Chapter 4 — Examples

Wolfgang Förstner

July 11, 2017

Contents

1 Summary

We give three examples for parameter estimation and evaluation

1. Gauss–Markov model for linear regression with two unknowns with evaluation.
2. Gauss–Markov model for linear regression for similarity transformation with evaluation.
3. Gauss–Helmert model for linear regression for similarity transformation with evaluation.

We provide the specific models and equations for the examples.

The discussion includes general hints how to use the evaluation methods in other applications and how to report evaluation results in publications.

Partly we refer to problems addressed in the book. The results with concrete numbers derived with the MATLAB-code, are given in example boxes. The background color green indicates recommended procedures, red indicates pitfalls, other text has light grey background.

For the theory we refer to Chap. 4. References to sections are given as 'PCV-NUMBER', e.g., PCV-4.2, references to equation as 'PCV-(NUMBER)', e.g., PCV-(4.138). For the software we refer to the home page of the book http://www.ipb.uni-bonn.de/book-pcv/.

2 GMM for linear regression with two unknowns with evaluation

The scope of this example is to demonstrate the estimation and the evaluation in the linear Gauss-Markov model. The MATLAB-script file is `GMM/DEMOS-GMM/demos_GMM_regression.m` under http://www.ipb.uni-bonn.de/book-pcv/#cod.

2.1 The Model and the Estimates

The observations $l_n, n = 1, ..., N$ depend linearly on the time t. The intercept x_1 and the slope x_2 are unknown, see Fig. 1.

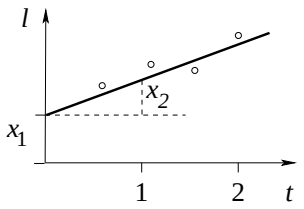


Figure 1: Model for Example 1: linear regression

We assume all observations to be uncorrelated and have the same standard deviation $\sigma := \sigma_{l_n}$:

$$\mathbb{E}(\underline{l}_n) = x_1 + x_2 t_n\,, \qquad \mathbb{D}(\underline{l}_n) = \sigma^2\,. \tag{1}$$

Collecting the observations, parameters and coefficients in the corresponding vectors and matrices, namely

$$\boldsymbol{l} = \begin{bmatrix} l_1 \\ \cdots \\ l_n \\ \cdots \\ l_N \end{bmatrix}\,, \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{2}$$

and the partitioned design matrix

$$\underset{N\times 2}{A} = \begin{bmatrix} 1 & t_1 \\ \cdots & \\ 1 & t_n \\ \cdots & \\ 1 & t_N \end{bmatrix} = [\mathbf{1}\ \boldsymbol{t}] := [C, D] \quad \text{with} \quad C = \mathbf{1} \quad \text{and} \quad D = \boldsymbol{t}\,. \tag{3}$$

the model reads as

$$\mathbb{E}(\underline{\boldsymbol{l}}) = A\boldsymbol{x}\,, \qquad \mathbb{D}(\underline{\boldsymbol{l}}) = \sigma^2 I_N = w^{-1} I_N\,, \tag{4}$$

see PCV-(4.34). This is the same model we use for the fitting line in Chap. 10.5 for deriving the uncertainty of the 2D line through given points. The normal equation matrix can be given explicitly

$$\mathsf{N} = \begin{bmatrix} Nw & w\sum_{n=1}^{N} x_n \\ w\sum_{n=1}^{N} x_n & w\sum_{n=1}^{N} x_n^2 \end{bmatrix}, \tag{5}$$

see PCV-(4.40). The estimated residuals are

$$\widehat{\boldsymbol{v}} = (\widehat{x}_1 + \widehat{x}_2 t_i) - \boldsymbol{l}\,, \tag{6}$$

see PCV-(4.41). A numerical example is given in the box on page 4. The estimated sum of the squared residuals and the estimated variance factor are

$$\Omega = w\sum_{n=1}^{N} v_n^2 \quad \text{and} \quad \widehat{\sigma}_0^2 = \frac{\Omega}{N-2}\,, \tag{7}$$

see PCV-(4.81) and PCV-(4.80).

The theoretical and the empirical covariance matrices are

$$\Sigma_{\widehat{x}\widehat{x}} = \sigma_0^2 \mathsf{N}^{-1} \quad \text{and} \quad \widehat{\Sigma}_{\widehat{x}\widehat{x}} = \widehat{\sigma}_0^2 \mathsf{N}^{-1}\,. \tag{8}$$

The covariance matrix of the residuals is

$$\Sigma_{vv} = \Sigma_{ll} - \mathsf{A}\Sigma_{\widehat{x}\widehat{x}}\mathsf{A}^{\mathsf{T}}\,. \tag{9}$$

For getting insight into the structure of the result, we reduce the times t_i to their *centroid*

$$\bar{t}_n = t_n - \mu_t \quad \text{with} \quad \mu_t = \frac{\sum_{n=1}^{N} t_n}{N} \tag{10}$$

and obtain the design matrix for the centred model

$$\mathsf{A}^{(c)} = [1 \;\; \bar{t}_n]_{n=1,\ldots,N}\,. \tag{11}$$

Hence the new mode reads as

$$\mathbb{E}(\underline{l}_n) = x_1^{(c)} + x_2 t_i^{(c)}\,, \tag{12}$$

where the intercept refers to the abscissa at centroid μ_t. A numerical example is given in the following box. The the covariance matrix of the unknown parameters then is diagonal

$$\Sigma_{\widehat{x}\widehat{x}} = \frac{\sigma_0^2}{wN\sum_{n=1}^{N})\bar{t}_n^2} \begin{bmatrix} \sum_{n=1}^{N} \bar{t}_n^2 & 0 \\ 0 & N \end{bmatrix} = \frac{\sigma_0^2}{w} \begin{bmatrix} \frac{1}{N} & 0 \\ 0 & \frac{1}{\sum_{n=1}^{N} \bar{t}_n^2} \end{bmatrix} \tag{13}$$

Hence the standard deviation of the estimated intercept in the centred model, which is at the centroid, is

$$\sigma_{\widehat{\bar{x}}_1} = \frac{\sigma}{\sqrt{N}}\,, \tag{14}$$

which decreases with increasing number N of observations.

Example 2.1: ***Linear regression with two parameters (1).*** Figure 2 shows the result of an example generated with `demos_GMM_regression.m` and initialization of the random number generator with `init_rand=15`. We will refer to this numerical example in the following.
The true values are given by:

$$\tilde{\boldsymbol{x}} = \begin{bmatrix} 0.5 \\ 1.0 \end{bmatrix}, \quad \sigma = 0.5\,, \quad \boldsymbol{t} = \begin{bmatrix} -1 \\ 1 \\ 2 \\ 14 \end{bmatrix}, \quad \tilde{\boldsymbol{l}} = \begin{bmatrix} -0.5 \\ 1.5 \\ 2.5 \\ 14.5 \end{bmatrix}, \quad \tilde{\boldsymbol{e}} = \begin{bmatrix} -0.6271 \\ -0.3932 \\ -0.5604 \\ 3.5861 \end{bmatrix}, \quad \boldsymbol{l} = \begin{bmatrix} 0.1271 \\ 1.8932 \\ 3.0604 \\ 10.9139 \end{bmatrix}. \tag{15}$$

The true errors $\tilde{\boldsymbol{e}}$ result from sampling from $\mathcal{N}(0, \sigma^2)$ with $\sigma = 0.5$. We also introduce an outlier in order to demonstrate the difficulty to identify outliers. Observation l_4 is changed by the error $\nabla l_4 = -4$; this error is 8 times the standard deviation of the assumed observational noise. The estimated parameters and estimated residuals

$$\widehat{\boldsymbol{x}} = \begin{bmatrix} 1.1956 \\ 0.7008 \end{bmatrix}, \quad \widehat{\boldsymbol{v}} = \begin{bmatrix} 0.3678 \\ 0.0032 \\ -0.4633 \\ 0.0922 \end{bmatrix}. \tag{16}$$

Observe, this result can be obtained by just assuming the observations have the same standard deviation, though this needs not be known.

If we would have ground truth, i.e., the true values for the parameters, we could report the differences

$$\widehat{\boldsymbol{x}} - \tilde{\boldsymbol{x}} = \begin{bmatrix} 0.1956 \\ 0.2008 \end{bmatrix}. \tag{17}$$

Without knowing anything about the observational process, i.e., the structure of the problem and the level of the observational noise, this difference cannot be evaluated. Moreover, if we – as a reader of such a result – would have a different experimental setup, using the same functional model (here a linear regression with two parameters), e.g., more observations, possibly distributed differently, then we would not be able to predict the performance in our situation. This indicates, that even if we give the differences $\widehat{\boldsymbol{x}} - \tilde{\boldsymbol{x}}$ of the estimates to some ground truth, the reader does not learn something from this difference, if not provided with more information; this will be discussed below.

