Selected Notes

Wolfgang Förstner

March 12, 2024

Preamble

This is a selection of notes I wrote for publications or for communicating with collegues on interesting topics. They are meant as an appendix for the DGPF-2024 tutorial on approximate and rigorous methods for estimation.

Wolfgang Förstner, Bonn
March 12, 2024

Contents

I Technical Notes on Statistics and Estimation Theory **5**

1 Gauss–Helmert Model as Optimization Problem **6**

The Gauss–Helmert generalizes the well-known Gauss–Markov model by allowing implicit relations between the observations and the unknown parameters. The classical derivation of the estimation procedure refers to the statistical nature of the Maximum-Likelihood optimization. The note separates the description of the model and the optimization function from the generally iterative numerical optimization procedure, in order to elucidate the non-statistical properties of the intermediate steps before treating point of convergence as final estimate.

2 Pre-calibration and in-situ Self-calibration with Correlated Observations **22**

Deformation analysis based on point clouds taken at different times may require to take into account both preclalibration and in situ self-calibration of the used instruments. We analyse the mutual effect of pre-calibration and in-situ self-calibration w.r.t. (1) the necessity to exploit the full covariance structure of the point cloud induced by the pre-calibration and (2) the possibility of increasing the computational efficiency during the self-calibration.

3 The Mean of Correlated Observations **37**

For uncorrelated observations the accuracy of the mean increases with the number of observations. In case they are correlated, there is an upper limit for the accuracy. The note analyses the situation for constant correlation and for exponentially decaying correlation, autoregressive noise.

4 Accuracy of the Mean when using a Wrong Covariance Matrix **43**

Suboptimal, i.e., approximate solutions often are used or needed when estimating parameters. One of such simplifications refers to the stochastical model, especially the covariance matrix of the observations, which often is assume to be a multiple of a unit matrix, implicitly assuming all observations have the same weight and are mutually uncorrelated. This note provides the general relation between the accuracy of the estimated parameters when using an approximate covariance matrix and exemplifies this using the mean of repeated observations.

II Technical Notes on Estimation of Geometric Entities **48**

5 Motions and their Uncertainty **49**

We addresses the ambiguity of representing uncertain motions. We analyse the relation between an exponential representation with a homogeneous 4x4 matrix and the representation with a rotation matrix, also represented exponentially, and a translation vector. The rotation parts turns out to be identical, while the translation parts differ, why a transparent documentation of the motion representation is necessary. As a sideline, the note addresses the inversion, the concatenation, and the difference of uncertain rotations and uncertain motions.

6 Centroid Form of an Uncertain Plane **81**
A plane can be represented in various manners. We especially discuss the centroid form of an uncertain plane, which naturally results from estimating a plane from a given point set. We discuss the representation, its recursive estimation assuming isotropic point uncertainty and optimal estimation.

7 Planes from Points **100**
We describe the statistically optimal estimation of a single and of multiple planes from a point cloud, where the full covariance matrix of all scene coordinates is available, e.g., from bundle adjustment. This procedure might be used to derive ground truth data for plane extraction or for homography estimation.

8 Direct Solutions for the Similarity from Plane Pairs **106**
We collect some direct solutions for determining the similarity (or motion) from corresponding plane pairs, representing point clouds. Some of the solutions are able to handle the case, where the sign of the normals are not consistent.

III Technical Notes on Bundle Adjustment and Surface Reconstruction **111**

9 Rule of Thumb for Precision of Points from Multiview Triangulation **112**
For planning bundle adjustment configurations, the expected accuracy of triangulated points is an essential ingredient. We derive rules of thumb for the accuracy of multi-view triangulating by providing simple expressions for the depth and lateral accuracy of 3D points, for images arranged in a line, in a planar region and in a spherical region, covering the case of omnidirectional cameras.

10 Multi-View Triangulation with Directions **117**
We provide simple solution to the optimal triangulation of a scene point from multiple views assuming isotropic uncertainty of the directions. As a special case we provide a simple expression for the distance of the triangulated point in case of homogeneous directional uncertainty and small basis, expressed as a function of the effective base line, the viewing angle and the resolution of an omnidirectional camera and the matching accuracy in pixels.

Part I

Technical Notes on Statistics and Estimation Theory

1 Gauss–Helmert Model as Optimization Problem

The Gauss–Helmert generalizes the well-known Gauss–Markov model by allowing implicit relations between the observations and the unknown parameters. The classical derivation of the estimation procedure refers to the statistical nature of the Maximum-Likelihood optimization. The note separates the description of the model and the optimization function from the generally iterative numerical optimization procedure, in order to elucidate the non-statistical properties of the intermediate steps before treating point of convergence as final estimate.

1.1 Preface . . . 6
1.2 Motivation . . . 7
1.3 The Gauss–Helmert model for estimating parameters . . . 7
1.3.1 The mathematical model . . . 7
1.3.2 The task . . . 8
1.4 The solutions for linear models . . . 9
1.4.1 The solution for the basic linear Gauss–Helmert model . . . 9
1.4.2 The Gauss–Helmert model for general covariance matrix . . . 10
1.4.3 Gauss–Markov model . . . 11
1.4.4 Model with constraints between the observations only . . . 12
1.4.5 The quasi Gauss–Markov model . . . 12
1.4.6 Results using pseudo inverses . . . 13
1.5 The non-linear Gauss–Helmert model . . . 15
1.5.1 The algorithm for estimating the parameters . . . 17
1.5.2 Derivation of the procedure . . . 18

Remark: While throughout the notes we use one of the classical statistical notation (observations $\boldsymbol{y}$, and parameters $\boldsymbol{\theta}$) , in this note we adopt one of the notations used in Geodesy and Photogrammetry which better fits to the notation used by Boyd and Vandenberghe (2004), thus we name the observations $\boldsymbol{l}$ and the unknown parameters $\boldsymbol{x}$. ◇

1.1 Preface

This note (2021) describes the estimation within the Gauss–Helmert model as a specific optimization problem, making explicit the numerical character of the numerical process for determining the parameters, omitting the statistical interpretation of the intermediate steps within the optimization procedure. This clarifies (1) the role of the stochastical model at the beginning of statistical parameter estimation task and used for evaluating the uncertainty of the result, and (2) the non-statistical role of the numerical method for achieving the final parameters. It is common to derive the estimator for a parameter vector within a statistical framework, and not distinguish the different aspects of the whole task: (a) the specification of the model, (b) the specification of the optimization function, (c) the numerical process of optimization, and (d) the evaluation of the obtained parameters. This note is intended to separate these steps.

1.2 Motivation

Parameter estimation consists in determining unknown parameters from given observations. Its mathematical model consists of the functional model, relating the mean values of the observations to the unknown parameters, and the stochastical model which describes the uncertainty of observation process. We often categorize functional models according to their algebraic structure. The *Gauss–Markov model* is a functional model, where the mean observations are an *explicit* function $\mathbb{E}(\underline{\boldsymbol{l}}) = \boldsymbol{f}(\boldsymbol{x})$ of the parameters.

Gauss–Markov model

Here we discuss the mathematical model of an estimation task with a functional model, where the mean observations and the parameters are related by an *implicit* function This is called the *Gauss–Helmert model*. Given are N observations $\boldsymbol{l}$ together with the uncertainty of the observation process $\mathbb{D}(\underline{\boldsymbol{l}})$, implicitly assuming the measuring deviations are normally distributed. The mean values $\mathbb{E}(\underline{\boldsymbol{l}})$ of the observations are functionally related to unknown parameters $\boldsymbol{x}$ by G implicit equations

Gauss–Helmert model

$$\boldsymbol{g}(\boldsymbol{x}, \mathbb{E}(\underline{\boldsymbol{l}})) = \boldsymbol{0}\,. \tag{1.1}$$

The task is to find optimal estimates $\boldsymbol{x}$ for the unknown parameters.

The derivation, presented here, is based on the following assumptions.

- We consider the cases where the representation of the parameters and observations may be redundant, such as for normalized homogeneous coordinates or rotation matrices. Instead of including constraints, such as a length or an orthogonality constraint, we allow that the estimation refers to a minimal representation of the corrections, close to the approximate values of the parameters or the observations, namely in the tangent space defined by the individual constraints. As a consequence, the observations and parameters may be lists of individual groups of possibly redundantly represented entities, e.g., $\boldsymbol{x} := \{R, \boldsymbol{t}, \lambda\}$ for the rotation, the translation, and the scale of a spatial similarity, the corrections, however, are vectors of a locally minimal representation, e.g., $\Delta\boldsymbol{x} = [\Delta\boldsymbol{r}^\mathsf{T}, \Delta\boldsymbol{t}^\mathsf{T}, \Delta\lambda]^\mathsf{T}$, where $\Delta\boldsymbol{r}$ describes a small rotation with three parameters.
- We treat the expectation of the observations $\boldsymbol{y} = \mathbb{E}(\underline{\boldsymbol{l}})$ as unknowns. This is a consequence of the previous point and in contrast to classical setups, where the optimization function has the residuals as unknown. In the linearized model the corrections $\Delta\boldsymbol{y}$ and $\Delta\boldsymbol{x}$ to the expectation of the observations $\mathbb{E}(\underline{\boldsymbol{l}})$ and the parameters $\boldsymbol{x}$ are unknown, which allows us to update them in the original, non-linear model taking their algebraic properties, e.g., length or orthogonality, into account.

1.3 The Gauss–Helmert model for estimating parameters

We now describe the set-up of the estimation procedure with a Gauss–Helmert model as functional model, derive the optimization task, provide a solution for the case where the model is linear, finally provide the solution to the non-linear model using a linearized model within an iterative scheme.

1.3.1 The mathematical model

We start from N given observations, collected in the N-vector $\boldsymbol{l}$. We assume, they are a sample of a normal distribution, specified by the *unknown expectation vector and partially known dispersion matrix*. The *stochastical model* for the observation process therefore is given by

stochastical model

$$\underline{\boldsymbol{l}} \sim \mathcal{N}(\mathbb{E}(\underline{\boldsymbol{l}}), \mathbb{D}(\underline{\boldsymbol{l}}))\,, \tag{1.2}$$

The dispersion matrix of the observations

variance factor σ_0^2

$$\mathbb{D}(\underset{\sim}{\boldsymbol{l}}) = \sigma_0^2\,\Sigma_{ll}\,. \tag{1.3}$$

is specified by an approximate covariance matrix Σ_{ll} which differs from the true covariance matrix by an *unknown variance factor* σ_0^2. The *functional model* of the Gauss–Helmert assumes the U unknown parameters $\boldsymbol{x}$ and the N unknown mean values $\mathbb{E}(\underset{\sim}{\boldsymbol{l}})$ are constrained by the following G-dimensional implicit function[1][2]

functional model

$$\underset{G\times 1}{\boldsymbol{g}}\,(\underset{U\times 1}{\boldsymbol{x}}, \underset{N\times 1}{\mathbb{E}(\underset{\sim}{\boldsymbol{l}})}) = \mathbf{0}\,. \tag{1.4}$$

Observe, that (1.2) can be interpreted as the likelihood function of the unknown parameters $\boldsymbol{x}$

likelihood function

$$L(\boldsymbol{x}) := L(\boldsymbol{x}, \boldsymbol{g}) = p(\boldsymbol{l} \mid \boldsymbol{x}, \boldsymbol{g}) = \mathcal{M}(\mathbb{E}(\underset{\sim}{\boldsymbol{l}} \mid \boldsymbol{x}, \boldsymbol{g}), \mathbb{D}(\underset{\sim}{\boldsymbol{l}} \mid \boldsymbol{x}, \boldsymbol{g}))\,, \tag{1.5}$$

for given observations $\boldsymbol{l}$ and functions $\boldsymbol{g}$, where the distribution $\mathcal{M}$ is characterized by its first and second moment. In order to be able to determine the U parameters $\boldsymbol{x}$ we need to require there are at least as many constraints as unknowns:

$$G \geq U\,, \tag{1.6}$$

redundancy

or that the number of redundant constraints, i.e., the redundancy

$$R = G - U \geq 0\,. \tag{1.7}$$

is non-negative. Similarly, in order to have a guarantee that the implicit function (1.4) of $[\boldsymbol{x}^\mathsf{T}, \mathbb{E}(\underset{\sim}{\boldsymbol{l}}^\mathsf{T})] \in \mathbb{R}^{U+N}$ is not empty, the number G of constraints should not exceed $U + N$, hence

$$N \geq G - U \tag{1.8}$$

Therefore we have the following relation

$$N \geq R \geq 0 \tag{1.9}$$

as a necessary condition for the model setup.

1.3.2 The task

unknown parameters $\boldsymbol{x}$ and unknown mean observations $\boldsymbol{y}$

The goal is to find the maximum-likelihood estimates $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{y}}$ for the unknown parameters $\boldsymbol{x}$ and the unknown expectation of the observations, short, the *mean observations* $\boldsymbol{y} = \mathbb{E}(\underset{\sim}{\boldsymbol{l}})$ such that the weighted sum of the residuals[3]

$$\widehat{\boldsymbol{v}} = \widehat{\boldsymbol{y}} - \boldsymbol{l}\,, \tag{1.10}$$

namely

$$\Omega = \boldsymbol{v}^\mathsf{T}\Sigma_{ll}^{-1}\boldsymbol{v}\,, \tag{1.11}$$

becomes minimum and the estimates fulfil the constraints

$$\boldsymbol{g}(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}}) = \mathbf{0}\,. \tag{1.12}$$

Observe,

- the optimization function (1.11) does not depend on the variance factor σ_0^2.

[1] The Gauss–Markov model $\mathbb{E}(\underset{\sim}{\boldsymbol{l}}) = \boldsymbol{f}(\boldsymbol{x})$ therefore can be interpreted as a special case of the Gauss–Helmert model, setting $\boldsymbol{g}(\boldsymbol{x}, \mathbb{E}(\underset{\sim}{\boldsymbol{l}})) = -\mathbb{E}(\underset{\sim}{\boldsymbol{l}}) + \boldsymbol{f}(\boldsymbol{x})$

[2] The definition of the implicit function is different from Förstner and Wrobel (2016, Eq. (4.426)), where the two arguments of the implicit function $\boldsymbol{g}$ are exchanged

[3] We us the variable $\boldsymbol{y}$ for the mean observation, in order to avoid to define approximate values for the fitted observations $\boldsymbol{y}$ within the iteration loop, since the intermediate values in an iteration scheme have no statistical meaning.

- 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:[4] For given observations $\boldsymbol{l}$, constraints $\boldsymbol{g}$ and weight matrix $W_{ll} = \Sigma_{ll}^{-1}$ find values for $\boldsymbol{x}$ and $\boldsymbol{y}$ that

$$\begin{aligned} &\text{minimize} && (\boldsymbol{y}-\boldsymbol{l})^{\mathsf{T}} W_{ll} (\boldsymbol{y}-\boldsymbol{l}) && (1.13)\\ &\text{subject to} && \boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{0}\,, && (1.14) \end{aligned}$$

  where $\boldsymbol{y}$ stands for the unknown mean observation $\mathbb{E}(\underline{\boldsymbol{l}})$.

Remark: We may assume the observations appear in I statistically independent groups $\{\boldsymbol{l}_i, \Sigma_{l_i l_i}\}, i = 1, ..., I$, and if the dimension of these groups is the same, say d, we have $N = dI$. Furthermore, we often face the situation, that the constraints only refer to one group of observations. Then the functional model (1.12) can be written as

$$\boldsymbol{g}_i(\boldsymbol{x}, \boldsymbol{y}_i) = \boldsymbol{0}\,, \quad i = 1, ..., I\,. \tag{1.15}$$

Hence, if the number of constraints per group is constant, say c, then the number of constraints is $G = cI$. As an example, this situation holds for the model of a 3D similarity for two sets of 3D points, where we have groups of $d = 6$ observations, namely the 3D coordinates in the two systems, and $c = 3$ constraints per group relating these coordinates via a similarity transformation with their parameters $\boldsymbol{x}$. ◇

We first provide a solution for the linear Gauss–Helmert model. We specialize it for independent and identically distributed observations and derive the solution for the two basic models, namely the Gauss–Markov model and the model with constraints between observations only. We also show, that the Gauss–Helmert model can be solved by chosing adequate substitute observations leading to a Gauss–Markov model. Since in case the model is non-linear the coefficient matrices need to be updated during the iteration process, why this model is called a *quasi Gauss–Markov model*. In the next section we then handle the non-linear case. Finally, we provide a derivation via an equivalent Gauss–Markov model.

1.4 The solutions for linear models

1.4.1 The solution for the basic linear Gauss–Helmert model

We start with the linear Gauss–Helmert model with covariance matrix $\mathbb{D}(\underline{\boldsymbol{l}}) = \Sigma_{ll} = I$. We handle it as an algebraic, not a statistical optimization problem.

The original optimization problem reads as: for given observations $\boldsymbol{l} \in \mathbb{R}^N$, a regular $N \times N$ covariance matrix $\Sigma_{ll} = I_N$, full rank coefficient matrices. $X \in \mathbb{R}^{G \times U}$ and $Y \in \mathbb{R}^{G \times N}$ and a constant vector $\boldsymbol{b} \in \mathbb{R}^G$

$$\boxed{\begin{aligned} &\text{GHM:} && \text{minimize} && (\boldsymbol{y}-\boldsymbol{l})^{\mathsf{T}}(\boldsymbol{y}-\boldsymbol{l}) \\ &&& \text{subject to} && X\boldsymbol{x} + Y^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{b} = \boldsymbol{0}\,. \end{aligned}} \tag{1.16}$$

w.r.t. the unknown parameters $\boldsymbol{x}$ and the mean observations $\boldsymbol{y}$.

Hence, here we chose the constraint function

$$\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) = X\boldsymbol{x} + Y^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{b} \tag{1.17}$$

which is linear in the unknown parameters. The coefficient matrices often are called *design matrices*, since they specify the design of the observation process. They are assumed to be given and fixed.

Furthermore, for a compact representation of the solution we use the *substituted observations* $\boldsymbol{n}(\boldsymbol{l})$ together with their covariance matrix

$$\boldsymbol{n}(\boldsymbol{l}) = Y^{\mathsf{T}}\boldsymbol{l} + \boldsymbol{b} \quad \text{and} \quad \mathbb{D}(\underline{\boldsymbol{n}}) = \Sigma_{nn} = Y^{\mathsf{T}}Y\,. \tag{1.18}$$

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

We obtain the estimated parameters and the fitted observations from

$$\boxed{\begin{array}{l}\widehat{\boldsymbol{x}} = \; -(X^{\mathsf{T}}\Sigma_{nn}^{-1}X)^{-1}\, X^{\mathsf{T}}\Sigma_{nn}^{-1}\, \boldsymbol{n}(\boldsymbol{l}) \\ \widehat{\boldsymbol{y}} = \boldsymbol{l} - Y\Sigma_{nn}^{-1}\, \boldsymbol{g}(\widehat{\boldsymbol{x}}, \boldsymbol{l})\,.\end{array}} \tag{1.19}$$

Remark: Generally, the parameters are estimated based on the normal equations

$$(X^{\mathsf{T}}\Sigma_{nn}^{-1}X)\widehat{\boldsymbol{x}} + X^{\mathsf{T}}\Sigma_{nn}^{-1}\, \boldsymbol{n}(\boldsymbol{l}) = \boldsymbol{0} \tag{1.20}$$

which can be solved in any numerical manner, especially if we want ot exploit the sparsity of X, Σ_{ll}, or Σ_{nn}. ◇

Proof: Using Lagrangian multipliers we need to find the minimum of

$$\boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\lambda}) = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{l})^{\mathsf{T}}(\boldsymbol{y} - \boldsymbol{l}) + \boldsymbol{\lambda}^{\mathsf{T}}(X\boldsymbol{x} + Y^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{b})\,. \tag{1.21}$$

Necessary conditions are

$$\boldsymbol{0} = \frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{x}^{\mathsf{T}}} \;=\; X^{\mathsf{T}}\boldsymbol{\lambda} \tag{1.22}$$

$$\boldsymbol{0} = \frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{y}^{\mathsf{T}}} \;=\; \boldsymbol{y} - \boldsymbol{l} + Y\boldsymbol{\lambda} \tag{1.23}$$

$$\boldsymbol{0} = \frac{\partial \boldsymbol{\Phi}}{\partial \boldsymbol{\lambda}^{\mathsf{T}}} \;=\; X\boldsymbol{x} + Y^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{b}\,. \tag{1.24}$$

Multiplying (1.23) with Y^{T} from the left leads to

$$\boldsymbol{y} = \boldsymbol{l} - Y\boldsymbol{\lambda}\,. \tag{1.25}$$

Substituting this expression for $\boldsymbol{y}$ in (1.24) yields

$$\boldsymbol{0} = X\boldsymbol{x} + Y^{\mathsf{T}}(\boldsymbol{l} - Y\boldsymbol{\lambda}) + \boldsymbol{b}\,, \tag{1.26}$$

which allows to solve for $\boldsymbol{\lambda}$

$$\boldsymbol{\lambda} = (Y^{\mathsf{T}}Y)^{-1}(X\boldsymbol{x} + Y^{\mathsf{T}}\boldsymbol{l} + \boldsymbol{b})\,. \tag{1.27}$$

From (1.22) and (1.27) we obtain the normal equations for the estimates of the unknown parameters[5] $\boldsymbol{x}$

$$X(Y^{\mathsf{T}}Y)^{-1}X\widehat{\boldsymbol{x}} = -X(Y^{\mathsf{T}}Y)^{-1}(Y^{\mathsf{T}}\boldsymbol{l} + \boldsymbol{b})\,. \tag{1.28}$$

From (1.25) and (1.27) we finally obtain estimates $\widehat{\boldsymbol{y}}$ for the mean observations[6] $\boldsymbol{y}$,

$$\widehat{\boldsymbol{y}} = \boldsymbol{l} - Y(Y^{\mathsf{T}}Y)^{-1}(X\widehat{\boldsymbol{x}} + Y^{\mathsf{T}}\boldsymbol{l} + \boldsymbol{b})\,, \tag{1.29}$$

as a function of the estimated parameters $\widehat{\boldsymbol{x}}$ and the observations $\boldsymbol{l}$.

1.4.2 The Gauss–Helmert model for general covariance matrix

The Gauss–Helmert model with general covariance matrix reads as: for given observations $\boldsymbol{l}$, regular covariance matrix $\Sigma_{ll} = W_{ll}^{-1}$, and coefficient matrices X and Y

$$\boxed{\begin{array}{lll}\text{GHM}(\Sigma): & \text{minimize} & (\boldsymbol{y} - \boldsymbol{l})^{\mathsf{T}}\Sigma_{ll}^{-1}(\boldsymbol{y} - \boldsymbol{l}) \\ & \text{subject to} & X\boldsymbol{x} + Y^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{b} = \boldsymbol{0}\,.\end{array}} \tag{1.30}$$

[5] Observe, for given substitute observations $\boldsymbol{n} = Y^{\mathsf{T}}\boldsymbol{l} + \boldsymbol{b}$, this is the solution for the Gauss–Markov model minimizing $(X\boldsymbol{x} + \boldsymbol{n})\Sigma_{nn}^{-1}(X\boldsymbol{x} + \boldsymbol{n})$ w.r.t. the parameters $\boldsymbol{x}$

[6] Observe, for fixed $\widehat{\boldsymbol{x}}$, this is the solution of the problem with constraints for observations $\boldsymbol{l}$ only, minimizing $|X\widehat{\boldsymbol{x}} + Y^{\mathsf{T}}\boldsymbol{l} + \boldsymbol{b})|_2$ w.r.t. the observations $\boldsymbol{l}$, leading to fitted observations $\widehat{\boldsymbol{l}} = \widehat{\boldsymbol{y}}$.

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

$$\boldsymbol{n} = Y\boldsymbol{l} + \boldsymbol{b} \quad \text{and} \quad \mathbb{D}(\underline{\boldsymbol{n}}) = \Sigma_{nn} = Y^{\mathsf{T}}\Sigma_{ll}Y\,. \tag{1.31}$$

We obtain the estimated parameters and the fitted observations from

$$\boxed{\begin{array}{l}\widehat{\boldsymbol{x}} = \;\; -(X^{\mathsf{T}}\Sigma_{nn}^{-1}X)^{-1}\,X^{\mathsf{T}}\Sigma_{nn}^{-1}\,\boldsymbol{n}(\boldsymbol{l}) \\ \widehat{\boldsymbol{y}} = \boldsymbol{l} - \Sigma_{ll}Y\Sigma_{nn}^{-1}\,\boldsymbol{g}(\widehat{\boldsymbol{x}},\boldsymbol{l})\,.\end{array}} \tag{1.32}$$

Proof: We transfer this model to an unweighted Gauss–Helmert model. Especially, we eliminate the weights of the observations. For eliminating the weights, we use the substitutions

$$Y_g = \Sigma_{ll}^{1/2}Y \;\;,\quad \boldsymbol{l}_g = \Sigma_{ll}^{-1/2}\,\boldsymbol{l}\,,\quad \text{and} \quad \boldsymbol{y}_g = \Sigma_{ll}^{-1/2}\,\boldsymbol{y}\,. \tag{1.33}$$

Now, we need to solve the following unweighted Gauss–Helmert model: for given observations $\boldsymbol{l}_g$ and coefficient matrices X and Y_g,

$$\begin{array}{lll}\text{GHM(w):} & \text{minimize} & (\boldsymbol{y}_g - \boldsymbol{l}_g)^{\mathsf{T}}(\boldsymbol{y}_g - \boldsymbol{l}_g) \\ & \text{subject to} & X\boldsymbol{x} + Y_g^{\mathsf{T}}\boldsymbol{y}_g = \boldsymbol{0}\,,\end{array} \tag{1.34}$$

w.r.t. the unknown parameters $\boldsymbol{x}$ and the mean observations $\boldsymbol{y}_g$.

We thus obtain the normal equation system

$$X(Y_g^{\mathsf{T}}Y_g)^{-1}X\widehat{\boldsymbol{x}} = -X(Y_g^{\mathsf{T}}Y_g)^{-1}(Y_g^{\mathsf{T}}\boldsymbol{l}_g + \boldsymbol{b})\,. \tag{1.35}$$

or explicitly

$$X^{\mathsf{T}}(Y^{\mathsf{T}}\Sigma_{ll}Y)^{-1}X\;\widehat{\boldsymbol{x}} = -X^{\mathsf{T}}(Y^{\mathsf{T}}\Sigma_{ll}Y)^{-1}Y^{\mathsf{T}}\boldsymbol{l} \tag{1.36}$$

The fitted observations we obtain from

$$\widehat{\boldsymbol{y}}_g = \boldsymbol{l}_g - Y_g(Y_g^{\mathsf{T}}Y_g)^{-1}(X\widehat{\boldsymbol{x}} + Y_g^{\mathsf{T}}\boldsymbol{l}_g + \boldsymbol{b}))\,. \tag{1.37}$$

or finally

$$\widehat{\boldsymbol{y}} = \boldsymbol{l} - \Sigma_{ll}Y(Y^{\mathsf{T}}\Sigma_{ll}Y)^{-1}(X\widehat{\boldsymbol{x}} + Y^{\mathsf{T}}\boldsymbol{l} + \boldsymbol{b})\,. \tag{1.38}$$

1.4.3 Gauss–Markov model

The Gauss–Markov results from specializing the design matrix Y in the Gauss–Helmert model to

$$Y = -I\,, \tag{1.39}$$

leading to the constraint function

$$\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) = X\boldsymbol{x} - \boldsymbol{y} + \boldsymbol{b} \tag{1.40}$$

and substitute observations and their covariance matrix

$$\boldsymbol{n}(\boldsymbol{l}) = -\boldsymbol{l} + \boldsymbol{b} \quad \text{with} \quad \mathbb{D}(\underline{\boldsymbol{n}}) = \Sigma_{nn} = \Sigma_{ll}\,. \tag{1.41}$$

The Gauss–Markov model with covariance matrix $\Sigma_{ll} = W_{ll}^{-1}$ leads to the following general least squares optimization problem, for given observations $\boldsymbol{l}$, weight matrix $W_{ll} = \Sigma_{ll}^{-1}$, and coefficient matrix X

$$\boxed{\begin{array}{lll}\text{GMM}(\Sigma)\text{:} & \text{minimize} & (\boldsymbol{y} - \boldsymbol{l})^{\mathsf{T}}W_{ll}(\boldsymbol{y} - \boldsymbol{l}) \\ & \text{subject to} & \boldsymbol{y} = X\boldsymbol{x} + \boldsymbol{b}\,.\end{array}} \tag{1.42}$$

w.r.t. the unknown parameters $\boldsymbol{x}$ and mean observations $\boldsymbol{y}$. It yields the optimal parameters

$$\boxed{\begin{array}{l}\widehat{\boldsymbol{x}} = -(X^{\mathsf{T}}W_{ll}X)^{-1}X^{\mathsf{T}}W_{ll}\;\boldsymbol{n}(\boldsymbol{l}) \\ \widehat{\boldsymbol{y}} = \boldsymbol{l} - \boldsymbol{g}(\widehat{\boldsymbol{x}},\boldsymbol{l})\end{array}} \tag{1.43}$$

or explicitly in the classical form

$$\begin{array}{l}\widehat{\boldsymbol{x}} = (X^{\mathsf{T}}W_{ll}X)^{-1}X^{\mathsf{T}}W_{ll}(\boldsymbol{l} - \boldsymbol{b}) \\ \widehat{\boldsymbol{y}} = X\widehat{\boldsymbol{x}} + \boldsymbol{b}\,.\end{array} \tag{1.44}$$

1.4.4 Model with constraints between the observations only

The model with constraints between the observations only results from specializing the design matrix A in the Gauss–Helmert model to:

$$A = 0\,, \tag{1.45}$$

leading to the constraint function

$$\boldsymbol{g}(\boldsymbol{y}) = A^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{b}\,, \tag{1.46}$$

not depending on unknown parameters $\boldsymbol{x}$, and substitute observations and their covariance matrix

$$\boldsymbol{n}(\boldsymbol{l}) = A^{\mathsf{T}}\boldsymbol{l} + \boldsymbol{b} \quad \text{with} \quad \mathbb{D}(\underline{\boldsymbol{n}}) = \Sigma_{nn} = A^{\mathsf{T}}\Sigma_{ll}A^{\mathsf{T}}\,. \tag{1.47}$$

The model with constraints between the given observations $\boldsymbol{l}$ having covariance matrix Σ_{ll} leads to the following least squares problem

$$\boxed{\begin{array}{lll} \text{CONSTR}(\Sigma): & \text{minimize} & (\boldsymbol{y}-\boldsymbol{l})^{\mathsf{T}}\Sigma_{ll}^{-1}(\boldsymbol{y}-\boldsymbol{l}) \\ & \text{subject to} & A^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{b} = \boldsymbol{0}\,. \end{array}} \tag{1.48}$$

w.r.t. the mean observations $\boldsymbol{y}$. It yields the optimal estimates for the fitted observations

$$\boxed{\widehat{\boldsymbol{y}} = \boldsymbol{l} - \Sigma_{ll}A(A^{\mathsf{T}}\Sigma_{ll}A^{\mathsf{T}})^{-1}\,(A^{\mathsf{T}}\boldsymbol{l} + \boldsymbol{b})\,.} \tag{1.49}$$

1.4.5 The quasi Gauss–Markov model

As already indicated in the footnotes for (1.28) and (1.29) we can perform the estimation in the Gauss–Helmert model in two steps:

1. First we perform a Gauss–Markov model using the substitute observations

   $$\boldsymbol{n} = A^{\mathsf{T}} + \boldsymbol{b} \tag{1.50}$$

   hence

   $$\boldsymbol{n} = X\boldsymbol{x} \quad \text{with} \quad \mathbb{D}(\underline{\boldsymbol{n}}) = A^{\mathsf{T}}\Sigma_{ll}A\,, \tag{1.51}$$

   Using (1.43), this leads to the optimal estimates for the parameters $\boldsymbol{x}$ using the normal equations

   $$\widehat{\boldsymbol{x}} = (X^{\mathsf{T}}W_{nn}X)^{-1}\,X^{\mathsf{T}}W_{nn}\,\boldsymbol{n}\,. \tag{1.52}$$

2. Now, as we have the optimal estimates $\widehat{\boldsymbol{x}}$, we can treat them as fixed values. With the constant vector

   $$\boldsymbol{c}(\widehat{\boldsymbol{x}}) = X\widehat{\boldsymbol{x}} + \boldsymbol{b}\,, \tag{1.53}$$

   thus

   $$\boldsymbol{g}(\boldsymbol{l}) = A^{\mathsf{T}}\boldsymbol{l} + \boldsymbol{c}(\widehat{\boldsymbol{x}})\,, \tag{1.54}$$

   we can find the estimates for the fitted observations from the model for constraints between the observations only

   $$A^{\mathsf{T}}\boldsymbol{l} + \boldsymbol{c}(\widehat{\boldsymbol{x}}) = \boldsymbol{0} \quad \text{and} \quad \mathbb{D}(\underline{\boldsymbol{l}}) = \Sigma_{ll}\,. \tag{1.55}$$

   With (1.49), this leads to the estimates

   $$\widehat{\boldsymbol{y}} = \boldsymbol{l} - \Sigma_{ll}A(A^{\mathsf{T}}\Sigma_{ll}A)^{-1}\,\boldsymbol{g}(\boldsymbol{l}) \tag{1.56}$$

The Gauss–Markov model (1.51) is called the *quasi Gauss–Markov model* in the context of solving the parameters in the Gauss–Helmert model. In case the constraints are non-linear, the coefficient matrices are not fixed but need to be updated during the iteration process, which motivates the prefix *quasi*.

1.4.6 Results using pseudo inverses

The results can be written compactly using pseudo inverses. This is motivated from the least-squares solution of the simple Gauss–Markov model relating the mean observations to the unknown parameters via

$$\boldsymbol{y} = X\boldsymbol{x} \tag{1.57}$$

and minimizing $|\boldsymbol{y} - \boldsymbol{l}|_2$. This leads to the classical solution $\hat{\boldsymbol{x}} = (X^\mathsf{T}X)^{-1}X^\mathsf{T}\ \boldsymbol{l}$, which with the pseudo inverse

$$X^+ = (X^\mathsf{T}X)^{-1}X^\mathsf{T} \tag{1.58}$$

can be written as

$$\hat{\boldsymbol{x}} = X^+\boldsymbol{l} \tag{1.59}$$

This is an intuitive description of the inversion of (1.57), keeping in mind, that the inversion is not unique, since X is not regular, and regularization is enforced by the least squares principle.

Similarly, in case we minimize a weighted sum of squares $(\boldsymbol{y} - \boldsymbol{l})^\mathsf{T}W(\boldsymbol{y} - \boldsymbol{l})$ w.r.t. the parameters $\boldsymbol{x}$, with the weighted pseudo inverse

$$X_w^+ = (X^\mathsf{T}WX)^{-1}X^\mathsf{T}W \tag{1.60}$$

we obtain the solution

$$\hat{\boldsymbol{x}} = X_w^+\boldsymbol{l}\,. \tag{1.61}$$

We first define the properties of pseudo inverses and then provide the solutions of the different estimation problems.

1.4.6.1 Pseudo inverse and weighted pseudo inverse

For the regular $M \times N$ matrix A, with $M \geq N$ and $\mathrm{rk}(A) = N$ we use the pseudo inverse A^+:

$$A^+ := (A^\mathsf{T}A)^{-1}A^\mathsf{T} \tag{1.62}$$

It fulfils further the four relations:

$$AA^+A = A \quad A^+AA^+ = A^+ \quad (AA^+)^\mathsf{T} = AA^+ \quad A^+A = I\,. \tag{1.63}$$

Similarly, with the symmetric weight matrix U we use the weighted pseudo inverse (see Pepić (2010))

$$A_U^+ := (A^\mathsf{T}UA)^{-1}A^\mathsf{T}U \tag{1.64}$$

which fulfils the four relations

$$AA_U^+A = A \quad A_U^+AA_U^+ = A_U^+ \quad (UAA_U^+)^\mathsf{T} = UAA^+ \quad A_U^+A = I\,. \tag{1.65}$$

1.4.6.2 Solutions with pseudo inverses

We explicitly use the following inverses:

$$\begin{aligned}
X^+ &= (X^\mathsf{T}X)^{-1}\ X^\mathsf{T} && (1.66)\\
X_{\boldsymbol{w}_{ll}}^+ &= (X^\mathsf{T}W_{ll}X)^{-1}\ X^\mathsf{T}W_{ll} && (1.67)\\
X_{\boldsymbol{w}_{nn}}^+ &= (X^\mathsf{T}W_{nn}X)^{-1}\ X^\mathsf{T}W_{nn} && (1.68)\\
Y^+ &= (Y^\mathsf{T}Y)^{-1}\ Y^\mathsf{T} && (1.69)\\
Y_{\Sigma_{ll}}^+ &= (Y^\mathsf{T}\Sigma_{ll}Y)^{-1}\ Y^\mathsf{T}\Sigma_{ll} && (1.70)
\end{aligned}$$

Then we obtain the following solutions:

- Gauss–Markov model ($Y = -I$). Starting from the model

$$\boldsymbol{y} - \boldsymbol{b} = X\,\boldsymbol{x} \tag{1.71}$$

we obtain

$$\begin{aligned} \widehat{\boldsymbol{x}} &= -X^{+}_{\boldsymbol{w}_{ll}}\,\boldsymbol{n}(\boldsymbol{l}) \quad \text{and} \quad \widehat{\boldsymbol{y}} &= \boldsymbol{l} + \boldsymbol{g}(\widehat{\boldsymbol{x}}, \boldsymbol{l}) \qquad (1.72)\\ &= X^{+}_{\boldsymbol{w}_{ll}}\,(\boldsymbol{l} - \boldsymbol{b}) & = X\,\widehat{\boldsymbol{x}} + \boldsymbol{b}\,. \qquad (1.73) \end{aligned}$$

- Model with constraints between the observations only ($X = 0$). Starting from the model

$$Y^{\mathsf{T}}(\boldsymbol{y} - \boldsymbol{l}) + \boldsymbol{g}(\boldsymbol{l}) = \boldsymbol{0} \tag{1.74}$$

we arrive at the solution $\widehat{\boldsymbol{y}} - \boldsymbol{l} = -Y^{+\mathsf{T}}_{\Sigma_{ll}}\,\boldsymbol{g}(\boldsymbol{l})$, or

$$\widehat{\boldsymbol{y}} \;=\; \boldsymbol{l} - Y^{+\mathsf{T}}_{\Sigma_{ll}}\,\boldsymbol{g}(\boldsymbol{l}) \tag{1.75}$$

- Gauss–Helmert model. Starting from the model

$$X\boldsymbol{x} + \underbrace{Y^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{b}}_{\boldsymbol{n}(\boldsymbol{y})} = Y^{\mathsf{T}}(\boldsymbol{y} - \boldsymbol{l}) + \boldsymbol{g}(\boldsymbol{x}, \boldsymbol{l}) = \boldsymbol{0} \tag{1.76}$$

when first using $\boldsymbol{n}(\boldsymbol{l})$ as observations and then fixing the estimate for $\boldsymbol{x}$ we arrive at

$$\widehat{\boldsymbol{x}} = -X^{+}_{\boldsymbol{w}_{nn}}\,\boldsymbol{n}(\boldsymbol{l}) \qquad \text{and} \qquad \widehat{\boldsymbol{y}} \;=\; \boldsymbol{l} - Y^{+\mathsf{T}}_{\Sigma_{ll}}\,\boldsymbol{g}(\widehat{\boldsymbol{x}}, \boldsymbol{l}) \tag{1.77}$$

taking the covariance matrix Σ_{nn} of $\underline{\boldsymbol{n}}(\underline{\boldsymbol{l}})$ into account.

The solutions are collected in the following Table, starting with the Gauss–Helmert model with general covariance matrix and then showing the different specializations.

Table 1.1: Statistically optimal solutions in the linear model $(X, Y, \mathbb{D}(\underline{\boldsymbol{l}}))$ with its specializations: $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) = X\boldsymbol{x} + Y^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{b} = \boldsymbol{0}$ relating the mean $\boldsymbol{y} = \mathbb{E}(\underline{\boldsymbol{l}})$ of the observations $\underline{\boldsymbol{l}}$ to the unknown parameters $\boldsymbol{x}$ assuming a general covariance matrix and a unit matrix $\mathbb{D}(\underline{\boldsymbol{l}}) = \Sigma_{ll}$ and $\mathbb{D}(\underline{\boldsymbol{l}}) = I$, respectively. We use the substitute observations $\boldsymbol{n}(\boldsymbol{l}) = Y^{\mathsf{T}}\boldsymbol{l} + \boldsymbol{b}$ with their covariance matrix Σ_{nn}.
Rows 1 and 2: Gauss–Helmert model.
Rows 3 and 4: Gauss–Markov: $\boldsymbol{n}(\boldsymbol{l}) = -\boldsymbol{l} + \boldsymbol{b}$.
Rows 5 and 6: Model with constraints between the observations: $\boldsymbol{g}(\boldsymbol{y}) = Y^{\mathsf{T}}\boldsymbol{y} + \boldsymbol{b}$.

	model$(X, Y, \mathbb{D}(\underline{\boldsymbol{l}}))$	task		solution		
1	GHM(X, Y, Σ_{ll})	min.	$(\boldsymbol{y} - \boldsymbol{l})^{\mathsf{T}}\Sigma_{ll}^{-1}(\boldsymbol{y} - \boldsymbol{l})$	$\widehat{\boldsymbol{x}} = -$	$X^{+}_{\boldsymbol{w}_{nn}}$	$\boldsymbol{n}(\boldsymbol{l})$
		s.t.	$X\boldsymbol{x} + Y^{\mathsf{T}}\boldsymbol{y} = \boldsymbol{c}$	$\widehat{\boldsymbol{y}} = \boldsymbol{l} -$	$Y^{+\mathsf{T}}_{\Sigma_{ll}}$	$\boldsymbol{g}(\widehat{\boldsymbol{x}}, \boldsymbol{l})$
2	GHM(X, Y, I)	min.	$(\boldsymbol{y} - \boldsymbol{l})^{\mathsf{T}}(\boldsymbol{y} - \boldsymbol{l})$	$\widehat{\boldsymbol{x}} = -$	$X^{+}_{\boldsymbol{w}_{nn}}$	$\boldsymbol{n}(\boldsymbol{l})$
		s.t.	$X\boldsymbol{x} + Y^{\mathsf{T}}\boldsymbol{y} = \boldsymbol{c}$	$\widehat{\boldsymbol{y}} = \boldsymbol{l} -$	$Y^{+\mathsf{T}}$	$\boldsymbol{g}(\widehat{\boldsymbol{x}}, \boldsymbol{l})$
3	GMM$(X, -I, \Sigma_{ll})$	min.	$(\boldsymbol{y} - \boldsymbol{l})^{\mathsf{T}}\Sigma_{ll}^{-1}(\boldsymbol{y} - \boldsymbol{l})$	$\widehat{\boldsymbol{x}} = -$	$X^{+}_{\boldsymbol{w}_{ll}}$	$\boldsymbol{n}(\boldsymbol{l})$
		s.t.	$X\boldsymbol{x} - \boldsymbol{y} = \boldsymbol{c}$	$\widehat{\boldsymbol{y}} = \boldsymbol{l} -$		$\boldsymbol{g}(\widehat{\boldsymbol{x}}, \boldsymbol{l})$ [1)]
4	GMM$(X, -I, I)$	min.	$(\boldsymbol{y} - \boldsymbol{l})^{\mathsf{T}}(\boldsymbol{y} - \boldsymbol{l})$	$\widehat{\boldsymbol{x}} = -$	X^{+}	$\boldsymbol{n}(\boldsymbol{l})$
		s.t.	$X\boldsymbol{x} - \boldsymbol{y} = \boldsymbol{c}$	$\widehat{\boldsymbol{y}} = \boldsymbol{l} -$		$\boldsymbol{g}(\widehat{\boldsymbol{x}}, \boldsymbol{l})$ [1)]
5	CONSTR$(0, Y, \Sigma_{ll})$	min.	$(\boldsymbol{y} - \boldsymbol{l})^{\mathsf{T}}\Sigma_{ll}^{-1}(\boldsymbol{y} - \boldsymbol{l})$			
		s.t.	$Y^{\mathsf{T}}\boldsymbol{y} = \boldsymbol{c}$	$\widehat{\boldsymbol{y}} = \boldsymbol{l} -$	$Y^{+\mathsf{T}}_{\Sigma_{ll}}$	$\boldsymbol{g}(\boldsymbol{l})$
6	CONSTR$(0, Y, I)$	min.	$(\boldsymbol{y} - \boldsymbol{l})^{\mathsf{T}}(\boldsymbol{y} - \boldsymbol{l})$			
		s.t.	$Y^{\mathsf{T}}\boldsymbol{y} = \boldsymbol{c}$	$\widehat{\boldsymbol{y}} = \boldsymbol{l} -$	$Y^{+\mathsf{T}}$	$\boldsymbol{g}(\boldsymbol{l})$

[1)] This is equivalent to $\widehat{\boldsymbol{y}} = X\widehat{\boldsymbol{x}} + \boldsymbol{b}$

This closes the section on the estimation in the linear Gauss–Helmert model. We did not construct the solutions, but just proved they are correct. The generalization to non-linear constraints will also use the reduction to a Gauss–Markov model, but derive the iterative solution explicitly. Moreover, coefficient matrices A and B then depend on the current estimates of the parameters and the observations thus need to be updated in each iteration.

1.5 The non-linear Gauss–Helmert model

The functional model generally is non-linear. We assume we have approximate values $\boldsymbol{x}^a$ and $\boldsymbol{y}^a$ for the parameters $\boldsymbol{x}$ and the mean observations $\boldsymbol{y}$ and updates

$$\boldsymbol{x} := u_x(\boldsymbol{x}^a, \Delta\boldsymbol{x})) \quad \text{e.g.,} \quad \boldsymbol{x}^a := \boldsymbol{x}^a + \Delta\boldsymbol{x}\,. \tag{1.78}$$

and

$$\boldsymbol{y} := u_y(\boldsymbol{y}^a, \Delta\boldsymbol{y}) \quad \text{e.g.,} \quad \boldsymbol{y}^a := \boldsymbol{y}^a + \Delta\boldsymbol{y}\,. \tag{1.79}$$

These relations hold for small corrections $\Delta\boldsymbol{x}$ and $\Delta\boldsymbol{y}$. Given values for $\boldsymbol{x}$ and its approximations $\boldsymbol{x}^a$ we assume we can determine the corrections from

$$\Delta\boldsymbol{x} = u_x^{-1}(\boldsymbol{x}, \boldsymbol{x}^a) \quad \text{e.g.,} \quad \Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^a \tag{1.80}$$

Similarly, we assume there exist inverse functions for the mean observations

$$\Delta\boldsymbol{y} = u_y^{-1}(\boldsymbol{y}, \boldsymbol{y}^a) \quad \text{e.g.,} \quad \Delta\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{y}^a \tag{1.81}$$

Hence we have the update function with $m \geq n$, for small $\Delta\boldsymbol{x}$, especially for $m = n$

$$u_x : \mathbb{R}^n \mapsto \mathbb{R}^m \quad \Delta\boldsymbol{x} \mapsto \boldsymbol{x} = u_x(\Delta\boldsymbol{x}; \boldsymbol{x}^a) \quad \text{especially} \quad \boldsymbol{x} = \Delta\boldsymbol{x} + \boldsymbol{x}^a \tag{1.82}$$

$$u_x^{-1} : \mathbb{R}^m \mapsto \mathbb{R}^n \quad \boldsymbol{x} \mapsto \Delta\boldsymbol{x} = u_x^{-1}(\boldsymbol{x}; \boldsymbol{x}^a) \quad \text{especially} \quad \Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^a\,, \tag{1.83}$$

and similarly, for u_y.

Example: Non-linear update and its inversion for 3D rotations. Let the unknown parameters be a 3×3 rotation matrix R. We actually estimate a small 3-vector $\Delta\boldsymbol{r}$ of small rotation angles. The approximate rotation matrix R^a the can be corrected using

$$R = u_x(R^a, \Delta\boldsymbol{r}) = R(\Delta\boldsymbol{r})\, R^a\,. \tag{1.84}$$

where $R(\Delta\boldsymbol{r})$ is a rotation matrix depending on the 3-vector $\Delta\boldsymbol{r}$, e.g., using the exponential or the Cayley form

$$R(\Delta\boldsymbol{r}) = \exp(S(\Delta\boldsymbol{r})) \quad \text{or} \quad R(\Delta\boldsymbol{r}) = (I + S(\Delta\boldsymbol{r}/2))(I - S(\Delta\boldsymbol{r}/2))^{-1} \tag{1.85}$$

with the skew symmetric matrix $S(\boldsymbol{a})$ inducing the cross product $\boldsymbol{a} \times \boldsymbol{b} = S(\boldsymbol{a})\boldsymbol{b}$. In case we have given R and some approximation R^a, we may determine the correction vector $\Delta\boldsymbol{r}$ from

$$S(\Delta\boldsymbol{r}) = \log(R^\mathsf{T} R^a) \approx R^\mathsf{T} R^a - I\,, \tag{1.86}$$

thus taking the off diagonal terms of the product $R^\mathsf{T} R^a$ of the two rotation matrices as the sought 3-vector. This can compactly be written as

$$\Delta\boldsymbol{r} = u_x^{-1}(R, R^a) = s(R^\mathsf{T} R^a)\,. \tag{1.87}$$

where the function

$$s(A) = \frac{1}{2}\begin{bmatrix} A_{32} - A_{23} \\ A_{13} - A_{31} \\ A_{21} - A_{21} \end{bmatrix} \tag{1.88}$$

extracts the skew vector of the 3×3 rotation matrix A. ◇

Similarly, we have the updates and their inversion starting from $\boldsymbol{l}$, first for the approximations of the mean observations

$$\boldsymbol{y}^a = u_y(\boldsymbol{l}, \boldsymbol{v}^a) \quad \text{and} \quad \boldsymbol{v}^a = u_y^{-1}(\boldsymbol{y}^a, \boldsymbol{l}) = -u_y^{-1}(\boldsymbol{l}, \boldsymbol{y}^a)\,. \tag{1.89}$$

which for small residuals can be defined in either manner. Thus we have for the mean observations

$$\boldsymbol{y} = u_y(\boldsymbol{l}, \boldsymbol{v}) \quad \text{and} \quad \boldsymbol{v} = u_y^{-1}(\boldsymbol{y}, \boldsymbol{l}) = -u_y^{-1}(\boldsymbol{l}, \boldsymbol{y})\,. \tag{1.90}$$

For small values we have

$$\boldsymbol{v} = \boldsymbol{v}^a + \Delta\boldsymbol{y}\,, \tag{1.91}$$

see Fig. 1.1. Since the observations $\boldsymbol{l}$ and the residuals $\boldsymbol{v}$ may have a different structure, e.g., if the observations are rotation matrices and the residuals are rotation vectors, the covariance matrix Σ_{ll} refers to the residuals of the observations

Covariance matrix for rotation matrices. In the case of an observed rotation matrix R, we represent the uncertain rotation as

$$\underline{R} = R(\underline{\boldsymbol{r}})\,\mathbb{E}(\underline{R}) \quad \text{with} \quad \mathbb{D}(\underline{\boldsymbol{r}}) = \Sigma_{rr} \tag{1.92}$$

If R is observed, the we refer to the 3×3 matrix Σ_{rr} as the covariance matrix Σ_{ll} of the observed rotation. ◇

We are now prepared to derive a linear substitute problem used for iteratively determining the unknowns $\boldsymbol{y}$ and $\boldsymbol{x}$.

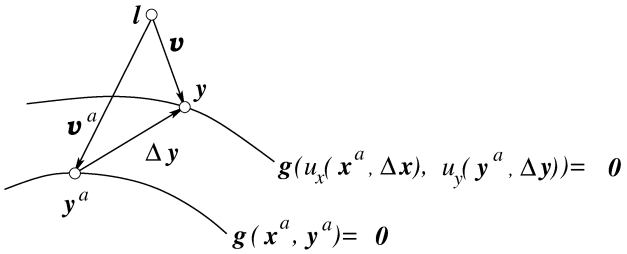

Figure 1.1: Update of the unknowns and the mean observations in the Gauss–Helmert model. The corrections $\Delta\boldsymbol{x} = u_x^{-1}(\boldsymbol{x}, \boldsymbol{x}^a)$ to the parameters and the corrections $\Delta\boldsymbol{y} = u_y^{-1}(\boldsymbol{y}, \boldsymbol{y}^a) = \boldsymbol{v} - \boldsymbol{v}^a$ to the mean observations and residuals are meant to converge to zero. The figure assumes the dimensions of the observations/parameters $(\boldsymbol{l}, \boldsymbol{y}, \boldsymbol{y}^a)$ and the dimensions of their residuals/corrections $(\boldsymbol{v}, \boldsymbol{v}^a, \boldsymbol{y}^a)$ are the same

corrections to mean observations and parameters

1. We define the corrections to the parameters and the mean observations

$$\Delta\boldsymbol{x} = u_x^{-1}(\boldsymbol{x}, \boldsymbol{x}^a) = \boldsymbol{x}^a - \boldsymbol{x} \quad \text{and} \quad \Delta\boldsymbol{y} = u_y^{-1}(\boldsymbol{y}, \boldsymbol{y}^a) = \boldsymbol{v} - \boldsymbol{v}^a\,, \tag{1.93}$$

   in order to iteratively improve the approximations $\boldsymbol{x}^a$ and $\boldsymbol{y}^a$ such that after convergence $\Delta\boldsymbol{x} = \boldsymbol{0}$ and $\Delta\boldsymbol{y} = \boldsymbol{0}$. Observe, that the approximate residuals $\boldsymbol{v}^a = \boldsymbol{y}^a - \boldsymbol{l}$ also are to be corrected by $\Delta\boldsymbol{y}$.

2. The optimization function then reads as[7]

$$\boxed{\Omega = (\boldsymbol{v}^a + \Delta\boldsymbol{y})^\mathsf{T}\,\Sigma_{ll}^{-1}\,(\boldsymbol{v}^a + \Delta\boldsymbol{y})\,.} \tag{1.94}$$

   where the covariance matrix Σ_{ll} refers to the corrections $\boldsymbol{v}$ of the observations.

   Remark: Observe, the optimization function (1.94) of this non-linear model results from the one (1.30) of the linear model using $\boldsymbol{y} - \boldsymbol{l} = \boldsymbol{y}^a + \Delta\boldsymbol{y} - \boldsymbol{l} = \boldsymbol{v}^a + \Delta\boldsymbol{y}$. ◇

nearized
nstraints

3. Linearization of the nonlinear implicit function (1.12) leads to the constraints

$$\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{g}(u_x(\boldsymbol{x}^a,\Delta\boldsymbol{x}),u_y(\boldsymbol{y}^a,\Delta\boldsymbol{y})) = \boldsymbol{g}(\boldsymbol{x}^a,\boldsymbol{y}^a) + \mathsf{X}\Delta\boldsymbol{x} + \mathsf{Y}^\mathsf{T}\Delta\boldsymbol{y} = \boldsymbol{0} \tag{1.95}$$

with the Jacobians

$$\boxed{\underset{G\times U}{\mathsf{X}} = \left.\frac{\partial \boldsymbol{g}}{\partial \Delta\boldsymbol{x}}\right|_{x=x^a,y=y^a} \quad \text{and} \quad \underset{G\times N}{\mathsf{Y}^\mathsf{T}} = \left.\frac{\partial \boldsymbol{g}}{\partial \Delta\boldsymbol{y}}\right|_{x=x^a,y=y^a},} \tag{1.96}$$

to be evaluated at the approximations of the mean observations and of the parameters.

Remark: Also the structure of the constraints of the linear Gauss–Helmert model is preserved, when replacing the unknowns $\boldsymbol{x}$ and $\boldsymbol{y}$ by their corrections $\Delta\boldsymbol{x}$ and $\Delta\boldsymbol{y}$ and the constant $\boldsymbol{b}$ by $\boldsymbol{g}(\boldsymbol{x}^a,\boldsymbol{y}^a)$. ◇

Therefore the *linear substitute problem* for determining the corrections $\Delta\boldsymbol{x}$ and $\Delta\boldsymbol{y}$ is: *linear substitute problem*

$$\text{minimize} \quad (\boldsymbol{v}^a + \Delta\boldsymbol{y})^\mathsf{T}\,\Sigma_{ll}^{-1}\,(\boldsymbol{v}^a + \Delta\boldsymbol{y}) \tag{1.97}$$

$$\text{subject to} \quad \boldsymbol{g}(\boldsymbol{x}^a,\boldsymbol{y}^a) + \mathsf{X}\Delta\boldsymbol{x} + \mathsf{Y}^\mathsf{T}\Delta\boldsymbol{y} = \boldsymbol{0}\,, \tag{1.98}$$

for given approximate values $\boldsymbol{y}^a$ and thus $\boldsymbol{v}^a = u_y^{-1}(\boldsymbol{y}^a,\boldsymbol{l})$, function $\boldsymbol{g}$, Jacobians X and Y, and covariance matrix Σ_{ll}.

We first will provide the algorithm and then its derivation.

1.5.1 The algorithm for estimating the parameters

We start from the observations $\{\boldsymbol{l},\Sigma_{ll}\}$, the implicit functions $\boldsymbol{g}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{0}$, and the approximate values $\boldsymbol{x}^a$ for the unknowns and $\boldsymbol{y}^a$ for the mean observations, which are initiated with $\boldsymbol{y}^a := \boldsymbol{l}$. We obtain the following algorithm for an iterative solution:

1. Iterate until convergence
   (a) Determine the Jacobians X and Y (1.96) at the current approximate values $(\boldsymbol{x}^a,\boldsymbol{y}^a)$. *Jacobians at current approximations*
   (b) Determine the contradictions $\boldsymbol{c}_g$ of the negative constraints at the appproximate values $\boldsymbol{x}^a$ of the unknown parameters together with their weight matrix [8] [9]

$$\boldsymbol{c}_g = -\boldsymbol{g}(\boldsymbol{x}^a,\boldsymbol{l}) \quad \text{and} \quad \mathsf{W}_{gg} = (\mathsf{Y}^\mathsf{T}\Sigma_{ll}\mathsf{Y})^{-1}\,. \tag{1.99}$$

contradictions of constraints given the parameters

   (c) Solve the normal equation system for the corrections $\Delta\boldsymbol{x}$ of the parameters

$$\mathsf{N}\Delta\boldsymbol{x} = \boldsymbol{m} \quad \text{with} \quad \mathsf{N} = \mathsf{X}^\mathsf{T}\mathsf{W}_{gg}\mathsf{X} \quad \text{and} \quad \boldsymbol{m} = \mathsf{X}^\mathsf{T}\mathsf{W}_{gg}\,\boldsymbol{c}_g\,. \tag{1.100}$$

normal equation system

   (d) Update the approximate parameters

$$\boldsymbol{x}^a := u_x(\boldsymbol{x}^a,\Delta\boldsymbol{x}) \quad \text{e.g.,} \quad \boldsymbol{x}^a := \boldsymbol{x}^a + \Delta\boldsymbol{x}\,. \tag{1.101}$$

   hence

$$-\boldsymbol{g}(\boldsymbol{x}^a,\boldsymbol{l}) := \boldsymbol{c}_g - \mathsf{X}\Delta\boldsymbol{x} \tag{1.102}$$

   (e) Determine the corrections for the mean observations

$$\Delta\boldsymbol{y} = \Sigma_{ll}\mathsf{Y}\mathsf{W}_{gg}(\boldsymbol{c}_g - \mathsf{X}\Delta\boldsymbol{x}) - \boldsymbol{v}^a\,. \tag{1.103}$$

   (f) Update the approximate mean observations *update of approximate mean observations*

$$\boldsymbol{y}^a := u_y(\boldsymbol{y}^a,\Delta\boldsymbol{y}) \quad \text{e.g.,} \quad \boldsymbol{y}^a := \boldsymbol{y}^a + \Delta\boldsymbol{y}\,. \tag{1.104}$$

2. Set the final estimates of the unknown parameters and of the mean observations, sometimes called the fitted observation $\widehat{\boldsymbol{l}} := \widehat{\boldsymbol{y}}$ *final estimates*

$$\widehat{\boldsymbol{x}} := \boldsymbol{x}^a \quad \text{and} \quad \widehat{\boldsymbol{y}} = \boldsymbol{y}^a \,. \tag{1.105}$$

covariance matrix of the estimated parameters

3. Determine the covariance matrix of the estimated parameters

$$\Sigma_{\hat{x}\hat{x}} = N^{-1} \,. \tag{1.106}$$

4. If we only know an approximate covariance matrix Σ_{ll}^a and we assume the covariance matrix Σ_{ll} differs from the approximation by an unknown variance factor σ_0^2

$$\Sigma_{ll} = \sigma_0^2 \Sigma_{ll}^a \quad \text{with} \quad W_{ll} = \Sigma_{ll} \,, \tag{1.107}$$

then we can perform the estimation with Σ_{ll}^a, instead of using Σ_{ll}, which has no effect onto the estimates. But then we can find an estimate

$$\widehat{\sigma}_0^2 = \frac{\boldsymbol{c}_l^\mathsf{T} W_{ll}^a \boldsymbol{c}_l}{G-U} \quad \text{or} \quad \widehat{\sigma}_0^2 = \frac{\boldsymbol{c}_g^\mathsf{T} W_{c_g c_g}^a \boldsymbol{c}_g}{G-U} \,. \tag{1.108}$$

estimated variance factor

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

$$\widehat{\Sigma}_{\hat{x}\hat{x}} = \widehat{\sigma}_0^2 \Sigma_{\hat{x}\hat{x}}^a \quad \text{with} \quad \Sigma_{\hat{x}\hat{x}}^a = (X^\mathsf{T}(Y^\mathsf{T}\Sigma_{ll}^a Y)^{-1}X)^{-1} \,. \tag{1.109}$$

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

Remark: If the observational noise is small and an approximate solution is acceptable, the steps 1.(e–f) can be omitted. Then the Jacobians X and Y are to be determined at $(\boldsymbol{x}^a, \boldsymbol{l})$ instead of at $(\boldsymbol{x}^a, \boldsymbol{y}^a)$. ◇

The complete procedure is given in the algorithm below. The green parts refer to the case, where the degrees of freedom of the parameters and observations is less than the number of elements of their representation.

