# On the role and evaluation of rigorous and approximate estimation methods

Wolfgang Förstner

Match 12, 2024

## Contents

# 1 Introduction

The paper addresses the role and evaluation of rigorous and approximate estimation methods.

The result of any method may be characterized by the following input-output relation:

$$\boldsymbol{l} \quad \overset{\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{0},\, \sigma_0^2\Sigma_{ll}^a}{\longmapsto} \quad \{\widehat{\sigma}_0^2, \widehat{\boldsymbol{x}}, \Sigma^a_{\widehat{x}\widehat{x}}, \widehat{\boldsymbol{y}}\} \quad \text{with} \quad \boldsymbol{y} = \mathbb{E}(\underline{\boldsymbol{l}}) \tag{1}$$

- The vector $\boldsymbol{l}$ collects the $N$ observations.
- The method internally has a mathematical model, which is composed of two parts.

  1. The stochastical model: the method assumes, the observations are a sample of a distribution $\mathcal{M}(\mathbb{E}(\underline{\boldsymbol{l}}), \mathbb{D}(\underline{\boldsymbol{l}}))$ with first and second moments

     $$\boldsymbol{y} = \mathbb{E}(\underline{\boldsymbol{l}}) \quad \text{and} \quad \mathbb{D}(\underline{\boldsymbol{l}}) = \sigma_0^2\Sigma_{ll}^a \tag{2}$$

     We sometimes refer to the expected value $\boldsymbol{y}$ of the observations as mean observations. The covariance matrix of the observations often is assumed to be known up to an unknown variance factor $\sigma_0^2$, since the methods often are invariant to a scaling of the covariance matrix $\Sigma_{ll}^a$. If the inital variance factor $\sigma_0 = 1$ is chosen, the matrix $\Sigma_{ll}^a$ can be interpreted as an approximate covariance matrix of the observations.
  2. The functional model: the method assumes the observations are related to the $U$ unknown parameters $\boldsymbol{x}$ by the $K$ generally nonlinear constraints

     $$\boldsymbol{k}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}\,. \tag{3}$$

     Sometimes, a part of the constraints only refers to the parameters. Then we partition the $K$ constraints into $G$ constraints $\boldsymbol{g}$ between the parameters and the observations and $H$ constraints $\boldsymbol{h}$ on the parameters only:

     $$\boldsymbol{k}(\boldsymbol{x}, \boldsymbol{y}) = \begin{bmatrix} \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) \\ \boldsymbol{h}(\boldsymbol{x}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}. \tag{4}$$

- The method, generally, provides three types of output
  - A measure for the consistency of the observations and the parameters with the model $\{\boldsymbol{g}(\boldsymbol{y}, \boldsymbol{x}) = 0, \sigma_0^2\Sigma_{ll}^a\}$. This is an estimate $\widehat{\sigma}_0^2$ for the variance factor $\sigma_0^2$, which it the model holds is close to 1. Of course, this factor is not available for minimal solutions, where the number of constraints and the number of parameters is the same. Not all methods provide this information, even not by ML-estimation procedures, though they easily could be made available.

– All methods yield the estimated parameters $\widehat{\boldsymbol{x}}$ as output.
– The covariance matrix $\Sigma_{\widehat{x}\widehat{x}}$ of the parameters. It may be derived also for a large class of approximate values.
– Sometimes the method also provides estimates $\widehat{\boldsymbol{y}}$ for the mean observations $\boldsymbol{y}$. Most often they are given implicitly by the estimated residuals

$$\widehat{\boldsymbol{v}} = \widehat{\boldsymbol{y}} - \boldsymbol{l} \tag{5}$$

useful for evaluating individual observations or groups of observations.

In all cases, these the variance factor and the covariance matrix of the parameters can be derived from repeatedly applying the estimation method to samples of the observations.

*Remark:* Observe, we did not refer to the classical Gauss–Markov model $\mathbb{E}(\boldsymbol{l}) = \boldsymbol{f}(\boldsymbol{x})$, since this is not general enough for many tasks in geometric computation. But it may easily be rewritten as $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) = \mathbb{E}(\boldsymbol{l}) - \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$. ◇

Hence, we focus on estimation methods, possibly with constraints, but not with inequality constraints.

The goal of this note is to discuss the mutual evaluation of different methods solving the same problem w.r.t. the accuracy of the derived parameters.

## 1.1 Maximum likelihood estimation for Gaussians

Maximum likelihood estimation takes that parameter vector $\widehat{\boldsymbol{x}}$ as estimate for which the likelihood

$$L(\boldsymbol{x}) = p(\boldsymbol{x}|\boldsymbol{l}) \tag{6}$$

is maximum, where the density function $p$ takes into account the constraints between the uncertain observations and the parameters. For normally distributed observations this can be formalized as follows. In order to simplify the notation, and avoid statistical terms within the optimization procedure as far as possible, we will also write the optimization problem as follows:[1], see (Förstner, 2024, p. 6 ff.) For simplicity, we assume the model is linear in the observations and the unknown parameters. Then, for given observations $\boldsymbol{l}$, regular covariance matrix $\Sigma_{ll}$, and coefficient matrices $A$ and $B$

$$\boxed{\begin{array}{lll} \text{GHM}(\Sigma): & \text{minimize} & (\boldsymbol{y}-\boldsymbol{l})^\mathsf{T}\Sigma_{ll}^{-1}(\boldsymbol{y}-\boldsymbol{l}) \\ & \text{subject to} & \boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) = A\boldsymbol{x} + B^\mathsf{T}\boldsymbol{y} + \boldsymbol{b} = \boldsymbol{0}\,. \end{array}} \tag{7}$$

w.r.t. the unknown parameters $\boldsymbol{x}$ and the mean observations $\boldsymbol{y}$. For a compact representation of the solution we use the substituted observations with their – now different – covariance matrix

$$\boldsymbol{n} = B^\mathsf{T}\boldsymbol{l} + \boldsymbol{b} \quad \text{and} \quad \mathbb{D}(\underline{\boldsymbol{n}}) = \Sigma_{nn} = B^\mathsf{T}\Sigma_{ll}B\,. \tag{8}$$

[1] This in the flavour of the problems discussed in Boyd and Vandenberghe (2004).

We obtain the estimated parameters and the fitted observations from

$$\boxed{\begin{array}{l}\widehat{\boldsymbol{x}} = \; -(A^{\mathsf{T}}\Sigma_{nn}^{-1}A)^{-1}\, A^{\mathsf{T}}\Sigma_{nn}^{-1}\, \boldsymbol{n}(\boldsymbol{l}) \\ \widehat{\boldsymbol{y}} = \boldsymbol{l} - \Sigma_{ll}B\Sigma_{nn}^{-1}\, \boldsymbol{g}(\widehat{\boldsymbol{x}}, \boldsymbol{l})\,.\end{array}} \tag{9}$$

*Cramer-Rao bound, covariance matrix of estimated parameters*

The theoretical covariance matrix of the estimated parameters can be derived from (9) by variance propagation, leading to the Cramer-Rao bound for the uncertainty

$$\boxed{\Sigma_{\widehat{x}\widehat{x}} = (A^{\mathsf{T}}(B^{\mathsf{T}}\Sigma_{ll}B)^{-1}A)^{-1}\,.} \tag{10}$$

This expression can generally provide the covariance matrix for a specific case. However, it also can be used to derive the uncertainty of the parameters as a function of the experimental design, in the most simple case as an algebraic function of number and the standard deviation of the observations, as for the mean of a set of observations (see Sections 4.1.1 and 4.1.2, but also for the uncertainty of 3D points from triangulation, see (Förstner, 2024, p. 112 ff.) and (Förstner, 2024, p. 122 ff.).

If we only know an approximate covariance matrix $\Sigma_{ll}^a$ and we assume the covariance matrix $\Sigma_{ll}$ differs from the approximation by an unknown variance factor $\sigma_0^2$

$$\Sigma_{ll} = \sigma_0^2\Sigma_{ll}^a\,, \tag{11}$$

then we can perform the estimation with $\Sigma_{ll}^a$, instead of using $\Sigma_{ll}$, which has no effect onto the estimates. But then, with $K = G + U$, we can find an estimate

$$\boxed{\widehat{\sigma}_0^2 = \frac{(\widehat{\boldsymbol{y}} - \boldsymbol{l})^{\mathsf{T}}\Sigma_{ll}^{-1}(\widehat{\boldsymbol{y}} - \boldsymbol{l})}{K - U}\,.} \tag{12}$$

for the estimated variance factor. Then we obtain an estimate for the covariance matrix of the estimated parameters

$$\widehat{\Sigma}_{\widehat{x}\widehat{x}} = \widehat{\sigma}_0^2\Sigma_{\widehat{x}\widehat{x}}^a \quad \text{with} \quad \Sigma_{\widehat{x}\widehat{x}}^a = (A^{\mathsf{T}}(B^{\mathsf{T}}\Sigma_{ll}^aB)^{-1}A)^{-1}\,. \tag{13}$$

the attribute *estimated* only referring to use of the estimated variance factor.

Fig. 1 visualizes the principle. For a given pair $\boldsymbol{l} = [l_1, l_2]^{\mathsf{T}}$ of observations and the constraint

$$\mathbb{E}(\underline{\boldsymbol{l}}) = \boldsymbol{a}x \quad \text{or} \quad \boldsymbol{g}(x, \boldsymbol{y}) = \boldsymbol{a}x - \boldsymbol{y} = \boldsymbol{0} \tag{14}$$

the task is to find the most probable point $\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{l}}$ on the straint line $\boldsymbol{y} = \boldsymbol{a}\,x$. The result depends on the assumption about the uncertainty of the observations. We consider the following two models:

1. [A] In case the observations are assumed to have the same standard deviation, say $\sigma^2$ and are uncorrelated, thus we use $\Sigma^{a,[A]} = \sigma^2 I_2$ as approximate covariance matrix, the best point is the footpoint on the straight

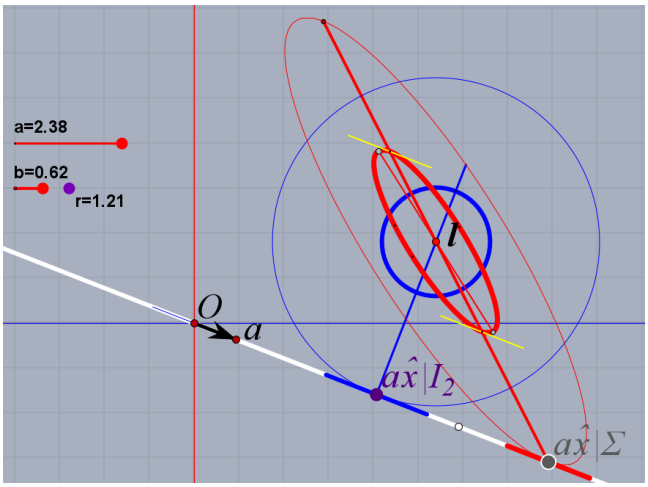


Figure 1: ML estimation for normally distributed observations. Model $\underline{\boldsymbol{l}} \sim \mathcal{N}(\boldsymbol{y}, \Sigma)$ with the constraint $\boldsymbol{a}x - \boldsymbol{y} = \boldsymbol{0}$, represented by the white line. If $\Sigma = \sigma^2 I_2$ ($\sigma = r = 1.26$) , indicated by the blue circular standard ellipse, the optimal point $\widehat{\boldsymbol{y}}$ lies on the footpoint $\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{l}} \mid I_2 = \boldsymbol{a}\widehat{x} \mid I_2$ of $\boldsymbol{l}$ onto the line $\mathbb{E}(\underline{\boldsymbol{l}}) = \boldsymbol{a}x$. If the covariance matrix $\Sigma$ is a general matrix, represented by the red standard ellipse (with semiaxes $a = 2.48$ and $b = 0.78$), then the optimal point $\widehat{\boldsymbol{l}} \mid \Sigma = \boldsymbol{a}\widehat{x} \mid \Sigma$ is the intersection of the (blue) line $\mathbb{E}(\underline{\boldsymbol{l}}) = \boldsymbol{a}x$ passing through $O$ and the (red) line, defined by the direction from $\boldsymbol{l}$ to that point of the ellipse, where the tangent (yellow) is parallel to $\boldsymbol{a}$

line $\boldsymbol{a}\widehat{x} \mid I_2$, independent on the standard deviation – the classical LS solution. It is the most probable point, since all contours of the isotropic Gaussian distribution are concentric circles. The smalles circle reaching the line is the thin blue circle, which touches the line at the best estimate, since all other points on the line would have a smaller likelihood $L_A(x) = p_A(\boldsymbol{a}x \mid \boldsymbol{l})$.