◇

The standard deviation of the slope is

$$\sigma_{\widehat{x}_2} = \frac{\sigma}{\sum_{n=1}^{N} \bar{t}_n^2}. \tag{18}$$

With the root mean square distance of the observed times from their centroid

$$\mathrm{RMSE}_t := \frac{1}{N} \sqrt{\sum_{n=1}^{N} \bar{t}_n^2}. \tag{19}$$

We hence have the standard deviation of the estimated slope

$$\sigma_{\widehat{x}_2} = \frac{\sigma}{\sqrt{N}} \frac{1}{\mathrm{RMSE}_t}. \tag{20}$$

A numerical example is given in box on page 7.

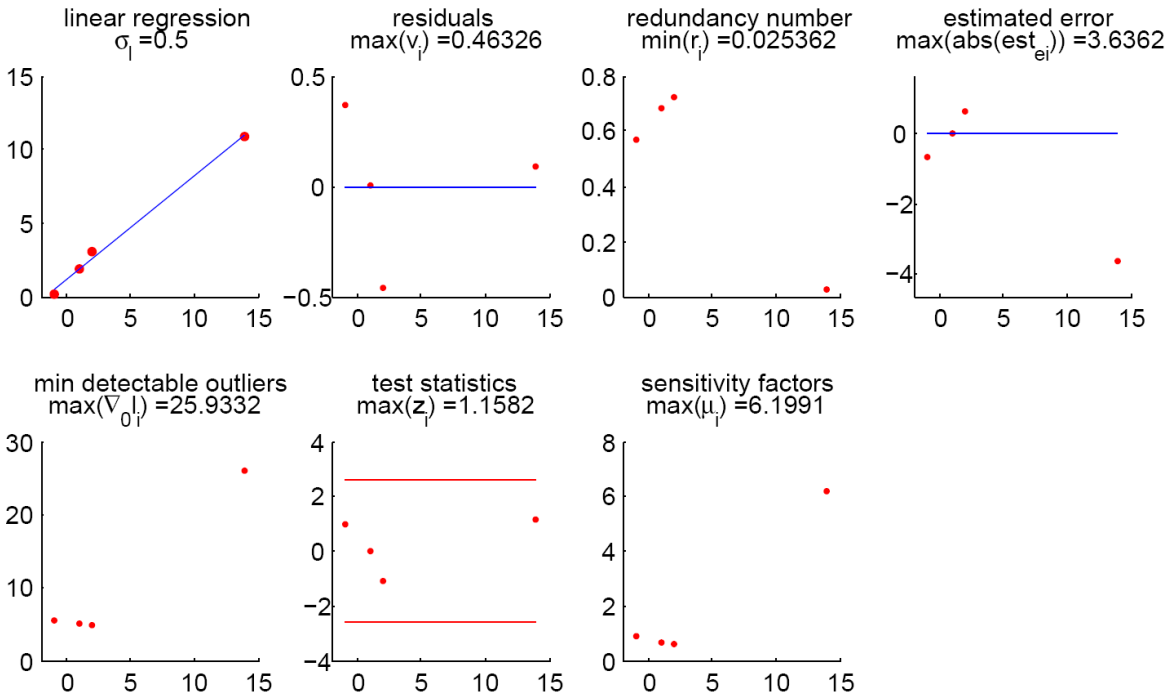


Figure 2: Result of regression. From top left to bottom right: (1) original data l_n; (2) estimated residuals $\widehat{v}_n$, (3) redundancy numbers r_n; (4) estimated errors ∇l_n; (5) minimal detectable outliers $\nabla_0 l_n$; (6) test statistics z_n; (7) sensitivity factors μ_n. Explanation see text.

2.2 The Estimation

The estimation is realized in the MATLABfunction `GaussMarkovModelLinear.m`. It in a first step follows Alg. 1, PCV-p.91. An additional routine `diagnostics_1d.m` performs the sensitivity analysis. Given a set r_U of parameters of interest it determines all diagnostic parameters of interest:

- the covariance matrix of the estimated parameters

$$\Sigma_{\widehat{x}\widehat{x}} = (A^{\mathsf{T}} W_{ll} A)^{-1} \tag{21}$$

  assuming the a priori variance factor is $\sigma_0^2 = 1$. It allows to derive the standard deviations of the estimated parameters $\sigma_{\widehat{x}_u} = \sqrt{\Sigma_{\widehat{x}_u \widehat{x}_u}}$ from the diagonal elements of the covariance matrix. It does not depend on real observations, but only on the mathematical model of the design, i.e., the geometric configuration and the assumed uncertainty of the observations.

- the residuals $\widehat{v}_i$ and the estimated varaince factor

$$\widehat{v}_n = \boldsymbol{a}_n^{\mathsf{T}} \widehat{\boldsymbol{x}} + a_n - l_n \quad \text{and} \quad \widehat{\sigma}_0^2 = \frac{\widehat{\boldsymbol{v}}^{\mathsf{T}} W_{ll} \widehat{\boldsymbol{v}}}{N - U}\,. \tag{22}$$

- the test statistics z_n (PCV-(4.284))

$$z_n = \frac{-\widehat{v}_n}{\sigma_{\widehat{v}_n}}\,. \tag{23}$$

- the redundancy numbers as diagonal elements of the redundancy matrix (PCV-(4.61))

$$r_n = R_{nn}\,. \tag{24}$$

- the minimum size $\nabla_0 l_i$ of detectable outliers (PCV-(4.304), (4.300))

$$\nabla_0 \boldsymbol{l}_i = \delta_0 \frac{\sigma_{l_i}}{\sqrt{r_i}}\,. \tag{25}$$

  We use $\delta_0 = 4.13$, see PCV-p.67, Table 3.2.

- the sensitivity factor w.r.t. all 4 parameters (PCV-(4.292))

$$\mu_{x,n} = \sqrt{\frac{1-r_n}{r_n}}\,; \tag{26}$$

- the sensitivity factor w.r.t. the selected set $\boldsymbol{\tau}_U$ of parameters (in PCV-(4.296) referred to as parameter set $\boldsymbol{k}$)

$$\mu_{x_1,n} = \sqrt{\frac{\overline{u}_n}{r_n}}\,. \tag{27}$$

2.3 Evaluating the Precision of the Estimates

2.3.1 Simulations vs. Theoretical Derivations

We have three methods to derive the theoretical precision of estimates, which in a first step are equivalent:

1. Using the Cramer-Rao bound based on the numerical determination of $\Sigma_{\widehat{x}\widehat{x}}$ for well selected cases, see PCV-(4.49). This requires only one simulation and estimation for each configuration.

2. Using the Cramer-Rao bound based on an algebraic derivation. In this case we derive algebraic expressions for the design matrix, the normal equation matrix and its inverse, as in PCV-13.3.6.1 done for the relative orientation of the image pair. This gives direct insight into the dependencies of the standard deviations of the parameters of the configuration.

3. Using sampling techniques as described in PCV-4.6.8.2: For each configuration (choice of the functional and mathematical model) this requires $K > 25$ samples and therefore estimates for obtaining an accuracy of better than 5%.

Depending on the complexity of the problem, we can choose between them.

Example 2.2: ***Linear regression with two parameters (2).*** The following information presumes, that some a priori standard deviation σ of the observations is known, i.e., the user of the estimation software knows how accurate the observations are.
All residuals are below σ, hence the result appears fine, though we know there is an outlier in the 4-th observation.

The estimated variance factor is

$$\widehat{\sigma}_0 = \frac{\sum_{n=1}^{N} \widehat{v}_n^2/\sigma_{l_n^2}}{N-U} = 0.8467\,. \tag{28}$$

For determining $\widehat{\sigma}_0$ we exploit the assumption, that the observations are mutually uncorrelated. The estimated variance factor $\widehat{\sigma}_0^2$ is not significantly deviating from the a priori value $\sigma_0^2 = 1$. However, since the redundancy $R = N - U = 4 - 2 = 2$ is very low, this value is very uncertain, see the discussion on the estimated variance factor in PCV-4.2.3.

The theoretical covariance matrix and the theoretical standard deviations of the parameters are

$$\Sigma_{\widehat{x}\widehat{x}} = \begin{bmatrix} +0.0915 & -0.0072 \\ -0.0072 & +0.0018 \end{bmatrix}, \quad \sigma_{\widehat{x}_1} = 0.3025\,, \quad \sigma_{\widehat{x}_2} = 0.0426\,. \tag{29}$$

This is the Cramer-Rao bound, the lower bound for the achievable precision in this experiment, i.e., for this design, the distribution $\boldsymbol{t}$ of the observations, the assumed model and the assumed noise level σ.