1.5.2 Derivation of the procedure

We now derive the procedure.

1.5.2.1 Estimating the parameters with a quasi Gauss–Markov model

We start from the constraint (1.95) rewritten as

$$-\boldsymbol{g}(\boldsymbol{x}^a, \boldsymbol{y}^a) - Y^\mathsf{T}\Delta\boldsymbol{y} = X\Delta\boldsymbol{x} \,. \tag{1.110}$$

In order to eliminate the dependeny of $\Delta\boldsymbol{y}$, we introduce the *contradiction of the constraints*, i.e., the value

$$\boldsymbol{c}_g = -\boldsymbol{g}(\boldsymbol{x}^a, \boldsymbol{l}) \tag{1.111}$$

choosing the negative sign for making the following equations more intuitive. With $\boldsymbol{g}(\boldsymbol{x}^a, \boldsymbol{y}^a) = \boldsymbol{g}(\boldsymbol{x}^a, \boldsymbol{l}) + Y^\mathsf{T}\boldsymbol{v}^a$ and $\boldsymbol{v} = \boldsymbol{v}^a + \Delta\boldsymbol{y}$ we have, up to first order, $\boldsymbol{g}(\boldsymbol{x}^a, \boldsymbol{y}^a) + Y^\mathsf{T}\Delta\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}^a, \boldsymbol{l}) + Y^\mathsf{T}\boldsymbol{v}$ and therefore we can rewrite (1.110) as

$$\boldsymbol{c}_g - Y^\mathsf{T}\boldsymbol{v} = X\Delta\boldsymbol{x} \,. \tag{1.112}$$

[7] Observe, we do not have the estimated residuals $\boldsymbol{v}$ in the optimization function, but their corrections $\boldsymbol{v} - \boldsymbol{v}^a = \Delta\boldsymbol{y}$ (1.93), in order to be able to handle observations, such as directions or rotations, where a non-linear update of the observations is more appropriate, replacing (1.93), see (7.19).

[8] Again we do not indicate, that $\boldsymbol{c}_g$ depends on approximate values thus omit a superscript ^{a}.

[9] If we have the special linear Gauss–Markov model $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) = X\boldsymbol{x} - \boldsymbol{y} = \boldsymbol{0}$, thus $Y = -I$, and use the approximate values $\boldsymbol{x}^a = \boldsymbol{0}$, then we have $\boldsymbol{c}_g = -\boldsymbol{g}(\boldsymbol{x}^a, \boldsymbol{l}) = \boldsymbol{l}$.

Algorithm 1: Estimation in the Gauss–Helmert model.
$[\hat{\boldsymbol{x}}, \Sigma_{\hat{x}\hat{x}}, \hat{\sigma}_0^2, R]$ = `GaussHelmertModell_D`$(\boldsymbol{l}, \Sigma_{ll}, c_g, u_x, u_y, \boldsymbol{x}^a, \boldsymbol{\sigma}_{\hat{x}}^a, T_\theta,$ `maxiter`$)$
Input: observed values $\{\boldsymbol{l}, \Sigma_{ll}\}$, number N,
constraint function $[\boldsymbol{c}_g, X, Y] = c_g(\boldsymbol{l}, \boldsymbol{y}^a, \boldsymbol{x}^a)$, number G,
update functions u_x and u_y for the parameters and mean observations.
approximate values $\boldsymbol{x}^a$, possibly $\sigma_{\hat{\theta}_u}^a$,
parameters T_θ, `maxiter` for controlling convergence.
Output: estimated parameters $\{\hat{\boldsymbol{x}}, \Sigma_{\hat{x}\hat{x}}\}$, variance factor $\hat{\sigma}_0^2$, redundancy R.

Redundancy $R = G - U$;
if $R < 0$ **then stop**, not enough constraints;
Initiate: iteration $\nu = 0$, approximate values $\boldsymbol{y}^a := \boldsymbol{l}$, stopping variable: $s = 0$;
repeat
  Constraints and Jacobians : $[\boldsymbol{c}_g, X, Y] = c_g(\boldsymbol{l}, \boldsymbol{y}^a, \boldsymbol{x}^a)$, see (7.14), (1.96);
  Weight matrix of constraints: $W_{gg} = (Y^\mathsf{T}\Sigma_{ll}Y)^{-1}$;
  Build normal equation system: $[N, \boldsymbol{m}]$, see (7.16);
  if N is singular **then stop**: normal equation matrix is singular;
  Updates of parameter vector: $\Delta\boldsymbol{x}$, see (1.78), $\boldsymbol{x}^a := u_x(\boldsymbol{x}^a, \Delta\boldsymbol{x})$;
  Corrections for fitted observations: $\Delta\boldsymbol{y}$, see (7.18);
  Update fitted observations: $\boldsymbol{y}^a = u_y(\boldsymbol{y}^a, \Delta\boldsymbol{y})$, see (7.19);
  Set iteration: $\nu := \nu + 1$;
  if $\max_u(|\widehat{\Delta x}_u|/\sigma_{\hat{x}_u}^a) < T_x$ or $\nu =$ `maxiter` **then** $s = 2$;
until $s \equiv 2$;
Estimated parameters $\hat{\boldsymbol{x}} := \hat{\boldsymbol{x}}^a$ and covariance matrix: $\Sigma_{\hat{x}\hat{x}}$, see (7.22);
if $R > 0$ **then** variance factor $\hat{\sigma}_0^2 = \boldsymbol{c}_g^\mathsf{T} W_{gg}\, \boldsymbol{c}_g / R$;
else $\hat{\sigma}_0^2 = 1$;

Now, we define the substitution

$$\boldsymbol{v}_g = -Y^\mathsf{T}\boldsymbol{v}\,. \tag{1.113}$$

This is that part of the residuals $\boldsymbol{v}$ of the observations $\boldsymbol{l}$, which is relevant for the constraints. Its uncertainty results from (1.112), since $\Delta\boldsymbol{x}$ is assumed to be fixed in this step,

$$\mathbb{D}(\underline{\boldsymbol{c}}_g) = \mathbb{D}(-\boldsymbol{g}(\boldsymbol{x}^a, \underline{\boldsymbol{l}})) = \Sigma_{gg} = Y^\mathsf{T}\Sigma_{ll}Y\,. \tag{1.114}$$

We thus arrive at a representation of the functional model which has the algebraic structure of a Gauss-Markov model with $\boldsymbol{c}_g$ as observations and $\Delta\boldsymbol{x}$ as unknowns

quasi Gauss-Markov model representing the Gauss–Helmert model

$$\boxed{\boldsymbol{c}_g + \boldsymbol{v}_g = X\Delta\boldsymbol{x} \quad \text{with} \quad \mathbb{D}(\underline{\boldsymbol{c}}_g) = \Sigma_{gg}\,.} \tag{1.115}$$

Starting from here we solve the optimization problem for determining the corrections $\Delta\boldsymbol{x}$

$$\text{minimize} \qquad \boldsymbol{v}_g\, \Sigma_{c_g c_g}^{-1}\, \boldsymbol{v}_g \tag{1.116}$$
$$\text{subject to} \qquad -(\boldsymbol{c}_g + \boldsymbol{v}_g) + X\Delta\boldsymbol{x} = \boldsymbol{0}\,, \tag{1.117}$$

for given the contradictions $\boldsymbol{c}_g$ of the constraints, the Jacobian X, and the covariance matrix $\Sigma_{c_g c_g}$. As we know from the estimation with the Gauss–Markov model, we obtain the normal equation system

$$\boxed{N\Delta\boldsymbol{x} = \boldsymbol{m} \quad \text{with} \quad N = X^\mathsf{T}(Y^\mathsf{T}\Sigma_{ll}Y)^{-1}X \quad \text{and} \quad \boldsymbol{m} = X^\mathsf{T}(Y^\mathsf{T}\Sigma_{ll}Y)^{-1}\,\boldsymbol{c}_g\,.} \tag{1.118}$$

Hence the updated parameters are

$$\boldsymbol{x}^a := u_x(\boldsymbol{x}^a, \Delta\boldsymbol{x})\,. \tag{1.119}$$

We, however, need to be aware of the following: both, the coefficient matrix X and – via the Jacobian Y – the covariance matrix Σ_{gg} in (1.114) generally depend on the current values $\boldsymbol{x}^a$ and $\boldsymbol{y}^a$, since the Jacobians have to be determined at these values, see (1.96). So, we need to determine updates $\Delta\boldsymbol{y}$ for the mean observations $\boldsymbol{y}$ within the iterative scheme.

Since its Jacobian and covariance matrix depend on the unknown parameters we call thus functional the *quasi Gauss–Markov model* replacing the implicit constraints in the Gauss–Helmert model.[10]

1.5.2.2 Update of approximate fitted observations

From (1.115) and (1.113) we have the residuals $\boldsymbol{v}_g$ at some point within the iteration scheme

$$\boldsymbol{v}_g = -\boldsymbol{c}_g + X\Delta\boldsymbol{x} = -Y^\mathsf{T}\boldsymbol{v}\,, \tag{1.120}$$

which result after finding the locally best corrections $\Delta\boldsymbol{x}$. If we could determine the residuals $\boldsymbol{v}$ of the original observations from the residuals $\boldsymbol{v}_g$, i.e., invert the relation $\boldsymbol{v}_g = -Y^\mathsf{T}\boldsymbol{v}$, we could derive the corrections

$$\Delta\boldsymbol{y} = -\boldsymbol{v}^a + \boldsymbol{v}\,. \tag{1.121}$$

We could use them to determine updates for the mean observations $\boldsymbol{y}$. We obviously cannot determine the residuals $\boldsymbol{v}$ of the original observations by inversion of (1.113), since the matrix Y in generally does not have full rank.

Therefore we determine those residuals $\boldsymbol{v}$ which fulfil the constraint (1.120) and minimize $\Omega = \boldsymbol{v}^\mathsf{T}\Sigma_{ll}^{-1}\boldsymbol{v}$. With the Lagrangian parameter vector $\boldsymbol{\lambda}$ we thus need to

$$\boxed{\begin{array}{lll}\text{CONSTR}(\Sigma): & \text{minimize} & \boldsymbol{v}^\mathsf{T}\Sigma_{ll}^{-1}\boldsymbol{v} \\ & \text{subject to} & Y^\mathsf{T}\boldsymbol{v} + \boldsymbol{v}_g = \boldsymbol{0}\,.\end{array}} \tag{1.122}$$

w.r.t. the residuals $\boldsymbol{v}$. Setting the partials of

$$\boldsymbol{\Phi}(\boldsymbol{v},\boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{v}^\mathsf{T}\Sigma_{ll}^{-1}\boldsymbol{v} + \boldsymbol{\lambda}^\mathsf{T}(Y^\mathsf{T}\boldsymbol{v} + \boldsymbol{v}_g) \tag{1.123}$$

to $\boldsymbol{0}$ yields the two necessary equations for $\boldsymbol{v}$

$$\frac{\partial\boldsymbol{\Phi}}{\partial\boldsymbol{v}^\mathsf{T}} = \Sigma_{ll}^{-1}\boldsymbol{v} + Y\boldsymbol{\lambda} = \boldsymbol{0} \quad \text{and} \quad \frac{\partial\boldsymbol{\Phi}}{\partial\boldsymbol{\lambda}^\mathsf{T}} = Y^\mathsf{T}\boldsymbol{v} + \boldsymbol{v}_g = \boldsymbol{0}\,. \tag{1.124}$$

From the first equation we obtain

$$\boldsymbol{v} = -\Sigma_{ll}Y\boldsymbol{\lambda} \tag{1.125}$$

which from the second equation leads to

$$\widehat{\boldsymbol{v}}_g - Y^\mathsf{T}\Sigma_{ll}Y\boldsymbol{\lambda} = \boldsymbol{0} \tag{1.126}$$

Therefore we have

$$\boldsymbol{\lambda} = (Y^\mathsf{T}\Sigma_{ll}Y)^{-1}\boldsymbol{v}_g\,, \tag{1.127}$$

which finally yields

$$\widehat{\boldsymbol{v}} = -\Sigma_{ll}Y(Y^\mathsf{T}\Sigma_{ll}Y)^{-1}\widehat{\boldsymbol{v}}_g \tag{1.128}$$

inverting the substitution in (1.113) in an intuitive manner. Hence, from (1.121) and (1.90) we obtain the corrections

$$\Delta\boldsymbol{y} = -u_y^{-1}(\boldsymbol{y}^a,\boldsymbol{l}) - \Sigma_{ll}Y(Y^\mathsf{T}\Sigma_{ll}Y)^{-1}\,\boldsymbol{g}(\boldsymbol{x}^a,\boldsymbol{l})\,. \tag{1.129}$$

[10] In the German geodetic literature on *adjustment theory* (equivalent to the estimation theory) the functional model (1.115) is called *'quasi vermittelnde Ausgleichung'*, derived from the German *'vermittelnde Ausgleichung'* representing the Gauss–Markov model. This motivates the English naming of this functional model, which only occurs as substitute for the linearized Gauss–Helmert model

If the residuals/corrections and the parameters/observations have the same dimension, e.g., for classical point coordinates, this simplifies to

$$\Delta \boldsymbol{y} = -\boldsymbol{y}^a + \boldsymbol{l} - \Sigma_{ll} Y (Y^\mathsf{T} \Sigma_{ll} Y)^{-1}\, \boldsymbol{g}(\boldsymbol{x}^a, \boldsymbol{l}))\,. \tag{1.130}$$

The update for the estimates of the mean observations then read as

$$\boldsymbol{y}^a := u_y(\boldsymbol{y}^a, \Delta \boldsymbol{y}) \quad \text{especially} \quad \widehat{\boldsymbol{y}}^a := \widehat{\boldsymbol{y}}^a + \Delta \boldsymbol{y}\,. \tag{1.131}$$

If the observations and the constraints, $\boldsymbol{l}$ and $\boldsymbol{g}$ are grouped as $\boldsymbol{l} = [\boldsymbol{l}_i]$ and $\boldsymbol{g} = [\boldsymbol{g}_i]$, such that each group $\boldsymbol{g}_i$ only refers to the corresponding group $\boldsymbol{l}_i$ and the observational groups are mutually independent, i.e., for $i \neq j$ we have $\mathbb{C}\text{ov}(\underline{\boldsymbol{l}}_i, \underline{\boldsymbol{l}}_j) = \boldsymbol{0}$, then with $\boldsymbol{v}_i^a$ the updates can be done group wise:

correction of estimated observations

$$\boxed{\Delta \boldsymbol{y}_i = u_y^{-1}(\boldsymbol{l}_i, \boldsymbol{y}_i^a) - \Sigma_{l_i l_i} Y_i (Y_i^\mathsf{T} \Sigma_{l_i l_i} Y_i)^{-1}\, \boldsymbol{g}(\widehat{\boldsymbol{x}}, \boldsymbol{l}_i)\,,} \tag{1.132}$$

with the individual updates

$$\widehat{\boldsymbol{y}}_i^a := u_y(\widehat{\boldsymbol{y}}_i^a, \Delta \boldsymbol{y}_i)\,. \tag{1.133}$$

If the observational noise is small, the Jacobian Y can be determined at the observations $\boldsymbol{l}$ instead of at the current value $\boldsymbol{y}$ of the mean observations. Hence the update step in (7.19) then would be omitted. The evaluation still can be based on the estimated variance factor, which can be based on $\boldsymbol{c}_g$ alone, and the covariance matrix of the estimated parameters.

1.5.2.3 Final estimates and evaluation

The final estimates are derived from the approximate values in the last iteration, assuming convergence is achieved. Hence we have the final estimates

$$\widehat{\boldsymbol{x}} := \boldsymbol{x}^a\,, \quad \widehat{\boldsymbol{y}} := \boldsymbol{y}^a\,, \quad \text{and} \quad \widehat{\boldsymbol{v}} = \boldsymbol{v}^a\,. \tag{1.134}$$

The estimated variance factor uses the value of the optimization function at the estimates and can be written in different ways

estimated variance factor

$$\widehat{\sigma}_0^2 = \frac{\Omega(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}})}{G-U} = \frac{\widehat{\boldsymbol{v}}^\mathsf{T} W_{ll} \widehat{\boldsymbol{v}}}{G-U} = \frac{\boldsymbol{c}_g^\mathsf{T} W_{gg} \boldsymbol{c}_g}{G-U} \tag{1.135}$$

The last relation can be derived at the point of convergence where $\Delta \boldsymbol{x} = \boldsymbol{0}$, $\Delta \boldsymbol{y} = \boldsymbol{0}$, and $\boldsymbol{g}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}$, using $\boldsymbol{c}_g = Y^\mathsf{T} \boldsymbol{v}$. Hence, the optimization can be based on the weighted sums of the squares of the estimated residuals $\boldsymbol{v}$ or the contradiction $\boldsymbol{c}_g$ of the constraints.

Finally, the theoretical covariance matrix of the estimated parameters can be derived from (1.118) by variance propagation, leading to the Cramer-Rao bound for the uncertainty

Cramer-Rao bound, covariance matrix of estimated parameters

$$\Sigma_{\widehat{x}\widehat{x}} = (X^\mathsf{T} (Y^\mathsf{T} \Sigma_{ll} Y)^{-1} X)^{-1}\,. \tag{1.136}$$

2 Pre-calibration and in-situ Self-calibration with Correlated Observations

Deformation analysis based on point clouds taken at different times may require to take into account both preclalibration and in situ self-calibration of the used instruments. We analyse the mutual effect of pre-calibration and in-situ self-calibration w.r.t. (1) the necessity to exploit the full covariance structure of the point cloud induced by the pre-calibration and (2) the possibility of increasing the computational efficiency during the self-calibration.

2.1 Preface . . . 22
2.2 Summary . . . 22
2.3 Introduction . . . 24
2.3.1 Motivation . . . 24
2.3.2 Rao's lemma . . . 24
2.3.3 Goal and result . . . 26
2.4 The setup . . . 26
2.4.1 The covariance matrix of the observations . . . 26
2.4.2 A: Self-calibration with independent points . . . 29
2.4.3 B: Self-calibration with fusion using independent points . . . 29
2.4.4 C: Self-calibration exploiting a priori calibration . . . 30
2.4.5 D: Self-calibration with fusion using correlated points . . . 32
2.5 Synopsis . . . 33
2.6 Concluding remarks . . . 34
2.7 Appendix: Covariance matrix for given design matrix, observations, estimate and covariance matrix of parameters . . . 34
2.7.1 Example: The mean of two values $y_i, i = 1, 2$ . . . 34

2.1 Preface

The note (2023) addresses the question how a priori pre-calibration result may influence a possible in-situ self-calibration, both concerning the achievable accuracy as well as the numerical effort. The result uses a lemma by Rao (1967, Lemma 5a) which states under which conditions the result of an estimation is invariant to a change in the assumed structure of the covariance matrix of the observations.

2.2 Summary

We analyse the computational and statistical efficiency of self-calibration when reconstructing a surface from point cloud taken with a laser scanner where we know the calibration result. We discuss fusing the prior calibration information with the one from the in-situ measurements and the effect of the uncertainty of the prior calibration $(\boldsymbol{c}_a, \Sigma_{c_a c_a})$ onto the covariance matrix $\mathbb{D}(\underline{\boldsymbol{\theta}})$ of the unknown parameters $\boldsymbol{\theta}$.

We address four cases, A to D, differing by their stochastical and their functional models:

- The a priori calibration result is not available (or used, cases (A,C)) or is integrated into the self-calibration with a priori information in a Bayesian manner (cases (B,D)). Hence, we have the two (linearized) functional models for estimating the parameters $\boldsymbol{y}$ of the object's form and the calibration parameters $\boldsymbol{c}$

$$(A,C):\ \mathbb{E}\left(\underline{\boldsymbol{y}}\right) = [\mathsf{B}\,,\,\mathsf{C}]\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{c}\end{bmatrix},\ \text{or}\ (B,D):\ \mathbb{E}\left(\begin{bmatrix}\underline{\boldsymbol{y}}\\ \underline{\boldsymbol{c}}_a\end{bmatrix}\right) = \begin{bmatrix}\mathsf{B} & \mathsf{C}\\ \mathsf{0} & \mathsf{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{c}\end{bmatrix}. \quad (2.1)$$

- The covariance matrix of the observations is assumed to be (a) block diagonal, assuming the **p**oints are mutually uncorrelated or to be (b) fully populated due to the joint effect of the uncertainty $\Sigma_{c_a c_a}$ of the a priori **c**alibration parameters $\boldsymbol{c}_a$ onto the observations $\mathsf{C}_a\boldsymbol{c}_a$. So we either use

$$\begin{aligned}(A):\ \mathbb{D}(\underline{\boldsymbol{y}}) &= \Sigma_{ll,p} =: \Sigma_0\,, \quad \text{or} \quad (C):\ \mathbb{D}(\underline{\boldsymbol{y}}) = \Sigma_0 + \mathsf{C}_a\Sigma_{c_a c_a}\mathsf{C}_a^\mathsf{T} =: \Sigma\,. \quad (2.2)\\ (B):\ \mathbb{D}\left(\begin{bmatrix}\underline{\boldsymbol{y}}\\ \underline{\boldsymbol{c}}_a\end{bmatrix}\right) &= \begin{bmatrix}\Sigma_0 & \mathsf{0}\\ \mathsf{0} & \Sigma_{c_a c_a}\end{bmatrix} \quad \text{or} \quad (D):\ \mathbb{D}\left(\begin{bmatrix}\underline{\boldsymbol{y}}\\ \underline{\boldsymbol{c}}_a\end{bmatrix}\right) = \begin{bmatrix}\Sigma_0 & \mathsf{0}\\ \mathsf{0} & \mathsf{0}\end{bmatrix} + \begin{bmatrix}\mathsf{C}_a\\ \mathsf{I}\end{bmatrix}\Sigma_{c_a c_a}[\mathsf{C}_a^\mathsf{T}\,,\,\mathsf{I}]\,.\end{aligned}$$

The four cases are analysed w.r.t. their estimates and covariance matrices, see Tab. 2.1.

	$\mathbb{D}(\underline{\boldsymbol{y}}) = \Sigma_0$		$\mathbb{D}(\underline{\boldsymbol{y}}) = \Sigma$
SC	$\widehat{\boldsymbol{\theta}} \mid A$ $\mathbb{D}(\widehat{\underline{\boldsymbol{\theta}}} \mid A) =$ $\begin{bmatrix}\mathsf{B}^\mathsf{T}\mathsf{W}_0\mathsf{B} & \mathsf{B}^\mathsf{T}\mathsf{W}_0\mathsf{C}\\ \mathsf{C}^\mathsf{T}\mathsf{W}_0\mathsf{B} & \mathsf{C}^\mathsf{T}\mathsf{W}_0\mathsf{C}\end{bmatrix}^{-1}$	$\equiv$	$\widehat{\boldsymbol{\theta}} \mid C$ $\mathbb{D}(\widehat{\underline{\boldsymbol{\theta}}} \mid C) =$ $\mathbb{D}(\widehat{\underline{\boldsymbol{\theta}}} \mid A) + \begin{bmatrix}\mathsf{0}\\ \mathsf{I}_C\end{bmatrix}\Sigma_{c_a c_a}[\mathsf{0}\,,\mathsf{I}_C]$
BSC	$\mathbb{D}(\widehat{\underline{\boldsymbol{\theta}}} \mid B) =$ $\begin{bmatrix}\mathsf{B}^\mathsf{T}\mathsf{W}_0\mathsf{B} & \mathsf{B}^\mathsf{T}\mathsf{W}_0\mathsf{C}\\ \mathsf{C}^\mathsf{T}\mathsf{W}_0\mathsf{B} & \mathsf{C}^\mathsf{T}\mathsf{W}_0\mathsf{C} + \mathsf{W}_{c_a c_a}\end{bmatrix}^{-1}$		$\mathbb{D}(\widehat{\underline{\boldsymbol{\theta}}} \mid D) =$ D: $\begin{bmatrix}(\mathsf{B}^\mathsf{T}\mathsf{W}_0\mathsf{B})^{-1} & \mathsf{0}\\ \mathsf{0} & \Sigma_{c_a c_a}\end{bmatrix}$

Table 2.1: Estimates and covariance matrices of the estimated parameters when using the four models for self-calibration and assuming $\mathsf{C} = \mathsf{C}_a$. SC: self-calibration without prior, BSC: Bayesian self-calibration

The main result of this note is the following: *If the matrix $\mathsf{C} \equiv \mathsf{C}_a$ is common to the stochastical model in (2.2) and the functional models in (2.1)*, then, following Rao (1967, Lemma 5a), *the estimates of model A and C coincide*, allowing to use model A for an efficient estimation of the parameters and their covariance matrix. Moreover, using Rao's lemma decorrelates and simplifies the solution for model D.

Especially, we have the following relations between the estimates in the model A and in the models B and C

$$\mathbb{D}(\widehat{\underline{\boldsymbol{\theta}}} \mid B) \le \mathbb{D}(\widehat{\underline{\boldsymbol{\theta}}} \mid A) \le \mathbb{D}(\widehat{\underline{\boldsymbol{\theta}}} \mid C)\,, \quad (2.3)$$

$$\mathbb{D}(\widehat{\underline{\boldsymbol{y}}} \mid D) \le \mathbb{D}(\widehat{\underline{\boldsymbol{y}}} \mid B) \le \mathbb{D}(\widehat{\underline{\boldsymbol{y}}} \mid A) = \mathbb{D}(\widehat{\underline{\boldsymbol{y}}} \mid C)\,. \quad (2.4)$$

An individual sensitivity analysis allows to determine the expected loss in quality, accuracy and reliability in Baarda's sense, without requiring actual observations.

2.3 Introduction

2.3.1 Motivation

Taking point clouds as observations for the estimation of object forms, for deformation analysis, or for calibration needs to take the stochastical properties of the coordinates of the points into account as far as necessary. The quality of the assumed stochastical model needs to be *acceptable*, not necessarily optimal, *for the envisaged application*.

Especially for deformation analysis, where the deformations are in the order of the measuring precision, a realistic stochastical model, taking all known dependencies into account, may be required.

Unfortunately, the points in a point cloud may be correlated due to the uncertainty of the instrumental calibration. This generally leads to a large fully populated covariance matrix Σ_{yy} of the N observations, collected in the vector $\boldsymbol{y}$. As a consequence any estimation minimizing the weighed squares of the residuals is confronted with using the inverse $W_{yy} = \Sigma_{yy}$, which often is called information matrix or precision matrix.

This note shows under which conditions it is possible to work with uncorrelated points, thus with a block matrix containing the 3×3 covariance matrices $\Sigma_{y_i y_i}$ of the I individual points, instead of a fully populated covariance matrix, without losing accuracy.

2.3.2 Rao's lemma

The idea is to exploit the Lemma 5a in Rao (1967) which states under which conditions the estimation with a covariance matrix containing certain additive variance components does not change the parameters. Especially, it starts from the given the linear Gauss-Markov model,

$$\boldsymbol{y} + \boldsymbol{v} = X\boldsymbol{\theta} \quad \text{with} \quad \Sigma_0 = \mathbb{D}(\boldsymbol{y})\,, \tag{2.5}$$

and the estimated parameters

$$\hat{\boldsymbol{\theta}}_0 = (X^\mathsf{T}\Sigma_0^{-1}X)^{-1}X^\mathsf{T}\Sigma_0^{-1}\boldsymbol{y}\,, \tag{2.6}$$

Then, when using the modified covariance matrix

$$\Sigma = X\boldsymbol{\Gamma}X^\mathsf{T} + \Sigma_{\mathbf{0}}Z\boldsymbol{\Theta}Z^\mathsf{T}\Sigma_{\mathbf{0}} + \Sigma_{\mathbf{0}} \quad \text{with} \quad Z^\mathsf{T}X = 0\,, \tag{2.7}$$

with arbitrary matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Theta}$ (which we will not need in the following) the estimate $\hat{\boldsymbol{\theta}}_0$ from (2.6) is identical to the estimate,

$$\hat{\boldsymbol{\theta}} = (X^\mathsf{T}\Sigma^{-1}X)^{-1}X^\mathsf{T}\Sigma^{-1}\boldsymbol{y}\,, \tag{2.8}$$

when using the full covariance matrix.

Fig. 2.1 shows the principle of least squares estimation with a unit matrix and an arbitrary covariance matrix for the observations in the simple model $\underline{\boldsymbol{y}} \sim \mathcal{N}(\boldsymbol{x}\theta, \Sigma)$. Fig. 2.2 visualizes the idea of Rao's lemma.

As can be seen by variance propagation its covariance matrix is

$$\Sigma_{\hat{\theta}\hat{\theta}} = (X^\mathsf{T}\Sigma^{-1}X)^{-1}\,, \tag{2.9}$$

hence, not $(X^\mathsf{T}\Sigma_0^{-1}X)^{-1}$, thus in principal needs to take the full covariance matrix Σ into account.[1]

Observe, the two first components in the covariance matrix (2.7) have the structure of a weighted block dyadic product XSX^T, similar to the 1D case $s\boldsymbol{x}\boldsymbol{x}^\mathsf{T}$.

[1] The result has as special case the mean of N values y_n in case the observations have the same variance σ^2 and are mutually correlated with the same correlation coefficient $\rho \in [-1/(N-1), +1]$. Then the normal arithmetic mean $\hat{\mu} = \sum_n l_n/N$ is the optimal estimator, but its variance is $\sigma_{\hat{\mu}}^2 = (1 + r(N-1)) \cdot \sigma^2/N$, but not σ^2/N. This can be shown using $A = \mathbf{1}_N$, thus a vector of N ones, and $\Sigma = \sigma^2[(1-\rho)I_N + \rho\mathbf{1}_N\mathbf{1}_N^\mathsf{T}]$.

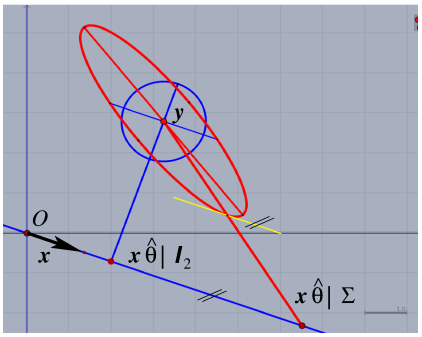

Figure 2.1: Least squares estimation. Model $\underline{\boldsymbol{y}} \sim \mathcal{N}(\boldsymbol{x}\theta, \Sigma)$. If $\Sigma = I_2$, indicated by the blue circular standard ellipse, the optimal point lies on the footpoint $\widehat{\boldsymbol{y}} \mid I_2 = \boldsymbol{x}\widehat{\theta} \mid I_2$ of $\boldsymbol{y}$ onto the line $\mathbb{E}(\underline{\boldsymbol{y}}) = \boldsymbol{x}\theta$. If the covariance matrix Σ is a general matrix, represented by the red standard ellipse, then the optimal point $\widehat{\boldsymbol{y}} \mid \Sigma = \boldsymbol{x}\widehat{\theta} \mid \Sigma$ is the intersection of the (blue) line $\mathbb{E}(\underline{\boldsymbol{y}}) = \boldsymbol{a}\theta$ passing through O and the (red) line, defined by the direction from $\boldsymbol{y}$ to that point of the ellipse, where the tangent (yellow) is parallel to $\boldsymbol{x}$

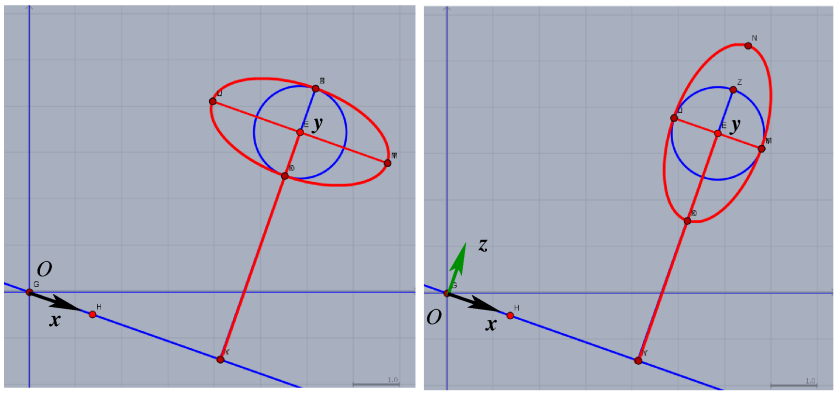

Figure 2.2: Visualization of Rao's lemma: Least squares estimation with modified covariance matrix. Model $\underline{\boldsymbol{y}} \sim \mathcal{N}(\boldsymbol{x}\theta, \Sigma)$. If $\Sigma = \sigma^2 I + \gamma \boldsymbol{x}\boldsymbol{x}^{\mathsf{T}}$ or if $\Sigma = \sigma^2 I + \theta \boldsymbol{z}\boldsymbol{z}^{\mathsf{T}}$, with $\boldsymbol{z} \perp \boldsymbol{x}$, hence generally, if $\Sigma = \sigma^2 I + \gamma \boldsymbol{x}\boldsymbol{x}^{\mathsf{T}} + \theta \boldsymbol{z}\boldsymbol{z}^{\mathsf{T}}$, the semi-axes of the standard ellipse are parallel or orthogonal to $\boldsymbol{x}$. Then, the least squares estimate for the generalized covariance matrix is the same as for $\Sigma = I_2$. However, the covariance matrix of the estimate depends on the modification, namely the factors σ^2 and γ.

Remark: It is well known[2], that changing the covariance matrix of the observations leads to an effect onto the estimated parameter, which is in the range of their standard deviations, unless the change of the covariance matrix or the weight matrix is very large. It mainly influences their covariance matrix. The result of Rao's Lemma addresses the extreme case, where the effect onto the parameters is zero, which requires that the change of the covariance matrix has a special structure. The effect of model errors has been discussed in the context of self-calibration in Förstner (1982). ◇

2.3.3 Goal and result

The idea is to choose the matrices X, Σ_0, $\mathbf{\Gamma}$, and $\mathbf{\Theta}$ in (2.7), such, that the estimation of the parameters for the object and the calibration

1. can be performed within self-calibration with a block diagonal matrix for the observed points, which increases computational efficiency
2. can use the parameters of a priori calibration for an in-situ self-calibration possibly improving these parameters, and
3. efficiently derive the uncertainty of the estimated parameters.

Computational efficiency also can be achieved, in case only a part of the calibration parameters is included in the self-calibration. The increase in efficiency refers to the estimation of the parameters, as well as to determination of their covariance matrix.

2.4 The setup

We now discuss the used stochastical model of the observations and then four mathematical models for the self-calibration

2.4.1 The covariance matrix of the observations

We assume two sources of measurement deviations, (1) caused by the object properties, yielding $\Sigma_{yy,p}$, and (2) caused by the prior calibration, yielding $\Sigma_{yy,c}$. Hence, we assume the complete uncertainty is described by

$$\Sigma_{yy} = \Sigma_{yy,p} + \Sigma_{yy,c}\,. \tag{2.10}$$

1. The covariance matrix $\Sigma_{yy,p}$ is assumed to be block diagonal

   $$\Sigma_{yy,p} = \mathrm{Diag}(\{\Sigma_{y_iy_i,p}\})\,. \tag{2.11}$$

   and has full rank. The individual 3×3 covariance matrices $\Sigma_{y_iy_i,p}$ are assumed to reflect those parts of the directional and distance uncertainties, which are independent for each point, including those parts which depend on the surface point, e.g., its material and the impact angle.

2. The covariance matrix $\Sigma_{yy,c}$ is assumed to contain all uncertainties of the a priori calibration which effect all points of a scan simultaneously. We do not assume other types of correlations, e.g., caused by the atmosphere. Using the primary error concept the effect of the C calibration parameters[3] $\boldsymbol{c}$ onto the observations is assumed to be describable by

   $$\boldsymbol{y}_c = C_a \boldsymbol{c}_a\,. \tag{2.12}$$

[2] See Koch (1999), where eq. (3.108) shows the effect of using a slightly changed weight matrix $W+\Delta W$ instead of W, and with (3.32) reads as $\hat{\boldsymbol{\theta}} \mid (W+\Delta W) \approx \hat{\boldsymbol{\theta}} \mid W - (X^\mathsf{T} W X)^{-1} X^\mathsf{T} \Delta W\, \hat{\boldsymbol{e}}$, with the estimated residuals $\hat{\boldsymbol{e}} = \boldsymbol{y} - X\hat{\boldsymbol{\theta}}$.

[3] We assume a perfectly constructed instrument would lead to $\boldsymbol{c} = 0$.

Assuming a linear model appears to be reasonable as the effects are small. The estimated parameters $\boldsymbol{c}_a := \hat{\boldsymbol{c}}_a$ of the a priori calibration will be uncertain[4]

$$\underline{\boldsymbol{c}}_a \sim \mathcal{N}(\boldsymbol{\mu}_{c_a}, \Sigma_{c_a c_a}) \,. \tag{2.13}$$

This leads to the uncertain effects of the calibration onto the observations

$$\underline{\boldsymbol{y}}_c \sim \mathcal{N}(\boldsymbol{\mu}_{y_c}, \Sigma_{yy,c_a}) \quad \text{with} \quad \boldsymbol{\mu}_{y_c} = \mathsf{C}_a \boldsymbol{\mu}_{c_a} \quad \text{and} \quad \Sigma_{yy,c_a} = \mathsf{C}_a \Sigma_{c_a c_a} \mathsf{C}_a^\mathsf{T} \,. \tag{2.14}$$

The covariance matrix has a low rank $C = \mathrm{rk}(\mathsf{C}_a)$, but generally is fully populated.

Hence, also the covariance matrix Σ_{yy} will be fully populated since it has the structure

$$\Sigma_{yy} = \Sigma_{yy,p} + \mathsf{C}_a \Sigma_{c_a c_a} \mathsf{C}_a^\mathsf{T} \,, \tag{2.15}$$

where the first part is sparse, namely block-diagonal, and the second part has the structure of a block dyadic product.

On notation: In the following we denote the inverses of covariance matrices as weight/precision/- or information matrices:

$$\mathsf{W} = \Sigma^{-1} \,, \quad \mathsf{W}_0 = \Sigma_0^{-1} \,, \quad \mathsf{W}_{yy,p} = \Sigma_{yy,p}^{-1} \,, \quad \text{and} \quad \mathsf{W}_{c_0 c_0} = \Sigma_{c_0 c_0}^{-1} \,. \tag{2.16}$$

We now discuss four cases for the self-calibration, which simultaneously determines the parameters of the object and calibration parameters. We assume two alternatives for the functional model of the self-calibration and two alternatives for the stochastical model for the observations. Hence, we arrive at the following models

A. Uncorrelated points for self-calibration

B. Uncorrelated points for self-calibration with fusion of the prior calibration

C. Correlated points for self-calibration

D. Correlated points for self-calibration with fusion of the prior calibration

1. The a priori calibration result (A,C) is not available or used or (B,D) is fusing the self-calibration with a priori information. Hence, we have the two (linearized) functional models for estimating the parameters $\boldsymbol{y}$ of the object's form and the calibration parameters $\boldsymbol{c}$

$$(A,C): \; \mathbb{E}\left(\underline{\boldsymbol{y}}\right) = [\mathsf{B}\,,\mathsf{C}] \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{c} \end{bmatrix} \,, \quad \text{or} \quad (B,D): \; \mathbb{E}\left(\begin{bmatrix} \underline{\boldsymbol{y}} \\ \underline{\boldsymbol{c}}_a \end{bmatrix} \right) = \begin{bmatrix} \mathsf{B} & \mathsf{C} \\ \mathsf{0} & \mathsf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{c} \end{bmatrix} . \tag{2.17}$$

We call models (A,C) *self-calibration* and models (B,D) *self-calibration with fusion* in the following.
Depending on the context, the self-calibration may refer to only a subset of parameters used in the prior calibration, e.g., only those which are to be expected to be determinable within the self-calibration. Similarly, self-calibration with fusion (2) may only refer to those parameters which are expected to change over time. Hence, the models may have the same coefficient matrix B but different coefficient matrices C.

2. The covariance matrix of the observations is assumed to be block diagonal, assuming the **p**oints are mutually uncorrelated or to be fully populated due to the joint effect of the uncertainty $\Sigma_{c_a c_a}$ of the a priori **c**alibration parameters $\boldsymbol{c}_a$ onto the observations $\mathsf{C}_a \boldsymbol{c}_a$.

[4] Random variables are underscored.

Since the observation vectors in models (A,C) are different from thos in models (B,D) we need to consider them separately.
In case of models (A,C) we either use

$$\mathbb{D}(\boldsymbol{y} \mid A) = \Sigma_{yy,p} =: \Sigma_0\,, \tag{2.18}$$

or

$$\mathbb{D}(\boldsymbol{y} \mid C) = \Sigma_{yy,p} + \Sigma_{yy,c} = \Sigma_{yy,p} + C_a \Sigma_{c_a c_a} C_a^{\mathsf{T}} =: \Sigma\,. \tag{2.19}$$

Stochastical model (A) is a special case of model (C), so, when used, leads to sub-optimal estimates, if the observations actually are correlated.
In case of models B the a priori information ($\boldsymbol{c}_a, \Sigma_{c_a c_a}$ used in the self-calibration with fusion we reasonably may assume the prior information is independent of the observed points, hence we have

$$\mathbb{D}\left(\begin{bmatrix} \underline{\boldsymbol{y}} \\ \underline{\boldsymbol{c}}_a \end{bmatrix} \mid B\right) = \begin{bmatrix} \Sigma_{yy,p} & 0 \\ 0 & \Sigma_{c_a c_a} \end{bmatrix}. \tag{2.20}$$

In case D we assume the observed coordinates are mutually correlated due to the common calibration uncertainty. But, then also the calibration parameters $\boldsymbol{c}_a$ will be correlated with the observed points, since we have

$$\begin{bmatrix} \underline{\boldsymbol{y}} \\ \underline{\boldsymbol{c}}_a \end{bmatrix} = \begin{bmatrix} \underline{\boldsymbol{y}}_g + C_a \underline{\boldsymbol{c}}_a \\ \underline{\boldsymbol{c}}_a \end{bmatrix} = \begin{bmatrix} I & C_a \\ 0 & I \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{\theta}} \\ \underline{\boldsymbol{c}} \end{bmatrix}. \tag{2.21}$$

Hence we obtain the joint covariance matrix

$$\begin{aligned} \mathbb{D}\left(\begin{bmatrix} \underline{\boldsymbol{y}} \\ \underline{\boldsymbol{c}}_a \end{bmatrix} \mid D\right) &= \begin{bmatrix} I & C \\ 0 & I \end{bmatrix} \begin{bmatrix} \Sigma_{yy,p} & \\ & \Sigma_{c_a c_a} \end{bmatrix} \begin{bmatrix} I & 0 \\ C^{\mathsf{T}} & I \end{bmatrix} && (2.22) \\ &= \begin{bmatrix} \Sigma_{yy,p} + C_a \Sigma_{c_a c_a} C_a^{\mathsf{T}} & C_a \Sigma_{c_a c_a} \\ \Sigma_{c_a c_a} C_a^{\mathsf{T}} & \Sigma_{c_a c_a} \end{bmatrix} && (2.23) \\ &= \begin{bmatrix} \Sigma_{yy,p} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} C_a \\ I \end{bmatrix} \Sigma_{c_a c_a} [C_a^{\mathsf{T}}, I] && (2.24) \end{aligned}$$

There is a profound difference when fusing the uncorrelated and the correlated observations in models B and D.
We first look at the models A and C. As can be seen from (2.18) and (2.19), the uncertainty does not decrease when taking the correlations into account:

$$\mathbb{D}(\boldsymbol{y} \mid C) - \mathbb{D}(\boldsymbol{y} \mid A) = \Sigma - \Sigma_0 = C_a \Sigma_{c_a c_a} C_a^{\mathsf{T}} \geq 0 \tag{2.25}$$

Hence, the uncertainty in model C generally is higher than in model A. Hence, we can expect, the results using model B are worse (not better) than those with model A. Since the two groups of observations is not independent the model does represent a Bayesian estimation of the parameters.
This contrasts to the relation between the uncertainties in models B and D. Here we have

$$\begin{aligned} \mathbb{D}(\boldsymbol{y} \mid D) - \mathbb{D}(\boldsymbol{y} \mid B) &= \begin{bmatrix} \Sigma_{yy,p} + C_a \Sigma_{c_a c_a} C_a^{\mathsf{T}} & C_a \Sigma_{c_a c_a} \\ \Sigma_{c_a c_a} C_a^{\mathsf{T}} & \Sigma_{c_a c_a} \end{bmatrix} - \begin{bmatrix} \Sigma_{yy,p} & 0 \\ 0 & \Sigma_{c_a c_a} \end{bmatrix} \\ &= \begin{bmatrix} C_a \Sigma_{c_a c_a} C_a^{\mathsf{T}} & C_a \Sigma_{c_a c_a} \\ \Sigma_{c_a c_a} C_a^{\mathsf{T}} & 0 \end{bmatrix} \gtreqless 0\,. && (2.26) \end{aligned}$$

Hence, the accuracy difference is indefinite. This indicates, that model D will not generally lead to better results than model B. Also, since the two groups of observations is not independent the model does not represent a Bayesian estimation of the parameters.

We now discuss the four different models in more detail.

2.4.2 A: Self-calibration with independent points

The most simple model, case A, is in-situ self-calibration *without* having access (or using) to the result of a prior calibration. It reads as

$$\mathbb{E}(\underline{\boldsymbol{y}} \mid A) = \boldsymbol{X}\boldsymbol{\theta} \quad \mathbb{D}(\underline{\boldsymbol{y}} \mid A) = \Sigma_0 \quad \text{with} \quad \boldsymbol{X} = [\boldsymbol{B}, \boldsymbol{C}], \quad \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{c} \end{bmatrix} \tag{2.27}$$

where

- the Y parameters $\boldsymbol{y}$ are used to describe the object, e.g., using splines, and
- the C parameters $\boldsymbol{c}$ are calibration parameters within the self-calibrating estimation. They generally need not be the same as in a pre-calibration.
- Since we do not use or have access to a prior calibration, we need to assume the covariance matrix of the observations is block diagonal.

The uncertainty of the estimated parameters results from the normal equations

$$\boldsymbol{N}_0\hat{\boldsymbol{\theta}} = \boldsymbol{n}_0 \tag{2.28}$$

with

$$\boldsymbol{N}_0 = \begin{bmatrix} \boldsymbol{N}_{11,0} & \boldsymbol{N}_{12,0} \\ \boldsymbol{N}_{21,0} & \boldsymbol{N}_{22,0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}^\mathsf{T}\boldsymbol{W}_0\boldsymbol{B} & \boldsymbol{B}^\mathsf{T}\boldsymbol{W}_0\boldsymbol{C} \\ \boldsymbol{C}^\mathsf{T}\boldsymbol{W}_0\boldsymbol{B} & \boldsymbol{C}^\mathsf{T}\boldsymbol{W}_0\boldsymbol{C} \end{bmatrix} \tag{2.29}$$

and

$$\boldsymbol{n}_0 = \begin{bmatrix} \boldsymbol{n}_{1,0} \\ \boldsymbol{n}_{2,0} \end{bmatrix} = \begin{bmatrix} \boldsymbol{B}^\mathsf{T}\boldsymbol{W}_0\boldsymbol{y} \\ \boldsymbol{C}^\mathsf{T}\boldsymbol{W}_0\boldsymbol{y} \end{bmatrix}. \tag{2.30}$$

The index 0 stands using the block-diagonal matrix Σ_0. In case model A holds we have the covariance matrix

$$\mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid A) = \boldsymbol{N}_0^{-1}. \tag{2.31}$$

This model is useful, since *the normal equation system can be setup point by point*, and *it will generally be sparse*, since each point only influences the coordinates of the neighbouring knots/control points of a spline surface. The sparsity of $\boldsymbol{N}_0$ has two positive numerical effects:

(i) The solution of the normal equation system can exploit the sparsity, and therefore can be performed numerically efficient.

(ii) Though the covariance matrix $\Sigma_{\hat{\theta}\hat{\theta},0} = \boldsymbol{N}_0^{-1}$ will be generally full, one may efficiently determine those elements of the covariance matrix, where the normal equation matrix is non-zero, without needing to determine the other elements of the covariance matrix, see Takahashi et al. (1973, cf MATLAB-code `sparseinv.m`) and Vanhatalo and Vehtari (2008).

This model certainly is too simplified, since neither possible correlations between the observations nor some, possibly available, a priori information is taken into account.

2.4.3 B: Self-calibration with fusion using independent points

In model B, we now want to fuse some a priori results $(\boldsymbol{c}_a, \Sigma_{c_a c_a})$ within the self-calibration from (2.27). As we discussed above, this corresponds to a Bayesian estimation of the parameter vector, with prior on the calibration parameters. This then just leads to additional observations $\underline{\boldsymbol{c}}_a \sim \mathcal{M}(\boldsymbol{c}_a, \Sigma_{c_a c_a})$ and thus the model

$$\mathbb{E}\left(\begin{bmatrix} \underline{\boldsymbol{y}} \\ \underline{\boldsymbol{c}}_a \end{bmatrix} \mid B\right) = \begin{bmatrix} \boldsymbol{B} & \boldsymbol{C} \\ \boldsymbol{0} & \boldsymbol{I}_c \end{bmatrix} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{c} \end{bmatrix} \tag{2.32}$$

with the extended covariance matrix of the joint observation vector $(\boldsymbol{y}_p, \boldsymbol{c}_a)$

$$\mathbb{D}\left(\left[\begin{array}{c} \underline{\boldsymbol{y}} \\ \underline{\boldsymbol{c}}_a \end{array}\right] \mid B\right) = \left[\begin{array}{cc} \Sigma_0 & 0 \\ 0 & \Sigma_{c_a c_a} \end{array}\right]. \tag{2.33}$$

The normal equation system reads as

$$M_0 \hat{\boldsymbol{\theta}} = \boldsymbol{m}_0 \tag{2.34}$$

with

$$M_0 = N_0 + \left[\begin{array}{c} 0 \\ I_C \end{array}\right] W_{c_a c_a} [0\,,\, I_C] = \left[\begin{array}{cc} B^\mathsf{T} W_0 B & B^\mathsf{T} W_0 C \\ C^\mathsf{T} W_0 B & C^\mathsf{T} W_0 C + W_{c_a c_a} \end{array}\right] \tag{2.35}$$

and similarly

$$\boldsymbol{m}_0 = \boldsymbol{n}_0 + \left[\begin{array}{c} \mathbf{0} \\ W_{c_a c_a} \boldsymbol{c}_a \end{array}\right]. \tag{2.36}$$

Also here, the normal equation matrix will be sparse, allowing to increase numerical efficiency, both during the solution as well as for determining the covariance matrix of the parameters. This is the main motivation for using this model.

Remark: 1. Though this model formally is correct, in the context of in-situ self-calibration it contains a contradiction: The prior calibration result $(\boldsymbol{c}_s, \Sigma_{c_a c_a}$ is used explicitely, but the observed points are assumed to be uncorrelated, though they are assume to be measured by the same instrument, thus should be treated as mutually dependent. ◇

Remark: 2. In case parameters $\boldsymbol{c}$ are partitioned, namely $\boldsymbol{c}_{a2}$ of $\boldsymbol{c}_a$, e.g., if

$$\boldsymbol{c}_a = \left[\begin{array}{c} \boldsymbol{c}_{a1} \\ \boldsymbol{c}_{a2} \end{array}\right] \quad \text{and} \quad \boldsymbol{y}_c = [C_1\,,\, C_2] \left[\begin{array}{c} \boldsymbol{c}_{a1} \\ \boldsymbol{c}_{a2} \end{array}\right] = C_1 \boldsymbol{c}_{a1} + C_2 \boldsymbol{c}_{a2} \tag{2.37}$$

where the parameters $\boldsymbol{c}_{a1}$ are just fixed values, used for correcting the observations, then we can rewrite the model as

$$(\underline{\boldsymbol{y}}_p + C_1 \boldsymbol{c}_{a1} + C_2 \boldsymbol{c}_{a2}) + \underline{\boldsymbol{v}} = B\boldsymbol{y} + C\boldsymbol{c}\,, \tag{2.38}$$

Now, since the effect of the parameters $\boldsymbol{c}_{a2}$ onto the observations is the same as those of $\boldsymbol{c}$, the coefficient matrices C and C_2 coincide, why we obtain the model

$$(\underline{\boldsymbol{y}}_p + C_1 \boldsymbol{c}_{a1}) + \underline{\boldsymbol{v}} = B\boldsymbol{y} + C(\boldsymbol{c} - \boldsymbol{c}_{a2})\,, \tag{2.39}$$

Hence if we only are able to estimate the difference $\Delta\boldsymbol{c} = \boldsymbol{c} - \boldsymbol{c}_{a2}$, i.e., for given $\boldsymbol{c}_{a2}$ the corrections $\Delta\boldsymbol{c}$. ◇

Though this model takes into account the result of a prior calibration it still assumes a too simplistic covariance matrix Σ_0 for the observations, thus is statistically suboptimal, in case correlations between the points exist.

2.4.4 C: Self-calibration exploiting a priori calibration

2.4.4.1 The model

In model C, we instead of fusing the result of the a priori calibration with the current measurements, we correct take into account that the observations due to the uncertainty of the a priori calibration are correlated. Hence, we have the same functional model as in case A,

$$\mathbb{E}(\underline{\boldsymbol{y}} \mid C) = \mathbb{E}(\underline{\boldsymbol{y}}_p + C_a \underline{\boldsymbol{c}}_a \mid C) = X\boldsymbol{\theta} \quad \text{with} \quad X = [B, C]\,, \quad \boldsymbol{\theta} = \left[\begin{array}{c} \boldsymbol{y} \\ \boldsymbol{c} \end{array}\right], \tag{2.40}$$

but now assume the covariance matrix of the observations is

$$\mathbb{D}(\underline{\boldsymbol{y}} \mid C) = \Sigma_{yy,p} + C_a \Sigma_{c_a c_a} C_a^\mathsf{T} =: \Sigma\,. \tag{2.41}$$

Observe, we generally do not enforce, the self-calibration determines/corrects the same parameters as the a priori calibration, which is reasonable, in case we only want to improve the results of some calibration parameters.

Though the design matrix A is sparse, the resulting normal equation system will not be sparse. Thus – without further constraints – no numerically efficient solution is possible.

2.4.4.2 Exploiting Rao's result

This changes, if we assume the two matrices C and C_a coincide. Then Rao's lemma can be applied.

If we refer to (2.7), then, when assuming

$$C \equiv C_a\,, \quad \Sigma_0 = \Sigma_{yy,p}\,, \quad \mathbf{\Gamma} = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_{c_a c_a} \end{bmatrix} \quad \text{and} \quad \mathbf{\Theta} = 0\,, \tag{2.42}$$

we obtain $\Sigma := \Sigma_{yy}$ of (2.15), and therefore can conclude: *under the mentioned conditions, using the block-diagonal matrix $\Sigma_{yy,p}$ during estimation leads to the same estimates as when using the full covariance matrix Σ_{yy}.*

Explicitly, the estimated parameters following from the model

$$\underline{\boldsymbol{y}} + \underline{\boldsymbol{v}} = [A, C] \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{c} \end{bmatrix}\,, \quad \text{and} \quad \Sigma := \Sigma_{yy,p} + C\Sigma_{c_a c_a} C^{\mathsf{T}}\,. \tag{2.43}$$

are identical to those following from model A

$$\underline{\boldsymbol{y}} + \underline{\boldsymbol{v}} = [A, C] \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{c} \end{bmatrix}\,, \quad \text{and} \quad \Sigma_0 := \Sigma_{yy,p}\,. \tag{2.44}$$

independent on whether we correct the observations for their calibration errors $C\boldsymbol{c}_a$, as discussed above, thus

$$\boxed{\hat{\boldsymbol{\theta}} \mid C = \hat{\boldsymbol{\theta}} \mid A} \tag{2.45}$$

The covariance matrix of the estimates now results from

$$\hat{\underline{\boldsymbol{\theta}}} = (X^{\mathsf{T}} W_0 X)^{-1} X^{\mathsf{T}} W_0\, \underline{\boldsymbol{y}} \tag{2.46}$$

We obtain the uncertainty of the parameters by variance propagation as

$$\begin{aligned} \mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid C) &= (X^{\mathsf{T}} W_0 X)^{-1} X^{\mathsf{T}} W_0 (\Sigma_0 + C\Sigma_{c_a c_a} C^{\mathsf{T}}) W_0 X (X^{\mathsf{T}} W_0 X)^{-1} && (2.47) \\ &= (X^{\mathsf{T}} W_0 X)^{-1} + (X^{\mathsf{T}} W_0 X)^{-1} X^{\mathsf{T}} W_0 C \Sigma_{c_a c_a} C^{\mathsf{T}} W_0 X (X^{\mathsf{T}} W_0 X)^{-1} && (2.48) \end{aligned}$$

But since

$$(X^{\mathsf{T}} W_0 X)^{-1} [X^{\mathsf{T}} W_0 B\,,\, X^{\mathsf{T}} W_0 C] = \left[\begin{bmatrix} I_Y \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ I_C \end{bmatrix} \right] \tag{2.49}$$

we arrive at

$$\mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid C) = (X^{\mathsf{T}} W_0 X)^{-1} + \begin{bmatrix} 0 \\ I_C \end{bmatrix} \Sigma_{c_a c_a} [0\,, I_C] \tag{2.50}$$

or

$$\boxed{\mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid C) = \mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid A) + \begin{bmatrix} 0 \\ I_C \end{bmatrix} \Sigma_{c_a c_a} [0\,, I_C]} \tag{2.51}$$

or explicitely

$$\mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid C) = \begin{bmatrix} \Sigma_{\hat{y}\hat{y},0} & \Sigma_{\hat{y}\hat{c},0} \\ \Sigma_{\hat{c}\hat{y},0} & \Sigma_{\hat{c}\hat{c},0} + \Sigma_{c_a c_a} \end{bmatrix} \tag{2.52}$$

Hence, using model C allows arrive at computational efficient estimation of the parameters, as well determination of their covariance matrix.

This observation gives some insight into the ability of this model to compensate for information of some prior calibration, which allows to exploit Rao's result to increase computational efficiency for determining the parameters.

However, this observation also indicates that the prior information is not fully integrated.

2.4.5 D: Self-calibration with fusion using correlated points

Model D now integrates the prior calibration and the in-situ measurements in a Bayesian self-calibration. The model now is

$$\mathbb{E}\left(\begin{bmatrix} \underline{\boldsymbol{y}} \\ \underline{\boldsymbol{c}}_a \end{bmatrix} \mid D\right) = \underbrace{\begin{bmatrix} B & C \\ 0 & I \end{bmatrix}}_{X} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{c} \end{bmatrix} \tag{2.53}$$

We now have to take into account that the observations $\underline{\boldsymbol{y}}$ and the prior values $\underline{\boldsymbol{c}}_a$ are correlated and use the joint covariance matrix from (2.22):

$$\begin{aligned} \mathbb{D}\left(\begin{bmatrix} \underline{\boldsymbol{y}} \\ \underline{\boldsymbol{c}}_a \end{bmatrix} \mid D\right) &= \begin{bmatrix} \Sigma_{yy,p} + C_a \Sigma_{c_a c_a} C_a^{\mathsf{T}} & C_a \Sigma_{c_a c_a} \\ \Sigma_{c_a c_a} C_a^{\mathsf{T}} & \Sigma_{c_a c_a} \end{bmatrix} && (2.54) \\ &= \begin{bmatrix} \Sigma_{yy,p} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} C_a \\ I \end{bmatrix} \Sigma_{c_a c_a} [C_a^{\mathsf{T}}, I]\,. && (2.55) \end{aligned}$$

This model appears to enables the invocation of to Rao's Lemma if $C = C_a$, since the second column of X is common to the functional and the stochastical model.

