Localisation précise par moyens spatiaux

Advanced Parameter Estimation in Celestial Mechanics

Gerhard Beutler

Table of Contents: Advanced Parameter Estimation

1. Least Squares (LSQ): a review (based on lecture notes by W. Gurtner)
2. LSQ with Conditions
3. Collocation Methods
4. From LSQ to Kalman filtering: Deterministic systems
5. From LSQ to Kalman filtering: Stochastic systems

Least squares: a review

Part of this section is based on lecture notes by Prof. Werner Gurtner, in particular figures and formulas.

Contents

- 1. The mathematical model of an adjustment
- 2. Least squares
- 3. More general adjustment methods

Least squares: A review

Repeated measurements of one and the same object (e.g., length of a rod) show that measurements contain errors, where the distinction is made between

- blunders
- Regular or systematic errors
- Irregular or random errors

If the differences between the measurements are far above the expected precision of a method, one speaks of blunders.

Systematic errors always influence a measurement in the same sense, e.g., instrument errors, environmental influences, parallaxes).

Random errors are based on the imperfection of the measurement process. They have to be treated statistically (keyword: expectation values, variances).

The total error consists of the sum of all random errors and of all systematic errors.

An outlier may be a random error with very small probability or a blunder – the treatment of outliers is based on experience (!).

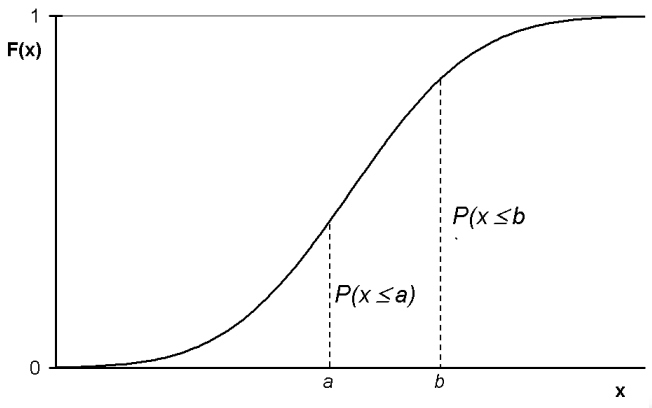
Least squares: A review

Random variables and their characteristics:

- A random or stochastic variable is a function, assigning a distinct real value to each result of a random experiment.
- The probability that a random variable X assumes a value x is denoted by P(X=x).
- The probability for X assuming a value in the interval (a,b] is denoted by $P(a < X \leq b)$.

Probability function F(x):

The probability function F(x) provides the probability that X assumes values ≤ x:

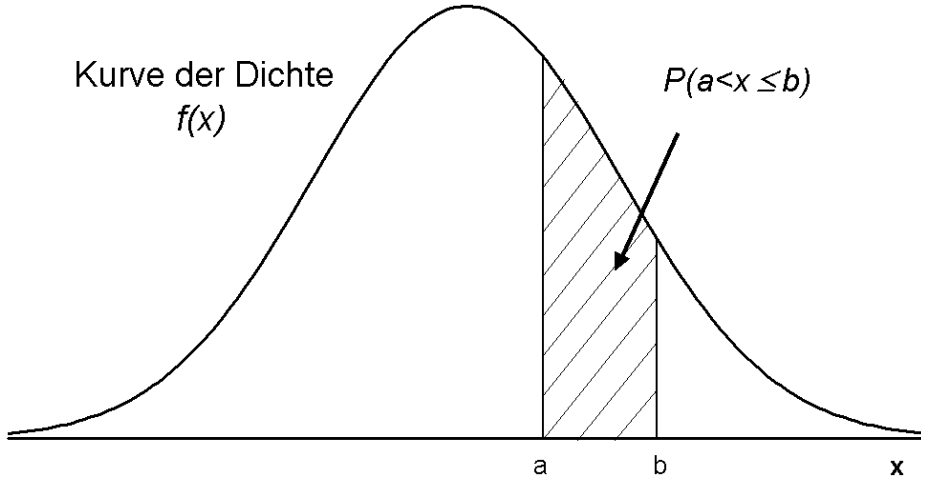


F(x) grows monotonically.

F(-"infty") = 0, F(+"infty") = 1

$P(a < X \leq b) = F(b) - F(a)$

Least squares: A review



The probability density function f(x) is the first derivative w.r.t x of the probabilty function, i.e., f(x) = F'(x).

Properties:

- $F(x) = \int_{-\infty}^{\cdot} f(v)\,dv$
- $F(\infty) = \int_{-\infty}^{\infty} f(v)dv = 1$
- $P(a < X \le b) = F(b) - F(a) = \int_{a}^{b} f(v)dv$

Least squares: A review

Parameters of a distribution:

Expectation values E(x) or mean values μ for continuous and discrete variables:

$$\mu = E(X) = \int x\, f(x)\, dx$$

$$\mu = E(X) = \sum_{n} x_i\, p(x_i)$$

Parameters of a distribution:

Standard deviation σ and/or variance for continuous and discrete variables:

$$\sigma^2 = V(X) = E([X-\mu]^2) = \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)\, dx$$

Least squares: A review

The variance is a special case (for k=2) of central moments of order k:

$$E([X-\mu]^k) = E([X-E(X)]^k) = \int_{-\infty}^{+\infty} (x-\mu)^k f(x)dx$$

For the variance we have the following important relationship:

$$E([X-\mu]^2) = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu E(X) + \mu^2$$

$$E([X-\mu]^2) = E(X^2) - \mu^2$$

Least squares: A review

Gaussian or normal distribution:

Introduced by C.F. Gauss (1777-1855) in the context of the theory of measurement errors. It is the most frequently used distribution in practice, because:

- Many random variables encountered in practice are normally distributed.
- The sum of many arbitrarily distributed random variables results in good approximation in a normal distribution.

A normal distribution with σ =1 and μ = 0 is a standard normal distribution.

The probability density function of a normal distribution reads as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (-\infty < x < \infty,\ \sigma > 0)$$

Least squares: A review

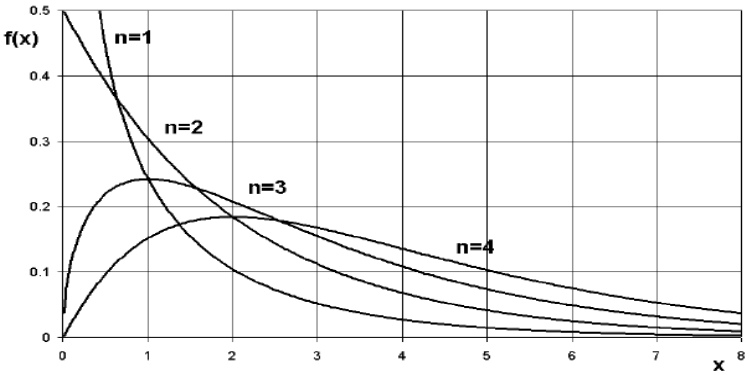
χ^2 Distribution: The sum of the squares of *n* independent normally distributed random Variables X_i, *i=1,2,…, n* (with σ =1 and μ = 0) is called χ^2

The probability distribution of this new random variable is called *Chi-square distribution*. The corresponding probability density is:

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n-2)/2} \cdot e^{-x/2}$$

The χ^2 distribution has mean value μ = n and variance σ^2 = 2n.

The χ^2 asymptotically is normally distributed with μ = n und $\sigma^2 = 2n$.



Figur 1.6: Dichte der χ^2-Verteilung

Least squares: A review

For statistical tests we may formulate the following

Theorem: Let:

- X*i, i=1,2,….,n* normally distributed random variables with mean values μ and variance σ^2.
- An empirical estimation of the mean value and the corresponding variance is obtained by:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$$

where the random variable Z is χ^2 distributed with degree of freedom *f=n-1:*

$$Z = (n-1)\frac{S^2}{\sigma^2}$$

Least squares: A review

The mean value and the variance are also called parameters of a distribution (there are others).

From a sample of *n* elements these parameters may be estimated. The corresponding formulas are called estimators.

We thus define functions $U = g(X_1,X_2,...,X_n)$ of the random variables X_i.

These functions in turn are random variables with own probabilty distributions and parameters.

An estimator is called unbiased, if its expectation value equals the expectation value of the corresponding parameter (defined by its density function).

The empirical estimators for mean and variance on the previous page are unbiased.

An estimator is called consistent, if its expectation value converges to the true value for large values *n*.

Least squares: A review

Probability distribution with more than one random variable: Often one has to deal not only with one random variable X but with several, e.g., with measuring pressure and temperature, latitude and longitude, coordinates of a point in $\boldsymbol{E}^2$ or $\boldsymbol{E}^3$, etc. One thus has to define the probability that (X,Y,...) assume at maximum the values (x,y,...):

$$F(x,y,\ldots) = P(X \leq x, Y \leq y, \ldots)$$

F(x,y,..) is the distribution function of the multi-dimensional distribution of the random variables (X,Y,...).

In two dimensions the density function f(x,y) is defined by:

$$F(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(x,y)\, dx\, dy$$

The probability for the values (x,y) lying in a rectangle is defined by:

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x,y)\, dx\, dy$$

Least squares: A review

Covariances: The following definitions are provided for dimension 2. A generalization to higher dimensions is evident.

The variance σ^2 of a sum of two random variables is:

$$\sigma^2 = E[(X+Y)^2] - [E(X+Y)]^2$$

$$\sigma^2 = E(X^2) + 2E(XY) + E(Y^2) - [E(X)+E(Y)]^2$$

from where:

$$\sigma^2 = E(X^2) + 2E(XY) + E(Y^2) - [E(X)]^2 - 2E(X)E(Y) - [E(Y)]^2$$

$$\sigma^2 = E(X^2)-[E(X)]^2 + E(Y^2)-[E(Y)]^2 + 2[E(XY)-E(X)E(Y)]$$

$$\sigma^2 = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}$$

$$\boxed{\sigma_{XY} = E(XY) - E(X)E(Y) = E([X - \mu_X][Y - \mu_Y])}$$

σ_{xy} is the covariance of the random variables X and Y.

Least squares: A review

The correlation coefficient of X und Y is defined by:

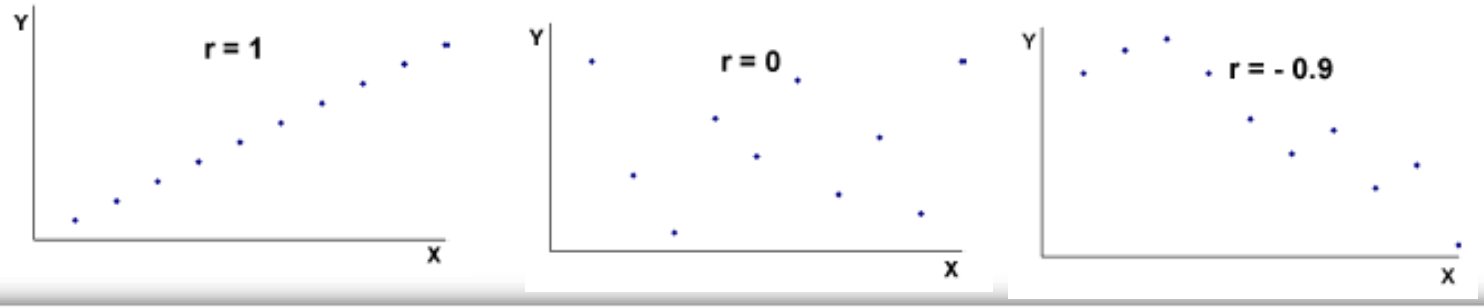
$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Estimators for covariance and correlation coefficient:

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})$$

$$r = \frac{S_{XY}}{S_X S_Y}$$

ρ und r may assume values between –1 and +1:



Least squares: A review

Independence: Two random variables are called independent, if for all (x,y):

$$F(x,y) = F_x(x)\, F_y(y).$$

From where we may conclude

$$f(x,y) = f_x(x)\, f_y(y).$$

For independent random variables we have:

$$E(XY) = E(X)\, E(Y)$$

Implying that the covariance is

$$\sigma_{xy} = 0$$

Consequently the correlation coefficient is:

$$\rho = 0$$

Implying that independent random variables are uncorrelated.

Least squares: A review

Error propagation: Let $X^T = (X_1, X_2, ..., X_n)$ a random vector for which all variances und and covariances are known. The corresponding information is conveniently reported in the covariance matrix:

$$\mathbf{K}_{XX} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{pmatrix}$$

Let $Z^T = (Z_1, Z_2, ..., Z_m)$ a linear function of X:

$$Z = \mathbf{C} \cdot X,$$

where **C** is a matrix with m rows and n columns. The covariance matrix K_{zz} of Z simply is:

$$\mathbf{K}_{zz} = \mathbf{C} \cdot \mathbf{K}_{xx} \cdot \mathbf{C}^T$$

If the functional dependency of X und Z is not linear, a linearization must take place before!

Least squares: A review

Covariance and co-factor matrix: If the vector X corresponds to a series of *independent* measurements, $\mathbf{K}_{xx}$ is diagonal.

If all measurements have the same variance σ_0^2, we have

$$\mathbf{K}_{xx} = \sigma_0^2 \cdot \mathbf{E}, \quad \mathbf{E} = \text{Identity matrix.}$$

Even if this is not true, it makes in general sense to use one measurement as unit (let us assume with variance σ_0^2).

The matrix Q_{xx}, obtained from K_{xx} by division by σ_0^2 is called matrix of cofactors:

$$\mathbf{K}_{xx} = \sigma_0^2 \cdot \mathbf{Q}_{xx}, \quad \sigma_i^2 = \sigma_0^2 \cdot q_{ii}, \quad \sigma_{ik} = \sigma_0^2 \cdot q_{ik}$$

Least squares: A review

Weights: Let $X^T = (X_1, X_2, ..., X_n)$ a random vector of independent random variables with $\mu=0$ and variances $(\sigma_1, \sigma_2, ... , \sigma_n,)$. Its covariance matrix $\mathbf{K}_{xx}$ is diagonal, with $K_{xx}(i,i) = \sigma_i^2$.

With $Z^T = (\sigma_0/\sigma_1 \ X_1, \sigma_0/\sigma_2 \ X_2, ..., \sigma_0/\sigma_n \ X_n)$ we define a random vector with: $\mathbf{K}_{zz} = \sigma_0^2 \cdot E$.

Obviousy $\mathbf{K}_{zz} = \mathbf{P} \cdot \mathbf{K}_{xx}$ with **P** diagonal $[P_{ii} = p_i = (\sigma_0/\sigma_i)^2]$.

p_i ist the weight of the random variable with No. i.

If the covariance matrix is not diagonal, the weight matrix is defined by:

$$\mathbf{P}_{xx} = \sigma_0^2 \cdot \mathbf{K}_{xx}^{-1} = \mathbf{Q}_{xx}^{-1}$$

Least squares: A review

In order to determine parameters from observations (measurements) we need:

- A functional model of the observations, i.e., the functional dependence of the observed quantities from the parameters to be determined,
- the statistical model of the observations as well as
- a statistical criterion to estimate the parameters.

The measured quantities L_i are treated as random variables. Their statistical properties are described by their covariance matrix, where σ_i^2 is the variance of observation L_i , σ_{ik} the covariance of observations L_i and L_k:

$$\mathbf{K} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & & \sigma_{2n} \\ \vdots & & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{pmatrix}$$

Least squares: A review

Adjustment of intermediate observation equations represents the *n* observed quantities L_i as functions of *u* parameters X_i :

$$L_1 = F_1(X_1,X_2,\ldots,X_u)$$

$$L_2 = F_2(X_1,X_2,\ldots,X_u)$$

$$\ldots$$

$$L_n = F_n(X_1,X_2,\ldots,X_u)$$

Alternatively, the method of conditioned observations introduces $r < n$ condition equations between the observations, which have to be met precisely. The *n* observations assume the role of parameters. The condition equations may be written as:

$$F_1(L_1,L_2,\ldots,L_n) = 0$$

$$F_2(L_1,L_3,\ldots,L_n) = 0$$

$$\ldots$$

$$F_r(L_1,L_2,\ldots,L_n) = 0$$

Least squares: A review

In practice adjustment with intermediary observations is much more frequently met than adjustment based on condition equations.

Criterion of parameter estimation is needed whenever the number *n* of observations exceeds the number *u* of parameters.

- *n* > *u* is no luxury:
  - Outliers may be identified and eliminated.
  - Results become more accurate.
  - The accuracy of results may be estimated.
- Adjustment wants to obtain, with small corrections (residuals), applied to the observations, a system of *u* linear equations for the *u* parameters.
- Several criteria have been proposed:
  - Minimize the largest residual (Tschebyscheff)
  - Minimize sum of residuals (Laplace), L1 norm
  - Minimize sum of residuals square → Method of Least Squares (LSQ), (Gauss), L2 norm.

Question: How to establish an L1-norm with LSQ (poor (wo)man's solution)?

Least squares: A review

MONATLICHE

CORRESPONDENZ

ZUR BEFÖRDERUNG

DER

ERD- UND HIMMELS-KUNDE.

SEPTEMBER, 1809.

XVII.

Summariſche Überſicht der zur Beſtimmung der Bahnen der beyden neuen Hauptplaneten angewandten Methoden.

Vom

Hrn. Prof. *Gauſs*. *)

Die von Kreis- und Parabel-Hypotheſen unabhängige Beſtimmung der Bahn eines Himmelskörpers aus einer kurzen Reihe von Beobachtungen beruhet auf zwey Forderungen: I. Muſs man Mittel haben, die Bahn zu finden, die drey gegebenen vollſtändigen Beobachtungen Genüge thut. II. Muſs man die ſo gefundene Bahn ſo verbeſſern können, daſs die Differenzen der Rechnung von dem ganzen Vorrath der Beobachtungen ſo gering als möglich werden.

Left: Portrait of C.F. Gauss, Right: Orbit determination & improvement – probably the first application of LSQ. In addition first orbit determination is discussed.

Least squares: A review

Method of least squares:

v is the vector of residuals to be added to the observations ℓ, $P_{\ell\ell}$ the weight matrix associated with the observations. The LSQ criterion in its most general form asks to minimize the weighted sum of residuals:

$$\boxed{\mathbf{v}^T \mathbf{P}_{\ell\ell} \mathbf{v} = \text{minimal}}$$

where:

$$\mathbf{P}_{\ell\ell} = \mathbf{Q}_{\ell\ell}^{-1} = \sigma_0^2 \cdot \mathbf{K}_{\ell\ell}^{-1}$$

With independent observations the covariance matrix $\mathbf{K}_{\ell\ell}$ becomes diagonal, in consequence also the matrices $\mathbf{Q}_{\ell\ell}$ und $\mathbf{P}_{\ell\ell}$. The diagonal element p_i of $\mathbf{P}_{\ell\ell}$ is the weight of observation No. *i*:

$$p_i = \frac{\sigma_0^2}{\sigma_i^2}$$

The LSQ criterion is then reduced to:

$$\sum_n p_i \cdot v_i^2 = [pvv] = \text{minimal}$$

Least squares: A review

Justification of LSQ:

- With normally distributed observations one obtains the parameters with the highest probabilities.
- Independent of the distribution of the observations one obtains parameters of maximum accuracy.
- The results are easily calculated (toolkit of linear algebra).
- LSQ qives plausible results (e.g., arithmetic mean when measuring a quantity *n* times with the same method).

