denmark
- Helmut Moritz (1933–), Austria
- Weikko A. Heiskanen (1895–1971), Finland
- C. Christian Tscherning (1942–2014), Denmark
- Danie G. Krige (1919–2013), South Africa, described in his master's thesis the technique that became popularly known in geostatistics as "*kriging*". Equivalent to least-squares collocation, also known more formally as *Wiener-Kolmogorov prediction*.

Aalto University
School of Engineering

Spatial covariance

The use of sparse spatial data to make meaningful inferences is possible because of long-range *spatial correlation*. This property is described by the *spatial covariance function*. The figure is the covariance function for gravity anomalies formulated by R.A. Hirvonen in 1962, based on data from Ohio and Finland.



It is characterized by two parameters: the *variance* C_0, describing the variability of gravity anomalies, and the *correlation length* d, describing the distance over which gravity anomalies are typically still somewhat correlated.

Aalto University
School of Engineering

Least-squares collocation

In least-squares collocation, the spatial covariance function is used to generate the *coefficient matrix* used for predicting, or estimating, unknown values $\widehat{\ell}_P$ in locations P where no measurement was taken from the measurements $\underline{\ell}_1, \underline{\ell}_2, \ldots, \underline{\ell}_N$ that are available.

Like this (observations assumed error free):

$$\widehat{\ell}_P = \left[\begin{array}{cccc} C(s_{P1}) & C(s_{P2}) & \cdots & C(s_{PN}) \end{array} \right] \left[\begin{array}{cccc} C(0) & C(s_{21}) & \cdots & C(s_{N1}) \\ C(s_{12}) & C(0) & \cdots & C(s_{N2}) \\ \vdots & \vdots & \ddots & \vdots \\ C(s_{1N}) & C(s_{2N}) & \cdots & C(0) \end{array} \right]^{-1} \left[\begin{array}{c} \ell_1 \\ \underline{\ell}_2 \\ \vdots \\ \underline{\ell}_N \end{array} \right],$$

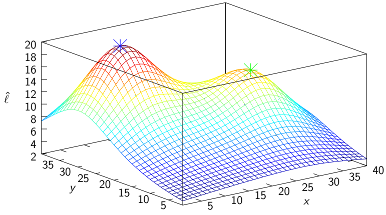
where the covariance function is $C(s_{ij})$, with s_{ij} being the (spherical) distance between points i and j.

For the Hirvonen function, we have

$$C(s) = \frac{C_0}{1 + \left({}^{s}/_{d} \right)^2}.$$

Aalto University
School of Engineering

Least-squares collocation (2)



The predicted values of a function given in two points shown. For this special case (error free observations) this is an interpolation technique:

- Close to the data points, the predicted values are close to the data-point values.
- *At* a data point, the predicted value equals the data value of the point (this applies only for error free observations!)
- Far away from the data points, the predicted values go smoothly to zero: lacking real info, zero is a "best guess".
- The "prediction surface" is a linear combination of Hirvonen "bell-curve surfaces" like shown in the earlier figure.

Aalto University
School of Engineering

The space age: real-time least squares

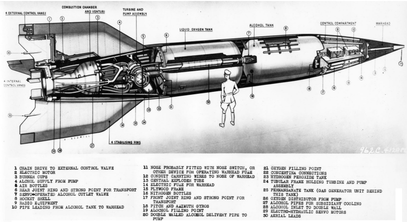
- Motivation
- The Kalman filter and the Extended Kalman filter
- The particle filter
- Applications

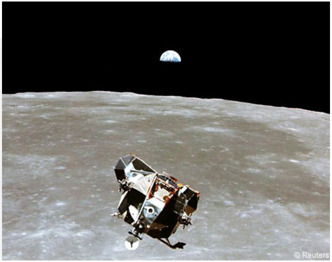
Aalto University
School of Engineering

Unmanned flight started with missiles. The first modern missile was the German V2: it contained *a guidance system* based on inertial techniques. This creates a need to be able to calculate the path, either on the ground or on-board, while the flight was ongoing.

Missiles became satellite carriers, satellites spacecraft, and the paths of all needed to be determined from observations during flight.