The model then leads to the same estimated parameters as when using the covariance matrix

$$\mathbb{D}\left(\begin{bmatrix} \underline{\boldsymbol{y}} \\ \underline{\boldsymbol{c}}_a \end{bmatrix}\right) = \begin{bmatrix} \Sigma_{yy,p} & 0 \\ 0 & 0 \end{bmatrix}, \tag{2.56}$$

However, this implies, that the parameters $\boldsymbol{c}_a$ from the prior calibration have variance zero, thus are taken as fixed values, just correcting the observations $\boldsymbol{y} - C\boldsymbol{c}_a$ in *a non-self-calibrating model*

$$\boldsymbol{y} - C\boldsymbol{c}_a = B\boldsymbol{y} \quad \text{with} \quad \mathbb{D}(\underline{\boldsymbol{y}}) = \Sigma_0\,. \tag{2.57}$$

Again, using the actual (fully populated) covariance matrix of the joint observation vector into consideration, we obtain the covariance matrix of the estimates by variance propagation from

$$\begin{aligned} \begin{bmatrix} \hat{\underline{\boldsymbol{y}}} \\ \hat{\underline{\boldsymbol{c}}} \end{bmatrix} &= \begin{bmatrix} (B^{\mathsf{T}} W_0 B)^{-1} B^{\mathsf{T}} W_0 & 0 \\ 0 & I_C \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{y}} - C\underline{\boldsymbol{c}}_a \\ \underline{\boldsymbol{c}}_a \end{bmatrix} && (2.58) \\ &= \begin{bmatrix} (B^{\mathsf{T}} W_0 B)^{-1} B^{\mathsf{T}} W_0 & -(B^{\mathsf{T}} W_0 B)^{-1} B^{\mathsf{T}} W_0 C \\ 0 & I_C \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{y}} \\ \underline{\boldsymbol{c}}_a \end{bmatrix}. && (2.59) \end{aligned}$$

Using

$$N_{11} = B^{\mathsf{T}} W_0 B \quad \text{and} \quad N_{12} = B^{\mathsf{T}} W_0 C \tag{2.60}$$

neglecting the index zero, this reads as

$$\begin{bmatrix} \hat{\underline{\boldsymbol{y}}} \\ \hat{\underline{\boldsymbol{c}}} \end{bmatrix} = \begin{bmatrix} N_{11}^{-1} B^{\mathsf{T}} W_0 & -N_{11}^{-1} N_{12} \\ 0 & I_C \end{bmatrix} \begin{bmatrix} \underline{\boldsymbol{y}} \\ \underline{\boldsymbol{c}}_a \end{bmatrix} \tag{2.61}$$

and we obtain

$$\begin{aligned} \mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid D) &= \begin{bmatrix} N_{11}^{-1} B^{\mathsf{T}} W_0 & -N_{11}^{-1} N_{12} \\ 0 & I_C \end{bmatrix} \begin{bmatrix} \Sigma_0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} W_0 B N_{11}^{-1} & 0 \\ -N_{21} N_{11}^{-1} & I_C \end{bmatrix} && (2.62) \\ &\quad + \begin{bmatrix} N_{11}^{-1} B^{\mathsf{T}} W_0 & -N_{11}^{-1} N_{12} \\ 0 & I_C \end{bmatrix} \begin{bmatrix} C \\ I \end{bmatrix} \Sigma_{c_a c_a} [C^{\mathsf{T}}, I] \begin{bmatrix} W_0 B N_{11}^{-1} & 0 \\ -N_{21} N_{11}^{-1} & I_C \end{bmatrix} \\ &= \begin{bmatrix} (B^{\mathsf{T}} W_0 B)^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ I_C \end{bmatrix} \Sigma_{c_a c_a} [0, I_C] && (2.63) \\ &= \begin{bmatrix} (B^{\mathsf{T}} W_0 B)^{-1} & 0 \\ 0 & \Sigma_{c_a c_a} \end{bmatrix}. && (2.64) \end{aligned}$$

This is a stunning result: The fusion of the prior information $(\boldsymbol{c}_a, \Sigma_{c_a c_a})$ from the calibration does neither improve the calibration parameters, the correlations assumed for the joint observation vector $(\boldsymbol{y}, \boldsymbol{c}_e)$ (1) have no effect onto the calibration parameters, (2) are not needed to determine the object parameters $\boldsymbol{y}$, and (3) decorrelate the estimation of the parameter for object and calibration.

2.5 Synopsis

The following Table 2.2 collects the results, especially the covariance matrices

$$\mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid k)\,, \quad k = A, B, C, D\,. \tag{2.65}$$

for the four cases.

	$\mathbb{D}(\underline{\boldsymbol{y}}) = \Sigma_0$		$\mathbb{D}(\underline{\boldsymbol{y}}) = \Sigma$
SC	$\hat{\underline{\boldsymbol{\theta}}} \mid A$ $\mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid A) =$ $\begin{bmatrix} B^\mathsf{T}W_0B & B^\mathsf{T}W_0C \\ C^\mathsf{T}W_0B & C^\mathsf{T}W_0C \end{bmatrix}^{-1}$ (2.29)	$\equiv$	$\hat{\underline{\boldsymbol{\theta}}} \mid C$ $\mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid C) =$ $\mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid A) + \begin{bmatrix} 0 \\ I_C \end{bmatrix} \Sigma_{c_a c_a} [0\,, I_C]$ (2.35)
BSC	$\mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid B) =$ $\begin{bmatrix} B^\mathsf{T}W_0B & B^\mathsf{T}W_0C \\ C^\mathsf{T}W_0B & C^\mathsf{T}W_0C + W_{c_a c_a} \end{bmatrix}^{-1}$ (2.51)		$\mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid D) =$ D: $\begin{bmatrix} (B^\mathsf{T}W_0B)^{-1} & 0 \\ 0 & \Sigma_{c_a c_a} \end{bmatrix}$ (2.64)

Table 2.2: The covariance matrices of the estimated parameters when using the four models for self-calibration, SC: self-calibration without prior, BSC: Bayesian self-calibration

First, the estimated parameters for model A and C are the same, see (2.45):

$$\hat{\boldsymbol{\theta}} \mid C = \hat{\boldsymbol{\theta}} \mid A\,. \tag{2.66}$$

Second, we compare the accuracy achievable in the different models:

1. the influence of changing the covariance matrix onto the accuracy can be determined for models A and C. Since the uncertainty of the observations in model A are assumed to be not larger than that in model C, hence because $\Sigma - \Sigma_0 \geq 0$ the uncertainty of the parameters in model C generally is larger than that of models A:

   $$\mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid C) \geq \mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid A)\,. \tag{2.67}$$

   However, the accuracy of the object parameters for models A and C is the same:

   $$\mathbb{D}(\hat{\underline{\boldsymbol{y}}} \mid C) = \mathbb{D}(\hat{\underline{\boldsymbol{y}}} \mid A)\,. \tag{2.68}$$

2. the influence of the fusion of prior and in-situ self-calibration can be determined for models A and B. Since the model B includes additional, independent information compared to model A, the uncertainty generally increased by the fusion process:

   $$\mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid A) \geq \mathbb{D}(\hat{\underline{\boldsymbol{\theta}}} \mid B)\,. \tag{2.69}$$

   Observe, this holds for both, the parameters $\boldsymbol{y}$ of the object as well as the calibration parameters $\boldsymbol{c}$, which easily can be seen using the Schur complements of the two diagonal block matrices of the covariance matrices.

3. the accuracy of the estimated parameters in model D cannot be compared to the others in general, since it is not a generalization of one of them. However, the accuracy of the estimated object parameters can be compared. We especially have

   $$\mathbb{D}(\hat{\underline{\boldsymbol{y}}} \mid D) \leq \mathbb{D}(\hat{\underline{\boldsymbol{y}}} \mid B) \leq \mathbb{D}(\hat{\underline{\boldsymbol{y}}} \mid A) = \mathbb{D}(\hat{\underline{\boldsymbol{y}}} \mid C)\,, \tag{2.70}$$

   again using the Schur complements of the corresponding covariance matrices.

2.6 Concluding remarks

Generally, these result only are valid, if Rao's lemma can be applied, i.e., if the calibration parameters $\boldsymbol{c}$ determined in the self-calibration are the same which cause the correlations between the points, formally if the coefficient matrix C in the functional model is the same as the one C_a used in the stochastical model, hence if $\mathsf{C}_a = \mathsf{C}$. This may, be enforced by assuming the calibration parameters not corrected in the self-calibration have zero effect onto the observed points, e.g., of one assumes the these parameters, which are determined in the prior calibration, have small enough variance, to assume it to be zero.

2.7 Appendix: Covariance matrix for given design matrix, observations, estimate and covariance matrix of parameters

On can show, that there is a set of covariance matrices Σ_{yy} if the following is given:

1. the linear model $\mathbb{E}(\underline{y}) = \mathsf{X}\boldsymbol{\theta}$,
2. the value of the estimate and its covariance matrix $\{(\widehat{\boldsymbol{\theta}}, \Sigma_{\widehat{\theta}\widehat{\theta}}) = (\boldsymbol{\theta}, \mathsf{V})\}$ of the parameters, and
3. a vector $\boldsymbol{y}$ of observations,

such that the estimated parameters and their covariance matrix follow from a weighted least squares estimation.

2.7.1 Example: The mean of two values $y_i, i = 1, 2$

Given are two observations $\boldsymbol{y} = [y_i]$ and an estimate $\theta = \widehat{\theta}$ for the mean with variance $v = \sigma^2_{\widehat{\theta}}$. The covariance matrix of the observations is to be chosen adequately.

2.7.1.1 A special solution

We have the following model

$$\mathbb{E}(\underline{\boldsymbol{y}}) = \mathsf{X}\theta \quad \text{with} \quad \mathsf{X} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \mathbf{1}_2 \tag{2.71}$$

and need to choose, say in the form, containing the correlation coefficient $\rho \in [-1, +1]$

$$\mathbb{D}(\underline{\boldsymbol{y}}) = \sigma^2 \begin{bmatrix} 1 & \rho k \\ \rho k & k^2 \end{bmatrix} \quad \text{with} \quad \sigma_{y_1} = \sigma \quad \text{and} \quad \sigma_{y_2} = k\sigma\,, \tag{2.72}$$

such that the two constraints

$$\theta = \widehat{\theta} = (\mathsf{X}^\mathsf{T}\Sigma_{yy}^{-1}\mathsf{X})^{-1}\,\mathsf{X}^\mathsf{T}\Sigma_{yy}^{-1}\boldsymbol{y}\,, \tag{2.73}$$

$$v = \sigma^2_{\widehat{\theta}} = (\mathsf{X}^\mathsf{T}\Sigma_{yy}^{-1}\mathsf{X})^{-1}\,. \tag{2.74}$$

This are two constraints for the three not yet specified parameters σ, k, and ρ.

Explicitely, we obtain

$$\mathsf{W}_{yy} = \frac{1}{k^2\sigma^2(1-\rho^2)} \begin{bmatrix} k^2 & -\rho k \\ -\rho k & 1 \end{bmatrix} \tag{2.75}$$

and

$$\widehat{\theta} = \frac{l_2 + k^2\, l_1 - k\, l_1\, \rho - k\, l_2 \rho}{k^2 - 2\rho k + 1} = \frac{k(k-\rho)l_1 - (k\rho - 1)l_2}{k^2 - 2\rho k + 1} \tag{2.76}$$

$$\widehat{\sigma}^2_{\widehat{\theta}} = \frac{k^2(1-\rho^2)}{k^2 - 2\rho k + 1}\,\sigma^2 \tag{2.77}$$

From the two constraints

$$\hat{\theta} = \theta \quad \text{and} \quad \hat{\sigma}^2_{\hat{\theta}} = v \tag{2.78}$$

we obtain the two parameters σ^2 and ρ as a function of k and the given observations:

$$\begin{aligned}
\sigma^2 &= \frac{v\ (l_1 - l_2)\ (l_1 + l_2 - 2\,x)}{k^2\,{l_1}^2 - 2\,k^2\,l_1\,x + k^2\,x^2 - {l_2}^2 + 2\,l_2\,x - x^2} && (2.79)\\
&= \frac{v\ (l_1 - l_2)\ (l_1 + l_2 - 2\,x)}{((l_2 - x) + kl_1 - kx)(-(l_2 - x) + kl_1 - kx)} && (2.80)\\
\rho &= \frac{l_2 - x + k^2\,l_1 - k^2\,x}{k\ (l_1 + l_2 - 2\,x)} && (2.81)\\
&= \frac{(l_2 - k^2 l_1) + (1 - k^2)x}{k\ (l_1 + l_2 - 2\,x)} && (2.82)
\end{aligned}$$

For the special case

$$l_1 = 1\,, \quad l_2 = 0\,, \quad x = -1 \tag{2.83}$$

we obtain

$$\sigma^2 = \frac{3}{4k^2 - 1} v \quad \text{and} \quad \rho = \frac{2k^2 + 1}{3k} \quad \text{for} \quad k \in (0.5, 1) \tag{2.84}$$

2.7.1.2 A generalizable solution

We use the following Fig. 2.3, assuming $\boldsymbol{x} = X = \mathbf{1}_2$, and we observe the following:

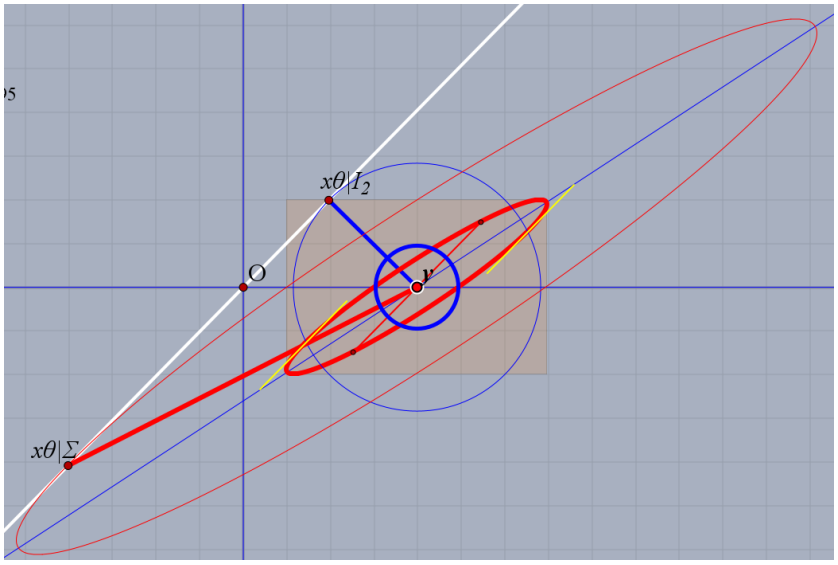

Figure 2.3: The generalized mean

1. The observed point $\boldsymbol{y}$ is slantly projected to $\boldsymbol{x}\hat{\theta}|\boldsymbol{\Sigma}$ on the line $\boldsymbol{x}\theta$.
2. The length of radius of the standard ellipse parallel to the line leads to the standard interval of the estimated point $\boldsymbol{x}\hat{\theta}$, in the figure half of length of the yellow tangent segment.
3. The length of the conjugate diameter is irrelevant for both, the position and the standard deviation of the estimate.

Hence we can specify the set of covariance matrices by mapping the reference covariance matrix $\boldsymbol{\Sigma}_0 = \boldsymbol{I}_2$ to $\boldsymbol{\Sigma}$ by applying the mapping the two unit vectors $\boldsymbol{e}_i, i = 1, 2$ to the tow

conjugate diameters $\boldsymbol{d}_i, i = 1, 2$ of the standard ellipse of $\boldsymbol{\Sigma}$. The two conjugate diameters are

$$\boldsymbol{d}_1 = \frac{\boldsymbol{x}}{|\boldsymbol{x}|}\sigma \quad \text{and} \quad \boldsymbol{d}_2(f) = \frac{\boldsymbol{x}\theta - \boldsymbol{y}}{|\boldsymbol{x}\theta - \boldsymbol{y}|}f \text{ for some arbitrary standard deviation} f > 0\,. \tag{2.85}$$

Hence we obtain the set of covariance matrices, parametrized by f from

$$\begin{aligned}
\boldsymbol{\Sigma}(f) &= [\boldsymbol{d}_1\ \boldsymbol{d}_2(f)] \begin{bmatrix} \boldsymbol{d}_1^\mathsf{T} \\ \boldsymbol{d}_2^\mathsf{T}(f) \end{bmatrix} = \boldsymbol{d}_1\boldsymbol{d}_1^\mathsf{T} + \boldsymbol{d}_2(f)\boldsymbol{d}_2^\mathsf{T}(f) && (2.86)\\
&= \frac{\boldsymbol{x}\boldsymbol{x}^\mathsf{T}}{\boldsymbol{x}^\mathsf{T}\boldsymbol{x}}\sigma^2 + \frac{(\boldsymbol{x}\theta - \boldsymbol{y})(\boldsymbol{x}\theta - \boldsymbol{y})^\mathsf{T}}{(\boldsymbol{x}\theta - \boldsymbol{y})^\mathsf{T}(\boldsymbol{x}\theta - \boldsymbol{y})}f^2 && (2.87)\\
&= \frac{\boldsymbol{x}\boldsymbol{x}^\mathsf{T}}{\boldsymbol{x}^\mathsf{T}\boldsymbol{x}}\sigma^2 + \frac{\boldsymbol{r}\boldsymbol{r}^\mathsf{T}}{\boldsymbol{r}^\mathsf{T}\boldsymbol{r}}f^2 \quad \text{with} \quad \boldsymbol{r} = \boldsymbol{x}\theta - \boldsymbol{y} && (2.88)
\end{aligned}$$

3 The Mean of Correlated Observations

For uncorrelated observations the accuracy of the mean increases with the number of observations. In case they are correlated, there is an upper limit for the accuracy. The note analyses the situation for constant correlation and for exponentially decaying correlation, autoregressive noise.

3.1 Preface . . . 37
3.2 Goal . . . 37
3.3 Random constant bias . . . 38
3.3.1 The model . . . 38
3.3.2 The solution . . . 38
3.3.3 Alternative derivation . . . 39
3.3.4 Covariance of arithmetic mean with correlated observations . . . 40
3.3.5 Using a more general covariance matrix . . . 40
3.4 Random autoregressive noise . . . 41
3.4.1 The model . . . 41
3.4.2 The solution . . . 41
3.5 The general case . . . 42

3.1 Preface

The arithmetic mean in many cases can be used as a proxy for a more general estimation problem. Here, we analyse the effect of correlations onto the accuracy of the estimated mean. The Note 4 generalizes the results.

3.2 Goal

We derive the precision of the correlated mean, by generatively model the observed values as a mean value which additively is distorted by a random effect with zero mean. We discuss two cases:

1. The noise in the measurements $\underline{y}_i$ consists of a uncorrelated part $\underline{d}_i$ and a correlated part $\underline{b}$, where the correlated part $\underline{b}$ is describable as a noisy bias. Namely, we have:

$$\underline{y}_i = \mu + \underline{b} + \underline{d}_i\,, \quad i = 1, ..., N \quad \text{with} \quad \underline{d}_i \sim \mathcal{M}(0, \sigma_d^2) \quad \text{and} \quad \underline{b} \sim \mathcal{M}(0, \sigma_b^2) \tag{3.1}$$

   We obtain the following result:

   (a) The estimated mean is independent of the bias

$$\hat{\mu} = \frac{1}{N} \sum_n y_n\,. \tag{3.2}$$

   (b) The variance of the estimated mean cannot become smaller than the variance of the bias

$$\sigma_{\hat{\mu}}^2 = \frac{\sigma_d^2}{N} + \sigma_b^2\,. \tag{3.3}$$

2. The noise is an autoregressive process of first order AR(1), namely

$$\underline{y}_i = \mu + \underline{b}_i\,, \quad n = 1, ..., N \tag{3.4}$$

and the AR(1)-process with parameter a

$$b_i = ab_{i-1} + \underline{e}_i \quad \text{with} \quad n > 1 \tag{3.5}$$

starting with

$$\underline{e}_i \sim \mathcal{M}(0, \sigma_e^2) \quad \text{and} \quad \underline{e}_1 = \mathcal{M}\left(0, \frac{\sigma_e^2}{1-a^2}\right) \tag{3.6}$$

The variance of the estimated mean is

$$\sigma_{\hat{\mu}}^2 = \frac{1+a}{1+(1-2/N)a}\frac{\sigma_e^2}{N} \quad \text{with} \quad |a| < 1\,. \tag{3.7}$$

3.3 Random constant bias

3.3.1 The model

We can write the generative model as

$$\underline{\boldsymbol{y}} = \mathbf{1}\mu + \mathbf{1}\underline{b} + \underline{\boldsymbol{n}} \tag{3.8}$$

leading to the covariance matrix

$$\boldsymbol{\Sigma}_{yy} = \sigma_d^2 \boldsymbol{I}_N + \mathbf{1}\mathbf{1}^\mathsf{T}\sigma_b^2\,. \tag{3.9}$$

Thus, the observed values have the variance and covariance

$$\sigma_{y_i}^2 = \sigma_d^2 + \sigma_b^2 \quad \text{and} \quad \sigma_{y_i y_j} = \sigma_b^2 \text{ for } i \neq j \tag{3.10}$$

hence have correlation coefficient

$$\rho_{ij} = \frac{\sigma_b^2}{\delta_{ij}\sigma_d^2 + \sigma_b^2} > 0 \text{ for } i \neq j\,. \tag{3.11}$$

Therefore, with the common correlation coefficient

$$\rho = \frac{\sigma_b^2}{\sigma_d^2 + \sigma_b^2} \text{ for } i \neq j\,, \tag{3.12}$$

the covariance matrix explicitly reads

$$\boldsymbol{\Sigma}_{yy} = (\sigma_d^2 + \sigma_b^2)\begin{bmatrix} 1 & \dots & \rho & \dots & \rho \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \rho & \dots & 1 & \dots & \rho \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \rho & \dots & \rho & \dots & 1 \end{bmatrix}. \tag{3.13}$$

3.3.2 The solution

The Gauss-Markov model reads

$$\boldsymbol{y} + \boldsymbol{v} = \mathbf{1}\mu \quad \text{with} \quad \boldsymbol{\Sigma}_{ll} = \sigma_d^2 \boldsymbol{I}_N + \sigma_b^2 \mathbf{1}\mathbf{1}^\mathsf{T} \tag{3.14}$$

The weight matrix of the observations has the structure

$$\boldsymbol{W}_{ll} = a\boldsymbol{I}_N + b\mathbf{1}\mathbf{1}^\mathsf{T}\,. \tag{3.15}$$

Therefore, we can determine a and b from

$$(\sigma_d^2 \mathsf{I} + \sigma_b^2 \mathbf{1}\mathbf{1}^\mathsf{T})(a\mathsf{I} + b\mathbf{1}\mathbf{1}^\mathsf{T}) = \underbrace{a\sigma_d^2}_{=1}\mathsf{I} + \underbrace{(a\sigma_b^2 + b\sigma_d^2 + bN\sigma_b^2)}_{!=0}\mathbf{1}\mathbf{1}^\mathsf{T}. \tag{3.16}$$

We obtain

$$a = \frac{1}{\sigma_d^2} \quad \text{and} \quad b = -\frac{a\sigma_b^2}{\sigma_d^2 + N\sigma_b^2} = -\frac{\sigma_b^2}{\sigma_d^2(\sigma_d^2 + N\sigma_b^2)} \tag{3.17}$$

Therefore the weight matrix is

$$\mathsf{W}_{yy} = \frac{1}{\sigma_d^2}\mathsf{I}_N - \frac{\sigma_b^2}{\sigma_d^2(\sigma_d^2 + N\sigma_b^2)}\mathbf{1}\mathbf{1}^\mathsf{T} = \frac{1}{\sigma_d^2}\left(\mathsf{I} - \frac{\sigma_b^2}{\sigma_d^2 + N\sigma_b^2}\mathbf{1}\mathbf{1}^\mathsf{T}\right) \tag{3.18}$$

The normal equation system is

$$\mathsf{N}\hat{\boldsymbol{\theta}} = \boldsymbol{n} \tag{3.19}$$

with

$$\mathsf{N} = \mathbf{1}^\mathsf{T}\mathsf{W}_{ll}\mathbf{1} = \frac{N}{\sigma_d^2} - \frac{N^2\sigma_b^2}{\sigma_d^2(\sigma_d^2 + N\sigma^2)} = \frac{N}{\sigma_d^2}\left(1 - \frac{N\sigma_b^2}{\sigma_d^2 + N\sigma_b^2}\right) \tag{3.20}$$

and

$$\boldsymbol{n} = \mathbf{1}^\mathsf{T}\mathsf{W}_{ll}\,\boldsymbol{y} = \frac{1}{\sigma_d^2}\left(\mathbf{1}^\mathsf{T} - \frac{N\sigma_b^2}{\sigma_d^2 + N\sigma_b^2}\mathbf{1}^\mathsf{T}\right)\boldsymbol{y} = \frac{1}{\sigma_d^2}\left(1 - \frac{N\sigma_b^2}{\sigma_d^2 + N\sigma^2}\right)\sum_n y_n\,. \tag{3.21}$$

The solution for the mean is

$$\hat{\mu} = \frac{1}{N}\sum_n y_n\,. \tag{3.22}$$

Hence we have the result: **The correlated arithmetic mean is independent on the correlation coefficient.**

The variance of the estimated mean is

$$\sigma_{\hat{\mu}}^2 = \frac{\sigma_d^2}{N}\frac{\sigma_d^2 + N\sigma_b^2}{\sigma_d^2} = \frac{\sigma_d^2}{N} + \sigma_b^2 \tag{3.23}$$

Hence we have the result: **The variance of the estimated mean of correlated observations diminishes with increasing N but cannot be smaller than the variance of the bias.** Hence, in case the variance of the bias is much larger than the variance of the noise, the variance of the mean is close to the variance of the bias.

3.3.3 Alternative derivation

We assume the model

$$\begin{bmatrix} \boldsymbol{y} \\ b_0 \end{bmatrix} + \begin{bmatrix} \boldsymbol{v} \\ v_{b_0} \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{1} \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \mu \\ b \end{bmatrix} \quad \text{with} \quad \mathbb{D}\left(\begin{bmatrix} \boldsymbol{y} \\ \underline{b}_0 \end{bmatrix}\right) = \begin{bmatrix} \sigma_d^2\mathsf{I} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & \sigma_b^2 \end{bmatrix} \tag{3.24}$$

with $b_0 = 0$, since we assumed $\underline{b} \sim \mathcal{M}(0, \sigma_b^2)$. The normal equation matrix

$$\mathsf{N} = \begin{bmatrix} \mathbf{1}^\mathsf{T} & 0 \\ \mathbf{1}^\mathsf{T} & 1 \end{bmatrix}\begin{bmatrix} w_d\mathsf{I} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & w_b \end{bmatrix}\begin{bmatrix} \mathbf{1} & \mathbf{1} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} w_d N & w_d N \\ w_d N & w_d N + w_b \end{bmatrix} \tag{3.25}$$

with its inverse

$$\mathsf{N}^{-1} = \frac{1}{N w_d w_b}\begin{bmatrix} w_d N + w_b & -w_d N \\ -w_d N & w_d N \end{bmatrix} \tag{3.26}$$

Hence the variances of the estimate $\hat{\mu}$ is

$$\sigma_{\hat{\mu}}^2 = \frac{\sigma_d^2}{N} + \sigma^2\,, \tag{3.27}$$

as above.

We also can directly determine the variance of $\hat{\mu}$ using the Schur-complement of N in (3.25):

$$\sigma_{\hat{\mu}}^2 = \left(w_d N - \frac{w_d^2 N^2}{w_d N + w_b} \right)^{-1} \tag{3.28}$$

$$= \left(\frac{w_d^2 N^2 + w_d w_b N - w_d^2 N^2}{w_d N + w_b} \right)^{-1} = \frac{\sigma_d^2}{N} + \sigma_b^2 \,. \tag{3.29}$$

3.3.4 Covariance of arithmetic mean with correlated observations

The simple arithmetic mean assumes $\Sigma_0 := \sigma_d^2 I$. Then the estimate is

$$\hat{\mu} = \frac{\mathbf{1}^\mathsf{T}}{N} \boldsymbol{y} \,, \tag{3.30}$$

with covariance matrix, assuming Σ_0 holds

$$\sigma_{\hat{\mu}_0}^2 = \frac{\sigma_d^2}{N} \,. \tag{3.31}$$

If this arithmetic mean is taken, but the actual covariance matrix is $\sigma_d^2 I_N + \sigma_b^2 \mathbf{1}\mathbf{1}^\mathsf{T}$ variance propagation of (3.30) yields

$$\sigma_{\hat{\mu}}^2 = \frac{\mathbf{1}^\mathsf{T}}{N} (\sigma_d^2 I_N + \sigma_b^2 \mathbf{1}\mathbf{1}^\mathsf{T}) \frac{\mathbf{1}}{N} \tag{3.32}$$

This can be simplified to

$$\sigma_{\hat{\mu}}^2 = \frac{\sigma_d^2}{N} + \sigma_b^2 \,, \tag{3.33}$$

which coinsides with (3.23).

3.3.5 Using a more general covariance matrix

The observations up to now have been assumed to be positively correlated, see (3.12)

$$\rho = \frac{\sigma_b^2}{\sigma_d^2 + \sigma_b^2} \quad \text{or} \quad \frac{\sigma_b^2}{\sigma_d^2} = \frac{\rho}{1-\rho} \,. \tag{3.34}$$

However, they also may have negative correlation. Of course, this then cannot be explained by a stochastic bias term anymore.

Therefore we assume

$$\Sigma_{yy} = \sigma_l^2 \begin{bmatrix} 1 & \dots & \rho & \dots & \rho \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \rho & \dots & 1 & \dots & \rho \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \rho & \dots & \rho & \dots & 1 \end{bmatrix} = \sigma_l^2 (1-\rho) I_N + \sigma_l^2 \rho \mathbf{1}\mathbf{1}^\mathsf{T} \,. \tag{3.35}$$

If $\rho < 0$ this is implicitly assuming $\sigma_b^2 < 0$. Therefore, we use the derivation above, which is valid also for $\sigma_b^2 < 0$. We now realize, that the correlation cannot have an arbitrary negative value ≥ -1, since the variance (3.23) of the mean needs to be positive. This leads to the following constraint, first formally

$$\frac{\sigma_d^2}{N} \geq \sigma_b^2 \quad \text{or} \quad \frac{1}{N} \geq \frac{\sigma_b^2}{\sigma_d^2} \tag{3.36}$$

then using (3.34)

$$\frac{1}{N} \geq \frac{\rho}{1-\rho} \quad \text{or} \quad \frac{\rho}{\rho-1} \geq \frac{-1}{N}\,, \tag{3.37}$$

which finally leads to a constraint on the correlation coefficient

$$\rho \geq \frac{-1}{N-1}\,. \tag{3.38}$$

For example: for two observations $N = 2$ the correlation coefficient may be arbitrary in the range $[-1, +1]$, but for three observations $N = 3$ the correlation coefficient needs to be larger than -50%.

3.4 Random autoregressive noise

3.4.1 The model

The N observations result from

$$\underline{\boldsymbol{y}} = \mathbf{1}\mu + \underline{\boldsymbol{b}} \tag{3.39}$$

with the following covariance matrices

$$\mathbb{D}(\underline{\boldsymbol{b}}) = \boldsymbol{\Sigma}_{bb} = [\Sigma_{b_i b_j}] = \frac{\sigma_e^2}{1-a^2}\left[a^{|i-j|}\right] \quad \text{with} \quad |a| < 1\,. \tag{3.40}$$

Hence the covariance matrix of the noise is

$$\boldsymbol{\Sigma}_{yy} = \boldsymbol{\Sigma}_{bb} \tag{3.41}$$

We have the inverse of the covariance matrix of the bias, which is a tridiagonal matrix:

$$\boldsymbol{W}_{yy} = \boldsymbol{\Sigma}_{bb}^{-1} = \frac{1}{\sigma_e^2}\,\mathrm{Tri}[1, 1+a^2, ..., 1+a^2, 1][-a, ..., -a] \tag{3.42}$$

3.4.2 The solution

The Gauss–Markov model reads as

$$\hat{\boldsymbol{y}} + \boldsymbol{v} = \mathbf{1}\mu \quad \text{with} \quad \boldsymbol{\Sigma}_{yy} = \boldsymbol{\Sigma}_{bb} = \boldsymbol{W}_{bb}^{-1}\,. \tag{3.43}$$

The normal equation system now is

$$\boldsymbol{N}\hat{\boldsymbol{\theta}} = \boldsymbol{n} \tag{3.44}$$

with (canceling the common factor $\sigma_e^2/(1-a)$)

$$\boldsymbol{N} = \mathbf{1}^\mathsf{T}\boldsymbol{W}_{yy}\mathbf{1}\,\frac{\sigma_e^2}{1-a} = \frac{2 + (N-2)(1+a^2) - 2(N-1)a}{1-a} = (N - (N-2)a) \tag{3.45}$$

and

$$\boldsymbol{n} = \mathbf{1}^\mathsf{T}\boldsymbol{W}_{yy}\boldsymbol{x}\,\frac{\sigma_e^2}{1-a} = \sum_{n=1}^{N} x_i - a\sum_{i=2}^{N-1} x_i\,. \tag{3.46}$$

The variance of the estimated mean is

$$\sigma_{\hat{\mu}}^2 = \frac{1}{(1-a)(N-(N-2)a)}\sigma_e^2 \tag{3.47}$$

which can be rewritten as

$$\sigma_{\hat{\mu}}^2 = \frac{1}{(1-a)(1-(1-2/N)a)}\,\frac{\sigma_e^2}{N}\,. \tag{3.48}$$

We have the following limits.

- For large N we achieve
$$\lim_{N\to\infty} \sigma^2_{\hat{\mu}} = \frac{1}{1-a^2}\frac{\sigma_e^2}{N}\,, \tag{3.49}$$
thus the variance is larger by a factor $1/(1-a^2)$ compared to the uncorrelated mean.
- For $a = 0$ we obtain the result of the uncorrelated mean.
- For $a = 1$, the noise process is semi-stationary, and we obtain
$$\lim_{a\to 1} \sigma^2_{\hat{\mu}} = \infty \tag{3.50}$$
independent on the number of observations.

3.5 The general case

The situation of the mean with constant correlation is a special case discussed in Rao (1967) in Lemma 5a. The estimated parameters of the model ($\boldsymbol{y} = X\boldsymbol{\theta}, \Sigma$) are the same if instead of Σ the covariance matrix $\Sigma + X\boldsymbol{\Gamma}X^\mathsf{T}$ is used. In our case we used $\boldsymbol{\Gamma} = \sigma^2$. The efficiency of the estimate, thought being the same, is reduced due to the correlations induced by $\boldsymbol{b}$.

Generalizing (3.24), we use the model, assuming $\boldsymbol{\Gamma} = \Sigma_{bb}$,
$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{b}_0 \end{bmatrix} + \begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{v}_b \end{bmatrix} = \begin{bmatrix} X & X \\ 0^\mathsf{T} & I \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{b} \end{bmatrix} \quad \text{with} \quad \mathbb{D}\left(\begin{bmatrix} \boldsymbol{y} \\ \underline{\boldsymbol{b}}_0 \end{bmatrix} \right) = \begin{bmatrix} \Sigma_{yy} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & \Sigma_{bb} \end{bmatrix} \tag{3.51}$$
again assuming $\boldsymbol{b}_0 = \mathbf{0}$. The normal equation matrix is
$$N = \begin{bmatrix} M & M \\ M & M + W_{bb} \end{bmatrix} \tag{3.52}$$
$$= \begin{bmatrix} X^\mathsf{T} & 0 \\ X^\mathsf{T} & I \end{bmatrix} \begin{bmatrix} W_{yy} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & W_{bb} \end{bmatrix} \begin{bmatrix} X & X \\ \mathbf{0} & I \end{bmatrix} \tag{3.53}$$
$$= \begin{bmatrix} X^\mathsf{T}W_{yy}X & X^\mathsf{T}W_{yy}X \\ X^\mathsf{T}W_{yy}X & X^\mathsf{T}W_{yy}X + W_{bb} \end{bmatrix} \tag{3.54}$$
with its inverse
$$\Sigma_{\hat{p}\hat{p}} = N^{-1} = \begin{bmatrix} \Sigma_{\hat{\mu}\hat{\mu}} & \Sigma_{\hat{\mu}\hat{b}} \\ \Sigma_{\hat{b}\hat{\mu}} & \Sigma_{\hat{b}\hat{b}} \end{bmatrix}. \tag{3.55}$$
Now we use the inverse of Schur-complement,
$$\Sigma_{\hat{\mu}\hat{\mu}} = (M - M(M + W_{bb})^{-1}M)^{-1}\,, \tag{3.56}$$
and the Woodbury identity,
$$(A + CBC^\mathsf{T})^{-1} = A^{-1} - A^{-1}C(B^{-1} + C^\mathsf{T}A^{-1}C)^{-1}C^\mathsf{T}A^{-1}\,, \tag{3.57}$$
with $A = M^{-1}$, $B = \Sigma_{bb}$ and $C = I$, and obtain
$$\Sigma_{\hat{\mu}\hat{\mu}} = M^{-1} + \Sigma_{bb} \tag{3.58}$$
hence,
$$\Sigma_{\hat{\mu}\hat{\mu}} = (X^\mathsf{T}\Sigma_{yy}^{-1}X)^{-1} + \Sigma_{bb}\,. \tag{3.59}$$
The prior, influencing all parameters the same way, leads to an increase of the covariance matrix.

Again, the result, using $\boldsymbol{\Gamma} = \Sigma_{bb}$, also holds if $\boldsymbol{\Gamma} < \mathbf{0}$, but only if
$$(X^\mathsf{T}\Sigma_{yy}^{-1}X)^{-1} \geq \Sigma_{bb} \tag{3.60}$$

4 Accuracy of the Mean when using a Wrong Covariance Matrix

Suboptimal, i.e., approximate solutions often are used or needed when estimating parameters. One of such simplifications refers to the stochastical model, especially the covariance matrix of the observations, which often is assume to be a multiple of a unit matrix, implicitly assuming all observations have the same weight and are mutually uncorrelated. This note provides the general relation between the accuracy of the estimated parameters when using an approximate covariance matrix and exemplifies this using the mean of repeated observations.

4.1 Preface . . . 43
4.2 Summary . . . 43
4.3 Problem . . . 44
4.4 The accuracy of the approximate solution . . . 44
4.5 The weighted arithmetic mean . . . 44
4.5.1 The effect of using equal weights . . . 45
4.5.2 Modeling the weights using the Gamma-distribution . . . 45
4.6 An example . . . 46

4.1 Preface

The arithmetic mean in many cases can be used as a proxy for a more general estimation problem. Here, we analyse the loss in accuracy of the estimated mean when using a wrong covariance matrix. The note generalizes the results from Note **??**.

4.2 Summary

The note shows the effect of using a wrong covariance matrix when estimating parameters. Especially we obtain the following results for estimating the mean from N values l_n:

1. If the mean is estimated assuming, that all values have the same weight $w = 1/\sigma^2$, thus $\widehat{\theta} = \sum_n l_n/N$, but the values really have individual weights

$$w_n = \frac{1}{\tilde{\sigma}_n^2} \tag{4.1}$$

then the variance of the approximately determined mean is larger by a factor

$$\lambda = \overline{\tilde{\sigma}_n^2} \cdot \overline{1/\tilde{\sigma}_n^2} = \frac{\mu_{\tilde{\sigma}^2}^{(a)}}{\mu_{\tilde{\sigma}^2}^{(h)}} \tag{4.2}$$

thus the ratio of the arithmetic mean and the harmonic mean of the variances. The factor λ only is 1, in case the variances are identical for all values l_n.

2. In the special case, that the weights are assumed to be randomly taken from a Gamma distribution and their relative variation is

$$c = \frac{\sigma_w}{\mu_w} < 1\,, \tag{4.3}$$

then the factor is given by

$$\lambda = \frac{1}{1-c^2}\,. \tag{4.4}$$

If $c \geq 1$ the factor is not limited.

4.3 Problem

If the estimation is performed in a Gauss-Markov model $\mathbb{E}(\underline{\boldsymbol{y}}) = \boldsymbol{X\theta}$ with $\boldsymbol{\Sigma}_{yy} = \boldsymbol{\Sigma}$ but the true covariance matrix of the observations is $\widetilde{\boldsymbol{\Sigma}_{yy}} = \widetilde{\boldsymbol{\Sigma}}$, then the covariance matrix of the estimated parameters is

$$\boldsymbol{\Sigma}_{\hat{\theta}\hat{\theta}} = (\boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\ \widetilde{\boldsymbol{\Sigma}}\ \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\,, \tag{4.5}$$

which follows from $\hat{\underline{\boldsymbol{\theta}}} = (\boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\underline{\boldsymbol{y}} - \boldsymbol{x})$. Observe, only if $\boldsymbol{\Sigma} = \widetilde{\boldsymbol{\Sigma}}$ do we obtain the classical result

$$\widetilde{\boldsymbol{\Sigma}}_{\hat{\theta}\hat{\theta}} = (\boldsymbol{X}^\mathsf{T}\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\,. \tag{4.6}$$

4.4 The accuracy of the approximate solution

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

$$\lambda(\boldsymbol{\Sigma}_{\hat{\theta}\hat{\theta}}\widetilde{\boldsymbol{\Sigma}}^{-1}_{\hat{\theta}\hat{\theta}}) = \lambda\left((\boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\ \widetilde{\boldsymbol{\Sigma}}\ \boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}^\mathsf{T}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\ \boldsymbol{X}^\mathsf{T}\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}\right)\,. \tag{4.7}$$

Equations (4.5) and (4.7) can be used to investigate the effect of choosing a simplified stochastical model, e.g., when using $\boldsymbol{\Sigma}_{yy} = \sigma^2\boldsymbol{I}_N$ instead of $\widetilde{\boldsymbol{\Sigma}}$.

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

$$\lambda(\boldsymbol{\Sigma}_{\hat{\theta}\hat{\theta}}\widetilde{\boldsymbol{\Sigma}}^{-1}_{\hat{\theta}\hat{\theta}}) = \lambda\left((\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^{-1}\boldsymbol{X}^\mathsf{T}\ \widetilde{\boldsymbol{\Sigma}}\ \boldsymbol{X}(\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^{-1}\ \boldsymbol{X}^\mathsf{T}\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}\right)\,, \tag{4.8}$$

obviously, independent on the scaling of the covariance matrices.

With the hat matrix

$$\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^\mathsf{T}\boldsymbol{X})^{-1}\ \boldsymbol{X}^\mathsf{T} \tag{4.9}$$

this is equivalent to analysing

$$\lambda(\boldsymbol{\Sigma}_{\hat{\theta}\hat{\theta}}\widetilde{\boldsymbol{\Sigma}}^{-1}_{\hat{\theta}\hat{\theta}}) = \lambda(\boldsymbol{H}\widetilde{\boldsymbol{\Sigma}}\boldsymbol{H}\widetilde{\boldsymbol{\Sigma}}^{-1}) \geq 1\,, \tag{4.10}$$

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.

4.5 The weighted arithmetic mean

We want to investigate the effect of using a wrong covariance matrix in case of diagonal covariance matrices,

$$\widetilde{\boldsymbol{\Sigma}} = \mathrm{Diag}([\widetilde{\sigma_i^2}]) \quad \text{and} \quad \boldsymbol{\Sigma} = \boldsymbol{I}\,, \tag{4.11}$$

4.5.1 The effect of using equal weights

We start with a simple example, the weighted arithmetic mean of N observations. The design matrix for the arithmetic mean is

$$X = \mathbf{1}\,. \tag{4.12}$$

Then with $X^\mathsf{T}X = \mathbf{1}^\mathsf{T}\mathbf{1} = N$ Eq. (4.10) reduces to

$$\begin{aligned}
\lambda(\Sigma_{\hat{\theta}\hat{\theta}}\tilde{\Sigma}^{-1}_{\hat{\theta}\hat{\theta}}) &= \frac{\sigma^2_{\hat{\theta}}}{\tilde{\sigma}^2_{\hat{\theta}}} && (4.13)\\
&= \frac{1}{N^2}\lambda\left(\mathbf{1}\mathbf{1}^\mathsf{T}\tilde{\Sigma}\mathbf{1}\mathbf{1}^\mathsf{T}\tilde{\Sigma}^{-1}\right) && (4.14)\\
&= \frac{1}{N^2}\lambda\left(\mathbf{1}^\mathsf{T}\tilde{\Sigma}\mathbf{1}\cdot\mathbf{1}^\mathsf{T}\tilde{\Sigma}^{-1}\mathbf{1}\right) && (4.15)\\
&= \frac{1}{N^2}\mathrm{tr}\tilde{\Sigma}\cdot\mathrm{tr}\tilde{\Sigma}^{-1} && (4.16)\\
&= \frac{\sum_{n=1}^N \tilde{\sigma}^2_n}{N}\cdot\frac{\sum_{n=1}^N w_n}{N} && (4.17)\\
&= \overline{\tilde{\sigma}^2_n}\cdot\overline{w_n} && (4.18)\\
&= \frac{\mu^{(a)}_{\tilde{\sigma}^2}}{\mu^{(h)}_{\tilde{\sigma}^2}} \geq 1\,. && (4.19)
\end{aligned}$$

or the ratio of the arithmetic mean $\mu^{(a)}_{\tilde{\sigma}^2} = \overline{\tilde{\sigma}^2_n}$ and the harmonic mean $\mu^{(h)}_{\tilde{\sigma}^2} = \left(\overline{1/\tilde{\sigma}^2_n}\right)^{-1}$ of the variances or of the weights. This ratio always is larger than 1 except all variances are identical.

4.5.2 Modeling the weights using the Gamma-distribution

The Gamma-distribution is a useful model for the weights, since it is the conjugate prior for the precision $w = 1/\sigma^2$ of the Gaussian distribution with known mean.

Let the weights be Gamma distributed

$$\underline{w}_n \sim \mathrm{Gamma}(\alpha,\beta) = \mathrm{Gamma}(k,\theta) \tag{4.20}$$

where the two parametrizations are related by

$$k = \alpha \quad \text{and} \quad \theta = \frac{1}{\beta}\,. \tag{4.21}$$

The mean and the variance are given by

$$\mathbb{E}(\underline{w}_n) = \frac{\alpha}{\beta} = k\theta \quad \text{and} \quad \mathbb{V}(\underline{w}_n) = \frac{\alpha}{\beta^2} = k\theta^2\,. \tag{4.22}$$

So, given a mean weight μ_w and a variance of the weights σ^2_w we may choose the parameters

$$\alpha = \frac{\mu^2_w}{\sigma^2_w} \quad \text{and} \quad \beta = \frac{\sigma^2_w}{\mu_w}\,. \tag{4.23}$$

The inverse weights, thus the variances follow an inverse Gamma distribution

$$\underline{\sigma}^2_n \sim \mathrm{invGamma}(a,b) \tag{4.24}$$

with the same parameters. Their mean is

$$\mathbb{E}(\underline{\sigma}^2_n) = \frac{\beta}{\alpha-1} \quad \text{for } \alpha > 1 \quad \text{and} \quad \mathbb{V}(\underline{\sigma}^2_n) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}\,. \tag{4.25}$$

For values $\alpha \leq 1$ the inverse Gamma distribution has no finite mean, similar to the variance of the Cauchy distribution. This is plausible, since then the likelihood of small weights thus large variances is very high.

Hence the product of the means of the variances and the weights is given by

$$\lambda = \frac{\alpha}{\beta} \cdot \frac{\beta}{\alpha - 1} = \frac{\alpha}{\alpha - 1} \geq 1 \tag{4.26}$$

So, in case the weights on an average are μ_w and have a standard deviation of $\sigma_w = c \cdot \mu_w$, thus

$$c = \frac{\sigma_w}{\mu_w}\,, \tag{4.27}$$

we obtain

$$\lambda = \frac{\sigma^2_{\hat{\theta}}}{\tilde{\sigma}^2_{\hat{\theta}}} = \frac{\frac{\mu_w^2}{\sigma_w^2}}{\frac{\mu_w^2}{\sigma_w^2} - 1} = \frac{1}{1 - c^2}\,. \tag{4.28}$$

For values $c \geq 1$ the ratio of the variances is unlimited.

4.6 An example

We take as an example the mean of two points in the plane, and compare the arithmetic mean with the statistically optimal mean.

The Fig. 4.1 4.1 shows the arithmetic mean and the weighted mean (centroids) of two

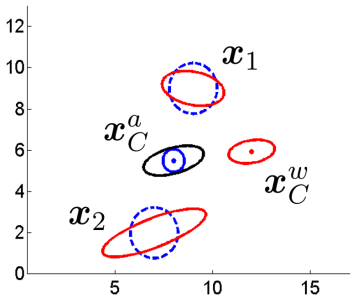

Figure 4.1: Simple mean $\boldsymbol{x}_C^a$ and weighted mean $\boldsymbol{x}_C^w$ of two points $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ with strongly anisotropic uncertainty (red standard ellipses). The centroid determined as weighted mean clearly lies outside the line joining the two points.

points. They are assumed to be mutually independent. Their uncertainty is different and anisotropic (red standard ellipses). The centroids result from the two models

$$\begin{bmatrix} \underline{\boldsymbol{x}}_1 \\ \underline{\boldsymbol{x}}_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} I_2 \\ I_2 \end{bmatrix} \boldsymbol{x}_C^a, \sigma^2 I_4 \right) \tag{4.29}$$

and

$$\begin{bmatrix} \underline{\boldsymbol{x}}_1 \\ \underline{\boldsymbol{x}}_1 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} I_2 \\ I_2 \end{bmatrix} \boldsymbol{x}_C^w, \mathrm{Diag}(\{\Sigma_{11}, \Sigma_{22}\}) \right)\,. \tag{4.30}$$

The variance σ^2 in model (4.29) was assumed to be the mean of the two variances in model (4.30), s. the two blue circles.

Explicitly, the centroids are

$$\hat{\boldsymbol{x}}_C^a = \frac{1}{2}(\boldsymbol{x}_1 + \boldsymbol{x}_2) \quad \text{und} \quad \hat{\boldsymbol{x}}_C^w = (\Sigma_{11}^{-1} + \Sigma_{22}^{-1})^{-1}(\Sigma_{11}^{-1}\boldsymbol{x}_1 + \Sigma_{22}^{-1}\boldsymbol{x}_2)\,. \tag{4.31}$$

The simple arithmetic mean lies in the middle of the two points $\boldsymbol{x}_1$ und $\boldsymbol{x}_2$ on the connecting line. The weighted mean, however, lies significantly off the connecting line. The

uncertainty of the two given points allows, that the centroid may be more easily shifted in the direction of the major axes of the standard ellipses.

The standard ellipses around represent centroids are

- the covariance matrix (blue circle) of the arithmetic mean, assuming the same isotropic accuracy (blue dashed circles) of the two points. It clearly overestimates its accuracy, compared to
- the covariance matrix (black ellipse) of the arithmetic mean, assuming the anisotropic accuracy (red circles around the points), and
- the covariance matrix (red ellipse) of the weighted mean, which is smaller than the accuracy of the arithmetic mean, when assuming the known uncertainty of the point during estimation.

Part II

Technical Notes on Estimation of Geometric Entities

5 Motions and their Uncertainty

We addresses the ambiguity of representing uncertain motions. We analyse the relation between an exponential representation with a homogeneous 4x4 matrix and the representation with a rotation matrix, also represented exponentially, and a translation vector. The rotation parts turns out to be identical, while the translation parts differ, why a transparent documentation of the motion representation is necessary. As a sideline, the note addresses the inversion, the concatenation, and the difference of uncertain rotations and uncertain motions.

5.1 Preface . . . 50
5.2 Motivation . . . 50
5.3 Overview . . . 52
5.4 Uncertain Rotations . . . 54
5.4.1 General setup . . . 54
5.4.2 Representing rotations . . . 55
5.4.3 Relations between the representations . . . 56
5.4.4 The rotation in exponential representation . . . 56
5.5 Uncertain Motions . . . 59
5.5.1 Representations . . . 59
5.5.2 Comparing the two representations . . . 60
5.5.3 The motion in exponential representation . . . 60
5.5.4 The motion in partially exponential representation . . . 62
5.5.5 Evaluating the covariance matrix of estimated motions . . . 64
5.6 Examples . . . 66
5.6.1 Estimating motion parameters . . . 66
5.6.2 Example for comparing absolute and relative poses in multi-view analysis . . . 68
5.7 MATLAB Software . . . 69
5.8 Appendix . . . 71
5.8.1 Epipolar constraint using motion matrices . . . 71
5.8.2 Differential relation between Euler angles and the exponential representation . . . 72
5.8.3 Adjoint motion matrix in exponential representation . . . 73
5.8.4 Adjoint motion in partially exponential representation . . . 74
5.8.5 Uncertain inverse rotation . . . 75
5.8.6 Uncertain inverse motion in exponential representation . . . 75
5.8.7 Uncertain inverse motion in partially exponential representation . . . 76
5.8.8 Uncertain concatenated motions in exponential representation . . . 76
5.8.9 Uncertain concatenated motions in partially exponential representation . . . 77
5.8.10 Uncertain relative motion in exponential representation . . . 78
5.8.11 Uncertain relative motion in partially exponential representation . . . 79

5.1 Preface

The note initially (2009) was motivated by the need to concatenate uncertain motions. Later, in 2017 an extension was motivated by the search for an error in a program for estimating the motion between two point clouds based on corresponding planar regions. The error turned out to be a conceptual one: the generation of the test data and the check of the estimated motions was inconsistent, since one used the exponential form of a motion, while the other used the exponential form of a rotation and the translation. For a detailed discussion see Solà et al. (2018).

5.2 Motivation

This note is motivated by a problem when handling uncertain 3D motions or poses: The two classical representations, the one what we call the *exponential* representation and the other what we call *partially exponential* representation, may both be used for estimating motions or poses, but lead to different covariance matrices of the translation component. The note aims at clarifying the mutual relations between the different representations.

The *exponential representation* of an uncertain motion with mean rotation R and mean translation $\boldsymbol{Z}$, exploits the Lie group structure of the noise component of the motion using what is called a twist vector $\underline{\boldsymbol{s}}$, which contains the noise components $\underline{\boldsymbol{r}}$ and $\underline{\boldsymbol{t}}$ for rotation and translation,[1] in the form

$$^{s}\underline{\mathsf{M}} = \exp(\mathsf{A}(\underline{\boldsymbol{s}}))\ \mathsf{M}\,, \tag{5.1}$$

with

$$\mathsf{M} = \begin{bmatrix} R & \boldsymbol{Z} \\ \boldsymbol{0}^{\mathsf{T}} & 1 \end{bmatrix}\,,\quad \boldsymbol{s} = \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{t} \end{bmatrix}\,,\quad \mathsf{A}(\boldsymbol{s}) = \begin{bmatrix} S(\boldsymbol{r}) & \boldsymbol{t} \\ \boldsymbol{0}^{\mathsf{T}} & 0 \end{bmatrix}\,, \tag{5.2}$$

and

$$S(\boldsymbol{r}) = \begin{bmatrix} 0 & -r_3 & r_2 \\ r_3 & 0 & -r_1 \\ -r_2 & r_1 & 0 \end{bmatrix}\,. \tag{5.3}$$

The matrix $\mathsf{A}(\boldsymbol{s})$ is close to zero, such that the motion matrix $\exp(\mathsf{A}(\underline{\boldsymbol{s}}))$ is close to I_4.

The *partially exponential representation* directly integrates the noise components $\underline{\boldsymbol{\rho}}$ and $\underline{\boldsymbol{\tau}}$ for a small rotation and translation

$$^{\zeta}\underline{\mathsf{M}} = \begin{bmatrix} \exp(S(\underline{\boldsymbol{\rho}}))R & \boldsymbol{Z} + \underline{\boldsymbol{\tau}} \\ \boldsymbol{0}^{\mathsf{T}} & 1 \end{bmatrix}\,, \tag{5.4}$$

into M thus only applies the exponential map to the noise component of the rotation. Again, since $\underline{\boldsymbol{\rho}}$ is small, the rotation matrix $\exp(S(\underline{\boldsymbol{\rho}}))$ is close to I_3.

The following example shows the effect of the different representations. Given are 100 random samples of an uncertain motion together with the true motion. From this sample we may obtain two covariance matrices Σ_1 and Σ_2 with the following vectors of standard deviations for the rotational and the translational component:

$$\boldsymbol{\sigma}_1 = \begin{bmatrix} 0.5109 \\ 0.4803 \\ 0.2760 \\ 0.4696 \\ 0.3324 \\ 0.3591 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\sigma}_2 = \begin{bmatrix} 0.5109 \\ 0.4803 \\ 0.2760 \\ 0.6125 \\ 0.6189 \\ 0.7433 \end{bmatrix}\,. \tag{5.5}$$

The rotation parameters have the same standard deviations, while the standard deviations of the translational component significantly differ. Without further information we cannot judge, which covariance matrix is the correct one. We would prefer the first one,

[1] Random variables are underscored.

since it shows smaller standard deviations for the translation component, and thus is more likely to be the Cramér-Rao lower bound for the uncertainty of the parameters. Actually, the motion matrices were simulated using the partially exponential representation and the covariance matrices Σ_1 and Σ_2 were derived from the sample assuming the partially exponential representation and the exponential representation, respectively. This demonstrates, the meaning of the two vectors $(\boldsymbol{r}, \boldsymbol{t})$ and $(\boldsymbol{\rho}, \boldsymbol{\tau})$ are different.

Both representations are useful, as the following examples demonstrate:

1. On one hand, concatenating uncertain motions appears to be easier with the exponential representation, where the (differential of the) twist vector $\boldsymbol{s}$ of the concatenated motion $\mathsf{M} = \mathsf{M}_2\mathsf{M}_1$ is given by

$$\mathrm{d}\boldsymbol{s} = \mathrm{Ad}(\mathsf{M}_2)\ \mathrm{d}\boldsymbol{s}_1 + \mathrm{d}\boldsymbol{s}_2 \quad \text{with} \quad \mathrm{Ad}(\mathsf{M}_2) = \begin{bmatrix} R_2 & 0 \\ S(\boldsymbol{Z}_2)R_2 & R_2 \end{bmatrix}, \tag{5.6}$$

   or

$$\mathrm{d}\boldsymbol{r} = R_2\ \mathrm{d}\boldsymbol{r}_1 + \mathrm{d}\boldsymbol{r}_2 \quad \text{and} \quad \mathrm{d}\boldsymbol{t} = R_2\ \mathrm{d}\boldsymbol{t}_1 + \mathrm{d}\boldsymbol{t}_2 + S(\boldsymbol{Z}_2)R_2\mathrm{d}\boldsymbol{r}_2\,. \tag{5.7}$$

   Observe, the matrix $\mathrm{Ad}(\mathsf{M}_2)$ only depends on one of the two motions. In contrast, the partially exponential representation yields the joined rotation and translation components

$$\mathrm{d}\boldsymbol{\rho} = R_2\ \mathrm{d}\boldsymbol{\rho}_1 + \mathrm{d}\boldsymbol{\rho}_2 \quad \text{and} \quad \mathrm{d}\boldsymbol{\tau} = R_2\ \mathrm{d}\boldsymbol{\tau}_1 + \mathrm{d}\boldsymbol{\tau}_2 - S(R_2^\mathsf{T}\boldsymbol{Z}_1)\mathrm{d}\boldsymbol{\rho}_2\,, \tag{5.8}$$

   which looks very similar. But the relation cannot be written using a matrix only depending on one of the two motions, which is a clear disadvantage when concatenating multiple motions.

2. On the other hand the epipolar constraint for two calibrated images using partially exponential representation directly refers to the uncertain rotation and translation component

$$\mathbf{x}'^\mathsf{T}\ S(\boldsymbol{Z} + \underline{\boldsymbol{\tau}})\ R(\underline{\boldsymbol{\rho}})R\ \mathbf{x}'' = 0 \tag{5.9}$$

   whereas with the exponential representation it is given by

$$\mathbf{x}'^\mathsf{T}[I_3 \mid 0]\ \mathrm{Ad}(\underline{\mathsf{M}})^{-\mathsf{T}} \begin{bmatrix} 0 \\ I_3 \end{bmatrix} \mathbf{x}'' = 0\,, \tag{5.10}$$

   with the adjoint motion matrix $\mathrm{Ad}(\mathsf{M})$ (5.73) or more explicitly by

$$\mathbf{x}'^\mathsf{T} S(R(\underline{\boldsymbol{r}})\boldsymbol{Z} + \underline{\boldsymbol{t}})\ R(\underline{\boldsymbol{r}})R\ \mathbf{x}'' = 0 \tag{5.11}$$

   which is more cumbersome to handle, see (5.60).