The evaluation of the deviations of the estimates $\widehat{\boldsymbol{x}} = (1.1956, 0.7008)$ from the ground truth $\tilde{\boldsymbol{x}} = (0.5, 1.0)$ now can be related to the theoretical covariance matrix, which depends on both, the design of the experiment and the assumed noise level of the observations. This deviation is significant, since the test statistic (the Mahalanobis distance of $\widehat{\boldsymbol{x}}$ from $\tilde{\boldsymbol{x}}$)

$$X = (\widehat{\boldsymbol{x}} - \tilde{\boldsymbol{x}})^\mathsf{T} \Sigma_{\widehat{x}\widehat{x}}^{-1} (\widehat{\boldsymbol{x}} - \tilde{\boldsymbol{x}}) = 53.45 > \chi^2_{2,0.99} = 9.21\,. \tag{30}$$

is larger than the tolerance, see the test PCV-(3.32).
Such a comparison is valuable for both the author (having performed the experiment and publishing this in a paper) and the reader (of a conference or journal paper): It tells whether all information of the observations is exploited. For the author this indicates, that there appear not to be any hidden systematic errors left. For the reader this indicates, that the method appears to be adequately designed. This of course has to seen in the context of the size of the experiment, which here is too small.

◇

Before trusting the Cramer-Rao bound and the algebraic derivations, it is useful to perform comparisons between these measures and the result of simulations, in order to get experience for which type of problems simulations appear necessary.

The checks of the implementation as discussed in PCV-4.6.8 are based on the required coherence between the simulations and the other "one-shot" methods using the Cramer-Rao. The evaluation can be based on statistical tests, which allow evaluation as a function also of K. The test only work, if the underlying model is linear enough, i.e., second order effects do not disturb. In order to avoid confusion between different sources for deviations, very small standard deviations (for avoiding second order effects) and large number of iterations (in order to avoid biased estimates) need to be chosen.

A numerical example is given in box on page 8.

Example 2.3: ***Linear regression with two parameters (3).*** We want to check the correctness of the implemented software. For this we refer to PCV-4.6.8 and generate a sufficiently large number of samples for the observations for a fixed parameter vector, following the mathematical model of the estimation procedure, hence without outliers. We choose $K = 25$ in order to obtain accurate results for the check. We refer to the result with `init_rand=15`.

The check leads to the following results:

- The mean of the estimated variance factors is $s^2 = 0.6654$. This appears small compared with the expected value $\mathbb{E}(\widehat{\underline{\sigma}_0}^2) = 1$. However, the confidence interval $[T_l, T_u]$ for a significance level is $[0.4981, 1.6983]$. Hence we have

$$s^2 = 0.7672 \in [0.5593, 1.5923]\,, \tag{31}$$

  and the alternative hypotheses, that the estimated variance factors significantly deviate from 1, is to be rejected.

- The theoretical covariance matrix and the empirical covariance matrix, derived from $K = 25$ estimated parameter vectors $\widehat{\boldsymbol{x}}_k$ are

$$\boldsymbol{\Sigma}_{xx} = \begin{bmatrix} +0.0915 & -0.0072 \\ -0.0072 & +0.0018 \end{bmatrix}, \quad \mathbb{D}(\widehat{\underline{\boldsymbol{x}}}) = \begin{bmatrix} 0.0834 & -0.0065 \\ -0.0065 & 0.0021 \end{bmatrix}. \tag{32}$$

  The test statistic X^2 (see (4.358)) for checking, whether the estimated covariance matrix significantly deviates from the theoretical covariance matrix, is within the confidence interval:

$$X^2_{\boldsymbol{\Sigma}} = 0.7528 \in [0.0717, 12.8382]\,. \tag{33}$$

- Finally, we check whether the estimated parameters are biased using PCV-(4.360). The mean of the estimated parameters is $\widehat{\boldsymbol{m}}_{\widehat{x}} = [0.5479, 1.0070]^\mathsf{T}$. The Mahalanobis distance from the true parameter vector $[0.5, 1.0]$ also lies within the confidence region

$$X^2_{\text{bias}} = 1.0756 \in [0.0100, 10.5966]\,. \tag{34}$$

Hence, we have no reason to assume the implementation has errors.

◇

2.3.2 The Ideal Dependencies

The theoretical precision of the result is representative for many estimation problems. We summarize and interprete these results, and discuss its relevance for other estimation problems.

Recall, the theoretical precision of the estimated parameters is:

$$\sigma_{\widehat{\widehat{x}}_1} = \frac{\sigma}{\sqrt{N}} \quad \text{and} \quad \sigma_{\widehat{x}_2} = \frac{\sigma}{\sqrt{N}} \frac{1}{\text{RMSE}_t} \tag{35}$$

- The standard deviations of the estimates σ_{x_u} linearly increase with the standard deviation σ of the observations.

- The standard deviations decreases with the square root of the number N of observations. This strictly only holds for the centroid. The standard deviation of the slope only decreases with $\sqrt{N}$ if the average distance of the observations from the centroid remain unchanged. This holds (approximately) if the density of the observation over time is changed, but the time interval $t_N - t_1$ remains constant. This often is a reasonable model: for example when analysing the absolute or relative orientation of images using well distributed points in the images, then the average spread (RMSE_x) would characterize

the distribution of the observed image points independent of the number of image points. Then the standard deviations of the pose parameters will approximately decrease with $1/\sqrt{N}$, where N is he number of points in the image.

The graph of $y = 1\sqrt{N}$ however visually is similar to the graphs of $a = 1/N$ or $y = 1/\log N$. Showing the decay of the standard deviations of the parameters therefore should be accompanied by a graph, showing

$$\sqrt{N}\sigma_{\widehat{x}_2} = \frac{\sigma_{\widehat{x}_2}}{\mathrm{RMSE}_t} \tag{36}$$

This ideally does not depend on the number of observations, if the configuration does not change, only the density of the observations. Deviations easily can be seen.

- The standard deviation of the slope decreases linearly with the width RMSE_t of the data. This is typical for geometric problems, where the observed features "carry" the information: The larger the width of the data, the more precise the solution. The width in structure from motion problems may refer to
  - the coverage of the image area,
  - the viewing angle,
  - the length of a straight line segment, or
  - the area of a planar regions covered by 3D points.

  The special structure of a geometric problem may also lead to other dependencies of the width of the data: As an example: the standard deviation of the rotation angles (ω and φ) of a camera across the viewing direction decrease quadratically with the width d of the image area covered by image features. Here, a plot of $d^2\sigma_\omega$ for varying d should show no dependency on d.

Numerical examples are given in the box on page 10.

2.3.3 Causes for Deviations from the Ideal Dependencies

Often these dependencies are derived by simulations to demonstrate the "robustness" of the solution (actually the theoretical precision): showing the uncertainty of the estimated parameters as a function of the noise added to the observations. This is derived by repeating the estimation K times, and reporting the RMSE of the parameters as a function of σ. If the number K of samples is large enough the linear dependency should be visible in the graph.

Deviations from the linearity may either be have different causes, e.g., :

- a too low number K of samples. The relative precision of the estimated standard deviation is appr. $\sqrt{1/K}$. For achieving a 5% accuracy at least $K = 25$ samples need to be taken.
- the influence of the linearization of a non-linear model, see the discussion in Sect. 2.7.6.
- a lack of convergence of an iterative estimation scheme. This may even occur for a linear problem, if no direct solution, e.g., by Gaussian elimination is used to solve the normal equations, but e.g., a conjugate gradient method.

Example 2.4: ***Linear regression with two parameters (4).*** We want to demonstrate dependencies of the noise level σ using simulated data, and discuss how to visualize such results.

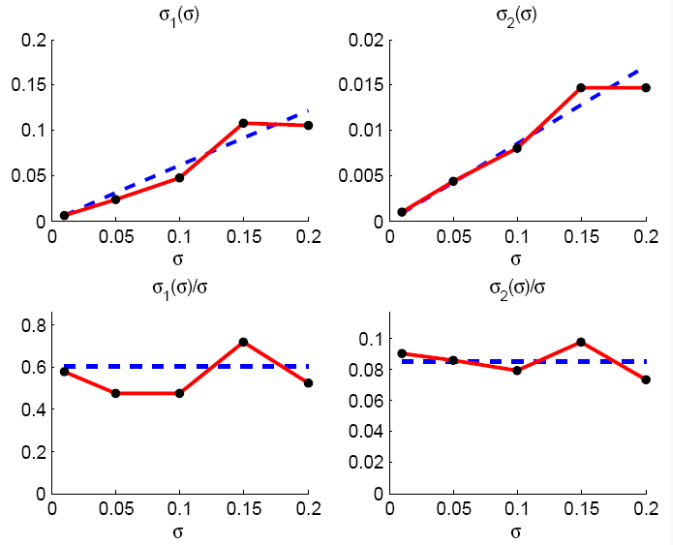
We repeat the simulations used for checking the correctness of the implementation for the following noise levels:

$$\sigma \in [0.01, 0.05, 0.1, 0.15, 0.20]\,, \tag{37}$$

and visualize the corresponding standard deviations from the empirically determined covariance matrices (32), right. There, for $\sigma = 0.5$ we would obtain $\widehat{\sigma}_{\widehat{x}_1} = 0.2889$ and $\widehat{\sigma}_{\widehat{x}_2} = 0.0453$. The red lines in the top row of the figure shows the standard deviations $\widehat{\sigma}_{\widehat{x}_i}(\sigma), i = 1, 2$ for the two parameters. They approximately increase linearly with σ; the theoretical increase is shown as dashed blue line. The discrepancies are due to the number $K = 25$ of samples used for the simulation, which causes a relative error of 5%. This blue dashed curve usually is not known, unless for a single choice of σ the theoretical covariance matrix $\Sigma_{\widehat{x}\widehat{x}}$ is determined, which for equally weighted observations is

$$\Sigma_{\widehat{x}\widehat{x}} = \sigma^2 (A^{\mathsf{T}} A)^{-1}\,, \tag{38}$$

see (PCV-(4.49)). A linear dependency easily can be mistaken for an affine dependency, where there is an offset at $\sigma = 0$, if the simulations do not start with a very small sigma.