2. [B] In case the observations are assumed to have different accuracy in different directions, measured by the semiaxes $(a, b)$ and and the direction of the largest semiaxes, collected in the covariance matrix thus we use $\Sigma^{a,[B]} = \Sigma$ as approximate covariance matrix, indicated by the red standard ellipse, then we obtain a different estimate $\boldsymbol{a}\widehat{x} \mid \Sigma$. Also here, the density function of the Gaussian has concentric and similar elliptic contours. The smallest ellipse (thin red) touches the straight line at the optimal estimate, again, since all other points on the line have smaller likelihood $L_B(x) = p_B(\boldsymbol{a}x \mid \boldsymbol{l})$.

Besides the observation vector and the assumed covariance matrices, the figure shows to further elements of the output of the estimation method:

- The uncertainty of the estimates, namely the blue and red segment on the straight line. They are the projection of the standard ellipses onto the line in the direction from $\boldsymbol{l}$ to $\widehat{\boldsymbol{l}}$.

- The inconsistency of the assumed model, i.e., the common constraint and the two alternative covariance matrices, with the given observation. The standard ellipses of the assumed covariance matrices are given by

$$(\boldsymbol{l}-\boldsymbol{y})^{\mathsf{T}}\mathsf{W}^a(\boldsymbol{l}-\boldsymbol{y}) = 1 \quad \text{with} \quad \mathsf{W}^a = \sigma^{-2}\mathsf{I}_2 \quad \text{or} \quad \mathsf{W}^a = \boldsymbol{\Sigma}^{-1} \tag{15}$$

  If we would take slightly larger covariance matrices, namely $\sigma_0^{[A]^2}(\sigma^2\mathsf{I}_2)$ or $\sigma_0^{[B]^2}\boldsymbol{\Sigma}^a$, specifically those represented by the thin blue circle or the thin red ellipse, the uncertainty of the observations would be consistent with the fitted observations $\widehat{\boldsymbol{l}} = \widehat{\boldsymbol{y}}$. The factor is nothing else than the estimated variance factor

$$\widehat{\sigma}_0^2 = \frac{(\boldsymbol{l}-\widehat{\boldsymbol{y}})^{\mathsf{T}}\mathsf{W}^a(\boldsymbol{l}-\widehat{\boldsymbol{y}})}{(\boldsymbol{l}-\widehat{\boldsymbol{y}})^{\mathsf{T}}\mathsf{W}(\boldsymbol{l}-\widehat{\boldsymbol{y}})} \quad \text{with} \quad \widehat{\sigma}_0^2 = \frac{(\boldsymbol{l}-\widehat{\boldsymbol{y}})^{\mathsf{T}}\mathsf{W}^a(\boldsymbol{l}-\widehat{\boldsymbol{y}})}{K-U} \tag{16}$$

  where $\boldsymbol{\Sigma} = \mathsf{W}^{-1}$ are represented by the thin blue circle and the thin red ellipse. Of course, we can only use the right equation for calculating the estimated variance factor. However, using the improved covariance matrices

$$\boldsymbol{\Sigma} = \widehat{\sigma}_0^2\,\boldsymbol{\Sigma}^a\,, \tag{17}$$

  we have the guarantee that the estimated point, in this case $\widehat{\boldsymbol{y}} = \widehat{\boldsymbol{l}} = \boldsymbol{a}\widehat{x}$, lies on both the constraint line and the standard ellipse of the improved covariance matrix.

  The method can be generalized to nonlinear constraints and to cases where some constraints only refer to the unknown parameters.

We will use this interpretation of the ML-estimation in the next sections.

## 1.2 Notion of an approximate method

What is an approximate method? A method which does not lead to an optimal result. A method, leading to an optimal result, usually is called rigorous.

However, we often have a hierarchy of models, one being a special case of the previous one. Let us take the four classical estimation principles, which differ in the assumption abput the observational noise $\boldsymbol{e} = \mathbb{E}(\underline{\boldsymbol{l}}) - \boldsymbol{l}$:

1. Maximum-likelihood (ML) estimation: It assumes the distribution of the observational noise $\underline{\boldsymbol{e}}$ is known, but arbitrary.

2. Generalized least squares (GLS) estimation: It can be derived from ML-estimation, assuming the observational noise is normally distributed, thus $\underline{\boldsymbol{e}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{ll})$.

3. Weighted least squares (WLS) estimation: It can be derived from GLS-estimation, assuming the observational noise elements are jointly normally distributed but uncorrelated, thus $\underline{\boldsymbol{e}} \sim \mathcal{N}(\mathbf{0}, \mathrm{Diag}([\sigma_{l_n}^2]))$.

4. Simple least squares (LS) estimation: It can be derived from WLS-estimation, assuming the observational noise elements are jointly normally distributed, uncorrelated, and have the same variance, thus $\underline{\boldsymbol{e}} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \boldsymbol{I}_N)$.

Then the result of all models may be called rigorous, if we assume – not simultaneously – the assumed covariance matrix is a good approximation for the uncertainty of our data. However, models B can also be termed suboptimal or approximate, in case we know that model A, is a good approximation for the data's uncertainty. The same holds for model C w.r.t. model B. Hence, model B may be both, a reference for the estimation with model C, or an approximation w.r.t. model A. This clearly shows, that the notion rigorous is only meaningful in a well-defined context. The same holds when including model D in the discussion.

We will call ML-estimates based on normally distributed observations with an arbitrary covariance matrix as *statistically optimal estimates* if the assumed type of the distribution, including the used constraints and the covariance matrix, in an acceptable manner approximates the real situation. This on one hand – in a first step – identifies all methods, which do not integrate the uncertainty of the observations in some explicit way, as approximate methods. On the other hand, it leaves enough freedom for classifying methods as statistically optimal, though the method does not explicitly refers to some distribution of the observations, but the individual weighting of observations can be motivated by some acceptable assumption about the variance covariance matrix of the observations. As an example, LS estimates can be statistically optimal, if all observations can be assumed to have the same standard deviation and are mutually uncorrelated. In both cases, naturally the decision on the acceptability lies in the hand of the user of the method.

Of course, there are other types of approximations, which cannot be included in some hierarchy, some of which may be identified easily:

- The method optimizes some other optimization function. **Example:** Instead of minimizing the (RMSE) of the reprojection error using some – possibly iterative – LS method, the method minimizes an algebraically motivated error, e.g., only the RMSE of some constraints, as many methods for reconstructing geometric entities.

- The method is iterative, and the iteration is stopped before convergence. **Example:** Only applying one iteration may be motivated by requiring a method has a constant computing time, thus not depending on the quality of some approximate values.

- The method, though optimizing the correct optimization function, does not fulfil all constraints. **Example:** Estimating the essential matrix using

the 8-point algorithms – in a first step – does not lead to a matrix of rank two.

- The method does not use all information provided by the data. **Example:** RANSAC type procedures may provide the solution of the best smallest tuple of observations, neglecting all other inliers.

- An iterative method uses approximations of the required Jacobians. **Example:** (a) Minimizing the Sampson error in an iterative scheme uses the Jacobian evaluated at the observations instead at the fitted observations, (b) Using the Jacobian of the first iteration during the further iterations, both cases meant to save computation time.

- Accepting a non-negligible bias in the estimate.

We now discuss the role of approximate and rigorous methods in the view of users and authors of methods.

## 1.3 The user's perspective on rigorous and approximate Methods

Eventually, the user decides on whether to use a specific method. There are various reasons why to choose or not to choose a specific method. We want to collect some of these reasons.

Reasons for preferring an approximate method may be:

- The quality of the result of an approximate method is acceptable for the application. **Example:** Some approximate methods may be – in a certain scenario – a factor two or three worse (in standard deviation), but the high accuracy of the given data leads to still acceptable results.

- The method allows to predict computation time, which is needed for real time applications (robotics, methods used during user interaction). **Example:** All methods which do not require iterations (direct methods) or are guaranteed to succeed with a constant number of iterations fall into this class.

- No approximate values for the parameters are available. Then direct solutions[2] are appropriate, but most of them are approximate. **Example:** The 8-point algorithm for estimating the relative pose $(R, \boldsymbol{t})$. **Counterexample:** The best fitting line or plane through a set of points if their uncertainty is isotropic.

Reasons for preferring a rigorous method may be:

[2] methods which do not require such approximate values

- They may lead to sufficient estimates, exploiting all information of the given data. **Example:** Classical ML-estimates are sufficient, if the uncertainty of the given data is taken into account.
- They provide adequate information about the quality of the result. This may refer just to the uncertainty of the estimated parameters but also to the diagnostic information, e.g., for outlier detection. **Example:** Classical ML-estimates may provide a covariance matrix for the estimated parameters, the Cramer-Rao bound. ML-type robust estimators may be able to cope with an acceptable large percentage of outliers.
- Usually their optimization principle is easy to explain. **Example:** LS, GLS, ML are well established and well understood optimization criteria.

## 1.4 The author's perspective on rigorous and approximate Methods

Authors of methods either propose a new/refined method to solve a given problem aiming at a scientific publication, and/or aiming at solving a hitherto unsolved/suboptimally solved class of problems encountered in a practical application. Authors may or may not be aware of the user's perspective on a method. I only discuss the situation where an author of a method aims at a scientific publication.

Motivations to publish a rigorous method may be:

- Establish a reference method for the solution of a hitherto unsolved problem. **Example:** Integrating camera/systems calibration into bundle adjustment leading to self-calibrating bundle adjustment, e.g., for a new type of camera/camera system.
- Establishing a computational more efficient method compared with the state of the art, possibly exploiting prior knowledge or reducing the need for prior knowledge. **Example:** Transforming a batch solution to an incremental (Kalman-filter type) solution, transforming/inventing a method which works for 3D, instead of 2D, for continuous flow of data instead of a sequence of data.
- Exploiting the potential of knowledge about the underlying optimization principles. **Example:** ML-estimation allows predicting the performance using the Cramer-Rao bound, and allows statistical testing of hypotheses referring to the observations or parameters.

Motivations to publish an approximate method may be:

- Providing a direct solution of a problem for a smaller/larger number of observation than previous methods: **Examples:** Providing a hitherto not available minimal solution (such as Nister's 5-point algorithm), providing a hitherto not available redundant solution (such as a Khoshelham's motion estimation from multiple plane matches)

- Providing an approximate method with proven and acceptable loss in accuracy w.r.t. a rigorous solution (such as a direct solution followed by a one-iteration ML-solution)

There are also motivations why one should not publish a method:

- The new method does not lead to useful improvements. **Examples:** (1) The decrease in standard deviation is less than a small amount, e.g., less than 20%. The reason may be the following: the decrease in standard deviation is too small to reduce the number of iterations of a subsequent rigorous method. (2) The decrease in computation time is not really more than 50% (wait a year! c.f. Moore's law).