When estimating a motion matrix from observed points, lines or planes using a maximum likelihood approach we basically obtain three types of numbers, which can be checked, i.e., statistically tested, using simulated data, which should lead to the following statements: (a) there are no reasons to believe the *parameters* are biased, (b) there are no reasons to believe the *variance factor*[2] deviates from 1, and (c) there are no reasons to believe the *theoretical* covariance matrix differs from the empirical covariance matrix, see (Förstner and Wrobel, 2016, Sect. 4.6.8). The test on the parameters and the covariance matrix may be performed for rotations and translations separately. Depending on how the representation for the motion is chosen and how the empirical tests are realized, the parameters usually show no bias, the variance factor does not show a deviation from 1, the covariance matrix of the rotation parameters coincide but there may be discrepancies in the covariance matrix of the translation parameters.

[2] The variance factor measures the distance of the assumed model and the given data. It is Fisher distributed, if the model holds.

The dependency on the representation of motions or poses has a direct effect on (1) checking their covariance matrices empirically, either using real or synthetic data,[3] on (2) reporting covariance matrices for motions or poses, and on (3) using them in subsequent analysis steps.

This note especially we will show:

- The Lie group property of matrix groups can, with some slight modifications be applied to the definition and use of uncertain motions represented with the partially exponential representation.
- We will discuss the variance propagation for inverse, concatenated and relative motions.
- We derive the relations for rotations as most simple case, and for motions in the mentioned two representations.
- We give two examples: (1) for estimating motions from corresponding points, and (2) deriving relative motions from bundle adjustment taking the full covariance matrix of the resulting pose parameters into account.

Basic material on Lie groups for representing uncertain transformations has been collected by Eade (2014), however, the note does not provide proofs. The most recent paper on handling uncertain motions which are correlated is by **?**, which appears not to always give the most intuitive expressions. Both papers do not address the second representation with the pair $(\mathsf{R}, \boldsymbol{Z})$, only.

The note is organized as follows. We first give a summary of the relations, assuming the reader is acquainted with the basic concepts. Then we will provide the relations in more detail, first for rotations – as special motions –, and then for the two types of motion representations. We will compare the two motion representations and, finally, give examples for estimating motions and analysing the relative pose derived from a bundle adjustment. The proofs will be found in an appendix.

On notation. Matrices are written in capital sans-serif letters, homogeneous 4×4 matrices in upright letters, 3×3 matrices in slanted letters, such as M, A and R, S. Vectors are written in boldface times, 3-vectors representing 3D points in upright, such as $\boldsymbol{Z}$ vectors representing (numerically), numerically small entities are written in small vectors, such as $\boldsymbol{r}$ or $\boldsymbol{m}$. Stochastic entities are underscored, such as a stochastic 3×3 matrix $\underline{\mathsf{R}}$ or a stochastic 3-vector $\underline{\boldsymbol{r}}$. Names of entities are written in calligraphic letters, in order to be able to express different representations, e.g., $\mathcal{M}(\mathsf{M})$ and $\mathcal{M}(\mathsf{R}, \boldsymbol{Z})$. If the entity is assumed to be uncertain we underscore its name, e.g., the uncertain motion may be defined as $\underline{\mathcal{M}}(\underline{\mathsf{M}})$. For clarity, we sometimes use the multiplication dot between matrices, e.g., in the expression $\mathrm{Ad}(\mathsf{M}) \cdot \boldsymbol{s}$, which is not the multiplication dot for the scalar multiplication in $\boldsymbol{a} \cdot \boldsymbol{b}$.

5.3 Overview

We assume the following notation for Lie groups, which in our context refer to groups of regular matrices:

- A Lie group $\mathcal{G}$ has elements $g, h \in \mathcal{G}$, an operation $f = g \circ h \in \mathcal{G}$ and an inverse element such that $g^{-1} \circ g = g \circ g^{-1} = e$, with the unit element e. The dimension n of the Lie group is the number of degrees of freedom for representing an element. An element of $\mathcal{G}$ also is called an action (rotation, motion), as it is meant to operate on a vector.

[3] The author, not being aware of the difference in the two representations, spent one month finding an error in his software on estimating a motion from corresponding planes, see (Förstner and Khoshelham, 2017)

In our case we discuss rotations $\mathsf{R} \in \mathrm{SO}(3)$ and motions $\mathsf{M} \in \mathrm{SE}(3)$, having dimension $n = 3$ and $n = 6$ respectively.

- The corresponding Lie algebra $\mathcal{g}$ spans the tangent space at the unit element, its elements are n-vectors $x \in \mathbb{R}^n$ or – equivalently – matrices $X = x^\wedge$ (read: "wedge")[4] linearly depending on x, y and having the same size as the elements of the Lie group. The inverse relation is $x = X^\vee$ used for deriving the n-vector from the corresponding matrix.
  In our case the elements are 3-vectors $\boldsymbol{r} \in \mathcal{g} = \mathbb{R}^3$, also called rotation vectors, and the – not necessarily small – 6-vectors $\boldsymbol{m} \in \mathcal{g} = \mathbb{R}^6$ also called twist vectors, concatenating the rotation and the translation components of the motion. As an example for the matrix X, we have the element $\mathsf{S}(\boldsymbol{r}) = \boldsymbol{r}^\wedge$, being the skew matrix of the rotation vector $\boldsymbol{r} \in g = \mathbb{R}^3$.
- The basic relation between the Lie algebra and the Lie group is the exponential map

$$\mathcal{g} \mapsto \mathcal{G} : g = \exp(x^\wedge) \tag{5.12}$$

  which describes the elements g of $\mathcal{G}$ around the unit element e.
  As an example, we have the exponential $\mathsf{R}(\boldsymbol{r}) = \exp(\mathsf{S}(\boldsymbol{r}))$, being the rotation matrix as element of $\mathcal{G} = \mathrm{SO}(3)$. The unit element $e \in \mathcal{G}$ of the rotation group $\mathcal{G} = \mathrm{SO}(3)$ here is the unit matrix $\mathsf{I}_3 = \exp(\mathsf{S}(\mathbf{0}))$ and corresponds to the 3-vector $\mathbf{0}$, i.e., $x = 0 \in \mathcal{g}$ in the Lie algebra $\mathcal{g} = so(3)$.
  If we write $\exp(x)$, where x is an element of the Lie algebra, we actually mean $\exp(x^\wedge)$:

$$\exp(x) := \exp(x^\wedge)\,. \tag{5.13}$$

The two tables 5.1 and 5.2 collect the main algebraic relations for rotations, and motions in exponential and partially exponential representation. They are derived and discussed more in detail in the next section. The collected relations are useful in the following situations:

- Representing rotations $\mathcal{R}(\mathsf{R})$ and motions $\mathcal{M}(\mathsf{M})$ (row 1).
- Generating uncertain rotations $\underline{\mathcal{R}}(\underline{\mathsf{R}})$ and uncertain motions $\underline{\mathcal{M}}(\underline{\mathsf{M}})$ (rows 6, with 2 and 3), assuming the small elements have mean 0 and some covariance matrix. Here the difference between the exponential representation $\underline{\mathcal{M}}({}^s\underline{\mathsf{M}})$ and the partially exponential representation $\underline{\mathcal{M}}({}^\zeta\underline{\mathsf{M}})$ become visible.
- Deriving the small left rotation or motion from a small right rotation or motion leading to the same uncertain rotation or motion (rows 4 and 5), e.g., in the form $\mathsf{M}(\mathrm{Ad}(\mathsf{M})\cdot\underline{\boldsymbol{s}})\ \mathsf{M} = \mathsf{M}(\underline{\boldsymbol{s}}_{\mathrm{ad}})\cdot\mathsf{M} = \mathsf{M}\cdot\mathsf{M}(\underline{\boldsymbol{s}})$, derived from the adjoint action $\exp(x_{\mathrm{ad}}) = g\ \exp(x)g^{-1}$. Observe, the adjoint matrix is not used in other relations of the partially exponential representation.
- Deriving small deviations between estimated and true rotations and motions (row 7) using $\mathsf{V}(\boldsymbol{r}) \approx \mathsf{I}_3$ and $\mathsf{R}(\mathrm{d}\boldsymbol{r}) \approx \mathsf{I}_3 + \mathsf{S}(\mathrm{d}\boldsymbol{r})$.
- Switching between the two motion representations (row 8, columns 3 and 4).
- Deriving the mean and covariance matrix of the inverse (rows 10 and 11).
- Deriving mean and the covariance matrix of the concatenation (rows 12 and 13), of two possibly correlated rotations or motions.
- Deriving mean and the covariance matrix of the relative rotation or motion (rows 14 and 15), of two possibly correlated rotations or motions.

Comparing the relations for the two motion representations in columns 3 and 4, we observe great similarities, partially identical relations. Specifically, the two differential motions $\mathrm{d}\boldsymbol{s}$

[4] The notation results from the outer product of two vectors, which in the special case of 3-vectors reduces to the cross product. Thus we have $\boldsymbol{x} \wedge \boldsymbol{y} = \boldsymbol{x}^\wedge\, \boldsymbol{y}$ equivalent to $\boldsymbol{x} \times \boldsymbol{y} = [\boldsymbol{x}]_\times \boldsymbol{y} = \mathsf{S}(\boldsymbol{x})\boldsymbol{y}$

of the exponential representation and $\mathrm{d}\boldsymbol{\zeta}$ of the partially exponential representation are related by a linear transformation. This can be interpreted as a change of the basis of the three axes in the tangent space of the Lie group which refer to the translation component. On the other hand, is it obvious, that the relations for the exponential representation are simpler and more mutually connected. As mentioned above, e.g., the concatenation of two differential motions (row 12, column 3) only uses the adjoint matrix $\mathrm{Ad}(\mathsf{M}_2)$ of the second motion, whereas the term $-S(R_2\boldsymbol{Z}_1)$ with the skew matrix in the expression for the translation component (row 4) depends on both motions.

	1 ↓ object \ $\mathcal{G}$ →	2 SO(3)	3 SE(3), $\boldsymbol{s}$	4 SE(3), $\boldsymbol{\zeta}$
1	action, group element $g \in \mathcal{G}$	R	$\mathsf{M} = \begin{bmatrix} R & \boldsymbol{Z} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}$	$\mathsf{M} = \begin{bmatrix} R & \boldsymbol{Z} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}$
2	small algebra element $x \in \mathfrak{g}$	$\boldsymbol{r}$	$\boldsymbol{s} = \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{t} \end{bmatrix}$	$\boldsymbol{\zeta} = \begin{bmatrix} \boldsymbol{\rho} \\ \boldsymbol{\tau} \end{bmatrix}$
3	log of small action $X = x^\wedge$	$S(\boldsymbol{r})$ (5.3)	$\mathsf{A}(\boldsymbol{s}) = \begin{bmatrix} S(\boldsymbol{r}) & \boldsymbol{t} \\ \mathbf{0}^\mathsf{T} & 0 \end{bmatrix}$ (5.2)	(log of row 7, column 4)
4	adjoint action at $e \in \mathcal{G}$ $\exp(x_{\mathrm{ad}}) = g\exp(x)g^{-1}$	$R(\boldsymbol{r}_{\mathrm{ad}}) = RR(\boldsymbol{r})R^\mathsf{T}$ (5.29)	$\mathsf{M}(\boldsymbol{s}_{\mathrm{ad}}) = \mathsf{M}\,\mathsf{M}(\boldsymbol{s})\,\mathsf{M}^{-1}$ (5.70)	$\mathsf{M}(\boldsymbol{\zeta}_{\mathrm{ad}}) = \mathsf{M}\,\mathsf{M}(\boldsymbol{\zeta})\,\mathsf{M}^{-1}$ (5.91)
5	adjoint matrix for $\mathrm{d}x$ X_{ad}, $\mathrm{d}x_{\mathrm{ad}} = X_{\mathrm{ad}}\,\mathrm{d}x$	$R_{\mathrm{ad}} = R$ (5.34)	$\mathrm{Ad}(\mathsf{M}) = \begin{bmatrix} R & 0 \\ S(\boldsymbol{Z})R & R \end{bmatrix}$ (5.73)	$\mathrm{Ad}(\mathsf{M}) = \begin{bmatrix} R & 0 \\ S(\boldsymbol{Z})R & R \end{bmatrix}$ (5.94)
6	uncertain group element $\underline{g} = \exp(\underline{x})\, g \in \mathcal{G}$, $\underline{x}$ small	$\underline{R} = R(\underline{\boldsymbol{r}})R$ (5.21)	${}^s\underline{\mathsf{M}} = \exp(\mathsf{A}(\underline{\boldsymbol{s}}))\,\mathsf{M}$ (5.52)	${}^\zeta\underline{\mathsf{M}} = \begin{bmatrix} R(\underline{\boldsymbol{\rho}})R & \boldsymbol{Z} + \underline{\boldsymbol{\tau}} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}$ (5.54)
7	multiplicative noise element $\exp(\underline{x}) = \underline{g}\,g^{-1} \in \mathcal{G}$	$R(\underline{\boldsymbol{r}})$ (5.56)	$\exp(\mathsf{A}(\underline{\boldsymbol{s}})) = \begin{bmatrix} R(\underline{\boldsymbol{r}}) & V(\underline{\boldsymbol{r}})\underline{\boldsymbol{t}} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}$ (5.56)	$\begin{bmatrix} R(\underline{\boldsymbol{\rho}}) & (I_3 - R(\underline{\boldsymbol{\rho}}))\boldsymbol{Z} + \underline{\boldsymbol{\tau}} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}$ (5.90)
8	differential noise element $\mathrm{d}x^{-1} \in \mathfrak{g}$	$\mathrm{d}\boldsymbol{r}$ (5.67)	$\mathrm{d}\boldsymbol{s} = \begin{bmatrix} I_3 & 0 \\ S(\boldsymbol{Z}) & I_3 \end{bmatrix} \mathrm{d}\boldsymbol{\zeta}$ (5.67)	$\mathrm{d}\boldsymbol{\zeta} = \begin{bmatrix} I_3 & 0 \\ -S(\boldsymbol{Z}) & I_3 \end{bmatrix} \mathrm{d}\boldsymbol{s}$ (5.67)

Table 5.1: Lie group elements (1/2): actions, adjoints, noisy elements, inverses, concatenations and relative actions

5.4 Uncertain Rotations

5.4.1 General setup

In all cases we represent the uncertain linear transformation $\underline{\mathcal{X}}$ by the mean transformation matrix of size $m \times m$

$$\underline{\mathcal{X}}: \quad \{X, \Sigma_{\Delta x \Delta x}\} \tag{5.14}$$

and a stochastic n-vector $\underline{\Delta\boldsymbol{x}}$, which captures the noise of the transformation, and has zero mean and a covariance matrix as second moments

$$\underline{\Delta\boldsymbol{x}} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\Delta x \Delta x})\,. \tag{5.15}$$

I.e. we assume the distribution is uni-modal and can be represented sufficiently well by the first two moments. We do not assume the distribution to be a normal distribution, unless we want to perform statistical testing. Then, we assume higher order of the nonlinear relations effects are small enough to be acceptable for the application.

The dimension n of the vector $\Delta\boldsymbol{x}$ is identical to the degrees of freedom of the transformation, in order to have regular covariance matrix in general. Hence, the two matrices

	1	2	3	4
	$\downarrow$ object $\setminus$ $\mathcal{G} \rightarrow$	SO(3)	SE(3), $\boldsymbol{s}$	SE(3), $\boldsymbol{\zeta}$
9	uncertain inverse $\underline{g}^{(-1)} \in \mathcal{G}$	$\underline{R}^{-1} = R^{\mathsf{T}}\left(\underline{\boldsymbol{r}}^{(-1)}\right) R^{\mathsf{T}}$ (5.36)	${}^{s}\underline{\mathsf{M}}^{-1} = \mathsf{M}^{-1}\left(\underline{\boldsymbol{s}}^{(-1)}\right)\mathsf{M}^{-1}$ (5.77)	${}^{\zeta}\underline{\mathsf{M}}^{-1} = \begin{bmatrix} R(\underline{\boldsymbol{\rho}}^{(-1)})R^{\mathsf{T}} & -R^{\mathsf{T}}\boldsymbol{Z} + \underline{\boldsymbol{\tau}}^{(-1)} \\ \boldsymbol{0}^{\mathsf{T}} & 1 \end{bmatrix}$ (5.95)
10	differential inverse $\mathrm{d}x^{(-1)},\ \mathrm{d}x^{(-1)} \in \mathcal{g}$	$\mathrm{d}\boldsymbol{r}^{(-1)} = -R^{\mathsf{T}}\mathrm{d}\boldsymbol{r}$ (5.38)	$\mathrm{d}\boldsymbol{s}^{(-1)} = -\mathsf{M}_{\mathrm{ad}}^{-1}\mathrm{d}\boldsymbol{s}$ (5.78)	$\begin{bmatrix} \mathrm{d}\boldsymbol{\rho}^{(-1)} \\ \mathrm{d}\boldsymbol{\tau}^{(-1)} \end{bmatrix} = -\begin{bmatrix} R^{\mathsf{T}} & \boldsymbol{0} \\ R^{\mathsf{T}}S(\boldsymbol{Z}) & R^{\mathsf{T}} \end{bmatrix}\begin{bmatrix} \mathrm{d}\boldsymbol{\rho} \\ \mathrm{d}\boldsymbol{\tau} \end{bmatrix}$ (5.97)
11	concatenation $g = g_2 \circ g_1 \in \mathcal{G}$	$R = R_2 R_1$	$\mathsf{M} = \mathsf{M}_2\mathsf{M}_1$	$\mathsf{M} = \mathsf{M}_2\mathsf{M}_1$
12	differential concatenation $\mathrm{d}x = \mathrm{d}(x_2 \circ x_1)$	$\mathrm{d}\boldsymbol{r} = \boldsymbol{R}_{2,\mathrm{ad}}\,\mathrm{d}\boldsymbol{r}_1 + \mathrm{d}\boldsymbol{r}_2$ (5.42)	$\mathrm{d}\boldsymbol{s} = \mathrm{Ad}(\mathsf{M}_2)\,\mathrm{d}\boldsymbol{s}_1 + \mathrm{d}\boldsymbol{s}_2$ (5.238)	$\mathrm{d}\boldsymbol{\rho} = \mathrm{d}\boldsymbol{\rho}_2 + R_2\mathrm{d}\boldsymbol{\rho}_1$ $\mathrm{d}\boldsymbol{\tau} = \mathrm{d}\boldsymbol{\tau}_2 + R_2\mathrm{d}\boldsymbol{\tau}_1 - S(R_2\boldsymbol{Z}_1)\mathrm{d}\boldsymbol{\rho}_2$ (5.244)
13	relative action $g = g_1^{-1} \circ g_2 \in \mathcal{G}$	$R = R_1^{\mathsf{T}}R_2$	$\mathsf{M} = \mathsf{M}_1^{-1}\mathsf{M}_2$	$\mathsf{M} = \mathsf{M}_1^{-1}\mathsf{M}_2$
14	differential relative action $\mathrm{d}x = \mathrm{d}(x_1^{-1} \circ x_2)$	$\mathrm{d}\boldsymbol{r} = R_1^{\mathsf{T}}(\mathrm{d}\boldsymbol{r}_2 - \mathrm{d}\boldsymbol{r}_1)$ (5.49)	$\mathrm{d}\boldsymbol{s}_{12} = \mathsf{M}_{1,\mathrm{ad}}^{-1}\,(\mathrm{d}\boldsymbol{s}_2 - \mathrm{d}\boldsymbol{s}_1)$ (5.87)	$\mathrm{d}\boldsymbol{\rho} = R_1^{\mathsf{T}}(\mathrm{d}\boldsymbol{\rho}_2 - \mathrm{d}\boldsymbol{\rho}_1)$ $\mathrm{d}\boldsymbol{\tau} = R_1^{\mathsf{T}}S(\boldsymbol{Z}_2 - \boldsymbol{Z}_1)\mathrm{d}\boldsymbol{\rho}_1 + R_1^{\mathsf{T}}\mathrm{d}(\boldsymbol{\tau}_2 - \boldsymbol{\tau}_1)$ (5.107), (5.108)

Table 5.2: Lie group elements (2/2): actions, adjoints, noisy elements, inverses, concatenations and relative actions

X and $\Sigma_{\Delta x \Delta x}$ have different dimension in general. The matrix X may be the mean motion $X := \mathbb{E}(\underline{\mathsf{X}})$, or an estimated motion $X := \widehat{X}$, depending on the context. If we use a minimal representation, it also may be the vector $\boldsymbol{x}$ specifying the motion. The way how X is related to $\Delta\boldsymbol{x}$ needs to be specified, and even may vary for the same type of transformation. In all cases we might exploit the fact that transformations build a Lie group, i.e., a continuous group, and can be written as matrix exponential. We warm up with rotations as special motions.

5.4.2 Representing rotations

There are many ways to represent rotations. We only address three of them.

1. We start with the classical definition of rotations using Euler angles, say $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$. We generally have the uncertain rotation

$$\underline{\mathcal{R}}: \quad \{\boldsymbol{\alpha}, \Sigma_{\Delta\alpha\Delta\alpha}\}\,. \tag{5.16}$$

   e.g., specified by

$$\underline{R} = R_2(\underline{\alpha}_3)R_2(\underline{\alpha}_2)R_1(\underline{\alpha}_1) \quad \text{with} \quad \underline{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + \underline{\Delta\boldsymbol{\alpha}}\,, \tag{5.17}$$

   where the indices of the rotation matrices indicate the rotation axes. In whatever sequenced the angles are applied, and what ever axis sequence is chosen, the representation for some angles will have a singularity, what is called the *gimbal lock*.

2. Therefore the Rodriguez form, depending on a rotation vector $\boldsymbol{\vartheta}$, often is preferred. Here we have the uncertain rotation

$$\underline{\mathcal{R}}: \quad \{\boldsymbol{\vartheta}, \Sigma_{\Delta\vartheta\Delta\vartheta}\}\,. \tag{5.18}$$

   It is given by the exponential map of the skew matrix S_ϑ of the rotation vector $\boldsymbol{\vartheta}$:

$$R = \exp(S(\underline{\boldsymbol{\vartheta}})) = I_3 + \frac{\sin\|\underline{\boldsymbol{\vartheta}}\|}{\|\underline{\boldsymbol{\vartheta}}\|}\underline{S}_\vartheta + \frac{1-\cos\|\underline{\boldsymbol{\vartheta}}\|}{\|\underline{\boldsymbol{\vartheta}}\|^2}\underline{S}_\vartheta^2 \quad \text{with} \quad \underline{\boldsymbol{\vartheta}} = \boldsymbol{\vartheta} + \underline{\Delta\boldsymbol{\vartheta}}\,, \tag{5.19}$$

3. Finally, we also can adopt the multiplicative definition of an uncertain rotation. Here the uncertain rotation is given by

$$\underline{\mathcal{R}}: \quad \{R, \Sigma_{rr}\}, \tag{5.20}$$

specified by

$$\underline{R} = \exp(S(\underline{\boldsymbol{r}}))\ R = R(\underline{\boldsymbol{r}})\ R. \tag{5.21}$$

5.4.3 Relations between the representations

When comparing the three definitions of the uncertain rotations, we need to have explicit expressions for the derivatives of R w.r.t. the elements of the noise component, either $\underline{\Delta\boldsymbol{\alpha}}$, $\underline{\Delta\boldsymbol{\theta}}$, or $\underline{\boldsymbol{r}}$.

Unfortunately, the expressions for the derivatives of the exponential $\exp(S(\boldsymbol{\vartheta}))$ w.r.t. $\boldsymbol{\vartheta}$ at some arbitrary – not necessarily small – vector, e.g., at $\boldsymbol{\vartheta} = \mathbb{E}(\underline{\boldsymbol{\vartheta}})$ are quite cumbersome. Therefore in the following we will not use the definition of an uncertain transformation using the exponential of some matrix depending on arbitrary parameters. This excludes choice 2 for defining uncertain rotations.

However, we can derive the Jacobian of the angles $\boldsymbol{r}$ in the multiplicative exponential representation w.r.t. Euler angles $\boldsymbol{\alpha}$. We specifically have

$$J_{r\alpha} = \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\alpha}} = [R_2(\alpha_3)R_2(\alpha_2)\boldsymbol{e}_1 \mid R_2(\alpha_3)\boldsymbol{e}_2 \mid \boldsymbol{e}_3]\ , \tag{5.22}$$

see Appendix 5.8.2. Since $|J_{r\alpha}| = \cos\alpha_2$ we have

$$J_{\alpha r} = J_{r\alpha}^{-1} \quad \text{if} \quad \cos\alpha_2 \neq 0\,. \tag{5.23}$$

This not only makes the Gimbal lock of the representation with Euler angles explicit, but shows, that we can choose either representation if we avoid the Gimbal lock.

Since all minimal representations for rotations show singularities for specific rotations or are not unique, we only discuss the option 3, with the multiplicative way to represent an uncertain rotation.

5.4.4 The rotation in exponential representation

We now discuss the adjoint rotation, the inverse, the concatenated, and the relative rotation.

5.4.4.1 The adjoint rotation

Let us for a moment define an uncertain rotation by first applying a small random rotation $R(\underline{\boldsymbol{q}})$ and then a fixed large rotation, e.g., $R := \mathbb{E}(\underline{R})$:

$$\underline{R} = R\ R(\underline{\boldsymbol{q}}) \qquad \mathbb{D}(\underline{\boldsymbol{q}}) = \Sigma_{qq}\,. \tag{5.24}$$

Applying it to a vector $\boldsymbol{x}$ we obtain a stochastic vector

$$\underline{\boldsymbol{y}} = \underline{R}\,\boldsymbol{x} = R\ R(\underline{\boldsymbol{q}})\,\boldsymbol{x}\,. \tag{5.25}$$

Now, let us choose another small rotation via what is called the *adjoint* rotation vector $\boldsymbol{q}_{\text{ad}}$ [5]

$$R(\boldsymbol{q}_{\text{ad}}) = R\ R(\boldsymbol{q})\ R^{-1} = R\ R(\boldsymbol{q})\ R^{\mathsf{T}}\,. \tag{5.26}$$

If we apply this small rotation with $\boldsymbol{q}_{\text{ad}}$ to $\boldsymbol{y} = R\boldsymbol{x}$ we obtain

$$R(\underline{\boldsymbol{q}}_{\text{ad}})\ R\,\boldsymbol{x} = R\ R(\underline{\boldsymbol{q}})\,\boldsymbol{x}\,. \tag{5.27}$$

[5] the use of the name $\boldsymbol{q}$ for a rotation vector, should not be confused with the common naming of quaternions, which do not play a role in this note.

Hence, if we first perturb $\boldsymbol{x}$ by a small rotation $\underline{\boldsymbol{q}}$ and then rotate the perturbed vector $R(\underline{\boldsymbol{q}})\,\boldsymbol{x}$ we obtain the same uncertain vector as when first rotating $\boldsymbol{x}$ and then perturbing the rotated vector $\boldsymbol{y} = R\boldsymbol{x}$ with the adjoint rotation vector $\underline{\boldsymbol{q}}_{\mathrm{ad}}$. We could also have written the relation – neutrally w.r.t. order – as

$$R(\underline{\boldsymbol{q}}_L)\; R\; \boldsymbol{x} = R\; R(\underline{\boldsymbol{q}}_R)\; \boldsymbol{x}\,, \tag{5.28}$$

the indices standing for left and right hand rotation. Hence the adjoint rotation $\underline{\boldsymbol{q}}_{\mathrm{ad}} =: \underline{\boldsymbol{q}}_L$ leads to the same result if applied to the left of a rotation as the original rotation $\boldsymbol{q}$ applied to the right of a rotation.[6]

Thus we have for any rotation vector $\boldsymbol{r}$ the adjoint rotation

$$R(\underline{\boldsymbol{r}}_{\mathrm{ad}}) = R\; R(\boldsymbol{r})\; R^{-1} \tag{5.29}$$

or the relation

$$\boxed{R(\underline{\boldsymbol{r}}_{\mathrm{ad}})\; R = R\; R(\boldsymbol{r})\,.} \tag{5.30}$$

Now, we express the *differential* adjoint rotation vector $\mathrm{d}\boldsymbol{r}_{\mathrm{ad}}$ directly as a function of the differential vector $\mathrm{d}\boldsymbol{r}$. We have

$$\mathrm{d}R(\boldsymbol{r}_{\mathrm{ad}})\; R = R\; \mathrm{d}R(\boldsymbol{r}) \tag{5.31}$$

or

$$S(\mathrm{d}\boldsymbol{r}_{\mathrm{ad}})\; R = R\; S(\mathrm{d}\boldsymbol{r}) = S(R\mathrm{d}\boldsymbol{r})\; R \tag{5.32}$$

hence

$$\boxed{\mathrm{d}\boldsymbol{r}_{\mathrm{ad}} = R\; \mathrm{d}\boldsymbol{r}\,.} \tag{5.33}$$

We observe: the differential rotation vector $\boldsymbol{r}$ and its differential adjoint rotation vector $\mathrm{d}\boldsymbol{r}_{\mathrm{ad}}$ are *linearly* related by the rotation matrix R. Since, due to $R(\boldsymbol{r}) = I_3 + S(\boldsymbol{r}) + O(r_i^2)$, the vector $\boldsymbol{r}$ spans the tangent space of a rotation at the unit rotation. But $\boldsymbol{r}_{\mathrm{ad}}$ also defines a basis, just a different one in this 3-dimensional tangent space.

Later we will see that the rotation matrix in (5.33) actually is the adjoint rotation matrix, which in this case simplifies to

$$R_{\mathrm{ad}} = R\,, \tag{5.34}$$

see (5.73).

5.4.4.2 The uncertain inverse rotation

Let now the uncertain rotation be given by

$$\boxed{\underline{R} = R(\underline{\boldsymbol{r}})\; R\,.} \tag{5.35}$$

The inverse rotation is represented the same way

$$\underline{R}^{-1} = R(\underline{\boldsymbol{r}}^{(-1)})\; R^{-1}\,. \tag{5.36}$$

The mean of the uncertain inverse is the inverse of the mean rotation:

$$\mathbb{E}(\underline{R}^{-1}) = (\mathbb{E}(\underline{R}))^{-1}\,. \tag{5.37}$$

The differential rotation vector $\mathrm{d}\boldsymbol{r}^{(-1)}$ of the inverse rotation can be shown to be

$$\boxed{\mathrm{d}\boldsymbol{r}^{(-1)} = -R^{\mathsf{T}}\mathrm{d}\boldsymbol{r}\,,} \tag{5.38}$$

see Appendix 5.8.5.

[6] Following this interpretation of the adjoint rotation it would have been straight forward to define an uncertain rotation by $\underline{R} = R\;\exp(S(\underline{\boldsymbol{r}}))$. However, most authors use the original definition of an uncertain rotation, where the noise component of the rotation is applied after the mean rotation. Unfortunately the definiton of a similarity transformation or conjugation of matrices $B = X^{-1}AX$ (https://mathworld.wolfram.com/SimilarMatrices.html) is just using the inverse operation sequence as the adjoint action $\exp(x_{\mathrm{ad}}) = g\;\exp(x)\;g^{-1}$ in a Lie group; however, see https://mathworld.wolfram.com/SimilarityTransformation.html.

5.4.4.3 The uncertain concatenated rotation

Let a *possibly correlated* rotation pair be given by

$$\left\{\underline{\mathcal{R}}_1, \underline{\mathcal{R}}_2\right\}: \quad \left\{[R_1, R_2], \mathbb{D}\left(\left[\begin{array}{c}\underline{\boldsymbol{r}}_1\\ \underline{\boldsymbol{r}}_2\end{array}\right]\right)\right\}. \tag{5.39}$$

The concatenated rotation is

$$\underline{\mathcal{R}} = \underline{\mathcal{R}}_2\underline{\mathcal{R}}_1: \quad \underline{R} = \underline{R}_2\underline{R}_1 = R(\underline{\boldsymbol{r}})R\,. \tag{5.40}$$

The mean of the concatenated rotations is

$$\mathbb{E}(\underline{R}) = \mathbb{E}(\underline{R}_2)\,\mathbb{E}(\underline{R}_1)\,. \tag{5.41}$$

The differential of the rotation vector $\boldsymbol{r}$ of the concatenated rotations is given by

$$\boxed{\mathrm{d}\boldsymbol{r} = R_2\mathrm{d}\boldsymbol{r}_1 + \mathrm{d}\boldsymbol{r}_2\,.} \tag{5.42}$$

This is a special case of the concatenated motions, see Appendix 5.8.8.

Observe, Eq. (5.42) allows to derive the uncertainty of a correlated rotation pair $\left\{\underline{\mathcal{R}}_1, \underline{\mathcal{R}}_2\right\}$

$$\Sigma_{rr} = J\,\Sigma_{pp}\,J^{\mathsf{T}}\,, \tag{5.43}$$

with

$$J = [R_2 \mid I_3] \quad \text{and} \quad \Sigma_{pp} = \mathbb{D}\left(\left[\begin{array}{c}\underline{\boldsymbol{r}}_1\\ \underline{\boldsymbol{r}}_2\end{array}\right]\right) = \left[\begin{array}{cc}\Sigma_{r_1r_1} & \Sigma_{r_1r_2}\\ \Sigma_{r_2r_1} & \Sigma_{r_2r_2}\end{array}\right]. \tag{5.44}$$

5.4.4.4 The uncertain relative rotation

We want to determine the relative rotation

$$R_{12} = R_1^{-1}R_2 \tag{5.45}$$

in case all rotations are uncertain and possibly correlated. Let the uncertain rotations be given by

$$\underline{R}_1 = R(\underline{\boldsymbol{r}}_1)\,R_1 \quad \text{and} \quad \underline{R}_2 = R(\underline{\boldsymbol{r}}_2)\,R_2\,. \tag{5.46}$$

Then the uncertain relative rotation is

$$R(\underline{\boldsymbol{r}}_{12})\,R_{12} = (R(\underline{\boldsymbol{r}}_1)\,R_1)^{-1}R(\underline{\boldsymbol{r}}_2)\,R_2\,. \tag{5.47}$$

The mean of the relative rotations is

$$\mathbb{E}(\underline{R}) = \mathbb{E}(\underline{R}_1)^{-1}\,\mathbb{E}(\underline{R}_2)\,. \tag{5.48}$$

The differential $\mathrm{d}\boldsymbol{r}_{12}$ of the rotation vector of the relative rotation is

$$\boxed{\mathrm{d}\boldsymbol{r}_{12} = R_1^{\mathsf{T}}(\mathrm{d}\boldsymbol{r}_2 - \mathrm{d}\boldsymbol{r}_1)\,.} \tag{5.49}$$

The result is a special case of the relative motion, see Appendix 5.8.10

The result in (5.49) can be derived using the relation (5.36) for the inverse and the relation (5.42) for the concatenation.

5.5 Uncertain Motions

5.5.1 Representations

The uncertainty of a motion is captured in the uncertain twist vector $\mathit{m}(\underline{\Delta \boldsymbol{m}})$

$$\underline{\Delta \boldsymbol{m}} \sim \mathcal{N}\left(\mathbf{0}, \Sigma_{\Delta m \Delta m}\right) . \tag{5.50}$$

We address the following two representations

1. The exponential representation with the twist vector

$$\mathit{m}: \quad \boldsymbol{s} = \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{t} \end{bmatrix} . \tag{5.51}$$

is given by

$$^{s}\underline{\mathcal{M}}: \quad ^{s}\underline{\mathsf{M}} = \exp(\mathsf{A}(\underline{\boldsymbol{s}}))\ \mathsf{M} \quad \text{with} \quad \mathsf{A}(\boldsymbol{s}) = \begin{bmatrix} S(\boldsymbol{r}) & \boldsymbol{t} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix} . \tag{5.52}$$

2. The partially exponential representation with the twist vector

$$\mathit{m}: \quad \boldsymbol{\zeta} = \begin{bmatrix} \boldsymbol{\rho} \\ \boldsymbol{\tau} \end{bmatrix} . \tag{5.53}$$

is given by

$$^{\zeta}\underline{\mathcal{M}}: \quad ^{\zeta}\underline{\mathsf{M}} = \begin{bmatrix} ^{\zeta}\underline{R} & ^{\zeta}\underline{\boldsymbol{Z}} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \tag{5.54}$$

with

$$^{\zeta}R = \exp(S(\boldsymbol{\rho}))\ R \quad \text{and} \quad ^{\zeta}\boldsymbol{Z} = \boldsymbol{Z} + \boldsymbol{\tau} \tag{5.55}$$

It appears obvious, that both representations are useful. However, they differ in the meaning of the twist vector, as we will see.

Observe, we have

$$R(\boldsymbol{r}) = \exp(S(\boldsymbol{r})) \quad \text{and} \quad \mathsf{M}(\boldsymbol{s}) = \exp(\mathsf{A}(\boldsymbol{s})) = \begin{bmatrix} R(\boldsymbol{r}) & V(\boldsymbol{r})\boldsymbol{t} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \tag{5.56}$$

with

$$R(\boldsymbol{r}) = I_3 + \sum_{n=1}^{\infty} \frac{S^n(\boldsymbol{r})}{n!} = I_3 + \frac{\sin\|\boldsymbol{r}\|}{\|\boldsymbol{r}\|} S_r + \frac{1-\cos\|\boldsymbol{r}\|}{\|\boldsymbol{r}\|^2} S_r^2 \tag{5.57}$$

and

$$V(\boldsymbol{r}) = I_3 + \sum_{n=1}^{\infty} \frac{S^n(\boldsymbol{r})}{(n+1)!} = I_3 + \frac{1-\cos\|\boldsymbol{r}\|}{\|\boldsymbol{r}\|^2} S_r + \frac{1-\sin\|\boldsymbol{r}\|}{\|\boldsymbol{r}\|^3} S_r^2 . \tag{5.58}$$

see Leonardos et al. (2015, eq. (19)). Thus for small values of $\|\boldsymbol{r}\|$ we may use the first order approximation

$$V(\boldsymbol{r}) \approx I_3 . \tag{5.59}$$

Therefore we have for an uncertain motion in exponential representation with small $\boldsymbol{s}$

$$^{s}\underline{\mathsf{M}} \approx \begin{bmatrix} R(\underline{\boldsymbol{r}}) & \underline{\boldsymbol{t}} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} R & \boldsymbol{Z} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} = \begin{bmatrix} R(\underline{\boldsymbol{r}})R & R(\underline{\boldsymbol{r}})\boldsymbol{Z} + \underline{\boldsymbol{t}} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} . \tag{5.60}$$

The corresponding expression for the partially exponential representation is lengthy.

5.5.2 Comparing the two representations

We now compare the two representations with the two twist vectors

$$\boldsymbol{s} = \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{t} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\zeta} = \begin{bmatrix} \boldsymbol{\rho} \\ \boldsymbol{\tau} \end{bmatrix} \tag{5.61}$$

for defining the uncertain motions as

$${}^{s}\underline{\mathsf{M}} = \exp(\mathsf{A}(\underline{\boldsymbol{s}}))\ \mathsf{M} \quad \text{and} \quad {}^{\zeta}\underline{\mathsf{M}} = \begin{bmatrix} \exp(S(\underline{\boldsymbol{\zeta}}))\ R & \boldsymbol{Z} + \underline{\boldsymbol{\tau}} \\ \boldsymbol{0}^{\mathsf{T}} & 1 \end{bmatrix}. \tag{5.62}$$

Assuming the two uncertain motions are statistically equivalent, we can relate the differentials of the twist vectors. We obtain the total differential for the two motions from:

- for the exponential representation

$${}^{s}\mathrm{d}\mathsf{M} = \begin{bmatrix} S(\mathrm{d}\boldsymbol{r}) & \mathrm{d}\boldsymbol{t} \\ \boldsymbol{0}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} R & \boldsymbol{Z} \\ \boldsymbol{0}^{\mathsf{T}} & 1 \end{bmatrix} = \begin{bmatrix} S(\mathrm{d}\boldsymbol{r})R & S(\mathrm{d}\boldsymbol{r})\boldsymbol{Z} + \mathrm{d}\boldsymbol{t} \\ \boldsymbol{0}^{\mathsf{T}} & 1 \end{bmatrix} \tag{5.63}$$

  and

- for the partially exponential model

$${}^{\zeta}\mathrm{d}\mathsf{M} = \begin{bmatrix} S(\mathrm{d}\boldsymbol{\rho}) & \mathrm{d}\boldsymbol{\tau} \\ \boldsymbol{0}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} R & \boldsymbol{Z} \\ \boldsymbol{0}^{\mathsf{T}} & 1 \end{bmatrix} = \begin{bmatrix} S(\mathrm{d}\boldsymbol{\rho})R & \mathrm{d}\boldsymbol{\tau} \\ \boldsymbol{0}^{\mathsf{T}} & 1 \end{bmatrix} \tag{5.64}$$

If the uncertain motions are the same, the two differentials must be identical, and we obtain the relations

$$\mathrm{d}\boldsymbol{r} = \mathrm{d}\boldsymbol{\rho} \qquad \text{or} \qquad \mathrm{d}\boldsymbol{\rho} = \mathrm{d}\boldsymbol{r} \tag{5.65}$$

$$\mathrm{d}\boldsymbol{t} = \mathrm{d}\boldsymbol{\tau} + S(\boldsymbol{Z})\mathrm{d}\boldsymbol{\rho} \qquad \text{or} \qquad \mathrm{d}\boldsymbol{\tau} = \mathrm{d}\boldsymbol{t} - S(\boldsymbol{Z})\mathrm{d}\boldsymbol{r} \tag{5.66}$$

$$\begin{bmatrix} \mathrm{d}\boldsymbol{r} \\ \mathrm{d}\boldsymbol{t} \end{bmatrix} = \begin{bmatrix} I_3 & 0 \\ S(\boldsymbol{Z}) & I_3 \end{bmatrix} \begin{bmatrix} \mathrm{d}\boldsymbol{\rho} \\ \mathrm{d}\boldsymbol{\tau} \end{bmatrix} \qquad \text{or} \qquad \begin{bmatrix} \mathrm{d}\boldsymbol{\rho} \\ \mathrm{d}\boldsymbol{\tau} \end{bmatrix} = \begin{bmatrix} I_3 & 0 \\ -S(\boldsymbol{Z}) & I_3 \end{bmatrix} \begin{bmatrix} \mathrm{d}\boldsymbol{r} \\ \mathrm{d}\boldsymbol{t} \end{bmatrix}$$

between the twist vectors $\boldsymbol{s}$ and $\boldsymbol{\zeta}$. Hence, we have the relations

$$\boxed{\mathrm{d}\boldsymbol{s} = J_{s\zeta}\mathrm{d}\boldsymbol{\zeta} \quad \text{and} \quad \mathrm{d}\boldsymbol{\zeta} = J_{\zeta s}\mathrm{d}\boldsymbol{s}} \tag{5.67}$$

with

$$J_{s\zeta} = \begin{bmatrix} I_3 & 0 \\ S(\boldsymbol{Z}) & I_3 \end{bmatrix} \quad \text{and} \quad J_{\zeta s} = J_{s\zeta}^{-1} = \begin{bmatrix} I_3 & 0 \\ -S(\boldsymbol{Z}) & I_3 \end{bmatrix}. \tag{5.68}$$

This allows us to transfer the covariance matrices of the twist vectors

$$\Sigma_{s\xi} = J_{s\zeta}\Sigma_{\zeta\zeta}J_{s\zeta}^{\mathsf{T}} \quad \text{and} \quad \Sigma_{\zeta\zeta} = J_{s\zeta}\Sigma_{ss}J_{s\zeta}^{\mathsf{T}}. \tag{5.69}$$

between both representations.

As a result, we find: the uncertain rotation components $\mathrm{d}\boldsymbol{r}$ and $\mathrm{d}\boldsymbol{\rho}$ of both representations are identical but the uncertain translation components $\mathrm{d}\boldsymbol{t}$ and $\mathrm{d}\boldsymbol{\tau}$ differ by the effect of the uncertain rotation applied to the full translation $\boldsymbol{Z}$.

5.5.3 The motion in exponential representation

We now discuss the adjoint, the inverse, the concatenated, and the relative motion in exponential representation.

5.5.3.1 The adjoint motion for the exponential representation

The adjoint motion $\mathsf{M}(\boldsymbol{s}_{\mathrm{ad}})$ is defined with, what is called the *adjoint* motion vector $\boldsymbol{s}_{\mathrm{ad}}$,

$$\mathsf{M}(\boldsymbol{s}_{\mathrm{ad}}) = \mathsf{M}\,\mathsf{M}(\boldsymbol{s})\,\mathsf{M}^{-1}\,. \tag{5.70}$$

For proofs we often use it in the differential form

$$\mathsf{A}(\boldsymbol{s}_{\mathrm{ad}})\,\mathsf{M} = \mathsf{M}\,\mathsf{A}(\boldsymbol{s})\,, \tag{5.71}$$

allowing to exchange the differential of the perturbing noise matrix A and the motion matrix M. Also here we obtain a simple linear relation between the differentials of the twist vectors

$$\boxed{\mathrm{d}\boldsymbol{s}_{\mathrm{ad}} = \mathrm{Ad}(\mathsf{M})\,\mathrm{d}\boldsymbol{s}\,,} \tag{5.72}$$

with the adjoint motion matrix relating the two 6-vectors

$$\boxed{\mathrm{Ad}(\mathsf{M}) = \begin{bmatrix} R & 0 \\ S(\boldsymbol{Z})R & R \end{bmatrix},} \tag{5.73}$$

and its inverse

$$\mathsf{M}_{\mathrm{ad}}^{-1} = \begin{bmatrix} R^{\mathsf{T}} & 0 \\ R^{\mathsf{T}}\,S^{\mathsf{T}}(\boldsymbol{Z}) & R^{\mathsf{T}} \end{bmatrix}. \tag{5.74}$$

Eq. (5.73) can also be written as

$$\boxed{\mathsf{A}\,(\mathrm{Ad}(\mathsf{M})\cdot\mathrm{d}\boldsymbol{s})\quad \mathsf{M} = \mathsf{M}\,\mathsf{A}(\mathrm{d}\boldsymbol{s})\,.} \tag{5.75}$$

The proof is given in Appendix 5.8.3. We observe: the differential rotation vector $\boldsymbol{s}$ and its differential adjoint motion vector $\mathrm{d}\boldsymbol{s}_{\mathrm{ad}}$ are linearly related by the adjoint motion matrix $\mathrm{Ad}(\mathsf{M})$. The relation between the small motion vectors only holds for differential motions. This is sufficient for all practical cases, where the relative precision of the motion parameters is high enough. Observe, when restricting to rotations we have

$$R_{\mathrm{ad}} = R\,, \tag{5.76}$$

The simplicity of this relation does not reveil the strength of the concept for more general transformations.

5.5.3.2 The inverse motion in exponential representation

Similarly as for rotations, we can derive the relation between the differential twist vector of the inverse motion to the one of the original motion.

We have the basic relation

$$\underline{\mathsf{M}}^{-1} = \mathsf{M}\left(\underline{\boldsymbol{s}}^{(-1)}\right)\cdot\mathsf{M}^{-1} = (\mathsf{M}(\underline{\boldsymbol{s}})\cdot\mathsf{M})^{-1} = \mathsf{M}^{-1}\cdot\mathsf{M}^{-1}(\underline{\boldsymbol{s}})\,. \tag{5.77}$$

Using the adjoint motion we can derive the following relation between the differential twist vectors:

$$\boxed{\mathrm{d}\boldsymbol{s}^{(-1)} = -\mathrm{Ad}(\mathsf{M})^{-1}\,\mathrm{d}\boldsymbol{s}\,,} \tag{5.78}$$

see the proof in the Appendix 5.8.6

5.5.3.3 The concatenated motion in exponential representation

Let a possibly correlated motion pair be given by

$$\{{}^{s}\underline{\mathcal{M}}_1, {}^{s}\underline{\mathcal{M}}_2\}: \quad \left\{[\mathsf{M}_1, \mathsf{M}_2], \mathbb{D}\left(\begin{bmatrix} \underline{\boldsymbol{s}}_1 \\ \underline{\boldsymbol{s}}_2 \end{bmatrix}\right)\right\}. \tag{5.79}$$

The concatenated motion is

$$ {}^s\underline{\mathcal{M}} = {}^s\underline{\mathcal{M}}_2 \, {}^s\underline{\mathcal{M}}_1 : \quad {}^s\underline{\mathsf{M}} = {}^s\underline{\mathsf{M}}_2 {}^s\underline{\mathsf{M}}_1 = \mathsf{M}(\underline{\boldsymbol{s}})\mathsf{M} \, . \tag{5.80} $$

The mean of the concatenated motion is

$$ \mathbb{E}({}^s\underline{\mathsf{M}}) = \mathbb{E}({}^s\underline{\mathsf{M}}_2) \, \mathbb{E}({}^s\underline{\mathsf{M}}_1) \, . \tag{5.81} $$

The differential of the rotation vector $\boldsymbol{s}$ of the concatenated rotation is given by

$$ \boxed{\mathrm{d}\boldsymbol{s} = \mathsf{M}_2 \mathrm{d}\boldsymbol{s}_1 + \mathrm{d}\boldsymbol{s}_2 \, .} \tag{5.82} $$

see Appendix 5.8.8.

5.5.3.4 The relative motion in exponential representation

We want to determine the relative motion

$$ \mathsf{M}_{12} = \mathsf{M}_1^{-1} \mathsf{M}_2 \tag{5.83} $$

in case all motions are uncertain. Let the uncertain motions be given by

$$ {}^s\underline{\mathsf{M}}_1 = \mathsf{M}(\underline{\boldsymbol{s}}_1) \, \mathsf{M}_1 \quad \text{and} \quad {}^s\underline{\mathsf{M}}_2 = \mathsf{M}(\underline{\boldsymbol{s}}_2) \, \mathsf{M}_2 \, . \tag{5.84} $$

Then the uncertain relative motion is

$$ \mathsf{M}(\underline{\boldsymbol{s}}_{12}) \, \mathsf{M}_{12} = (\mathsf{M}(\underline{\boldsymbol{s}}_1) \, \mathsf{M}_1)^{-1} \mathsf{M}(\underline{\boldsymbol{s}}_2) \, \mathsf{M}_2 \, ., \tag{5.85} $$

The mean of the relative rotations is

$$ \mathbb{E}({}^s\underline{\mathsf{M}}) = \mathbb{E}({}^s\underline{\mathsf{M}}_1)^{-1} \, \mathbb{E}({}^s\underline{\mathsf{M}}_2) \, . \tag{5.86} $$

The differential $\mathrm{d}\boldsymbol{s}_{12}$ of the rotation vector of the relative rotation is

$$ \boxed{\mathrm{d}\boldsymbol{s}_{12} = \mathsf{M}_1^{\mathsf{T}} (\mathrm{d}\boldsymbol{s}_2 - \mathrm{d}\boldsymbol{s}_1) \, .} \tag{5.87} $$

see Appendix 5.8.10

5.5.4 The motion in partially exponential representation

The uncertain motion is defined as

$$ {}^{\zeta}\underline{\mathsf{M}} = \begin{bmatrix} R(\boldsymbol{\rho})R & \boldsymbol{Z} + \underline{\boldsymbol{\tau}} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} , \qquad \mathbb{D}(\underline{\boldsymbol{\zeta}}) = \Sigma_{\zeta\zeta} \, , \tag{5.88} $$

We also can write this as a multiplication of a motion with a small random motion

$$ {}^{\zeta}\underline{\mathsf{M}} = \mathsf{M}(\underline{\boldsymbol{\zeta}}) \, \mathsf{M} \quad \text{with} \tag{5.89} $$

with the small motion

$$ \mathsf{M}(\boldsymbol{\zeta}) = \begin{bmatrix} R(\boldsymbol{\rho}) & (I_3 - R(\boldsymbol{\rho}))\boldsymbol{Z} + \boldsymbol{\tau} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \quad \text{and} \quad \mathsf{M} = \begin{bmatrix} R & \boldsymbol{Z} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \tag{5.90} $$

5.5.4.1 The adjoint motion for the partially exponential representation

Since the adjoint motion transfers small motions, we also can define an adjoint motion in case of the partially exponential representation. It is defined as the motion depending on the adjoint twist vector $\boldsymbol{\zeta}_{\mathrm{ad}}$

$$\mathsf{M}(\boldsymbol{\zeta}_{\mathrm{ad}}) = \mathsf{M}\,\mathsf{M}(\boldsymbol{\zeta})\,\mathsf{M}^{-1}\,. \tag{5.91}$$

Thus we have the form which can be used in proofs

$$\boxed{\mathsf{M}(\underline{\boldsymbol{\zeta}}_{\mathrm{ad}})\,\mathsf{M} = \mathsf{M}\,\mathsf{M}(\boldsymbol{\zeta})\,.} \tag{5.92}$$

Interestingly, also here we have a linear relationship between the differential adjoint twistvector $\mathrm{d}\boldsymbol{\zeta}_{\mathrm{ad}}$ and the differential original twist vector $\mathrm{d}\boldsymbol{\zeta}$:

$$\mathrm{d}\boldsymbol{\zeta}_{\mathrm{ad}} = {}^{\zeta}\mathsf{M}_{\mathrm{ad}}\;\mathrm{d}\boldsymbol{\zeta} \tag{5.93}$$

with the adjoint motion matrix

$$\boxed{{}^{\zeta}\mathsf{M}_{\mathrm{ad}} = \begin{bmatrix} R & 0 \\ S(\boldsymbol{Z})\,R & R \end{bmatrix},} \tag{5.94}$$

Observe, that the two adjoint matrices ${}^{s}\mathsf{M}_{\mathrm{ad}}$ in (5.73) and ${}^{\zeta}\mathsf{M}_{\mathrm{ad}}$ in (5.94) are identical. This results from the fact, that the adjoint motion for a differential twist has translation component zero, hence the two adjoint twist vectors do not differ if the original twist vectors are the same: The Jacobians in (5.68) then are unit matrices. This is the reason, why we did not indicate the difference in the naming of the adjoint matrices in Table 5.1 in row 5, columns 3 and 4.

5.5.4.2 The inverse motion in partially exponential representation

The uncertain inverse in partially exponential representation is defined as

$${}^{\zeta}\underline{\mathsf{M}}^{-1} = \begin{bmatrix} R(\underline{\boldsymbol{\rho}}^{(-1)})R^{\mathsf{T}} & -R^{\mathsf{T}}\boldsymbol{Z} + \underline{\boldsymbol{\tau}}^{(-1)} \\ 0 & 1 \end{bmatrix}, \tag{5.95}$$

and depends on the stochastic twist vector

$$\underline{\boldsymbol{\zeta}}^{(-1)} = \begin{bmatrix} \underline{\boldsymbol{\rho}}^{(-1)} \\ \underline{\boldsymbol{\tau}}^{(-1)} \end{bmatrix}. \tag{5.96}$$

As we saw in the last section, the differential adjunct twists are related to their twists via the adjoint motion matrix, which is identical for both cases. Therefore also the differential of the inverse twist vector in the partially exponential representation is given by

$$\boxed{\begin{bmatrix} \mathrm{d}\boldsymbol{\rho}^{(-1)} \\ \mathrm{d}\boldsymbol{\tau}^{(-1)} \end{bmatrix} = -\begin{bmatrix} R^{\mathsf{T}} & 0 \\ -R^{\mathsf{T}}S^{\mathsf{T}}(\boldsymbol{Z}) & R^{\mathsf{T}} \end{bmatrix}\begin{bmatrix} \mathrm{d}\boldsymbol{\rho} \\ \mathrm{d}\boldsymbol{\tau} \end{bmatrix},} \tag{5.97}$$

see Appendix 5.8.7. Observe, this is not the negative inverse of the adjunct motion matrix, since we have

$$\mathrm{Ad}(\mathsf{M})\;\mathrm{Ad}(\mathsf{M})^{-1} = \begin{bmatrix} R & 0 \\ S(\boldsymbol{Z})\,R & R \end{bmatrix}\begin{bmatrix} R^{\mathsf{T}} & 0 \\ -R^{\mathsf{T}}S(\boldsymbol{Z}) & R^{\mathsf{T}} \end{bmatrix} = \begin{bmatrix} I_3 & 0 \\ 0 & I_3 \end{bmatrix}, \tag{5.98}$$

and the second factor differs in the sign of the (2,1)-submatrix.

5.5.4.3 The concatenated motion in partially exponential representation

Let a possibly correlated motion pair be given by

$$\{{}^{\zeta}\underline{\mathcal{M}}_1, {}^{\zeta}\underline{\mathcal{M}}_2\}: \quad \left\{[\mathsf{M}_1, \mathsf{M}_2], \mathbb{D}\left(\begin{bmatrix} \underline{\zeta}_1 \\ \underline{\zeta}_2 \end{bmatrix}\right)\right\}. \tag{5.99}$$

The concatenated motion is

$${}^{\zeta}\underline{\mathcal{M}} = {}^{\zeta}\underline{\mathcal{M}}_2 {}^{\zeta}\underline{\mathcal{M}}_1: \quad {}^{\zeta}\underline{\mathsf{M}} = {}^{\zeta}\underline{\mathsf{M}}_2 \, {}^{\zeta}\underline{\mathsf{M}}_1. \tag{5.100}$$

We find the mean values of the concatenated motion is

$$\mathbb{E}({}^{\zeta}\underline{\mathsf{M}}) = \mathbb{E}\left({}^{\zeta}\underline{\mathsf{M}}_2\right) \; \mathbb{E}\left({}^{\zeta}\underline{\mathsf{M}}_1\right). \tag{5.101}$$

The differentials of the twist vectors also are linearly related by

$$\boxed{\mathrm{d}\boldsymbol{\rho} = \mathrm{d}\boldsymbol{\rho}_2 + R_2 \mathrm{d}\boldsymbol{\rho}_1 \quad \text{and} \quad \mathrm{d}\boldsymbol{\tau} = \mathrm{d}\boldsymbol{\tau}_2 + R_2 \mathrm{d}\boldsymbol{\tau}_1 - S(R_2 \boldsymbol{Z}_1)\mathrm{d}\boldsymbol{\rho}_2} \tag{5.102}$$

Observe, the rotation component transforms as for the exponential represenation and the translation component has a different term with the skew matrix. Moreover, and much more important: this matrix depends on both motions via $\boldsymbol{Z}_1$ and R_2, which complicates multiple concatenations.

5.5.4.4 The relative motion in partially exponential representation

Let a possibly correlated motion pair be given by

$$\{{}^{\zeta}\underline{\mathcal{M}}_1, {}^{\zeta}\underline{\mathcal{M}}_2\}: \quad \left\{[\mathsf{M}_1, \mathsf{M}_2], \mathbb{D}\left(\begin{bmatrix} \underline{\zeta}_1 \\ \underline{\zeta}_2 \end{bmatrix}\right)\right\}. \tag{5.103}$$

Then the relative pose can be determined by

$${}^{\zeta}\underline{\mathsf{M}}_{12} = {}^{\zeta}\underline{\mathsf{M}}_1^{-1} {}^{\zeta}\underline{\mathsf{M}}_2 = \begin{bmatrix} {}^{\zeta}\underline{R}_{12} & {}^{\zeta}\underline{\boldsymbol{Z}}_{12} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} = \begin{bmatrix} R(\underline{\boldsymbol{\rho}}_{12})R_{12} & \boldsymbol{Z}_{12} + \underline{\boldsymbol{\tau}}_{12} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix}. \tag{5.104}$$

or from

$${}^{\zeta}\underline{R}_{12} = {}^{\zeta}\underline{R}_1^{\mathsf{T}} {}^{\zeta}\underline{R}_2 \quad \text{and} \quad {}^{\zeta}\underline{\boldsymbol{Z}}_{12} = {}^{\zeta}\underline{R}_1^{\mathsf{T}}({}^{\zeta}\underline{\boldsymbol{Z}}_2 - {}^{\zeta}\underline{\boldsymbol{Z}}_1). \tag{5.105}$$

We obtain the mean relative motion as

$$\mathbb{E}({}^{\zeta}\underline{\mathsf{M}}_{12}) = \mathbb{E}({}^{\zeta}\underline{\mathsf{M}}_1)^{-1}\mathbb{E}({}^{\zeta}\underline{\mathsf{M}}_2) \tag{5.106}$$

Using the result from the uncertain relative rotation the differentials of the rotation and the translation vector are related by

$$\boxed{\mathrm{d}\boldsymbol{\rho}_{12} = R_1^{\mathsf{T}}(\mathrm{d}\boldsymbol{\rho}_2 - \mathrm{d}\boldsymbol{\rho}_1)} \tag{5.107}$$

and by variance propagation from (5.105)

$$\boxed{\mathrm{d}\boldsymbol{\tau}_{12} = R_1^{\mathsf{T}} S(\boldsymbol{Z}_2 - \boldsymbol{Z}_1)\mathrm{d}\boldsymbol{\rho}_1 + R_1^{\mathsf{T}}\mathrm{d}(\boldsymbol{\tau}_2 - \boldsymbol{\tau}_1).} \tag{5.108}$$

5.5.5 Evaluating the covariance matrix of estimated motions

We now discuss how to evaluate whether a theoretical covariance matrix is consistent with an empirical one.