Classical LSQ dealt with independent measurements. Helmert (around 1900), Tienstra (around 1950) and others have generalized LSQ to correlated observations.

Least squares: A review

Each observed quantity ℓ_i is written as a function of the parameters x_i :

$$\ell_i = F(x_1,...,x_u) \quad , \quad i = 1,...,n$$

in matrix notation written as:

$$L = F(X)$$

Due to the measurement errors the actual measurements ℓ_i' may not be met exactly. This why the quantities v_i, associated with each of the observations are introduced to exactly meet the observation EQs:

$$\bar{\ell} = \boldsymbol{\ell}' + \mathbf{v} = F(\bar{\mathbf{x}})$$

where $\boldsymbol{\ell}$' stands for the actual observation, the same symbol with bar for the adjusted observation, **v** for the array of residuals, and **x** (with bar) for the estimated parameter array.

Least squares: A review

Assuming that the observational EQs are linear in the parameters, they may be written in the form:

$$\mathbf{v} = \mathbf{A} \cdot \mathbf{x} - \boldsymbol{\ell}'$$

Adopting the LSQ criterion (left with correlations, right without) implies:

$$\mathbf{v}^T \mathbf{P}\, \mathbf{v} = \mathbf{v}^T \mathbf{Q}^{-1}\, \mathbf{v} = \text{minimal}$$

$$\sum_{i=1}^{n} p_i \cdot v_i^2 = [pvv] = \text{minimal}$$

The minimum is assumed, if the partial derivatives of the left-hand side of the above condition w.r.t. to all *u* parameters are zero:

$$\frac{\partial [pvv]}{\partial x_1} = 0 \quad ; \quad \frac{\partial [pvv]}{\partial x_2} = 0 \quad ; \quad \ldots \quad ; \quad \frac{\partial [pvv]}{\partial x_u} = 0$$

$$\frac{\partial [pvv]}{\partial x_1} = \sum_{k=1}^{n} 2 p_k v_k \cdot \frac{\partial v_k}{\partial x_1} = \sum_{k=1}^{n} 2 p_k v_k \cdot a_k = 2[pav] = 0$$

.....

Least squares: A review

We thus obtain the normal equation system (NEQs) with *u* EQs and *u* parameters:

$$
\begin{aligned}
[paa]\, x_1 + [pab]\, x_2 + [pac]\, x_3 + \ldots - [pa\ell] &= 0 \\
[pab]\, x_1 + [pbb]\, x_2 + [pbc]\, x_3 + \ldots - [pb\ell] &= 0 \\
[pac]\, x_1 + [pbc]\, x_2 + [pcc]\, x_3 + \ldots - [pc\ell] &= 0 \\
\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots &
\end{aligned}
$$

In matrix notation: $N \cdot x - f = 0$

$$
\mathbf{N} = \begin{pmatrix} [paa] & [pab] & [pac] & \ldots \\ [pab] & [pbb] & [pbc] & \ldots \\ [pac] & [pbc] & [pcc] & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}; \qquad \mathbf{f} = \begin{pmatrix} [pa\ell] \\ [pb\ell] \\ [pc\ell] \\ \vdots \end{pmatrix}
$$

$$
\mathbf{N} = \mathbf{A}^T\mathbf{P}\mathbf{A} \; ; \quad \mathbf{f} = \mathbf{A}^T\mathbf{P} \cdot \boldsymbol{\ell}'
$$

Least squares: A review

$$\mathbf{N} \cdot \mathbf{x} - \mathbf{f} = \mathbf{0}$$

is called NEQs, **N** normal equation matrix. **N** is a quadratic, symmetrical and positive definite matrix.

With a fully populated weight matrix the NEQs is obtained by

$$\Omega = \mathbf{v}^T \mathbf{P}\, \mathbf{v} = \text{minimum}, \quad \mathbf{v} = \mathbf{A}\,\mathbf{x} - \boldsymbol{\ell}'$$

$$\Omega = \mathbf{x}^T\mathbf{A}^T\mathbf{P}\,\mathbf{A}\,\mathbf{x} - \mathbf{x}^T\mathbf{A}^T\mathbf{P}\,\boldsymbol{\ell}' - \boldsymbol{\ell}'^T\mathbf{P}\,\mathbf{A}\,\mathbf{x} + \boldsymbol{\ell}'^T\mathbf{P}\,\boldsymbol{\ell}'$$

$$\frac{\partial \Omega}{\partial \mathbf{x}} = 2\,\mathbf{A}^T\mathbf{P}\,\mathbf{A}\,\mathbf{x} - \mathbf{A}^T\mathbf{P}\,\boldsymbol{\ell}' - \mathbf{A}^T\mathbf{P}\,\boldsymbol{\ell}' = \mathbf{0}$$

$$= 2\,\mathbf{A}^T\mathbf{P}\,\mathbf{A}\,\mathbf{x} - 2\,\mathbf{A}^T\mathbf{P}\,\boldsymbol{\ell}' = \mathbf{0}$$

$$\boxed{\mathbf{A}^T\mathbf{P}\,\mathbf{A}\,\mathbf{x} - \mathbf{A}^T\mathbf{P}\,\boldsymbol{\ell}' = \mathbf{0}}$$

The solution of the NEQs reads as:

$$\boxed{\mathbf{x} = (\mathbf{A}^T\,\mathbf{P}\,\mathbf{A})^{-1}\,\mathbf{A}^T\,\mathbf{P}\,\boldsymbol{\ell}'}$$

Least squares: A review

Mean errors of the parameters: With the supposedly known covariance matrix of the observations one my derive the covariance matrix of the parameters using the law of error propagation:

$$\mathbf{Q}_{xx} = (\mathbf{A}^T\mathbf{P}\,\mathbf{A})^{-1}\cdot\mathbf{A}^T\mathbf{P}\,\mathbf{A}\cdot(\mathbf{A}^T\mathbf{P}\,\mathbf{A})^{-1} = (\mathbf{A}^T\mathbf{P}\,\mathbf{A})^{-1} = \mathbf{N}^{-1}$$

$$\mathbf{K}_{xx} = \sigma_0^2\cdot\mathbf{Q}_{xx}$$

$$\sigma_i^2 = \sigma_0^2\cdot q_{ii} \quad ; \quad \sigma_{ik} = \sigma_0^2\cdot q_{ik}$$

$$\rho_{ik} = \frac{\sigma_{ik}}{\sigma_i\cdot\sigma_k} = \frac{q_{ik}}{\sqrt{q_{ii}\cdot q_{kk}}}$$

With the inverted NEQ matrix $\mathbf{Q}_{xx}$ and with the standard deviation σ_0 of the observations the mean errors, the covariances and correlation coefficients between the parameters may be calculated.

The covariance matrix of linear functions of the parameters may be computed as well.

Least squares: A review

Mean error a posteriori of weight unit: The LSQ estimate of σ_0 is computes as:

$$m_0^2 = \frac{\sum_{i=1}^{n} (\ell_i' - \bar{\ell}_i)^2}{n-u} = \frac{\mathbf{v}^T \mathbf{P} \mathbf{v}}{n-u} \qquad \text{wobei:} \qquad E[m_0^2] = \sigma_0^2$$

The following formula is important, because it provides m_0 without the need to explicitly calculate the residuals:

$$\mathbf{v}^T\mathbf{P}\mathbf{v} = (\mathbf{x}^T\mathbf{A}^T - \ell'^T)\,\mathbf{P}\,(\mathbf{A}\,\mathbf{x} - \ell')$$

$$= \mathbf{x}^T(\mathbf{A}^T\mathbf{P}\,\mathbf{A}\,\mathbf{x} - \mathbf{A}^T\mathbf{P}\,\ell') - \ell'^T\mathbf{P}\,\mathbf{A}\,\mathbf{x} + \ell'^T\mathbf{P}\,\ell'$$

$$\text{———} = 0 \text{———}$$

$$\boxed{\mathbf{v}^T\mathbf{P}\mathbf{v} = \ell'^T\mathbf{P}\,\ell' - \mathbf{x}^T\cdot\mathbf{A}^T\mathbf{P}\,\ell'}$$

Least squares: A review

Note that $\mathbf{A}^T\mathbf{P}\,\ell' = f$ is the RHS of the NEQs, which is why:

$$v^T\mathbf{P}v = \ell'^T P\,\ell' - x^T f$$

When summing up the NEQs, one only has to calculate in addition $\ell'^T P\,\ell'$ to be able to use the above formula.

If the elements of ℓ' are large in absolute value, numerical problems may occur.

Non-linear observation equations have to be linearized prior to the adjustment. The success depends on the quality of $\mathbf{x}_0$.

$$\bar{\ell} = F(\bar{\mathbf{x}}) \quad \rightarrow \quad \bar{\ell} = F(\mathbf{x}_0) + \mathbf{A}\,\mathbf{x}, \qquad \bar{\mathbf{x}} = \mathbf{x}_0 + \mathbf{x}$$

$$\mathbf{A} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_u} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_u} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_u} \end{pmatrix}$$

Least squares: A review

When dealing with linearized observation EQs one was to replace the observations by the difference “observations – observed functions calculated with the a priori parameter values”:

$$v = A x - \ell \qquad \text{where:} \quad \ell = \ell' - \ell_0$$

The elements of $\ell' - \ell_0$ are given the attribute observed – computed.

Matrix **A** is called the first design matrix.

The linearization process neglects the terms of the higher than the first order in the parameter increments.

→

The parameter estimation problem in principle has to solved iteratively.

In practice one tries to use high quality a priori values allowing for a solution in one iteration step.

Least squares: A review

Test of the mathematical model: The ratio of the variances a posteriori and a priori may be used for this purpose (n and u are the numbers of observations and parameters, respectively):

$$z = \frac{(n-u)m_0^2}{\sigma_0^2} = \frac{\mathbf{v}^T\mathbf{P}\mathbf{v}}{\sigma_0^2}$$

z is a random variable with degree of freedom n-u and a chi-square distribution, implying that the following random variable has a chi-square/(n-u) distribution and expectation value 1:

$$\frac{m_0^2}{\sigma_0^2} = \frac{z}{(n-u)}$$

For n-u >> 1 the ratio should be close to 1. If this is not the case one either has outliers in the system or the mathematical and/or statistical model describing the observations is inadequate.

Least squares with conditions

Adjustment of intermediate observation equations with conditions between parameters: A linear(ized) system of n observation equations is augmented by r<u linear condition equations between the parameters.

The condition equations might be used to eliminate r parameters. The resulting system is, however, often difficult to handle.

The combined system may be written as:

$$
\begin{aligned}
v_1 &= a_1 x_1 + b_1 x_2 + c_1 x_3 + \ldots - l_1 \\
v_2 &= a_2 x_1 + b_2 x_2 + c_2 x_3 + \ldots - l_2 \\
&\vdots \\
v_n &= a_n x_1 + b_n x_2 + c_n x_3 + \ldots - l_n \\
\\
0 &= f_1 x_1 + g_1 x_2 + h_1 x_3 + \ldots + w_1 \\
0 &= f_2 x_1 + g_2 x_2 + h_2 x_3 + \ldots + w_2 \\
&\vdots \\
0 &= f_r x_1 + g_r x_2 + h_r x_3 + \ldots + w_r
\end{aligned}
$$

or

$$
\begin{aligned}
\underset{n.1}{\mathbf{v}} &= \underset{n.u}{\mathbf{A}} \cdot \underset{u.1}{\mathbf{x}} - \underset{n.1}{\ell} \\
\underset{r.1}{\mathbf{0}} &= \underset{r.u}{\mathbf{C}} \cdot \underset{u.1}{\mathbf{x}} + \underset{r.1}{\mathbf{w}_c}
\end{aligned}
$$

Least squares with conditions

The weighted sum of residuals squares must be minimized, where r conditions have to be met exactly. The following minimum principle does the job ($\mathbf{k}_c$ is the array of *r* Lagrange multipliers):

$$\Omega = \mathbf{v}^T\mathbf{P}\mathbf{v} + 2\,\mathbf{k}_c^T(\mathbf{C}\,\mathbf{x} + \mathbf{w}_c)$$

$$\Omega = (\mathbf{x}^T\mathbf{A}^T - \boldsymbol{\ell}^T)\,\mathbf{P}\,(\mathbf{A}\,\mathbf{x} - \boldsymbol{\ell}) + 2\,\mathbf{k}_c^T\,(\mathbf{C}\,\mathbf{x} + \mathbf{w}_c)$$

$$\Omega = \mathbf{x}^T\mathbf{A}^T\mathbf{P}\,\mathbf{A}\,\mathbf{x} - \mathbf{x}^T\mathbf{A}^T\mathbf{P}\,\boldsymbol{\ell} - \boldsymbol{\ell}^T\mathbf{P}\,\mathbf{A}\,\mathbf{x} + \boldsymbol{\ell}^T\mathbf{P}\,\boldsymbol{\ell} + 2\,\mathbf{k}_c^T\mathbf{C}\,\mathbf{x} + 2\,\mathbf{k}_c^T\mathbf{w}_c$$

The partial derivatives w.r.t. all parameters and multipliers have to be 0:

$$\begin{array}{|rcrcrcl|}\hline \mathbf{A}^T\mathbf{P}\mathbf{A} & \mathbf{x} & + & \mathbf{C}^T\mathbf{k}_c & - & \mathbf{A}^T\mathbf{P}\,\boldsymbol{\ell} & = 0 \\ \mathbf{C} & \mathbf{x} & & & + & \mathbf{w}_c & = 0 \\ \hline\end{array}$$

The system may be brought into a form looking like a NEQs:

$$\hat{\mathbf{N}} = \begin{pmatrix} \mathbf{A}^T\mathbf{P}\mathbf{A} & \mathbf{C}^T \\ \mathbf{C} & \mathbf{0} \end{pmatrix}; \quad \underset{u+r.1}{\hat{\mathbf{x}}} = \begin{pmatrix} \mathbf{x} \\ \mathbf{k}_c \end{pmatrix}; \quad \hat{\mathbf{f}} = \begin{pmatrix} \mathbf{A}^T\mathbf{P}\,\boldsymbol{\ell} \\ -\mathbf{w}_c \end{pmatrix}$$

$$\boxed{\hat{\mathbf{N}}\cdot\hat{\mathbf{x}} - \hat{\mathbf{f}} = 0} \qquad \boxed{\hat{\mathbf{x}} = \hat{\mathbf{N}}^{-1}\cdot\hat{\mathbf{f}}}$$

Least squares with conditions

The mean error of the weight unit is given by:

$$v^T P v = x^T A^T P A x - x^T A^T P \ell - \ell^T P A x + \ell^T P \ell$$

$$= x^T (A^T P A x - A^T P \ell) - \ell^T P A x + \ell^T P \ell$$

$$----- = - C^T k_c ------$$

$$\boxed{v^T P v = \ell^T P \ell - x^T A^T P \ell + k_c{}^T w_c}$$

$$\boxed{m_0 = \sqrt{\frac{v^T P v}{n - (u - r)}}}$$

Least squares with conditions

There are practical problems associated with LSQ with conditions:

- The number of parameters grows with the number *r* of scalar condition equations. This may be cumbersome if *r* is a big number.
- A modified minimum principle has to be applied.

There is a "poor woman's/man's solution" to this problem. Instead of asking the condition equations to be met precisely, we may

- interpret the condition equations as observation equations
- form a NEQ system from these NEQs using a diagonal (or even unit) weight matrix
- superimpose this "artificial" NEQ scaled with a big scaling factor $(\sigma_0/\sigma_1)^2$ to the NEQ resulting from the "real" observations.

This alternative is in practice in most cases "good enough". (The Austrians would say "passt").

It is an advantage of this poor woman's/man's solution that the number of parameters is not growing and that the conventional LSQ formulas apply!

Pre-elimination of parameters

Large-scale adjustment problems often may be split up in separate parts, with part-specific and general parameters.

When, e.g., determining the gravity field of the Earth in daily batches (daily NEQs) using kinematic satellite positions as observations the gravity field parameters are general in nature, the orbit parameters part-specific.

The complete NEQ including all parameters has the following structure:

$$\left(\begin{array}{c|c|c|c|c} N & C_1 & C_2 & \cdots & C_n \\ \hline C_1^T & M_1 & 0 & \cdots & 0 \\ \hline C_2^T & 0 & M_2 & \cdots & 0 \\ \hline \vdots & \vdots & \vdots & \ddots & \vdots \\ \hline C_n^T & 0 & 0 & \cdots & M_n \end{array}\right) \cdot \begin{pmatrix} x \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} g \\ h_1 \\ h_2 \\ \vdots \\ h_n \end{pmatrix} = 0$$

Pre-elimination of parameters

The solution of such problems is simple:

- Let us focus on the part *i* of the NEQs and pre-eliminate the parameters $\mathbf{y}_i$.
- Provided that matrix $\mathbf{M}_i$ is regular, the following procedure works and and results for each part in a reduced NEQ.
- Eventually, all reduced NEQs have to be superimposed:

$$N_i x + C_i y_i = g_i$$

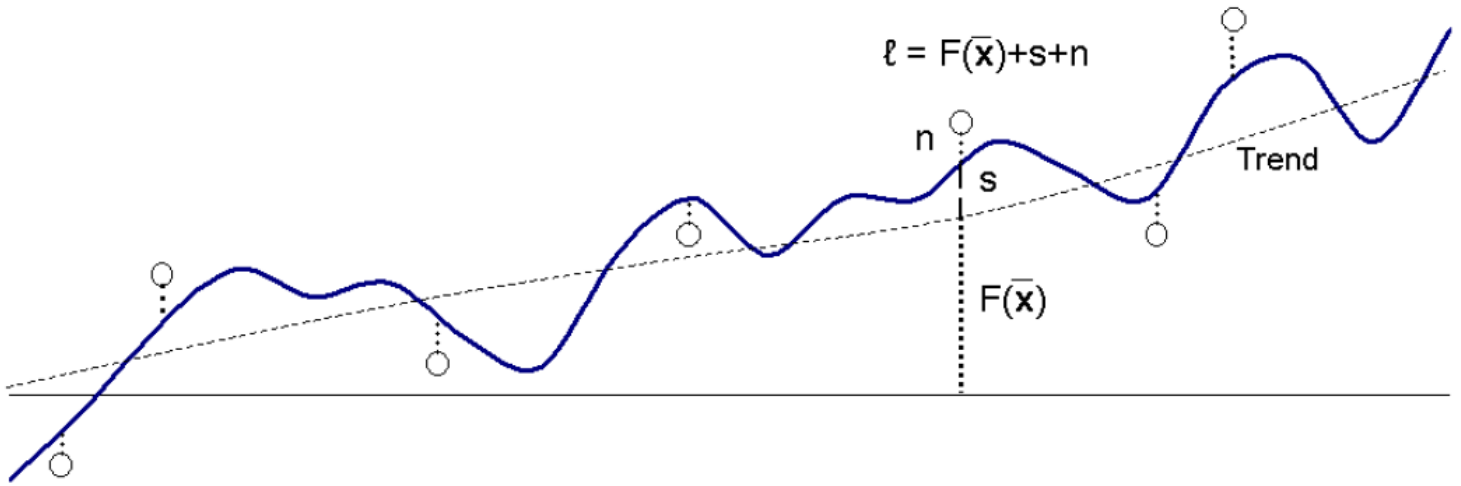
$$C_i^T x + M_i y_i = h_i$$

$$y_i = M_i^{-1}\left(h_i - C_i^T x\right)$$

$$\left(N_i - C_i M_i^{-1} C_i^T\right) x = g_i - C_i M_i^{-1} h_i \rightarrow$$

$$\sum_{i=1}^{n}\left(N_i - C_i M_i^{-1} C_i^T\right) x = \sum_{i=1}^{n}\left(g_i - C_i M_i^{-1} h_i\right)$$

LSQ Collocation



In practice one often encounters the problem that the observed functions cannot be represented with measurement accuracy with a “still reasonable” number *u* of parameters x_i, i=1,2,...,u of a deterministic model.