- The preconditions under which the method works are not made explicit. **Example:** The required accuracy of approximate values for a rigorous method is not analysed. – If the method still submitted for being reviewed, the reviewer might reject the publication, due to the lack of clarity on the conditions under which the method works well.

- The method does not provide any information about the quality of the result. **Example:** (empirically nearly) all new direct methods fall into this class, if they do not report on success rates, how the accuracy can be evaluated, or how to detect/identify singularities.

The discussion clearly shows: the author of a (publication of a) method needs to provide a data sheet characterizing the quality of the proposed method, which allows a user/reader to decide, whether the proposed method may be acceptable in his/her application area. This augments the neccessary, but not sufficient experimental benchmarks.

## 1.5 Goals

We address the following questions:

- Can approximate methods be upgraded to provide uncertainty information about their estimates? Yes, this is possible for a large class of methods provided they only contain differentiable steps, see Sect. 2.

- Can direct methods be upgraded to methods which are rigorous but still guarantee fixed computing time (for given number of data)? Yes, this is possible in case a direct solution exists which leads to parameters, which are close enough the optimal ones, such that a single iteration with a ML-estimation leads to acceptable results. **Example:** We will show this for the registration of point clouds using plane correspondences.

- How to characterize the loss in quality of approximate methods compared to rigorous methods? We mainly discuss how to characterize accuracy loss. **Examples:** Registration, and bundle adjustment will demonstrate this.

We need some basic techniques:

- Implicit variance propagation for minimal and non-minimal direct solutions
- Algebraically parametrized variances/covariances for understanding the influence of the experimental design onto the accuracy.
- Different measures for comparing theoretical and empirical covariance matrices.

### 1.6 Previous work

tbd.

## 2 Approximate methods and their accuracy

We assume we have given $N$ observed values $\boldsymbol{l}$, which generally are noisy thus follow

$$\boldsymbol{y} \sim \mathcal{M}(\mathbb{E}(\underline{\boldsymbol{l}}), \Sigma_{ll}) \quad \text{with} \quad \underline{\boldsymbol{l}} + \underline{\boldsymbol{v}} = \mathbb{E}(\underline{\boldsymbol{l}})\,. \tag{18}$$

We discuss how to algebraically and numerically derive covariance matrices for a large class of direct solutions, minimal and non-minimal ones. This allows to replace simulation studies onto the accuracy of these methods by the analysis of the predicted covariance matrices.

### 2.1 Minimal solutions

Minimal solutions for a set of $U$ unknown parameters $\boldsymbol{x}$ are characterized by a combination of $G \leq U$ constraints $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{l}) = \mathbf{0}$ involving the observational values $\boldsymbol{l}$ and, possibly, further $U - G \geq 0$ constraints $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$ only referring to the unknown parameters. Thus, the problem minimal solvers address is to determine the parameters $\boldsymbol{x}$ from given observations $\boldsymbol{l}$ using the constaints:

$$\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) = \mathbf{0} \quad \text{and} \quad \boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}\,. \tag{19}$$

**Example 2.1:** ***Minimal solution for the fundamental matrix.*** The fundamental matrix $\mathsf{F}$ has nine unknown parameters $\boldsymbol{x} := \boldsymbol{f} = \text{vec}\mathsf{F}$. Since the matrix $\mathsf{F}$ is homogeneous and is singular, we only need $G = 7$ correspondences $(\mathbf{x}_i', \mathbf{x}_i''), i = 1, ..., 7$ collected in a 42-vector $\boldsymbol{l}$ fulfilling the epipolar constraint and – in addition – the length constraint and the singularity constraint for the $U = 9$ vector $\boldsymbol{x} = \mathbf{f}$. With the the vector $\mathbf{f}^\circ = \text{vec}\mathsf{F}$ of the cofactor matrix[3] $\mathsf{F}^\circ$ we therefore have the minimal problem

$$\underset{7\times 1}{\boldsymbol{g}} = [\mathbf{x}_i{}'^{\mathsf{T}} \mathsf{F} \mathbf{x}_i''] = [(\mathbf{x}_i{}''^{\mathsf{T}} \otimes \mathbf{x}_i^{\mathsf{T}})\mathbf{f}] = \mathbf{0} \quad \text{and} \quad \underset{2\times 1}{\boldsymbol{h}} = \begin{bmatrix} \mathbf{f}^{\mathsf{T}}\mathbf{f} - 1 \\ \mathbf{f}^{\mathsf{T}}\mathbf{f}^\circ \end{bmatrix} = \mathbf{0} \tag{20}$$

[3] For a $3 \times 3$ matrix $A = [\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3]$ the cofactor matrix is $A^\circ = [\boldsymbol{a}_2 \times \boldsymbol{a}_3, \boldsymbol{a}_3 \times \boldsymbol{a}_1, \boldsymbol{a}_1 \times \boldsymbol{a}_2, ]$. It fulfils $A^\circ = |A| A^{-\mathsf{T}}$

for determining the vector $\boldsymbol{f}$ form the observed coordinates, see (Förstner and Wrobel, 2016, p. 572). ◇

The minimal solution $\boldsymbol{x} = \boldsymbol{x}(\boldsymbol{l})$ usually is based on some algebraic derivation, which might be highly complex. In case all steps in the solution are differentiable, and the number of used constraints is minimal,[4] we may easily derive the covariance matrix of the parameter vector $\boldsymbol{x}$.

Let the $U$ constraints for the $U$ parameters $\boldsymbol{x}$ – possibly out of several solution vectors – be available. Following Barath et al. (2020, eq. (26)), we can derive the covariance matrix of $\underline{\boldsymbol{x}}$ by implicit variance propagation, see Förstner and Wrobel (2016, Sect. 2.7.5). For this, we concatenate the constraints and use their Jacobians

$$\underset{U\times 1}{\boldsymbol{k}(\boldsymbol{x},\boldsymbol{y})} = \begin{bmatrix} \boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) \\ \boldsymbol{h}(\boldsymbol{x}) \end{bmatrix} \quad \text{with} \quad \underset{U\times U}{A} = \left.\frac{\partial \boldsymbol{k}}{\partial \boldsymbol{x}^\mathsf{T}}\right|_{x,l}, \; \underset{U\times N}{B^\mathsf{T}} = \left.\frac{\partial \boldsymbol{k}}{\partial \boldsymbol{l}^\mathsf{T}}\right|_{x,l}, \tag{21}$$

we obtain

$$\Sigma_{xx} = A^{-1} B^\mathsf{T} \Sigma_{ll} B A^{-\mathsf{T}}. \tag{22}$$

if no singularity exists. Observe, if we have multiple, say $T$, solutions $\boldsymbol{x}_t$, we obtain a covariance matrix for each of the $T$ solutions, since the Jacobians $A$ and $B$ need to be evaluated at $\boldsymbol{x} = \boldsymbol{x}_t$.

Summarizing: For minimal solutions, which do only contain differential algebraic steps, we can derive the covariance matrix for each of several solutions, exploiting the Jacobians of the used constraints, without following the algebraic steps of the direct solution. Furthermore, we may numerically identify singularities by analysing the condition number of the Jacobian $A$ or of the covariance matrix $\Sigma_{xx}$.

## 2.2 Models linear in homogeneous parameters

We assume the given $N$ observations are related to the $U$ unknown inhomogeneous parameter vector $\mathbf{x}$ by the $G$ constraints, which are linear in $\mathbf{x}$, and the length constraint

$$\boldsymbol{g}(\mathbf{x},\boldsymbol{y}) = A(\boldsymbol{y})\mathbf{x} = \mathbf{0} \quad \text{and} \quad |\mathbf{x}|^2 = 1. \tag{23}$$

The matrix $G \times U$ matrix $A(\boldsymbol{y})$ generally has rank $U-1$ thus is singular with a rank deficiency of 1.

We are interested in some estimate $\widehat{\mathbf{x}}$ for the parameters. The classical approach exploits the linearity of the constraints w.r.t. $\mathbf{x}$ and minimizes the length of the residual of the constraints when applied to the noisy observations

$$|A(\boldsymbol{l})\,\mathbf{x}|^2 = \mathbf{x}^\mathsf{T} A^\mathsf{T}(\boldsymbol{l})\, A(\boldsymbol{l})\,\mathbf{x} \quad \text{with} \quad |\mathbf{x}|^2 = 1. \tag{24}$$

[4] Some solvers allow redundant constraints, and, depending on some criteria only use a minimal set of constraints. Then, the following derivation applies ot the selected constraints.

w.r.t. the parameters $\mathbf{x}$. This leads to the classical singular value solution

$$\widehat{\mathbf{x}} = \boldsymbol{v}_U \quad \text{with} \quad A = UDV^\mathsf{T}\,,\; V = [\boldsymbol{v}_u]\,, u = 1, ..., U \quad \text{and} \quad d_1 \geq d_2 \geq ... d_U\,. \tag{25}$$

the matrices $U$ and $V$ being orthogonal with $U^\mathsf{T}U = V^\mathsf{T}V = I_{U-1}$.

Accuracy. We interpret this optimization using a quasi-Gauss–Markov model, and start from the linearization at the mean/true values, using $\mathbf{x} = \tilde{\mathbf{x}} + \Delta\mathbf{x}$,

$$\boldsymbol{g} = \boldsymbol{g}(\mathbf{x}, \boldsymbol{l}) = A(\boldsymbol{l})\,\mathbf{x} = \underbrace{A(\boldsymbol{y})\,\tilde{\mathbf{x}}}_{=\mathbf{0}} + A(\boldsymbol{y})\,\Delta\mathbf{x} - \underbrace{B^\mathsf{T}(\tilde{\mathbf{x}}, \boldsymbol{y})\,\underline{\boldsymbol{v}}}_{=\boldsymbol{v}_g} = A(\boldsymbol{y})\,\Delta\mathbf{x} - \boldsymbol{v}_g \tag{26}$$

with the corrections

$$\boldsymbol{v}_g = B^\mathsf{T}(\tilde{\mathbf{x}}, \boldsymbol{y})\boldsymbol{v} \tag{27}$$

of the values $\boldsymbol{g}$ and take into account the linearized length constraint to find an estimate $\widehat{\Delta\mathbf{x}}$ for $\Delta\mathbf{x}$ of the linear quasi Gauss–Markov model with constraint

$$\boldsymbol{g} + \boldsymbol{v}_g = A\Delta\,\mathbf{x} \quad \text{with} \quad \mathbf{x}^\mathsf{T}\Delta\mathbf{x} = 0\,. \tag{28}$$

This model is called a quasi-Gauss–Markov model, since its form is a Gauss–Markov model, but the matrix $A(\boldsymbol{l} + \boldsymbol{v})$ depends on the unknown corrections $\boldsymbol{v}$ and the vector $\mathbf{x}$ in the length constraint is unknown. While we can use the estimate $\widehat{\mathbf{x}}$ instead of $\mathbf{x}$, we hhave no access to the corrections, but in case the noise level is low, the approximation of of $A(\boldsymbol{y})$ by $A(\boldsymbol{l})$ will be acceptable. The solution $\widehat{\Delta\mathbf{x}}$ for the corrections of the parameters result from

$$\begin{bmatrix} A^\mathsf{T}A & \mathbf{x} \\ \mathbf{x}^\mathsf{T} & 0 \end{bmatrix} \begin{bmatrix} \widehat{\Delta\mathbf{x}} \\ 0 \end{bmatrix} = \begin{bmatrix} A^\mathsf{T}\boldsymbol{g} \\ 0 \end{bmatrix}. \tag{29}$$