Standard deviation of the the parameters as a function of the noise level. The empirical dependency is given in red, the theoretical dependency is given in dashed blue. **Top left:** Standard deviation $\widehat{\sigma}_{\widehat{x}_1}(\sigma)$. The dependencies should linearly increase. **Top right:** Standard deviation $\widehat{\sigma}_{\widehat{x}_1}(\sigma)$. **Bottom left:** Standard deviation $\widehat{\sigma}_{\widehat{x}_2}(\sigma)/\sigma$. **Bottom right:** Standard deviation $\widehat{\sigma}_{\widehat{x}_2}(\sigma)/\sigma$. The dependencies here should be a constant

If not a very small noise level σ for the observations is included in the simulations, it is recommended to visualize the ratio

$$r_1(\sigma) = \frac{\widehat{\sigma}_{\widehat{x}_1}(\sigma)}{\sigma}\,, \tag{39}$$

which should be a constant. This easily can be checked visually; see the bottom row.

◇

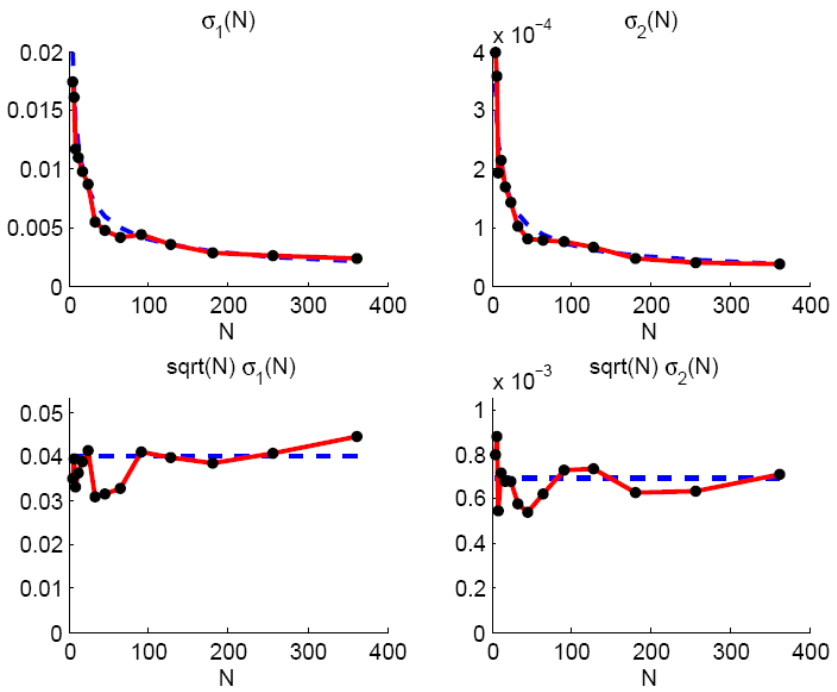
- a Levenberg-Marquardt solution is used in the presence of a singular normal equation system. Hence, the geometry of the problem represents a degenerate configuration. Then the unknown parameters are not estimable, and the regularizer enforces a solution close to the one yielding a minimum norm for the covariance matrix of the unknown parameters. Then the estimates depend on the approximate values. This might lead to a deviation of the linear relationship between the RMSE for the unknown parameters and the assumed noise standard deviation.

Example 2.5: ***Linear regression with two parameters (5).*** Similarly as for the noise standard deviation, we want to demonstrate dependencies on the density of the observations using simulated data, and discuss how to visualize such results. We therefore assume the observations to be regularly spaced in a fixed interval, and vary N. We assume this interval to be 100, and the noise standard deviation to be 0.2.

We assume the observations are taken in a fixed interval, and vary N. For the sequence

$$N \in [4, 6, 8, 11, 16, 23, 32, 45, 64, 91, 128, 181, 256, 362] \tag{40}$$

the red line in the top row in Fig. 2.3.3 shows the estimated $\widehat{\sigma}_{\widehat{x}_i}, i = 1, 2$. The standard deviations are decaying, as expected.



Standard deviation of the the parameters as a function of observational density The empirical dependency is given in red, the theoretical dependency is given in dashed blue. **Top left:** Standard deviation $\widehat{\sigma}_{\widehat{x}_1}(N)$. **Top right:** Standard deviation $\widehat{\sigma}_{\widehat{x}_2}(N)$. The theoretical dependencies follow approximately $1/\sqrt{N}$. **Bottom left:** Standard deviation $\sqrt{N}$
$est\sigma_{\widehat{x}_1}(N)$. **Bottom right:** Standard deviation $\sqrt{N}\ \ \widehat{\sigma}_{\widehat{x}_2}(N)$. The dependencies here should be a constant

In order to confirm the dependency on N, namely a decay with $1/\sqrt{N}$, we show

$$r_2(N) = \sqrt{N}\ \ \widehat{\sigma}_{\widehat{x}_i}(N)\,, \tag{41}$$

in the bottom row. It should be a constant.
Often the level of the theoretical precision is not known or difficult to obtain, e.g., since the software does not provide the standard deviations of the estimates. In this case the blue dashed curves in top row could be replaced by a best fitting function $\sigma_{\widehat{x}_i} = a_i/\sqrt{N}$ in order to visually prove the type of dependency, and avoid the normalize plots in the bottom row.

◇

- inconsistencies between the simulated data and the used model. Hence the simulation checks both: the program for generating artificial data and the estimation routine.
- suboptimal implementation of the simulation or the estimation.

2.4 Testing and the Sensitivity of the Estimation

The evaluation of the result can be based on useful measures for identifying outliers and weaknesses in the geometric configuration, see the box on page 14 and Fig. 2, p. 5.

2.4.1 Detectability of Outliers and Testing

The largest residual is $\widehat{\boldsymbol{v}}_3 = -0.463$, much larger in magnitude than $\widehat{v}_4 = 0.092$, though we know that there is an outlier in l_4.

To obtain insight into the geometry of the observational design we investigate the redundancy matrix,[1] which shows how deviations in the observation influence the residuals.

Relating the observations to the centroid the redundancy matrix is

$$R = I - A(A^\mathsf{T}A)^{-1}A = I - \frac{\mathbf{1}\mathbf{1}^\mathsf{T}}{\mathbf{1}^\mathsf{T}\mathbf{1}} - \frac{\bar{\boldsymbol{t}}\,\bar{\boldsymbol{t}}^\mathsf{T}}{\bar{\boldsymbol{t}}^\mathsf{T}\bar{\boldsymbol{t}}}\,. \tag{42}$$

or with the mean position RMSE_t

$$R = I - \frac{1}{N}\mathbf{1}\mathbf{1}^\mathsf{T} + \frac{1}{N}\frac{\bar{\boldsymbol{t}}\,\bar{\boldsymbol{t}}^\mathsf{T}}{\mathrm{RMSE}_t^2}\,. \tag{43}$$

In our case we obtain

$$\nabla\widehat{\boldsymbol{v}} = -R\nabla\boldsymbol{l} \quad \text{with} \quad R = \begin{bmatrix} 0.5688 & -0.3587 & -0.3225 & 0.1123 \\ -0.3587 & 0.6848 & -0.2935 & -0.0326 \\ -0.3225 & -0.2935 & 0.7210 & -0.1051 \\ 0.1123 & -0.0326 & -0.1051 & 0.0254 \end{bmatrix}. \tag{44}$$

Hence, even if the outlier in observation l_4 would have a much larger size, the effect onto the residual of observation l_3 would be larger than the effect onto $\widehat{v}_4$. It can be seen from the elements of the redundancy matrix: Since the off-diagonal term $r_{34} = -0.1051$ is approximately 4-times larger than the redundancy number, namely $r_4 = 0.0254|r_{43}| > r_4$, residual $\widehat{v}_3$ is more influence by an outlier in l_4, than the corresponding residual $\widehat{v}_4$.