Collocation works with interim over-parameterization: Two new parameters, namely the „s_i“ and the noise „n_i“ are introduced for each measurement.

If *n* is the number of actual measurements we end up with *u + 2 n* parameters (!).

In the above figure the symbols „o“ represents the actual measurements, the dotted line the deterministic model, and the blue solid line the „deterministic model + signal“.

Collocation

The observation EQs assume the form

$\mathbf{A\,x} + \mathbf{s} + \mathbf{n} = \mathbf{w}$, where $\mathbf{w} = \ell' - \mathbf{F}(x_{01}, x_{02}, \ldots x_{0u})$.

The observation EQs would be those of intermediary adjustment, if "**s**" and "**n**" are replaced by "**-v**".

Introducing matrix **B** the observation EQs may be written as:

$$\mathbf{B\,v + A\,x - w = 0}\ , \text{ where}$$

$$\underset{m.2m}{\mathbf{B}} = (\mathbf{E} \mid \mathbf{E}) = \left(\begin{array}{ccccc|ccccc} 1 & 0 & . & . & 0 & 1 & 0 & . & . & 0 \\ 0 & 1 & . & . & . & 0 & 1 & . & . & . \\ . & . & 1 & . & . & . & . & 1 & . & . \\ . & . & . & . & . & . & . & . & . & . \\ 0 & . & . & . & 1 & 0 & . & . & . & 1 \end{array}\right)$$

$$\underset{1.2m}{\mathbf{v}^T} = \left(\mathbf{s}^T \mid \mathbf{n}^T\right)$$

Collocation

The “observation equations” must hold precisely and therefore should be called condition equations. The weighted sum of residuals shall be minimized under the additional *n* conditions that the observation equations hold precisely:

$$\mathbf{v}^T \mathbf{P}\, \mathbf{v} + 2\, \mathbf{k}\, (\mathbf{B}\, \mathbf{v} + \mathbf{A}\, \mathbf{x} - \mathbf{w}) = \min$$

The *2n* elements of **v** and the *u* elements of **x** are the parameters.

k is the array of Lagrange multipliers. They are in principle “not interesting”, but have to be determined in order to calculate the parameters of interest.

The partial derivatives of the above EQ w.r.t. all elements of arrays **v** and **x** gives the following system of *2n+u* scalar EQs:

$$\mathbf{P}\, \mathbf{v} + \mathbf{B}^T \mathbf{k} = \mathbf{0}$$

$$\mathbf{A}^T \mathbf{k} = \mathbf{0}$$

P is the weight matrix, the inverse of the covariance matrix **C** of **v**.

Collocation

Assuming that **s** und **n** are independent, we have:

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{ss} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{nn} \end{pmatrix}$$

If matrices $\mathbf{C}_{ss}$ and $\mathbf{C}_{nn}$ are regular, matrix $\mathbf{P} = \mathbf{C}^{-1}$ is regular, as well.

We may therefore express **v** with *2*n elements by the vector **k** (with *n* elements) :

$$\mathbf{v} = - \mathbf{P}^{-1} \mathbf{B}^T \mathbf{k} = - \mathbf{C} \mathbf{B}^T \mathbf{k}$$

Using this result in the observation EQs we obtain:

$$- \mathbf{B} \mathbf{C} \mathbf{B}^T \mathbf{k} + \mathbf{A} \mathbf{x} - \mathbf{w} = \mathbf{0} \quad \rightarrow \quad \mathbf{k} = (\mathbf{C}_{ss} + \mathbf{C}_{nn})^{-1} (\mathbf{A} \mathbf{x} - \mathbf{w})$$

and eventually:

$$\mathbf{A}^T (\mathbf{C}_{ss} + \mathbf{C}_{nn})^{-1} \mathbf{A} \mathbf{x} = \mathbf{A}^T (\mathbf{C}_{ss} + \mathbf{C}_{nn})^{-1} \mathbf{w}$$

Collocation

The previous result is formally identical to an adjustment of *n* inter-mediate observation equations with weight matrix $\mathbf{P} = \mathbf{C}_{zz}^{-1} = (\mathbf{C}_{ss}+\mathbf{C}_{nn})^{-1}$.

x may be determined by resolving the previous equation and one easily shows that **s** and **n** may now be determined, as well:

$$\mathbf{v} = \mathbf{C}\mathbf{B}^{\mathsf{T}}\mathbf{k} = -\mathbf{C}\mathbf{B}^{\mathsf{T}}\mathbf{C}_{zz}^{-1}\cdot(\mathbf{A}\mathbf{x}-\mathbf{w})$$

$$= \begin{pmatrix}\mathbf{s}\\ \mathbf{n}\end{pmatrix} = \begin{pmatrix}\mathbf{C}_{ss}\\ \mathbf{C}_{nn}\end{pmatrix}\mathbf{C}_{zz}^{-1}\cdot(\mathbf{w}-\mathbf{A}\mathbf{x})$$

$$\mathbf{n} = \mathbf{C}_{nn}\,\mathbf{C}_{zz}^{-1}(\boldsymbol{\ell} - \mathbf{F}(\bar{\mathbf{x}}))$$

or

$$\mathbf{s} = \mathbf{C}_{ss}\,\mathbf{C}_{zz}^{-1}(\boldsymbol{\ell} - \mathbf{F}(\bar{\mathbf{x}}))$$

Collocation

The matrices $\mathbf{C}_{ss}$ und $\mathbf{C}_{nn}$ decide about the separation of noise and signal.

Note that this separation is not required, if one is only interested in the deterministic parameters x.

How does one obtain the matrix $\mathbf{C}_{zz}$? The following method is often encountered in practice:

- The parameter estimation problem is solved conventionally with the method of intermediary observations *without* signal and noise parts.
- By analyzing the residuals, one determines an empirical covariance function $\mathbf{C}_{zz}$.

Assuming equidistant observations the expectation values $E(r(t)\cdot r(t+k\Delta t))$ of the residuals $r(t_i)$ at times are formed as:

$$E(r(t)\, r(t+k\Delta t)) = \Sigma\, r_i\, r_{i+k} / (n-k),$$

where the sum has to include the indices i=1,2, ...,n-k.

If the observations are not equidistant, “distance classes” must be formed

Collocation

The (symmetric) empirical covariance matrix reads as:

$$C_{zz} = \begin{pmatrix} E(r(t)^2) & E(r(t)\cdot r(t+\Delta t)) & E(r(t)\cdot r(t+2\Delta t)) & \dots & \dots & \dots \\ --- & E(r(t)^2) & E(r(t)\cdot r(t+\Delta t)) & \dots & \dots & \dots \\ --- & --- & E(r(t)^2) & \dots & \dots & \dots \\ \dots & \dots & \dots & & & \\ \dots & \dots & \dots & \dots & \dots & \dots \\ & & & & & E(r(t)^2) \end{pmatrix}$$

The first line of this matrix contains the so-called empirical covariance function, referring to epochs k Δt, k=0,1,2,...,*n-1*, where *n* is the number of observations.

The upper diagonal part of line *j* contains the first n-(*j-1*) values of the first line.

The other elements are obtained by exploiting the symmetry of the matrix.

From LSQ to Kalman Filtering

1. Introduction
2. Deterministic filtering
3. The dynamical system equations
4. Two versions of the filter algorithm
5. Adding the stochastic component → the true Kalman filter

Introduction

So far we assumed:

- All observations are available prior to the adjustment.
- The system is described explicitly by linear EQs.
- The functional model of the observed functions is deterministic.

We first drop the first two, but not the third assumption.

We assume that the observations become available one after the other, as a function of time, and that the parameters shall be determined after each measurement epoch *t*, using all the measurements available up to *t.* The measurements shall be determined using the LSQ method.

A single parameter determination is thus replaced by a parameter estimation process in time. Let us furthermore assume that

- the measured functions are described in the most general case by a particular solution of a non-linear ordinary DEQs
- and that the parameters define a particular solution of the DEQs.

Introduction

The dynamical system is described by the function $\boldsymbol{x}(t) = \boldsymbol{x}(t;x_0)$ of time t. $\boldsymbol{x}(t)$ has d elements (parameters) x_{0i} , $i=1,2,...,d$.

$\boldsymbol{x}(t) = \boldsymbol{x}(t;x_0)$ is the state vector of the system at time t).

At epochs t_j a vector function $\boldsymbol{y}(\boldsymbol{x}(t_j))$ is measured. The function has d' elements, where usually d' << d.

The task might be generalized by allowing for different kinds and numbers of measurements at each epoch. We abstain from this generalization.

The parameter estimation process in time shall give at a particular time t the best possible estimate of the state vector $\boldsymbol{x}(t)$ using all observations available up to this point in time.

It goes without saying (which is why it is said) that the best possible use shall be made for the estimate at t_j of the estimate at t_{j-1}.

Introduction

The state vector of the system $\boldsymbol{x}(t_j;x_0)$, which is a function of $\boldsymbol{x}_0$, shall be estimated of each t_j .

For the time being we will estimate $\boldsymbol{x}_0$ using LSQ.

At time t_j the distinction is made between three tasks:

- Filtering at t_j: Determine $\boldsymbol{x}(t_j;\boldsymbol{x}_0)$ using all measurements l_k' , $k=1,2,\ldots,j$ up to t_j.
- Prediction at t_j: Determine $\boldsymbol{x}(t_j;\boldsymbol{x}_0)$ using all measurements l_k' , $k=1,2,\ldots,j-m$ up to t_{j-m} , $m>0$.
- Smoothing at time t_j: Determine $\boldsymbol{x}(t_j;\boldsymbol{x}_0)$ using all measurements l_k' , $k=1,2,\ldots,j+m$ up to t_{j-m} , $m>0$.

Filtering, prediction and smoothing at t_j are also called estimations.

Note that filtering at the last epoch $t=t_n$ coincides with the classical “en bloc”LSQ solution of the parameter estimation problem.

Subsequently, we will uniquely deal with filtering.

Introduction

Let us write the observation EQs in their usual form:

$$\boldsymbol{l}_j' + \boldsymbol{v}_j = \boldsymbol{y}(\boldsymbol{x}(t_j;\boldsymbol{x}_0)) \;,\; j=0,1,2,...N' \quad (4.1) \text{ , where}$$

- t time,
- t_j measurement epoch No j ,
- $\boldsymbol{l}_j$ array of measurements at t_j ,
- d' dimension (length) of array l_j' ,
- $\boldsymbol{x}(t_j;\boldsymbol{x}_0)$ state vector of the system at t_j ,
- $\boldsymbol{x}_0$ state vector at t_0 , parameter array,
- d dimension *of* $\boldsymbol{x}$ *a*nd $\boldsymbol{x}_0$
- $\boldsymbol{v}_j$ array of residuals at t_j, and
- $N=N'+1$ total number of measurement epochs.

In general $\boldsymbol{y}(...)$ is a non-linear function of the state vector $\boldsymbol{x}(t_j;\boldsymbol{x}_0)$, which is in turn a non-linear function of $\boldsymbol{x}_0$.

If $d' = 1$, $N = n$ is the number of observations (measurements). In general we have $n = d' N$.

Introduction

Assuming that computation time (cpu) is no issue we may solve each filter problem conventionally, i.e., using the LSQ formalism.

The observation EQs. have to be linearlized:

$$\boldsymbol{v}_j = \boldsymbol{A}(t_j)\ \Delta\boldsymbol{x}_0 - (\boldsymbol{l}_j' - \boldsymbol{l}_{j0})\ ,\ j=0,1,2,...N' \qquad (4.2)$$

$\boldsymbol{l}_{j0} = y(x(t_j;\boldsymbol{x}_{00}))$ is calculated using $\boldsymbol{x}_{00}$, the a priori value of $\boldsymbol{x}_0$ known at t_j. $\Delta\boldsymbol{x}_0$ is the increment to these values: $\boldsymbol{x}_0 = \boldsymbol{x}_{00} + \Delta\boldsymbol{x}_0$.

$\boldsymbol{A}$ is a matrix with d' rows and d columns. Element A_{ik} represents the partial derivative of element No. i *of* $\boldsymbol{y}(x_0)$ w.r.t. element k of $\boldsymbol{x}_0$.

$\boldsymbol{A}$, the first design matrix, has to be established using the chain rule, i.e., one first has to take the partial derivatives of $\boldsymbol{y}_j$ w.r.t. the elements of $\boldsymbol{x}(t)$, then the partial derivatives of the elements of $\boldsymbol{x}(t)$ w.r.t. the elements of $\boldsymbol{x}_0(t)$.

System equations

We assume that the dynamical system is defined by an explicit DEQs of 1st order ($\boldsymbol{x}^{(1)}(t)$ stands for the 1st time derivative of $\boldsymbol{x}(t)$):

$$\boldsymbol{x}^{(1)} = \boldsymbol{f}(t; \boldsymbol{x}) \qquad (4.3)$$

$\boldsymbol{x}(t)$ and $\boldsymbol{f}(t,...)$ are arrays with d elements. d is the dimension of the DEQs.

As every DEQs of a higher than the 1st order may be reduced to one of order one, we do not lose generality by considering only DEQs of 1st order.

Equation (4.3) does not yet define a particular solution. We need additional information, e.g., the initial values:

$$\boldsymbol{x}(t_0) = \boldsymbol{x}_0 \qquad (4.4)$$

As the DEQs (4.3) is non-linear, the associated parameter estimation process is non-linear, as well.

→ The DEQs (4.3) has to be linearized first.

System equations

The particular solution of a non-linear DEQs may be linearized, provided approximate initial values $\boldsymbol{x}_{00}$ and thus an approximate solution $\xi(t)$ are available:

$$\xi^{(1)} = f(t;\, \xi)\ ,\ \xi(t_0) = x_{00} \qquad (4.5)$$

Let $\Delta\boldsymbol{x}(t) = \boldsymbol{x}(t) - \xi(t)$. We may obviously write:

$$\Delta\boldsymbol{x}^{(1)} = \boldsymbol{f}(t;\boldsymbol{x}) - \boldsymbol{f}(t;\xi) \qquad (4.6)$$

$$\Delta\boldsymbol{x}\,(t_0) = \boldsymbol{x}_0 - \boldsymbol{x}_{00} = \Delta\boldsymbol{x}_0 \qquad (4.7)$$

Function $f(t;x)$ is developed around $(t;\xi)$ into a Taylor series in $\Delta\boldsymbol{x}(t) = \boldsymbol{x}(t) - \xi(t)$ and truncated after the terms of 1st order:

$$\boldsymbol{f}(t;\boldsymbol{x}) = \boldsymbol{f}(t;\xi) + \mathcal{A}(t)\ \Delta\boldsymbol{x}(t) \qquad (4.8)$$

$\mathcal{A}(t)$ is a matrix with dimension $(d \times d)$, the elements $\mathcal{A}_{ik}$ of which are the partial derivatives of element f_i of array $\boldsymbol{f}(t,\xi)$ *w.r.t.* x_k at (t,ξ).

We thus obtain the linearized DEQs for $\Delta x(t)$:

$$\Delta\boldsymbol{x}^{(1)} = \mathcal{A}(t)\ \Delta\boldsymbol{x} \qquad (4.9)$$

System equations

We have thus shown that every non-linear DEQs may be approximated by a linear, homogeneous DEQs, *provided* an approximate solution $\boldsymbol{\xi}(t)$ of the original DEQs is known.

The theory of digital filters always assumes that the dynamical system may be described by a linear DEQs.

In order to avoid restrictions, and to be compatible with general filter theory, we also allow for inhomogeneous system equations:

$$\Delta \boldsymbol{x}^{(1)} = \boldsymbol{\mathcal{A}}(t)\, \Delta \boldsymbol{x} + \boldsymbol{g}(t) \qquad (4.10)$$

System equations

Important properties of linear DEQs:

- The general solution of the inhomogeneous DEQs (4.10) is obtained as the general solution of the associated homogeneous DEQs and a subsequent „variation of constants“.
- The homogeneous DEQs associated with DEQ (4.10) reads as:
  - $\boldsymbol{x}_h^{(1)} = \mathcal{A}(t)\, \boldsymbol{x}_h \;\; , \boldsymbol{x}_h(t_0) = \Delta \boldsymbol{x}_0$ (4.11)
- The solution is a linear combination of the initial state vector:
  $\boldsymbol{x}_h(t) = \boldsymbol{H}(t,t_0)\, \Delta \boldsymbol{x}_0$ (4.12)
- $\boldsymbol{H}(t,t_0)$ solves the DEQs:
  - $\boldsymbol{H}^{(1)}(t,t_0) = \mathcal{A}(t)\, \boldsymbol{H}(t,t_0)$ (4.13)
- As $\boldsymbol{x}_h(t)$ at t_0 assumes the value $\boldsymbol{x}_0$, we have:
  - $\boldsymbol{H}(t_0,t_0) = \boldsymbol{E}$ (4.14)
- where $\boldsymbol{E}$ is the identity matrix of dimension d.
- Note that matrix $\boldsymbol{H}(t,t_0)$ does not depend on $\Delta \boldsymbol{x}_0$ – and thus may be calculated prior to solving the filter problem.

System equations

The d columns of matrix $\boldsymbol{H}(t,t_0)$ represent a complete system of solutions of the homogeneous DEQs (4.11) implying that any solution of EQ (4.11) may be written as a LC of $\boldsymbol{H}(t,t_0)$.

EQ. (4.12) says that the solution of the homogeneous DEQs (4.11) is a linear combination of the initial state vector at t_0, an important property, because one is usually interested in

$$x(t_j) = x_j = H(t_j, t_0)\, x_0 \,,$$

and not in $\boldsymbol{x}_0$.

As we also allowed for inhomogeneous linear DEQs in Eq. (4.10) we now have to provide the inhomogeneous solution.