Evaluating whether the theoretical covariance matrix $\Sigma_{\hat{\theta}\hat{\theta}}$ of estimated parameters $\boldsymbol{\theta}$ is trustworthy, it can be compared with the empirical covariance matrix $\widehat{\Sigma}_{\hat{\theta}\hat{\theta}}$ derived from

a sample of $\{\hat{\boldsymbol{\theta}}_k, k = 1, ..., K\}$, when knowing the true value $\tilde{\boldsymbol{\theta}}$, e.g., when using simulated data

$$\hat{\Sigma}_{\hat{\theta}\hat{\theta}} = \frac{1}{K} \sum_{k=1}^{K} (\hat{\boldsymbol{\theta}}_k - \tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_k - \tilde{\boldsymbol{\theta}})^\mathsf{T} . \tag{5.109}$$

In our context we, instead of the differences $\hat{\boldsymbol{\theta}}_k - \tilde{\boldsymbol{\theta}}$ of the estimated and the true parameters we use the estimated twist vectors $\widehat{\boldsymbol{m}}_k$, since their means are zero.

When evaluating the covariance matrix of estimated motions from a sample $\hat{\mathsf{M}}_k, k = 1, ..., K$ and a given true motion M we need to distinguish how we determine the empirical covariance matrix of the twist vector.

- In the case of the exponential representation we use the small matrices

$$\begin{aligned}
\mathsf{L}_k &= \hat{\mathsf{M}}_k \mathsf{M}^{-1} && (5.110)\\
&= \begin{bmatrix} \hat{R}_k & \hat{\boldsymbol{Z}}_k \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} R^\mathsf{T} & -R^\mathsf{T}\boldsymbol{T} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} && (5.111)\\
&= \begin{bmatrix} \hat{R}_k R^\mathsf{T} & \hat{\boldsymbol{Z}}_k - \hat{R}_k R^\mathsf{T} \boldsymbol{Z} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} && (5.112)\\
&\approx I_4 + \begin{bmatrix} S(\hat{\boldsymbol{r}}_k) & \hat{\boldsymbol{t}}_k \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} . && (5.113)
\end{aligned}$$

  and derive the small twist vectors $\hat{\boldsymbol{s}}_k = (\hat{\boldsymbol{r}}_k, \hat{\boldsymbol{t}}_k)$ from

$$\hat{\boldsymbol{r}}_k = \begin{bmatrix} L_{k23} \\ L_{k12} \\ L_{k31} \end{bmatrix} = \begin{bmatrix} (\hat{R}_k R^\mathsf{T})_{23} \\ (\hat{R}_k R^\mathsf{T})_{12} \\ (\hat{R}_k R^\mathsf{T})_{31} \end{bmatrix} \quad \text{and} \quad \hat{\boldsymbol{t}}_k = \begin{bmatrix} L_{k14} \\ L_{k24} \\ L_{k34} \end{bmatrix} = \hat{\boldsymbol{Z}}_k - \hat{R}_k R^\mathsf{T} \boldsymbol{Z} \tag{5.114}$$

  This also could be written compactly as

$$\hat{\boldsymbol{s}}_k = \log \left(\hat{\mathsf{M}}_k \mathsf{M}^{-1} \right)^\vee , \tag{5.115}$$

  the operator $^\vee$ (read: "vee") being the inverse of the operator $^\wedge$, thus, if $X = x^\wedge$ we have $x = X^\vee$.
  Then the empirical covariance matrix of $\hat{\underline{\boldsymbol{s}}}$ is

$$\hat{\Sigma}_{\hat{s}\hat{s}} = \frac{1}{K} \sum_k \boldsymbol{s}_k \boldsymbol{s}_k^\mathsf{T} . \tag{5.116}$$

- In the case of the partially multiplicative model we use

$$G_k = \hat{R}_k R^\mathsf{T} \approx I_3 + S(\boldsymbol{\rho}_k) \quad \text{and} \quad \boldsymbol{h}_k = \hat{\boldsymbol{Z}}_k - \boldsymbol{Z} = \boldsymbol{\tau}_k \tag{5.117}$$

  This leads to the elements of the small twist vector $\boldsymbol{\zeta}_k = (\boldsymbol{\rho}_k, \boldsymbol{\tau}_k)$

$$\boldsymbol{\rho}_k = \begin{bmatrix} G_{k23} \\ G_{k12} \\ G_{k31} \end{bmatrix} = \begin{bmatrix} (\hat{R}_k R^\mathsf{T})_{23} \\ (\hat{R}_k R^\mathsf{T})_{12} \\ (\hat{R}_k R^\mathsf{T})_{31} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\tau}_k = \boldsymbol{h}_k = \hat{\boldsymbol{Z}}_k - \boldsymbol{Z} . \tag{5.118}$$

  Then the empirical covariance matrix of $\hat{\underline{\boldsymbol{s}}}$ is

$$\hat{\Sigma}_{\hat{\zeta}\hat{\zeta}} = \frac{1}{I} \sum_k \boldsymbol{\zeta}_k \boldsymbol{\zeta}_k^\mathsf{T} . \tag{5.119}$$

As a result, linearizing the given model and deriving the empirical deviations of the estimated motions from the true motion need to be consistent.

In both cases we use a statistical test to check whether the expectation of the covariance matrix from the sample is identical to the theoretical covariance matrix, see (Förstner and Wrobel, 2016, Sect. 4.6.8.2).

5.6 Examples

We discuss two applications:

- Estimating motion parameters,
- Comparing absolute and relative poses.

5.6.1 Estimating motion parameters

Let us assume we have given I corresponding 3D points $\{\boldsymbol{X}, \boldsymbol{Y}\}_i, i = 1, ..., I$, where the coordinates $\boldsymbol{X}_i$ are fixed given values, and the coordinates $\boldsymbol{Y}_i$ are noisy observations of the corresponding moved points $\boldsymbol{X}_i$, having covariance matrices Σ_{ii}. We assume the correspondences are mutually independent, hence $\Sigma_{ii'} = \boldsymbol{0}$. Then, with the homogeneous coordinates

$$\mathbf{X}_i = \begin{bmatrix} \boldsymbol{X}_i \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{Y}_i = \begin{bmatrix} \boldsymbol{Y}_i \\ 1 \end{bmatrix} \tag{5.120}$$

we have the non-linear Gauss-Markov model (stochastical variables are underscored)

$$\mathbb{E}(\underline{\mathbf{Y}}_i) = \mathsf{M}\ \mathbf{X}_i \quad \text{and} \quad \mathbb{D}(\underline{\mathbf{Y}}_i) = \begin{bmatrix} \Sigma_{ii} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 0 \end{bmatrix} \quad \text{with} \quad i = 1, ..., I\,. \tag{5.121}$$

or, with the residuals (corrections),

$$\mathbf{Y}_i + \mathbf{v}_i = \mathsf{M}\ \mathbf{X}_i\,. \tag{5.122}$$

We assume we have an approximate motion matrix M^a. The model needs to be linearized, which depends on the type of representation.

5.6.1.1 Linearization with the exponential representation

With the exponential representation we have

$$\mathbf{Y}_i + \mathbf{v}_i = \mathsf{M}(\boldsymbol{s})\ \mathsf{M}^a\ \mathbf{X}_i = \mathsf{M}(\boldsymbol{s})\ \mathbf{X}_i^a \tag{5.123}$$

with the approximately *moved* coordinates

$${}^s\mathbf{X}_i^a = \mathsf{M}^a\ \mathbf{X}_i\,. \tag{5.124}$$

The goal is to estimate the twist vector $\boldsymbol{s}$ form the I correspondences. Linearization leads to

$$\mathbf{Y}_i + \mathbf{v}_i = (I_4 + A(\boldsymbol{s}))\ {}^s\mathbf{X}_i^a \tag{5.125}$$

where $\mathbf{v}_i$ are the residuals of (corrections to) the coordinates $\mathbf{Y}_i$. With the linearized observations

$${}^s\Delta\mathbf{y} = \mathbf{Y}_i - {}^s\mathbf{X}_i^a \tag{5.126}$$

this can be rewritten as

$$\begin{aligned}
{}^s\Delta\mathbf{y}_i + \mathbf{v}_i &= A(\boldsymbol{s}))\ {}^s\mathbf{X}_i^a && (5.127)\\
{}^s\Delta\mathbf{y}_i + \mathbf{v}_i &= \begin{bmatrix} S(\boldsymbol{r}) & \boldsymbol{t} \\ \mathbf{0}^\mathsf{T} & 0 \end{bmatrix} \begin{bmatrix} {}^s\boldsymbol{X}_i^a \\ 1 \end{bmatrix} && (5.128)\\
{}^s\Delta\boldsymbol{y}_i + \boldsymbol{v}_i &= S(\boldsymbol{r}){}^s\boldsymbol{X}_i^a + \boldsymbol{t} && (5.129)\\
& && (5.130)
\end{aligned}$$

thus finally

$${}^s\Delta\boldsymbol{y}_i + \mathbf{v}_i = [-S({}^sX_i^a) \mid I_3] \begin{bmatrix} \boldsymbol{s} \\ \boldsymbol{t} \end{bmatrix} \tag{5.131}$$

or the linear substitute model

$$\Delta \boldsymbol{y}_i + \mathbf{v}_i = {}^{s}\mathsf{X}_i \, \Delta\boldsymbol{\theta} \tag{5.132}$$

with the design matrix for each point and the unknown parameters

$${}^{s}\mathsf{X}_i = [-\mathsf{S}({}^{s}\boldsymbol{X}_i^a) \mid \mathsf{I}_3] \quad \text{and} \quad \Delta\boldsymbol{\theta} = \boldsymbol{s}\,. \tag{5.133}$$

The update of the parameters within the ν-th iteration is

$${}^{s}\mathsf{M}^{(\nu+1)} = \begin{bmatrix} \exp(\mathsf{S}(\hat{\boldsymbol{r}}^{(\nu)})) & \hat{\boldsymbol{t}}^{(\nu)} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \mathsf{M}^{(\nu)}\,. \tag{5.134}$$

5.6.1.2 Linearization with the partially exponential representation

With the exponential representation we have

$$\mathbf{Y}_i + \mathbf{v}_i = {}^{\zeta}\mathsf{M}\ \mathbf{X}_i = \begin{bmatrix} R(\boldsymbol{\rho})\ R & \boldsymbol{Z} + \boldsymbol{\tau} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_i \\ 1 \end{bmatrix} \tag{5.135}$$

Linerization leads to

$$\mathbf{Y}_i + \mathbf{v}_i = \begin{bmatrix} (\mathsf{I}_3 + \mathsf{S}(\boldsymbol{\rho}))\ R^a & \boldsymbol{Z}^a + \boldsymbol{\tau} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_i \\ 1 \end{bmatrix} \tag{5.136}$$

$$\mathbf{Y}_i + \mathbf{v}_i = \begin{bmatrix} R^a\ \boldsymbol{X}_i + \mathsf{S}(\boldsymbol{\rho})\ R^a\ \boldsymbol{X}_i + \boldsymbol{Z}^a + \boldsymbol{\tau} \\ 1 \end{bmatrix} \tag{5.137}$$

With the approximately *rotated* coordinates

$${}^{\zeta}\boldsymbol{X}_i^a = R^a\ \boldsymbol{X}_i \tag{5.138}$$

and the linearized observations

$${}^{\zeta}\Delta\boldsymbol{y} = \boldsymbol{Y}_i - (R^a\ \boldsymbol{X}_i + \boldsymbol{Z}^a) \tag{5.139}$$

we have the linearized model

$${}^{\zeta}\Delta\boldsymbol{y} + {}^{\zeta}\boldsymbol{v}_i = \mathsf{S}(\boldsymbol{\rho}){}^{\zeta}\boldsymbol{X}_i^a + \boldsymbol{\tau} \tag{5.140}$$

or finally

$${}^{\zeta}\Delta\boldsymbol{y} + {}^{\zeta}\boldsymbol{v}_i = {}^{\zeta}\mathsf{X}_i \, \Delta\boldsymbol{\theta} \tag{5.141}$$

with

$${}^{\zeta}\mathsf{X}_i = [-\mathsf{S}({}^{\zeta}\boldsymbol{X}_i^a) \mid \mathsf{I}_3] \quad \text{and} \quad \Delta\boldsymbol{\theta} = \boldsymbol{\zeta}\,. \tag{5.142}$$

The update of the parameters within the ν-th iteration is

$${}^{\zeta}\mathsf{M}^{(\nu+1)} = \begin{bmatrix} R(\hat{\boldsymbol{\rho}}^{(\nu)})\ R^{(\nu)} & \boldsymbol{Z}^{(\nu)} + \hat{\boldsymbol{\tau}}^{(\nu)} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix}. \tag{5.143}$$

5.6.1.3 Comparison

The design matrices differ in the argument of the skew matrix. For the exponential model we have explicitly

$${}^{s}\mathsf{X}_i = [-\mathsf{S}(R^a\boldsymbol{X}_i + \boldsymbol{Z}^a) \mid \mathsf{I}_3] \tag{5.144}$$

while for the partially exponential model we have

$${}^{\zeta}\mathsf{X}_i = [-\mathsf{S}(R^a\boldsymbol{X}_i) \mid \mathsf{I}_3] \tag{5.145}$$

Hence, the normal equation matrices

$${}^{s}\mathsf{N} = \sum_i {}^{s}\mathsf{X}_i^{\mathsf{T}}\ \Sigma_{ii}^{-1}\ {}^{s}\mathsf{X}_i \quad \text{and} \quad {}^{\zeta}\mathsf{N} = \sum_i {}^{\zeta}\mathsf{X}_i^{\mathsf{T}}\ \Sigma_{ii}^{-1}\ {}^{\zeta}\mathsf{X}_i \tag{5.146}$$

are differing in the rotation component and therefore also the inverse normal equation matrices, i.e., the covariance matrices of the estimated parameters.

Observe, in an extended Kalman filter for the motion parametrized by $\boldsymbol{x}$ with innovation of measurement residual $\boldsymbol{y}_k = \boldsymbol{z}_k - \boldsymbol{h}(\hat{\boldsymbol{x}}_{k|k-1})$ the Jacobian $H = \partial \boldsymbol{h}/\partial \boldsymbol{x}$ depends on the representation of the motion in the function $\boldsymbol{h}$, which may use one of the representations discussed in this note. The resulting covariance matrices will of course differ, depending on the choice of the representation.

5.6.2 Example for comparing absolute and relative poses in multi-view analysis

Let aus assume a free bundle block adjustment with two cameras at $\boldsymbol{Z}_t, t = 1, 2$ and 6 scene points $\boldsymbol{X}_i, i = 1..6$, as shown in Fig 5.1. The basis points towards the scene points, mimicking a docking situation. We are interested in precision of the relative motion.

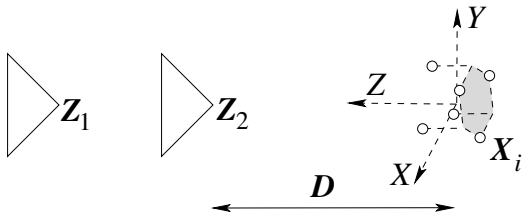

Figure 5.1: Relative motion from free bundle adjustment. The basis is 1 m. The distance D to the scene points is 2 m. The distance difference of the scene points is 0.3 m. The uncertainty of the image rays is 0.1 mrad

The free bundle adjustment with the software package BACS[7] (Schneider and Förstner, 2013) yields the covariance matrix of all pose parameters fixing the gauge in the centroid of the given scene points. The covariance matrix of the 12 parameters of the two twists is given by

$$\Sigma_{\hat{p}\hat{p}} = \mathbb{D}\left(\begin{bmatrix} \underline{\zeta}_1 \\ \underline{\zeta}_2 \end{bmatrix}\right) = SRS \tag{5.147}$$

where the diagonal matrix $S = \mathrm{Diag}([\sigma_{p_u}])$ contains the standard deviations, and the matrix $R = [\rho_{u'u''}]$ the correlations between the parameters. As an example we obtain the standard deviations for the rotations in [rad] and for the translations in [m]

$$\begin{bmatrix} \sigma_{\rho 11} & \sigma_{\tau 11} & \sigma_{\rho 21} & \sigma_{\tau 21} \\ \sigma_{\rho 12} & \sigma_{\tau 12} & \sigma_{\rho 22} & \sigma_{\tau 22} \\ \sigma_{\rho 13} & \sigma_{\tau 13} & \sigma_{\rho 23} & \sigma_{\tau 23} \end{bmatrix} = \begin{bmatrix} 0.0141 & 0.0425 & 0.0137 & 0.0278 \\ 0.0141 & 0.0425 & 0.0137 & 0.0278 \\ 0.0004 & 0.0121 & 0.0003 & 0.0078 \end{bmatrix} \tag{5.148}$$

The correlation matrix R is given by

$$\frac{1}{1000}\begin{bmatrix} 1000 & 0 & 0 & 0 & -1000 & 0 & 995 & 0 & 0 & 0 & -995 & 0 \\ 0 & 1000 & 0 & 1000 & 0 & 0 & 0 & 995 & 0 & 995 & 0 & 0 \\ 0 & 0 & 1000 & 0 & 0 & 0 & 0 & 0 & 7 & 0 & 0 & 0 \\ 0 & 1000 & 0 & 1000 & 0 & 0 & 0 & 995 & 0 & 995 & 0 & 0 \\ -1000 & 0 & 0 & 0 & 1000 & 0 & -995 & 0 & 0 & 0 & 995 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1000 & 0 & 0 & 0 & 0 & 0 & 992 \\ 995 & 0 & 0 & 0 & -995 & 0 & 1000 & 0 & 0 & 0 & -1000 & 0 \\ 0 & 995 & 0 & 995 & 0 & 0 & 0 & 1000 & 0 & 1000 & 0 & 0 \\ 0 & 0 & 7 & 0 & 0 & 0 & 0 & 0 & 1000 & 0 & 0 & 0 \\ 0 & 995 & 0 & 995 & 0 & 0 & 0 & 1000 & 0 & 1000 & 0 & 0 \\ -995 & 0 & 0 & 0 & 995 & 0 & -1000 & 0 & 0 & 0 & 1000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 992 & 0 & 0 & 0 & 0 & 0 & 1000 \end{bmatrix}$$

For symmetry reasons the rotations around and the translations along the X- and the Y-axes have the same standard deviation. Observe the position of the cameras w.r.t. scene is only 3 to 4 cm. Also, the rotation angles around the Y-and the Y-axis are 0.014 [rad] or appr. 0.8°. Also there are very high correlations between the two sets of pose parameters, some numerically nearly 1.

[7] *b*undle *a*djustment for *c*ameras *s*ystems

If we now determine the relative pose $\mathsf{M} := \mathsf{M}_{12} = \mathsf{M}_1^{-1}\mathsf{M}_2$ we obtain the following set of standard deviations of the twist vector of the relative pose using (5.107) and (5.108)

$$\begin{bmatrix} \sigma_{\rho_1} & \sigma_{\tau_1} \\ \sigma_{\rho_2} & \sigma_{\tau_2} \\ \sigma_{\rho_3} & \sigma_{\tau_3} \end{bmatrix} = \begin{bmatrix} 0.0014 & 0.0028 \\ 0.0014 & 0.0028 \\ 0.0005 & 0.0045 \end{bmatrix} \tag{5.149}$$

The precision of the rotations around and the translations along the X- and Y-axes are approximately 10-times more precise, which is cause by the high correlations of the corresponding pose parameters of the two cameras. The correlation matrix of the relative pose parameters is

$$\begin{bmatrix} 1.0000 & 0 & 0 & 0 & -0.4963 & 0 \\ 0 & 1.0000 & 0 & 0.4963 & 0 & 0 \\ 0 & 0 & 1.0000 & 0 & 0 & 0 \\ 0 & 0.4963 & 0 & 1.0000 & 0 & 0 \\ -0.4963 & 0 & 0 & 0 & 1.0000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1.0000 \end{bmatrix} \tag{5.150}$$

showing no correlations above 50 %.

5.7 MATLAB Software

The main routines are available as MATLAB-functions.

1	`calc_A_from_s.m`	$\mathsf{A} = \begin{bmatrix} S(\boldsymbol{r}) & \boldsymbol{t} \\ \mathbf{0}^\mathsf{T} & 0 \end{bmatrix}$
2	`calc_concatenated_M_s.m`	$\mathsf{M}(\underline{\boldsymbol{s}}) = \mathsf{M}(\underline{\boldsymbol{s}}_2) \cdot \mathsf{M}(\underline{\boldsymbol{s}}_1)$
4	`calc_concatenated_M_z.m`	$\mathsf{M}(\underline{\boldsymbol{\zeta}}) = \mathsf{M}(\underline{\boldsymbol{\zeta}}_2) \cdot \mathsf{M}(\underline{\boldsymbol{\zeta}}_1)$
5	`calc_concatenated_R.m`	$R(\underline{\boldsymbol{r}}) = R_2(\underline{\boldsymbol{r}}_2) \cdot R(\underline{\boldsymbol{r}}_1)$
6	`calc_inverse_M_s.m`	$\mathsf{M}(\underline{\boldsymbol{s}}^{(-1)}) = \mathsf{M}(\underline{\boldsymbol{s}})$
7	`calc_inverse_M_z.m`	$\mathsf{M}(\underline{\boldsymbol{\zeta}}^{(-1)}) = \mathsf{M}(\underline{\boldsymbol{\zeta}})$
8	`calc_inverse_R.m`	$R(\underline{\boldsymbol{r}}^{(-1)}) = R(\underline{\boldsymbol{r}})$
9	`calc_relative_M_s.m`	$\mathsf{M}(\underline{\boldsymbol{s}}) = \mathsf{M}^{-1}(\underline{\boldsymbol{s}}_1) \cdot \mathsf{M}(\underline{\boldsymbol{s}}_2)$
10	`calc_relative_M_z.m`	$\mathsf{M}(\underline{\boldsymbol{\zeta}}) = \mathsf{M}^{-1}(\underline{\boldsymbol{\zeta}}_1) \cdot \mathsf{M}(\underline{\boldsymbol{\zeta}}_2)$
11	`calc_relative_R.m`	$M(\underline{\boldsymbol{r}}) = R^\mathsf{T}(\underline{\boldsymbol{r}}_1) \cdot R(\underline{\boldsymbol{r}}_2)$
12	`calc_s_from_A.m`	$\mathsf{A} = \begin{bmatrix} S(\boldsymbol{r}) & \boldsymbol{t} \\ \mathbf{0}^\mathsf{T} & 0 \end{bmatrix} \rightarrow \boldsymbol{s} = \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{t} \end{bmatrix}$
13	`calc_z_from_M_M0.m`	$\mathsf{M} = \mathsf{M}(\boldsymbol{\zeta}) \cdot \mathsf{M}_0 \rightarrow \boldsymbol{\zeta}$

Table 5.3: MATLAB routines for rotations and motions in exponential and partially exponential representation

The variables for rotations and motions are structs:

$$\{\texttt{R.R, R.C}\} \quad \{\texttt{M.Ms, M.Cs}\} \quad \{\texttt{M.Mz, M.Cz}\}\,. \tag{5.151}$$

with the covariance matrices `*.C*` having the sizes 3×3, 6×6, and 6×6. For the input of the concatenated and relative rotations and motions we have structs for the transformation pairs:

$$\{\texttt{Rp.Rp, Rp.Cp}\} \quad \{\texttt{Mp.Msp, Mp.Csp}\} \quad \{\texttt{Mp.Mzp, Mp.Czp}\}\,. \tag{5.152}$$

Here the transformations are concatenated leading to

$$\texttt{Rp.Rp} = \texttt{[R1.R,R2.R]} \tag{5.153}$$
$$\texttt{Msp.Msp} = \texttt{[M1s.Ms,M2s.Ms]} \tag{5.154}$$
$$\texttt{Mzp.Mzp} = \texttt{[M1z.Mz,M2z.Mz]}\,. \tag{5.155}$$

The covariance matrices of the pairs allow for correlated transformation parameters, i.e.,

$$\texttt{Rp.Cp} := \begin{bmatrix} \Sigma_{r_1 r_1} & \Sigma_{r_1 r_2} \\ \Sigma_{r_2 r_1} & \Sigma_{r_2 r_2} \end{bmatrix} \tag{5.156}$$
$$\texttt{Msp.Csp} := \begin{bmatrix} \Sigma_{s_1 s_1} & \Sigma_{s_1 s_2} \\ \Sigma_{s_2 s_1} & \Sigma_{s_2 s_2} \end{bmatrix} \tag{5.157}$$
$$\texttt{Mzp.Czp} := \begin{bmatrix} \Sigma_{\zeta_1 \zeta_1} & \Sigma_{\zeta_1 \zeta_2} \\ \Sigma_{\zeta_2 \zeta_1} & \Sigma_{\zeta_2 \zeta_2} \end{bmatrix}. \tag{5.158}$$

In addition we have two routines for each representation to check the implementation:

- `check_basics_rotations.m` and `check_simulated_rotation.m`,
- `check_basics_motion_s.m` and `check_simulated_motion_s.m`, and
- `check_basics_motion_z.m` and `check_simulated_motion_z.m`.

One checks the basic relations:

- vector of adjoint transformtion,
- vector of inverse transformtion,
- function for inverse transformtion,
- vector of concatenated transformtion,
- vector of relative transformtion,
- difference transformation $\mathcal{T}_{12}$ as concatenation of $\mathcal{T}_1^{-1}$ and $\mathcal{T}_2$, hence $\mathcal{T}_{12} = \mathcal{T}_1^{-1} \circ \mathcal{T}_2$.

The output are differences between entities derived in two different manners, which therefore should be numerically small. If no relation fails the numerical test, the transformation is classified as `ok`.

The other checks the whether the mean parameters and their covariance matrix derived from a sample is identical to the given (theoretical) mean and covariance matrix. The output provides the test statistics for the covariance matrix and the mean and the corresponding critical region. E.g. for the exponentially represented motion we obtain:

```
Checks for motions s
Number U of unknown parameters = 6
Redundancy R                   = 6
Number K of samples            = 100
-------------------------------------------------
covariance matrix C_xx  ok: lambda   = 23.9562  in [5.8957,49.0108]
mean of parameters x    ok: mean(dx) =  6.2880  in [0.2994,24.1028]
```

If the prespecified noise standard deviation `sigma_n` is small, generally no test fails. If it is set to `sigma_n`= 0., it is likely that the tests fail due to neglected second order effects. Also, if the number `K` of samples is large, the statistical test becomes more sensitive, such that test statistics may lie outside the critical region.

Finally, the covariance matrices derived with the partially exponential and the exponential representations are compared assuming the motions have been generated with the partially exponential representation. The comparison shows, that the rotations together with their covariance matrix do not significantly differ, but the mean values do:

```
Checks for rotations s|z
Number U of unknown parameters = 3
Redundancy R                   = 3
Number K of samples            = 100
-------------------------------------------------
covariance matrix C_xx  ok: lambda   = 11.8448     in [0.2994,24.1028]
mean of parameters x    ok: mean(dx) =  1.6323     in [0.0153,17.7300]
+++++++++++++++++++++++++++++++++++++++++++++++++
Checks for translations s|z
Number U of unknown parameters = 3
Redundancy R                   = 3
Number K of samples            = 100
-------------------------------------------------
covariance matrix C_xx not ok: lambda   = 108.5578 not in [0.2994,24.1028] *****
mean of parameters x       ok: mean(dx) =   0.5210     in [0.0153,17.7300]
```

5.8 Appendix

5.8.1 Epipolar constraint using motion matrices

If the two images can be modelled as (see Förstner and Wrobel (2016, Eq. (12.34)), PCV)

$$\mathbf{x}' = [\mathsf{K}_1 \mid \mathbf{0}]\mathsf{M}_1^{-1}\mathbf{X} \quad \text{and} \quad \mathbf{x}'' = [\mathsf{K}_2 \mid \mathbf{0}]\mathsf{M}_2^{-1}\mathbf{X}$$

the projection rays are (see PCV (12.76))

$$\mathbf{l}_{\mathrm{x}'} = \mathsf{Q}_1\,\mathbf{L} \quad \text{and} \quad \mathbf{l}_{\mathrm{x}''} = \mathsf{Q}_2\,\mathbf{L}$$

with the projection matrices for lines

$$\mathsf{Q}_1 = [\mathit{0} \mid \mathsf{K}_1^{\circlearrowleft}]\;\mathsf{M}_{L,1}^{-1} \quad \text{and} \quad \mathsf{Q}_2 = [\mathit{0} \mid \mathsf{K}_2^{\circlearrowleft}]\;\mathsf{M}_{L,2}^{-1}$$

The motion matrix for lines and its inverse are given by (see PCV (12.75))

$$\mathsf{M}_L = \begin{bmatrix} \mathit{R} & \mathit{0} \\ \mathit{S}(\boldsymbol{Z})\mathit{R} & \mathit{R} \end{bmatrix} \quad \text{and} \quad \mathsf{M}_L^{-1} = \begin{bmatrix} \mathit{R}^{\mathsf{T}} & \mathit{0} \\ \mathit{R}^{\mathsf{T}}\mathit{S}^{\mathsf{T}}(\boldsymbol{Z}) & \mathit{R}^{\mathsf{T}} \end{bmatrix}$$

and identical to the adjoint motion matrix, see Table row 5:

$$\boxed{\mathsf{M}_L \equiv \mathrm{Ad}(\mathsf{M})} \tag{5.159}$$

Hence we have the line projection matrices

$$\mathsf{Q}_1 = [\mathit{0} \mid \mathsf{K}_1^{\circlearrowleft}]\;\mathsf{M}_{\mathrm{ad},1}^{-1} \quad \text{and} \quad \mathsf{Q}_2 = [\mathit{0} \mid \mathsf{K}_2^{\circlearrowleft}]\;\mathsf{M}_{\mathrm{ad},2}^{-1}$$

Two lines $\mathbf{L}_i, i = 1, 2$ intersect if $\mathbf{L}_1^{\mathsf{T}}\mathsf{D}\mathbf{L}_2 = 0$ (see PCV (7.100)), which is the basis for the definition of the fundamental matrix (see PCV (13.70))

$$\mathsf{F} = \mathsf{Q}_1\mathsf{D}\mathsf{Q}_2^{\mathsf{T}} = [\mathit{0} \mid \mathsf{K}^{\circlearrowleft}]\;\mathsf{M}_{\mathrm{ad},1}^{-1}\begin{bmatrix} \mathit{0} & \mathit{I} \\ \mathit{I} & \mathit{0} \end{bmatrix}\mathsf{M}_{\mathrm{ad},2}^{-\mathsf{T}}\begin{bmatrix} \mathit{0} \\ \mathsf{K}^{\circlearrowleft\mathsf{T}} \end{bmatrix}$$

which specializes to the essential matrix assuming the coordinate system in the left image and the motion M from the left to the right camera

$$\boxed{\mathsf{E} = [\mathit{I}_3 \mid \mathit{0}]\mathsf{M}_{\mathrm{ad},2}^{-\mathsf{T}}\begin{bmatrix} \mathit{0} \\ \mathit{I}_3 \end{bmatrix} = \mathit{S}(\boldsymbol{Z})\mathit{R}\,.} \tag{5.160}$$

5.8.2 Differential relation between Euler angles and the exponential representation

A rotation can be represented by Euler angles with the vector

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \tag{5.161}$$

e.g., as

$$R(\boldsymbol{\alpha}) = R_3(\alpha_3)R_2(\alpha_2)R_1(\alpha_1) \tag{5.162}$$

and by a multiplicative representation with a small vector

$$\boldsymbol{r} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \end{bmatrix} \tag{5.163}$$

as

$$R(\boldsymbol{r}, R^a) := R(\boldsymbol{r})R^a\,. \tag{5.164}$$

The task is to derive the Jacobian

$$J_{r\alpha} = \frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\alpha}}\,. \tag{5.165}$$

We start from the identity of the total derivative

$$\mathrm{d}R = \mathrm{d}R(\boldsymbol{\alpha}) = \mathrm{d}R(\boldsymbol{r}, R^a)\,. \tag{5.166}$$

and aim at finding a relation between $\mathrm{d}\boldsymbol{\alpha}$ and $\mathrm{d}\boldsymbol{r}$ under the assumption $R = R^a$, i.e., differential vectors $\mathrm{d}\boldsymbol{\alpha}$ and $\mathrm{d}\boldsymbol{r}$. We first obtain

$$\begin{aligned} \mathrm{d}R(\boldsymbol{\alpha}) &= \mathrm{d}\left(R_3(\alpha_3)R_2(\alpha_2)R_1(\alpha_1)\right) \qquad (5.167)\\ &= \mathrm{d}R_3(\alpha_3)\left(R_2(\alpha_2)R_1(\alpha_1)\right) + R_3(\alpha_3)\,\mathrm{d}R_2(\alpha_2)\,R_1(\alpha_1) + \left(R_3(\alpha_3)R_2(\alpha_2)\right)\mathrm{d}R_1(\alpha_1) \end{aligned}$$

Now we observe, e.g., for α_1

$$\begin{aligned} \mathrm{d}R_1(\alpha_1) &= \mathrm{d}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha_1 & -\sin\alpha_1 \\ 0 & \sin\alpha_1 & \cos\alpha_1 \end{bmatrix} && (5.168)\\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\sin\alpha_1 & -\cos\alpha_1 \\ 0 & \cos\alpha_1 & -\sin\alpha_1 \end{bmatrix}\mathrm{d}\alpha_1 && (5.169)\\ &= \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & +1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha_1 & -\sin\alpha_1 \\ 0 & \sin\alpha_1 & \cos\alpha_1 \end{bmatrix}\mathrm{d}\alpha_1 && (5.170)\\ &= S(\boldsymbol{e}_1)R_1(\alpha_1)\mathrm{d}\alpha_1\,, && (5.171) \end{aligned}$$

or generally

$$\mathrm{d}R_i(\alpha) = S(\boldsymbol{e}_i)R_i(\alpha)\mathrm{d}\alpha_i\,. \tag{5.172}$$

Similarly we thus have

$$\mathrm{d}R_2(\alpha_2) = S(\boldsymbol{e}_2)R_2(\alpha_2)\,\mathrm{d}\alpha_2 \quad \text{and} \quad \mathrm{d}R_3(\alpha_3) = S(\boldsymbol{e}_3)R_3(\alpha_3)\,\mathrm{d}\alpha_3 \tag{5.173}$$

This leads to

$$\begin{aligned} \mathrm{d}R(\boldsymbol{\alpha}) &= S(\boldsymbol{e}_3)\,R_3(\alpha_3)\,R_2(\alpha_2)\,R_1(\alpha_1)\,\mathrm{d}\alpha_3 + && (5.174)\\ &\quad R_3(\alpha_3)\,S(\boldsymbol{e}_2)\,R_2(\alpha_2)\,R_1(\alpha_1)\,\mathrm{d}\alpha_2 + && (5.175)\\ &\quad R_3(\alpha_3)\,R_2(\alpha_2)\,S(\boldsymbol{e}_1)\,R_1(\alpha_1)\,\mathrm{d}\alpha_1 && (5.176) \end{aligned}$$

We now use the relation $R(\boldsymbol{a} \times \boldsymbol{b}) = R\boldsymbol{a} \times R\boldsymbol{b}$ which is valid for all $\boldsymbol{b}$ in the form

$$RS(\boldsymbol{a}) = S(R\boldsymbol{a})R \quad \text{or} \quad RS(\boldsymbol{a})R^{\mathsf{T}} = S(R\boldsymbol{a})\,. \tag{5.177}$$

Then we obtain

$$\begin{aligned} \mathrm{d}R(\boldsymbol{\alpha}) &= S(\boldsymbol{e}_3)R\,\mathrm{d}\alpha_3 + && (5.178)\\ &\quad S(R_3(\alpha_3)\boldsymbol{e}_2)R\,\mathrm{d}\alpha_2 + && (5.179)\\ &\quad S(R_3(\alpha_3)R_2(\alpha_2)\boldsymbol{e}_1)R\,\mathrm{d}\alpha_1 && (5.180) \end{aligned}$$

or the skew symmetric matrix

$$\begin{aligned} \mathrm{d}R(\boldsymbol{\alpha})R^{\mathsf{T}} &= S(\boldsymbol{e}_3\mathrm{d}\alpha_3) + && (5.181)\\ &\quad S(R_3(\alpha_3)\boldsymbol{e}_2\mathrm{d}\alpha_2) + && (5.182)\\ &\quad S(R_3(\alpha_3)R_2(\alpha_2)\boldsymbol{e}_1\mathrm{d}\alpha_1) && (5.183) \end{aligned}$$

Now the total differential of $R(\boldsymbol{r}; R^a)$ is given by

$$\mathrm{d}R(\boldsymbol{r}, R^a) = S(\mathrm{d}\boldsymbol{r})R^a \tag{5.184}$$

Hence we have

$$\mathrm{d}R(\boldsymbol{r}, R^a)R^{a\mathsf{T}} = S(\mathrm{d}\boldsymbol{r}) \tag{5.185}$$

Since the approximate rotation matrix is the point of linearization, we have the constraint

$$\mathrm{d}R(\boldsymbol{\alpha})R^{\mathsf{T}} = \mathrm{d}R(\boldsymbol{r}, R^a)R^{\mathsf{T}} \tag{5.186}$$

Therefore the two skew symmetric matrices (5.181) and (5.185) need to be identical. From this we follow

$$\boldsymbol{e}_3\mathrm{d}\alpha_3 + R_3(\alpha_3)\boldsymbol{e}_2\mathrm{d}\alpha_2 + R_3(\alpha_3)R_2(\alpha_2)\boldsymbol{e}_1\mathrm{d}\alpha_1 = \mathrm{d}\boldsymbol{r} \tag{5.187}$$

or

$$\mathrm{d}\boldsymbol{r} = J_{r\alpha}\,\mathrm{d}\boldsymbol{\alpha} \tag{5.188}$$

with the Jacobian

$$\boxed{J_{r\alpha} = [R_3(\alpha_3)R_2(\alpha_2)\boldsymbol{e}_1 \mid R_3(\alpha_3)\boldsymbol{e}_2 \mid \boldsymbol{e}_3]} \tag{5.189}$$

The determinant of the Jacobian is

$$|J_{r\alpha}| = \cos\alpha_2\,. \tag{5.190}$$

This is why for $\cos\alpha_2 = 0$ or for $\alpha_2 = \pm 90°$ there is no unique relation between $\mathrm{d}\boldsymbol{r}$ and $\mathrm{d}\boldsymbol{\alpha}$, which is known as the Gimbal lock.

5.8.3 Adjoint motion matrix in exponential representation

We prove (5.73):

$$\mathrm{Ad}(\mathsf{M}) = \begin{bmatrix} R & 0 \\ S(\boldsymbol{Z})R & R \end{bmatrix}. \tag{5.191}$$

For this, we express the differential $\mathrm{d}\boldsymbol{s}_{\mathrm{ad}}$ of the small motion vector $\boldsymbol{s}_{\mathrm{ad}}$directly as a function of the differential $\mathrm{d}\boldsymbol{s}$. We start from (5.71)

$$\mathrm{d}\mathsf{M}(\boldsymbol{s}_{\mathrm{ad}})\,\mathsf{M} = \mathsf{M}\,\mathrm{d}\mathsf{M}(\boldsymbol{s}) \tag{5.192}$$

with its differential

$$\mathsf{A}(\mathrm{d}\boldsymbol{s}_{\mathrm{ad}})\,\mathsf{M} = \mathsf{M}\mathsf{A}(\mathrm{d}\boldsymbol{s}) \tag{5.193}$$

With the vector

$$\mathrm{d}\boldsymbol{s}_{\mathrm{ad}} = \begin{bmatrix} \mathrm{d}\boldsymbol{r}_{\mathrm{ad}} \\ \mathrm{d}\boldsymbol{t}_{\mathrm{ad}} \end{bmatrix} \tag{5.194}$$

this explicitly yields

$$\begin{bmatrix} S(\mathrm{d}\boldsymbol{r}_{\mathrm{ad}}) & \mathrm{d}\boldsymbol{t}_{\mathrm{ad}} \\ \mathbf{T} & 0 \end{bmatrix} \begin{bmatrix} R & \boldsymbol{Z} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} = \begin{bmatrix} R & \boldsymbol{Z} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} S(\mathrm{d}\boldsymbol{r}) & \mathrm{d}\boldsymbol{t} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix} \tag{5.195}$$

$$\begin{bmatrix} S(\mathrm{d}\boldsymbol{r}_{\mathrm{ad}})\, R & S(\mathrm{d}\boldsymbol{r}_{\mathrm{ad}})\, \boldsymbol{Z} + \mathrm{d}\boldsymbol{t}_{\mathrm{ad}} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix} = \begin{bmatrix} R\, S(\mathrm{d}\boldsymbol{r}) & R\mathrm{d}\boldsymbol{t} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix} \tag{5.196}$$

$$\begin{bmatrix} S(\mathrm{d}\boldsymbol{r}_{\mathrm{ad}})\, R & S(\mathrm{d}\boldsymbol{r}_{\mathrm{ad}})\, \boldsymbol{Z} + \mathrm{d}\boldsymbol{t}_{\mathrm{ad}} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix} = \begin{bmatrix} S(R\mathrm{d}\boldsymbol{r})\, R & R\mathrm{d}\boldsymbol{t} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix} \tag{5.197}$$

hence by comparing the upper left submatrices

$$\mathrm{d}\boldsymbol{r}_{\mathrm{ad}} = R\, \mathrm{d}\boldsymbol{r}\,, \tag{5.198}$$

and therefore

$$\mathrm{d}\boldsymbol{t}_{\mathrm{ad}} = R\, \mathrm{d}\boldsymbol{t} + S(\boldsymbol{Z})\mathrm{d}\boldsymbol{r}_{\mathrm{ad}}\,. \tag{5.199}$$

Compound this reads as

$$\boxed{\mathrm{d}\boldsymbol{s}_{\mathrm{ad}} = \mathrm{Ad}(\mathsf{M})\, \mathrm{d}\boldsymbol{s} \quad \text{with} \quad \mathrm{Ad}(\mathsf{M}) = \begin{bmatrix} R & 0 \\ S(\boldsymbol{Z})R & R \end{bmatrix}.} \tag{5.200}$$

with the adjoint motion matrix $\mathrm{Ad}(\mathsf{M})$.

5.8.4 Adjoint motion in partially exponential representation

We prove (5.94)

$$\mathrm{d}\boldsymbol{\zeta}_{\mathrm{ad}} = {}^{\zeta}\mathsf{M}_{\mathrm{ad}}\, \mathrm{d}\boldsymbol{\zeta} \quad \text{with} \quad {}^{\zeta}\mathsf{M}_{\mathrm{ad}} = \begin{bmatrix} R & 0 \\ S(\boldsymbol{Z})\, R & R \end{bmatrix} \tag{5.201}$$

with the vector

$$\mathrm{d}\boldsymbol{\zeta}_{\mathrm{ad}} = \begin{bmatrix} \mathrm{d}\boldsymbol{\rho}_{\mathrm{ad}} \\ \mathrm{d}\boldsymbol{\tau}_{\mathrm{ad}} \end{bmatrix}. \tag{5.202}$$

We have start from

$$\mathsf{M}(\mathrm{d}\boldsymbol{\zeta}_{\mathrm{ad}}) = \mathsf{M}\, \mathsf{M}(\mathrm{d}\boldsymbol{\zeta})\, \mathsf{M}^{-1} \tag{5.203}$$

The differential reads

$$\begin{aligned} \begin{bmatrix} S(\mathrm{d}\boldsymbol{\rho}_{\mathrm{ad}}) & \mathrm{d}\boldsymbol{\tau}_{\mathrm{ad}} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix} &= \begin{bmatrix} R & \boldsymbol{Z} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} S(\mathrm{d}\boldsymbol{\rho}) & \mathrm{d}\boldsymbol{\tau} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} R^{\mathsf{T}} & -R^{\mathsf{T}}\boldsymbol{Z} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \\ &= \begin{bmatrix} R\, S(\mathrm{d}\boldsymbol{r}) & R\mathrm{d}\boldsymbol{\tau} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} R^{\mathsf{T}} & -R^{\mathsf{T}}\boldsymbol{Z} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} && (5.204) \\ &= \begin{bmatrix} RS(\mathrm{d}\boldsymbol{\rho})R^{\mathsf{T}} & -RS(\mathrm{d}\boldsymbol{\rho})R^{\mathsf{T}}\boldsymbol{Z} + R\mathrm{d}\boldsymbol{\tau} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix} && (5.205) \\ &= \begin{bmatrix} S(R\mathrm{d}\boldsymbol{\rho}) & -S(R\mathrm{d}\boldsymbol{\rho})\boldsymbol{Z} + R\mathrm{d}\boldsymbol{\tau} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix} && (5.206) \\ &= \begin{bmatrix} S(R\mathrm{d}\boldsymbol{\rho}) & S(\boldsymbol{Z})R\mathrm{d}\boldsymbol{\rho} + R\mathrm{d}\boldsymbol{\tau} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{bmatrix} && (5.207) \end{aligned}$$

From the upper left sub-matrix we conclude

$$\mathrm{d}\boldsymbol{\rho}_{\mathrm{ad}} = R\, \mathrm{d}\boldsymbol{\rho}\,. \tag{5.208}$$

With this relation we obtain from the upper right part

$$\mathrm{d}\boldsymbol{\tau}_{\mathrm{ad}} = R\, \mathrm{d}\boldsymbol{\tau} + S(\boldsymbol{Z})R\, \mathrm{d}\boldsymbol{\rho}\,. \tag{5.209}$$

Joined this can be written as

$$\boxed{\mathrm{d}\boldsymbol{\zeta}_{\mathrm{ad}} = {}^{\zeta}\mathsf{M}_{\mathrm{ad}}\,\mathrm{d}\boldsymbol{\zeta} \quad \text{with} \quad {}^{\zeta}\mathsf{M}_{\mathrm{ad}} = \begin{bmatrix} R & 0 \\ S(\boldsymbol{Z})\,R & R \end{bmatrix}}\,. \tag{5.210}$$

with the adjoint motion matrix ${}^{\zeta}\mathsf{M}_{\mathrm{ad}}$, which is the same as for the exponential representation.

5.8.5 Uncertain inverse rotation

We prove (5.36). We have the relation

$$R(\underline{\boldsymbol{r}}^{(-1)})\,R^{-1} = (R(\underline{\boldsymbol{r}})\,R)^{-1} = R^{\mathsf{T}}R^{\mathsf{T}}(\underline{\boldsymbol{r}})\,. \tag{5.211}$$

Taking the total differential we obtain

$$S(\mathrm{d}\boldsymbol{r}^{(-1)})\,R^{-1} = R^{\mathsf{T}}S^{\mathsf{T}}(\mathrm{d}\boldsymbol{r})\,. \tag{5.212}$$

This yields

$$S(\mathrm{d}\boldsymbol{r}^{(-1)}) = R^{\mathsf{T}}S^{\mathsf{T}}(\mathrm{d}\boldsymbol{r})R = S(-R^{\mathsf{T}}\mathrm{d}\boldsymbol{r})\,. \tag{5.213}$$

Thus we obtain the Jacobian

$$\boxed{J_{r^{(-1)}r} = \frac{\partial \boldsymbol{r}^{(-1)}}{\partial \boldsymbol{r}} = -R^{\mathsf{T}}\,.} \tag{5.214}$$

Remark: If we would have defined the uncertain rotation with a noisy rotation from the right $\underline{R} = R\,R(\underline{\boldsymbol{r}})$, we would have obtained:

$$R^{-1}\,R(\underline{\boldsymbol{r}}^{(-1)}) = (R\,R(\underline{\boldsymbol{r}}))^{-1} = R^{\mathsf{T}}(\underline{\boldsymbol{r}})\,R^{\mathsf{T}}\,, \tag{5.215}$$

thus the differential

$$S(\mathrm{d}\boldsymbol{r}^{(-1)}) = R\,S^{\mathsf{T}}(\mathrm{d}\boldsymbol{r})\,R^{\mathsf{T}} = S^{\mathsf{T}}(R\mathrm{d}\boldsymbol{r})\,, \tag{5.216}$$

thus

$$\mathrm{d}\boldsymbol{r}^{(-1)} = -R_{\mathrm{ad}}\,\mathrm{d}\boldsymbol{r} = -R\,\mathrm{d}\boldsymbol{r}\,. \tag{5.217}$$

This relation is slightly more intuitive than (5.214). ◇

5.8.6 Uncertain inverse motion in exponential representation

We prove (5.78)

$$\mathrm{d}\boldsymbol{s}^{(-1)} = -\mathrm{Ad}(\mathsf{M})^{-1}\,\mathrm{d}\boldsymbol{s}\,. \tag{5.218}$$

We have the basic relation

$$\exp(\mathsf{A}(\boldsymbol{s}^{(-1)}))\cdot\mathsf{M}^{-1} = (\exp(\mathsf{A}(\boldsymbol{s}))\cdot\mathsf{M})^{-1} = \mathsf{M}^{-1}(\exp(\mathsf{A}(\boldsymbol{s}))^{-1}\,. \tag{5.219}$$

Taking the total differential, and using the first order approximation of $(\exp(X))^{-1} = I - X + 1/2X^2 - ...$ we obtain by taking the total differential

$$\mathsf{A}\left(\mathrm{d}\boldsymbol{s}^{(-1)}\right)\,\mathsf{M}^{-1} = -\mathsf{M}^{-1}\,\mathsf{A}(\mathrm{d}\boldsymbol{s})\,. \tag{5.220}$$

This yields

$$\mathsf{A}\left(\mathrm{d}\boldsymbol{s}^{(-1)}\right) = -\mathsf{M}^{-1}\,\mathsf{A}(\mathrm{d}\boldsymbol{s})\mathsf{M}\,, \tag{5.221}$$

or

$$\mathsf{M}\mathsf{A}\left(\mathrm{d}\boldsymbol{s}^{(-1)}\right)\mathsf{M}^{-1} = -\mathsf{A}(\mathrm{d}\boldsymbol{s}) \tag{5.222}$$

thus using (5.75)

$$\mathsf{A}\left(\mathrm{Ad}(\mathsf{M})\mathrm{d}\boldsymbol{s}^{(-1)}\right) = -\mathsf{A}(\mathrm{d}\boldsymbol{s}) \tag{5.223}$$

Therefore, we obtain the Jacobian

$$J_{s^{(-1)}s} = \frac{\partial \boldsymbol{s}^{(-1)}}{\partial \boldsymbol{s}} = -\mathsf{M}_{\mathrm{ad}}^{-1}\,, \tag{5.224}$$

which yields

$$\boxed{\mathrm{d}\boldsymbol{s}^{(-1)} = -\mathrm{Ad}(\mathsf{M})^{-1}\,\mathrm{d}\boldsymbol{s}\,.} \tag{5.225}$$

5.8.7 Uncertain inverse motion in partially exponential representation

We prove (5.97)

$$\begin{bmatrix} \mathrm{d}\boldsymbol{\rho}^{(-1)} \\ \mathrm{d}\boldsymbol{\tau}^{(-1)} \end{bmatrix} = -\begin{bmatrix} R^{\mathsf{T}} & 0 \\ -R^{\mathsf{T}}S^{\mathsf{T}}(\boldsymbol{Z}) & R^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \mathrm{d}\boldsymbol{\rho} \\ \mathrm{d}\boldsymbol{\tau} \end{bmatrix}. \tag{5.226}$$

It should hold

$$\begin{bmatrix} R(\underline{\boldsymbol{\rho}})R & \boldsymbol{Z}+\underline{\boldsymbol{\tau}} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} R(\underline{\boldsymbol{\rho}}^{(-1)})R^{\mathsf{T}} & -R^{\mathsf{T}}\boldsymbol{Z}+\underline{\boldsymbol{\tau}}^{(-1)} \\ 0 & 1 \end{bmatrix} = I_4\,, \tag{5.227}$$

or

$$\begin{bmatrix} R(\underline{\boldsymbol{\rho}})RR(\underline{\boldsymbol{\rho}}^{(-1)})R^{\mathsf{T}} & R(\underline{\boldsymbol{\rho}})R(-R^{\mathsf{T}}\boldsymbol{Z}+\underline{\boldsymbol{\tau}}^{(-1)})+\boldsymbol{Z}+\underline{\boldsymbol{\tau}} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} = I_4\,. \tag{5.228}$$

The differential of the upper left submatrix is

$$S(\mathrm{d}\boldsymbol{\rho}) + RR(\mathrm{d}\boldsymbol{\rho}^{(-1)})R^{\mathsf{T}} = S(\mathrm{d}\boldsymbol{\rho}) + R(R\mathrm{d}\boldsymbol{\rho}^{(-1)}) = 0\,. \tag{5.229}$$

Therefore we obtain

$$\mathrm{d}\boldsymbol{\rho}^{(-1)} = -R^{\mathsf{T}}\,\mathrm{d}\boldsymbol{\rho}\,. \tag{5.230}$$

The differential of the upper right matrix is

$$-S(\mathrm{d}\boldsymbol{\rho})\boldsymbol{Z} + R\mathrm{d}\boldsymbol{\tau}^{(-1)} + \mathrm{d}\boldsymbol{\tau} = 0 \tag{5.231}$$

This yields

$$\mathrm{d}\boldsymbol{\tau}^{(-1)} = R^{\mathsf{T}}S(\mathrm{d}\boldsymbol{\rho})\boldsymbol{Z} - R^{\mathsf{T}}\mathrm{d}\boldsymbol{\tau} = -R^{\mathsf{T}}S(\boldsymbol{Z})\mathrm{d}\boldsymbol{\rho} - R^{\mathsf{T}}\mathrm{d}\boldsymbol{\tau}\,. \tag{5.232}$$

This can be written as

$$\begin{bmatrix} \mathrm{d}\boldsymbol{\rho}^{(-1)} \\ \mathrm{d}\boldsymbol{\tau}^{(-1)} \end{bmatrix} = -\begin{bmatrix} R^{\mathsf{T}} & 0 \\ R^{\mathsf{T}}S(\boldsymbol{Z}) & R^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \mathrm{d}\boldsymbol{\rho} \\ \mathrm{d}\boldsymbol{\tau} \end{bmatrix} \tag{5.233}$$

5.8.8 Uncertain concatenated motions in exponential representation

We prove (5.82)

$$\mathrm{d}\boldsymbol{s} = \mathsf{M}_2\mathrm{d}\boldsymbol{s}_1 + \mathrm{d}\boldsymbol{s}_2\,. \tag{5.234}$$

We start from the total differential of $\underline{\mathsf{M}} = \underline{\mathsf{M}}_2\underline{\mathsf{M}}_1$:

$$\mathsf{A}(\underline{\boldsymbol{s}})\mathsf{M} = \mathsf{A}(\underline{\boldsymbol{s}}_2)\mathsf{M} + \mathsf{M}_2\mathsf{A}(\underline{\boldsymbol{s}}_1)\mathsf{M}_1 \tag{5.235}$$

or multiplying with $\mathsf{M}^{-1} = \mathsf{M}_1^{-1}\mathsf{M}_2^{-1}$ from the right

$$\mathsf{A}(\underline{\boldsymbol{s}}) = \mathsf{M}_2\mathsf{A}(\underline{\boldsymbol{s}}_1)\mathsf{M}_2^{-1} + \mathsf{A}(\underline{\boldsymbol{s}}_2)\,. \tag{5.236}$$

With (5.75) we thus obtain

$$\mathsf{A}(\underline{\boldsymbol{s}}) = \mathsf{A}(\mathsf{M}_{2\mathrm{ad}}\,\underline{\boldsymbol{s}}_1) + \mathsf{A}(\underline{\boldsymbol{s}}_2)\,. \tag{5.237}$$

This allows to express the differential motion parameters as

$$\mathrm{d}\boldsymbol{s} = \mathsf{M}_{2\mathrm{ad}}\mathrm{d}\boldsymbol{s}_1 + \mathrm{d}\boldsymbol{s}_2\,. \tag{5.238}$$

5.8.9 Uncertain concatenated motions in partially exponential representation

We prove (5.244)

$$\mathrm{d}\boldsymbol{\rho} = \mathrm{d}\boldsymbol{\rho}_2 + \mathrm{d}R_2\boldsymbol{\rho}_1 \quad \text{and} \quad \mathrm{d}\boldsymbol{\tau} = \mathrm{d}\boldsymbol{\tau}_2 + R_2\mathrm{d}\boldsymbol{\tau}_1 - S(R_2\boldsymbol{Z}_1)\mathrm{d}\boldsymbol{\rho}_2\,. \tag{5.239}$$

We explicitly have

$$\underline{\mathsf{M}} \;=\; \begin{bmatrix} R(\underline{\boldsymbol{\rho}})R & \boldsymbol{Z}+\underline{\boldsymbol{\tau}} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} = \begin{bmatrix} R(\underline{\boldsymbol{\rho}}) & (I_3 - R(\underline{\boldsymbol{\rho}}))\boldsymbol{Z} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} R & \boldsymbol{Z} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \tag{5.240}$$

and similarly

$$\underline{\mathsf{M}}_i \;=\; \begin{bmatrix} R(\underline{\boldsymbol{\rho}}_i)R_i & \boldsymbol{Z}_i+\underline{\boldsymbol{\tau}}_i \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix}$$

Therefore

$$\underline{\mathsf{M}} \;=\; \begin{bmatrix} R(\underline{\boldsymbol{\rho}}_2)R_2 & \boldsymbol{Z}_2+\underline{\boldsymbol{\tau}}_2 \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} R(\underline{\boldsymbol{\rho}}_1)R_1 & \boldsymbol{Z}_1+\underline{\boldsymbol{\tau}}_1 \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \tag{5.241}$$

$$= \begin{bmatrix} R(\underline{\boldsymbol{\rho}}_2)R_2R(\underline{\boldsymbol{\rho}}_1)R_1 & R(\underline{\boldsymbol{\rho}}_2)R_2(\boldsymbol{Z}_1+\underline{\boldsymbol{\tau}}_1)+\boldsymbol{Z}_2+\underline{\boldsymbol{\tau}}_2 \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \tag{5.242}$$

We now linearize, multiplicatively for R, additively for $\boldsymbol{Z}$:

$$\begin{bmatrix} \left(S(\mathrm{d}\boldsymbol{\rho}_2) + R_2S(\mathrm{d}\boldsymbol{\rho}_1)R_2^{\mathsf{T}}\right) R & (R_2\boldsymbol{Z}_1+\boldsymbol{Z}_2) + S(\mathrm{d}\boldsymbol{\rho}_2)R_2\boldsymbol{Z}_1 + R_2\mathrm{d}\boldsymbol{\tau}_1 + \mathrm{d}\boldsymbol{\tau}_2 \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix}$$

$$= \begin{bmatrix} (S(\mathrm{d}\boldsymbol{\rho}_2 + R_2\mathrm{d}\boldsymbol{\rho}_1))\;R & \boldsymbol{Z} + S(\mathrm{d}\boldsymbol{\rho}_2)R_2\boldsymbol{Z}_1 + R_2\mathrm{d}\boldsymbol{\tau}_1 + \mathrm{d}\boldsymbol{\tau}_2 \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \tag{5.243}$$

By comparison with (5.240) we find

$$\mathrm{d}\boldsymbol{\rho} = \mathrm{d}\boldsymbol{\rho}_2 + \mathrm{d}R_2\boldsymbol{\rho}_1 \quad \text{and} \quad \mathrm{d}\boldsymbol{\tau} = \mathrm{d}\boldsymbol{\tau}_2 + R_2\mathrm{d}\boldsymbol{\tau}_1 - S(R_2\boldsymbol{Z}_1)\mathrm{d}\boldsymbol{\rho}_2 \tag{5.244}$$

Relation to the concatenated motion with exponential representation. We can write (5.244) as

$$\mathrm{d}\boldsymbol{\zeta} = \mathsf{M}_{\mathrm{con1}}\,\mathrm{d}\boldsymbol{\zeta}_1 + \mathsf{M}_{\mathrm{con2}}\mathrm{d}\boldsymbol{\zeta}_2\,, \tag{5.245}$$

with

$$\mathsf{M}_{\mathrm{con1}} = \begin{bmatrix} R_2 & 0 \\ 0 & R_2 \end{bmatrix} \quad \text{and} \quad \mathsf{M}_{\mathrm{con2}} = \begin{bmatrix} I_3 & 0 \\ -S(R_2\boldsymbol{Z}_1) & I_3 \end{bmatrix}. \tag{5.246}$$

Using the Jacobians $J_{s\zeta}$ for switching between the representations, see (5.67) we can show, that this leads to the form

$$\mathrm{d}\boldsymbol{s} = \mathsf{M}_{2\mathrm{ad}}\,\mathrm{d}\boldsymbol{s}_1 + \mathrm{d}\boldsymbol{\zeta}_2 \quad \text{with} \quad \mathsf{M}_{2\mathrm{ad}} = \begin{bmatrix} R_2 & 0 \\ S(\boldsymbol{Z}_2)R_2 & R_2 \end{bmatrix}. \tag{5.247}$$