Using the relative distances

$$d_n = \frac{\bar{t}_n}{\mathrm{RMSE}_t} \quad \text{with} \quad |d_n| \le \sqrt{N-1}\,. \tag{45}$$

the redundancy numbers r_n thus are

$$r_n = 1 - \frac{1 + d_n^2}{N} = 1 - u_n \in [0, 1]\,. \tag{46}$$

They obviously sum to the redundancy $R = N - U = N - 2$.

The redundancy numbers show two extreme distribution.

- If the observations are equally spaced, e.g., $t_n = t_0 + n\Delta t$ then the redundancy of the middle observation (assuming N is odd) $r_{(N+1)/2} = (N-1)/N$, whereas the redundancy number of the first (or the last) observation is $r_1 = (N-1)/N\ (N-2)/(N+1) < r_{(N+1)/2}$. Hence if the number of observations is larger than 40, all redundancy numbers

[1] The redundancy matrix only is symmetric if all observations have the same weight.

are above 0.9. This simplifies the analysis, and allows to work with approximations. The first approximation for the redundancy numbers

$$r_n \approx \overline{r_n} = \frac{R}{N} \tag{47}$$

just assumes the redundancy numbers do not vary too much; this approximation is assumed in the variance analysis using the triangulation, see PCV-15.4.1.3, Eq. (15.132). The second approximation is $r_n \approx 1$ assuming $U \ll N$. Then also $\mathsf{R} \approx \mathsf{I}$, which simplifies the analysis of the residuals, as they are assumed to be uncorrelated.

- If $N-1$ of the observations are clustered and one observation is far off, we obtain the maximum redundancy number $r_n \approx 11/(N-1)$ for the observations in the cluster and $r_n \approx 0$ for the observation far off, since it is necessary for determining the slope of the line. Hence, no approximation of the redundancy numbers can be derived, and they need to be used for a reliable analysis.

Testing the residuals is mandatory if outliers are to be expected. The standardized residuals

$$z_n = \frac{-v_n}{\sigma_{v_n}} = \frac{-v_n}{\sigma_{l_n}\sqrt{r_n}}\,, \tag{48}$$

via the redundancy number r_n take the geometry into account. They are more sensitive if the redundancy number is small, i.e., at the borders of the observations. A less sensitive test statistic for outlier detection is the normalized residual

$$z_n^* = \frac{-v_n}{\sigma_{l_n}}\,. \tag{49}$$

The detectability of outliers can be characterized by the minimum size of an outlier which can be detected reliably by a statistical test. Following PCV-(4.285) and PCV-(4.289) we have this minimum size of a detectable outlier for the two tests

$$\nabla_0 l_n = \delta_0 \frac{\sigma_{l_n}}{\sqrt{r_n}} \quad \text{and} \quad \nabla_0^* l_n = \delta_0 \frac{\sigma_{l_n}}{r_n} = \frac{1}{\sqrt{r_n}} \nabla_0 l_n\,. \tag{50}$$

They differ by a factor $1/\sqrt{r_n} \geq 1$. Hence is r_n is small, say below 0.1, we not only see just 10% of the causing outlier in the residuals, but – in our example – instead of $\nabla_0 l_n \approx 12\sigma_n$ for the statistical test we can only find outliers larger than $\nabla_0^* l_n \approx 30\sigma_l$.

2.4.2 The Theoretical Sensitivity w.r.t. all Parameters

We now analyse the sensitivity of the result w.r.t. to possible outliers.

The sensitivity factor w.r.t. all parameters is

$$\mu_{x,n} = \sqrt{\frac{1-r_n}{r_n}} = \sqrt{\frac{1+d_n^2}{N-(1+d_n^2)}}\,. \tag{51}$$

Leverage Points: The effect of observations, besides the number N of observations, essentially depends on the relative distance of the observation to the centroid. Fig. 3 shows this dependency for the case $N = 10$. Obviously, observations with small redundancy number have a large influence onto the estimated parameters. Such points are called *leverage points, see PCV p. 127.*

Example 2.6: ***Linear regression with two parameters (6).*** We collect the decisive numbers w.r.t. outlier detection. We have the following indicators (see Fig. (2), p. 5)

n	v_n	r_n	$\widehat{\nabla l_n}$	z_n	z_n^*	$\nabla_0 l_n$	$\nabla_0^* l_n$
1	0.3678	0.5688	-0.6466	0.9754	0.7356	2.6518	3.5159
2	0.0032	0.6848	-0.0047	0.0078	0.0064	2.4169	2.9206
3	**-0.4633**	0.7210	0.6425	-1.0911	**-0.9265**	2.3554	2.7739
4	0.0922	**0.0254**	**-3.6362**	**1.1582**	0.1844	**12.5584**	**78.8571**

Evaluation of outlier detection. We assumed $\delta_0 = 4$, see PCV, Table 3.2, p. 67

Observe the effect of using a suboptimal, non-sufficient test statistic z_n^*: (1) the test statistic z_n^* (in this example) points towards a wrong observation, (2) outliers must be large by a factor of at least 79 of their standard deviation to be detectable.
Reporting these numbers (except z_n^* and $\nabla_0^* l_n$) for visual inspection of the result may be appropriate for problems with a not too large number of observations. A summarizing report however is useful, where the extreme values are collected together with an indicator whether they are acceptable. These extreme numbers are part of a self-diagnosis of the estimation procedure.

The minimum redundancy number belongs to the 4-th observation:

$$r_4 = 0.025\,. \tag{52}$$

It indicates, that if the observation is changed by some amount, the effect onto the corresponding residual is only approximately 2.5% of that amount. The other 97.5% of this amount influence the parameters, as we will see, when analysing the sensitivity of the estimates.

The estimated size of a possible outlier in this observation is

$$\widehat{\nabla l_4} = \frac{-\widehat{v}_4}{r_4} = 3.64\,, \tag{53}$$

which is in the right order of magnitude.

A statistical test, does not indicate an outlier: the maximum test statistic occurs at the 4th observation:

$$z_4 = -\frac{\widehat{v}_4}{\sigma_{\widehat{v}_4}} = 1.16 \leq 2.58\,. \tag{54}$$

It correctly points towards the erroneous observation, though it is not significant; for a significance level of $S = 99\%$ the two-sided test has a non-rejection region $[-2.58, +2.58]$.

The largest size $\nabla_0 l_n$ of a detectable outlier, when using a statistical test with z_n, is in observation l_4, namely

$$\nabla_0 l_4 = 12.56\,. \tag{55}$$

An outlier in this observation needs to be larger than 25 times (!) the standard deviation of $\sigma = 0.5$ to be detectable with a minimum probability of 80%. This three times larger, than the outlier of size $\nabla l_4 = -4$ we introduced.

◇

2.4.3 The Theoretical Sensitivity w.r.t. Centroid

We now investigate the sensitivity of then result for the case, that we are only interested in one of the two parameters. We start with the sensitivity w.r.t. the centroid, i.e., the value $f(\mu_t)$ with $f(t) = x_1 + x_2 t$ and $\mu_t = \sum_n t_n / N$.

The question is: How much influence does a non detectable outlier in one of the observations have onto the centroid. Hence the slope of the line is of no interest and treated as a nuisance parameter. This is like we would be only interested in the position of an object in 3D space, and not interested in his orientation (rotation matrix).

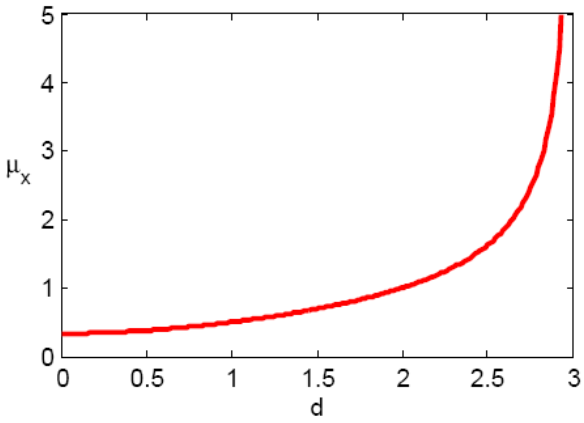


Figure 3: Sensitivity factor $\mu_{x,n}$ as a function of the relative distance d of the observation from the centroid. For leverage points, i.e., single points lying far apart from the others the relative distance is large. Changes in the corresponding observation have a large influence on the parameters

We eliminate the scale following PCV-(4.122) and obtain the part $\overline{C}$ of the reduced design matrix:

$$\overline{C} = C = \mathbf{1} \quad \text{and} \quad \Sigma_{x_1 x_1} = \sigma^2 \frac{1}{N} \,. \tag{56}$$

We now need the value $\overline{u}_{x_1 n}$, see PCV-(4.128),

$$\overline{u}_{x_1,n} = \frac{1}{N} \,. \tag{57}$$

Therefore the sensitivity factor w.r.t. to the centroid is, see PCV-(4.296)

$$\mu_{x_1,n} = \sqrt{\frac{\overline{u}_n}{r_n}} = \sqrt{\frac{1}{N - (1 + d_n^2)}} \leq \mu_{x,n} \,. \tag{58}$$

Figure 4 shows the dependency of the sensitivity factor $\mu_{x_1,n}$ on the relative distance of an observation to the centroid. It is significantly smaller than $\mu_{x,n}$, since parts of the non detectable errors are absorbed by the slope, which is a nuisance parameter.