System equations

The inhomogeneous DEQs (4.10) is solved by the method of variation of constants, where it is assumed that the initial state vector (4.12) is a function of time:

$$\Delta \boldsymbol{x}(t) = \boldsymbol{H}(t,t_0)\ \Delta \boldsymbol{x}_0(t) \qquad (4.15)$$

Replacing $\Delta \boldsymbol{x}(t)$ in EQ (4.10) by EQ (4.15) results eventually in:

$$\boldsymbol{H}(t,t_0)\ \Delta \boldsymbol{x}_0^{(1)}(t) = \boldsymbol{g}(t) \qquad (4.16)$$

Multiplying this EQ by $[\boldsymbol{H}(t,t_0)]^{-1} = \boldsymbol{H}(t_0,t)$, integrating both sides from t_0 to t, and meeting the original initial conditions at t_0 gives:

$$\Delta x_0(t) = \Delta x_0 + \int_{t_0}^{t} H^{-1}(t_0,t') \cdot g(t')dt' \qquad (4.17)$$

The solution of the original inhomogeneous DEQs thus reads as:

$$\Delta \boldsymbol{x}(t) = \boldsymbol{H}(t,t_0)\ \Delta \boldsymbol{x}_0 + \boldsymbol{q}(t,t_0) \qquad (4.18)$$

where:

$$q(t,t_0) = \int_{t_0}^{t} H^{-1}(t,t') \cdot g(t') \cdot dt' \qquad (4.19)$$

Observation equations

In order to linearize the observation EQs (4.1) we write:

$$\boldsymbol{y}(\boldsymbol{x}(t))= \boldsymbol{y}(\xi(t)+ \Delta\boldsymbol{x}(t)) =\boldsymbol{y}(\xi(t)) + \boldsymbol{A}' \, \Delta\boldsymbol{x}(t) \qquad (4.20)$$

Matrix $\boldsymbol{A}'$ has dimensions (d' x d). Its elements are the partial derivatives of the elements of $\boldsymbol{y}(\boldsymbol{x}(t))$ w.r.t. the elements of $\boldsymbol{x}(t)$ at $\boldsymbol{x}(t) = \xi(t)$.

Taking into account EQ (4.18) we obtain:

$$\boldsymbol{y}(\boldsymbol{x}(t)) = \boldsymbol{y}(\xi(t)) + \boldsymbol{A}' \, [\, \boldsymbol{H}(t,t_0) \, \Delta\boldsymbol{x}_0 + \boldsymbol{q}(t,t_0) \,] \qquad (4.21)$$

The observation EQs are thus linear in the parameters $\boldsymbol{x}_0$:

$$\boldsymbol{v}_j = \boldsymbol{y}(\boldsymbol{x}(t_j;\boldsymbol{x}_0)) - \boldsymbol{l}_j' \, , \; j=0,1,2,...N'$$

$$\boldsymbol{v}_j = \boldsymbol{A}' \, \boldsymbol{H}(t,t_0) \, \Delta\boldsymbol{x}_0 - (\boldsymbol{l}_j' - \boldsymbol{y}(\xi(t_j)) - \boldsymbol{A}' \, \boldsymbol{q}(t_j,t_0)) \, , \; j=0,1,2,...N' \qquad (4.22)$$

$$\boldsymbol{v}_j = \boldsymbol{A} \, \Delta\boldsymbol{x}_0 - (\boldsymbol{l}_j' - \boldsymbol{l}_{j0}) \, , \; j=0,1,2,...N' \qquad (4.23)$$

$$\boldsymbol{v}_j = \boldsymbol{A} \, \Delta\boldsymbol{x}_0 - \Delta\boldsymbol{l}_j \, , \; j=0,1,2,...N' \qquad (4.24)$$

where:

$$\boldsymbol{A} = \boldsymbol{A}' \, \boldsymbol{H}(t,t_0), \quad \boldsymbol{l}_{j0} = \boldsymbol{y}(\xi(t_j)) + \boldsymbol{A}' \, \boldsymbol{q}(t_j,t_0), \quad \Delta\boldsymbol{l}_j = (\boldsymbol{l}_j' - \boldsymbol{l}_{j0}) \qquad (4.25)$$

Filtering at time t_j, Version 1

We assume that measurements referring to different epochs are uncorrelated. Measurements made at the same epoch may, however, be correlated.

Let $\boldsymbol{P}_j$ be the weight matrix associated with epoch t_j. The index „*j*“ indicates that observations at different epochs may have different accuracies.

Based on the observation EQs (4.24) the NEQs of the parameter estimation process (and its solution) read as::

$$N_j \Delta x_{0j} = b_j \qquad \Delta x_{0j} = Q_j b_j$$

$$N_j = (A^T P A)_j = \sum_{k=0}^{j} A_k^T P_k A_k \doteq \sum_{k=0}^{j} \Delta N_k \qquad Q_j = N_j^{-1}$$

$$b_j = (A^T P \Delta l)_j = \sum_{k=0}^{j} A_k^T P_k \Delta l_k \doteq \sum_{k=0}^{j} \Delta b_k \qquad (4.26)$$

Index „*j*“ shall indicate that all measurements up to and including those of epoch t_j are used.

Filtering at time t_j, Version 1

For the NEQs one immediately obtains the following simple recursive formulas:

$$N_{j+1}\Delta x_{0,j+1} = b_{j+1}$$
$$N_{j+1} = N_j + \Delta N_{j+1} \qquad (4.27)$$
$$b_{j+1} = b_j + \Delta b_{j+1}$$

EQs (4.27) may be applied to all problems consisting of piecewise uncorrelated observations. $\Delta \boldsymbol{N}_{j+1}$ and $\Delta \boldsymbol{b}_{j+1}$ are defined by EQs (4.26).

EQs (4.27) are very flexible in application. The solution vector and the associated covariance matrix only have to be calculated when needed.

Filtering at time t_j, Version 1

The question arises whether recursive formulas exist for the inverted NEQs and the solution vector, as well?:

$$N_{j+1}\Delta x_{0,j+1} = b_{j+1}$$

$$\Delta x_{0,j+1} = Q_{j+1}\left(b_j + \Delta b_{j+1}\right) = Q_{j+1}\left(N_j \Delta x_{0j} + \Delta b_{j+1}\right)$$

$$\Delta x_{0,j+1} = Q_{j+1}\left(\left\{N_{j+1} - \Delta N_{j+1}\right\}\Delta x_{0j} + \Delta b_{j+1}\right)$$

$$\Delta x_{0,j+1} = \Delta x_{0j} + Q_{j+1}A_{j+1}^T P_{j+1}\left\{\Delta l_{j+1} - A_{j+1}\Delta x_{0j}\right\}$$

$$\Delta x_{0,j+1} = \Delta x_{0j} + K_{j+1}\left\{\Delta l_{j+1} - A_{j+1}\Delta x_{0j}\right\} \qquad (4.28)$$

Eq. (4.28) represents the initial state vector, estimated with all observations up to epoch t_{j+1}, as a function of the previous estimate (filter value at time t_j) and the array of residuals pertaining to epoch t_{j+1}, both based on the observations up to epoch t_j .

Filtering at time t_j, Version 1

Matrix $\boldsymbol{K}_j$ is also called gain matrix, the expression in parentheses {....} array of measurement innovation.

The algorithm (4.28) is extremely well suited for real-time applications, where the new measurements may be quickly checked for plausibility (screening of observations).

The simplest version of a data screening simply compares the array of measurement innovation with the (hopefully) known mean errors a priori σ of the measurements and removes outliers based, e.g., on an $(n \cdot \sigma)$–criterion .

Algorithm (4.28) is already a close relative of the Kalman filter. Only two aspects are missing:

- Transformation to a new parameter array for each t_j.
- Replace deterministic by stochastic system EQs.

Both aspects will be dealt with in due time.

Filtering at time t_j, Versions 1 and 2

Algorithm (4.28) has a minor flaw: The gain matrix $\boldsymbol{K}_j$ contains matrix $\boldsymbol{Q}_{j+1}$ for which no recurrence relation was provided. This matrix is calculated using EQ (4.27), via

$$\boldsymbol{N}_{j+1} = \boldsymbol{N}_j + \Delta\boldsymbol{N}_{j+1}\ , \quad \boldsymbol{Q}_{j+1}=(\boldsymbol{N}_{j+1})^{-1} \qquad (4.28a)$$

If the dimension d of the NEQs is a large number and d' a small number, i.e., if $d >> d'$ the above formula cannot be recommended.

This case motivates a recurrence relation for matrix $\boldsymbol{Q}_j$, as well. The algorithm (4.28) will be the same, only the algorithm to calculate $\boldsymbol{Q}_{j+1}$, will change.

Let us point out that we have reached our goal to represent the solution vector at time t_{j+1} recursively as a function of the solution vector at t_j.

Filtering at time t_j, Version 2

Starting from Eq. (4.27) we may write:

$$N_{j+1} = N_j + \Delta N_{j+1} = N_j + A_{j+1}^T P_{j+1} A_{j+1}$$

$$N_{j+1} = N_j \left\{ E + Q_j A_{j+1}^T P_{j+1} A_{j+1} \right\}$$

$$Q_{j+1} = \left\{ E + Q_j A_{j+1}^T P_{j+1} A_{j+1} \right\}^{-1} Q_j \qquad (4.28b)$$

Let us propose to calculate {...}$^{-1}$ as:

$$\left\{ E + Q_j A_{j+1}^T P_{j+1} A_{j+1} \right\}^{-1} = \left[E + Q_j A_{j+1}^T P_{j+1} X_{j+1} A_{j+1} \right] \qquad (4.28c)$$

Obviously {...} [...] = ***E***. One easily verifies that the (*d' x d'*) matrix $\boldsymbol{X}_{j+1}$ must be calculated as:

$$X_{j+1} = -\left(E + A_{j+1} Q_j A_{j+1}^T P_{j+1} \right)^{-1} \qquad (4.28d)$$

Filtering at time t_j, Version 2

Using EQ (28d) in EQ (28c) and the resulting expression in EQ (28b), one obtains the recurrence relation for the co-factor matrix ***Q***:

$$Q_{j+1} = Q_j - Q_j A_{j+1}^T P_{j+1} \left(E + A_{j+1} Q_j A_{j+1}^T P_{j+1} \right)^{-1} A_{j+1} Q_j$$

$$Q_{j+1} = Q_j - Q_j A_{j+1}^T \left(P_{j+1}^{-1} + A_{j+1} Q_j A_{j+1}^T \right)^{-1} A_{j+1} Q_j \qquad (4.28e)$$

Matrix $\boldsymbol{K}_j$ in EQ (4.28) may be calculated either using EQ (4.28a) or EQ (4.28e).

The choice of the formula depends on the particular problem: When using EQ (4.28a) one has to take the inverse of a matrix of dimension (*dxd*)- in EQ (4.28e) one of dimension (*d'xd'*).

(4.28e) shows that it is in principle possible to solve the filter problem without matrices $\boldsymbol{N}_j$.

Drawback: How to initialize $\boldsymbol{Q}_1$?

Filtering at time t_j, Parameter transformation

Algorithms (4.28,3.28a) and (4.28, 3.28e) assume that one and the same array $\Delta \boldsymbol{x}_0$ of parameters is used throughout the filtering process.

We will subsequently replace the parameter array $\Delta \boldsymbol{x}_0$ at epoch t_j by the array $\Delta \boldsymbol{x}_j = \Delta \boldsymbol{x}(t_j)$ using the transformation EQs (4.18) (with $t=t_j$).

The resulting algorithm only needs only minor modifications.

In order to solve the problem we will treat the general problem to perform a parameter transformation in a LSQ environment.

Filtering at time t_j, Parameter transformation

Let us assume that the observation EQs read as:

$$\boldsymbol{v} = \boldsymbol{A}\ \Delta\boldsymbol{x} - \Delta\boldsymbol{l} \qquad (4.29a)$$

The NEQs shall be written as:

$$\boldsymbol{N}\ \Delta\boldsymbol{x} = \boldsymbol{b} \qquad (4.29b)$$

Let a “new” parameterarray $\Delta\boldsymbol{y}$ be defined by the following transformation with the original one, namely $\Delta\boldsymbol{x}$:

$$\Delta\boldsymbol{x} = \boldsymbol{H}\ \Delta\boldsymbol{y} + \boldsymbol{q} \qquad (4.29c)$$

Replacing $\Delta\boldsymbol{x}$ in EQ (4.29a) by (4.29c) results in

$$\boldsymbol{v} = \boldsymbol{AH}\ \Delta\boldsymbol{y} - (\Delta\boldsymbol{l} - \boldsymbol{A}\ \boldsymbol{q}) \qquad (4.29d)$$

Filtering at time t_j, Parameter transformation

The NEQs expressed in the new parameters reads as:

$$(\boldsymbol{AH})^T \boldsymbol{P} (\boldsymbol{AH}) \Delta\boldsymbol{y} = (\boldsymbol{AH})^T \boldsymbol{P} (\Delta\boldsymbol{l} - \boldsymbol{A}\,\boldsymbol{q})$$

$$\boldsymbol{H}^T (\boldsymbol{A}^T\boldsymbol{P}\boldsymbol{A}) \boldsymbol{H}\, \Delta\boldsymbol{y} = \boldsymbol{H}^T \boldsymbol{A}^T\boldsymbol{P} (\Delta\boldsymbol{l} - \boldsymbol{A}\,\boldsymbol{q})$$

$$\boldsymbol{H}^T \boldsymbol{N}\, \boldsymbol{H}\, \Delta\boldsymbol{y} = \boldsymbol{H}^T (\boldsymbol{b} - \boldsymbol{N}\,\boldsymbol{q}) \qquad (4.29e)$$

$$\boldsymbol{N}'\, \Delta\boldsymbol{y} = \boldsymbol{b}' \qquad (4.29f)$$

$$\Delta\boldsymbol{y} = (\boldsymbol{N}')^{-1}\, \boldsymbol{b}' = \boldsymbol{H}^{-1}\boldsymbol{Q}(\boldsymbol{H}^T)^{-1}\, \boldsymbol{H}^T (\boldsymbol{b} - \boldsymbol{N}\,\boldsymbol{q}) = \boldsymbol{H}^{-1} (\Delta\boldsymbol{x} - \boldsymbol{q}) \qquad (4.29g)$$

By comparing EQs (4.29e) and (4.29f) we obtain:

$$\boldsymbol{N}' = \boldsymbol{H}^T \boldsymbol{N}\, \boldsymbol{H} \quad \text{and} \quad \boldsymbol{b}' = \boldsymbol{b} - \boldsymbol{H}^T \boldsymbol{N}\, \boldsymbol{q} \qquad (4.29h)$$

The inverted NEQs is obtained from EQ (4.29g):

$$\boldsymbol{Q}' = \boldsymbol{H}^{-1}\boldsymbol{Q}(\boldsymbol{H}^T)^{-1} \quad \text{and} \quad \Delta\boldsymbol{y} = \boldsymbol{H}^{-1} (\Delta\boldsymbol{x} - \boldsymbol{q}) \qquad (4.29i)$$

The sum $\Sigma\, \Delta\boldsymbol{l}^T\boldsymbol{P}\Delta\boldsymbol{l}$ has to be modified as well:

$$\Sigma\, \Delta\boldsymbol{l}'^T\boldsymbol{P}\Delta\boldsymbol{l}' = \Sigma\, \Delta\boldsymbol{l}^T\boldsymbol{P}\Delta\boldsymbol{l} + 2\, \boldsymbol{q}\, \boldsymbol{b} + \boldsymbol{q}^T\, \boldsymbol{N}\, \boldsymbol{q} \qquad (4.29k)$$

Filtering at time t_j, Algorithm 1

Algorithm I, based on NEQ matrices N_j (best suited for $d << d'$):

(1) Initialize: $\boldsymbol{N}_{d-1} = \Sigma\, \boldsymbol{A}_k^T \boldsymbol{P}_k \boldsymbol{A}_k$, $\boldsymbol{b}_{d-1} = \Sigma\, \boldsymbol{A}_k^T \Delta\boldsymbol{l}_k$, $k=0,1,...,d-1$

$\boldsymbol{Q}_{d-1} = (\boldsymbol{N}_{d-1})^{-1}$, $\Delta\boldsymbol{x}_{0,d-1} = \boldsymbol{Q}_{d-1}\, \boldsymbol{b}_{d-1}$

(2) Filter step $j \rightarrow j+1$, $j= d, d+1, ..., N'$

(1) Optionally: transformation of solution vector $\Delta\boldsymbol{x}_0$

(2) Define $\boldsymbol{A}_{j+1}$, $\boldsymbol{P}_{j+1}$, $\Delta\boldsymbol{l}_{j+1}$

(3) Predict and screen using current estimate $\Delta\boldsymbol{x}_{0j}$

Array of residuals for t_{j+1} with $\Delta\boldsymbol{x}_{0j}$: $\boldsymbol{r}_{j+1} := \Delta\boldsymbol{l}_{j+1} - \boldsymbol{A}_{j+1}\, \Delta\boldsymbol{x}_{0j}$

Check plausibility of r_{j+1} (outlier rejection)

(4) Update solution

$\boldsymbol{N}_{j+1} = \boldsymbol{N}_j + \boldsymbol{A}_{j+1}^T \boldsymbol{P}_{j+1} \boldsymbol{A}_{j+1}$

$\boldsymbol{Q}_{j+1} = (\boldsymbol{N}_{j+1})^{-1}$

$\boldsymbol{K}_{j+1} = \boldsymbol{Q}_{j+1} \boldsymbol{A}_{j+1}^T \boldsymbol{P}_{j+1}$

$\Delta\boldsymbol{x}_{0\,,\,j+1} = \Delta\boldsymbol{x}_{0j} + \boldsymbol{K}_{j+1}\, \boldsymbol{r}_{j+1}$

(3) Optionally: calculate residuals with $\Delta\boldsymbol{x}_{0,j+1}$: $\boldsymbol{r}_{j+1}' := \Delta\boldsymbol{l}_{j+1} - \boldsymbol{A}_{j+1}\, \Delta\boldsymbol{x}_{0,j+1}$

Filtering at time t_j, the Algorithm, Version 2

Algorithm II, based on co-factor matrix $\boldsymbol{Q}_j$ (best suited for $d' << d$):

(1) Initialize: $\boldsymbol{N}_{d-1} = \Sigma\, \boldsymbol{A}_k^T \boldsymbol{P}_k \boldsymbol{A}_k$, $\boldsymbol{b}_{d-1} = \Sigma\, \boldsymbol{A}_k^T \Delta\boldsymbol{l}_k$, $k=0,1,...,d-1$

$\boldsymbol{Q}_{d-1} = (\boldsymbol{N}_{d-1})^{-1}$, $\Delta\boldsymbol{x}_{0,d-1} = Q_{d-1}\, b_{d-1}$ *)

(2) Filter step $j \rightarrow j+1$, $j= d, d+1, ..., N'$

(1) Optionally: Transformation of solution vector Δx_0

(2) Define $\boldsymbol{A}_{j+1}$, $\boldsymbol{P}_{j+1}$, $\Delta\boldsymbol{l}_{j+1}$

(3) Predict and screen using $\Delta\boldsymbol{x}_{0j}$

Array of residuals at t_{j+1} using $\Delta\boldsymbol{x}_{0j}$: $\boldsymbol{r}_{j+1} := \Delta\boldsymbol{l}_{j+1} - \boldsymbol{A}_{j+1}\, \Delta\boldsymbol{x}_{0j}$

Check plausibility of r_{j+1} (outlier rejection)

(4) Update solution

$\boldsymbol{Q}_{j+1} = \boldsymbol{Q}_j - \boldsymbol{Q}_j \boldsymbol{A}_{j+1}^T (\boldsymbol{P}_{j+1}^{-1} + \boldsymbol{A}_{j+1}\, \boldsymbol{Q}_j\, \boldsymbol{A}_{j+1}^T)^{-1}\, \boldsymbol{A}_{j+1}\, \boldsymbol{Q}_j$

$\boldsymbol{K}_{j+1} = \boldsymbol{Q}_{j+1}\, \boldsymbol{A}_{j+1}^T\, \boldsymbol{P}_{j+1}$

$\Delta\boldsymbol{x}_{0\,,\,j+1} = \Delta\boldsymbol{x}_{0j} + \boldsymbol{K}_{j+1}\, \boldsymbol{r}_{j+1}$

(3) Optionally: calculate residuals with $\Delta\boldsymbol{x}_{0,j+1}$: $\boldsymbol{r}_{j+1}' := \Delta\boldsymbol{l}_{j+1} - \boldsymbol{A}_{j+1}\, \Delta\boldsymbol{x}_{0,j+1}$

*) Alternatively: „Plausible“ assumption for $\boldsymbol{Q}_{-1}$ (!!!???!!!)