In detail we have

$$\mathrm{d}\boldsymbol{\zeta} \;=\; \mathsf{M}_{\mathrm{con1}}\,\mathrm{d}\boldsymbol{\zeta}_1 + \mathsf{M}_{\mathrm{con2}}\mathrm{d}\boldsymbol{\zeta}_2 \tag{5.248}$$

$$J_{\zeta s}\mathrm{d}\boldsymbol{s} \;=\; \mathsf{M}_{\mathrm{con1}}\,J_{1,\zeta s}\mathrm{d}\boldsymbol{s}_1 + \mathsf{M}_{\mathrm{con2}}J_{2,\zeta s}\mathrm{d}\boldsymbol{s}_2 \tag{5.249}$$

$$\mathrm{d}\boldsymbol{s} \;=\; J_{\zeta s}^{-1}\mathsf{M}_{\mathrm{con1}}\,J_{1,\zeta s}\mathrm{d}\boldsymbol{s}_1 + J_{\zeta s}^{-1}\mathsf{M}_{\mathrm{con2}}J_{2,\zeta s}\mathrm{d}\boldsymbol{s}_2 \tag{5.250}$$

Now we use

$$J_{\zeta s} = \begin{bmatrix} I_3 & 0 \\ -S(R_2\boldsymbol{Z}_1+\boldsymbol{Z}_2) & I_3 \end{bmatrix} \quad \text{and} \quad J_{i,\zeta s} = \begin{bmatrix} I_3 & 0 \\ -S(\boldsymbol{Z}_i) & I_3 \end{bmatrix}$$

and first obtain

$$J_{\zeta s}^{-1}\mathsf{M}_{\text{con1}}\, J_{1,\zeta s} = \begin{bmatrix} R_2 & 0 \\ S(R_2\boldsymbol{Z}_1+\boldsymbol{Z}_2)R_2 & R_2 \end{bmatrix} J_{1,\zeta s} \tag{5.251}$$

$$= \begin{bmatrix} R_2 & 0 \\ S(R_2\boldsymbol{Z}_1+\boldsymbol{Z}_2)R_2 - R_2S(\boldsymbol{Z}_1) & R_2 \end{bmatrix} \tag{5.252}$$

$$= \begin{bmatrix} R_2 & 0 \\ S(\boldsymbol{Z}_2)R_2 + \underbrace{S(R_2\boldsymbol{Z}_1)R_2 - R_2S(\boldsymbol{Z}_1)}_{=0} & R_2 \end{bmatrix} \tag{5.253}$$

$$= \begin{bmatrix} R_2 & 0 \\ S(\boldsymbol{Z}_2)R_2 & R_2 \end{bmatrix}. \tag{5.254}$$

Similarly we have

$$J_{\zeta s}^{-1}\mathsf{M}_{\text{con2}}J_{2,\zeta s} = \begin{bmatrix} I_3 & 0 \\ S(R_2\boldsymbol{Z}_1+\boldsymbol{Z}_2)-S(R_2\boldsymbol{Z}_1) & I_3 \end{bmatrix} J_{2,\zeta s} = \begin{bmatrix} I_3 & 0 \\ 0 & I_3 \end{bmatrix},$$

which yields

$$\mathrm{d}\boldsymbol{s} = \mathsf{M}_{2\text{ad}}\, \mathrm{d}\boldsymbol{s}_1 + \mathrm{d}\boldsymbol{s}_2\,. \tag{5.255}$$

5.8.10 Uncertain relative motion in exponential representation

We prove (5.87)

$$\mathrm{d}\boldsymbol{s}_{12} = \mathsf{M}_{1,\text{ad}}^{-1}\,(\mathrm{d}\boldsymbol{s}_2 - \mathrm{d}\boldsymbol{s}_1)\,. \tag{5.256}$$

The uncertain relative motion is

$$\mathsf{M}(\underline{\boldsymbol{s}}_{12})\,\mathsf{M}_{12} = (\mathsf{M}(\underline{\boldsymbol{s}}_1)\,\mathsf{M}_1)^{-1}\mathsf{M}(\underline{\boldsymbol{s}}_2)\,\mathsf{M}_2\,., \tag{5.257}$$

or

$$\mathsf{M}(\mathrm{d}\boldsymbol{s}_{12})\,\mathsf{M}_{12} = \mathsf{M}_1^{-1}\,\mathsf{M}^{-1}(\mathrm{d}\boldsymbol{s}_1)\mathsf{M}(\mathrm{d}\boldsymbol{s}_2)\,\mathsf{M}_2\,. \tag{5.258}$$

Taking the total differential, we obtain

$$\mathsf{A}(\mathrm{d}\boldsymbol{s}_{12})\,\mathsf{M}_{12} = \mathsf{M}_1^{-1}\,\mathsf{A}(-\mathrm{d}\boldsymbol{s}_1)\,\mathsf{M}_2 + \mathsf{M}_1^{-1}\,\mathsf{A}(\mathrm{d}\boldsymbol{s}_2)\,\mathsf{M}_2\,. \tag{5.259}$$

or

$$\mathsf{A}(\mathrm{d}\boldsymbol{s}_{12}) = \mathsf{M}_1^{-1}\,\mathsf{A}(\mathrm{d}\boldsymbol{s}_2 - \mathrm{d}\boldsymbol{s}_1)\,\mathsf{M}_1 \tag{5.260}$$

or

$$\mathsf{M}_1\mathsf{A}(\mathrm{d}\boldsymbol{s}_{12})\mathsf{M}_1^{-1} = \mathsf{A}(\mathrm{d}\boldsymbol{s}_2 - \mathrm{d}\boldsymbol{s}_1) \tag{5.261}$$

Hence, with

$$\mathsf{M}_1\mathsf{A}(\mathrm{d}\boldsymbol{s}_{12})\mathsf{M}_1^{-1} = \mathsf{A}(\mathrm{Ad}(\mathsf{M}_1)\,\mathrm{d}\boldsymbol{s}_{12}) \tag{5.262}$$

Therefore we finally have the relation

$$\mathrm{d}\boldsymbol{s}_{12} = \mathsf{M}_{1,\text{ad}}^{-1}\,(\mathrm{d}\boldsymbol{s}_2 - \mathrm{d}\boldsymbol{s}_1)\,. \tag{5.263}$$

Check using the inverse and the concatenation We start from the concatenation

$$\mathsf{M} = \mathsf{M}_2\mathsf{M}_1\,, \tag{5.264}$$

use (5.78) and (5.238)

$$\mathrm{d}\boldsymbol{s}^{(-1)} = -\mathsf{M}_{\text{ad}}^{-1}\,\mathrm{d}\boldsymbol{s} \quad \text{and} \quad \mathrm{d}\boldsymbol{s} = \mathrm{Ad}(\mathsf{M}_2)\,\mathrm{d}\boldsymbol{s}_1 + \mathrm{d}\boldsymbol{s}_2\,, \tag{5.265}$$

and apply this to

$$\mathsf{M}_{12} = \mathsf{M}_1^{-1}\mathsf{M}_2\,. \tag{5.266}$$

This yields

$$\mathrm{d}\boldsymbol{s}_{12} = \mathsf{M}_{1,\text{ad}}^{-1}\,\mathrm{d}\boldsymbol{s}_2 - \mathsf{M}_{1,\text{ad}}^{-1}\,\mathrm{d}\boldsymbol{s}_1 = \mathsf{M}_{1,\text{ad}}^{-1}\,(\mathrm{d}\boldsymbol{s}_2 - \mathrm{d}\boldsymbol{s}_1)\,. \tag{5.267}$$

5.8.11 Uncertain relative motion in partially exponential representation

We prove (5.107) and (5.108)

$$\mathrm{d}\boldsymbol{\rho}_{12} = R_1^{\mathsf{T}}\mathrm{d}(\boldsymbol{\rho}_2 - \boldsymbol{\rho}_1) \quad \text{and} \quad \mathrm{d}\boldsymbol{\tau}_{12} = R_1^{\mathsf{T}} S(\boldsymbol{Z}_2 - \boldsymbol{Z}_1)\mathrm{d}\boldsymbol{\rho}_1 + R_1^{\mathsf{T}}\mathrm{d}(\boldsymbol{\tau}_2 - \boldsymbol{\tau}_1)\,. \tag{5.268}$$

We start from

$$\underline{\mathsf{M}}_i := \begin{bmatrix} R(\underline{\boldsymbol{\rho}}_i)R_i & Z_i + \underline{\boldsymbol{\tau}}_i \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \tag{5.269}$$

and obtain

$$\begin{aligned} \underline{\mathsf{M}}_{12} &= \underline{\mathsf{M}}_1^{-1}\,\underline{\mathsf{M}}_2 && (5.270)\\ &= \begin{bmatrix} R(\underline{\boldsymbol{\rho}}_1)R_1 & Z_1 + \underline{\boldsymbol{\tau}}_1 \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix}^{-1} \begin{bmatrix} R(\underline{\boldsymbol{\rho}}_2)R_2 & Z_2 + \underline{\boldsymbol{\tau}}_2 \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} && (5.271)\\ &= \begin{bmatrix} R_1^{\mathsf{T}} R^{\mathsf{T}}(\underline{\boldsymbol{\rho}}_1) & -(R_1^{\mathsf{T}} R^{\mathsf{T}}(\underline{\boldsymbol{\rho}}_1))(Z_1 + \underline{\boldsymbol{\tau}}_1) \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} R(\underline{\boldsymbol{\rho}}_2)R_2 & Z_2 + \underline{\boldsymbol{\tau}}_2 \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \\ &= \begin{bmatrix} R_1^{\mathsf{T}} R^{\mathsf{T}}(\underline{\boldsymbol{\rho}}_1)R(\underline{\boldsymbol{\rho}}_2)R_2 & R_1^{\mathsf{T}} R^{\mathsf{T}}(\underline{\boldsymbol{\rho}}_1)(Z_2 + \underline{\boldsymbol{\tau}}_2) - (R_1^{\mathsf{T}} R^{\mathsf{T}}(\underline{\boldsymbol{\rho}}_1))(Z_1 + \underline{\boldsymbol{\tau}}_1) \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \end{aligned}$$

Linearizing the rotation multiplicatively and the translation additively we have

$$\begin{aligned} \mathsf{M}_{12} &\approx \begin{bmatrix} R_1^{\mathsf{T}}(S^{\mathsf{T}}(\mathrm{d}\boldsymbol{\rho}_1) + S(\mathrm{d}\boldsymbol{\rho}_2))R_2 & R_1^{\mathsf{T}} S^{\mathsf{T}}(\mathrm{d}\boldsymbol{\rho}_1)\boldsymbol{Z}_2 + R_1^{\mathsf{T}}\mathrm{d}\boldsymbol{\tau}_2) - R_1^{\mathsf{T}} S^{\mathsf{T}}(\mathrm{d}\boldsymbol{\rho}_1))\boldsymbol{Z}_1 - R_1^{\mathsf{T}}\mathrm{d}\boldsymbol{\tau}_1) \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \\ &= \begin{bmatrix} R_1^{\mathsf{T}} S(\mathrm{d}\boldsymbol{\rho}_2 - \mathrm{d}\boldsymbol{\rho}_2) & R_1^{\mathsf{T}} S(\boldsymbol{Z}_2 - \boldsymbol{Z}_1)\mathrm{d}\boldsymbol{\rho}_1(Z_2 + \mathrm{d}\boldsymbol{\tau}_2) + R_1^{\mathsf{T}}(\mathrm{d}\boldsymbol{\tau}_2 - \mathrm{d}\boldsymbol{\tau}_1) \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \end{aligned} \tag{5.272}$$

Check using the inverse and the concatenation We start from the concatenation

$$\mathsf{M} = \mathsf{M}_l \mathsf{M}_r\,, \tag{5.273}$$

use (5.97) and (5.244)

$$\mathrm{d}\boldsymbol{\zeta}^{(-1)} = \begin{bmatrix} \mathrm{d}\boldsymbol{\rho}^{(-1)} \\ \mathrm{d}\boldsymbol{\tau}^{(-1)} \end{bmatrix} = \begin{bmatrix} -R^{\mathsf{T}} & 0 \\ -R^{\mathsf{T}} S(\boldsymbol{Z}) & -R^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \mathrm{d}\boldsymbol{\rho} \\ \mathrm{d}\boldsymbol{\tau} \end{bmatrix} \tag{5.274}$$

and

$$\begin{bmatrix} \mathrm{d}\boldsymbol{\rho} \\ \mathrm{d}\boldsymbol{\tau} \end{bmatrix} = \begin{bmatrix} R_l \mathrm{d}\boldsymbol{\rho}_r + \mathrm{d}\boldsymbol{\rho}_l \\ R_l \mathrm{d}\boldsymbol{\tau}_r + \mathrm{d}\boldsymbol{\tau}_l - S(R_l \boldsymbol{Z}_r)\mathrm{d}\boldsymbol{\rho}_l \end{bmatrix}. \tag{5.275}$$

and apply this to

$$\mathsf{M}_{lr} := \underbrace{\mathsf{M}_1^{-1}}_{\mathsf{M}_l}\,\underbrace{\mathsf{M}_2}_{\mathsf{M}_r} = \begin{bmatrix} R_1^{\mathsf{T}} & -R_1^{\mathsf{T}}\boldsymbol{Z}_1 \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \begin{bmatrix} R_2 & \boldsymbol{Z}_2 \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} = \begin{bmatrix} \underbrace{R_1^{\mathsf{T}} R_2}_{R_{12}} & \underbrace{R_1^{\mathsf{T}}(\boldsymbol{Z}_2 - \boldsymbol{Z}_1)}_{\boldsymbol{Z}_{12}} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix}. \tag{5.276}$$

We obtain

$$\underline{\mathsf{M}}_{lr} := \begin{bmatrix} R(\underline{\boldsymbol{\rho}}_{12})R_{12} & Z_{12} + \underline{\boldsymbol{\tau}}_{12} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \tag{5.277}$$

We use

$$\begin{bmatrix} \mathrm{d}\boldsymbol{\rho}_r \\ \mathrm{d}\boldsymbol{\tau}_r \end{bmatrix} := \begin{bmatrix} \mathrm{d}\boldsymbol{\rho}_2 \\ \mathrm{d}\boldsymbol{\tau}_2 \end{bmatrix} \tag{5.278}$$

and

$$\begin{bmatrix} \mathrm{d}\boldsymbol{\rho}_l \\ \mathrm{d}\boldsymbol{\tau}_l \end{bmatrix} := \begin{bmatrix} -R_1^{\mathsf{T}} & 0 \\ -R_1^{\mathsf{T}} S(\boldsymbol{Z}_1) & -R_1^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} \mathrm{d}\boldsymbol{\rho}_1 \\ \mathrm{d}\boldsymbol{\tau}_1 \end{bmatrix} = \begin{bmatrix} -R_1^{\mathsf{T}}\mathrm{d}\boldsymbol{\rho}_1 \\ -R_1^{\mathsf{T}} S(\boldsymbol{Z}_1)\mathrm{d}\boldsymbol{\rho}_1 - R_1^{\mathsf{T}}\mathrm{d}\boldsymbol{\tau}_1 \end{bmatrix} \tag{5.279}$$

Now, we have

$$\begin{bmatrix} \mathrm{d}\boldsymbol{\rho}_{lr} \\ \mathrm{d}\boldsymbol{\tau}_{lr} \end{bmatrix} = \begin{bmatrix} \mathrm{d}\boldsymbol{\rho}_{12}^{(-1)} \\ \mathrm{d}\boldsymbol{\tau}_{12}^{(-1)} \end{bmatrix} \tag{5.280}$$

$$= \begin{bmatrix} R_1^\mathsf{T}\mathrm{d}\boldsymbol{\rho}_2 - R_1^\mathsf{T}\mathrm{d}\boldsymbol{\rho}_1 \\ R_1^\mathsf{T}\mathrm{d}\boldsymbol{\tau}_2 - R_1^\mathsf{T}\mathrm{d}\boldsymbol{\tau}_1 - R_1^\mathsf{T}S(\boldsymbol{Z}_1)\mathrm{d}\boldsymbol{\rho}_1 - S(R_1^\mathsf{T}\boldsymbol{Z}_2)(-R_1^\mathsf{T}\mathrm{d}\boldsymbol{\rho}_1) \end{bmatrix}$$

$$= \begin{bmatrix} R_1^\mathsf{T}(\mathrm{d}\boldsymbol{\rho}_2 - \mathrm{d}\boldsymbol{\rho}_1) \\ R_1^\mathsf{T}(\mathrm{d}\boldsymbol{\tau}_2 - \mathrm{d}\boldsymbol{\tau}_1) + R_1^\mathsf{T}S(-\boldsymbol{Z}_1 + \boldsymbol{Z}_2)\mathrm{d}\boldsymbol{\rho}_1 \end{bmatrix} \tag{5.281}$$

6 Centroid Form of an Uncertain Plane

A plane can be represented in various manners. We especially discuss the centroid form of an uncertain plane, which naturally results from estimating a plane from a given point set. We discuss the representation, its recursive estimation assuming isotropic point uncertainty and optimal estimation.

6.1 Problem . . . 81
6.2 Centroid Representation of a Plane . . . 82
6.2.1 The Representation . . . 82
6.2.2 Covariance Matrix of the Plane Parameters . . . 83
6.3 Uncertain Plane from 3D Points . . . 84
6.3.1 Fitting a plane through 3D points with isotropic uncertainty . . . 84
6.3.2 Fitting a plane through a set of 3D points with arbitrary covariance matrix . . . 86
6.3.3 Checking a Set of Points for Planarity . . . 90
6.4 Estimating a Mean Plane . . . 90
6.4.1 Estimating the mean plane using moments . . . 91
6.4.2 Approximate estimating the mean plane using plane parameters . . . 92
6.4.3 An optimal solution based on the centroid representation . . . 93
6.5 Motion from Plane to Plane correspondences . . . 95
6.5.1 Problem Statement . . . 95
6.5.2 Minimal Solution for the Motion from Three Plane Correspondences 96
6.5.3 An Iterative Solution . . . 96
6.5.4 Theoretical Accuracy of the Motion . . . 97

6.1 Problem

This note (2020) collects methods for representing and estimating uncertain planes. It focusses on the geometrically intuitive centroid representation, naturally resulting from fitting a plane through a point cloud. We collect methods for estimating a plane from scene points, for averaging uncertain planes and for estimating a motion for plane correspondences.

The statistically rigorous estimation, discussed here in Sect. 6.3.2, has the advantage of giving insight into the uncertainty structure, whereas the solution based on spherically normalized homogeneous plane coordinates in Note 7 is technically more elegant, and easily generalizes to the estimation of multiple planes.

A natural representation of an uncertain plane is its centroid form

$$\mathcal{A}: \quad \{\boldsymbol{X}_0, \boldsymbol{Q}; \sigma_q, \sigma_\phi, \sigma_\psi\} , \tag{6.1}$$

see Fig. 6.1. This representation can directly be derived from a set of 3D points $\boldsymbol{X}_i, i = 1, ..., I$ with isotropic uncertainty $\Sigma_{X_i X_i} = \sigma_i^2 I_3$.

This note addresses three problems, namely

1. the estimation of a plane from uncertain points,
2. the estimation of a spatial motion from plane-to-point correspondences, and
3. the estimation of a spatial motion from plane-to-plane correspondences.

6.2 Centroid Representation of a Plane

6.2.1 The Representation

The centroid representation of a plane is given by (see Fig. 6.1)

$$\boxed{\mathcal{A}: \quad \{\boldsymbol{X}_0, Q; \sigma_q, \sigma_\phi, \sigma_\psi\}\ .} \tag{6.2}$$

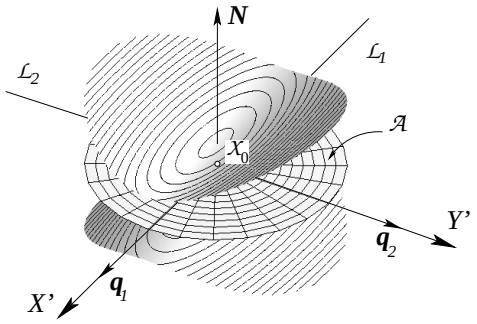

Figure 6.1: Uncertain plane $\mathcal{A}$. Its center is $\mathcal{X}_0$; the center is that point of the plane where the uncertainty across the (perpendicular to the) plane is smallest; it is uncertain along the normal by σ_q. Its normal is $\boldsymbol{N}$; its rotational uncertainty is composed of two independent uncertain rotations around $\mathcal{L}_1$ and $\mathcal{L}_2$ which are mutually perpendicular. The standard deviations σ_ϕ and σ_ψ are the uncertainties of the X'- and Y'-components of the normal $\boldsymbol{N}$. The three directions form an orthonormal tripod $Q = [\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{N}]$

Here we have:

- the coordinates of the centroid $\mathcal{Z}$;
- the rotation matrix
$$Q = [\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3] \tag{6.3}$$
with its normal
$$\boldsymbol{N} = \boldsymbol{q}_3 = Q\boldsymbol{e}_3 \tag{6.4}$$
and the local coordinate system $[\boldsymbol{q}_1, \boldsymbol{q}_2]$ in the plane, where $\boldsymbol{q}_1$ is the major axis, and $\boldsymbol{q}_2$ is the minor axis of the moment matrix point cloud, when projected into the plane.
- the variance σ_q^2 across the plane;
- the variances σ_ϕ^2 of the normal around $\boldsymbol{q}_2$ and σ_ψ^2 around $\boldsymbol{q}_1$.

The point $\mathcal{Z}_0(\boldsymbol{Z}_0)$ closest to the origin is given by

$$\boldsymbol{Z}_0 = D\boldsymbol{N}\,. \tag{6.5}$$

We will represent the coordinates $\boldsymbol{X}_0$ of the centroid $\mathcal{X}_0$ as the sum of two orthogonal vectors $\boldsymbol{Z}_0$ and $\boldsymbol{M}$

$$\boldsymbol{X}_0 = \boldsymbol{Z}_0 + \boldsymbol{M} = Q(D\boldsymbol{N}'' + \boldsymbol{M}'')\,. \tag{6.6}$$

see Fig. 6.2, and – represented in the rotated coordinate system (see Fig. 6.2 right) –

$$\boldsymbol{N}'' = \boldsymbol{N}' = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \boldsymbol{e}_3 \quad \boldsymbol{M}'' = \begin{bmatrix} M_X'' \\ M_Y'' \\ 0 \end{bmatrix}\ . \tag{6.7}$$

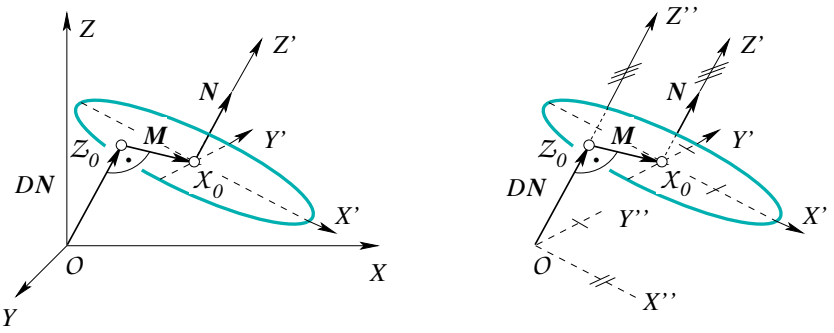

Figure 6.2: Representation of uncertain plane. **Left:** Relation between global frame (XYZ) and the local frame $(X'Y'Z')$. **Right:** Relation between the global system rotated by Q $(X''Y''Z'')$ and the local frame $(X'Y'Z')$, which are parallel

6.2.2 Covariance Matrix of the Plane Parameters

The standard deviations can be derived by transforming the points into the coordinate system $(X'Y'Z')$ of their weighted centroid. Then we only have three uncertain parameters collected in the 3-vector

$$\boldsymbol{A}^\circ = \begin{bmatrix} A_1^\circ \\ A_2^\circ \\ A_3^\circ \end{bmatrix}. \tag{6.8}$$

We have

- the uncertain Z'-coordinate A_1° of the centroid, and
- the uncertain X'- and Y'-coordinates (A_2°, A_3°) of the normal.

Hence we represent the uncertainty of the plane by

$$\mathbb{D}(\underline{\boldsymbol{A}}^\circ) = \begin{bmatrix} \sigma_q^2 & & \\ & \sigma_\phi^2 & \\ & & \sigma_\psi^2 \end{bmatrix}. \tag{6.9}$$

The three parameters are related to the centroid and the normal by

$$\Delta\boldsymbol{A}^* := \begin{bmatrix} \Delta\boldsymbol{X}_0 \\ \Delta\boldsymbol{N} \end{bmatrix} = \begin{bmatrix} Q & \begin{bmatrix} 0 \\ 0 \\ \Delta A_1^\circ \end{bmatrix} \\ Q & \begin{bmatrix} \Delta A_2^\circ \\ \Delta A_3^\circ \\ 0 \end{bmatrix} \end{bmatrix} = J_r(Q)\Delta\boldsymbol{A}^\circ \tag{6.10}$$

with

$$\underset{6\times 3}{J_r(Q)} = \begin{bmatrix} Q & \\ & Q \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_3 & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{q}_1 & \boldsymbol{q}_2 \end{bmatrix}. \tag{6.11}$$

The covariance matrices of the centroid and the normal then can be given directly. The centroid and the normal are statistically uncorrelated.

The centroid $\underline{\boldsymbol{X}}_0$ is only uncertain across the plane, hence in the direction of the normal

$$\boxed{\Sigma_{X_0X_0} = Q \begin{bmatrix} 0 & & \\ & 0 & \\ & & \sigma_q^2 \end{bmatrix} Q^\mathsf{T} = \sigma_q^2 \boldsymbol{N}\boldsymbol{N}^\mathsf{T}.} \tag{6.12}$$

The uncertainty of the normal $\boldsymbol{N}$ is

$$\boxed{\Sigma_{NN} = \mathsf{Q} \begin{bmatrix} \sigma_\phi^2 & & \\ & \sigma_\psi^2 & \\ & & 0 \end{bmatrix} \mathsf{Q}^\mathsf{T} = \sigma_\phi^2 \boldsymbol{q}_1 \boldsymbol{q}_1^\mathsf{T} + \sigma_\psi^2 \boldsymbol{q}_2 \boldsymbol{q}_2^\mathsf{T} .} \tag{6.13}$$

Hence the direction of the major uncertainty of the normal of the plane is coded in the covariance matrix. Eqs. (6.12) and (6.13) clarify, why we only need the rotation matrix Q and the three standard deviations σ_q, σ_ϕ, and σ_ψ for representing the uncertainty of the plane. The rotation matrix Q this is responsible for both, the normal and the covariance matrix of the plane.

The inverse relation is

$$\Delta \boldsymbol{A}^\circ = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathsf{Q}^\mathsf{T} & \\ & \mathsf{Q}^\mathsf{T} \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{X}_0 \\ \Delta \boldsymbol{N} \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_3^\mathsf{T} & \boldsymbol{0}^\mathsf{T} \\ \boldsymbol{0}^\mathsf{T} & \boldsymbol{q}_1^\mathsf{T} \\ \boldsymbol{0}^\mathsf{T} & \boldsymbol{q}_2^\mathsf{T} \end{bmatrix} \begin{bmatrix} \Delta \boldsymbol{X}_0 \\ \Delta \boldsymbol{N} \end{bmatrix} =: J_r^\mathsf{T}(\mathsf{Q}) \Delta \boldsymbol{A}^* . \tag{6.14}$$

which has covariance matrix

$$\Sigma_{A^\circ A^\circ} = \begin{bmatrix} \sigma_q^2 & & \\ & \sigma_\phi^2 & \\ & & \sigma_\psi^2 \end{bmatrix} . \tag{6.15}$$

Hence, if a point $\mathcal{X}_i$ lies on the plane $\mathcal{A}$, then the point $\mathcal{X}_i'([X_i', Y_i', 0])$ lies on the plane $\mathcal{A}'$, which is the $X'Y'$-plane. The points $\mathcal{X}_i$ and $\mathcal{X}_i'$ are related by

$$\boldsymbol{X}_i = \mathsf{Q} \boldsymbol{X}_i' + \boldsymbol{X}_0 \quad \text{or} \quad \boldsymbol{X}_i' = \mathsf{Q}^\mathsf{T} (\boldsymbol{X}_i - \boldsymbol{X}_0) . \tag{6.16}$$

6.3 Uncertain Plane from 3D Points

6.3.1 Fitting a plane through 3D points with isotropic uncertainty

Given are I uncertain 3D points $\mathcal{X}_i, i = 1, ..., I$, with $\{\boldsymbol{X}_i, \sigma_i^2 I_3\}$.

1. We can show that the best fitting plane $\mathcal{A}(\mathbf{A})$ with

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{N} \\ -D \end{bmatrix} \tag{6.17}$$

passes through the weighted centroid $\boldsymbol{X}_0$, that its normal $\boldsymbol{N}$ is the eigenvector of the (unweighted) moment matrix belonging to the smallest eigenvalue, and that it is given by $\boldsymbol{N}^\mathsf{T}(\boldsymbol{X} - \boldsymbol{X}_0) = 0$.
The moment matrix is

$$M = \sum_i w_i (\boldsymbol{X}_i - \boldsymbol{X}_0)(\boldsymbol{X}_i - \boldsymbol{X}_0)^\mathsf{T} = \mathsf{Q} \boldsymbol{\Lambda} \mathsf{Q}^\mathsf{T} = \lambda_1 \boldsymbol{q}_1 \boldsymbol{q}_1^\mathsf{T} + \lambda_2 \boldsymbol{q}_2 \boldsymbol{q}_2^\mathsf{T} + \lambda_3 \boldsymbol{q}_3 \boldsymbol{q}_3^\mathsf{T} \tag{6.18}$$

with

$$\boxed{w_i = \frac{1}{\sigma_i^2} \quad \text{and} \quad \boldsymbol{X}_0 = \frac{\sum_i w_i \boldsymbol{X}_i}{\sum w_i}} \tag{6.19}$$

and the rotation matrix,

$$\mathsf{Q} = [\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3] , \tag{6.20}$$

and the diagonal matrix

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix} , \tag{6.21}$$

where the eigenvalues are sorted in decreasing order. The normal is

$$\boxed{\boldsymbol{N} = \boldsymbol{q}_3 = \boldsymbol{e}_3^\mathsf{T} \mathsf{Q} .} \tag{6.22}$$

2. (Exercise) Show the *theoretical variances* of the parameters of a plane through I equally weighted ($w_i = 1$) 3D points $\mathcal{X}_i$ with standard deviation σ for all coordinates can be determined from

$$\sigma_q^2 = \frac{\sigma^2}{I} \qquad \sigma_\phi^2 = \frac{\sigma^2}{\lambda_1} \qquad \sigma_\psi^2 = \frac{\sigma^2}{\lambda_2}\,, \tag{6.23}$$

where σ_q^2 is the variance of the position of the plane in the direction of the normal and σ_ϕ^2 and σ_ψ^2 are the variances of rotations around the two principle axes of the point set.
Hint: Translate the point cloud into the origin and rotate it such that the two major axes of the moment matrix fall into the X- and the Y-coordinate axes. Then apply the reasoning from the chapter on the best fitting 2D line.
Using the weighted moment matrix, for general weights this generalizes to

$$\boxed{\sigma_q^2 = \frac{1}{I\bar{w}} \quad \sigma_\phi^2 = \frac{1}{\lambda_1} \quad \sigma_\psi^2 = \frac{1}{\lambda_2}} \tag{6.24}$$

3. (Exercise) Show that the *estimated variance* of the plane's position q perpendicular to the plane and the two principle normal directions are given by

$$\sigma_q^2 = \frac{1}{I-3}\,\frac{\lambda_3}{I} \qquad \sigma_\phi^2 = \frac{1}{I-3}\,\frac{\lambda_3}{\lambda_1} \qquad \sigma_\psi^2 = \frac{1}{I-3}\,\frac{\lambda_3}{\lambda_2}\,. \tag{6.25}$$

Using the weighted moment matrix, for general weights this generalizes to

$$\boxed{\hat{\sigma}_q^2 = \frac{1}{I-3}\frac{\lambda_3}{I\bar{w}} \quad \hat{\sigma}_\phi^2 = \frac{1}{I-3}\frac{\lambda_3}{\lambda_1} \quad \hat{\sigma}_\psi^2 = \frac{1}{I-3}\frac{\lambda_3}{\lambda_2}\,.} \tag{6.26}$$

6.3.1.1 Relation to moments and recursive estimation

Now we observe, that the parameters specifying an uncertain plane can be uniquely derived from the non-central moments. They allow a simple and possibly recursive estimation of the mean of several planes.

The non-central moments are

$$m_{kln} = \sum_i w_i X_i^k Y_i^l Z_i^n \quad \text{with} \quad k+l+n \in \{0,1,2\} \tag{6.27}$$

namely

$$m_{000} = \sum_i w_i \tag{6.28}$$
$$m_{100} = \sum_i w_i X_i \tag{6.29}$$
$$m_{010} = \sum_i w_i Y_i \tag{6.30}$$
$$m_{001} = \sum_i w_i Z_i \tag{6.31}$$
$$m_{200} = \sum_i w_i X_i^2 \tag{6.32}$$
$$m_{110} = \sum_i w_i X_i Y_i \tag{6.33}$$
$$m_{101} = \sum_i w_i X_i Z_i \tag{6.34}$$
$$m_{020} = \sum_i w_i Y_i^2 \tag{6.35}$$
$$m_{011} = \sum_i w_i Y_i Z_i \tag{6.36}$$
$$m_{002} = \sum_i w_i Z_i^2 \tag{6.37}$$

together with the number of points

$$I = \sum_i 1 \,. \tag{6.38}$$

Especially we have

$$I\overline{w} = m_{000} \tag{6.39}$$
$$X_0 = m_{100}/m_{000} \tag{6.40}$$
$$Y_0 = m_{010}/m_{000} \tag{6.41}$$
$$Z_0 = m_{001}/m_{000} \tag{6.42}$$
$$\mu_{200} = m_{200}/m_{000} - X_0^2 \tag{6.43}$$
$$\mu_{110} = m_{110}/m_{000} - X_0 Y_0 \tag{6.44}$$
$$\mu_{101} = m_{101}/m_{000} - X_0 Z_0 \tag{6.45}$$
$$\mu_{020} = m_{020}/m_{000} - Y_0^2 \tag{6.46}$$
$$\mu_{011} = m_{011}/m_{000} - Y_0 Z_0 \tag{6.47}$$
$$\mu_{002} = m_{002}/m_{000} - Z_0^2 \tag{6.48}$$
$$M = \begin{bmatrix} \mu_{200} & \mu_{110} & \mu_{101} \\ \mu_{110} & \mu_{020} & \mu_{011} \\ \mu_{101} & \mu_{011} & \mu_{002} \end{bmatrix} . \tag{6.49}$$

The eigenvalues of the moment matrix yield the variances of the position and the normal via (6.26). Hence we have a mapping from the moments $\boldsymbol{m}$ (including the number of points I) to the centroid form $\boldsymbol{c}$ of the plane

$$\boxed{\boldsymbol{m} \mapsto \boldsymbol{c}: \qquad \boldsymbol{c} = \boldsymbol{c}(\boldsymbol{m}) \quad \text{or} \quad \left\{ \boldsymbol{X}_0, Q; \sigma_q^2, \sigma_\phi^2, \sigma_\psi^2 \right\} \leftarrow \{m_{000}, ..., m_{002}, I\} \,.} \tag{6.50}$$

6.3.2 Fitting a plane through a set of 3D points with arbitrary covariance matrix

We can assume to have approximate values, thus only need to update these using an iterative scheme, where often only one iteration is necessary.

6.3.2.1 An iterative solution

We start from the nonlinear constraints

$$g_i(\widehat{\boldsymbol{X}}_i; \widehat{\boldsymbol{N}}, \hat{D}) = \widehat{\boldsymbol{N}}^{\mathsf{T}} \widehat{\boldsymbol{X}}_i - \hat{D} = 0 \tag{6.51}$$

where D is the distance of the plane to the origin. We will later find the centroid $\boldsymbol{X}_0$ on the plane. In addition, we have the length constraint for the normal

$$h(\widehat{\boldsymbol{N}}) = \frac{1}{2}\left(|\widehat{\boldsymbol{N}}|^2 - 1\right) = 0 \tag{6.52}$$

Starting from approximate values for the unknown parameters and the fitted observations we thus have the linearized model

$$g_i(\widehat{\boldsymbol{X}}_i; \widehat{\boldsymbol{N}}, \hat{D}) = \widehat{\boldsymbol{N}}^{a\mathsf{T}} \widehat{\boldsymbol{X}}_i^a - \hat{D}^a + \widehat{\boldsymbol{N}}^{a\mathsf{T}} \widehat{\Delta \boldsymbol{X}}_i + \widehat{\boldsymbol{X}}_i^{a\mathsf{T}} \widehat{\Delta \boldsymbol{N}} + \widehat{\Delta D} = 0 \tag{6.53}$$

or

$$g_i(\widehat{\boldsymbol{X}}_i; \widehat{\boldsymbol{N}}, \hat{D}) = g_i(\widehat{\boldsymbol{X}}_i^a; \widehat{\boldsymbol{N}}^a, \hat{D}^a) + \boldsymbol{a}_i^{\mathsf{T}} \widehat{\Delta \boldsymbol{\theta}} + \boldsymbol{b}_i^{\mathsf{T}} \widehat{\Delta \boldsymbol{y}} = 0 \tag{6.54}$$

with the corrections to the unknown parameters, collected in a 4-vector

$$\widehat{\Delta \boldsymbol{\theta}} := \begin{bmatrix} \widehat{\Delta \boldsymbol{N}} \\ \widehat{\Delta D} \end{bmatrix}, \quad \widehat{\Delta \boldsymbol{y}_i} := \widehat{\Delta \boldsymbol{X}}_i, \quad \boldsymbol{a}_i := \begin{bmatrix} \widehat{\boldsymbol{X}}_i^a \\ 1 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{b}_i := \boldsymbol{N}^a. \tag{6.55}$$

Therefore we have the normal equations

$$M \widehat{\Delta \boldsymbol{p}} = \boldsymbol{m} \quad \text{or} \quad \begin{bmatrix} N & H \\ H^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \widehat{\Delta \boldsymbol{\theta}} \\ \lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{n} \\ c_h \end{bmatrix} \tag{6.56}$$

with

$$N = \sum_i w_{q_i} \boldsymbol{a}_i \boldsymbol{a}_i^{\mathsf{T}}, \tag{6.57}$$

$$\boldsymbol{n} = \sum_i w_{q_i} (\boldsymbol{a}_i (-g_i + \boldsymbol{b}_i^{\mathsf{T}} (\widehat{\boldsymbol{X}}_i - \boldsymbol{X}_i)), \tag{6.58}$$

$$H = \widehat{\boldsymbol{N}}^a, \tag{6.59}$$

$$c_g = -(|\widehat{\boldsymbol{N}}^a|^2 - 1) \tag{6.60}$$

$$w_{q_i} = \frac{1}{\boldsymbol{b}_i^{\mathsf{T}} \Sigma_{y_i y_i} \boldsymbol{b}_i} := \frac{1}{\boldsymbol{N}^{a\mathsf{T}} \Sigma_{X_i X_i} \boldsymbol{N}^a} \tag{6.61}$$

We use the following update for the normal

$$\widehat{\boldsymbol{N}} = \mathsf{N}(\widehat{\boldsymbol{N}}^a + \widehat{\Delta \boldsymbol{N}}) \tag{6.62}$$

The covariance matrix of the parameters results from the inverse of the normal equation matrix, or, when eliminating the Lagrangian parameter from the 4×4 matrix

$$\Sigma_{\hat{\theta}\hat{\theta}} = (N + HH^{\mathsf{T}})^{-1} - HH^{\mathsf{T}} = \begin{bmatrix} \Sigma_{\widehat{N}\widehat{N}} & \Sigma_{\widehat{N}\hat{D}} \\ \Sigma_{\hat{D}\widehat{N}} & \Sigma_{\hat{D}\hat{D}} \end{bmatrix}, \tag{6.63}$$

which has rank 3, and generally is a full matrix.

6.3.2.2 Choosing the Local Coordinate System

We now choose the points reduced to some reference frame with center $\boldsymbol{X}_0$ and axes Q

$$\boldsymbol{X}_i = Q \boldsymbol{X}_i' + \boldsymbol{X}_0 \quad \text{or} \quad \boldsymbol{X}_i' = Q^{\mathsf{T}} (\boldsymbol{X}_i - \boldsymbol{X}_0), \tag{6.64}$$

In homogeneous coordinates this is

$$\begin{bmatrix} \boldsymbol{X}_i \\ 1 \end{bmatrix} = \begin{bmatrix} Q & \boldsymbol{X}_0 \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{X}'_i \\ 1 \end{bmatrix} \tag{6.65}$$

The plane therefore transforms as

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{A}_h \\ A_0 \end{bmatrix} = \begin{bmatrix} Q & \boldsymbol{X}_0 \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}^\mathsf{-} = \begin{bmatrix} \boldsymbol{A}'_h \\ A'_0 \end{bmatrix} = \mathbf{A}' \tag{6.66}$$

or explicitly

$$\begin{bmatrix} \boldsymbol{A}_h \\ A_0 \end{bmatrix} = \begin{bmatrix} Q & \mathbf{0} \\ -\boldsymbol{X}_0^\mathsf{T} Q & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{A}'_h \\ A'_0 \end{bmatrix}. \tag{6.67}$$

The normal $\boldsymbol{N} = \boldsymbol{A}_h$ therefore is transformed as

$$\boldsymbol{N} = Q\boldsymbol{N}' \quad \text{or} \quad \boldsymbol{N}' = Q^\mathsf{T}\boldsymbol{N}. \tag{6.68}$$

The distance $D = -A_0$ to the origin is transferred as

$$D = D' + \boldsymbol{X}_0^\mathsf{T} Q\boldsymbol{N}' = D' + \boldsymbol{X}_0^\mathsf{T}\boldsymbol{N} \quad \text{or} \quad D' = D - \boldsymbol{N}^\mathsf{T}\boldsymbol{X}_0. \tag{6.69}$$

The covariance matrices transform as

$$\begin{aligned} \mathbb{D}(\mathbf{A}) &= \begin{bmatrix} \Sigma_{NN} & \Sigma_{ND} \\ \Sigma_{DN} & \sigma_D^2 \end{bmatrix} && (6.70)\\ &= \begin{bmatrix} Q & 0 \\ \boldsymbol{X}_0^\mathsf{T} Q & 1 \end{bmatrix} \begin{bmatrix} \Sigma_{N'N'} & \mathbf{0} \\ \mathbf{0} & \sigma_{D'}^2 \end{bmatrix} \begin{bmatrix} Q^\mathsf{T} & Q^\mathsf{T}\boldsymbol{X}_0 \\ 0 & 1 \end{bmatrix} && (6.71)\\ &= \begin{bmatrix} Q\Sigma_{N'N'}Q^\mathsf{T} & Q\Sigma_{N'N'}Q^\mathsf{T}\boldsymbol{X}_0 \\ \boldsymbol{X}_0^\mathsf{T} Q\Sigma_{N'N'}Q^\mathsf{T} & \boldsymbol{X}_0^\mathsf{T} Q\Sigma_{N'N'}Q^\mathsf{T}\boldsymbol{X}_0 + \sigma_{D'}^2 \end{bmatrix} && (6.72) \end{aligned}$$

and

$$\begin{aligned} \mathbb{D}(\mathbf{A}') &= \begin{bmatrix} \Sigma_{N'N'} & \Sigma_{N'D'} \\ \Sigma_{D'N'} & \sigma_{D'}^2 \end{bmatrix} && (6.73)\\ &= \begin{bmatrix} Q^\mathsf{T} & 0 \\ -\boldsymbol{X}_0^\mathsf{T} & 1 \end{bmatrix} \begin{bmatrix} \Sigma_{NN} & \mathbf{0} \\ \mathbf{0} & \sigma_D^2 \end{bmatrix} \begin{bmatrix} Q^\mathsf{T} & -\boldsymbol{X}_0 \\ 0 & 1 \end{bmatrix} && (6.74)\\ &= \begin{bmatrix} Q^\mathsf{T}\Sigma_{NN}Q & -Q^\mathsf{T}\Sigma_{NN}\boldsymbol{X}_0 \\ -\boldsymbol{X}_0^\mathsf{T}\Sigma_{NN}Q & \boldsymbol{X}_0^\mathsf{T}\Sigma_{NN}\boldsymbol{X}_0 + \sigma_D^2 \end{bmatrix} && (6.75) \end{aligned}$$

We first choose Q such that the covariance matrix

$$\Sigma_{\widehat{N}\widehat{N}} = Q\Sigma_{\widehat{N}'\widehat{N}'}Q^\mathsf{T} = Q \begin{bmatrix} \sigma_{\hat{\phi}}^2 & & \\ & \sigma_{\hat{\psi}}^2 & \\ & & 0 \end{bmatrix} Q^\mathsf{T} \tag{6.76}$$

of the normal is diagonal, which can be achieved by an eigenvalue decomposition of $\Sigma_{\widehat{N}\widehat{N}}$. Then the normal is (6.4)

$$\underline{\widehat{\boldsymbol{N}}}' = N\left(\begin{bmatrix} \underline{\hat{N}}'_X \\ \underline{\hat{N}}'_Y \\ 1 \end{bmatrix}\right) = N\left(\begin{bmatrix} \underline{\widehat{\boldsymbol{N}}}'_r \\ 1 \end{bmatrix}\right) = Q^\mathsf{T}\underline{\boldsymbol{N}}. \tag{6.77}$$

After the diagonalization we obtain the covariance matrix, where the distance D'' is in the rotated and not yet translated system:

$$\mathbb{D}\left(\begin{bmatrix} \underline{\widehat{\boldsymbol{N}}}' \\ \underline{\hat{D}}'' \end{bmatrix}\right) = \begin{bmatrix} \mathrm{Diag}([\sigma_{\hat{\phi}}^2, \sigma_{\hat{\psi}}^2]) & \mathbf{0} & \Sigma_{\widehat{N}'_r\hat{D}''} \\ \mathbf{0}^\mathsf{T} & 0 & 0 \\ \Sigma_{\hat{D}''\widehat{N}'_r} & 0 & \sigma_{\hat{D}''}^2 \end{bmatrix} \tag{6.78}$$

with

$$\mathbb{D}\left(\left[\begin{array}{c}\underline{N}'_X\\ \underline{N}'_Y\end{array}\right]\right) = \mathrm{Diag}([\sigma^2_{\hat{\phi}}, \sigma^2_{\hat{\psi}}]) = \left[\begin{array}{c}\boldsymbol{q}_1^\mathsf{T}\\ \boldsymbol{q}_2^\mathsf{T}\end{array}\right] \Sigma_{\widehat{N}\widehat{N}} \, [\boldsymbol{q}_1, \boldsymbol{q}_2] \tag{6.79}$$

and

$$\mathbb{C}\mathrm{ov}\left(\left[\begin{array}{c}\underline{N}'_X\\ \underline{N}'_Y\end{array}\right], \hat{D}''\right) = \left[\begin{array}{c}\sigma_{\widehat{N}'_X\hat{D}''}\\ \sigma_{\widehat{N}'_Y\hat{D}''}\end{array}\right] = \left[\begin{array}{c}\boldsymbol{q}_1^\mathsf{T}\\ \boldsymbol{q}_2^\mathsf{T}\end{array}\right] \Sigma_{\widehat{N}\hat{D}} \, . \tag{6.80}$$

Next we choose $\boldsymbol{X}_0$ such that the uncertainty of a point across the plane is minimum. An arbitrary point $\mathcal{X}(\boldsymbol{X})$ has the distance

$$\underline{D}_X = \underline{\boldsymbol{N}}^\mathsf{T}\boldsymbol{X} - \underline{D} = [\boldsymbol{X}^\mathsf{T}, -1]\left[\begin{array}{c}\underline{\widehat{\boldsymbol{N}}}\\ -\underline{\hat{D}}\end{array}\right] . \tag{6.81}$$

Its variance is

$$\sigma^2_{D_X} = \boldsymbol{X}^\mathsf{T}\Sigma_{\widehat{N}\widehat{N}}\boldsymbol{X} - 2\Sigma_{\hat{D}\widehat{N}}\boldsymbol{X} + \sigma^2_{\hat{D}} \, . \tag{6.82}$$

From its derivative w.r.t. dX

$$\frac{\partial\sigma^2_{D_X}}{\partial\boldsymbol{X}} = 2\Sigma_{\widehat{N}\widehat{N}}\boldsymbol{X} - 2\Sigma_{\hat{D}\widehat{N}} \tag{6.83}$$

In the rotated system we have

$$\frac{\partial\sigma^2_{D''_X}}{\partial\boldsymbol{X}''} = 2\Sigma_{\widehat{N}'\widehat{N}'}\boldsymbol{X}'' - 2\Sigma_{\hat{D}''\widehat{N}''} = 2\left[\begin{array}{ccc}\sigma^2_{\hat{\phi}} & 0 & 0\\ 0 & \sigma^2_{\hat{\psi}} & 0\\ 0 & 0 & 0\end{array}\right]\left[\begin{array}{c}M''_X\\ M''_X\\ 0\end{array}\right] . \tag{6.84}$$

from which we obtain

$$\left[\begin{array}{c}M''_X\\ M''_Y\end{array}\right] = -\left[\begin{array}{cc}\sigma^2_{\hat{\phi}} & 0\\ 0 & \sigma^2_{\hat{\psi}}\end{array}\right]^{-1}\left[\begin{array}{c}\sigma_{\widehat{N}'_X\hat{D}''}\\ \sigma_{\widehat{N}'_Y\hat{D}''}\end{array}\right] = -\left[\begin{array}{cc}\sigma^2_{\hat{\phi}} & 0\\ 0 & \sigma^2_{\hat{\psi}}\end{array}\right]^{-1}\left[\begin{array}{c}\boldsymbol{q}_1^\mathsf{T}\\ \boldsymbol{q}_2^\mathsf{T}\end{array}\right]\Sigma_{\widehat{N}\hat{D}} \tag{6.85}$$

Finally, we have the centroid

$$\boldsymbol{X}_0 = \widehat{\boldsymbol{N}} - \boldsymbol{Q}\left[\begin{array}{ccc}\sigma^2_{\hat{\phi}} & 0 & 0\\ 0 & \sigma^2_{\hat{\psi}} & 0\\ 0 & 0 & 0\end{array}\right]^+ \boldsymbol{Q}^\mathsf{T}\Sigma_{\widehat{N}\hat{D}} \, . \tag{6.86}$$

or

$$\boxed{\boldsymbol{X}_0 = \widehat{\boldsymbol{N}} - \Sigma^+_{\widehat{N}\hat{D}}\Sigma_{\widehat{N}\hat{D}} \, .} \tag{6.87}$$

Remark: This is in full analogy to the centroid of the 2D line when using the covariance matrix of the normal

$$\Sigma_{\hat{n}\hat{n}} = \sigma^2_{\hat{\phi}}\boldsymbol{n}^\perp\boldsymbol{n}^{\perp\mathsf{T}} \quad \text{and} \quad \Sigma^+_{\hat{n}\hat{n}} = \sigma^{-2}_{\hat{\phi}}\boldsymbol{n}^\perp\boldsymbol{n}^{\perp\mathsf{T}} \quad \text{with} \quad \boldsymbol{n} = \left[\begin{array}{c}\sin\alpha\\ \cos\alpha\end{array}\right] \quad \text{and} \quad \boldsymbol{n}^\perp = \left[\begin{array}{c}\cos\alpha\\ -\sin\alpha\end{array}\right] \tag{6.88}$$

and the covariance with the distance

$$\Sigma_{\hat{n}\hat{d}} = \sigma_{\phi d}\boldsymbol{n}^\perp \tag{6.89}$$

since

$$\begin{aligned}
\boldsymbol{x}_0 &= \left[\begin{array}{cc}\cos\alpha & \sin\alpha\\ -\sin\alpha & \cos\alpha\end{array}\right]\left[\begin{array}{c}m_0\\ d\end{array}\right] && (6.90)\\
&= \left[\begin{array}{c}\sin\alpha\\ \cos\alpha\end{array}\right]d + \left[\begin{array}{c}\cos\alpha\\ -\sin\alpha\end{array}\right]m_0 && (6.91)\\
&= \boldsymbol{n} - \left[\begin{array}{c}\cos\alpha\\ -\sin\alpha\end{array}\right](-\sigma^{-2}_{\phi}\sigma_{\phi d}) && (6.92)\\
&= \boldsymbol{n} - \Sigma^+_{\hat{n}\hat{n}}\Sigma_{\hat{n}\hat{d}} \, . && (6.93)
\end{aligned}$$

$\diamond$

6.3.3 Checking a Set of Points for Planarity

6.3.3.1 Assuming the uncertainty of the points is known

We test whether the surface consisting of I points is planar, testing the null hypothesis

$$H_{01}: \quad \hat{\sigma}_0^2 = 1 \tag{6.94}$$

versus the alternative hypothesis

$$H_{a1}: \quad \hat{\sigma}_0^2 > 1 \tag{6.95}$$

using the chi-square test statistic

$$\underline{X}|H_{01} = \underline{\Omega}|H_{01} \sim \chi_R^2 \,. \tag{6.96}$$

which is χ_R^2-distributed under the null hypothesis. If the test is rejected, this may be caused

- by a too small standard deviation of the given points, or
- by a significant deviation of the surface from a plane, or
- both.

Remark: The degrees of freedom R should not be taken too large, since otherwise the null-hypothesis always will be rejected, see the discussion in Förstner and Wrobel (2016), around Eq. (4.88). ◇

6.3.3.2 Assuming an estimate of the uncertainty of the points of the plane is not known

We assume, the variance factor $\hat{\sigma}_{0a}^2$ of all given points may be taken from a robust estimate of all variance factors. Its degrees of freedom is assumed to be R_0.

We test the null hypothesis for the current plane

$$H_{02}: \quad \hat{\sigma}_0^2 = \hat{\sigma}_{0a}^2 \tag{6.97}$$

against the alternative hypothesis

$$H_{a2}: \quad \hat{\sigma}_0^2 > \hat{\sigma}_{0a}^2 \tag{6.98}$$

using the Fisher test statistic

$$\underline{F}|H_{02} = \frac{\underline{\hat{\sigma}}_0^2}{\underline{\hat{\sigma}}_{0a}^2}|H_{02} \sim F_{R,R_0} \tag{6.99}$$

which is F_{R,R_0}-distributed under the null hypothesis.

6.4 Estimating a Mean Plane

Given are I planes $\mathcal{A}_i$, the task is to find the best estimate for the mean plane $\mathbf{A}$.

We discuss three solutions:

1. A solution based on moments of the point cloud, assuming isotropic uncertainty.
2. A statistically suboptimal solution for the.
3. A statistically optimal solution based on the centroid representation.

6.4.1 Estimating the mean plane using moments

Let us assume we have J patches, represented by their moment vector $\boldsymbol{m}_j$.

Obviously, it is simple to derive the mean plane. We just need to add all non-central moments. Hence:

$$\boxed{\boldsymbol{m} = \sum_{j=1}^{J} \boldsymbol{m}_j \,.} \tag{6.100}$$

The parameters of the uncertain mean plane can then be derived from $\boldsymbol{c}(\boldsymbol{m})$, see (6.50). We need to observe:

1. Eq. (6.100) allows a recursive estimation of the plane. Let the mean plane derived from the first j patches be

   $$\boldsymbol{m}^{(j)} = \sum_{k=1}^{j} \boldsymbol{m}_k \,. \tag{6.101}$$

   Then adding the $(j+1)$-th patch leads to

   $$\boxed{\boldsymbol{m}^{(j+1)} = \boldsymbol{m}^{(j)} + \boldsymbol{m}_{j+1} \,.} \tag{6.102}$$

2. In a similar manner a patch k can be deleted if $k \in \{1, ..., j\}$:

   $$\boldsymbol{m}^{(j\backslash k)} = \boldsymbol{m}^{(j)} - \boldsymbol{m}_k \,. \tag{6.103}$$

3. Before inserting a patch into the list of patches, a statistical test could be performed. This can be based on the difference vector of the new patch $\mathcal{A}_{j+1}$ and the current mean plane $\mathcal{A}^{(j)}$

   $$\boldsymbol{d} = J_r^{\mathsf{T}}(\boldsymbol{\mu}_A)(\boldsymbol{A}_{j+1} - \boldsymbol{A}^{(j)}) \tag{6.104}$$

   and its covariance matrix

   $$\Sigma_{dd} = J_r^{\mathsf{T}}(\boldsymbol{\mu}_A) \left(\Sigma_{A_{j+1}A_{j+1}} + \Sigma_{A^{(j)}A^{(j)}}\right) J_r(\boldsymbol{\mu}_A) \tag{6.105}$$

   leading to the test statistic

   $$\underline{T} = \underline{\boldsymbol{d}}^{\mathsf{T}} \Sigma_{dd}^{-1} \underline{\boldsymbol{d}} \sim \chi^2_3 \,. \tag{6.106}$$

   Observe, the vector $\boldsymbol{d}$ in (6.104) is the difference $\boldsymbol{A}_{j+1,r} - \boldsymbol{A}_r^{(j)}$ of the reduced plane coordinates assuming the common tangent plane is given by $\boldsymbol{\mu}_A$. The argument $\boldsymbol{\mu}_A$ of $J(\boldsymbol{\mu}_A)$ best is chosen as the current mean plane $\boldsymbol{\mu}_A := \mathbf{A}^{(j)}$.

   Here we assume, the planes are Euclideanly normalized, see (6.17), i.e., the normal has length 1. Then the projection matrix $J_r(\mathbf{A})$ is given by

   $$\boxed{\underset{4\times 3}{J_r(\mathbf{A})} = \begin{bmatrix} \underset{3\times 2}{J_r(\boldsymbol{N})} & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & 1 \end{bmatrix} \quad \text{with} \quad J_r(\boldsymbol{N}) = \mathsf{null}(\boldsymbol{N}^{\mathsf{T}}) \,.} \tag{6.107}$$

4. All moments need to refer to the same coordinate system. Therefore, it might be useful to condition all coordinates before determining and fusing all patches.

5. There is no non-linearity involved in the recursive *estimation* involved, if we only consider the moments. The non-linearity only refers to deriving the centroid or other parameters of the planes. Especially no directions or angles are involved. A recursive determination of the variances would be difficult, without going back to the moments.

6. Eq. (6.102) can also be specialized to including a single point.

7. The whole procedure could once be repeated with modified weights. If the weights are reduced to 0, this is equivalent to deleting previously included patches, which can be done using (6.103).

Hence, the moments, the 11 parameters including the number of points, can be interpreted as the memory generating the current version of the plane. In statistical terms, the moments are sufficient test statistics, i.e., no other information is necessary to perform the estimation.

6.4.2 Approximate estimating the mean plane using plane parameters

Let us assume we have J patches, represented by their homogeneous vector $\mathbf{A}_j = [\boldsymbol{N}_j^\mathsf{T}, -D_j]^\mathsf{T}$ together with the covariance matrix of the reduced vector, namely $\left\{\left(A_j, \Sigma_{A_{jr}A_{jr}}\right)\right\}$.

The constraint, that the individual patch is identical to the mean plane is given by

$$\overline{\mathsf{\Pi}}(\mathbf{A}_j)\mathbf{A} = -\overline{\mathsf{\Pi}}(\mathbf{A})\mathbf{A}_j = \mathbf{0} \quad \text{with} \quad \overline{\mathsf{\Pi}}(\mathbf{A}) = \begin{bmatrix} 0 & -N_Z & N_Y & 0 \\ N_Z & 0 & -N_X & 0 \\ -N_Y & N_X & 0 & 0 \\ -D & 0 & 0 & -N_X \\ 0 & -D & 0 & -N_Y \\ 0 & 0 & -D & -N_Z \end{bmatrix}, \tag{6.108}$$

see Förstner and Wrobel (2016, Eq. (7.41)). Since a plane has only three degrees of freedom, we need to select three constraints from the six constraints in (6.108). If the coordinate system is chosen such that all distances D_j are non-zero, the last three constraints may be used. leading to

$$C_j^\mathsf{T}\overline{\mathsf{\Pi}}(\mathbf{A}_j)\mathbf{A} = -C_j^\mathsf{T}\overline{\mathsf{\Pi}}(\mathbf{A})\mathbf{A}_j = \mathbf{0} \quad \text{with} \quad C_j^\mathsf{T} = [\mathsf{0}_3 \mid \mathsf{I}_3]\,, \tag{6.109}$$

or

$$\boldsymbol{g}(\mathbf{A}_j, \mathbf{A}) = \mathsf{X}_j\mathbf{A} = \mathsf{Z}_j\mathbf{A}_j = \mathbf{0} \quad \text{with} \quad \mathsf{X}_j^\mathsf{T} = [-D_j\mathsf{I}_3 \mid -\boldsymbol{N}_j] \quad \text{and} \quad \mathsf{Z}_j = [D\mathsf{I}_3 \mid \boldsymbol{N}]\,. \tag{6.110}$$

Observe, the Jacobians of the constraint $\boldsymbol{g}(\boldsymbol{A}_j, \boldsymbol{A})$ w.r.t. the unknown parameters and the observations are X_j and Z_j, the last matrix is not depending on j.