With the inverse of the $2 \times 2$ block matrix

$$\begin{bmatrix} A^\mathsf{T}A & \mathbf{x} \\ \mathbf{x}^\mathsf{T} & 0 \end{bmatrix} = \begin{bmatrix} (A^\mathsf{T}A)^+ & \mathbf{x} \\ \mathbf{x}^\mathsf{T} & 0 \end{bmatrix} \tag{30}$$

see (Förstner and Wrobel, 2016, eq. (A.131)), we therefore can solve for the unknown parameters,

$$\widehat{\Delta\mathbf{x}} = (A^\mathsf{T}A)^+A^\mathsf{T}\boldsymbol{g} = A^+\boldsymbol{g}\,, \tag{31}$$

since

$$(A^\mathsf{T}A)^+A = (VDU^\mathsf{T}UDV^\mathsf{T})^+VDU^\mathsf{T} = VD^+U\,. \tag{32}$$

Hence the covariance matrix of the parameters results from

$$\Sigma_{\widehat{xx}} = A^+(\boldsymbol{y})\,\Sigma_{gg}\,A^{+\mathsf{T}}(\boldsymbol{y})\,. \tag{33}$$

Since we do not have $A(\boldsymbol{y})$, having rank $U - 1$, we use the rank $(U - 1)$-approximation $A_1$ of $A(\boldsymbol{l})$ from

$$A_1 = UD_1V^\mathsf{T} \quad \text{with} \quad D_1 = \text{Diag}([d_1, ..., d_{U-1}, 0]) \tag{34}$$

which fulfills $A_1(\boldsymbol{l})\widehat{\boldsymbol{x}} = \mathbf{0}$, and use it for determining the pseudo inverse $A^+ := A_1^+$. The approximation will be acceptable if the smallest singular value $d_u$ is small, indicating the noise of the given data $\boldsymbol{l}$ is small. Due to (27) we have

$$\Sigma_{gg} = B^\mathsf{T}(\tilde{\mathbf{x}}, \boldsymbol{y})\, \Sigma_{ll}\, B(\tilde{\mathbf{x}}, \boldsymbol{y}) \tag{35}$$

This finally leads to

$$\Sigma_{\widehat{xx}} = A^+ B^\mathsf{T} \Sigma_{ll} B A^{+\mathsf{T}}\,. \tag{36}$$

*Remark:* In case we have additional constraints for the parameters, say $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$ with its linearized form $H^\mathsf{T}\Delta\boldsymbol{x} = \mathbf{0}$, we can enforce these constraints onto the covariance matrix $\Sigma^{(1)}_{\widehat{xx}}$ from (36) by:

$$\Sigma^{(h)}_{\widehat{xx}} = J(\boldsymbol{h})\ \Sigma^{(1)}_{\widehat{xx}}\ J^\mathsf{T}(\boldsymbol{h}) \quad \text{with} \quad J(\boldsymbol{h}) = I_U - H(H^\mathsf{T}H)^{-1}H^\mathsf{T}\,. \tag{37}$$

see Förstner and Wrobel (2016, eq. (4.224)). ◇

Summarizing, we observe: In spite of having a singular value based approximate solution, we are able to predict the covariance matrix of the estimated parameters in a simple manner. However, the solution contains an approximation, since the matrix $A$ is evaluated at the given noisy observations instead of at the estimates for the expectation of the observations.

# 3 Evaluating covariance matrices

When evaluating a covariance matrix $\Sigma$ by comparing it to some other covariance matrix, say $C$, we are confronted with at least the following questions:

1. Checking correctness of the covariance matrix of a given method.

   The classical situation is to check, whether the covariance matrix $\Sigma_{\widehat{xx}}$ provided by a method actually is a useful prediction of the uncertainty of the estimated parameters. Such a check should be always be applied for a new method, in order to prove that the user of the method can take the reported covariance matrix as a reliable predictor for the accuracy of the method.

2. Checking superiority or acceptability of a new method compared to a reference method or a reference covariance matrix w.r.t. accuracy of the derived parameters.

   Two classical situations require such a check:

   (a) Given a new method, we want to know whether it is better than a reference method. Such a reference method either may be a state of the art method or – within the progress of a research activity – an older version of the solution of a given task.

   (b) Given a method, we want to know whether it is acceptable for some application, which is specified by a lower bound for its accuracy.

In both cases, the idea is to check, whether functions of the parameters have a smaller variance when using the covariance matrix $\Sigma$ of a new method $C$ than when using the covariance matrix $C$ of some reference method. The evaluation may be based on the mean or the maximum loss in accuracy.

3. Measuring the difference or the similarity of a method and a reference method w.r.t. the accuracy of the derived parameters.

   The classical situation is to check, whether a new method leading to estimates with covariance matrix $\Sigma$ leads to results of similar accuracy, when compared to a given method leading to estimates with covariance matrix $C$.

We only discuss measures for checking the superiority or the acceptability, case 2.

## 3.1 On ambiguities when comparing motions

Motions mostly are represented as a homogeneous motion matrix

$$\mathsf{M}(R, \boldsymbol{T}) = \begin{bmatrix} R & \boldsymbol{T} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} \tag{38}$$

Based on this convention, determining the difference of two motions, say $\mathsf{M}_1$ and $\mathsf{M}_2$ still is ambiguous, since we need to collect the small differences in the rotation and the translation, each with a 3-vector. This can be done in at least two ways, both are quite common:

1. We may take the quotient of the two matrices

$$\mathsf{M}_2\mathsf{M}_1^{-1} \approx I + \Delta Q \quad \text{with} \quad \Delta\mathsf{Q}(\boldsymbol{s}) = \begin{bmatrix} 0 & -s_3 & s_2 & s_4 \\ s_3 & 0 & -s_1 & s_5 \\ -s_2 & s_3 & 0 & s_6 \end{bmatrix} \tag{39}$$

   which should be close to a unit matrix. The 6-vector contains the joint difference of the rotation and the translation difference

$$\boldsymbol{s} = \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{t} \end{bmatrix} \tag{40}$$

2. Alternatively, we may apply the division scheme only to the rotation,

$$\mathsf{R}_2\mathsf{R}_1^{-1} \approx I + \Delta R \quad \text{with} \quad \Delta R(\boldsymbol{\tau}) = \begin{bmatrix} 0 & -\rho_3 & \rho_2 \\ \rho_3 & 0 & -\rho_1 \\ -\rho_2 & \rho_3 & 0 \end{bmatrix} \tag{41}$$

   and apply the classical difference scheme to the translation

$$\boldsymbol{T}_2 - \boldsymbol{T}_1 = \begin{bmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \end{bmatrix} \tag{42}$$

and collect the two 3-vectors in the 6-vector

$$\zeta = \begin{bmatrix} \boldsymbol{\rho} \\ \boldsymbol{\tau} \end{bmatrix} \tag{43}$$

As can be shown, the two vectors differ in the translation part, thus

$$\boldsymbol{r} = \boldsymbol{\rho} \quad \text{and} \quad \boldsymbol{\tau} \neq \boldsymbol{t}\,. \tag{44}$$

see **?**, Mangelson et al. (2019), and (Förstner, 2024, p. 49 ff.). As a consequence, empirically derived covariance matrices may differ, due to the scheme of determining the 6-vectors characterizing motion differences.

In the following, we assume all motion parameters are evaluated using the same scheme, and name the 6-vector characterizing motion differences $\boldsymbol{s}$.

## 3.2 Maximum and mean loss in accuracy

When comparing two covariance matrices we need to take into account that in many practical cases the units of the parameters involved are not homogeneous, such as when evaluating the quality of a motion, where some parameters refer to the rotations and others refer to the translation, thus the parameter vector contains elements with unit radians and units m. Especially for motions, we need to take care on the way, translations are compared. In addition, we need to provide measures which are invariant to the chosen coordinate system.

If a covariance matrices $\Sigma$ and a reference covariance matrix $C$ differ to some degree the quotient

$$\Sigma C^{-1} \approx I + \Delta\Sigma \tag{45}$$

should be close to a unit matrix, or the eigenvalues of that quotient should be close to 1

$$\lambda_u(\Sigma C^{-1}) \approx 1, \text{ for } u = 1, ..., U\,. \tag{46}$$

If the covariance matrices would be diagonal, the quotient would indicate the ratios of the corresponding standard deviations

$$r_u := \sqrt{(\Sigma C^{-1})_{uu}} = \frac{\sigma^{\Sigma}_{x_u}}{\sigma^{C}_{x_u}}\,. \tag{47}$$

This gives an intuition into the structure of the quotient of the two covariance matrices.

There are several ways to come to an intuitive description of the accuracy loss, in all cases discussing a maximum and a mean loss in accuracy.

- Using the ratios certainly is the first choice, since their interpretation is easy to explain. We might report the vector

$$\boldsymbol{r} = [r_1, ..., r_u, ..., r_U]^{\mathsf{T}} \tag{48}$$

or the mean ratio, using the quadratic mean

$$\overline{r} = \sqrt{1/U \sum_u r_u^2} = \frac{1}{\sqrt{U}} |\boldsymbol{r}|_2 \tag{49}$$

or the maximum ratio

$$r_{\max} = \max_u (r_u) \,. \tag{50}$$

All measures are interpreted as ratios of standard deviations.

- Since the correlations of the parameters are not taken into account we might similarly us the eigenvalues instead of the ratios of the variances. Let the eigenvalues be

$$\lambda_u = \mu_u^2 \tag{51}$$

such that the $\mu_u$ again can be interpreted as ratios of standard deviations We might report the vector

$$\boldsymbol{\mu} = [\mu_1, ..., \mu_u, ..., \mu_U]^\mathsf{T} \tag{52}$$

or the mean loss, using the quadratic mean

$$\overline{\mu} = \sqrt{1/U \sum_u \mu_u^2} = \frac{1}{\sqrt{U}} |\boldsymbol{\mu}|_2 \tag{53}$$

or the maximum loss

$$\max_\mu = \max_u (\mu_u) \,. \tag{54}$$

Also here, all measures are interpreted as ratios of standard deviations, namely in the direction of the eigenvectors.

The covariance matrix $\Sigma$ to be evaluated, may either be the Cramer-Rao bound, i.e., the uncertainty predicted by the method, or it may be derived form a sample of estimates, say $\widehat{\boldsymbol{x}}_j, j = 1, ..., J$, derived from a sample of the observations, say $\boldsymbol{l}_i, j = 1, ..., J$. The we obtain the the empirical covariance matrix from

$$\widehat{\Sigma}_{\widehat{x}\widehat{x}} = \frac{1}{J-1} \sum_j (\widehat{\boldsymbol{x}}_j - \widehat{\boldsymbol{m}})^\mathsf{T} (\widehat{\boldsymbol{x}}_j - \widehat{\boldsymbol{m}}) \quad \text{with} \quad \widehat{\boldsymbol{m}} = \frac{1}{J} \sum_j \widehat{\boldsymbol{x}}_j \tag{55}$$

If the number of parameters is large, we might not want to do sampling. Alternatively, to determining the quotient $\Sigma C^{-1}$ we may directly integrate this division into a scalar measure which assumes we have some reference parameters, say from a ML-estimate, via

$$F = \frac{1}{U} (\widehat{\boldsymbol{x}} - \tilde{\boldsymbol{x}})^\mathsf{T} C^{-1} (\widehat{\boldsymbol{x}} - \tilde{\boldsymbol{x}}) \mid H_0 \sim F(U, \infty) \tag{56}$$

which in case the zero-hypothesis holds, i.e., the method is optimal, the differences should be 0, is $F$-distributed with $U$ and $\infty$ degrees of freedom. In case

the method to be evaluated is approximate, solution will be biased, and the Fisher-test statistic will be larger than 1. Therefore, it might be meaningful to report the loss in accuracy by the quadratic difference of $F$ and 1:

$$\Delta F = \sqrt{F-1} = \frac{\sigma_b}{\sigma_x} \tag{57}$$

which tells how much the bias, say with standard deviation $\sigma_b$ increases the standard deviation $\sigma_x$ of the estimated parameters on an average.