Stochastic Systems

1. Introduction
2. Stochastic motion of a satellite
3. Treating stochastic systems with LSQ
4. Filtering with stochastic Systems

Introduction

From now on the functional model of observations will contain stochastic components, implying that a system is no longer described by one set of parameters (which are estimated with different sets of observations), but by different set of parameters which themselves define a stochastic process.

In the most general case the physical system is described by a stochastic DEQs – where we will assume linear systems.

The R.H.S. of stochastic DEQs contain random vectors.

We will only deal with simple special cases of stochastic DEQs.

We will in particular study the motion of a satellite in the field of a spherically symmetric Earth perturbed by the high-pass filtered part of drag – which hardly can be treated by a model with few parameters.

Introduction

EQs of motion of a satellite:

$$\ddot{\mathbf{r}} = -GM\frac{\mathbf{r}}{r^3} + \delta\mathbf{f}(t, \mathbf{r}, \dot{\mathbf{r}}) \qquad (5.1)$$

Let $\delta\boldsymbol{f}(t)$ represent the high-pass filtered accelerometer measurement of GRACE-A in along-track direction *S.* The measured accelerations were made a factor of 100 larger.

The "true" orbit is generated by numerical integration of EQs (5.1)

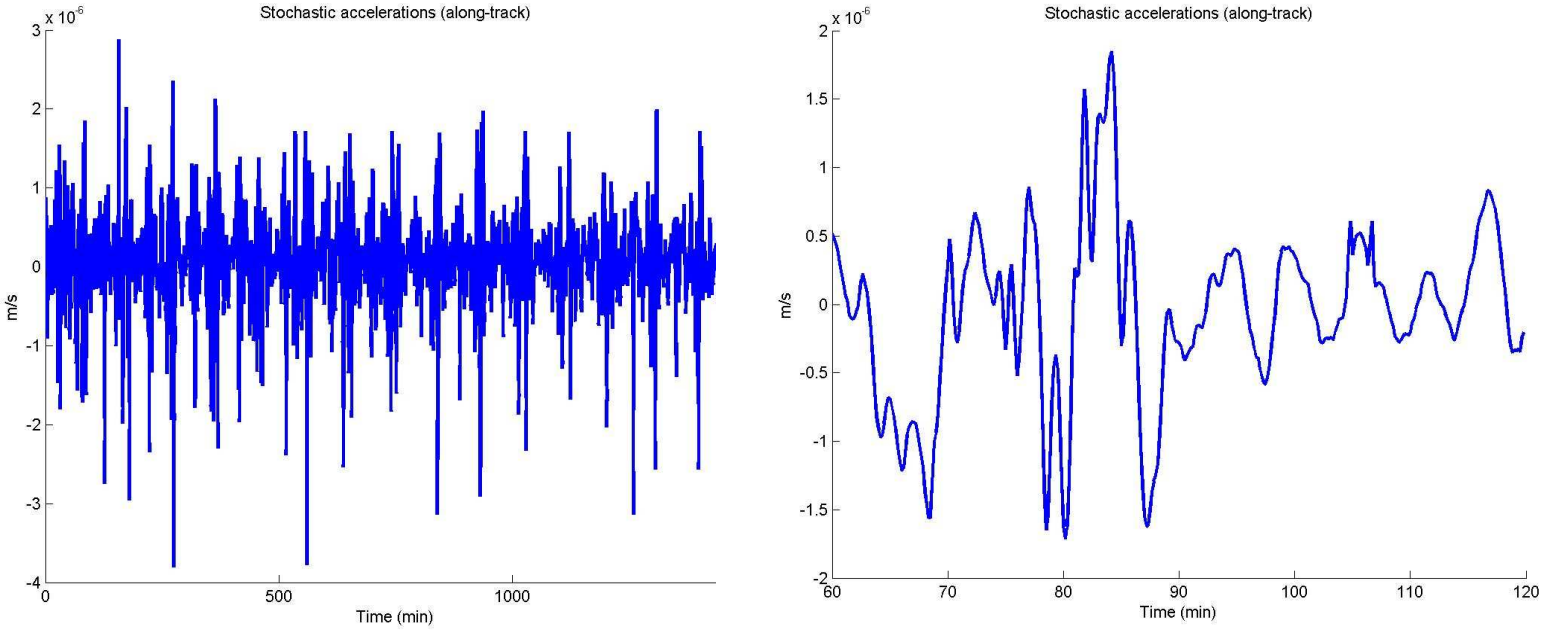
The osculating elements of the true orbit are obtained from the numerically integrated $\boldsymbol{r}(t)$ und $\boldsymbol{v}(t)$:

$$r(t), \dot{r}(t) \Rightarrow a(t), e(t), i(t), \Omega(t), \omega(t), T_0(t) \qquad (5.2)$$

We assume that the Cartesian components of $\boldsymbol{r}(t)$ are observed (GOCE orbit determination!). The spacing of "observations" is 30^s.

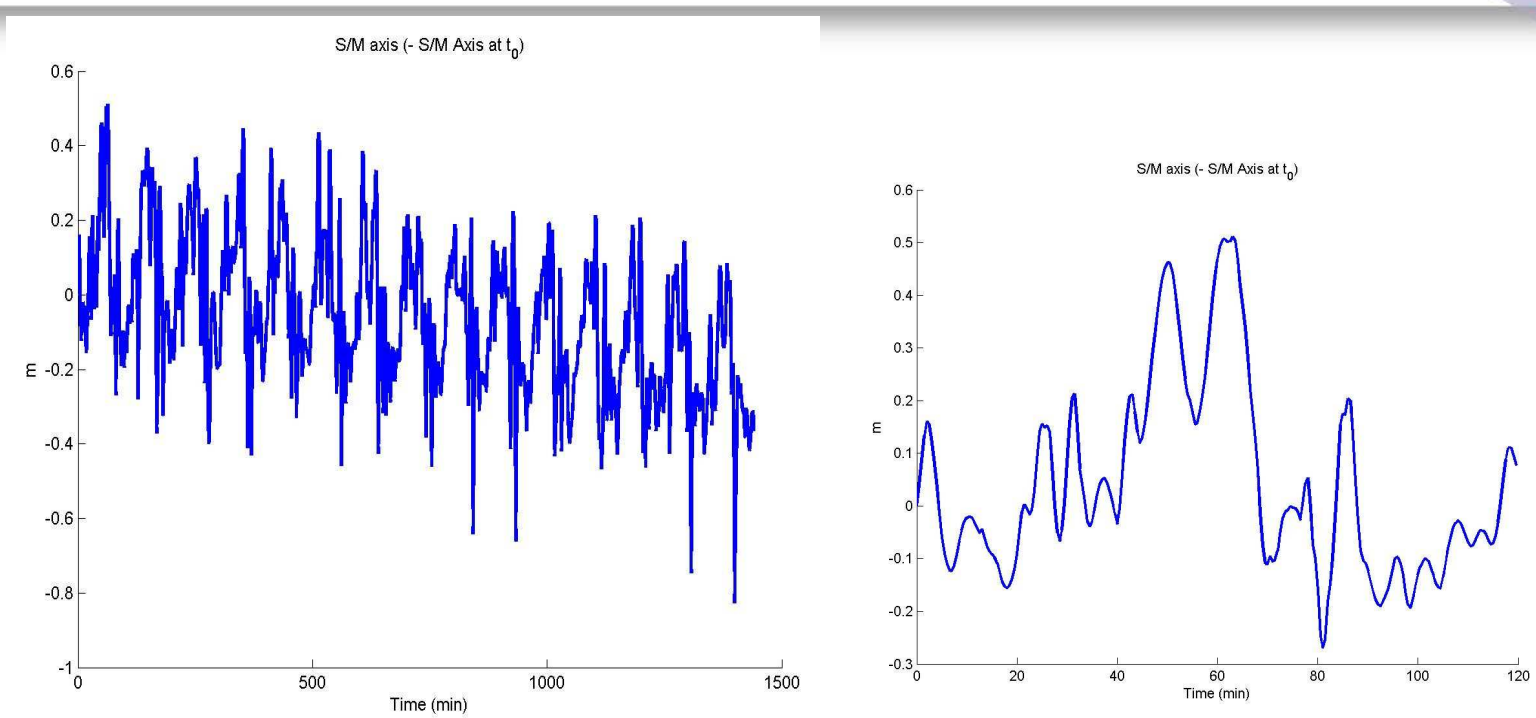
The measurement errors are assumed to be normally distributed with expectation values "0" and $\sigma = 1$ m in each coordinate.

Introduction



$\delta \boldsymbol{f}$ as taken from GRACE (amplified by a factor of 100) – corresponding to a satellite flying about at 300 km.
Left: Data over one day, Right: Data over 2 hours

Introduction

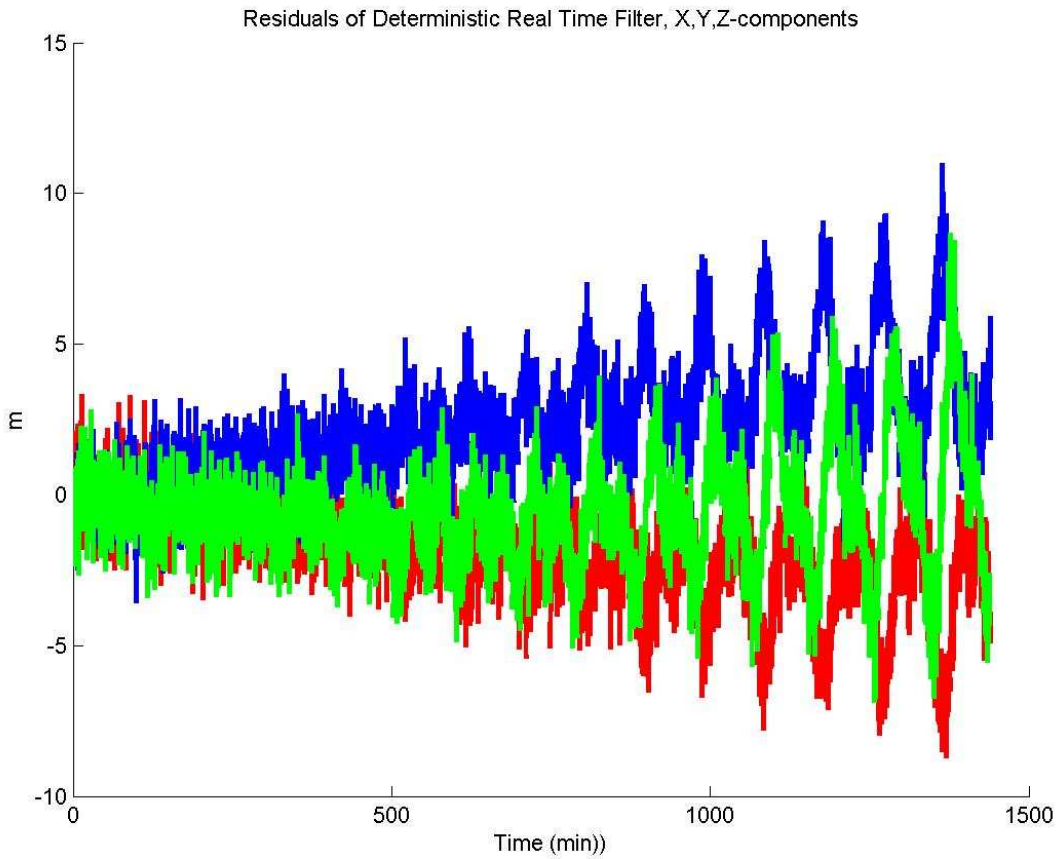


Semi-major axis *a* of the simulated orbit. The deviations are small (sub-m) and chaotic in nature.

If accelerometers are available, such deviations are taken care of.

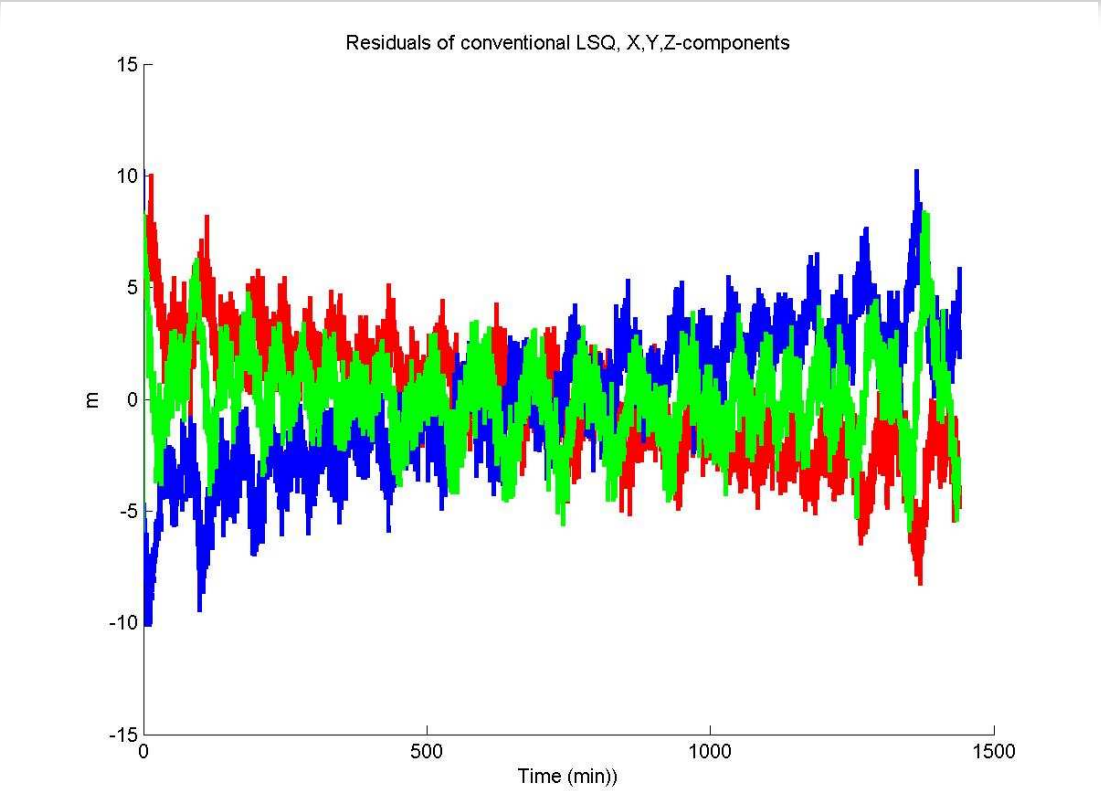
A LEO without accelerometers has a problem ...

Introduction



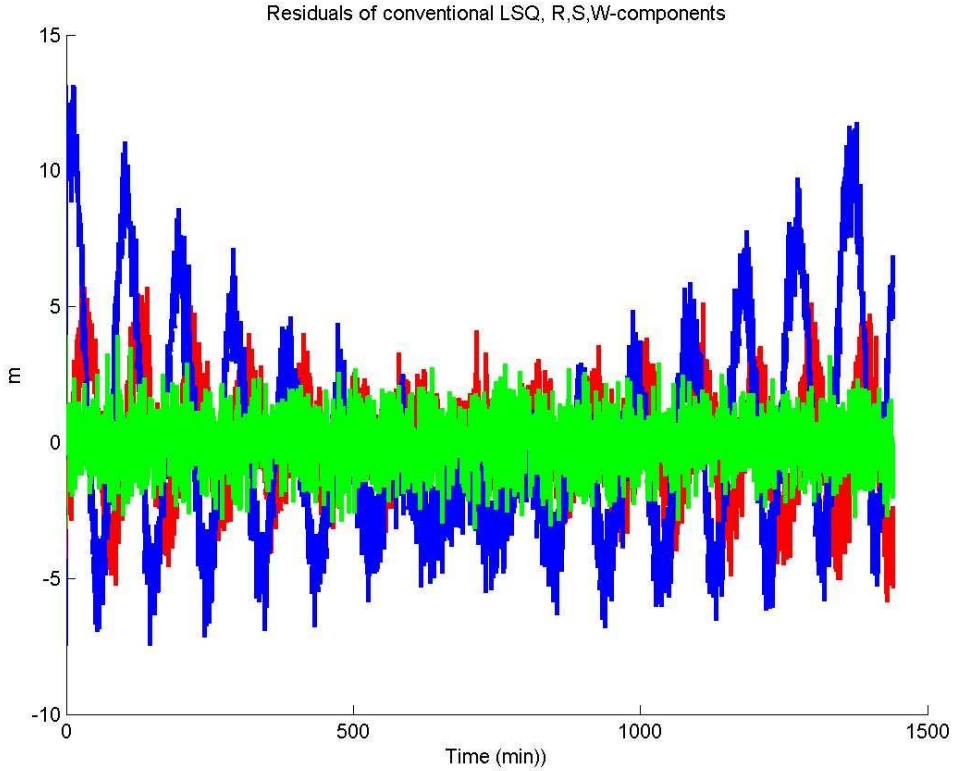
Residuals in X-, Y- und Z- coordinates, as seen by a *deterministic filter*. We see a “filter divergence” (here due to a model deficiency).

Introduction



Residuals of a classical orbit improvement in X, Y, Z are clearly different from the „filtered residuals“. The “wings” are typical for model deficits. Note that the filtered and normal residuals agree towards the end of the day (?).