2.4.4 The Theoretical Sensitivity w.r.t. Slope

In a similar manner we can analyse the sensitivity of the estimated slope. This is similar to analysing the sensitivity of the estimated rotation of an object, observed by a motion capture system, taking the 3D coordinates of the centre of gravity as nuisance parameters.

Here we reduce the normal equation system to the slope, and obtain the reduced design matrix

$$\overline{D} = D = \boldsymbol{t} \quad \text{and} \quad \Sigma_{x_2 x_2} = \sigma^2 \frac{1}{N \text{RMSE}_t^2} \,. \tag{59}$$

Here we now need $\overline{u}_{x_2 n}$:

$$\overline{u}_{x_2,n} = \frac{\bar{t}_n^2}{N \text{RMSE}_t^2} = \frac{d^2}{N} \,. \tag{60}$$

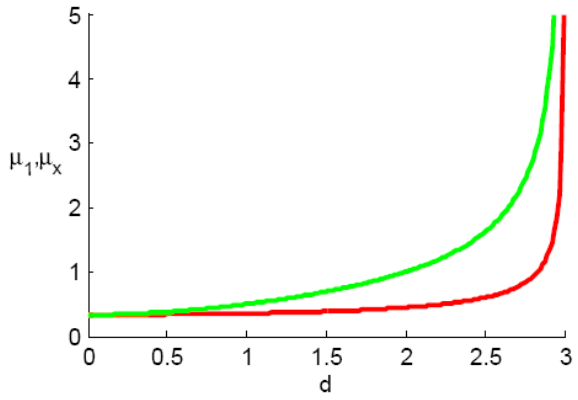


Figure 4: Sensitivity w.r.t. the centroid (red) compared to the sensitivity w.r.t. all parameters (green). Moderate leverage points mainly influence the slope, hence have only a limited influence onto the centroid. Only in extreme situations, where the distance of a point is very far off, non-detectable outliers are large enough to still have an influence on the centroid

Therefore the sensitivity factor w.r.t. to the slope is

$$\mu_{x_2,n} = \sqrt{\frac{d^2}{N - (1 + \delta_n^2)}} \leq \mu_{x,n} \tag{61}$$

The dependency of the sensitivity factor $\mu_{x_2,n}$ on the relative distance of an observation to the centroid is shown in Fig. 4. Obviously the difference is largest for points close to the centroid, reducing the sensitivity factor to 0: this is plausible, since these observations have no influence on to the slope at all.

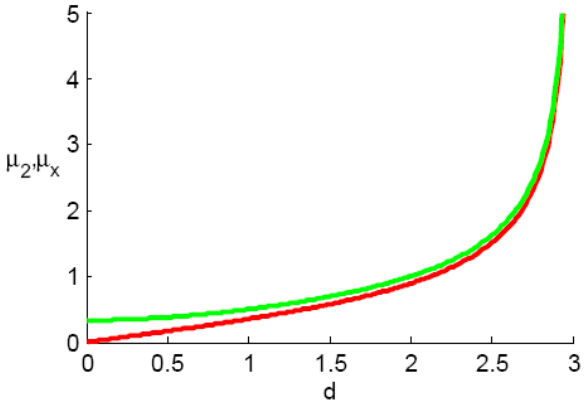


Figure 5: Sensitivity w.r.t. the slope (red) compared to the sensitivity w.r.t. all parameters (green). Points close to the centroid have no influence on the slope, as to be expected

3 Gauss–Markov Model for Planar Similarity Transformation with Evaluation

This section gives more details on the estimation of a similarity transformation used for generating Fig. 4.11. It at the same time explains the corresponding MATLAB source file

`fig_4_11_test_sensitivity_factors_GMM_similarity.m`.

3.1 The Mathematical Model

The geometric model is the following

$$\begin{bmatrix} x'_i \\ y'_i \end{bmatrix} = \begin{bmatrix} ax_i - by_i + c \\ bx_i + ay_i + d \end{bmatrix} . \tag{62}$$

It holds for the true or expected values.

We assume the coordinates $\boldsymbol{x}_i = [x_i, y_i]^\mathsf{T}$ are given fixed values, the transformed coordinates $\boldsymbol{x}'_i = [[x'_i, y_i]^\mathsf{T}$ are observed and the 4 parameters $[a, b, c, d]$ are unknown. We assume the observed coordinates have the same uncertainty, with covariance matrix $\Sigma_{x'_i x'_i} = \sigma^2 I_2$. Figure 6 shows

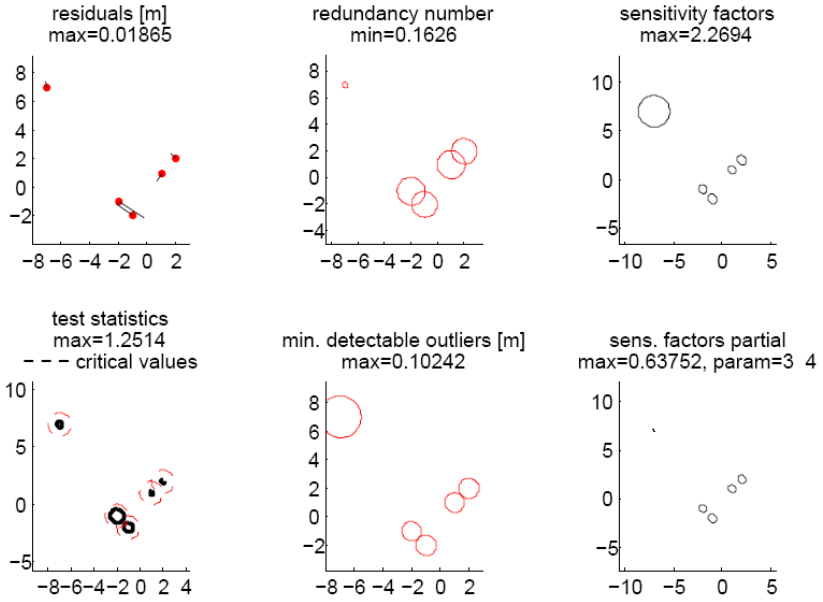


Figure 6: Results and diagnostic parameters for a similarity transformation, $\sigma = 0.01$ m.

We collect the $N = 2I$ observations (lines 93/94) and the $U = 4$ unknown parameters (lines 69) in the vectors

$$\boldsymbol{l} := \begin{bmatrix} \boldsymbol{x}'_1 \\ \dots \\ \boldsymbol{x}'_i \\ \dots \\ \boldsymbol{x}'_I \end{bmatrix} \quad \text{and} \quad \boldsymbol{x} := \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} . \tag{63}$$

The $N \times U$ design matrix is (see lines 95/96) is

$$A = [A_i^\mathsf{T}] = \left[\underbrace{\begin{bmatrix} x_i & -y_i & 1 & 0 \\ y_i & x_i & 0 & 1 \end{bmatrix}}_{A_i^\mathsf{T}} \right]_{i=1,\dots,I} . \tag{64}$$

The Gauss–Markov model reads

$$\boldsymbol{l} + \boldsymbol{v} = A\boldsymbol{x} + \boldsymbol{a}\,, \qquad \Sigma_{ll} = \sigma^2 I_N\,, \tag{65}$$

with the constant vector $\boldsymbol{a} = \boldsymbol{0}$. Observation wise this is

$$\boldsymbol{l}_i + \boldsymbol{v}_i = A_i^\mathsf{T}\boldsymbol{x}\,, \qquad \Sigma_{l_i l_i} = \sigma^2 I_2\,, \tag{66}$$