### 3.3 Effect of model errors

We discuss the effect of several types of errors in the method onto the result.

Since our topic is *approximate methods* we focus on typical simplifications made for generating approximate methods. There essentially are three types of simplifications:

1. Using a simplified covaraince matrix, e.g., (1) by neglecting mutual correlations or (2) by replacing the inhomogeenous uncertainty by a more homogeneous uncertainty,

$$\Sigma = [\sigma_{mn}] \mapsto \text{Diag}([\sigma_n^2]) \quad \text{or} \quad \text{Diag}([\sigma_n^2]) \mapsto \sigma^2 I_N \,. \tag{58}$$

   Neglecting correlations usually is motivated by computational efficiency, since the inversion of non-diagonal matrices generally leads to full covariance matrices, or requires the solution of equation systems with non-diagonal entries. Sometimes the correlations of observations are not knwon. Replacing inhomogeneous uncertainty by a more homogeneous also may be motivated by the lack of knowledge.

2. Replacing the Jacobians within an iteration process by matrices, which are more simple to calculate, e.g., (1) evaluating the Jacobains not at the fitted observations $\boldsymbol{y}$, which might not be available for the method, but at the given, noisy observation $\boldsymbol{l}$, or – within an iteration process – (2) taken the Jacobian, of the first iteration also in the following iterations:

$$A(\boldsymbol{l}) \mapsto A(\boldsymbol{y}) \quad \text{or} \quad A^{(\nu)} \mapsto A^{(1)} \tag{59}$$

   Evaluating Jacobians at the observations is the classical idea of Sampson's proposal. Fixing the Jacobain corresponds to classical gradient descent methods, in contrast to Gauss–Newton methods.

3. Stopping the iteration sequence before convergence.

We give some insight what effect to expect from the first type of approximation, and refer to (Förstner, 2024, p. 44 ff.).

If the estimation is performed in a Gauss-Markov model $\mathbb{E}(\underline{\boldsymbol{l}}) = A\boldsymbol{x}$ with $\Sigma_{ll} = \Sigma$ but the true covariance matrix of the observations is $\widetilde{\Sigma_{ll}} = \widetilde{\Sigma}$, then the covariance matrix of the estimated parameters is

$$\Sigma_{\widehat{xx}} = (A^\mathsf{T}\Sigma^{-1}A)^{-1}A^\mathsf{T}\Sigma^{-1}\,\widetilde{\Sigma}\,\Sigma^{-1}A(A^\mathsf{T}\Sigma^{-1}A)^{-1}\,, \tag{60}$$

which follows from $\widehat{\underline{\boldsymbol{x}}} = (A^{\mathsf{T}}\Sigma^{-1}A)^{-1}A^{\mathsf{T}}\Sigma^{-1}(\underline{\boldsymbol{l}} - \boldsymbol{a})$. Observe, only if $\Sigma = \widetilde{\Sigma}$ do we obtain the classical result

$$\tilde{\Sigma}_{\widehat{xx}} = (A^{\mathsf{T}}\widetilde{\Sigma}^{-1}A)^{-1}\,. \tag{61}$$

The relation between both covariance matrices can be derived from the eigenvalues of the quotient

$$\lambda(\Sigma_{\widehat{xx}}\tilde{\Sigma}^{-1}_{\widehat{xx}}) = \lambda\left((A^{\mathsf{T}}\Sigma^{-1}A)^{-1}A^{\mathsf{T}}\Sigma^{-1}\;\widetilde{\Sigma}\;\Sigma^{-1}A(A^{\mathsf{T}}\Sigma^{-1}A)^{-1}\;A^{\mathsf{T}}\widetilde{\Sigma}^{-1}A\right)\,. \tag{62}$$

Equations (60) and (62) can be used to investigate the effect of choosing a simplified stochastical model, e.g., when using $\Sigma_{ll} = \sigma^2 I_N$ instead of $\widetilde{\Sigma}$.

For $\Sigma = \sigma^2 I$ we would obtain

$$\lambda(\Sigma_{\widehat{xx}}\tilde{\Sigma}^{-1}_{\widehat{xx}}) = \lambda\left((A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}\;\widetilde{\Sigma}\;A(A^{\mathsf{T}}A)^{-1}\;A^{\mathsf{T}}\widetilde{\Sigma}^{-1}A\right)\,, \tag{63}$$

obviously, independent on the scaling of the covariance matrices.

With the hat matrix

$$H = A(A^{\mathsf{T}}A)^{-1}\;A^{\mathsf{T}} \tag{64}$$

this is equivalent to analysing

$$\lambda(\Sigma_{\widehat{xx}}\tilde{\Sigma}^{-1}_{\widehat{xx}}) = \lambda(H\widetilde{\Sigma}H\widetilde{\Sigma}^{-1}) \geq 1\,, \tag{65}$$

which is a unitless quantity. Due to the Gauss–Markov theorem his quantity always is not smaller than 1, i.e., – as to be expected – the approximate solution generally is less accurate than the optimal.

# 4 Examples

## 4.1 Example: mean of observations

### 4.1.1 1D: Neglecting common correlations

The first example is meant to show, that changes in the covariance matrix my not influence the estimates at all.

We assume the $N$ observations $\boldsymbol{l} = [l_n]$ have joint covariance matrix

$$\mathbb{D}(\underline{\boldsymbol{l}}): \quad \sigma_d^2 I_N + b\mathbf{1}\mathbf{1}^{\mathsf{T}} \mapsto \sigma_d^2 I_N\,. \tag{66}$$

or in detail.

$$\Sigma_{ll} = (\sigma_d^2 + b)\begin{bmatrix} 1 & \dots & \rho & \dots & \rho \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \rho & \dots & 1 & \dots & \rho \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \rho & \dots & \rho & \dots & 1 \end{bmatrix} \quad \text{and} \quad \rho = \frac{b}{\sigma_d^2 + b} \tag{67}$$

Then, neglecting the correlations has no effect onto the estimate $\widehat{\mu}$, but only onto its variance $\sigma^2_{\widehat{\mu}}$:

$$\sigma^2_{\widehat{mu}} = \frac{\sigma_d^2}{N} + b \tag{68}$$

see (Förstner, 2024, p. 37 ff.).

Tis is a special case a Lemma by Rao (1967): If the covariance matrix of the observations is changed in a specific manner, the resultant estimated parameters are the same, but with a different covariance matrix, see (Förstner, 2024, p. 24) and Fig. 2.

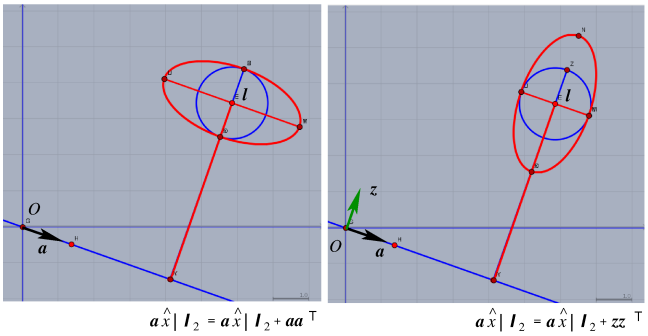


Figure 2: Visualization of Rao's lemma: Least squares estimation with modified covariance matrix. Model $\underline{\boldsymbol{l}} \sim \mathcal{N}(\boldsymbol{a}x, \Sigma)$. If $\Sigma = \sigma^2 \boldsymbol{I} + \gamma \boldsymbol{a}\boldsymbol{a}^\mathsf{T}$ or if $\Sigma = \sigma^2 \boldsymbol{I} + \theta \boldsymbol{z}\boldsymbol{z}^\mathsf{T}$, with $\boldsymbol{z} \perp \boldsymbol{a}$, hence generally, if $\Sigma = \sigma^2 \boldsymbol{I} + \gamma \boldsymbol{a}\boldsymbol{a}^\mathsf{T} + \theta \boldsymbol{z}\boldsymbol{z}^\mathsf{T}$, the semi-axess of the standard ellipse are parallel or orthogonal to $\boldsymbol{a}$. Then, the least squares estimate for the generalized covariance matrix is the same as for $\Sigma = \boldsymbol{I}_2$. However, the covariance matrix of the estimate depends on the modification, namely the factors $\sigma^2$ and $\gamma$.

### 4.1.2 1D: Simplification of standard deviations

This example is meant to show the effect onto the weighted mean when replacing varying weights by constant weights. In all cases the accuracy deteriorates, essentially depending on the scatter of the weights, see (Förstner, 2024, p. 43 ff.).

Estimating the weighted mean of $N$ observations may be simplified by approximating it by the arithmetic mean.

$$\mathbb{D}(\underline{\boldsymbol{l}}): \quad \widetilde{\Sigma} = \mathrm{Diag}([\widetilde{\sigma}_n^2]) \mapsto \boldsymbol{I}_N \,. \tag{69}$$

Effect onto estimate and onto variance by a factor

$$\lambda = \frac{\sigma^2_{\widehat{\mu}} \mid \widetilde{\Sigma}}{\sigma^2_{\widehat{\mu}} \mid \boldsymbol{I}} = \frac{\mu^{(a)}_{\tilde{\sigma}^2}}{\mu^{(h)}_{\tilde{\sigma}^2}} \geq 1 \tag{70}$$

being the ratio of the arithmetic mean and the harmonic mean of the variances, always being not smaller than 1. E.g. ist the weights are taken from a Gamma-distribution with mean weight $\mu_w$ and standard deviation

$$\sigma_w = c\mu_w\,, \quad c \leq 1 \tag{71}$$

of the weights, then the factor is

$$\lambda = \frac{\sigma^2_{\widehat{\mu}} \mid \widetilde{\Sigma}}{\sigma^2_{\widehat{\mu}} \mid I} = \frac{1}{1-c^2} \geq 1\,. \tag{72}$$

### 4.1.3 1D: Free choice of covariance matrix for finding estimate $x$

This example is meant to show, that for any estimated mean $\widehat{\mu}$ there exists a covariance matrix for two observations leading to this mean as optimal estimate. This is an indication, that postulating a method is optimum is meaningless without further specification.

Given are two observations. If we assume they are normally distributed, with covariance matrix

$$\mathbb{D}(\underline{\boldsymbol{l}}) = \sigma^2 \begin{bmatrix} 1 & \rho k \\ \rho k & k^2 \end{bmatrix} \tag{73}$$

we may specify the estimated mean to be $\widehat{x} = x$ and the variance of the estimated mean to be $\sigma^2_{\widehat{x}} = v$ and derive algebraic expressions for $\rho$ and $k$. In the special case $\boldsymbol{l} = [1, 0]^{\mathsf{T}}$ and $x = -1$ we require, that the mean of the two observations lies outside the interval between the two observations. We find

$$\sigma^2 = \frac{3}{4k^2-1} v \quad \text{and} \quad \rho = \frac{2k^2+1}{3k} \quad \text{for} \quad k \in (0.5, 1)\,. \tag{74}$$

As an example, for

$$\boldsymbol{l} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \Sigma_{ll} = \sigma^2 \begin{bmatrix} 27 & 17 \\ 17 & 12 \end{bmatrix} \quad \frac{\sigma_{l_1}}{\sigma_{l_1}} = \frac{2}{3} \quad \rho = \frac{17}{18} \approx 0.9444 \tag{75}$$

we obtain $\widehat{x} = -1$.