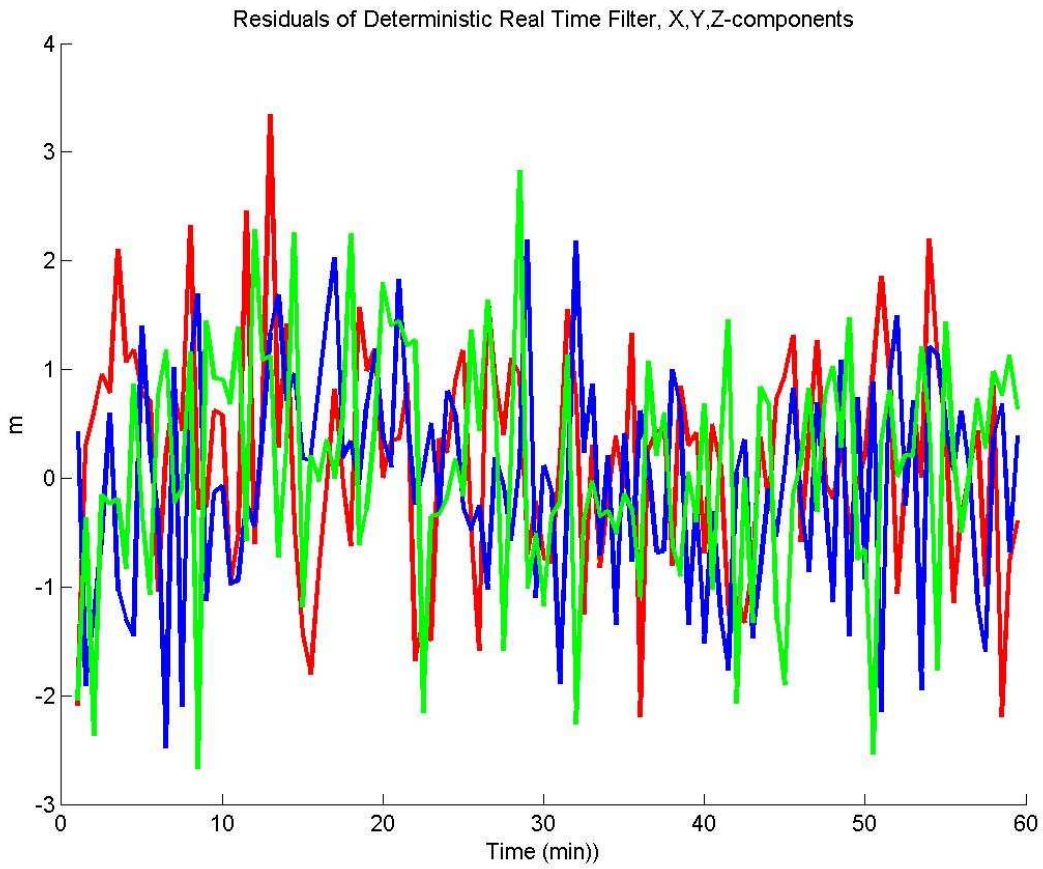
Introduction



The same residuals in R (Radial), S (= R x W), W (out of plane).

The W-component shows, as expected (?), normally distributed errors with σ = 1m).

Introduction



Filter residuals of the first hour. Divergence is not visible. →

Subdividing the one-day arcs into short arcs of 1^h solves the problem!?!?

Introduction

When setting up “new” osculating elements, let us say every 30^m, one automatically introduces a time series of orbit parameters (instead of one set of six parameters).

→ Short-arc methods represent very simple stochastic models.

Short-arc methods are well known and were/are rather successful in gravity field determination. They have the advantage that the intellectual work is minimized …

Problems associated with short arcs:

- Short-arc methods absorb model deficits in the orbit parameters.
- Short-arcs are not contiguous – a problem for some applications (e.g., atmosphere sounding).
- Short arc solutions weaken the entire solution of the parameter estimation problem. One day “needs” 48 x 6 = 288 parameters instead of 6 when introducing independent short arcs every 30^m.

Stochastic Systems

Herewith we dismiss the idea to model a complex dynamical system with only few parameters.

We lapse into the other extreme and introduce for each time interval $[t_i, t_{i+1}]$ defined by two subsequent observation epochs one parameter s_i, representing a constant acceleration in along-track direction. With a spacing of 30^s between observations we have thus introduced 2880 additional parameters (on top of the 6 initial osculating elements)

We are, however, able to constrain these parameters:

- Expectation value $E(s_i) = 0$
- Variance $E(s_i^2) = \sigma_s^2$

From the previous figures we take the estimate $\sigma_s = 0.5 \cdot 10^{-6}\ m/s^2$.

Stochastic Systems

The DEQs (5.1) is replaced for $t \in [t_i, t_{i+1})$ by:

$$\ddot{\mathrm{r}} = -GM \frac{\mathrm{r}}{r^3} + s_i \cdot \mathrm{e}_v \qquad \textbf{(5.2)}$$

EQ (5.2) may be viewed as a simple stochastic DEQ. Only the stochastic properties of the parameters s_i are known!

The task may be solved "easily" as a deterministic problem, if the partial derivatives w.r.t. the parameters s_i are known (consult Beutler (2005), Vol. 1 or *Beutler, Jaeggi et al. „Efficient satellite orbit modelling using pseudo-stochastic parameters", Journal of Geodesy, (2006) 80: pp 353-372*).

Here we circumvent this problem by replacing the s_i *by* stochastic pulses Δv_i defined by (Δt is the spacing between observations):

$$\Delta v_{i+1} = s_i \, \Delta t \qquad (5.3)$$

The acceleration s_i is thus replaced by an instantaneous velocity change $\Delta v_{i+1} \cdot \boldsymbol{e}_{v,i+1}$, where $\boldsymbol{e}_{v,i+1}$ is the unit vector along-track.

Stochastic Systems

The system EQs are those of the two-body problem

$$\ddot{\mathrm{r}} = GM\,\frac{\mathrm{r}}{r^3} \qquad (5.4)$$

The solution of the linearized NEQs at epoch $t=t_i$ read as:

$$\mathrm{r}(t_i) = \mathrm{r}_0(t_i) + \sum_{\mathrm{k}=1}^{6} \frac{\partial \mathrm{r}_0(t_i)}{\partial E_{0k}} \cdot \Delta E_{0k} + \sum_{k=1}^{i} \frac{\partial \mathrm{r}_0(t_i)}{\partial \Delta v_k} \cdot \Delta v_k \qquad (5.5)$$

At $t=t_i$ we thus have to solve for $6+i$ parameters. The number of parameters grows linearly with time t !

Note that the LSQ solution at $t=t_j$ will provide values for all pulses set up till $t=t_i$.

The pulses Δv_i are “normal” parameters of the parameter estimation process!

Stochastic Systems

How are the partial derivatives w.r.t. the Δv_i calculated? We already showed how this is done, bur repeat the procedure here.

By introducing a Δv_i, we define new initial values at t_i.

The linearized task is described by a homogeneous linear DEQs.

The new initial values may be represented as LC of the six partial derivatives w.r.t. the initial osculating elements.

→ The partial derivatives w.r.t. each Δv_i may be written as a LC of the partial derivatives w.r.t. six osculating elements:

$$\frac{\partial \mathrm{r}_0(t)}{\partial \Delta v_i} = \sum_{\mathrm{k}=1}^{6} \beta_{ik} \frac{\partial \mathrm{r}_0(t)}{\partial E_{0k}} \quad ; \quad \frac{\partial \dot{\mathrm{r}}_0(t)}{\partial \Delta v_i} = \sum_{\mathrm{k}=1}^{6} \beta_{ik} \frac{\partial \dot{\mathrm{r}}_0(t)}{\partial E_{0k}} \qquad (5.6)$$

Stochastic Systems

The coefficients β_{ik} are determined by conditions to be applied at t_i:

$$\Delta \mathrm{r}(t_i) = \frac{\partial \mathrm{r}_0(t_i)}{\partial \Delta v_i} \cdot \Delta v_i = \sum_{\mathrm{k}=1}^{6} \frac{\partial \mathrm{r}_0(t_i)}{\partial E_{0k}} \cdot \Delta E_{ik} = 0$$

$$\Delta \dot{\mathrm{r}}(t_i) = \frac{\partial \dot{\mathrm{r}}_0(t_i)}{\partial \Delta v_i} \cdot \Delta v_i = \sum_{\mathrm{k}=1}^{6} \frac{\partial \dot{\mathrm{r}}_0(t_i)}{\partial E_{0k}} \cdot \Delta E_{ik} = \Delta v_i \cdot \mathrm{e}_{v_i}$$

Implying that a system of six equations has to be solved to determine the coefficients β_{ik}:

$$\sum_{\mathrm{k}=1}^{6} \frac{\partial \mathrm{r}_0(t_i)}{\partial E_{0k}} \cdot \beta_{ik} = 0$$

$$\sum_{\mathrm{k}=1}^{6} \frac{\partial \dot{\mathrm{r}}_0(t_i)}{\partial E_{0k}} \cdot \beta_{ik} = \mathrm{e}_{v_i} \qquad (5.7)$$

$$\text{where}: \beta_{ik} = \Delta E_{ik} / \Delta v_i$$

Stochastic Systems

Note that each β_{ik} tells by how much the osculating element E_{0k} changes, when the velocity along-track changes at $t=t_i$ by $\Delta v_i = 1$ meter.

Let us now add the stochastic information concerning Δv_i:

From

- Expectation value $E(s_i) = 0$
- Variance $E(s_i^2) = \sigma_s^2$,
- with $\sigma_s = 0.5 \cdot 10^{-6}\ m/s^2$

we conclude, using $\Delta v_{i+1} = s_i\ \Delta t$, that we may define the observation equation

$$\Delta v_{i+1} = 0 \qquad (5.8a)$$

with the weight

$$P_{vi} = \sigma_0^2 / (\Delta t\ \sigma_s)^2 = 1/(30 \cdot 10^{-6})^2 \qquad (5.8b)$$

Solving the stochastic problem with LSQ

The LSQ solution proceeds in the following steps. We define …

- ... the number of observation epochs (avoid really large dimensions)
- ... the coefficients β_{ik}, *k=1,2,..,6,* for each pulse

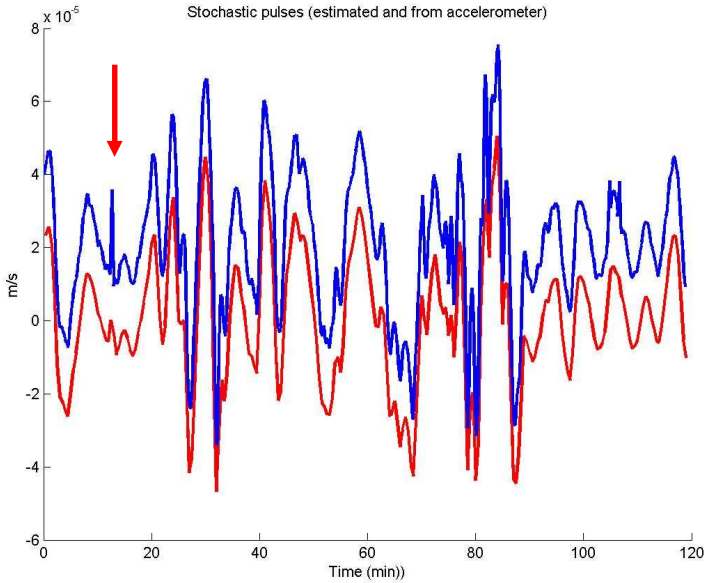
We set up

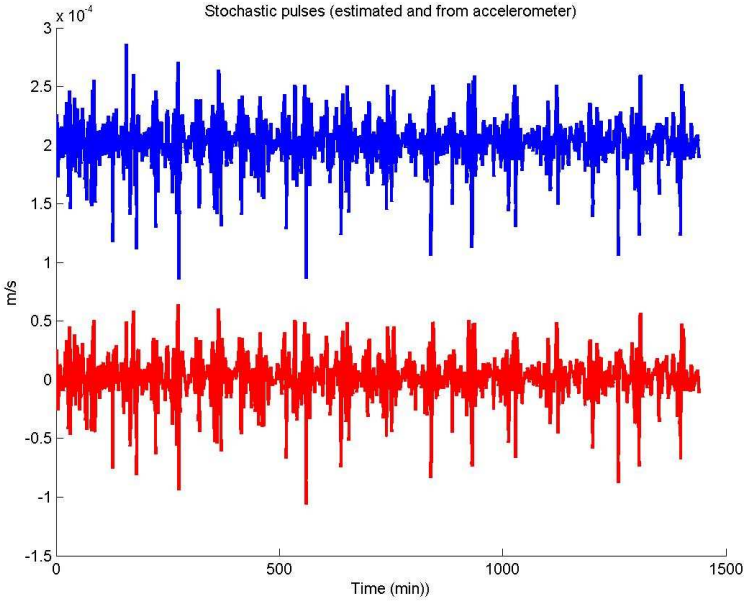
- ... set up the observation equations,
- ... set NEQs,

We superimpose …

- ... the NEQs for the individual Δv_i to the NEQs resulting from the "real" observations
- ... Solve the combined NEQs
- ... Calculate the residuals,
- ... Plot the pulses and compare them with the introduced accelerations.
- ... calculate the osculating elements associated with each observation epoch and compare them with the elements generated in the simulation.

Solving the stochastic problem with LSQ



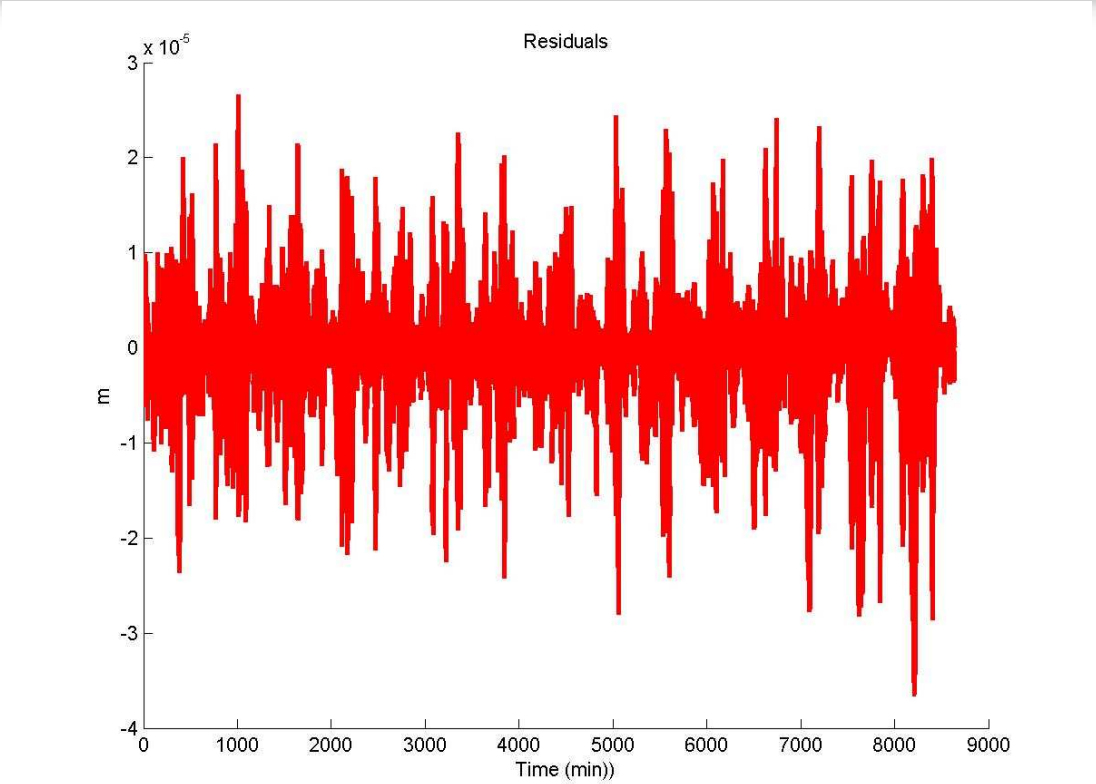


Experiment 1 – „error-free obs“: $\sigma_0 = 1 \cdot 10^{-7}$ m, analyze first two hours of the day. $P_{vi} = \sigma_0^2 / (\Delta t\, \sigma_s)^2 = 1.11 \cdot 10^{-5}$ is rather small.

The estimated pulses and the accelerometer-derived values are represented.

The pulses are recovered almost precisely (exception at „→).

Solving the stochastic problem with LSQ



Experiment 1 – „error-free obs“: $\sigma_0 = 1 \cdot 10^{-7}$ m, analyze first two hours of the day. $P_{vi} = \sigma_0^2 / (\Delta t\, \sigma_s)^2 = 1.11 \cdot 10^{-5}$ is rather small. The figure shows the residuals. No systematics.

Solving the stochastic problem with LSQ

After the adjustment the initial osculating elements and the pulses are known. The associated changes in the initial elements

$$\Delta E_i^T := (E_{i1}, E_{i2}, E_{i3}, E_{i4}, E_{i5}, E_{i6}),\ i=1,2,...,n-1 \qquad \textbf{(5.9a)}$$

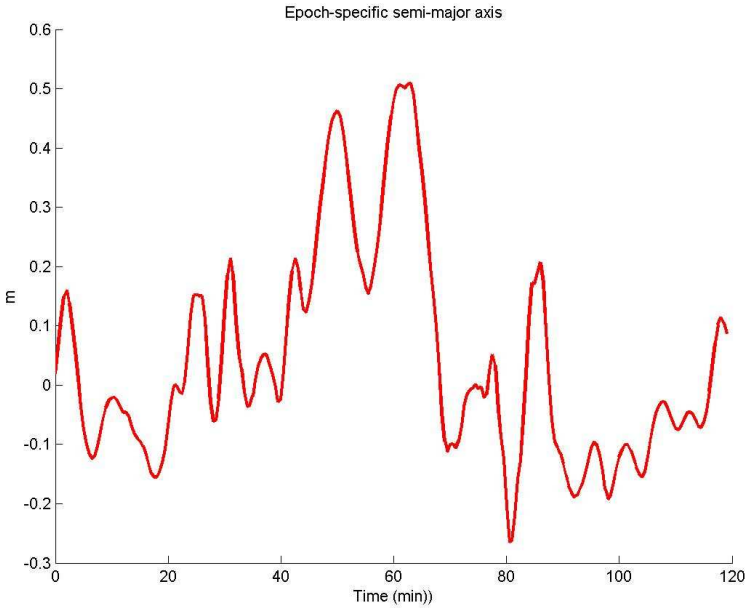
may now be calculated using the matrices **(5.9b)**

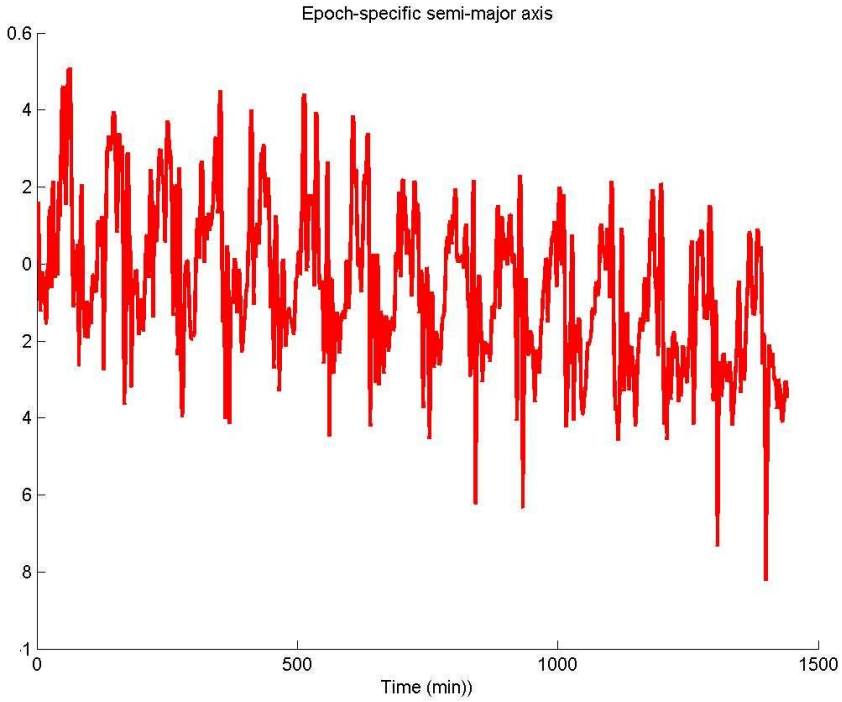
$$B_i^T = (\beta_{i1}, \beta_{i2}, \ldots, \beta_{i6}) :$$

$$\Delta E_i = \Delta E_0 + \sum_{k=1}^{i} \Delta v_k \cdot B_k \qquad \textbf{(5.9)}$$

These epoch-specific orbital elements may now be compared with the orbital elements emerging from the simulation-.