For estimating the plane, we concatenate all $3J$ constraints in the following form

$$\boldsymbol{g}(\{\mathbf{A}_j\}, \mathbf{A}) = \underset{3J\times 4}{\mathsf{X}}\,\mathbf{A} = \mathbf{0}\,. \tag{6.111}$$

The right singular vector of $\mathbf{X}$ belonging to the smallest singular value is the algebraically optimal mean plane, and can be determined using the SVD of X:

$$\boxed{\widehat{\mathbf{A}} = \mathsf{V}_{:,4} \quad \text{with} \quad \mathsf{X} = \mathsf{U}\mathsf{S}\mathsf{V}^\mathsf{T}\,.} \tag{6.112}$$

For deriving the covariance matrix of this solution, we start with the differential of $\boldsymbol{g}$:

$$\mathrm{d}\boldsymbol{g}(\boldsymbol{y}, \mathbf{A}) = \underbrace{\mathsf{X}\mathsf{J}_r(\mathbf{A})}_{3J\times 3}\,\underset{3\times 1}{\mathrm{d}\boldsymbol{A}_r} + \underbrace{\mathsf{Z}}_{3J\times 3J}\,\underset{3J\times 1}{\mathrm{d}\boldsymbol{y}} \tag{6.113}$$

with

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{A}_{1r} \\ \ldots \\ \boldsymbol{A}_{jr} \\ \ldots \\ \boldsymbol{A}_{Jr} \end{bmatrix}, \quad \boldsymbol{A}_{jr} = \mathsf{J}_r^\mathsf{T}(\mathbf{A})\,\mathbf{A}_j \quad \text{and} \quad \mathsf{Z} = \mathrm{Diag}(\{\underset{4\times 3}{\mathsf{Z}_j}\,\underset{3\times 4}{\mathsf{J}_r(\mathbf{A})}\})\,. \tag{6.114}$$

With the reduced coefficient matrix

$$\mathsf{X}_r = \mathsf{X}\mathsf{J}_r(\mathbf{A}) \tag{6.115}$$

we thus obtain the differential estimates

$$\mathrm{d}\widehat{\boldsymbol{A}}_r = -\mathsf{X}_r^+\,\mathsf{Z}\mathrm{d}\boldsymbol{y} = -(\mathsf{X}_r^\mathsf{T}\mathsf{X}_r)^{-1}\mathsf{X}_r^\mathsf{T}\mathsf{Z}\,\mathrm{d}\boldsymbol{y}\,. \tag{6.116}$$

Hence, we have the covariance matrix of the estimated reduced plane parameters

$$\boxed{\Sigma_{\hat{A}_r\hat{A}_r} = \mathsf{X}_r^+\,\mathsf{Z}\,\Sigma_{yy}\,\mathsf{Z}^\mathsf{T}\,\mathsf{X}_r^{+\mathsf{T}}\,.} \tag{6.117}$$

where

$$\Sigma_{yy} = \text{Diag}\left(\{\Sigma_{y_j y_j}\}\right) = \text{Diag}\left(\left\{J_r^{\mathsf{T}}(\mathbf{A})\Sigma_{A_j A_j} J_r(\mathbf{A})\right\}\right) . \tag{6.118}$$

Observe, the solution is suboptimal, since the pseudo inverse X_r^+ is taken instead of the weighted pseudo inverse $(X_{r,\mathsf{W}_{ll}})^+ = (X_r^{\mathsf{T}} W_{yy} X_r)^{-1} X_r^{\mathsf{T}} W_{yy}$. Finally, we obtain the covariance matrix of the estimated mean plane

$$\Sigma_{\hat{A}\hat{A}} = J_r(\mathbf{A})\, \Sigma_{\hat{A}_r\hat{A}_r}\, J_r^{\mathsf{T}}(\mathbf{A}) , \tag{6.119}$$

which has rank 3.

6.4.3 An optimal solution based on the centroid representation

We assume we have given the planes in centroid form,

$$\mathcal{A}_i : \quad \{\boldsymbol{X}_{0i}, Q_i; \sigma_{q_i}, \sigma_{\phi_i}, \sigma_{\psi_i}\} , \tag{6.120}$$

and want to determine the mean plane, also in centroid form

$$\mathcal{A} : \quad \{\boldsymbol{X}_0, Q; \sigma_q, \sigma_\phi, \sigma_\psi\} . \tag{6.121}$$

We use the following nonlinear constraints:

$$\widehat{\boldsymbol{N}} \times \widehat{\boldsymbol{N}}_i = \mathbf{0} , \tag{6.122}$$

which represents two degrees of freedom. We select two independent constraints:

$$M_i^{(s)} S(\widehat{\boldsymbol{N}}) \widehat{\boldsymbol{N}}_i = \mathbf{0} \tag{6.123}$$

and the translational constraint

$$\widehat{\boldsymbol{N}}^{\mathsf{T}} (\widehat{\boldsymbol{X}}_{0i} - \widehat{\boldsymbol{X}}_0) = 0 , \tag{6.124}$$

which represents the third degree of freedom. For proofs we will use

$$M_i^{(s)} = M^{(s)}(\boldsymbol{N}_i) = M^{(s)}(\boldsymbol{q}_{i3}) = \begin{bmatrix} \boldsymbol{q}_{i1}^{\mathsf{T}} \\ \boldsymbol{q}_{i2}^{\mathsf{T}} \end{bmatrix} \quad \text{with} \quad M_i^{(s)} M_i^{(s)\mathsf{T}} = I_2 , \tag{6.125}$$

hence

$$M_i^{(s)} S(\widehat{\boldsymbol{N}}_i) = \begin{bmatrix} \boldsymbol{q}_{i1}^{\mathsf{T}} \\ \boldsymbol{q}_{i2}^{\mathsf{T}} \end{bmatrix} S(\boldsymbol{q}_3) = \begin{bmatrix} \boldsymbol{q}_2^{\mathsf{T}} \\ -\boldsymbol{q}_1^{\mathsf{T}} \end{bmatrix} \tag{6.126}$$

Hence the nonlinear constraints are

$$\boldsymbol{g}_i(\widehat{\boldsymbol{N}}_i, \widehat{\boldsymbol{X}}_{0i}; \widehat{\boldsymbol{N}}, \widehat{\boldsymbol{X}}_0) = \begin{bmatrix} \widehat{\boldsymbol{N}}^{\mathsf{T}} (\widehat{\boldsymbol{X}}_{0i} - \widehat{\boldsymbol{X}}_0) \\ M_i^{(s)} S(\widehat{\boldsymbol{N}}) \widehat{\boldsymbol{N}}_i \end{bmatrix} = \mathbf{0} , \quad i = 1, \ldots, I . \tag{6.127}$$

6.4.3.1 The Iterative Solution

We also can assume approximate values, thus can update them using an iterative scheme. Linearization of the constraints yields

$$\begin{aligned} \boldsymbol{g}_i(\widehat{\boldsymbol{N}}_i, \widehat{\boldsymbol{X}}_{0i}; \widehat{\boldsymbol{N}}, \widehat{\boldsymbol{X}}_0) &= \boldsymbol{g}_i(\widehat{\boldsymbol{N}}_i^a, \widehat{\boldsymbol{X}}_{0i}^a; \widehat{\boldsymbol{N}}^a, \widehat{\boldsymbol{X}}_0^a) && (6.128) \\ &+ \begin{bmatrix} (\widehat{\boldsymbol{X}}_{0i} - \widehat{\boldsymbol{X}}_0)^{\mathsf{T}} \widehat{\Delta \boldsymbol{N}} + \widehat{\boldsymbol{N}}^{\mathsf{T}} \widehat{\Delta \boldsymbol{X}_{0i}} - \widehat{\boldsymbol{N}}^{\mathsf{T}} \widehat{\Delta \boldsymbol{X}_0} \\ -M^{(s)} S(\widehat{\boldsymbol{N}}_i) \widehat{\Delta \boldsymbol{N}} + M^{(s)} S(\widehat{\boldsymbol{N}}) \widehat{\Delta \boldsymbol{N}_i} \end{bmatrix} && (6.129) \end{aligned}$$

This can be written as

$$\boldsymbol{g}_i(\widehat{\boldsymbol{N}}_i, \widehat{\boldsymbol{X}}_{0i}; \widehat{\boldsymbol{N}}, \widehat{\boldsymbol{X}}_0) = \boldsymbol{g}_i(\widehat{\boldsymbol{N}}_i^a, \widehat{\boldsymbol{X}}_{0i}^a; \widehat{\boldsymbol{N}}^a, \widehat{\boldsymbol{X}}_0^a) + X_i \widehat{\Delta \boldsymbol{\theta}} + Z_i^{\mathsf{T}} \widehat{\Delta \boldsymbol{y}_i} \tag{6.130}$$

with

$$\widehat{\Delta\boldsymbol{\theta}} = J_r^{\mathsf{T}} \begin{bmatrix} \Delta\boldsymbol{X}_0 \\ \Delta\boldsymbol{N} \end{bmatrix} = \widehat{\Delta\boldsymbol{A}}^{\circ} \quad \text{and} \quad \widehat{\Delta\boldsymbol{y}}_i = J_r^{\mathsf{T}} \begin{bmatrix} \Delta\boldsymbol{X}_{0i} \\ \Delta\boldsymbol{N}_i \end{bmatrix} = \widehat{\Delta\boldsymbol{A}}_i^{\circ} \tag{6.131}$$

and, since $\widehat{\boldsymbol{N}}_i \equiv \widehat{\boldsymbol{N}}$,

$$X_i = \begin{bmatrix} -\widehat{\boldsymbol{N}}^{\mathsf{T}} & (\widehat{\boldsymbol{X}}_{0i} - \widehat{\boldsymbol{X}}_0)^{\mathsf{T}} \\ 0_{2\times 3} & -M^{(s)} S(\widehat{\boldsymbol{N}}) \end{bmatrix} J_r \quad \text{and} \quad Z_i^{\mathsf{T}} = \begin{bmatrix} \widehat{\boldsymbol{N}}^{\mathsf{T}} & \mathbf{0}^{\mathsf{T}} \\ 0_{2\times 3} & M^{(s)} S(\widehat{\boldsymbol{N}}) \end{bmatrix} J_r = \begin{bmatrix} 1 & \mathbf{0}^{\mathsf{T}} \\ \mathbf{0} & I_2 \end{bmatrix} = I_3 \,. \tag{6.132}$$

The weight matrix of the residuals therefore is

$$W_{c_i c_i} \quad = \quad (B_i^{\mathsf{T}} \Sigma_{y_i y_i} B_i)^{-1} = W_{A_i^{\circ} A_i^{\circ}} \,. \tag{6.133}$$

The normal equation matrix thus is

$$N = \sum_i A_i^{\mathsf{T}} W_{c_i c_i} A_i \tag{6.134}$$

$$= \sum_i J_r^{\mathsf{T}} \begin{bmatrix} -\widehat{\boldsymbol{N}} & 0_{3\times 2} \\ (\widehat{\boldsymbol{X}}_{0i} - \widehat{\boldsymbol{X}}_0) & S(\widehat{\boldsymbol{N}}) M^{(s)\mathsf{T}} \end{bmatrix} W_{A_i^{\circ} A_i^{\circ}} \begin{bmatrix} -\widehat{\boldsymbol{N}}^{\mathsf{T}} & (\widehat{\boldsymbol{X}}_{0i} - \widehat{\boldsymbol{X}}_0)^{\mathsf{T}} \\ 0_{2\times 3} & -M^{(s)} S(\widehat{\boldsymbol{N}}) \end{bmatrix} J_r \tag{6.135}$$

$$= \sum_i \begin{bmatrix} \boldsymbol{q}_3^{\mathsf{T}} & \mathbf{0}^{\mathsf{T}} \\ \mathbf{0}^{\mathsf{T}} & \boldsymbol{q}_1^{\mathsf{T}} \\ \mathbf{0}^{\mathsf{T}} & \boldsymbol{q}_2^{\mathsf{T}} \end{bmatrix} \begin{bmatrix} -\widehat{\boldsymbol{N}} & 0_{3\times 2} \\ (\widehat{\boldsymbol{X}}_{0i} - \widehat{\boldsymbol{X}}_0) & S(\widehat{\boldsymbol{N}}) M^{(s)\mathsf{T}} \end{bmatrix} W_{A_i^{\circ} A_i^{\circ}} \begin{bmatrix} -\widehat{\boldsymbol{N}}^{\mathsf{T}} & (\widehat{\boldsymbol{X}}_{0i} - \widehat{\boldsymbol{X}}_0)^{\mathsf{T}} \\ 0_{2\times 3} & -M^{(s)} S(\widehat{\boldsymbol{N}}) \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{q}_1 & \boldsymbol{q}_2 \end{bmatrix}$$

$$= \sum_i \begin{bmatrix} -1 & 0 & 0 \\ X''_{0i} - X''_0 & 1 & 0 \\ Y''_{0i} - Y''_0 & 0 & 1 \end{bmatrix} \begin{bmatrix} w_{q_i} & & \\ & w_{\phi_i} & \\ & & w_{\psi_i} \end{bmatrix} \begin{bmatrix} -1 & X''_{0i} - X''_0 & Y''_{0i} - Y''_0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{6.136}$$

$$= \sum_i \begin{bmatrix} -1 & 0 & 0 \\ X''_{0i} - X''_0 & 1 & 0 \\ Y''_{0i} - Y''_0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -w_{q_i} & w_{q_i}(X''_{0i} - X''_0) & w_{q_i}(Y''_{0i} - Y''_0) \\ & w_{\phi_i} & \\ & & w_{\psi_i} \end{bmatrix} \tag{6.137}$$

$$= \sum_i \begin{bmatrix} w_{q_i} & -w_{q_i}(X''_{0i} - X'_0) & -w_{q_i}(Y''_{0i} - Y''_0) \\ -w_{q_i}(X''_{0i} - X''_0) & w_{\phi_i} + w_{q_i}(X''_{0i} - X''_0)^2 & w_{q_i}(X''_{0i} - X'_0)(Y''_{0i} - Y''_0) \\ -w_{q_i}(Y''_{0i} - Y''_0) & w_{q_i}(X''_{0i} - X''_0)(Y''_{0i} - Y''_0) & w_{\psi_i} + w_{q_i}(Y''_{0i} - Y''_0)^2 \end{bmatrix} \tag{6.138}$$

The normal equation matrix is diagonal, if[1]

$$\sum_i w_{q_i}(X''_{0i} - X''_0) = \sum_i w_{q_i}(Y''_{0i} - Y''_0) = 0 \,. \tag{6.139}$$

Then we obtain

$$N = \sum_i \begin{bmatrix} w_{q_i} & 0 & 0 \\ 0 & w_{\phi_i} + w_{q_i} X'^2_{0i} & \\ 0 & & w_{\psi_i} + w_{q_i} Y'^2_{0i} \end{bmatrix} . \tag{6.140}$$

Hence, all entities have to be taken at their estimates. If we use the centroid of the fitted centroids

$$\boldsymbol{X}''_0 = \frac{\sum_i w_{q_i} \widehat{\boldsymbol{X}}''_{0i}}{\sum_i w_{q_i}} \tag{6.141}$$

and the individual centroids reduced to the common centroid

$$\boldsymbol{X}'_{0i} = \boldsymbol{X}''_{0i} - \boldsymbol{X}''_0 \,. \tag{6.142}$$

The right hand side of the normal equation system is

$$\boldsymbol{n} \quad = \quad \sum_i A_i^{\mathsf{T}} W_{c_i c_i} (-\boldsymbol{g}_i(\hat{\boldsymbol{\theta}}^a, \hat{\boldsymbol{y}}^a) + B_i(\hat{\boldsymbol{y}}_i - \boldsymbol{y}_i) \,. \tag{6.143}$$

[1] The original note said $\sum_i w_{\phi_i}(X''_{0i} - X''_0) = \sum_i w_{\psi_i}(Y''_{0i} - Y''_0) = 0$. But due to (6.141), this appears to be incorrect.

6.4.3.2 The Theoretical Precision

We obtain the variances for the three entities

$$\sigma^2_{\hat{q}} = \frac{1}{I}\,\frac{1}{\overline{w_q}} \tag{6.144}$$

$$\sigma^2_{\hat{\phi}} = \frac{1}{I}\,\frac{1}{\overline{w_\phi} + \overline{w_q}\,\overline{X^2_{0i}}} \tag{6.145}$$

$$\sigma^2_{\hat{\psi}} = \frac{1}{I}\,\frac{1}{\overline{w_\psi} + \overline{w_q}\,\overline{Y^2_{0i}}}\,. \tag{6.146}$$

If the I planes would have the same precision we would obtain

$$\sigma^2_{\hat{q}} = \frac{1}{I}\,\sigma^2_q\,, \quad \sigma^2_{\hat{\phi}} = \frac{1}{I}\,\frac{\sigma^2_q\sigma^2_\phi}{\sigma^2_q + \sigma^2_\phi\overline{X^2_{0i}}}\,, \quad \sigma^2_{\hat{\psi}} = \frac{1}{I}\,\frac{\sigma^2_q\sigma^2_\psi}{\sigma^2_q + \sigma^2_\psi\overline{Y^2_{0i}}}\,. \tag{6.147}$$

This is a plausible result: The precision of the normal of the average plane increases with the number I of the planes and with increasing scatter of the individual planes. Observe, if the standard deviation σ_q is 0, then the directions will also have standard deviation 0.

6.5 Motion from Plane to Plane correspondences

6.5.1 Problem Statement

Given are I correspondences $\{\mathcal{A}_i, \mathcal{A}'_j\}$ which are related by

$$\mathcal{M} : \mathcal{A}'_j \mapsto \mathcal{A}_i \qquad \mathcal{A}_i \equiv \mathcal{M}(\mathcal{A}'_j) \quad \text{for all } (ij) \in \mathcal{C}\,. \tag{6.148}$$

There are two options to establish the correspondences:

1. The planes $(\mathcal{A}_i, \mathcal{A}'_j)$ refer to the planar patches derived from some segmentation of two point clouds. Then each of the planes $\mathcal{A}_i$ or $\mathcal{A}'_j$ may have several correspondences, namely if there are coplanar planes one or both of the point clouds. The Jacobian B of the Gauss–Helmert model is block diagonal, each block B_k referring to the correspondence of coplanar planes $\{i_k\}$ and $\{j'_k\}$ in the two point clouds.
2. The planes $(\mathcal{A}_i, \mathcal{A}'_j)$ refer to aggregated coplanar planes in each point cloud. Then there is a one-to-one correspondence, and we may refer to the same index, thus refer to $(\mathcal{A}_k, \mathcal{A}'_k)$. In this case the partitioning of the point cloud has a final merge-step to find sets of coplanar points and to determine the average (ML-estimates) plane parameters.

We do not distinguish the two cases until we discuss the solution of the nonlinear Gauss–Helmert model.

We explicitly have

$$\mathcal{A}_i : \quad \left\{\boldsymbol{X}_0, \mathsf{Q}; \sigma^2_q, \sigma^2_\phi, \sigma^2_\psi\right\}_i \tag{6.149}$$

The constraint implies an unknown motion $\mathcal{M}$

$$\mathcal{M} : \quad \{\boldsymbol{T}, \mathsf{R}\} \tag{6.150}$$

which transforms the 3D points $\mathcal{X}_i$ into the coordinate system

$$\boldsymbol{X} = \mathsf{R}\boldsymbol{X}' + \boldsymbol{T}\,. \tag{6.151}$$

The corresponding transformation of the plane parameters is

$$\boldsymbol{X}_{0i} = \mathsf{R}\boldsymbol{X}'_{0j} + \boldsymbol{T} \tag{6.152}$$

and

$$Q_i = RQ'_j\,. \tag{6.153}$$

We need three constraints for the identity of two planes. These can be the following rotational constraint

$$\boxed{\boldsymbol{N}_i = R\boldsymbol{N}'_j\,,} \tag{6.154}$$

which represents two degrees of freedom, and the translational constraint

$$\boxed{\boldsymbol{N}_i^{\mathsf{T}}(R\boldsymbol{X}'_{0j} + \boldsymbol{T} - \boldsymbol{X}_{0i}) = 0\,,} \tag{6.155}$$

which represents the third degree of freedom.

From a counting argument we would need only two planes. However, then the translation along the intersecting 3D line is not determined. Therefore, we need at least three planes in general position for being able to determine the motion.

6.5.2 Minimal Solution for the Motion from Three Plane Correspondences

The three planes need to intersect in a 3D point $\mathcal{Y}$ not at infinity. Otherwise the translation in this direction is not determined.

Then the translation can be determined from the two intersection points $\boldsymbol{Y}$ and $\boldsymbol{Y}'$, and the rotation from the three normals.

If enough plane-plane correspondences are available the rotation may be derived from (6.154) in the form

$$N = RN' \tag{6.156}$$

Hence we have

$$H = N'^{\mathsf{T}}N = U\boldsymbol{\Lambda}V^{\mathsf{T}} \tag{6.157}$$

and thus

$$R = UV^{\mathsf{T}}\,. \tag{6.158}$$

Using this rotation the translation then can be determined from (6.155) in the form

$$\boldsymbol{N}_i^{\mathsf{T}}(R\boldsymbol{X}'_{0j} - \boldsymbol{X}_{0i}) = -\boldsymbol{N}_j^{\mathsf{T}}\boldsymbol{T} \tag{6.159}$$

which leads to the linear equation system

$$B^{\mathsf{T}}B\boldsymbol{T} = B^{\mathsf{T}}\boldsymbol{b} \tag{6.160}$$

with

$$B = -N = -[\boldsymbol{N}_i^{\mathsf{T}}] \quad \text{and} \quad \boldsymbol{b} = [\boldsymbol{N}_i^{\mathsf{T}}(R\boldsymbol{X}'_{0j} - \boldsymbol{X}_{0i})]\,. \tag{6.161}$$

Weighting is possible.

6.5.3 An Iterative Solution

We use the three constraints for each correspondence

$$\underset{3\times 1}{\boldsymbol{g}_{ij}}(\hat{\boldsymbol{T}}, \hat{R}(\hat{\boldsymbol{\theta}}); \widehat{\boldsymbol{X}}_{0i}, \widehat{\boldsymbol{N}}_i, \widehat{\boldsymbol{X}}'_{0j}, \widehat{\boldsymbol{N}}'_j) = \begin{bmatrix} \widehat{\boldsymbol{N}}_i^{\mathsf{T}}(\hat{R}\widehat{\boldsymbol{X}}'_{0j} + \hat{\boldsymbol{T}} - \widehat{\boldsymbol{X}}_{0i}) \\ M_i^{(s)}S(\widehat{\boldsymbol{N}}_i)\hat{R}\widehat{\boldsymbol{N}}'_j \end{bmatrix} = \boldsymbol{0}\,. \tag{6.162}$$

Where $M_i^{(s)}S(\widehat{\boldsymbol{N}}_i) \in \mathsf{null}^{\mathsf{T}}(\boldsymbol{N}_i^{\mathsf{T}})$ is a orthonormal 2×3 matrix which is achieved by selecting two independent rows of the skew symmetric matrix $S(\widehat{\boldsymbol{N}}_i)$.

The linearized model reads as

$$\boldsymbol{g}_{ij}(\hat{\boldsymbol{T}}, \hat{R}(\hat{\boldsymbol{\theta}}); \widehat{\boldsymbol{X}}_{0i}, \widehat{\boldsymbol{N}}_i, \widehat{\boldsymbol{X}}'_{0j}, \widehat{\boldsymbol{N}}'_j) = \boldsymbol{g}_{ij}(\hat{\boldsymbol{\theta}}^a, \hat{\boldsymbol{y}}^a) \tag{6.163}$$

$$+ \begin{bmatrix} (\hat{R}\widehat{\boldsymbol{X}}'_{0j} + \hat{\boldsymbol{T}} - \widehat{\boldsymbol{X}}_{0i})^{\mathsf{T}}\widehat{\Delta\boldsymbol{N}}_i - \widehat{\boldsymbol{N}}_i^{\mathsf{T}}S(\hat{R}\widehat{\boldsymbol{X}}'_{0j})\widehat{\Delta\boldsymbol{\theta}} + \widehat{\boldsymbol{N}}_i^{\mathsf{T}}\hat{R}\widehat{\Delta\boldsymbol{X}}'_{0j} + \widehat{\boldsymbol{N}}_i^{\mathsf{T}}\widehat{\Delta\boldsymbol{T}} - \widehat{\boldsymbol{N}}_i^{\mathsf{T}}\widehat{\Delta\boldsymbol{X}}_{0i} \\ -M_i^{(s)}S(\hat{R}\widehat{\boldsymbol{N}}'_j)\widehat{\Delta\boldsymbol{N}}_i - M_i^{(s)}S(\widehat{\boldsymbol{N}}_i)S(\hat{R}\widehat{\boldsymbol{N}}'_j)\widehat{\Delta\boldsymbol{\theta}} + M_i^{(s)}S(\widehat{\boldsymbol{N}}_i)\hat{R}\widehat{\Delta\boldsymbol{N}}'_j \end{bmatrix}^a$$

Hence we have

$$\boldsymbol{g}_{ij} = \boldsymbol{g}_{ij}(\hat{\boldsymbol{\theta}}^a, \hat{\boldsymbol{y}}^a) + \mathsf{X}_{ij}\widehat{\Delta\boldsymbol{\theta}} + \mathsf{Z}_{ij}^{\mathsf{T}}\widehat{\Delta\boldsymbol{y}} = \mathbf{0} \tag{6.164}$$

with

$$\underset{6\times 1}{\Delta\boldsymbol{\theta}} := \begin{bmatrix} \Delta\boldsymbol{T} \\ \Delta\boldsymbol{\theta} \end{bmatrix} \quad \text{and} \quad \underset{6\times 1}{\Delta\boldsymbol{y}} := \begin{bmatrix} \Delta\boldsymbol{A}_i^\circ \\ \Delta\boldsymbol{A}_j^\circ \end{bmatrix} \tag{6.165}$$

The Jacobians are

$$\underset{3\times 6}{\mathsf{X}_{ij}^{\mathsf{T}}} = \frac{\partial \boldsymbol{g}_{ij}}{\partial\hat{\boldsymbol{\theta}}} = \begin{bmatrix} -\widehat{\boldsymbol{N}}_i^{\mathsf{T}} \mathsf{S}(\hat{\mathsf{R}}\widehat{\boldsymbol{X}}'_{0j}) & \widehat{\boldsymbol{N}}_i^{\mathsf{T}} \\ -\mathsf{M}_i^{(s)} \mathsf{S}(\widehat{\boldsymbol{N}}_i)\mathsf{S}(\hat{\mathsf{R}}\widehat{\boldsymbol{N}}'_j) & \mathsf{0} \end{bmatrix}^a \tag{6.166}$$

and

$$\underset{3\times 6}{\mathsf{Z}_{ij}^{\mathsf{T}}} = \frac{\partial \boldsymbol{g}_{ij}}{\partial[\boldsymbol{A}_i^{\mathsf{T}}, \boldsymbol{A}_j'^{\mathsf{T}}]^{\mathsf{T}}} \tag{6.167}$$

$$= \frac{\partial \boldsymbol{g}_{ij}}{\partial[\boldsymbol{A}_i^{*\mathsf{T}}, \boldsymbol{A}_j'^{*\mathsf{T}}]^{\mathsf{T}}} \frac{\partial[\boldsymbol{A}_i^{*\mathsf{T}}, \boldsymbol{A}_j'^{*\mathsf{T}}]^{\mathsf{T}}}{\partial[\boldsymbol{A}_i^{\mathsf{T}}, \boldsymbol{A}_j'^{\mathsf{T}}]^{\mathsf{T}}} \tag{6.168}$$

$$= \underset{3\times 12}{\begin{bmatrix} -\widehat{\boldsymbol{N}}_i^{\mathsf{T}} & (\hat{\mathsf{R}}\widehat{\boldsymbol{X}}'_{0j} + \hat{\boldsymbol{T}} - \widehat{\boldsymbol{X}}_{0i})^{\mathsf{T}} & \widehat{\boldsymbol{N}}_i^{\mathsf{T}}\hat{\mathsf{R}} & \mathbf{0}^{\mathsf{T}} \\ \mathsf{0}_{2\times 3} & -\mathsf{M}_i^{(s)}\mathsf{S}(\hat{\mathsf{R}}\widehat{\boldsymbol{N}}'_j) & \mathsf{0}_{2\times 3} & \mathsf{M}_i^{(s)}\mathsf{S}(\widehat{\boldsymbol{N}}_i)\hat{\mathsf{R}} \end{bmatrix}}^a \begin{bmatrix} \mathsf{J}_r(\mathsf{Q}_i) & \\ & \mathsf{J}_r(\mathsf{Q}'_j) \end{bmatrix}^a \tag{6.169}$$

If each plane only is present in one constraint, hence we have $i = j$, the normal equations for the six unknown parameters read as

$$\mathsf{N}\widehat{\Delta\boldsymbol{\theta}} = \boldsymbol{n} \tag{6.169}$$

with

$$\underset{6\times 6}{\mathsf{N}} = \sum_i \underset{6\times 3}{\mathsf{A}_i} \underset{3\times 6}{(\mathsf{B}_i^{\mathsf{T}}} \underbrace{\mathrm{Diag}(\{\Sigma_{A_i^\circ A_i^\circ}, \Sigma_{A_i'^\circ A_i'^\circ}\})}_{6\times 6} \underset{6\times 3}{\mathsf{B}_i)^{-1}} \underset{3\times 6}{\mathsf{A}_i^{\mathsf{T}}} \tag{6.170}$$

$$\underset{6\times 1}{\boldsymbol{n}} = \sum_i \underset{6\times 3}{\mathsf{A}_i} \underset{3\times 6}{(\mathsf{B}_i^{\mathsf{T}}} \underbrace{\mathrm{Diag}(\{\Sigma_{A_i^\circ A_i^\circ}, \Sigma_{A_i'^\circ A_i'^\circ}\})}_{6\times 6} \underset{6\times 3}{\mathsf{B}_i)^{-1}} \underbrace{(-\boldsymbol{g}_i(\hat{\boldsymbol{\theta}}^a, \hat{\boldsymbol{y}}_i^a)) + \mathsf{B}_i(\hat{\boldsymbol{y}}_i^a - \boldsymbol{y}_i^a)}_{3\times 1} \tag{6.171}$$

The update of the translation and the rotation then is

$$\begin{bmatrix} \hat{\boldsymbol{T}}^{(\nu+1)} \\ \hat{\mathsf{R}}^{(\nu+1)} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{T}}^{(\nu)} + \widehat{\Delta\boldsymbol{T}} \\ \mathsf{R}(\widehat{\Delta\boldsymbol{\theta}})\hat{\mathsf{R}}^{(\nu)} \end{bmatrix}. \tag{6.172}$$

6.5.4 Theoretical Accuracy of the Motion

We assume the rotation and translation is an identity. We also assume the corresponding planes to have the same mean parameters and the same covariance matrix. This simplifies the expressions and allows us to derive the covariance matrix as a function of the planes.

We use the relations

$$\mathsf{D}_r = = \boldsymbol{r}\boldsymbol{r}^{\mathsf{T}} \tag{6.173}$$

$$\mathsf{S}^2(\boldsymbol{r}) = -(\mathsf{I}_3 - \mathsf{D}_r) \tag{6.174}$$

$$\mathsf{S}(\boldsymbol{r})\mathsf{R} = \mathsf{R}\mathsf{S}(\mathsf{R}^{\mathsf{T}}\boldsymbol{r}) \tag{6.175}$$

$$\boldsymbol{N} = \mathsf{Q}\boldsymbol{e}_3 \quad \text{or} \quad \boldsymbol{e}_3 = \mathsf{Q}^{\mathsf{T}}\boldsymbol{N}\,. \tag{6.176}$$

The Jacobians are (omitting the hats and assuming we always refer to the fitted values)

$$\mathsf{X}_i^{\mathsf{T}} = \frac{\partial \boldsymbol{g}_i}{\partial\hat{\boldsymbol{\theta}}} = \begin{bmatrix} -\widehat{\boldsymbol{N}}_i^{\mathsf{T}}\mathsf{S}(\boldsymbol{X}'_{0j}) & \boldsymbol{N}_i^{\mathsf{T}} \\ -\mathsf{M}_i^{(s)}\mathsf{S}(\boldsymbol{N}_i)\mathsf{S}(\boldsymbol{N}'_i) & \mathsf{0} \end{bmatrix} = \begin{bmatrix} (\boldsymbol{X}_{0i} \times \boldsymbol{N}_i)^{\mathsf{T}} & \boldsymbol{N}_i^{\mathsf{T}} \\ \mathsf{M}_i^{(s)}(\mathsf{I}_3 - \mathsf{D}_{N_i}) & \mathsf{0} \end{bmatrix} \tag{6.177}$$

and

$$Z_i^\mathsf{T} = \frac{\partial \boldsymbol{g}_{ij}}{\partial [\boldsymbol{A}_i^\mathsf{T}, \boldsymbol{A}_j'^\mathsf{T}]^\mathsf{T}} \tag{6.178}$$

$$= \begin{bmatrix} -\boldsymbol{N}_i^\mathsf{T} & \boldsymbol{0}^\mathsf{T} & \boldsymbol{N}_i^\mathsf{T} & \boldsymbol{0}^\mathsf{T} \\ \boldsymbol{0}_{2\times 3} & -M_i^{(s)} S(\boldsymbol{N}_i) & \boldsymbol{0}_{2\times 3} & M_i^{(s)} S(\boldsymbol{N}_i) \end{bmatrix} \begin{bmatrix} \begin{bmatrix} \boldsymbol{q}_3 & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{q}_{i1} & \boldsymbol{q}_{i2} \end{bmatrix} & \boldsymbol{0} \\ \boldsymbol{0} & \begin{bmatrix} \boldsymbol{q}_3 & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{q}_{i1} & \boldsymbol{q}_{i2} \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} -\boldsymbol{e}_1^\mathsf{T} & \boldsymbol{e}_1^\mathsf{T} \\ -M_i^{(s)}[\boldsymbol{0}, \boldsymbol{q}_{i2}, -\boldsymbol{q}_{i1}] & M_i^{(s)}[\boldsymbol{0}, \boldsymbol{q}_{i2}, -\boldsymbol{q}_{i1}] \end{bmatrix} \tag{6.179}$$

We now assume the covariance matrices of all planes to be identical and isotropic

$$\Sigma_{A_i^\circ A_i^\circ} = \Sigma_{A^\circ A^\circ} = \begin{bmatrix} \sigma_q^2 & & \\ & \sigma_\phi^2 & \\ & & \sigma_\phi^2 \end{bmatrix} \tag{6.180}$$

Remark: Better do not do this! ◇

Then we have

$$B_i^\mathsf{T} \Sigma_{A_i^\circ A_i^\circ} B_i = \begin{bmatrix} -\boldsymbol{e}_1^\mathsf{T} & \boldsymbol{e}_1^\mathsf{T} \\ -M_i^{(s)}[\boldsymbol{0}, \boldsymbol{q}_{i2}, -\boldsymbol{q}_{i1}] & M_i^{(s)}[\boldsymbol{0}, \boldsymbol{q}_{i2}, -\boldsymbol{q}_{i1}] \end{bmatrix} \begin{bmatrix} \Sigma_{A^\circ A^\circ} & \\ & \Sigma_{A^\circ A^\circ} \end{bmatrix} \begin{bmatrix} -\boldsymbol{e}_1^\mathsf{T} & \\ -M_i^{(s)}[\boldsymbol{0}, \boldsymbol{q}_{i2}, -\boldsymbol{q}_{i1}] & M \end{bmatrix}$$

$$= \begin{bmatrix} -\boldsymbol{e}_1^\mathsf{T} & \boldsymbol{e}_1^\mathsf{T} \\ -M_i^{(s)}[\boldsymbol{0}, \boldsymbol{q}_{i2}, -\boldsymbol{q}_{i1}] & M_i^{(s)}[\boldsymbol{0}, \boldsymbol{q}_{i2}, -\boldsymbol{q}_{i1}] \end{bmatrix} \begin{bmatrix} -\Sigma_{A^\circ A^\circ}\boldsymbol{e}_3 & -\Sigma_{A^\circ A^\circ} \begin{bmatrix} \boldsymbol{0}^\mathsf{T} \\ \boldsymbol{q}_{i2}^\mathsf{T} \\ -\boldsymbol{q}_{i1}^\mathsf{T} \end{bmatrix} M_i^{(s)\mathsf{T}} \\ \Sigma_{A^\circ A^\circ}\boldsymbol{e}_3 & \Sigma_{A^\circ A^\circ} \begin{bmatrix} \boldsymbol{0}^\mathsf{T} \\ \boldsymbol{q}_{i2}^\mathsf{T} \\ -\boldsymbol{q}_{i1}^\mathsf{T} \end{bmatrix} M_i^{(s)\mathsf{T}} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_q^2 & \boldsymbol{0} \\ \boldsymbol{0} & 2\sigma_\phi^2 \, M_i^{(s)} (\boldsymbol{q}_{i1}\boldsymbol{q}_{i1}^\mathsf{T} + \boldsymbol{q}_{i2}\boldsymbol{q}_{i2}^\mathsf{T}) M_i^{(s)^\mathsf{T}} \end{bmatrix}$$

$$= 2\Sigma_{A^\circ A^\circ}$$

Hence the normal equation matrix is

$$N = \frac{1}{2}\sum_i \begin{bmatrix} \boldsymbol{X}_{0i} \times \boldsymbol{N}_i & (I_3 - D_{N_i}) M_i^{(s)\mathsf{T}} \\ \boldsymbol{N}_i & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} w_q & \\ & w_\phi I_2 \end{bmatrix} \begin{bmatrix} (\boldsymbol{X}_{0i} \times \boldsymbol{N}_i)^\mathsf{T} & \boldsymbol{N}_i^\mathsf{T} \\ M_i^{(s)}(I_3 - D_{N_i}) & \boldsymbol{0} \end{bmatrix}$$

$$= \frac{1}{2}\sum_i \begin{bmatrix} \boldsymbol{X}_{0i} \times \boldsymbol{N}_i & (I_3 - D_{N_i}) M_i^{(s)\mathsf{T}} \\ \boldsymbol{N}_i & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} w_q (\boldsymbol{X}_{0i} \times \boldsymbol{N}_i)^\mathsf{T} & w_q \boldsymbol{N}_i^\mathsf{T} \\ w_\phi M_i^{(s)}(I_3 - D_{N_i}) & \boldsymbol{0} \end{bmatrix} \tag{6.184}$$

$$= \frac{1}{2}\sum_i \begin{bmatrix} w_q D(\boldsymbol{X}_{0i} \times \boldsymbol{N}_i) + w_\phi (I_3 - D_{N_i}) M_i^{(s)\mathsf{T}} M_i^{(s)} (I_3 - D_{N_i}) & (\boldsymbol{X}_{0i} \times \boldsymbol{N}_i) \boldsymbol{N}_i^\mathsf{T} \\ \boldsymbol{N}_i (\boldsymbol{X}_{0i} \times \boldsymbol{N}_i)^\mathsf{T} & w_q D(\boldsymbol{N}_i) \end{bmatrix}$$

$$= \frac{1}{2}\sum_i \begin{bmatrix} w_q D(\boldsymbol{X}_{0i} \times \boldsymbol{N}_i) + w_\phi (\boldsymbol{q}_{i1}\boldsymbol{q}_{i1}^\mathsf{T} + \boldsymbol{q}_{i2}\boldsymbol{q}_{i2}^\mathsf{T}) & w_q \, S(\boldsymbol{X}_{0i}) D(\boldsymbol{N}_i) \\ w_q \, D(\boldsymbol{N}) S(\boldsymbol{X}_{0i}) & w_q D(\boldsymbol{N}_i) \end{bmatrix} \tag{6.185}$$

$$= \frac{1}{2}\sum_i w_q \begin{bmatrix} (\boldsymbol{X}_{0i} \times \boldsymbol{N}_i)(\boldsymbol{X}_{0i} \times \boldsymbol{N}_i)^\mathsf{T} & (\boldsymbol{X}_{0i} \times \boldsymbol{N}_i) \boldsymbol{N}_i \\ \boldsymbol{N}_i (\boldsymbol{X}_{0i} \times \boldsymbol{N}_i)^\mathsf{T} & \boldsymbol{N}_i \boldsymbol{N}_i^\mathsf{T} \end{bmatrix} + w_\phi \begin{bmatrix} \boldsymbol{q}_{i1}\boldsymbol{q}_{i1}^\mathsf{T} + \boldsymbol{q}_{i2}\boldsymbol{q}_{i2}^\mathsf{T} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}$$

or generally

$$\boxed{N = \frac{1}{2}\sum_i w_{q_i} \begin{bmatrix} \boldsymbol{X}_{0i} \times \boldsymbol{N}_i \\ \boldsymbol{N}_i \end{bmatrix} [\boldsymbol{X}_{0i} \times \boldsymbol{N}_i \, , \, \boldsymbol{N}_i] + w_{\phi_i} \begin{bmatrix} \boldsymbol{q}_{i1} \\ \boldsymbol{0} \end{bmatrix} [\boldsymbol{q}_{i1}^\mathsf{T} \, , \, \boldsymbol{0}^\mathsf{T}] + w_{\psi_i} \begin{bmatrix} \boldsymbol{q}_{i2} \\ \boldsymbol{0} \end{bmatrix} [\boldsymbol{q}_{i2}^\mathsf{T} \, , \, \boldsymbol{0}^\mathsf{T}]} \tag{6.186}$$

Reducing the parameters to the translation yields the reduced normal equation matrix

$$\overline{N}_{TT} = \frac{1}{2}\sum_i \left(w_{q_i} D(\boldsymbol{X}_{0i} \times \boldsymbol{N}_i) + w_{\phi_i}(\boldsymbol{q}_{i1}\boldsymbol{q}_{i1}^\mathsf{T} + \boldsymbol{q}_{i2}\boldsymbol{q}_{i2}^\mathsf{T})\right) \tag{6.187}$$

$$- \left(\sum_i w_{q_i}\ D(\boldsymbol{N}) S(\boldsymbol{X}_{0i})\right) \left(\sum_i w_{q_i} D(\boldsymbol{N}_i)\right)^{-1} \left(\sum_i w_{q_i}\ S(\boldsymbol{X}_{0i}) D(\boldsymbol{N}_i)\right) \tag{6.188}$$

which can be determined if

$$\sum_i w_{q_i} D(\boldsymbol{N}_i) = \sum_i w_{q_i} \boldsymbol{N}_i \boldsymbol{N}_i^\mathsf{T} \tag{6.189}$$

is regular: Therefore at least three planes with non-coplanar normals are necessary for a solution.

7 Planes from Points

We describe the statistically optimal estimation of a single and of multiple planes from a point cloud, where the full covariance matrix of all scene coordinates is available, e.g., from bundle adjustment. This procedure might be used to derive ground truth data for plane extraction or for homography estimation.

7.1 Preface . . . 100
7.2 The Problem . . . 100
7.3 Formalization . . . 101
7.3.1 The incidence constraint . . . 101
7.3.2 The optimization problem . . . 101
7.3.3 Conditioning and approximate values . . . 101
7.3.4 The algorithm for estimating the parameters . . . 102
7.4 Multiple planes . . . 103
7.5 Outlier detection . . . 105

7.1 Preface

The note (2023) describes the statistically optimal estimation of a single and of multiple planes from a point cloud, where the full covariance matrix of the scene coordinates is available, e.g., from bundle adjustment. The solution for single planes differs from that of Sect. 6.3.2 in Ch. 6: There the plane is Euclideanly normalized, here they are spherically normalized, which leads to simpler expressions.

7.2 The Problem

Given are K sets $\{\{\mathcal{X}_i\}, i = i, ..., I\}_k, k = 1, ..., K$ of 3D points together with their complete covariance matrix $\Sigma = [\Sigma_{ik,ik}]$ the task is to finde the best fitting planes $\mathcal{A}_k$. We start with the derivation for a single plane and then generalize to multiple planes.

The motivation is to derive reference data for homographies for identified planes being seen in pairs of images, whose poses and scene points have been determined by bundle adjustment. Instead of including the plane constraints into the bundle adjustment, we propose to use the coordinates of the estimated scene points together with their full covariance matrix and determine the best fitting plane parameters. This can be seen as an estimation in steps (Kalman filtering) where in the second step the plane constraints are used to improve the estimates of the scene points, which in the first step have been determined without these constraints.

Though it is possible to estimate the planes individually, the resulting parameters are not optimal, since the mutual correlations between the scene points belonging to different planes are not taken into account.

We therefore just assume, the coordinates of the relevant scene points together with their full covariance matrix is available, e.g., when using the Ceres solver.

7.3 Formalization

We start with the case $K = 1$ and omit all indices referring tho the plane of interest.

7.3.1 The incidence constraint

We assume the points are given with their homogeneous coordinates $\mathbf{X}_i, i = 1, ..., I$ and their joint covariance matrix

$$\Sigma = [\Sigma_{\mathrm{X}_i\mathrm{X}_j}] = \left[\left[\begin{array}{cc} \Sigma_{X_iX_j} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 0 \end{array} \right]\right], \quad \text{with} \quad i, j = 1, ..., I\,. \tag{7.1}$$

and the plane $\mathcal{A}$ is represented by its spherically normalized homogeneous coordinates $\mathbf{A}$ with

$$|\mathbf{A}| = 1\,. \tag{7.2}$$

The a point $\mathcal{X}_i$ lies on the plane $\mathcal{A}$ if

$$\mathbf{X}_i^\mathsf{T}\mathbf{A} = 0\,. \tag{7.3}$$

7.3.2 The optimization problem

We now want to optimally estimate the plane parameters. The observations and unknown parameters in a Gauss-Helmert model with constraints are

$$\underset{N\times 1}{\boldsymbol{y}} := [\mathbf{X}_i],, \quad \underset{4\times 1}{\boldsymbol{\theta}} := \mathbf{A} \quad \text{and} \quad \underset{N\times 1=4I\times 1}{\boldsymbol{y}} := \mathbb{E}(\underline{\boldsymbol{y}}) \tag{7.4}$$

For achieving a ML-estimation we want minimize the residuals $\boldsymbol{y} - \boldsymbol{l}$ squared and weighted with the full weight matrix W

$$\Omega(\boldsymbol{\theta}, \boldsymbol{y}) = (\boldsymbol{y} - \boldsymbol{y})^\mathsf{T} W (\boldsymbol{y} - \boldsymbol{y}) \quad \text{with} \quad W = \left[\left[\begin{array}{cc} \Sigma^{-1}_{X_iX_j} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 0 \end{array} \right]\right] \tag{7.5}$$

subject to the constraints

$$\begin{array}{ll} \mathbf{0} = \boldsymbol{g}(\boldsymbol{\theta}, \boldsymbol{y}) := & [\mathbf{y}_i^\mathsf{T}\boldsymbol{\theta}]\,, \\ \mathbf{0} = \boldsymbol{h}(\boldsymbol{\theta}) := & \frac{1}{2}(|\boldsymbol{\theta}|^2 - 1)\,. \end{array} \tag{7.6}$$

7.3.3 Conditioning and approximate values

We assume the following:

- We have conditioned the given coordinates

$$\mathbf{X}_i^c = \mathsf{M}\mathbf{X}_i \quad \text{with} \quad \mathsf{M} = \left[\begin{array}{cc} \frac{1}{s} I_3 & -\frac{1}{s}\boldsymbol{\mu}_X \\ \mathbf{0}^\mathsf{T} & 1 \end{array} \right], \tag{7.7}$$

  and

$$s = \sqrt{\frac{1}{3}\mathrm{tr}(\mathbb{C}\mathrm{ov}(\boldsymbol{X}_i)}\,, \quad \text{and} \quad \boldsymbol{\mu}_X = \frac{1}{I}\sum_i \boldsymbol{X}_i \tag{7.8}$$

  since in non-homogeneous coordinates we have $\boldsymbol{X}_i^c = (\boldsymbol{X}_i - \boldsymbol{\mu}_X)/s$. Hence, we have the conditioned covariance matrix

$$\Sigma^c = [\mathsf{M}\Sigma_{ij}\mathsf{M}^\mathsf{T}] \tag{7.9}$$

  Since we determine the plane parameters $\hat{\boldsymbol{\theta}}^c = \hat{\mathbf{A}}^c$ in the conditioned coordinate system where we can uncondition the estimated plane parameters

$$\hat{\boldsymbol{\theta}} = \mathsf{M}\hat{\boldsymbol{\theta}}^c \quad \text{since} \quad \mathbf{A}^c = \mathsf{M}^{-1}\mathbf{A}\,. \tag{7.10}$$

  together with their covariance matrix

$$\Sigma_{\hat{\theta}\hat{\theta}} = \mathsf{M}\Sigma_{\hat{\theta}^c\hat{\theta}^c}\mathsf{M}^\mathsf{T} \tag{7.11}$$

- We have an approximate solution $\boldsymbol{\theta}^{c,a} := \mathbf{A}^{c,a}$ based on the conditioned 3D points assuming all have the same covariance matrix I_3.

7.3.4 The algorithm for estimating the parameters

We refer to PCV Sect. 8.3.2 and the note on the Gauss-Helmert model, Sect. 4.1 augmented by the constraints between the parameters. We omit all superscripts indicating that we have conditioned the data.

We start from the correlated observed I scene points in homogeneous coordinates $\{\boldsymbol{y}, \Sigma_{yy}\} := \{[\mathbf{X}_i], [\Sigma_{ij}]\}$, the constraints $\boldsymbol{g}(\boldsymbol{\theta}, \boldsymbol{y}) := [\mathbf{y}_i^\mathsf{T}\mathbf{A}] = \mathbf{0}$ and $h(\boldsymbol{\theta}) = 1/2(|\boldsymbol{\theta}|^2 - 1)$, and the approximate values $\boldsymbol{\theta}^a := \mathbf{A}^a$ for the unknowns and $\boldsymbol{y}^a := [\mathbf{X}_i]$ for the mean observations. We obtain the following algorithm for an iterative solution:

1. Iterate until convergence

   (a) *Jacobians at current approximations* Determine the Jacobians $\underset{I\times 4}{X}$ and $\underset{I\times 4I}{Z}$ at the current approximate values $(\boldsymbol{\theta}^a, \boldsymbol{y}^a)$. Here we have

   $$\underset{I\times 4}{X} = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\theta}} := Y^a = [\boldsymbol{y}_i^{a\mathsf{T}}], \quad \underset{I\times 4I}{Z^\mathsf{T}} = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{y}} = I_I \otimes \hat{\boldsymbol{\theta}}^{a\mathsf{T}} \quad \text{and} \quad \underset{1\times 4}{\boldsymbol{h}^\mathsf{T}} = \frac{\partial h}{\partial \boldsymbol{\theta}} := \boldsymbol{\theta}^{a,\mathsf{T}}. \tag{7.12}$$

   In the first iteration we have

   $$[\boldsymbol{y}_i^{(0)}] := [\mathbf{X}_i]. \tag{7.13}$$

   (b) *contradictions of constraints given the parameters* Determine the contradictions $\boldsymbol{c}_g$ and $\boldsymbol{c}_h$ of the negative constraints at the approximate values $\boldsymbol{\theta}^a$ and $\boldsymbol{y}$ of the unknown parameters together with their weight matrix [1]

   $$\underset{I\times 1}{\boldsymbol{c}_g} := -\underset{I\times 4}{[\boldsymbol{l}_i^\mathsf{T}]}\ \underset{4\times 1}{\boldsymbol{\theta}^a}, \quad \underset{I\times I}{W_{gg}} = (Z^\mathsf{T}\Sigma Z)^{-1} = \left(\left[\boldsymbol{\theta}^{a,\mathsf{T}}\Sigma_{ij}\boldsymbol{\theta}^a\right]\right)^{-1} \tag{7.14}$$

   and

   $$\underset{1\times 1}{c_h} = \frac{1}{2}(|\boldsymbol{\theta}^a|^2 - 1). \tag{7.15}$$

   (c) *normal equation system* Solve the normal equation system for the corrections $\Delta\boldsymbol{\theta}$ and $\Delta\boldsymbol{y}$ of the parameters

   $$\underbrace{\begin{bmatrix} X^\mathsf{T}W_{gg}X & \boldsymbol{h} \\ \boldsymbol{h}^\mathsf{T} & 0 \end{bmatrix}}_{N} \begin{bmatrix} \Delta\boldsymbol{\theta} \\ \mu \end{bmatrix} = \underbrace{\begin{bmatrix} X^\mathsf{T}W_{gg}\,\boldsymbol{c}_g \\ c_h \end{bmatrix}}_{\boldsymbol{m}}. \tag{7.16}$$

   (d) Update the approximate parameters

   $$\boldsymbol{\theta}^a := \mathsf{N}(\boldsymbol{\theta}^a + \Delta\boldsymbol{\theta}) \quad \text{with} \quad \mathsf{N}(\boldsymbol{x}) = \frac{\boldsymbol{x}}{|\boldsymbol{x}|}. \tag{7.17}$$

   (e) Determine the corrections for the mean observations

   $$\Delta\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{y}^a - \Sigma(I_I \otimes \boldsymbol{\theta}^{a,\mathsf{T}})W_{gg}\boldsymbol{g}(\hat{\boldsymbol{\theta}}^a\boldsymbol{y}). \tag{7.18}$$

   (f) *update of approximate mean observations* Update the approximate mean observations

   $$\boldsymbol{y}^a := [\mathsf{N}^e(\boldsymbol{y}_i + \Delta\boldsymbol{y}_i)] \quad \text{with} \quad \mathsf{N}^e(\mathbf{X}) = \frac{\mathbf{X}}{X_4}. \tag{7.19}$$

2. *final estimates* Set the final estimates of the unknown parameters and of the mean observations, sometimes called the fitted observation $\hat{\boldsymbol{y}} := \hat{\boldsymbol{y}}$

   $$\hat{\boldsymbol{\theta}} := \boldsymbol{\theta}^a \quad \text{and} \quad \hat{\boldsymbol{y}} = \boldsymbol{y}^a. \tag{7.20}$$

timated variance ctor

3. Determine the estimated variance factor

$$\hat{\sigma}_0^2 = \frac{\boldsymbol{c}_g^\mathsf{T} W_{c_g c_g} \boldsymbol{c}_g}{I-4} \,. \tag{7.21}$$

If the model holds its expectation is equal to 1.
Observe: Instead of minimizing the squared residuals $\boldsymbol{y}-\boldsymbol{l}$ weighted with W in (7.5), thus minimize $||\boldsymbol{y}-\boldsymbol{y}||_W$, we equivalently may minimize the weighted residuals of the squared constraints $\boldsymbol{c}_g = -\boldsymbol{g}(\boldsymbol{\theta}, \boldsymbol{y})$ weighted with their weight matrix $W_{c_g c_g}$, thus minimizing $||\boldsymbol{g}(\boldsymbol{\theta}, \boldsymbol{y})||_{W_{c_g c_g}}$, in both cases taking the constraints (7.6) into account.

4. Determine the covariance matrix of the estimated parameters

covariance matrix of the estimated parameters

$$\begin{bmatrix} X^\mathsf{T} W_{gg} X & \boldsymbol{h} \\ \boldsymbol{h}^\mathsf{T} & 0 \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma_{\hat{\theta}\hat{\theta}} & \cdot \\ \cdot & \cdot \end{bmatrix} . \tag{7.22}$$

Remark: If the observational noise is small and an approximate solution is acceptable, the steps 1.(e–f) can be omitted. Then the Jacobians X and Z are to be determined at $(\boldsymbol{\theta}^a, \boldsymbol{y})$ instead of at $(\boldsymbol{\theta}^a, \boldsymbol{y}^a)$. ◇

The complete procedure is given in the algorithm below.

Algorithm 2: Plane from correlated points, assuming conditioned values.
$[\hat{\mathbf{A}}, \Sigma_{\hat{A}\hat{A}}, \hat{\sigma}_0^2, R]$ = `CorrelatedPoints2Plane_D`$([\mathbf{X}_i], [\Sigma_{ij}], \mathbf{A}^a, T_\theta,$ `maxiter`$)$
Input: observed values $\boldsymbol{y} = [\boldsymbol{y}_i] := [\mathbf{X}_i]$, full covariance matrix $\Sigma = [\Sigma_{ij}]$
approximate values $\mathbf{A}^a$,
parameters T_θ, `maxiter` for controlling convergence.
Output: estimated parameters $\hat{\mathbf{A}}, \Sigma_{\hat{A}\hat{A}}$ for plane, variance factor $\hat{\sigma}_0^2$, redundancy R.

Redundancy $R = I - 3$;
if $R < 0$ **then stop**, not enough constraints;
Iteration $\nu = 0$, approx. values $\hat{\boldsymbol{\theta}}^a := \mathbf{A}^a$, $\boldsymbol{y}^a := [\mathbf{X}_i]$, stopping variable: $s = 0$;
repeat
  Jacobians: : $A = [\boldsymbol{y}_i^{a,\mathsf{T}}]$, $\boldsymbol{h} = \hat{\boldsymbol{\theta}}^a$;
  Constraints: $\boldsymbol{c}_g = -[\boldsymbol{y}_i^\mathsf{T}]\boldsymbol{\theta}^a$, $c_h = -1/2(|\boldsymbol{\theta}^a|^2 - 1)$;
  Weight matrix of constraints: $W_{gg} = [\boldsymbol{\theta}^{a,\mathsf{T}} \Sigma_{ij} \boldsymbol{\theta}^a]^{-1}$;
  Build normal equation system: $[N, \boldsymbol{m}]$, see (7.16);
  if N is singular **then stop**: normal equation matrix is singular;
  Updates of parameter vector $\boldsymbol{\theta}^a := \mathsf{N}(\boldsymbol{\theta}^a + \Delta\boldsymbol{\theta})$;
  Corrections for fitted observations: $\Delta\boldsymbol{y}$, see (7.18);
  Update fitted observations $\boldsymbol{y}^a = [\mathsf{N}^e(\boldsymbol{y}_i^a + \Delta\boldsymbol{y}_i)]$, see (7.19);
  Set iteration: $\nu := \nu + 1$;
  if $\max_u(|\widehat{\Delta\theta}_u|/\sigma^a_{\hat{\theta}_u}) < T_\theta$ or $\nu =$ `maxiter` **then** $s = 2$;
until $s \equiv 2$;
Estimated parameters $\hat{\mathbf{A}} := \hat{\boldsymbol{\theta}}^a$ and covariance matrix : $\Sigma_{\hat{A}\hat{A}}$, see (7.22);
if $R > 0$ **then** variance factor $\hat{\sigma}_0^2 = \boldsymbol{c}_g^\mathsf{T} W_{gg}\, \boldsymbol{c}_g / R$;
else $\hat{\sigma}_0^2 = 1$;

7.4 Multiple planes

We generalize the solution to the case of simultaneously estimating a set of K planes, in order to exploit all information for one bundle adjustment. This will yield different results due to the correlation between the scene points.

[1] We do not indicate, that $\boldsymbol{c}_g$ depends on approximate values thus omit a superscript ^{a}.