This confirms the — updated – intuition, that for any observation vector $\boldsymbol{l}$, any design matrix $\mathsf{A}$, and specified estimate $\boldsymbol{x}$ there exists (at least one) covariance matrix $\Sigma_{ll}$, such that the solution to the Gauss-Markov model $\boldsymbol{l}+\boldsymbol{v} = \mathsf{A}\boldsymbol{x}$ is $\widehat{\boldsymbol{x}} = \boldsymbol{x}$.

### 4.1.4 2D: Simplification of covariance matrices of points

The previous example can be easily generalized and visualized in 2D: In case we determine the statistically optimal mean of two 2D points with arbitrary covariance matrix, the mean point generally does not lie on the straight segment between the two points.

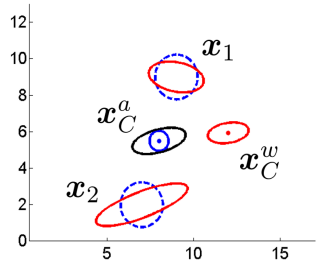


Figure 3: Simple mean $\boldsymbol{x}_C^a$ and weighted mean/centroid $\boldsymbol{x}_C^w$ of two uncertain points $\boldsymbol{x}_1$ und $\boldsymbol{x}_2$ with highly anisotropic uncertainty (red standard ellipses). The weighted centroid obviously lies not on the straight segment between the two points

The Fig. 3 shows the simple and the weighted mean of two 2D points. Their uncertainty is anisotropic, indicated by the red standard ellipses. The mean values result from the two Gauß-Markov models

$$\begin{bmatrix} \underline{\boldsymbol{x}}_1 \\ \underline{\boldsymbol{x}}_1 \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} I_2 \\ I_2 \end{bmatrix} \boldsymbol{x}_C^a, \sigma^2 I_4 \right) \tag{76}$$

and

$$\begin{bmatrix} \underline{\boldsymbol{x}}_1 \\ \underline{\boldsymbol{x}}_1 \end{bmatrix} \sim \mathcal{N}\left( \begin{bmatrix} I_2 \\ I_2 \end{bmatrix} \boldsymbol{x}_C^w, \mathrm{Diag}(\{\Sigma_{11}, \Sigma_{22}\}) \right) . \tag{77}$$

The variance $\sigma^2$ in model (76) was chosen to be the mean of the variances in model (77), see the blue circles.

Explicitly the centroids are

$$\widehat{\boldsymbol{x}}_C^a = \frac{1}{2}(\boldsymbol{x}_1 + \boldsymbol{x}_2) \quad \text{and} \quad \widehat{\boldsymbol{x}}_C^w = (\Sigma_{11}^{-1} + \Sigma_{22}^{-1})^{-1}(\Sigma_{11}^{-1}\boldsymbol{x}_1 + \Sigma_{22}^{-1}\boldsymbol{x}_2) . \tag{78}$$

The simple mean is the midpoint of the two given points, as expected in contrast to the weighted mean. The anisotropic uncertainty of the two points allows the mean to be shifted slightly in the direction of the two major semiaxes.

## 4.2 Motion from point and plane correspondences

Registration of point clouds is a standard problem in point cloud processing. It may be based on corresponding points, the classical starting point, on approximately corresponding point, as within the iterative closest point procedure, or point to plane correspondences in order to avoid the effect of non-existing point correspondences, or on corresponding plane segments, especially useful in man-made environments.

### 4.2.1 Registration from point correspondences

The classical method for registration of corresponding points goes back to Arun et al. (1987) and starts from the functional model

$$\mathbb{E}(\mathbf{X}'_i) = \mathsf{M}\mathbf{X}_i \quad \text{with} \quad \mathbf{X}_i = \begin{bmatrix} X_i \\ 1 \end{bmatrix}, \ \mathbf{X}'_i = \begin{bmatrix} X'_i \\ 1 \end{bmatrix}, \quad \text{and} \quad \mathsf{M} = \begin{bmatrix} R & \boldsymbol{T} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} \tag{79}$$

with the corresponding homogeneous coordinates $(\mathbf{X}_i, \mathbf{X}'_i), i = 1, ..., I$ and the rigid motion matrix $\mathsf{M}$, depending on the 3D rotation matrix $R$ and the 3D translation vector $\boldsymbol{T}$. The method assumes the points to have the same isotropic uncertainty, i.e., the covariance matrix $\Sigma_{ll} = I$. Though the method easily can be generalized to points with varying weight, but still isotropic uncertainty, the points in real point clouds show a severe inhomogeneity and strong isotropies, whether the points are derived from stereo pairs or directly observed by LiDAR scanners, as shown in Fig. 4. Hence, we may cope with the uncertainty cases

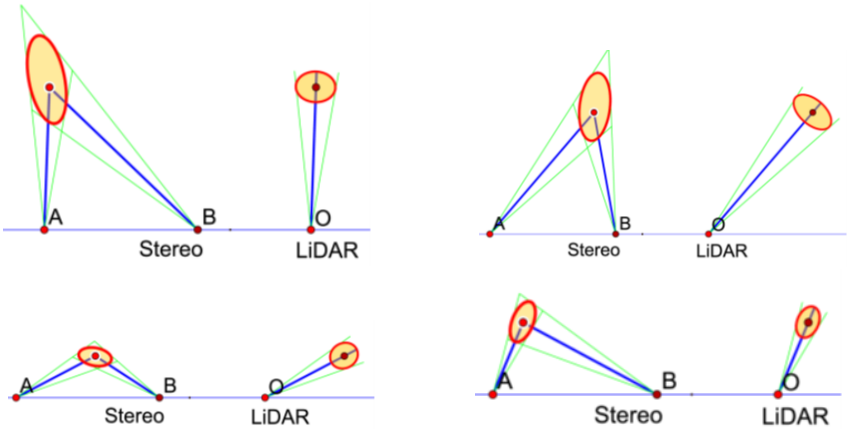


Figure 4: Point accuracy obtained with stereo and LiDAR systems. In both cases the uncertainty of the points is depending on the point coordinates and highly inhomogeneous and anisotropic

in Fig. 5. Only for the two cases in the lower row, direct solutions exist. In addition, the method assumes only the coordinates $\boldsymbol{X}'_i$ are uncertain, which generally does not reflect real situations.

### 4.2.2 Registration from plane correspondences

Khoshelham (2016) proposed a direct solution for determining the rigid motion between two point clouds based on plane correspondences using the model

$$\mathbb{E}(\mathbf{A}'_i) = \mathsf{M}(R, \boldsymbol{t})^{-\mathsf{T}}\mathbf{A}_i \quad \text{with} \quad \mathbf{A}_i = \begin{bmatrix} \boldsymbol{n}_i \\ -s_i \end{bmatrix}, \ \mathbf{A}'_i = \begin{bmatrix} \boldsymbol{n}'_i \\ -s'_i \end{bmatrix} \tag{80}$$

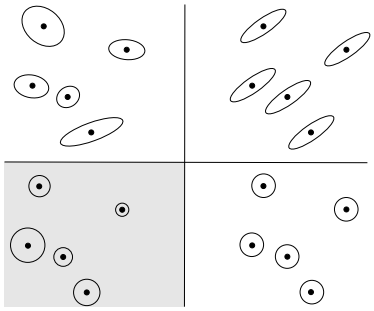

Figure 5: Uncertainty structure of point clouds **Top:** anisotropic, **Bottom:** isotrop. **Left:** inhomogeneus. **Right:** homogeneous. The direct solution of Arun et al. (1987) refers to the isotropic and homogeneous case, lower right, and – via weighting – can easily be generalized to the isotropic inhomogeneous case, lower left

with the homogeneous plane parameters $\mathbf{A}_i$ depending on the normalized normal vectors $\boldsymbol{N}_i$ and the distances $S_i$ of the plane from the origin. Rewriting the constraints as

$$\mathbf{0} = R\boldsymbol{n} - \boldsymbol{n}' \quad \text{and} \quad \mathbf{0} = \boldsymbol{n}^\mathsf{T}\boldsymbol{t} - s + s' \tag{81}$$

shows both constraints are linear in the elements of $R$ and $\boldsymbol{t}$. For a statistically optimal procedure we need the accuracy of the planes, derived from the point cloud, see (Förstner, 2024, p. 81 ff. and p. 100 ff). Some direct solutions for deriving motion parameters from corresponding planes a given in (Förstner, 2024, p. 106 ff.) Also this method is statistically suboptimal, but leads to sufficiently accurate approximations, similar to the algorithm by Arun et al. (1987). Förstner and Khoshelham (2017) proposed refinements of the algorithm by Khoshelham (2016) and compared it with the statistically optimal ML-estimate.

Here we augment these findings and report on the quality of approximate registration methods based on point correspondences and based on plane correspondences.

### 4.2.3 Approximating the stochastical model

We discuss the effect of simplifying the covariance matrix of the points in two point clouds with known correspondences onto the accuracy of the, now seven, parameters of a similarity transformation, thus including a scale parameter $s$. We choose three different stochastical models:

$$\Sigma^{(W)}_{x_i x_i} = \Sigma(\boldsymbol{x}_i)\,, \quad \Sigma^{(w)}_{x_i x_i} = \sigma^2_{x_i} I_3\,, \quad \Sigma^{(1)}_{x_i x_i} = \sigma^2 I_3 \tag{82}$$

We consider two cases, which differ in the angle of the viewing directions.

1. In the first case we assume the point lie in a spatial corridor. We assume $I = 100$ 3D points, which are observed in a rectangular box of size $200 \times 20 \times 900$ [m]$^3$ in an average distance of 550 m using a Leica RTC360 Scanner, positioned at a distance of 200 m.

   The data sheet of the scanner mentions the following accuracy of the instrument: The distance variance has a constant part of 1 [mm$^2$], and a distance dependent part of (10 [ppm])$^2$, the directional uncertainty is given with 18". Thus, we use

   $$\sigma_d = \sqrt{(0.001 \text{ [m] })^2 + (10^{-5} d)^2} \quad \text{and} \quad \sigma_\alpha = 18'', \tag{83}$$

   Fig. 6, left and middle show the observed coordinates with the corresponding viewing point and the corresponding standard ellipses (blown up). The uncertainty pattern is highly inhomogeneous and anisotropic. In Fig. 6, right we see the observed point together with the two viewing points. The ellipses indicat the joint uncertainty of the points in the left and the right point cloud. The ellipses indicate the uncertainty of the difference of the coordinates in the two coordinate systems, technically representing the covariance matrix $\Sigma_{nn} = B^\mathsf{T} \Sigma_{ll} B$, see eq. (8). Since the viewing points are quite close together, the ellipses only are a bit thicker than in the individual point clouds, and will be used for determining the similarity transformation.

   The following table shows the ratios of the standard deviations for the three solution pairs, based on the solution with constant weight (1), with individually weighted circular uncertainty, and with the rigorous determination of the covariance matrices (W).

| | $s$ | $r_1$ | $r_2$ | $r_3$ | $t_1$ | $t_2$ | $t_3$ | max |
|---|---|---|---|---|---|---|---|---|
| 1W | 1.4 | 1.44 | 1.53 | 1.74 | 2.48 | 1.74 | 1.60 | 2.48 |
| wW | 1.65 | 1.01 | 1.00 | 1.14 | 1.30 | 1.39 | 1.01 | 1.65 |
| 1w | 0.84 | 1.43 | 1.52 | 1.53 | 1.9 0 | 1.25 | 1.59 | 1.90 |

Table 1: Ratio of the standard deviations of the seven parameters $(s, \boldsymbol{r}, \boldsymbol{t})$ for the three comparisons of the approximate methods $(1, w)$ and the statistically optimal solution $(W)$. Case with narrow basis 200 m

2. Now, we assume that the same points are observed from two mutually far viewpoints, with a basis of 1000 m, such that the viewing directions intersect nearly with 90°.