A numerical example is given in the box on 18

Example 3.7: ***Similarity transformation (1).*** Figure 6 shows the result of a similarity transformation using `GaussMarkovModelLinear_groups.m` with initialization of the random numbers with `init_rand`= 15. We used this configuration for generating the images in PCV-Fig.4.11. We will refer to this figure in the following. The true values are given by:

$$\tilde{\boldsymbol{x}} = \begin{bmatrix} 2.0 \\ 0.5 \\ 3.0 \\ -2.0 \end{bmatrix}, \quad \sigma = 0.01\,, \quad [(x_i, y_i)] = \begin{bmatrix} -7.0 & 7.0 \\ 1.0 & 1.0 \\ 2.0 & 2.0 \\ -1.0 & -2.0 \\ -2.0 & -1.0 \end{bmatrix}, \quad \tilde{\boldsymbol{l}} = \begin{bmatrix} -14.5 & 8.5 \\ 4.5 & 0.5 \\ 6.0 & 3.0 \\ 2.0 & -6.5 \\ -0.5 & -5.0 \end{bmatrix}. \tag{67}$$

The true errors $\tilde{\boldsymbol{e}}$ are generated as sample from $\mathcal{N}(0, \sigma^2)$.

$$\tilde{\boldsymbol{e}} = \begin{bmatrix} 0.0159 & 0.0003 \\ 0.0033 & 0.0121 \\ 0.0026 & 0.0038 \\ 0.0108 & 0.0072 \\ -0.0082 & 0.0204 \end{bmatrix} \quad \boldsymbol{l} = \begin{bmatrix} -14.4841 & 8.5003 \\ 4.5033 & 0.5121 \\ 6.0026 & 3.0038 \\ 2.0108 & -6.4928 \\ -0.5082 & -4.9796 \end{bmatrix}. \tag{68}$$

The estimated parameters and the residuals, shown in Fig. 6 upper left, are

$$\widehat{\boldsymbol{x}} = \begin{bmatrix} 1.9986 \\ 0.4996 \\ 3.0023 \\ -1.9899 \end{bmatrix}, \quad \widehat{\boldsymbol{v}} = \begin{bmatrix} -0.0008 & 0.0030 \\ -0.0019 & -0.0038 \\ -0.0022 & 0.0027 \\ -0.0079 & 0.0061 \\ 0.0129 & -0.0080 \end{bmatrix}. \tag{69}$$

◇

3.2 The Estimation

The estimation is realized in the MATLAB function `GaussMarkovModelLinear_groups.m`. It in a first step follows Alg. 1, PCV-p.91. An additional routine `diagnostics_GMM_multi_d.m` performs the sensitivity analysis. Given a set r_U of parameters of interest it determines all diagnostic parameters of interest:

- the covariance matrix of the estimated parameters

$$\Sigma_{\widehat{x}\widehat{x}} = (A^\mathsf{T} W_{ll} A)^{-1} \tag{70}$$

  assuming the a priori variance factor is $\sigma_0^2 = 1$.

- the residuals $\widehat{\boldsymbol{v}}_i$ and the estimated variance factor

$$\widehat{\boldsymbol{v}}_i = A_i^\mathsf{T}\widehat{\boldsymbol{x}} + \boldsymbol{a}_i - \boldsymbol{l}_i \quad \text{and} \quad \widehat{\sigma}_0^2 = \frac{\widehat{\boldsymbol{v}}^\mathsf{T} W_{ll} \widehat{\boldsymbol{v}}}{N - U} \tag{71}$$

- the test statistics X_i (PCV-(4.302))

$$X_i = \widehat{\boldsymbol{v}}_i^{\mathsf{T}} \Sigma_{\widehat{v}_i \widehat{v}_i}^{-1} \widehat{\boldsymbol{v}}_i \,. \tag{72}$$

- the diagonal $d \times d$ block R_{ii} of the the redundancy matrix R (PCV-(4.299))

$$R_{ii} = \Sigma_{\widehat{v}_i \widehat{v}_i} W_{l_i l_i} \,. \tag{73}$$

- the minimum size $\nabla_0 \boldsymbol{l}_i$ of detectable outliers (PCV-(4.304), (4.300))

$$\nabla_0 \boldsymbol{l}_i = \delta_0 \sqrt{\lambda_{\max}(R_{ii}^{-1} \Sigma_{l_i l_i})} \,. \tag{74}$$

  We use $\delta_0 = 4.13$, independent on the dimension. This is a useful choice if the group size of the observations is not too large. As can be seen in Table 3.3, PCV-p.68, in our case of $d = 2$ this corresponds to applying a test with significance number $\alpha = 0.3\%$ and requiring a minimum power of $\beta_0 = 80\%$ for finding an outlier.

- the sensitivity factor w.r.t. all 4 parameters

$$\mu_{x,n} = \sqrt{\lambda_{\max}(R_{ii}^{-1} - I_d)} \,; \tag{75}$$

  From PCV-(4.310) we have $\Sigma_{\widehat{l_i l_i}} \Sigma_{\widehat{v}_i \widehat{v}_i}^{-1} = (\Sigma_{l_i l_i} - \Sigma_{\widehat{v}_i \widehat{v}_i}) \Sigma_{\widehat{v}_i \widehat{v}_i}^{-1} = \Sigma_{l_i l_i} \Sigma_{\widehat{v}_i \widehat{v}_i}^{-1} - I_2 = R_{ii}^{-1} - I_2$.

- the sensitivity factor w.r.t. the selected set r_U of parameters (in PCV-(4.315) referred to as parameter set $\boldsymbol{k}$)

$$\mu_{x_1,n} = \sqrt{\lambda_{\max}(\overline{U}_{r_U,ii}\, R_{ii}^{-1})} \,. \tag{76}$$

  This holds since $\overline{U}_{r_U,ii} = \overline{C}_i^{\mathsf{T}} \Sigma_{r_U r_U} \overline{C}_i W_{l_i l_i}$ (see PCV-(4.125))

The main results for the example are collected in the box on 21. It explicitly addresses tools for self-diagnosis:

- The covariance matrix or the standard deviations of the estimated parameters tell the sensitivity of the result w.r.t. randm errors in the observations.

- The estimated variance factor indicates the overall consistency of the model with the data. Here it can be determined from

$$\widehat{\sigma}_0^2 = \frac{\sum_{i=1}^N \widehat{\boldsymbol{v}}_i^{\mathsf{T}} W_{\widehat{v}_i \widehat{v}_i} \widehat{\boldsymbol{v}}_i}{N - U} = \frac{\sum_{i=1}^N |\widehat{\boldsymbol{v}}_i|^2 / \sigma^2}{2I - 4} \quad \text{with} \quad \underline{\widehat{\sigma}_0}^2 \sim F_{N-U,\infty} \,. \tag{77}$$

  The first expression for determining the estimated variance factor assumes the observational groups $\boldsymbol{l}_i$ to be uncorrelated, but may have individual and full covariance matrices $\Sigma_{l_i l_i} = W_{l_i l_i}^{-1}$. The second expression exploits the assumption that all points have the same isotropic uncertainty $\Sigma_{l_i l_i} = W_{l_i l_i}^{-1} = \sigma^2 I_2$. cFor a discussion on the evaluation of the estimated variance factor see PCV-4.2.3.

- The *maximal residual* $\max(\boldsymbol{v}_n)$ (or $\max(|\boldsymbol{v}_i|)$) should always be reported, though, if the design is not homogeneous it does not tell whether there are no outliers. Observe, if the

observations have the same standard deviation σ and are uncorrelated, the RMSE_v of the residuals (the dimension of the observational groups is d)

$$\mathrm{RMSE}_v = \sqrt{\frac{1}{N}\sum_{n=1}^{N} v_n^2} = \sqrt{\frac{1}{dI}\sum_{i=1}^{N} |\boldsymbol{v}_i|^2} \tag{78}$$

is related to the variance factor by

$$\widehat{\sigma}_0^2 = \frac{N}{N-U}\frac{1}{\sigma^2}\,\mathrm{RMSE}_v^2\,, \tag{79}$$

a result which allows us to *statistically test the root mean square error*. If the observations have different weight or are correlated, the RMSEdoes not follow a χ^2-distribution; it does not lead to a sufficient test statistic, since the prior knowledge about stochastical model is not used.

- The *maximal redundancy number* $\min(r_n)$ should be above 0.1.
- The *maximum test statistic* $\max(z_n)$ should always be reported in order to be sure that the statistical test dose not suggest an outlier to be present, if it remains in the non-rejection region which can be derived from the χ^2-distribution.
- The *maximum just detectable outlier* $\max(\nabla_0 l_n)$ indicates in a application oriented way whether the geometry allows to identify outliers. Observe, the value 0.102 is more than 10 times the standard deviation. This measure only is relevant if the goal of the estimation is to find outliers.
- The *maximum sensitivity factor* $\max(\mu_{x,n})$ should always be reported. It should be lass than5 or 10, since then non-detectable outliers have an influence of less than 20 or 40 times the standard deviation of the resultant estimates.
- The *maximum sensitivity factor* $\max\left(\mu_{x_{r_U},n}\right)$ is very useful if the goal of the estimation is to estimate the parameters in $\boldsymbol{r}_U$, and the user wants to have a guarantee that non-detectable outliers do not perturb the result. Observe, this sensitivity factor may be small, even if quite large outliers may stay undetected.