We consider K planes $\Pi_k, k = 1, ..., K$ with their I_k points $\mathcal{X}_{ik}, (ik) \in \mathcal{I}_k$. We assume the point sets for different planes are disjunct. We collect the I_k homogeneous coordinates of the observed scene points and their expectation for plane k in the $I_k \times 4$ matrices

$$\underset{I_K \times 4}{\mathsf{X}_k} = [\mathbf{X}_{ik}^\mathsf{T}] \quad \text{and} \quad \mathsf{Y}_k = \mathbb{E}(\mathsf{X}_k)\,. \tag{7.23}$$

Then we have the following

$$G = \sum_k I_k \tag{7.24}$$

constraints

$$\underset{G \times 1}{\boldsymbol{g}} = [\boldsymbol{g}_{ik}] = [\mathbb{E}(\mathsf{X}_k)\mathbf{A}_k] = [\mathbb{E}(\mathbf{X}_{ik}^\mathsf{T})\mathbf{A}_k] = \mathbf{0}\,, \quad h_k = \frac{1}{2}(|\mathbf{A}_k|^2 - 1) = 0 \quad k = 1, ..., K\,. \tag{7.25}$$

Witht the $4K$ unknown parameters, the $4G$ observations and their expectations

$$\underbrace{\boldsymbol{x}}_{4K \times 1} = [\boldsymbol{x}_k] := [\mathbf{A}_k]\,, \quad \underbrace{\boldsymbol{y}}_{\sum_k I_k} = [\boldsymbol{y}_{ik}] = [\mathbf{X}_{ik}] \quad \text{and} \qquad \text{and} \quad \boldsymbol{y} = \text{vec}(\mathsf{Y}^\mathsf{T}) = [\mathbb{E}(\mathbf{X}_{ik})] \tag{7.26}$$

the Jacobians X and Z are the following using the approximate values for $\boldsymbol{\theta}$ and Y

$$\mathsf{X} = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{x}} = \text{Diag}([\mathsf{X}_k]) := \text{Diag}([\mathsf{Y}_k]) \quad \text{and} \quad \mathsf{Z}^\mathsf{T} = \text{Diag}([\mathsf{Z}_{ik}^\mathsf{T}]) := \text{Diag}([\hat{\boldsymbol{\theta}}_k^{a\mathsf{T}}]) \tag{7.27}$$

The Jacobian for the constraints is

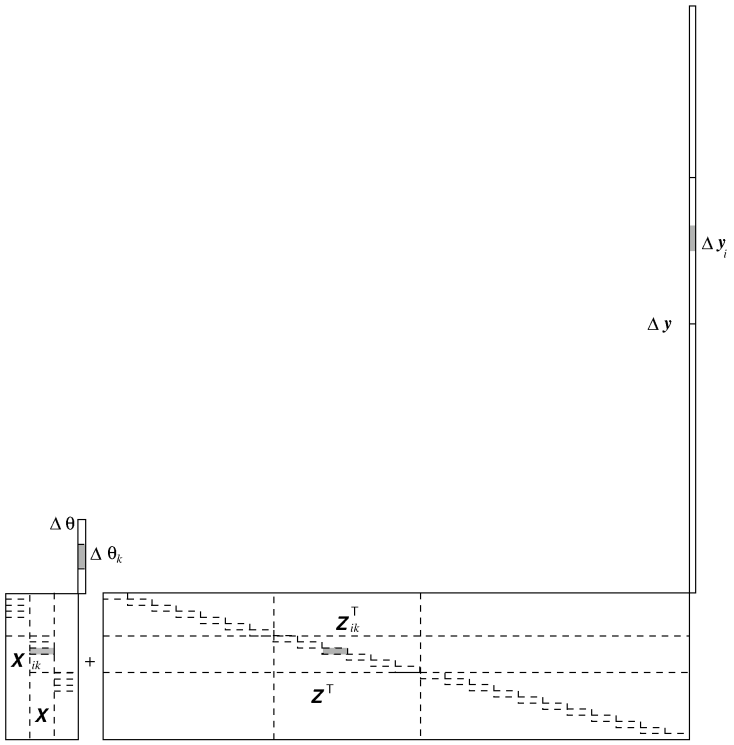

Figure 7.1: linearized constraints

$$\underset{4K \times K}{\mathsf{H}} = \text{Diag}([\boldsymbol{\theta}_k^a]) \tag{7.28}$$

Hence, with the approximate residuals

$$\boldsymbol{v}^a = \boldsymbol{y}^a - \boldsymbol{y} \tag{7.29}$$

we have the linearized optimization problem: Minimize

$$\Omega(\Delta\boldsymbol{\theta}, \Delta\boldsymbol{y}) = (\boldsymbol{y}^a + \Delta\boldsymbol{y})^{\mathsf{T}} W^{+} (\boldsymbol{y}^a + \Delta\boldsymbol{y}) \quad \text{with} \quad W = \left[\left[\begin{array}{cc} \Sigma^{-1}_{X_i X_j} & \mathbf{0} \\ \mathbf{0}^{\mathsf{T}} & 0 \end{array}\right]\right] \tag{7.30}$$

subject to the constraints

$$\begin{array}{ll} \mathbf{0} = \boldsymbol{g}(\Delta\boldsymbol{\theta}, \Delta\boldsymbol{y}) := & A\Delta\boldsymbol{\theta} + Z^{\mathsf{T}}\Delta\boldsymbol{y} - \boldsymbol{g}(\boldsymbol{\theta}^a, \boldsymbol{y}^a)\,, \\ \mathbf{0} = \boldsymbol{h}(\Delta\boldsymbol{\theta}) := & H^{\mathsf{T}}\Delta\boldsymbol{\theta} - \boldsymbol{h}(\Delta\boldsymbol{\theta}^a)\,. \end{array} \tag{7.31}$$

The full weight matrix of the constraints is

$$\underset{G\times G}{W_{gg}} = \left(\left[\boldsymbol{\theta}^{a,\mathsf{T}}_{ik} \Sigma_{ik,i'k'} \boldsymbol{\theta}^{a}_{i'k'}\right]\right)^{-1} \quad \text{with} \quad (ik) \in \mathcal{I}_k, k = 1, ..., K \tag{7.32}$$

hence, with the residual constraints

$$\boldsymbol{c}_g = -\boldsymbol{g}(\boldsymbol{\theta}^a, \boldsymbol{y}) = -[A_k]\boldsymbol{\theta}^a \quad \text{and} \quad \boldsymbol{c}_h = -\boldsymbol{h}(\boldsymbol{\theta}^a) \tag{7.33}$$

the normal equation system is

$$\underbrace{\left[\begin{array}{cc} A^{\mathsf{T}} W_{gg} A & H \\ H^{\mathsf{T}} & 0 \end{array}\right]}_{\underset{5K\times 5K}{N}} \left[\begin{array}{c} \underset{4K\times 1}{\Delta\boldsymbol{\theta}} \\ \underset{K\times 1}{\boldsymbol{\mu}} \end{array}\right] = \underbrace{\left[\begin{array}{c} A^{\mathsf{T}} W_{gg}\, \boldsymbol{c}_g \\ \boldsymbol{c}_h \end{array}\right]}_{\boldsymbol{m}}. \tag{7.34}$$

which, except for the block off-diagonal matrix H, is full. The algorithm above requires transparent adaptions.

Observe, the resulting plane parameters will be mutually correlated. But their individual 4×4 covariance matrix $\mathbb{D}(\underline{\mathbf{A}}_k)$ may be reported as uncertainty of the *ground truth*.

7.5 Outlier detection

It may be useful to eliminate individual scene points before a final plane estimation. The following test statistic can be used for outlier detection

$$X_{ik} = \boldsymbol{c}^{\mathsf{T}}_{g_{ik}} W_{g_{ik},g_{ik}} \boldsymbol{c}_{g_{ik}} = w_{g_{ik},g_{ik}} \mathbf{X}^{\mathsf{T}}_{ik} \hat{\boldsymbol{\theta}}_k\,. \tag{7.35}$$

hence we explicitly need the weight matrix W_{gg} in (7.32). If the given model is correct, especially if the covariance matrix of the scene points is correct, then the test statistic X_{ik} follows a χ^2_4-distribution.

In case, we normalize the test statistic by some estimate for the variance factor, its distribution is not known.

8 Direct Solutions for the Similarity from Plane Pairs

We collect some direct solutions for determining the similarity (or motion) from corresponding plane pairs, representing point clouds. Some of the solutions are able to handle the case, where the sign of the normals are not consistent.

8.1 Problem . . . 106
8.2 Minimal solutions . . . 106
8.2.1 A one-step direct solution of a similarity from four plane pairs . . . 107
8.2.2 A two-step solution for a motion from three planes . . . 107
8.3 Direct solutions the similarity from $I \geq 4$ plane pairs . . . 108
8.3.1 One step procedure . . . 108
8.3.2 Two step procedure . . . 108
8.4 Stability of the solution . . . 110

8.1 Problem

Given are plane pairs $\{\mathbf{A}_i, \mathbf{A}'_i\}_i = 1, ..., I$ which are assumed to be related by the similarity

$$\mathbf{A}'_i = \mathsf{H}^{-\mathsf{T}} \mathbf{A}_i \,. \tag{8.1}$$

determine a good estimate of H

$$\mathsf{H} = \begin{bmatrix} \lambda R & \boldsymbol{T} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} = \begin{bmatrix} R & \boldsymbol{T}/\lambda \\ \mathbf{0}^\mathsf{T} & 1/\lambda \end{bmatrix} . \tag{8.2}$$

We assume the planes to be Euclideanly normalized

$$\mathbf{A} = \begin{bmatrix} \boldsymbol{N} \\ -S \end{bmatrix} , \qquad \text{with} \qquad |\boldsymbol{N}| = 1. \tag{8.3}$$

In addition, we assume the coordinates to be conditioned, i. e. the distances of the planes to the origin should be less than 1. This can be achieved by a proper similarity transformation of coordinate system, such that the origin is in the center of all points and the distances S_i have absolute coordinates less than 1.

As the normals may not be consistent, as $\mathbf{A}$ and $-\mathbf{A}$ represent the same plane, we can distinguish two types of solutions, one which assumes the normals to be consistent, the other assuming they are not consistent.

In the following we first discuss solutions which do not exploit the full covariance structure or even do not refer to a statistical description of the uncertainty.

8.2 Minimal solutions

We discuss minimal a minimal solution for spatial similarity and for spatial motion.

Sect.	No. I of planes	normals	reflection
8.2.1	$I = 4$	consistent	allowed
8.2.2	$I = 3$	consistent	allowed
8.3.1	$I \geq 4$	not consistent	allowed
8.3.2	$I \geq 4$	consistent	allowed
		not consistent	not allowed

Table 8.1: Direct solution for the similarity from plane pairs

8.2.1 A one-step direct solution of a similarity from four plane pairs

The direct solution can be obtained from $\mathsf{H}^\mathsf{T}\mathbf{A}'_i = \mathbf{A}_i$ or $\mathbf{A}_i{}'^\mathsf{T}\mathsf{H} = \mathbf{A}_i^\mathsf{T}$, or

$$\mathsf{A} = \begin{bmatrix} \mathbf{A}_1^\mathsf{T} \\ \mathbf{A}_2^\mathsf{T} \\ \mathbf{A}_2^\mathsf{T} \\ \mathbf{A}_2^\mathsf{T} \end{bmatrix} = \begin{bmatrix} \mathbf{A}'^\mathsf{T}_1 \\ \mathbf{A}'^\mathsf{T}_2 \\ \mathbf{A}'^\mathsf{T}_3 \\ \mathbf{A}'^\mathsf{T}_4 \end{bmatrix} \begin{bmatrix} R & \boldsymbol{T}/\lambda \\ \mathbf{0}^\mathsf{T} & 1/\lambda \end{bmatrix} = \mathsf{A}'\mathsf{H} \tag{8.4}$$

Thus we directly obtain

$$\mathsf{H} = (\mathsf{A}')^{-1}\mathsf{A} \tag{8.5}$$

The matrix would be the correct result, if the data were noiseless. This is valid for both, a similarity and a motion.

Therefore, in general we enforce the matrix to be a similarity by enforcing the upper left 3×3-matrix to be a rotation and the lack of a projective component. With

$$\mathsf{H}(1:3,1:3) = UDV^\mathsf{T} \tag{8.6}$$

we therefore have the best estimate for a *similarity*

$$\mathsf{H} = \begin{bmatrix} |D|^{1/3}UV^\mathsf{T} & \mathsf{H}(1:3,4) \\ \mathbf{0}^\mathsf{T} & \mathsf{H}(4,4) \end{bmatrix} \tag{8.7}$$

This solution assumes the normals of the planes to be consistent. It allows for a mirroring.

8.2.2 A two-step solution for a motion from three planes

The two-step solution first determines the rotation from the three normals and then the translation from the intersection point.

Rotation. The rotation directly can be determined from the normals using

$$B' = [\boldsymbol{N}'_1, \boldsymbol{N}'_2, \boldsymbol{N}'_3] = R[\boldsymbol{N}_1, \boldsymbol{N}_2, \boldsymbol{N}_3] = RB \tag{8.8}$$

from

$$R = B^{-1}B' \tag{8.9}$$

which in case the data are noisy is no rotation. The best rotation is again obtained from the SVD of $R = UDV^\mathsf{T}$ from

$$\widehat{R} = UV^\mathsf{T} \tag{8.10}$$

If the data are related by a reflection, then $\det(\widehat{R}) = -1$.

Translation. The translation can easily be determined from the intersection point of the three planes.

Also, this solution assumes the normals of the planes to be consistent. The result allows the data to contain a reflection.

8.3 Direct solutions the similarity from $I \geq 4$ plane pairs

8.3.1 One step procedure

The basic constraint for each plane can be written as (see Heuel 2004, eq. (3.29) and sect. 3.3.1.6, tables 3.5 and 3.9)

$$\mathcal{A}_i \equiv \mathcal{A}_i' : \ \mathbf{A}_i \cap (\mathsf{H}^\mathsf{T}\mathbf{A}_i') = \overline{\mathsf{\Pi}}(\mathbf{A}_i)\mathsf{H}^\mathsf{T}\mathbf{A}_i' = \mathbf{0}\,, \tag{8.11}$$

or

$$(\overline{\mathsf{\Pi}}(\mathbf{A}_i) \otimes \mathbf{A}_i^\mathsf{T})\mathrm{vec}\mathsf{H} \stackrel{!}{=} \mathbf{0} \tag{8.12}$$

with the matrix

$$\underset{6\times 4}{\overline{\mathsf{\Pi}}(\mathbf{A})} = \begin{bmatrix} S(\boldsymbol{N}) & \mathbf{0} \\ -SI_3 & -\boldsymbol{N} \end{bmatrix} \tag{8.13}$$

containing the skew matrix $S(\boldsymbol{N})$ of the 3-vector $\boldsymbol{N}$. Observe, this constraint is independent on the sign of the plane vectors.

This gives rise to the direct solution

$$\underbrace{\begin{bmatrix} \overline{\mathsf{\Pi}}(\mathbf{A}_1) \otimes \mathbf{A}_1^\mathsf{T} \\ \cdots \\ \overline{\mathsf{\Pi}}(\mathbf{A}_i) \otimes \mathbf{A}_i^\mathsf{T} \\ \cdots \\ \overline{\mathsf{\Pi}}(\mathbf{A}_I) \otimes \mathbf{A}_I^\mathsf{T} \end{bmatrix}}_{\underset{6I\times 16}{B}} \mathbf{h} \stackrel{!}{=} \mathbf{0} \tag{8.14}$$

The best estimate for $\mathbf{h}$ is the right singular vector of the $6I \times 16$-matrix B belonging to the smallest singular value.

As each plane gives rise to three constraints, we need at least five planes. As we know that the elements $\mathsf{H}_{4,1:3}$ are zero, we can cancel the corresponding columns in the matrix B, then being of size $6I \times 12$ and can do with four planes minimum.

The result is an affinity

$$\mathsf{H} = \begin{bmatrix} A & \boldsymbol{T} \\ \mathbf{0}^\mathsf{T} & s \end{bmatrix} \tag{8.15}$$

which needs to be enforced to become a similarity, with

$$A = UDV^\mathsf{T} \tag{8.16}$$

leading to

$$\widehat{\mathsf{H}} = \begin{bmatrix} |D|^{1/3}/s\, UV^\mathsf{T} & \boldsymbol{T}/s \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix}. \tag{8.17}$$

Since only the deviation from the 0-constraints (8.14) is minimized, this solution allows the normals to be inconsistent. Again, if the data contain a reflection, the solution will be a reflection.

8.3.2 Two step procedure

We first determine the rotation, then rotate the planes and then determine translation and scale. Thus we assume the similarity to be

$$\widehat{\mathsf{H}} = \begin{bmatrix} \widehat{R} & \widehat{\boldsymbol{T}}/\widehat{\lambda} \\ \mathbf{0} & 1/\widehat{\lambda} \end{bmatrix} = \begin{bmatrix} I_3 & \widehat{\boldsymbol{T}}' \\ \mathbf{0}^\mathsf{T} & \widehat{\mu}' \end{bmatrix} \begin{bmatrix} \widehat{R} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 1 \end{bmatrix} \tag{8.18}$$

with

$$\widehat{\boldsymbol{T}} = \widehat{\boldsymbol{T}}'/\widehat{\mu}'\,, \qquad \widehat{\lambda} = 1/\widehat{\mu}' \tag{8.19}$$

8.3.2.1 Determining the rotation

Assuming consistency of the normals. For finding the optimal rotation we minimize the optimization function

$$\sum_i p_i |\boldsymbol{N}'_i - R\boldsymbol{N}_i|^2 \tag{8.20}$$

which is equivalent to maximize

$$\sum_i p_i \boldsymbol{N}'_i R\boldsymbol{N}_i = \mathrm{tr}(RH)\,, \qquad \mathsf{H} = \sum_i p_i \boldsymbol{N}_i \boldsymbol{N}'_i \tag{8.21}$$

The weights p_i can be approximated by

$$p_i = \frac{1}{\sigma^2_{\phi_i} + \sigma^2_{\phi'_i}} \approx \frac{N_i^3 N_i'^3}{N_i^3 + N_i'^3}\,. \tag{8.22}$$

The approximation is valid in case the planes have been determined from N_i and N'_i points, assuming the normals to have isotropic uncertainty. The solution can be found by using the SVD (or equivalently using quaternions)

$$H = UDV^\mathsf{T} \tag{8.23}$$

leading to the rotation

$$R = VU^\mathsf{T}\,. \tag{8.24}$$

If the data contain a reflection, then $\det R = -1$.

Not assuming consistency of the normals. From the constraints

$$\mathcal{N}'_i \equiv \mathcal{R}(\mathcal{N}_i):\ \boldsymbol{N}'_i \times R\boldsymbol{N}_i = S(\boldsymbol{N}'_i)R\boldsymbol{N}_i = (\boldsymbol{N}_i^\mathsf{T} \otimes S(\boldsymbol{N}'_i))\mathrm{vec}R \stackrel{!}{=} \mathbf{0} \tag{8.25}$$

we obtain the joint constraints

$$\begin{bmatrix} \boldsymbol{N}_1^\mathsf{T} \otimes S(\boldsymbol{N}'_1) \\ \cdots \\ \boldsymbol{N}_i^\mathsf{T} \otimes S(\boldsymbol{N}'_i) \\ \cdots \\ \boldsymbol{N}_I^\mathsf{T} \otimes S(\boldsymbol{N}'_I) \end{bmatrix} \boldsymbol{r} \stackrel{!}{=} \mathbf{0} \tag{8.26}$$

This yields an approximation for a rotation matrix, except for the sign. Hence, we are not able to allow for reflections. From $R = UDV^\mathsf{T}$ we obtain an estimate for the rotation

$$\widehat{R} = UV^\mathsf{T}\mathrm{sign}(|UV'|) \tag{8.27}$$

with $\det \widehat{R} = 1$.

8.3.2.2 Rotating the planes

We now rotate the planes, which just needs to be applied to the normals, therefore

$$\overline{\mathbf{A}}_i = \begin{bmatrix} \overline{\boldsymbol{N}}_i \\ -S_i \end{bmatrix} = \begin{bmatrix} \widehat{R}\boldsymbol{N}_i \\ -S_i \end{bmatrix}, \qquad \overline{\mathbf{A}}'_i = \begin{bmatrix} \overline{\boldsymbol{N}}'_i \\ -S'_i \end{bmatrix} = \begin{bmatrix} \widehat{R}\boldsymbol{N}'_i \\ -S'_i \end{bmatrix} \tag{8.28}$$

These planes only differ by scale and translation.

8.3.2.3 Estimating translation and scale

Transforming planes by translation $\boldsymbol{T}'$ and scale μ' is performed by

$$\overline{\mathbf{A}}_i = \begin{bmatrix} I_3 & \mathbf{0} \\ \boldsymbol{T}'^{\mathsf{T}} & \mu' \end{bmatrix} \overline{\mathbf{A}}'_i \tag{8.29}$$

thus only refers to the distances S_i and S'_i. We have the constraint

$$c_i = S_i - [\overline{\boldsymbol{N}}_i'^{\mathsf{T}} - S'_i] \begin{bmatrix} \boldsymbol{T}' \\ \mu \end{bmatrix} \stackrel{!}{=} 0 \tag{8.30}$$

with an approximate weight

$$w_i \approx \frac{1}{\sigma_{q_i}^2 + \sigma_{q'_i}^2} \approx \frac{N_i N'_i}{N_i + N'_i} \tag{8.31}$$

for the uncertainty of the position across the planar patches (but see the critics below).

Therefore, we can determine the scale and the translation from

$$\boldsymbol{S} = \begin{bmatrix} S_1 \\ \cdots \\ S_i \\ \cdots \\ S_I \end{bmatrix} \stackrel{!}{=} \underbrace{\begin{bmatrix} \overline{\mathbf{A}}'^{\mathsf{T}}_1 \\ \cdots \\ \overline{\mathbf{A}}'^{\mathsf{T}}_i \\ \cdots \\ \overline{\mathbf{A}}'^{\mathsf{T}}_I \end{bmatrix}}_{\substack{B \\ I \times 4}} \begin{bmatrix} \boldsymbol{T}' \\ \mu' \end{bmatrix} \tag{8.32}$$

The least squares solution for the translation and the scale is

$$\begin{bmatrix} \widehat{\boldsymbol{T}}' \\ \widehat{\mu}' \end{bmatrix} = (B^{\mathsf{T}} W B)^{-1} B W^{\mathsf{T}} \boldsymbol{S}\,, \qquad W = \mathrm{Diag}([w_1, ..., w_i, ..., w_I])\,. \tag{8.33}$$

which in the case of four planes reduces to

$$\begin{bmatrix} \widehat{\boldsymbol{T}}' \\ \widehat{\mu}' \end{bmatrix} = B^{-1} \boldsymbol{S} \tag{8.34}$$

The procedure cannot be based on some statistical model.

8.4 Stability of the solution

In case all planes are parallel the rotation cannot be determined.

In case the normals $\boldsymbol{A}_{hi}$ of the planes are coplanar, the translation cannot be determined.

In case the four planes intersect in one point the four plane vectors are linearly dependent and the matrices A and A' in (8.4) are singular or - in case of noise - close to singular. Then the scale cannot be determined.

In case the normals are well distributed the condition numbers

$$\kappa = \frac{\lambda_{\max}}{\lambda_{\min}} \tag{8.35}$$

of A and A' should be significantly less than the inverse standard deviation of the directions measured in radiants.

Part III

Technical Notes on Bundle Adjustment and Surface Reconstruction

9 Rule of Thumb for Precision of Points from Multiview Triangulation

For planning bundle adjustment configurations, the expected accuracy of triangulated points is an essential ingredient. We derive rules of thumb for the accuracy of multi-view triangulating by providing simple expressions for the depth and lateral accuracy of 3D points, for images arranged in a line, in a planar region and in a spherical region, covering the case of omnidirectional cameras.

9.1 Preface . . . 112
9.2 Problem . . . 112
9.3 Formal statement . . . 113
9.4 Linearization . . . 113
9.5 Special configurations . . . 114
9.5.1 Projection centers are on a straight line . . . 114
9.5.2 Projection centers are on a regular grid . . . 115
9.5.3 Projection centers on a spherical cap . . . 116

9.1 Preface

The note (2013) provides explicit expressions (rules of thumb) for the depth accuracy obtained from multi-view triangulation for three cases: (1) the projection centers lie in a line, (2) the projection centers lie in square, and (3) the projection centers are equally spaced on a spherical cap. The note is the basis for Förstner and Wrobel (2016, Sect. 15.7.1).

9.2 Problem

Given T images of a 3D point determine the precision of its position.

The standard deviation depends on

1. on whether the projection centers are in a row, in a rectangular grid, or on a spherical cap
2. the coordinate precision $\sigma_{x'}$ or the directional precision σ_α,
3. the principal distance c,
4. the baseline B or the diameter D of the set of projection centers, on the spherical cap δ measured in radiants, and
5. the common height Z above the unknown point or the radius Z of the spherical cap.

If the T projection centers are in a row we have for large T

$$\boxed{\sigma_{\widehat{W}}^{(1D)} = \frac{\sqrt{12}}{T^{3/2}} \frac{Z^2}{B} \frac{\sigma_{x'}}{c} = \sqrt{\frac{12}{T}} \frac{Z^2}{D} \frac{\sigma_{x'}}{c}.} \tag{9.1}$$

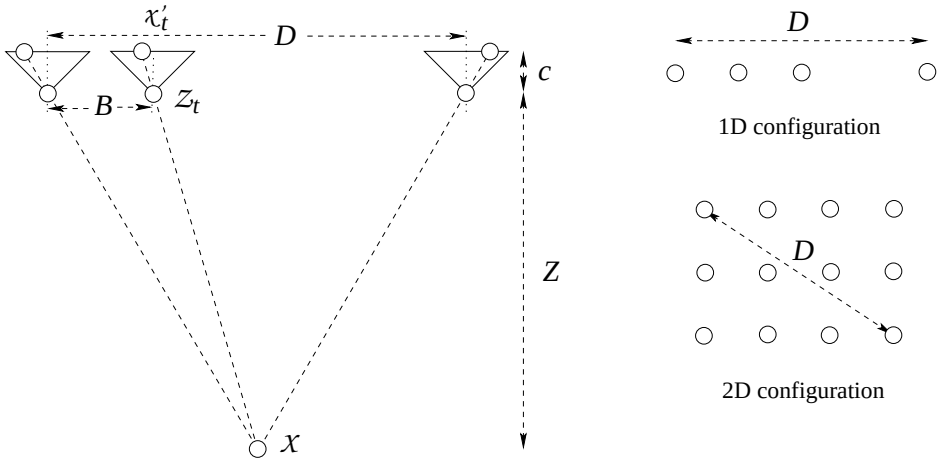

Figure 9.1: Ideal configuration for triangulation. Alternatively, the projection centers are on a sphere with radius Z regularly spaced in a spherical cap with diameter δ.

If the T projection centers are in a rectangular grid we have for large T

$$\boxed{\sigma_{\widehat{W}}^{(2D)} = \frac{\sqrt{6}}{T}\frac{Z^2}{B}\frac{\sigma_{x'}}{c} = \sqrt{\frac{12}{T}}\frac{Z^2}{D}\frac{\sigma_{x'}}{c}}, \tag{9.2}$$

If the T projection centers are evenly distributed on a spherical cap with diameter δ we have

$$\boxed{\sigma_{\widehat{W}}^{(\text{cap})} = \frac{\sqrt{3}}{\sqrt{T}}\frac{Z}{2-\cos\frac{\delta}{2}-\cos^2\frac{\delta}{2}}\sigma_\alpha\,.} \tag{9.3}$$

9.3 Formal statement

Without loss of generality the scene coordinate system sits close to the unknown scene point $\mathcal{X}([U,V,W])$. It is observed in T cameras, which for simplicity are assumed to be identical and are nadir views with $\mathsf{R} = \mathsf{I}_3$. Their common principal distance is c. Their projection centers $\mathcal{Z}_t$ are at $\boldsymbol{Z}_t, t = 1,...,T$. The projection matrices therefore are

$$\mathsf{P}_t = \text{Diag}([c,c,1])[\mathsf{I}_3|-\boldsymbol{Z}_t]\,. \tag{9.4}$$

We observe the T image points

$$\boldsymbol{x}'_t = c\begin{bmatrix} x_t \\ y_t \end{bmatrix} = c\frac{1}{U-Z_t}\begin{bmatrix} V-X_t \\ W-Y_t \end{bmatrix}. \tag{9.5}$$

The task is to estimate the unknown parameters $\boldsymbol{X}$.

9.4 Linearization

Using $\boldsymbol{X}^a = \boldsymbol{0}$, the linearized model reads as

$$\Delta\boldsymbol{x}'_t = c\begin{bmatrix} -\frac{1}{Z_t} & 0 & \frac{X_t}{Z_t^2} \\ 0 & -\frac{1}{Z_t} & \frac{Y_t}{Z_t^2} \end{bmatrix}\begin{bmatrix} \Delta U \\ \Delta V \\ \Delta W \end{bmatrix}. \tag{9.6}$$

With weights w_t for each point we obtain the normal equation matrix

$$\mathsf{N} = c^2\begin{bmatrix} \sum_t \frac{w_t}{Z_t^2} & 0 & -\sum_t \frac{w_t X_t}{Z_t^3} \\ 0 & \sum_t \frac{w_t}{Z_t^2} & -\sum_t \frac{w_t Y_t}{Z_t^3} \\ -\sum_t \frac{w_t X_t}{Z_t^3} & -\sum_t \frac{w_t X_t}{Z_t^3} & \sum_t \frac{w_t(X_t^2+Y_t^2)}{Z_t^4} \end{bmatrix} \tag{9.7}$$

If we assume the projection centers have the same Z-coordinate we obtain

$$\mathsf{N} = \frac{c^2}{Z^4}\begin{bmatrix} \sum_t w_t Z^2 & 0 & -\sum_t w_t X_t Z \\ 0 & \sum_t w_t Z_t^2 & -\sum_t w_t Y_t Z \\ -\sum_t w_t X_t Z & -\sum_t w_t X_t Z & \sum_t w_t (X_t^2 + Y_t^2) \end{bmatrix} \tag{9.8}$$

If we now assume the X- and Y-coordinates are centred with

$$\bar{X} = \frac{\sum_t w_t X_t}{\sum_t w_t} \quad \bar{Y} = \frac{\sum_t w_t Y_t}{\sum_t w_t} \tag{9.9}$$

and the weights are constant

$$w = \frac{1}{\sigma_{x'}^2} \tag{9.10}$$

the normal equation matrix is diagonal

$$\mathsf{N} = \frac{c^2}{Z^4 \sigma_{x'^2}}\begin{bmatrix} TZ^2 & 0 & 0 \\ & TZ^2 & 0 \\ 0 & 0 & \sum_t (X_t^2 + Y_t^2) \end{bmatrix}. \tag{9.11}$$

If we use the average distance of the projection center from its centroid

$$S = \sqrt{\frac{\sum_t (X_t^2 + Y_t^2)}{T}} \tag{9.12}$$

of the projection centers it reads as

$$\mathsf{N} = \frac{c^2}{Z^4 \sigma_{x'^2}}\begin{bmatrix} TZ^2 & 0 & 0 \\ & TZ^2 & 0 \\ 0 & 0 & S^2 T \end{bmatrix}. \tag{9.13}$$

Thus the variances of the 3D point are

$$\sigma_{\hat{U}} = \sigma_{\hat{V}} = \frac{Z}{\sqrt{T}}\frac{\sigma_{x'}}{c} \quad \text{and} \quad \sigma_{\widehat{W}} = \frac{Z^2}{S\sqrt{T}}\frac{\sigma_{x'}}{c} = \frac{Z}{S}\sigma_{\hat{U}}\,. \tag{9.14}$$

9.5 Special configurations

9.5.1 Projection centers are on a straight line

If the T projection centers are on a straight line with basis B in X-direction, their X_t-coordinates are

$$X_t = \left(t - \frac{T+1}{2}\right) B \qquad t = 1, ..., T \quad \text{with} \quad -X_1 = X_T = \frac{T-1}{2} B\,. \tag{9.15}$$

Then we have

$$S^2 = \frac{1}{12}(T^2 - 1)B^2\,. \tag{9.16}$$

Thus we obtain the standard deviation

$$\sigma_{\widehat{W}} = \frac{\sqrt{12}}{\sqrt{T(T^2-1)}}\frac{Z^2}{B}\frac{\sigma_{x'}}{c}\,. \tag{9.17}$$

For large T we can use the approximation

$$\boxed{\sigma_{\widehat{W}}^{(1D)} = \frac{\sqrt{12}}{T^{3/2}}\frac{Z^2}{B}\frac{\sigma_{x'}}{c}\,.} \tag{9.18}$$

If we use as reference the diameter D of the projection centers

$$D = (T-1)B \tag{9.19}$$

the average distance is

$$S^2 = \frac{1}{12}\frac{T+1}{T-1}D^2\,. \tag{9.20}$$

and the standard deviation is

$$\sigma_{\widehat{W}} = \sqrt{\frac{12(T-1)}{T(T+1)}}\frac{Z^2}{D}\frac{\sigma_{x'}}{c}\,. \tag{9.21}$$

which for large T simplifies to

$$\boxed{\sigma_{\widehat{W}}^{(1D)} = \sqrt{\frac{12}{T}}\frac{Z^2}{D}\frac{\sigma_{x'}}{c}\,.} \tag{9.22}$$

9.5.2 Projection centers are on a regular grid

If the $T = MN$ projection centers are on a regular grid with basis B_X in X- and B_Y in Y direction, their coordinates are

$$X_m = \left(m - \frac{M+1}{2}\right)B_X \qquad m = 1, ..., M \quad \text{and} \quad Y_n = \left(n - \frac{N+1}{2}\right)B_Y \qquad n = 1, ..., N\,. \tag{9.23}$$

Then we have

$$S^2 = S_X^2 + S_Y^2 = \frac{1}{12}((M^2-1)B_X^2 + (N^2-1)B_Y^2)\,. \tag{9.24}$$

We now assume the grid is quadratic with $B_X = B_Y$ and $T = N^2$. Then we obtain

$$S^2 = S_X^2 + S_Y^2 = \frac{1}{6}(N^2-1)B^2 = \frac{1}{6}(T-1)B^2\,. \tag{9.25}$$

Then the standard deviation is

$$\sigma_{\widehat{W}} = \frac{\sqrt{6}}{\sqrt{T(T-1)}}\frac{Z^2}{B}\frac{\sigma_{x'}}{c}\,. \tag{9.26}$$

For large T we can use the approximation

$$\boxed{\sigma_{\widehat{W}}^{(2D)} = \frac{\sqrt{6}}{T}\frac{Z^2}{B}\frac{\sigma_{x'}}{c}}\,. \tag{9.27}$$

Using the diameter

$$D = \sqrt{2}(N-1)B \tag{9.28}$$

we have the average distance squared

$$S^2 = \frac{1}{12}\frac{T-1}{\sqrt{T}-1}D^2 \tag{9.29}$$

which yields the standard deviation

$$\sigma_{\widehat{W}}^{(2D)} = \sqrt{\frac{12(\sqrt{T}-1)}{T-1}}\frac{Z^2}{D}\frac{\sigma_{x'}}{c} \tag{9.30}$$

which for large T simplifies to

$$\boxed{\sigma_{\widehat{W}}^{(2D)} = \sqrt{\frac{12}{T}}\frac{Z^2}{D}\frac{\sigma_{x'}}{c}}\,, \tag{9.31}$$

which is identical to the standard deviation if the projection centers are on a straight line.

9.5.3 Projection centers on a spherical cap

If the T projection centers are evenly distributed on a spherical cap with radius Z and angular diameter δ we use a slightly different model. We assume the uncertainty of the rays to be uniform in all directions with standard deviation σ_α, which corresponds to $\sigma_{x'}/c$ if the observed point is close to the principal point. Then the uncertainty of the ray at the observed image point is $\sigma_q = Z\sigma_\alpha$. The direction of the ray is

$$\boldsymbol{d} = \begin{bmatrix} \cos\lambda\sin\phi \\ \sin\lambda\sin\phi \\ \cos\phi \end{bmatrix} . \tag{9.32}$$

The normal equation matrix is (see PCV-A Sect. 9.5.3.2)

$$\mathsf{N} = \sum_t w_t (\mathsf{I}_3 - \boldsymbol{d}_t \boldsymbol{d}_t^\mathsf{T}) . \tag{9.33}$$

We again assume $w_t = 1/\sigma_q^2$.

We now replace the sum by an integral

$$\mathsf{N} = w_t \sum_t (\mathsf{I}_3 - \boldsymbol{d}_t \boldsymbol{d}_t^\mathsf{T}) \approx T \frac{1}{\sigma_q^2} \frac{\int_{\lambda,\phi\in C} (\mathsf{I}_3 - \boldsymbol{d}_t \boldsymbol{d}_t^\mathsf{T}) \cos\phi \, \mathrm{d}\lambda\mathrm{d}\phi}{\int_{\lambda,\phi\in C} \cos\phi \mathrm{d}\lambda\mathrm{d}\phi} . \tag{9.34}$$

For symmetry reason the normal equation matrix is diagonal:

$$N_{11} = N_{22} = \frac{1}{6\sigma_q^2}\left(4 + \cos^2\frac{\delta}{2} + \cos\frac{\delta}{2}\right) T \quad \text{and} \quad N_{33} = \frac{1}{3\sigma_q^2}\left(2 - \cos^2\frac{\delta}{2} - \cos\frac{\delta}{2}\right) T , \tag{9.35}$$

the second expression proving (9.3).

Observe for $d = 2\pi$ due to $\cos\frac{\delta}{2} = -1$ we obtain the fully isotropic configuration

$$N_{11} = N_{22} = N_{33} = \frac{2}{3\sigma_q^2} . \tag{9.36}$$

Thus the standard deviation for the ZW-coordinate is

$$\sigma_{\widehat{W}}^{(\mathrm{cap})} = \frac{\sqrt{3}}{\sqrt{T}} \frac{Z}{2 - \cos\frac{\delta}{2} - \cos^2\frac{\delta}{2}} \sigma_\alpha . \tag{9.37}$$

For small δ we obtain the approximation

$$\sigma_{\widehat{W}}^{(\mathrm{cap})} = \frac{\sqrt{8}}{\sqrt{T}} \frac{Z}{\delta} \sigma_\alpha . \tag{9.38}$$

Taking into account that then $\delta = D/Z$ and $\sigma_\alpha = \sigma_{x'}/c$ we obtain

$$\sigma_{\widehat{W}}^{(\mathrm{cap})} = \frac{\sqrt{8}}{\sqrt{T}} \frac{Z^2}{D} \frac{\sigma_{x'}}{c} . \tag{9.39}$$

The difference of the constants ($\sqrt{12}$ versus $\sqrt{8}$) result from the different roundness of the two figures (square versus circle).

10 Multi-View Triangulation with Directions

We provide simple solution to the optimal triangulation of a scene point from multiple views assuming isotropic uncertainty of the directions. As a special case we provide a simple expression for the distance of the triangulated point in case of homogeneous directional uncertainty and small basis, expressed as a function of the effective base line, the viewing angle and the resolution of an omnidirectional camera and the matching accuracy in pixels.

10.1 Preface 117
10.2 The Problem 117
10.3 The approximate Solution 118
10.4 The Solution with Different Uncertainties of the Distances 119
10.5 The Solution for Directional Observations with Different Uncertainty 120
10.6 Assuming Correlations between the Directions due to Least Squares Matching 120
10.6.1 The 2D Model 120
10.6.2 The 3D Model 121
10.7 Uncertainty of binocular triangulation with omnidirectional cameras 122

10.1 Preface

This note from 2007, and extended 2023, provides a simple solution to the optimal triangulation of a scene point from multiple views. It also provides a simple expression for the distance of the triangulated point in case of homogeneous directional uncertainty and small basis, expressed as a function of the effective base line, the viewing angle and the resolution of an omnidirectional camera and the matching accuracy in pixels.

10.2 The Problem

Given are K projection matrices $\mathsf{P}_k, k = 1, ..., K$ and corresponding image points $\boldsymbol{x}_k, k = 1, ..., K$. Triangulate a good 3D-point.k The idea is the following: The projection matrices together with the image point determine N projection rays, see Fig. 10.1. The optimal point $\boldsymbol{X}$ is the one closest to all these rays, where the notion distance needs to be specified and leads to different solutions.

We extend the approximate solution in three ways:

1. We handle the case where the distances are weighted individually.
2. We handle the case of isotropic and homogeneous uncertainty of the directions.
3. We handle the case of homogeneous mutual correlations between the directions.

In all cases we provide a rigorous solution.

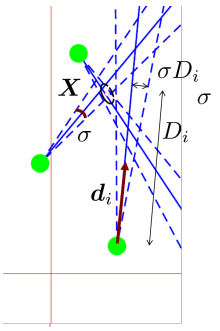

Figure 10.1: Optimal multi-view triangulation for directions. The problem is nonlinear in general, since the effect of directional uncertainties onto the 3D point depends on the unknown distances of the point to the given projection centers

10.3 The approximate Solution

The first solution just minimizes the sum of the squares of the distances of the rays to the 3D point.

The projection centers are

$$\boldsymbol{Z}_k = -\mathsf{H}_{n\infty}^{-1}\mathbf{h}_k \tag{10.1}$$

with

$$\mathsf{P}_k = [\mathsf{H}_{n\infty}|\mathbf{h}_k] \tag{10.2}$$

The projection lines have normalized direction

$$\mathbf{d}_k = \mathsf{N}\left(\mathsf{H}_{n\infty}^{-1}\mathbf{x}_k\right) \tag{10.3}$$

The 3D projection lines have Plücker coordinates

$$\mathbf{L}_k = \begin{bmatrix} \boldsymbol{L}_h \\ \boldsymbol{L}_0 \end{bmatrix}_k = \begin{bmatrix} \mathbf{d}_k \\ \boldsymbol{Z}_k \times \mathbf{d}_k \end{bmatrix} \tag{10.4}$$

The squared distances of the unknown point $\boldsymbol{X}$ to the lines are

$$\begin{aligned} d_{XL_k}^2 &= |\boldsymbol{L}_{0i} + \mathcal{S}(\boldsymbol{L}_{hi})\boldsymbol{X}|^2 && (10.5)\\ &= (\boldsymbol{Z}_k \times \mathbf{d}_k + \mathcal{S}(\mathbf{d}_k)\boldsymbol{X})^\mathsf{T}(\boldsymbol{Z}_k \times \mathbf{d}_k + \mathcal{S}(\mathbf{d}_k)\boldsymbol{X}) && (10.6)\\ &= |\boldsymbol{Z}_k \times \mathbf{d}_k|^2 + 2(\boldsymbol{Z}_k \times \mathbf{d}_k)^\mathsf{T}\mathcal{S}(\mathbf{d}_k)\boldsymbol{X} + \boldsymbol{X}^\mathsf{T}\mathcal{S}(\mathbf{d}_k)^\mathsf{T}\mathcal{S}(\mathbf{d}_k)\boldsymbol{X} && (10.7) \end{aligned}$$

The sum of the squared distances therefore is

$$\begin{aligned} \Omega &= \sum_k d_{XL_k}^2 && (10.8)\\ &= \sum_k |\boldsymbol{Z}_k \times \mathbf{d}_k|^2 + 2\sum_k (\boldsymbol{Z}_k \times \mathbf{d}_k)^\mathsf{T}\mathcal{S}(\mathbf{d}_k)\boldsymbol{X} + \boldsymbol{X}^\mathsf{T}\sum_k \mathcal{S}(\mathbf{d}_k)^\mathsf{T}\mathcal{S}(\mathbf{d}_k)\boldsymbol{X} && (10.9) \end{aligned}$$

The necessary condition for the minimum is

$$\frac{1}{2}\frac{\partial \Omega}{\partial \boldsymbol{X}} = \sum_k \mathcal{S}(\mathbf{d}_k)^\mathsf{T}(\boldsymbol{Z}_k \times \mathbf{d}_k) + \sum_k \mathcal{S}(\mathbf{d}_k)^\mathsf{T}\mathcal{S}(\mathbf{d}_k)\boldsymbol{X} = \mathbf{0} \tag{10.10}$$

Thus, the optimal point is given by

$$\widehat{\boldsymbol{X}} = \left(\sum_k \mathcal{S}(\mathbf{d}_k)^\mathsf{T}\mathcal{S}(\mathbf{d}_k)\right)^{-1} \sum_k \mathcal{S}(\mathbf{d}_k)^\mathsf{T}\mathcal{S}(\mathbf{d}_k)\boldsymbol{Z}_k \tag{10.11}$$

or

$$\boxed{\widehat{\boldsymbol{X}} = \left(\sum_k W_k\right)^{-1} \sum_k W_k \boldsymbol{Z}_k} \tag{10.12}$$

with

$$W_k = I_3 - \mathbf{d}_k \mathbf{d}_k^\mathsf{T} \tag{10.13}$$

in case $\mathbf{d}_k$ is normalized. Obviously, this is a weighted mean of the projection centers $\boldsymbol{Z}_k$ where the weight matrix is 0 in the direction of $\mathbf{d}_k$ and 1 otherwise. Thus, W_k is representing a cylindrical covariance matrix, with infinite uncertainty in the direction of the projection lines and standard deviation 1 perpendicular to the viewing direction.

The estimated variance of the distances of the fitted points to the projection lines can be obtained from

$$\boxed{\hat{\sigma}_d^2 = \frac{\Omega}{2I-3} \quad \text{with} \quad \Omega = \sum_k d^2(\mathcal{X}, \mathcal{L}_k) = \sum_k \left(\frac{\mathbf{d}_k^\mathsf{T}(\widehat{\boldsymbol{X}} - \boldsymbol{Z}_k)}{|\widehat{\boldsymbol{X}} - \boldsymbol{Z}_k|}\right)^2 .} \tag{10.14}$$

The theoretical covariance matrix of the estimated points is

$$\boxed{\Sigma_{\widehat{X}\widehat{X}} = \sigma_d^2 \left(\textstyle\sum_k W_k\right)^{-1}} \tag{10.15}$$

with some prior assumption about the standard deviation σ_d of the distances.

10.4 The Solution with Different Uncertainties of the Distances

Instead of (10.8) we optimize

$$\Omega = \sum_k \frac{d_{XL_k}^2}{\sigma_{d_k}^2} , \tag{10.16}$$

where the standard deviations of the distances are σ_{d_k}. We obtain the same solution (10.12) however instead of the weight-matrices in (10.13) we use

$$W_k = \frac{1}{\sigma_{d_k}^2} \left(I_3 - \mathbf{d}_k \mathbf{d}_k^\mathsf{T}\right) , \tag{10.17}$$

see PCV Eq. (10.174).

If the solution (10.12) is written with the normal equation matrix and the right-hand sides

$$N = \sum_k W_k \quad \text{and} \quad \boldsymbol{n} = \sum_k W_k \boldsymbol{Z}_k \tag{10.18}$$

(using the weights $eq : W - WLS$, assuming $\sigma_0 = 1$) we have the theoretical covariance matrix

$$\Sigma_{\widehat{X}\widehat{X}} = \sigma_0 N^{-1} . \tag{10.19}$$

Similarly, we obtain an estimate for the variance factor

$$\hat{\sigma}_0^2 = \frac{\Omega}{2I-3} \quad \text{with} \quad \Omega = \sum_k \left(\frac{\mathbf{d}_k^\mathsf{T}(\widehat{\boldsymbol{X}} - \boldsymbol{Z}_k)}{|\widehat{\boldsymbol{X}} - \boldsymbol{Z}_k|}\right)^2 . \tag{10.20}$$

10.5 The Solution for Directional Observations with Different Uncertainty

In case directions δ_k are observed, the uncertainty of the distances d_k of the unknown point to the given rays depend on the distances s_k of the point $\boldsymbol{x}$ to the projection centers $\boldsymbol{Z}_k$:

$$\sigma_{d_k} = s_k \sigma_{\delta_k} \quad \text{with} \quad s_k = |\boldsymbol{X} - \boldsymbol{Z}_k| \tag{10.21}$$

We cannot optimize (10.16) since the distances s_k depend on the unknown point.

However, see Fig. 10.2, we can iteratively update $\mathcal{X}$ by using (10.21) after an initialization with $s_k = 1$ in the first iteration. For not too large directional errors, say below 0.01 [rad] or 1 °, only a second iteration is necessary. This procedure can replace Algorithm 21

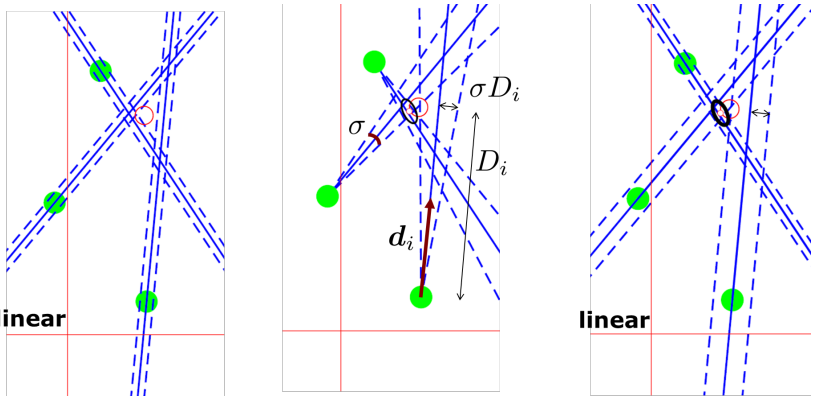

Figure 10.2: Optimal triangulation with isotropic directional uncertainties.

in PCV, in case it is clear that the 3D point is at finity and the rays do not diverge, or if some sufficiently good approximate value for $\boldsymbol{X}$ is known.

10.6 Assuming Correlations between the Directions due to Least Squares Matching

10.6.1 The 2D Model

We assume the position of the keypoint in one image is determined by some keypoint detector and the coordinate differences, i.e., parallaxes, to the other images are determined by least squares matching, like the Kanade-Lucas-Tracker. The reason simply is: the coordinates $\boldsymbol{x}_1$ of the detection usually is less accurate, say with standard deviation σ_x whereas the determination of the parallaxes $\boldsymbol{p}_k = \boldsymbol{x}_k - \boldsymbol{x}_0, k = 2, ..., K$ is highly accurate, say σ_{p_0}. Assuming a homogeneous configuration and enforcing the mean coordinate, derived from the parallaxes is $\mathbf{0}$ the uncertainty of the final coordinates $\boldsymbol{x}_k$ can be derived from

$$\underline{\boldsymbol{x}} = \begin{bmatrix} \underline{\boldsymbol{x}}_1 \\ \dots \\ \underline{\boldsymbol{x}}_k \\ \dots \\ \underline{\boldsymbol{x}}_K \end{bmatrix} = \begin{bmatrix} \mathbb{E}(\underline{\boldsymbol{x}}_1) \\ \dots \\ \mathbb{E}(\underline{\boldsymbol{x}}_k) \\ \dots \\ \mathbb{E}(\underline{\boldsymbol{x}}_K) \end{bmatrix} + \begin{bmatrix} \underline{\Delta \boldsymbol{x}_0} \\ \dots \\ \underline{\Delta \boldsymbol{x}_0} \\ \dots \\ \underline{\Delta \boldsymbol{x}_0} \end{bmatrix} + \begin{bmatrix} \underline{\Delta \boldsymbol{p}_1} \\ \dots \\ \underline{\Delta \boldsymbol{p}_k} \\ \dots \\ \underline{\Delta \boldsymbol{p}_K} \end{bmatrix} \tag{10.22}$$

with the covariance matrices for the detection $\underline{\Delta \boldsymbol{x}_0}$ and the parallaxes $\underline{\Delta \boldsymbol{p}} = [\underline{\Delta \boldsymbol{p}_k}]$:

$$\mathbb{D}(\underline{\Delta \boldsymbol{x}_0}) = \Sigma_{x_0 x_0} \quad \text{and} \quad \mathbb{D}(\underline{\Delta \boldsymbol{p}}) = (I_K - J_K/K) \otimes \Sigma_{pp} \quad \text{with} \quad J = \mathbf{1}_K \mathbf{1}_K^\mathsf{T} \tag{10.23}$$

see Förstner (1998). In the isotropic case we have

$$\Sigma_{x_0 x_0} = \sigma_{x_0}^2 I_2 \quad \text{and} \quad \Sigma_{pp} = \sigma_p^2 I_2 \,. \tag{10.24}$$

This yields the following covariance matrix for the K points

$$\Sigma_{xx} = \left(\sigma_{x_0}^2 \, J_K + \sigma_p^2 \, (I_K - J_K/K)\right) \otimes I_2 \tag{10.25}$$

We have the extreme case where the parallaxes are perfect: $\sigma_p = 0$:

$$\Sigma_{xx} = \sigma_{x_0}^2 \, \mathbf{1}\mathbf{1}^\mathsf{T} \otimes I_2 \tag{10.26}$$

Then all points are 100% correlated.

10.6.2 The 3D Model

We now want to extend the model to observed directions, namely assuming they are correlated. This extension is non-trivial, why we provide an approximate solution.

The reason is that the basic model (10.22) implicitely assumes the projection centers are coplanar, the viewing directions are parallel, the scene is fronto-parallel, and the image coordinates refer to a perspective model. Then a surface patch is mapped to identical image patches, allowing to use the result of Förstner (1998). As soon as the surface element is observed from different directions, this model does not hold anymore. This not only holds for tilted cameras but also for spherical cameras, where the addition in (10.22) cannot be easily replaced.

We therefore exploit the result of Förstner (1998) by modelling the situation is two steps:

1. In the first step, we assume the surface patch is seen along its normal, however, allowing the distance of the projection centers may vary. Then the setup of a simultaneous homogeneous least squares matching is possible. The resulting accuracies refer to the image coordinates ($\Delta \boldsymbol{p}_k$) refer to the scene, and, using the distances s_k to the projection centers can be transformed into individual directional uncertainties, which, due to the isotropy assumption, lead to isotropic directional uncertainties.
2. In the second step, we assume the directional accuracy approximately transfers to directions not being parallel to the normal. This is a valuable approximation if the deviation from the normal is not too large, since the deviation increases with $1/\cos(\alpha_k)$, where α_k is the angle between the observed direction and the normal of the surface patch. Neglecting this factor simulates the situation where the scene is assumed to consist of small spheres, whose relative direction is determined by least squares matching, which is an unlikely but not invalid assumption.

10.6.2.1 Observed Directions parallel to the Normal of a Surface Patch

The result of the previous subsection can directly be used for expressing the lateral uncertainty of the spatial deviations across the direction. Using (10.21) we find the directional uncertainty from

$$\sigma_{\delta_k} = \frac{\sigma_{d_k}}{s_k} \tag{10.27}$$

where the standard deviations σ_{d_k} correspond to the σ_{x_k} in the left bracket of (10.25). Hence we assume the directional errors $\underline{\boldsymbol{d}} = [\underline{\mathbf{d}}_k]$]are isotropic with

$$\Sigma_{dd} = \sigma_{x_0}^2 \, J_K + \sigma_p^2 \, (I_K - J_K/K) \tag{10.28}$$

Since we need the factors

$$w_{d_k=} = \frac{1}{\sigma_{d_k}^2} \tag{10.29}$$

in the weight matrices, which now are not independent we use the weight matrix

$$W_{dd} = \Sigma_{dd}^{-1} = [w_{kk'}] = \frac{1}{\sigma_p^2} I_K - \frac{K\sigma_{x_0}^2 - \sigma_p^2}{K^2 \sigma_p^2 \sigma_{x_0}^2} J_K \tag{10.30}$$

Since the individual weight matrix (10.17) for one direction can be written as

$$\mathsf{W}_k = \mathsf{S}(\mathbf{d}_k)\ (w_{d_k}\mathsf{I}_3)\mathsf{S}^\mathsf{T}(\mathbf{d}_k) \tag{10.31}$$

we obtain the full weight matrix as

$$\mathsf{W} = [\mathsf{W}_{kk'}] = \mathrm{Diag}(\mathsf{S}(\mathbf{d}_k))\,[w_{kk'}\mathsf{I}_3]\,\mathrm{Diag}^\mathsf{T}(\mathsf{S}(\mathbf{d}_k)) \tag{10.32}$$

or more explicit

$$\mathsf{W}_{kk'} = w_{kk'}\mathsf{S}(\mathbf{d}_k)\mathsf{S}(\mathbf{d}_{k'}) \tag{10.33}$$

Therefore, the solution for the 3D point reads as

$$\widehat{\boldsymbol{X}} = \mathsf{N}^{-1}\boldsymbol{n} \quad \text{with} \quad \mathsf{N} = [N_{ij}] = \sum_{k,k'} \mathsf{W}_{kk'} \quad \text{and} \quad \boldsymbol{n} = [n_j] = \sum_{k,k'} \mathsf{W}_{kk'}\boldsymbol{Z}_{k'}\,. \tag{10.34}$$

10.7 Uncertainty of binocular triangulation with omnidirectional cameras

Given is the configuration

- Distance D
- Basis B
- Angular range α
- Effective image diameter/width W
- Matching accuracy σ_δ referring to the direction

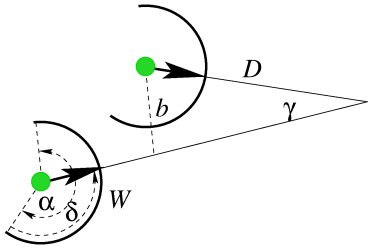

Figure 10.3: Configuration

Then we have for small δ

- The parallactic angle

$$\gamma = \frac{b}{D} \quad \text{or} \quad D\gamma = b \quad \text{and} \quad \mathrm{d}D\,\gamma + D\,\mathrm{d}\gamma = 0 \quad \text{and} \quad \frac{\sigma_\gamma}{\gamma} = \frac{\sigma_D}{D} \tag{10.35}$$

  thus

$$\sigma_D = \frac{D}{\gamma}\sigma_\gamma = \frac{D^2}{b}\sigma_\gamma \tag{10.36}$$

- The pixel size corresponding to direction elements $\Delta\delta$ in [rad] is

$$\Delta\delta = \frac{\alpha}{W} \tag{10.37}$$

  assuming a pixel distance corresponds to the same directional difference, which is an approximation.

- the uncertainty of the measured parallactic angle, as difference of two directions

$$\sigma_\gamma = \sqrt{2}\,\sigma_\delta \tag{10.38}$$

Hence, we finally have the distance accuracy

$$\sigma_D = \sqrt{2}\frac{D}{\gamma}\sigma_\delta \tag{10.39}$$

Since we usually describe the matching accuracy, i.e., the accuracy σ_p of the parallax in pixels, we need to take the resolution into account. Then we have

$$\sigma_\delta = \frac{\alpha}{W}\,\frac{\sigma_p}{\sqrt{2}} \tag{10.40}$$

Then we obtain for the distance

$$\boxed{\sigma_D = \frac{D}{\gamma}\,\frac{\alpha}{W}\,\sigma_p = \frac{D^2}{b}\,\frac{\alpha}{W}\,\sigma_p} \tag{10.41}$$

If we refer to the inverse depth

$$s = \frac{1}{D} \quad \text{with} \quad sD = 1 \quad \text{and} \quad \mathrm{d}s\,D + s\,\mathrm{d}D = 0 \quad \text{and} \quad \frac{\sigma_s}{s} = \frac{\sigma_D}{D} \tag{10.42}$$

we obtain

$$\sigma_s = \frac{s}{D}\sigma_D = \frac{s}{D}\,\frac{D^2}{b}\,\frac{\alpha}{W}\,\sigma_p = \frac{1}{b}\,\frac{\alpha}{W}\,\sigma_p \tag{10.43}$$

from which we may derive the matching accuracy

$$\sigma_p = b\,\frac{W}{\alpha}\,\sigma_s \tag{10.44}$$

if we know the camera, i.e., the viewing angle α, how its image is used (possibly reduced in resolution), i.e., the diameter of the image in pixels, and how large the effective baseline b is.

Bibliography

The numbers at the end of each reference are the pages where it is cited.

Boyd, S. and L. Vandenberghe (2004). *Convex optimization*. Cambridge University Press. 6, 9

Eade, E. (2014). Lie Groups for 2D and 3D Transformations. http://ethaneade.com/lie_groups.pdf, last visited 2.6.2016. 52

Förstner, W. (1982). Systematic errors in photogrammetric point determination. In *Proc. Survey Control Networks, International Federationof Surveyors (FIG), Meeting Study Group 5B*, Denmark, pp. 197–209. 26

Förstner, W. (1998). On the Theoretical Accuracy of Multi Image Matching, Restoration and Triangulation. In *Festschrift Univ. Prof. Dr.-Ing.-Dr, h. c. mult. Gottfried Konecny zur Emeritierung*, Number 227, pp. 105–118. Wiss. Arbeiten der Fachr. Vermessungswesen der Univ. Hannover. 120, 121

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*. 52

Förstner, W. and B. P. Wrobel (2016). *Photogrammetric Computer Vision – Statistics, Geometry, Orientation and Reconstruction*. Springer. 8, 51, 65, 71, 90, 92, 112

Koch, K.-R. (1999). *Parameter Estimation and Hypothesis Testing in Linear Models* (2nd ed.). Springer. 26

Leonardos, S., C. Allen-Blanchette, and J. Gallier (2015). The exponential map for the group of similarity transformations and applications to motion interpolation. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 377–382. 59

Pepić, S. H. (2010). Weighted Moore-Penrose Inverse: PHP vs. Mathematica. *Ser. Math. Infrom. 25*, 35–45. 13

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. 22, 23, 24, 42

Schneider, J. and W. Förstner (2013). Bundle adjustment and system calibration with points at infinity for omnidirectional camera systems. *Zeitschrift für Photogrammetrie, Fernerkundung und Geoinformation 4*, 309–321. 68

Solà, J., J. Deray, and D. Atchuthan (2018). A micro Lie theory for state estimation in robotics. *CoRR abs/1812.01537*. 50

Takahashi, K., J. Fagan, and M.-S. Chen (1973). Formation of a sparse bus impedance matrix and its application to short circuit study. In *IEEE Power Engineering Society*, Volume 7. 29

Vanhatalo, J. and A. Vehtari (2008). Modelling local and global phenomena with sparse Gaussian processes. In *Conference on Uncertainty in Artificial Intelligence*, pp. 571–578. 29