   Fig. 7, left and middle again show the observed coordinate with the individual viewing point and the corresponding standard ellipses. The ellipses in the right part again show the standard ellipses of the difference of the coordinate in both systems. Now, since the viewing directions are nearly perpendicular, taking differendes leads to nearly circular standard ellipses, though still inhomogeneous. Therefore, we can expect, that the differences

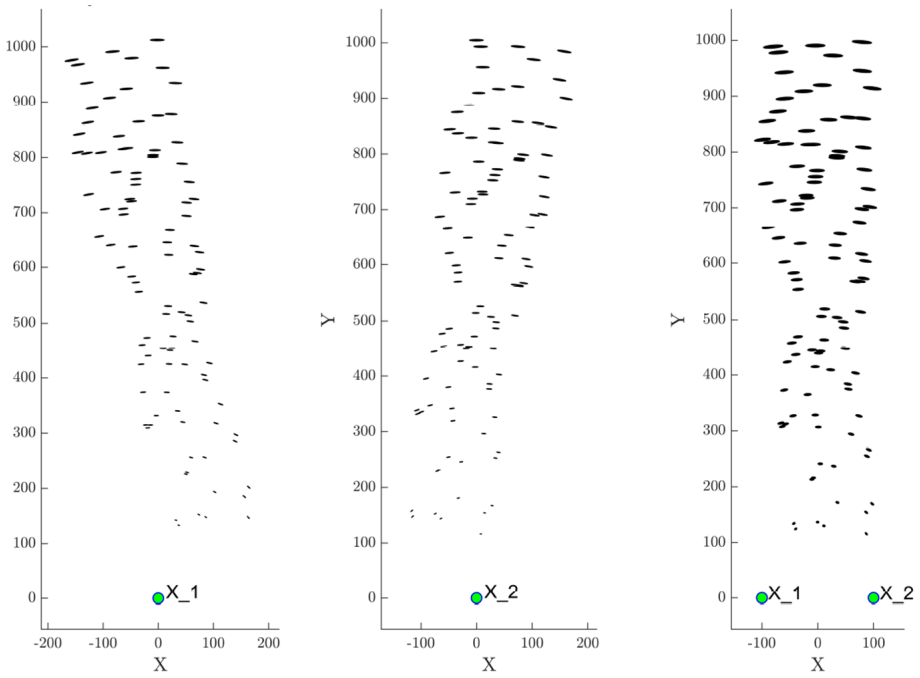


Figure 6: Observed coordinates with (blown up) standard ellipses. Small basis with 200 m

between the three methods will be smaller. This is confirmed by the values in the following table 2. The maximal ratios of the standard deviations are below 1.6D. The quadratic mean of the ratios of the standard deviations

| | $s$ | $r_1$ | $r_2$ | $r_3$ | $t_1$ | $t_2$ | $t_3$ | max |
|---|---|---|---|---|---|---|---|---|
| 1W | 1.02 | 1.10 | 1.12 | 1.49 | 1.02 | 1.58 | 1.12 | 1.58 |
| wW | 1.09 | 1.01 | 1.01 | 1.20 | 1.07 | 1.23 | 1.01 | 1.23 |
| 1w | 0.94 | 1.10 | 1.11 | 1.24 | 0.95 | 1.28 | 1.11 | 1.28 |

Table 2: Ratio of the standard deviations of the seven parameters $(s, \boldsymbol{r}, \boldsymbol{t})$ for the three comparisons of the approximate methods $(1, w)$ and the statistically optimal solution $(W)$. Case with wide basis 1000 m

are shown in Fig. 8 for the narrow and the wide basis. Obviously, the mean ratios of the standard deviations a much smaller if the point cloud is observed from different directions.

This type of analysis, can be made for any configuration, and used for planning purposes, in this case also the decision on an adequate stochastical model.

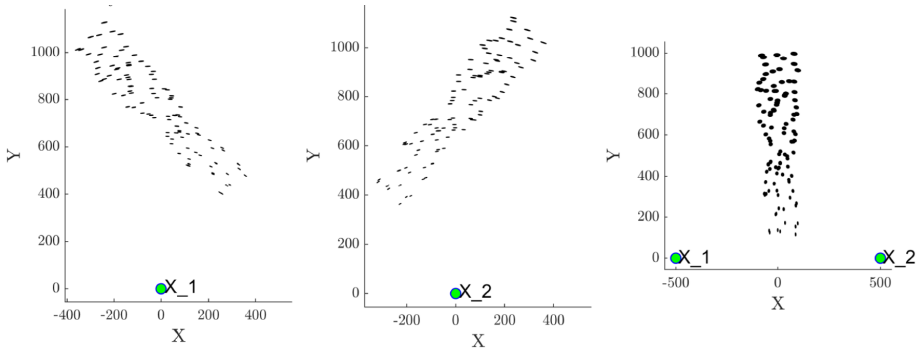


Figure 7: Observed coordinates with (blown up) standard ellipses. Wide basis with 1000 m

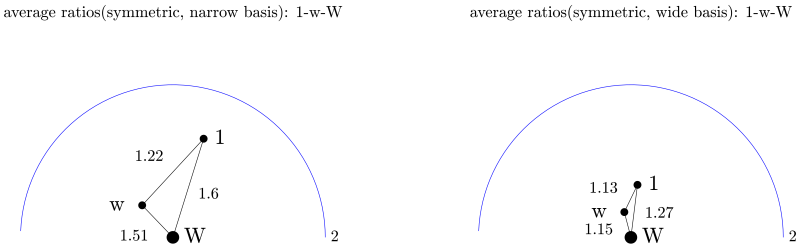


Figure 8: Mean relative standard deviations of parameters for the case with narrow and wide basis

### 4.2.4 Approximating the algorithm

We now compare three approximate methods with the ML-estimation for determining the rigid motion between two point clouds based on given plane correspondences, see Förstner and Khoshelham (2017). We take the following approximate methods

ALG the algebraic closed-form solution proposed by Khoshelham (2016). It minimizes the residual of the constraints

$$E^{[\mathrm{ALG}])} = |A(\boldsymbol{l})\boldsymbol{x}|^2 = (A(\boldsymbol{l})\boldsymbol{x})^\mathsf{T}\ A(\boldsymbol{l})\boldsymbol{x} \tag{84}$$

ALGw The algebraic solution with weighting the constraints. Since the constraints $\boldsymbol{g} = A(\boldsymbol{l})\boldsymbol{x}$ have a covariance matrix of $\Sigma_{gg} = B^\mathsf{T}\Sigma_{ll}B$, we minimize the weighted residual of the algebraic error

$$E^{[\mathrm{ALGw}])} = |A(\boldsymbol{l})\boldsymbol{x}|^2_{W_{gg}} = (A(\boldsymbol{l})\boldsymbol{x})^\mathsf{T}\ \Sigma_{gg}^{-1}\ A(\boldsymbol{l})\boldsymbol{x} \tag{85}$$

Since the matrix $B$ is now known at the beginning, the methods is a two-step method: (1) minimizing $E^{[\mathrm{ALG}])}$, (2) determining $B$ and minimizing $E^{[\mathrm{ALGw}])}$.

ML-1 Instead of refining the algebraic method, we apply a single iteration of a ML-estimation, based on the result of the algebraic minimization, thus also performing a two-step procedure.

We analyse compare the methods using simulated data, in order to be sure, that the sampling of the observations follows the assumed covariance matrices. The test refers to $J = 50$ plane pairs, randomly chosen in a cube $[-1, +1]^3$ and covariance matrices chosen as $\Sigma_{A_iA_i} = \sigma_A^2(I + UU^\mathsf{T})$ with random matrices having standard normally distributed elements, for details see (Förstner and Khoshelham, 2017, sect. 4.1). The average and the maximum loss in accuracy $\overline{\mu}$ and $\max_\mu$ following (53) and (54) is shown in Fig. 9 as a function of the number of corresponding planes. Obviously, the two refined approximate methods

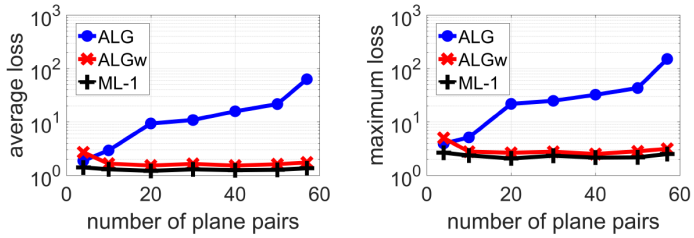


Figure 9: Loss in accuracy of the three direct solutions shown as a function of the number of plane pairs, realistic configuration, $J = 100$ samples. **Left:** Average loss $\overline{\mu}$. **Right:** Maximum loss $\max_\mu$. Zero-loss is achieved for values $1 = 10^0$, the bottom line of the graphs. The single iteration ML solution (ML-1) and the whitened algebraic solution (ALGw) have a significantly smaller loss of accuracy as compared to the algebraic solution (ALG) in terms of both measures

(ALGw) and (ML-1) consistently show no severe loss in accuracy, in contrast to the algebraic solution.

### 4.2.5 Comparison with ICP

Table 3 shows the result of the comparison between the direct plane-to-plane methods and the ICP variants in terms of loss of accuracy with respect to the optimal ML solution. While ICP-pl performs better than ICP-pt, both ICP variants have a significantly larger loss of accuracy as compared to ML-1 and ALGw which exploit the uncertainty of the planes. Using the purely algebraic solution also shows large losses in accuracy, being a factor 45 worse in standard deviation, compared to the ML-solution.

| Method | Average loss | Maximum loss |
|---|---|---|
| ML-1 | 1.24 | 2.04 |
| ALGw | 1.61 | 2.76 |
| ALG | 45.15 | 104.80 |
| ICP-pl | 3.66 | 6.09 |
| ICP-pt | 5.82 | 12.11 |

Table 3: Comparison of direct plane-to-plane methods with the point-to-point ICP (ICP-pt) and point-to-plane ICP (ICP-pl) in terms of loss of accuracy with respect to the optimal ML solution (realistic configuration, 57 plane pairs, $J = 100$ iterations)

#### 4.2.6 Computing times

Tab. 4 shows the computing times for a small and a large example. The computation times for 1000 (randomly generated) plane pairs shows that the estimation methods can be scaled up to larger scenes with more planes and perform in real time. The approximate methods ML-1 and ALGw, are significantly faster than the ML method. Especially, the ML-1 method, with a single addition ML-iteration appears to be the best choice.

However, this is only a tiny fraction of the CPU time for establishing the correspondences, as to be expected.

| # pairs | ML | ML-1 | ALGw | ALG |
|---|---|---|---|---|
| 57 | 0.085 | 0.012 | 0.021 | 0.010 |
| 1000 | 1.6247 | 0.1331 | 0.3088 | 0.0653 |

Table 4: CPU times (in seconds) for the optimal and direct motion estimation methods for 57 and 1000 plane pairs

### 4.3 Bundle adjustment with trilinear constraints

The last example shows the effect of different approximations possibly used in bundle adjustment and is taken from Schneider et al. (2017).

Instead of using collinearity equation, we use two epipolar and one trifocal constraint. The advantage is: the methods allows for points which are very far without leading to singularities. In detail we have

$$g_1(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{m}_i, \boldsymbol{m}_j) = 0\,, \quad g_2(\boldsymbol{x}_j, \boldsymbol{x}_k, \boldsymbol{m}_j, \boldsymbol{m}_k) = 0\,, \quad g_3(\boldsymbol{x}_i, \boldsymbol{x}_j, \boldsymbol{x}_k, \boldsymbol{m}_i, \boldsymbol{m}_j, \boldsymbol{m}_k) = 0\,, \tag{86}$$

with 6D vectors $\boldsymbol{m}_i$ representing the camera poses. Hence, the constraints between the images lead to non-zeros in the off diagonal terms of the Jacobian $B$ of the constraints w.r.t. observations.