Solving the stochastic problem with LSQ

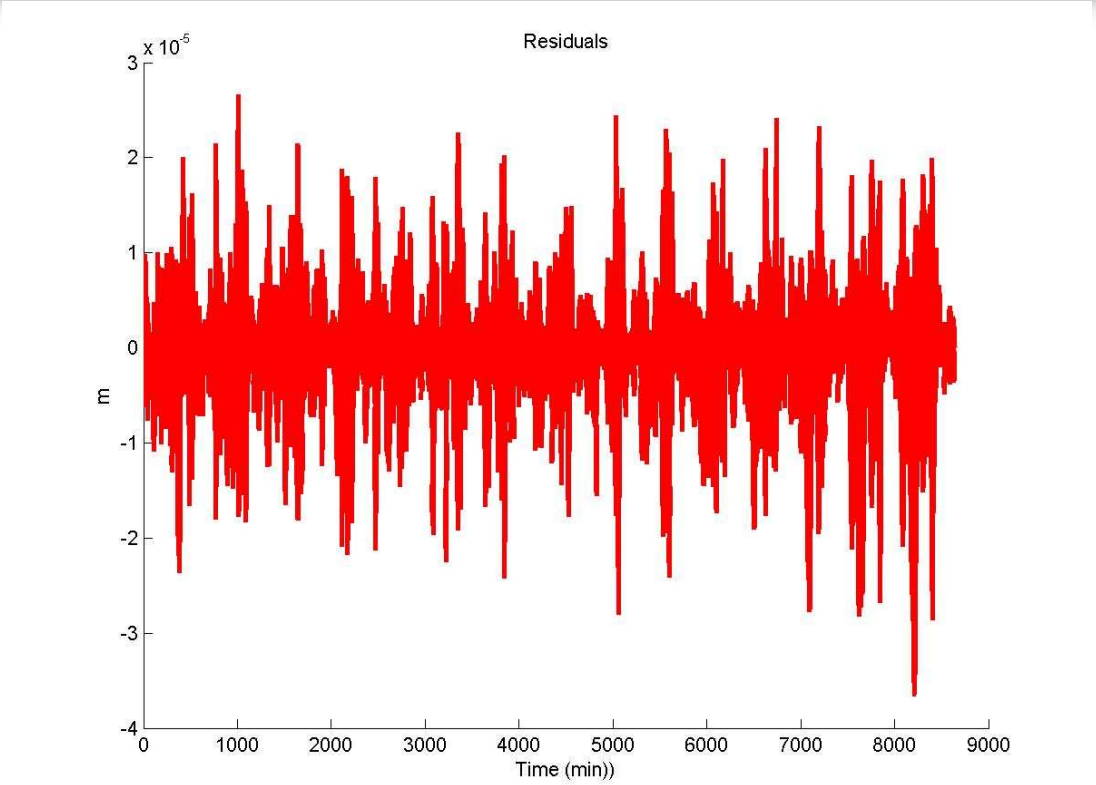




Experiment 1 – "error-free" obs: $\sigma_0 = 1 \cdot 10^{-7}$. The first two hours of the day are analyzed. Note that $P_{vi} = \sigma_0^2 / (\Delta t\, \sigma_s)^2 = 1.11 \cdot 10^{-5}$ is small.

The semi-major axes are reconstructed almost perfectly!

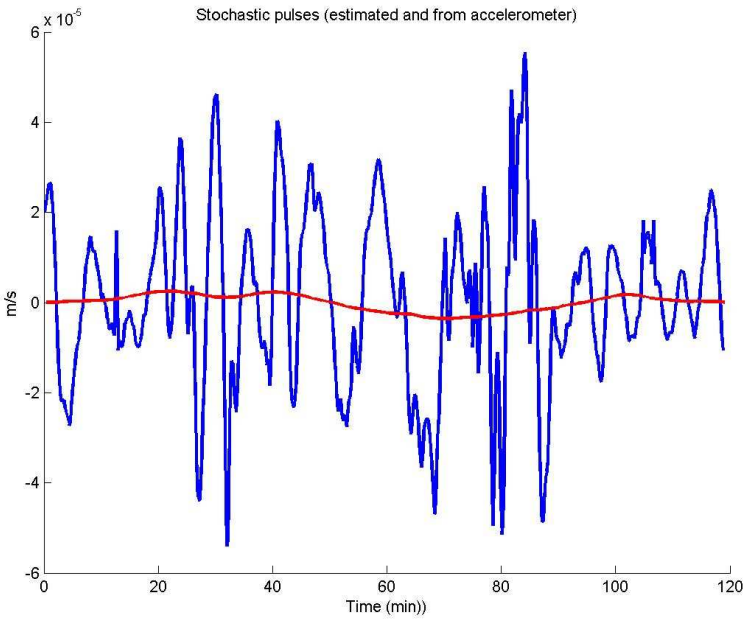
Solving the stochastic problem with LSQ

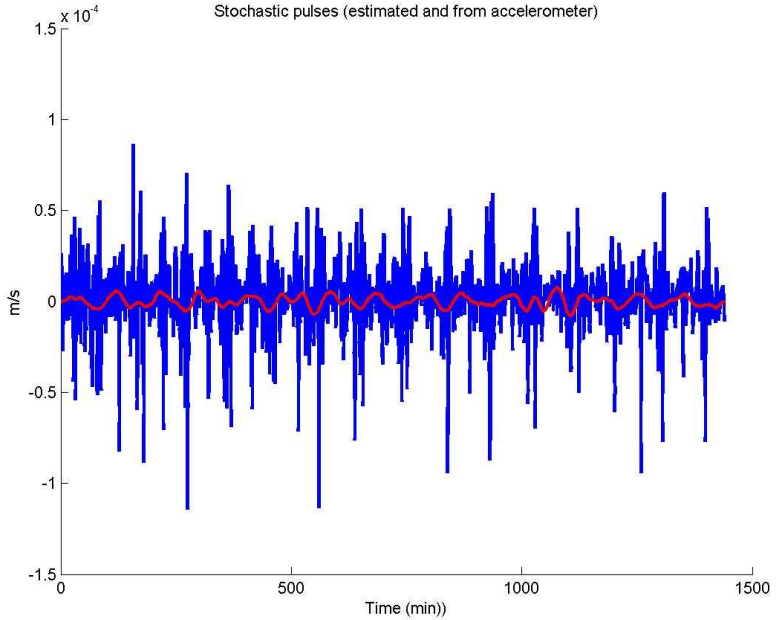


Experiment 1 – "error-free" obs: $\sigma_0 = 1 \cdot 10^{-7}$. The first two hours of the day are analyzed. Note that $P_{vi} = \sigma_0^2 / (\Delta t\, \sigma_s)^2 = 1.11 \cdot 10^{-5}$ is small.

The residuals do not show any systematics.

Solving the stochastic problem with LSQ

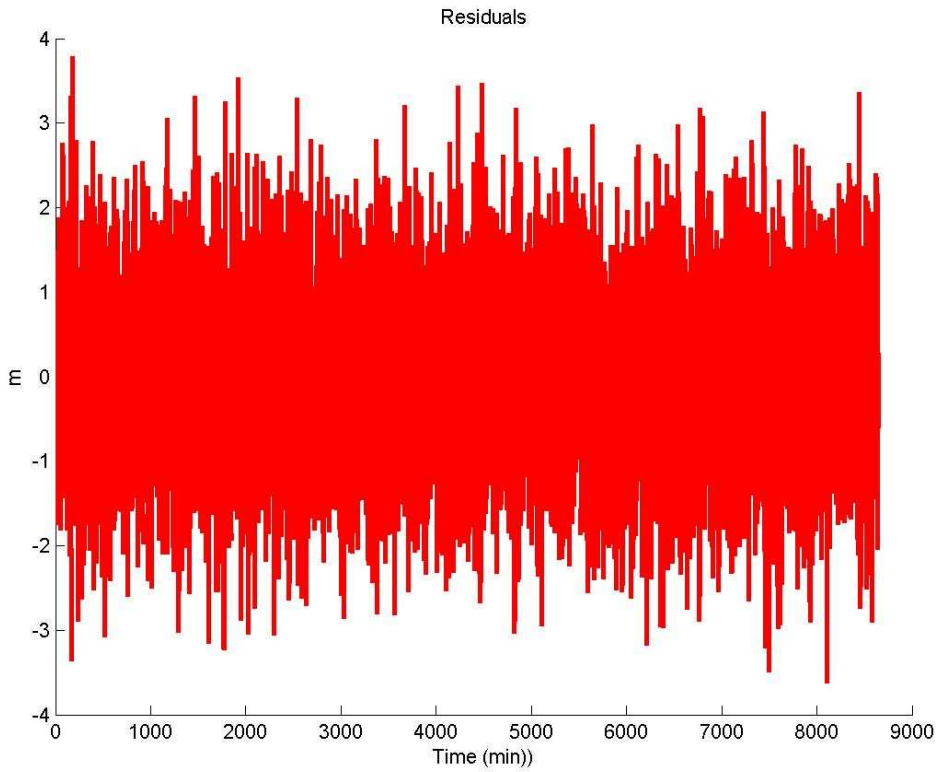




Experiment 2 – $\sigma_0 = 1m$: The first two hours and the entire days are analyzed. Note that $P_{vi} = \sigma_0^2 / (\Delta t\, \sigma_s)^2 = 1.11 \cdot 10^{+9}$ is big. The estimated and true pulses are shown..

With $\sigma_0 = 1m$ the individual pulses are heavily smoothed.

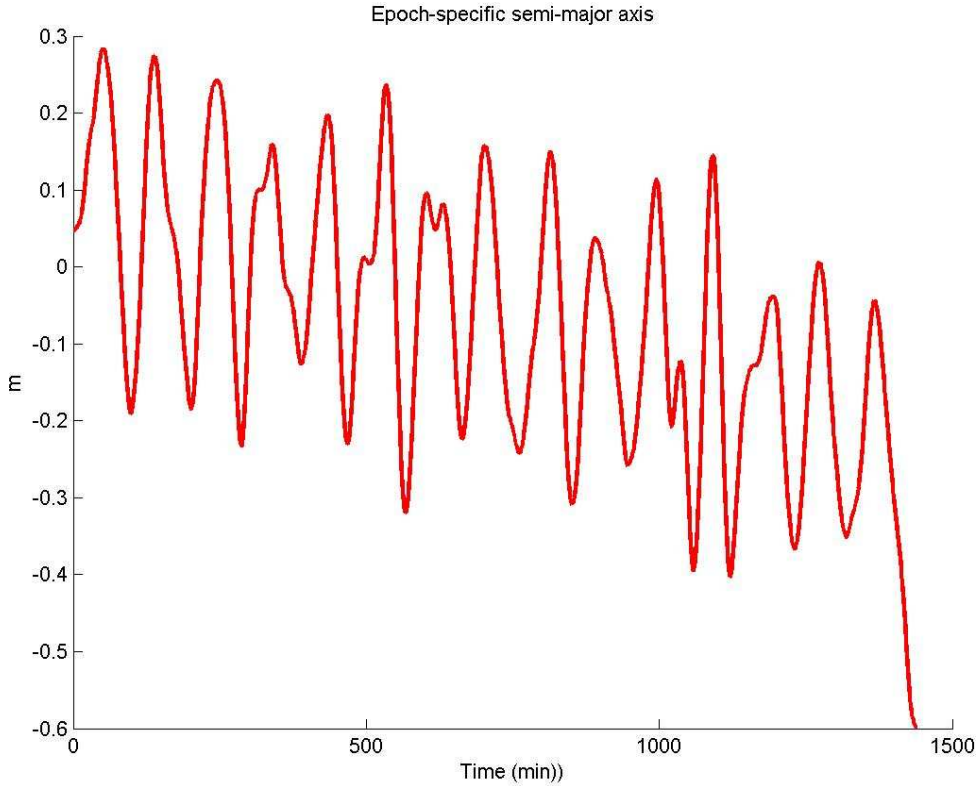
Solving the stochastic problem with LSQ



Experiment 2 – $\sigma_0 = 1m$: The first two hours and the entire days are analyzed. Note that $P_{vi} = \sigma_0^2 / (\Delta t\, \sigma_s)^2 = 1.11 \cdot 10^{+9}$ is big.

The residuals do not show obvious systematics

Solving the stochastic problem with LSQ



Experiment 2 – $\sigma_0 = 1m$: The first two hours of the day are analyzed. Note that $P_{vi} = \sigma_0^2 / (\Delta t\, \sigma_s)^2 = 1.11 \cdot 10^{+9}$ is big.

Note that the semi-major axes (as opposed to the pulses) are rather well reconstructed.

Solving the stochastic problem with LSQ

Summary:

The „error-free“ experiment shows that

- the pulses reconstruct (after division by Δt) the simulated accelerations very well,
- there are no systematics in the residuals,
- the orbital elements could be perfectly reconstructed and
- the observations of Δv_i had no impact on the results.

The experiment with $\sigma_0 = 1m$ shows that

- the estimated pulses have nothing to do with the simulated accelerations.
- there are no systematics in the residuals,
- the orbit elements were smoothed but that the major features correspond well to the simulations!

Filtering the stochastic problem

Our treatment of the stochastic DEQ was mathematically correct, but the resulting algorithm would be very inefficient. Computation time and disk space are considerable. For real-time applications the method is not useful.

The efficient solution to be presenting now will result in epoch-specific parameter estimation procedures with only six parameters!

At the end of the parameter estimation process the general parameters will be the same in both methods.

Drawback: The pulses Δv_i cannot be compared in the LSQ and the filtered method, because they are calculated only with the observations till t_i in the filter approach, whereas all values were used in the LSQ method.

A back-substitution process might be used to reconstruct the LSQ-values in the filter approach.

Filtering the stochastic problem

We perform a transformation of the initial osculating elements at each observation epoch, which absorbs the impact of the preceding stochastic pulses.

The basic relation was already provided in Eq. (5.9). From this EQ we may derive the following recurrence relation:

$$\Delta E_i = \Delta E_0 + \sum_{k=1}^{i} \Delta v_k \cdot B_k \rightarrow$$
$$\Delta E_i = \Delta E_{i-1} + B_i \cdot \Delta v_i \qquad (5.10)$$

The NEQs with parameters ΔE_{i-1},may be written as:

$$\boldsymbol{N}_{i-1}\ \Delta E_{i-1} = b_{i-1} \qquad (5.11)$$

Using Eq. (5.10) we now perform the following transformation

$$\Delta E_{i-1} = \Delta E_i - \boldsymbol{B}_i\ \Delta v_i \qquad (5.12)$$

Filtering the stochastic problem

The transformation allows it at t_i to superimpose the NEQ resulting from the pseudo-observation (5.8a) with (5.8b) for index *i* with the "old" NEQ $\boldsymbol{N}_i$.

Immediately afterwards Δv_i is pre-eliminated.

The impact of Δv_i from observations with index $k > i$ is taken care of in the new array of elements ΔE_i).

Let (where $\boldsymbol{U}_6$ is the identity matrix of dimension 6):

$$\Delta E_{i-1} = \left(U_6, -B_i\right)\begin{pmatrix}\Delta E_i \\ \Delta v_i\end{pmatrix} \qquad (5.13)$$

Filtering the stochastic problem

After the transformation (5.13) the NEQs (5.11) assumes the form:

$$\begin{pmatrix} U_6 \\ -B_i^T \end{pmatrix} N_{i-1} \begin{pmatrix} U_6, -B_i \end{pmatrix} \begin{pmatrix} \Delta E_i \\ \Delta v_i \end{pmatrix} = \begin{pmatrix} U_6 \\ B_i^T \end{pmatrix} b_{i-1}$$

$$\begin{pmatrix} N_{i-1} & -N_{i-1}B_i \\ -B_i^T N_{i-1} & B_i^T N_{i-1} B_i \end{pmatrix} \begin{pmatrix} \Delta E_i \\ \Delta v_i \end{pmatrix} = \begin{pmatrix} b_{i-1} \\ -B_i^T b_{i-1} \end{pmatrix} \qquad (5.13)$$

We may now add observation EQ (5.8a) with weight (5.8b)

$$\begin{pmatrix} N_{i-1} & -N_{i-1}B_i \\ -B_i^T N_{i-1} & B_i^T N_{i-1} B_i + P_{v_i} \end{pmatrix} \begin{pmatrix} \Delta E_i \\ \Delta v_i \end{pmatrix} = \begin{pmatrix} b_{i-1} \\ -B_i^T b_{i-1} \end{pmatrix} \qquad (5.14)$$

Filtering the stochastic problem

The last step consists of pre-eliminating parameter Δv_i – a procedure, which we already showed.

The result reads as:

$$N_{i-1}^{*}\Delta E_i = b_{i-1}^{*} \tag{5.15}$$

$$N_{i-1}^{*} = N_{i-1} - N_{i-1}B_i\left(B_i^T N_{i-1}B_i + P_{v_i}\right)^{-1} B_i^T N_{i-1} \tag{5.15a}$$

$$b_{i-1}^{*} = b_{i-1} - N_{i-1}B_i\left(B_i^T N_{i-1}B_i + P_{v_i}\right)^{-1} B_i^T b_{i-1} \tag{5.15b}$$

The transformation from EQ (5.11) to the system (5.15) thus consists of three steps:

(a) Parameter transformation $\Delta E_{i-1} \rightarrow \Delta E_i$

(b) Add information (5.8a,b) for Δv_i

(c) Pre-eliminate Δv_i

Filtering the stochastic problem

The NEQ contribution from the observations at t_i may now be conventionally added to Eq. (5.15):

$$
\begin{aligned}
&(N_{i-1}^{*} + A_i^T P_i A_i)\Delta E_i = b_{i-1}^{*} + A_i^T P_i \Delta l_i \rightarrow \\
&N_i \Delta E_i = b_i \qquad (5.16) \\
&\Delta E_i = N_i^{-1} b_i = Q_i b_i
\end{aligned}
$$

We have thus shown that the original parameter estimation task may be solved very efficiently with only six parameters.