Such a summarizing self-diagnosis is useful for a quick evaluation of the quality of the result of the estimation. Visualizing the individual numbers, as in Fig. 6, of course needs to be adapted to the individual estimation problem. If the estimation process is one module within a chain of modules, the characterizing numbers may be used by the subsequent module.

Example 3.8: ***Similarity transformation (2).*** The quality of the result of the estimation needs to be documented. The covariance matrix of the estimates is

$$\Sigma_{\widehat{x}\widehat{x}} = 10^{-4} \begin{bmatrix} 0.0102 & 0 & 0.0142 & -0.0142 \\ 0 & 0.0102 & 0.0142 & 0.0142 \\ 0.0142 & 0.0142 & 0.2398 & 0 \\ -0.0142 & 0.0142 & 0 & 0.2398 \end{bmatrix} \tag{80}$$

Observe, the parameters $\widehat{x}_1 = \widehat{a}$ and $\widehat{x}_2 = \widehat{b}$ representing scale and rotation are uncorrelated, also the two translation parameters $\widehat{x}_3 = \widehat{c}$ and $\widehat{x}_4 = \widehat{d}$ are uncorrelated. This is caused by the assumption that the observed points are uncorrelated and have the same standard deviation, thus the uncertainty of the point group is isotropic.

The standard deviations of the parameters are

$$\sigma_{\widehat{x}_1} = \sigma_{\widehat{x}_2} = 0.0010\,, \quad \sigma_{\widehat{x}_3} = \sigma_{\widehat{x}_4} = 0.0049\,. \tag{81}$$

The estimated variance factor is

$$\widehat{\sigma}_0^2 = 0.7863\,. \tag{82}$$

The quality of the observations can be characterized by the following statements, which are taken from the MATLAB output:

```
Maximal residual ......................  = 0.01520 at observation 5
Minimal redundancy number .............  = 0.16260 at observation 1
Maximal test statistic ................  = 1.25145 at observation 5
Maximum of minimal detectable outlier ..  = 0.10242 at observation 1
Maximal sensitivity factor ............  = 2.26936 at observation 1
Maximal sensitivity factor translation .  = 0.63752 at observation 4
```

Here we have assumed the user is also interested in the sensitivity of the estimated translation parameters x_3 and x_4 only, see PCV-Fig.4.11 right.

The quality measures indicate, that there is no reason to assume the model not to be consistent with the data and that the geoemtric configuration is acceptable.

◇

4 Gauss–Helmert Model for Planar Similarity Transformation

This section has two goals:

1. Demonstrate the sensitivity analysis for the Gauss–Helmert model.
2. Discuss the conditions for the equivalence of the Gauss–Helmert model and the Gauss–Markov model.

4.1 Sensitivity Analysis for the Gauss–Helmert Model

The sensitivity analysis aims at investigating the ability to find outliers in the observations and to determine the effect of non-detectable errors on the estimated parameters.

Let the group of observations related to the i-th constraint be collected in the vector $\widehat{\boldsymbol{v}}_i$. We assume that constraints do not share observations, see PCV-4.8.2.5. The direct approach would require the inversion of the covariance matrix $\Sigma_{\widehat{v}_i\widehat{v}_i}$ of the corresponding group of estimated residuals. Let the size of this group be N_i.

It can be derived from PCV-(4.456)

$$\Sigma_{\widehat{v}\widehat{v}} = \Sigma_{ll} B W_{gg} (B^{\mathsf{T}} \Sigma_{ll} B - A \Sigma_{\widehat{x}\widehat{x}} A^{\mathsf{T}}) W_{gg} B^{\mathsf{T}} \Sigma_{ll}\,. \tag{83}$$

Assuming $\Sigma_{ll} = \mathrm{Diag}(\{\Sigma_{l_i l_i}\})$ the covariance matrix related to the i-th group is

$$\underbrace{\Sigma_{\widehat{v}_i\widehat{v}_i}}_{N_i\times N_i} = \underbrace{\Sigma_{l_i l_i} B_i^{\mathsf{T}}}_{N_i\times G}\ \underbrace{Z}_{G\times G}\ \underbrace{B_i^{\mathsf{T}}\Sigma_{l_i l_i}}_{G\times N_i} \quad \text{with} \quad Z = W_{gg}(B^{\mathsf{T}}\Sigma_{ll}B - A\Sigma_{\widehat{x}\widehat{x}}A^{\mathsf{T}})W_{gg}\,. \tag{84}$$

Since generally $G < N_i$ the rank of this $N_i \times N_i$ is only G, thus it cannot be inverted.

However, testing the group $\boldsymbol{l}_i$ related to the i-th set of constraints is equivalent to testing the residual $\boldsymbol{c}_{g,i}$ of that constraint. This equivalent to use the residual

$$\boldsymbol{v}_g = B^{\mathsf{T}}\boldsymbol{v} \tag{85}$$

and use the linearized Gauss–Markov model (see PCV-457)

$$\boldsymbol{g}(\boldsymbol{l}, \widehat{\boldsymbol{x}}^a) + \widehat{\boldsymbol{v}}_g = -A\widehat{\Delta\boldsymbol{x}} \quad \text{with} \quad \mathbb{D}(\underline{\boldsymbol{v}}_g) = B^{\mathsf{T}}\Sigma_{ll}B\,, \tag{86}$$

for testing w.r.t. outliers in the values $\boldsymbol{c}_{g,i}$. This type of diagnosis is realized in the MATLAB function `diagnostics_GHM_constraints_multi_d.m`.

4.2 Estimation a Similarity Transformation using the Gauss–Helmert Model

If the observations can be expressed as functions of the unknown parameters, the Gauss–Markov model is the most appropriate model for estimation. If the similarity transformation has to be estimated from point pairs, which both are observed, we obtain constraints between the observations (x_i, y_i, x_i', y_i') and the unknown parameters (a, b, c, d):

$$\boldsymbol{g}_i := \begin{bmatrix} \tilde{x}_i' \\ \tilde{y}_i' \end{bmatrix} - \left(\begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} \tilde{x}_i \\ \tilde{y}_i \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix} \right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{87}$$

This constraint not easily can be transformed into a Gauss–Markov model. Therefore, here the Gauss–Helmert model is the appropriate choice for estimation.

The model is linear in the unknown parameters and, starting from approximate values zero, can be written as

$$\boldsymbol{g}(\widehat{\boldsymbol{l}}_i^a, \widehat{\boldsymbol{x}}^a) + A_i^{\mathsf{T}}\widehat{\Delta\boldsymbol{x}} + B_i\widehat{\Delta\boldsymbol{l}}_i = \boldsymbol{0}\,, \quad \mathbb{D}(\underline{\boldsymbol{l}}_i) = \Sigma_{l_i l_i}\,, \tag{88}$$

with

$$\boldsymbol{x} := \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}, \quad \boldsymbol{l}_i = \begin{bmatrix} x_i \\ y_i \\ x_i' \\ y_i' \end{bmatrix}, \tag{89}$$

and

$$\boldsymbol{g}_i := \boldsymbol{0}\,, \quad A_i^{\mathsf{T}} := \begin{bmatrix} -x_i & y_i & -1 & 0 \\ -y_i & -x_i & 0 & -1 \end{bmatrix}, \quad B_i^{\mathsf{T}} := \begin{bmatrix} -a & +b & 1 & 0 \\ -b & -a & 0 & 1 \end{bmatrix}. \tag{90}$$

The covariance matrix of the observations is assumed to be block diagonal, i.e., the observational groups are mutually uncorrelated. However, the covariance matrix of each group may contain arbitrary correlations. What is relevant in our context, the covariance matrix also may be singular, as long as the covariance matrix $\Sigma_{g_i g_i} = B_i^{\mathsf{T}}\Sigma_{l_i l_i}B_i$ is regular.

Hence we can simulate the model used in the previous section, where the coordinates (x_i, y_i) are assumed to be fixed, non-stochastic values by using the covariance matrix

$$\mathbb{D}(\boldsymbol{l}_i) = \mathbb{D}\left(\begin{bmatrix} \underline{x}_i \\ \underline{y}_i \\ \underline{x}'_i \\ \underline{y}'_i \end{bmatrix}\right) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma^2_{x'_i} & \sigma_{x'_i y'_i} \\ 0 & 0 & \sigma_{x'_i y'_i} & \sigma^2_{y'_i} \end{bmatrix}. \tag{91}$$

The estimation process does not need the inverse covariance matrix of the observations, therefore this way of modelling fixed observations does not lead to numerical difficulties.

The MATLAB function `demo_GHM_similarity.m` allows to simulate the result of the Gauss–Markov model with the boolean variable `simulate_GMM_similarity=true`, yielding the same result. Since the model is linear, only one iteration needs to be performed.