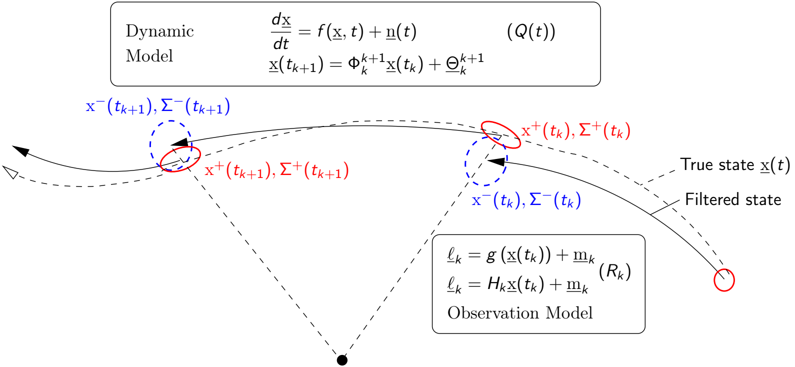
This is what the *Kalman filter* was developed for.





One critical application was the *rendez-vous problem*, two orbiting spacecraft finding each other in space and exchanging materials.

Aalto University
School of Engineering



- There is a *dynamic model* that propagates our knowledge of the *state* of the system forward in time using our knowledge of the system's behavour.
- There is an *observation model* that allows us to improve our knowledge of the state of the system as observations become available. This improvement is done in a least-squares way.
- Repeat. Kalman is a recursive filter.
- Kalman, R. E. (1960). "A New Approach to Linear Filtering and Prediction Problems". Journal of Basic Engineering. 82:35.

Aalto University
School of Engineering

Applications of the Kalman filter

The Kalman filter is used in many places outside spacecraft navigation. Examples:

- Integrated navigation using IMU (Inertial Measurement Unit) and GNSS (Global Navigation Satellite System) measurements, in aircraft, vessels, land vehicles, backpacks. The filter accepts measurements of both types and produces a navigation trajectory in real time.
- In the GNSS processing package GAMIT/GLOBK, the latter (GLOBK) is a Kalman filter.
- The Kalman filter is used in robotics for motion control,
- it is used in econometrics, and
- it is part of software used for automated trading on the stock exchange.

A really nice discription of the Kalman filter, with good pictures, can be found here.

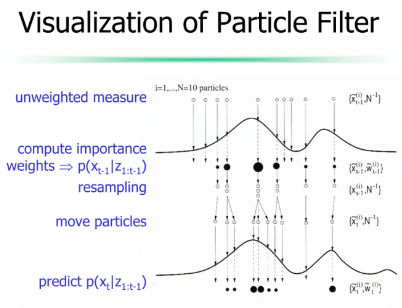
Other filters

- The *extended Kalman filter* involves *linearization*, and can be used in general non-linear situations. This applies to both the dynamic model and the observational model.
- For the Kalman filter, the best estimate is represented as a single state vector, and its statistical probability distribution as a single variance-covariance matrix of that state vector.
- It thus cannot handle *general* statistical distributions, just normal ones.
- (Non-normality will inevitably be produced by strong non-linearities in the system.)
- One solution to this is the *particle filter*.

Aalto University
School of Engineering

The particle filter

- The *particle filter* represents the probability distribution explicitly as a "cloud" of particles in state space, initially with each particle the same weight.



- As the system evolves and the particles move, *observations* will change (update) the weights of the particles. Every now and then the cloud must be "renormalized" – i.e., "heavy" particles split, "light" ones merged.
- The particle filter is especially suitable if the statistical distributions are known or suspected to be strongly non-normal.

Aalto University
School of Engineering

Comparison of least-squares techniques

- Adjustment:
  - Unknown is a vector of *real numbers* x_i.
  - Observed are real numbers ℓ_j.
- Collocation:
  - Unknown is a function f of one of more real arguments, like time t or place (x, y).
  - Observed are function values at discrete points or argument values $f(t_i)$.
- Kalman-filter:
  - Unknown is a vectorial "state" $x_i(t)$ of time.
  - Observed are values of a function of this state $F(x_i(t_j))$, at discrete times before the present time, $t_j < t$.
  - Estimation takes place in real time using only "past" observations.

Comparison of least-squares techniques

Needed in all are

1. *Functional models* describing the known physics of the situation and the observation process
2. *Statistical models* for all uncertainties afflicting the system and its observation process.

- *Linearization* will typically be necessary to make the problem tractable.
- *Results* obtained will be:
  - estimates of the unknowns, optimal in the least-squares sense
  - estimates of the uncertainties of the unknowns.

Summary; questions

- Practicalities of the course GIS-E3010 were discussed.
- The history, background and scope of least-squares computational methods in society, and more specifically in the geosciences, was presented.

Questions?

Thank you!

Aalto University
School of Engineering