Note that the parameters ΔE_i at t_i are defined by EQ (5.9). They define the state vector at t_i!

Filtering the stochastic problem

When combining EQs (5.11), (5.15a,b) and (5.16) we see that the update step in the case of stochastic filtering can be made formally equivalent to deterministic filtering (compare EQs (4-27)):

$$\boldsymbol{N}_{i+1}\ \Delta E_{i+1} = b_{i+1}$$

(5.17)

$$\boldsymbol{N}_{i}\ \ \Delta E_{i}\ \ \ = b_{i}$$

with:

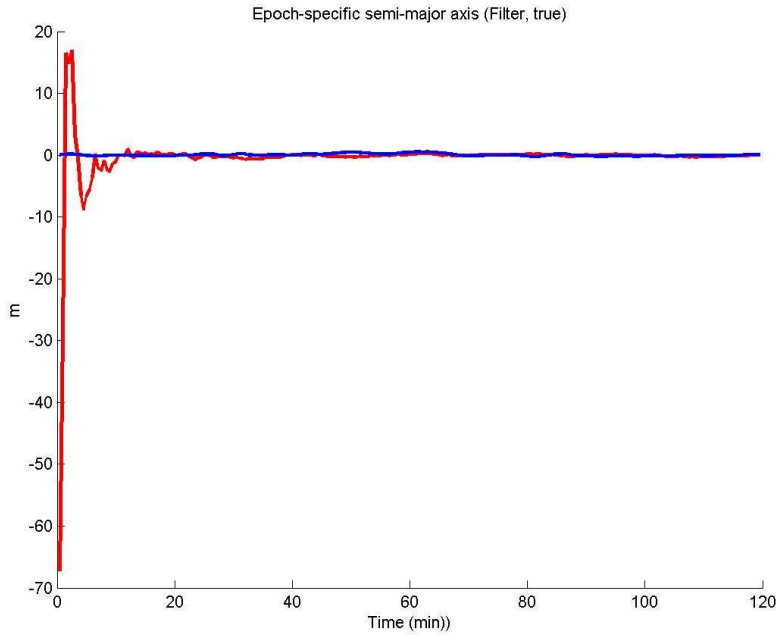
$$\boldsymbol{N}_{i+1} = \boldsymbol{N}_{i} + \Delta \boldsymbol{N}_{i}$$

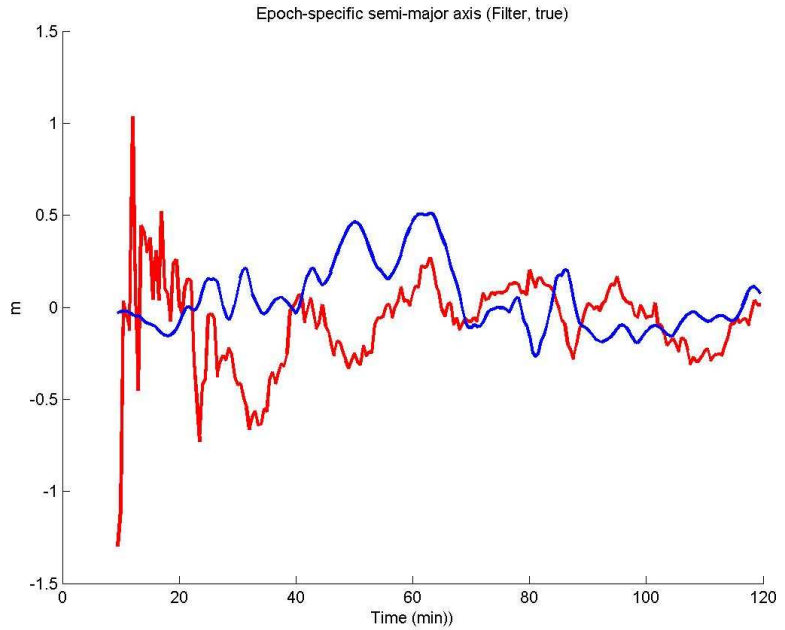
(5.18)

$$b_{i+1} = b_{i} + \Delta b_{i}$$

The only difference consists of the algorithm to calculate the increments. We are thus able to take over the formulas (4.28), (4.28a-e) of Chapter 4.

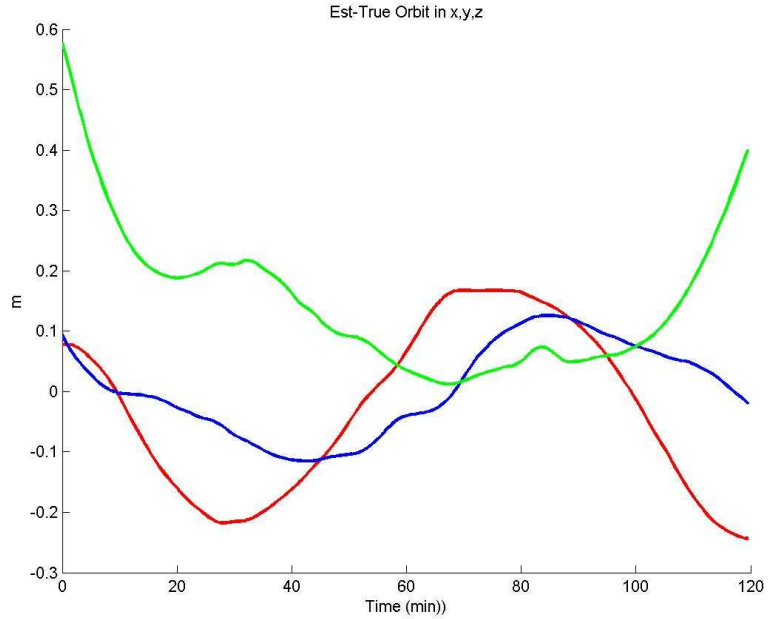
Filtering the stochastic problem

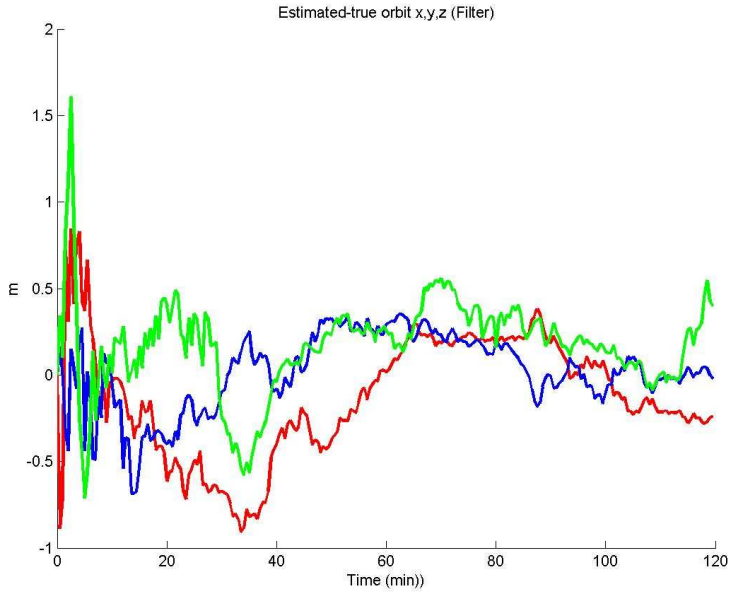




Filtered and true semi-major axis. The initial transient oscillation is clearly visible. Left: starting with the second, right: starting with observation 20 (after 10 min).

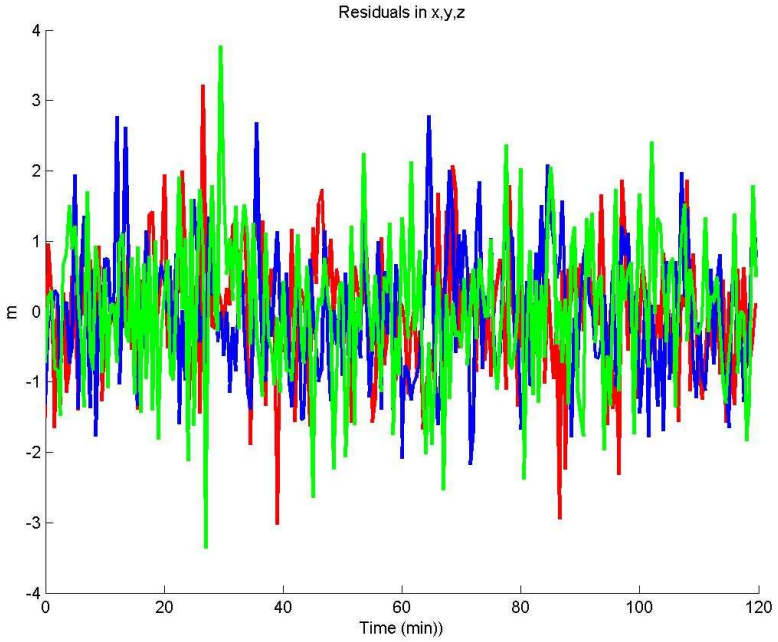
Filtering the stochastic problem

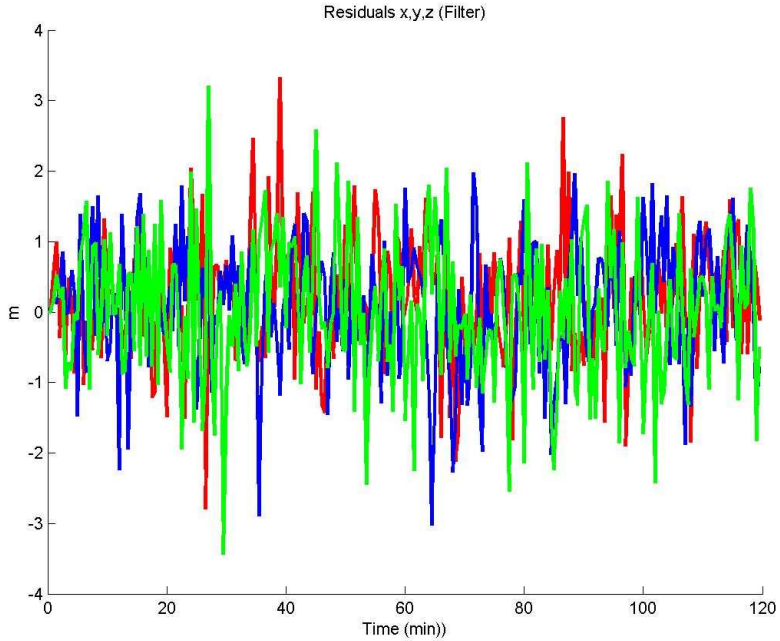




Estimated-true positions (x,y,z). Left: LSQ with 2 hours of data, Right: with filter. The LSQ solution is much more consistent.

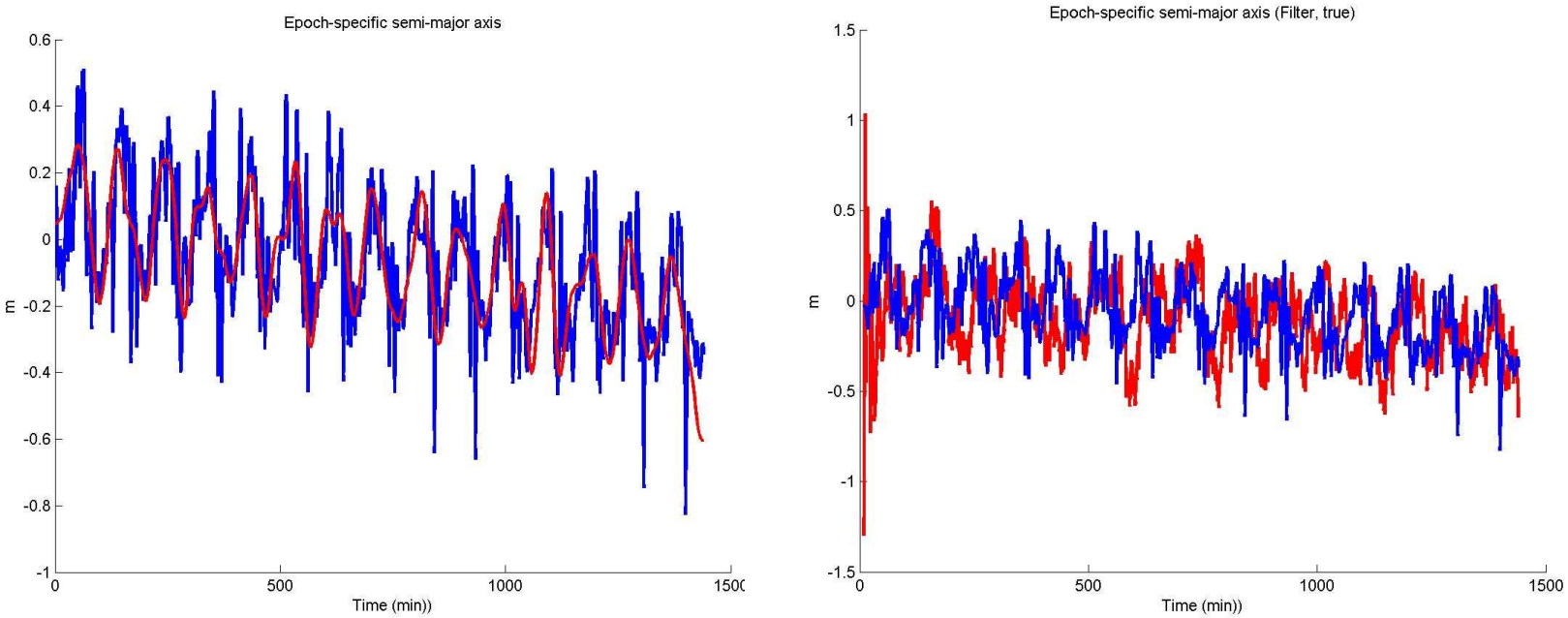
Filtering the stochastic problem





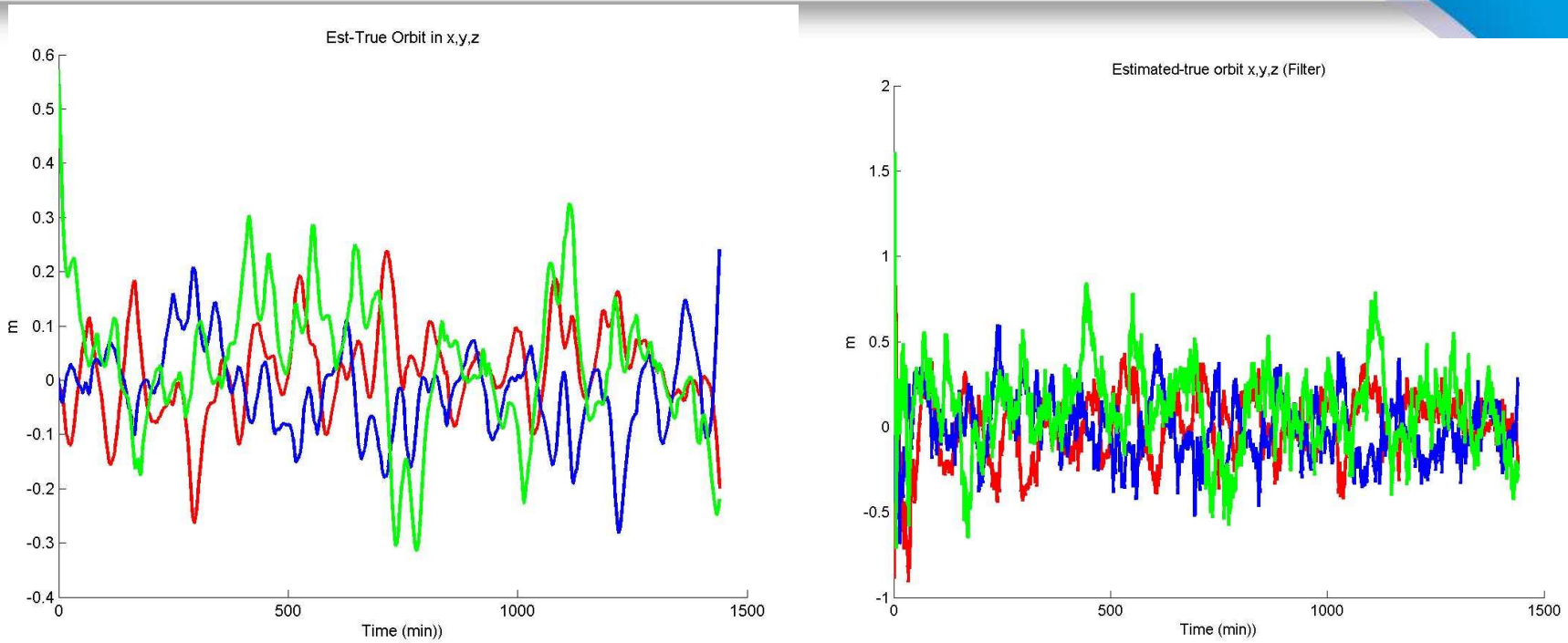
Residuals in x,y,z. Left: LSQ, Right: Filter. The residuals are of the same quality and therefore not suited to judge the quality of the solution.

Filtering the stochastic problem



Semi-major axis estimated by LSQ (left) and filter (right). The quality differences are obvious!

Filtering the stochastic problem



„Estimated-true“ positions (x,y,z). Left: LSQ, Right: Filter. The LSQ solution is better by about a factor of 2 even after the initial oscillation.

Note that the filter solutions are not as smooth as the LSQ solutions.

Filtering the stochastic problem

At the very last data point the LSQ and the filter solutions are identical, but the pseudo-stochastic parameters are not available when filtering.

A back-substitution step cures this problem when filtering. Back-substitution of course is not possible in real-time applications.

For more information we refer to *Beutler, Jäggi et al. „Efficient satellite orbit modelling using pseudo-stochastic parameters“, Journal of Geodesy, (2006) 80: pp 353-372.*

Filtering the stochastic problem

Summary:

In this chapter we dealt with parameter estimation problems, where the parameter array is in itself a stochastic process, defined in essence by a stochastic DEQs.

In order to solve the problem the epoch-specific parameters have to be constrained in a meaningful way.

We treated the epoch-specific parameters as random variables with expectation value “0” and weight $P_v = (\sigma_0 / \sigma_v)^2$. For P_v=0 the filter solution follows the observations very well, for large values of P_v we obtain a deterministic solution.

The concept introduced here is that of the Kalman filter (which would introduce stochastic information in all components).