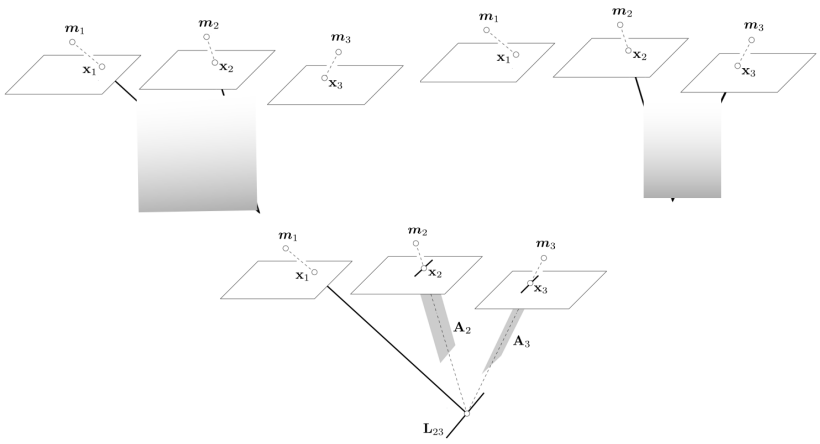


Figure 10: Two epipolar and one trifocal constraint per 3D point

The structure of normal equation system is the following:

$$N\Delta\boldsymbol{x} = \boldsymbol{n} \quad \text{with} \quad N = A^{\mathsf{T}}\underbrace{(B^{\mathsf{T}}\Sigma_{ll}B)^{-1}A}_{=C}, \quad \text{and} \quad \boldsymbol{n} = A^{\mathsf{T}}\underbrace{(B^{\mathsf{T}}\Sigma_{ll}B)^{-1}\Delta\boldsymbol{l}}_{=\boldsymbol{c}}. \tag{87}$$

Both matrices, $B$ and $N$ are sparse, but not diagonal. We need to avoid the inversion of $B^{\mathsf{T}}\Sigma_{ll}B$, since the inverse of a non-diagonal matrix usually is full. We may circumvent this by solving two sparse equation systems, first determining $[C, \boldsymbol{c}]$ from

$$B^{\mathsf{T}}\Sigma_{ll}B \;\; [C, \boldsymbol{c}] = [A, \Delta\boldsymbol{l}] \quad \text{and} \quad A^{\mathsf{T}}C \;\; \Delta\boldsymbol{x} = A^{\mathsf{T}}\boldsymbol{c}\,. \tag{88}$$

and then solving the generally sparse normal equation system.

We report on the effect of the following approximations, all referring to the weight matrix

$$W_{nn} = B^{\mathsf{T}}\Sigma_{ll}B \tag{89}$$

of the constraints.

A Approximating the Jacobians, by evaluating them at the observations instead of at the fitted observations. The result will depend on the noise level of the observations.

B Neglecting the correlations between the constraints, i.e., by using

$$W_{nn}^{[B]} = \mathrm{Diag}(W_{nn})\,. \tag{90}$$

C Applying both approximations A and B.

D Using the weight matrix $W_{nn}$ from the first iteration. The result will depend on the quality of the approximate values.

| name | image size | focel length | distance |
|---|---|---|---|
| BUILDING | 5 MPixel | 1589 pixel | approx 15 m - 60 m |
| FIELD | 12 MPixel | 2347 pixel | 100 m |

Table 5: Datasets with some characteristics

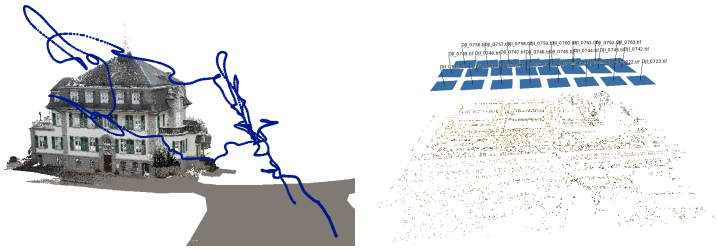

Figure 11: Data sets BUILDING and FIELD

The following experiments were performed with two data sets, taken with a UAV. The layout is shown in Fig. 11 Reference data are obtained from a ML estimation from which the estimated parameters $\widehat{\boldsymbol{x}}$ and the fitted observations $\widehat{l} = \widehat{\boldsymbol{y}}$ are taken as true values. The experiments are repeated 100 times, each time contaminating the observations with varying noise of level $\sigma_{0l}$. For case D, where the Jacobian is fixed after the first iteration, the approximate values are randomly distorted within a range of 0.1 % to 10 % of the relative precision $\sigma_{0x}$ of the pose parameters.

Fig. 12 shows the loss in accuracy measured by the difference $\Delta F$ from (57) of the Mahalanobis distance of the four cases of the approximated bundle adjustments, when compared to the statistically rigorous solution. We provide the loss for different moderate noise levels, from 0.1 mrad to 3.0 mrad directional standard deviation, for focal lengths of 2000 pixel this corresponds to standard deviations of 0.2 pixel to 6 pixel in the image.

In all cases the effect of the approximation increases linearly with the noise level. Obviously, the approximation of the Jacobian (case A, blue lines) by evaluating them at the observations instead of at the fitted observations leads to the smallest loss in accuracy. The effect of neglecting the correlations (case B, red lines) is larger. Hence, the combined effect (case C, green lines) shows an increase up to more than 40 %. The effect of fixing the Jacobian (case D, black line) is even larger.

The effect of the variation of the approximate values onto the result when fixing the Jacobian after the first iteration is shown in Tab. 6. The variation of the approximation is also shown in degrees: they indicate the variation of the relative poses of neighbouring cameras. Also here, the effect onto the accuracy reaches values up to 25 %.

The accuracy loss in these experiments are mostly below 30 %. This might be acceptable, if computing time is essential.

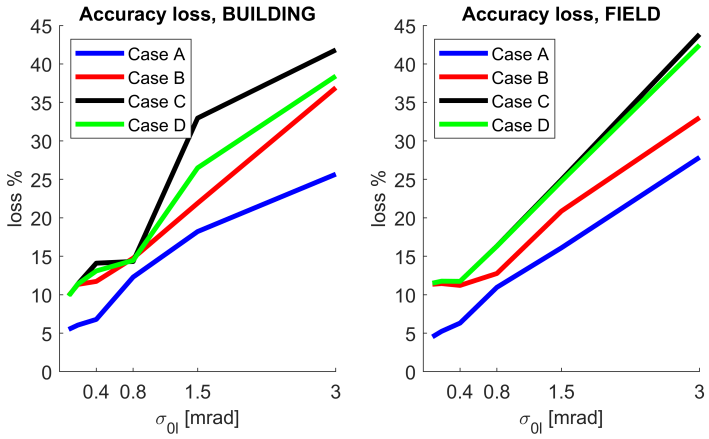


Figure 12: The loss in accuracy $\Delta F_{\text{Case}}$ in percent of estimated pose parameters induced by the individual approximations of Case A-D at different noise levels $\sigma_{0l}$ in radian. Cases A: approximate Jacobian, Cases B: neglecting correlations, Cases C: cases A and B, Cases D: fixing Jacobian of first iteration

| $\sigma_{0x}$ [mrad] | 1.0 | 3.0 | 10.0 | 30.0 | 100.0 |
|---|---|---|---|---|---|
| $\sigma_{0x}$ [°] | 0.06 | 0.17 | 0.57 | 1.72 | 5.73 |
| Building | | | | | |
| Case D | 9.82 | 10.52 | 17.16 | 20.97 | 31.44 |
| Field | | | | | |
| Case D | 11.50 | 13.28 | 14.52 | 16.94 | 25.36 |

Table 6: The loss in accuracy in percent of estimated pose parameters at noise level $\sigma_{0l} = 0.1$ mrad and different relative precision $\sigma_{0x}$ of approximate values, the variation of the approximation also given in degrees

# 5 Closure

We discussed several aspects of approximate and rigorous methods for parameter estimation:

- Since the notion of a rigorous method is linked to some specified optimization measure, characterizing a method as rigorous is not unique.
- Statistically optimal estimation methods often refer to maximum likelihood estimation, mostly including the assumption of normally distributed observations. This of course is not valid for robust methods, which we did not discuss.
- A large class of approximate methods may be upgraded (1) by providing a reliable prediction about the covariance matrix of the (suboptimally)

estimated parameters, and (2) by adding a single second algorithmic step to obtain a statistically nearly optimal method, making the CPU time predictable.

- The decision on which method to be applied, be it approximate or rigorous, lies in the hand of the user, since it depends on the criteria for accepting a possibly approximate solution in a specific application area.
- The author of (a paper on) a method, again be it approximate or rigorous, needs to provide some data sheet on the performance of the proposed method, which allows the user to evaluate, whether in his/her application area the method might be appropriate.
- We discussed several estimation problems which are regularly solved with approximate or statistically optimal methods. The range of effects of approximations onto the estimated parameters varies between 0 and infinite, thus needs to be investigated for each type of estimation problem and each class of application scenario. In many practical cases the effect is comparably small, sometimes leading to accuracy losses up to a factor 100, however, often below a factor 2.

The presented examples may stimulate further investigations into the characteristics of approximate methods.

The transfer of the discussed ideas on specifying and evaluating approximate and rigorous methods into the area of deep learning is difficult, since short cuts, such als algebraic derivations of covariance matrices, for evaluating the result not yet exist. The lack of a closed world assumption, which exists when estimating geometric entities, represents a real challenge.

## References


Arun, K. S., T. S. Huang, and S. B. Blostein (1987). Least-Squares Fitting of Two 3D Point Sets. *IEEE T-PAMI 9*(5), 698–700. 23, 24

Barath, D., M. Polic, W. Förstner, T. Sattler, T. Pajdla, and Z. Kukelova (2020). Making Affine Correspondences Work in Camera Geometry Computation Supplementary Material. In *Proc. of ECCV*. 12

Boyd, S. and L. Vandenberghe (2004). *Convex optimization*. Cambridge University Press. 3

Förstner, W. (2024). Selected Notes — Appendix to DGPF Tutorial on Approximate and Rigorous Estimation Methods. 3, 4, 16, 18, 20, 24

Förstner, W. and K. Khoshelham (2017). Efficient and Accurate Registration of Point Clouds with Plane to Plane Correspondences. In *3rd International Workshop on Recovering 6D Object Pose*. 24, 27, 28

Förstner, W. and B. P. Wrobel (2016). *Photogrammetric Computer Vision – Statistics, Geometry, Orientation and Reconstruction*. Springer. 12, 13, 14

Khoshelham, K. (2016, April). Closed-form solutions for estimating a rigid motion from plane correspondences extracted from point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing 114*, 78–91. 23, 24, 27

Mangelson, J. G., M. Ghaffari, R. Vasudevan, and R. M. Eustice (2019). Characterizing the Uncertainty of Jointly Distributed Poses in the Lie Algebra. *IEEE Transactions on Robotics 36*, 1371–1388. 16

Rao, R. C. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob., Vol. 1*, pp. 355–372. Univ. of Calif. Press. Lemma 5a. 20

Schneider, J., C. Stachniss, and W. Förstner (2017). On the quality and efficiency of approximate solutions to bundle adjustment with epipolar and trifocal constraints. In *Proceedings of the ISPRS Conference on Unmanned Aerial Vehicles in Geomatics (UAV-g)*. 29