2.3 ROBUST ESTIMATION

2.3.1 Terminology and Criteria

We have seen in section 2.2.3 that the LS estimate is very susceptible to outliers. Nevertheless, the concept of outliers and their possibly deteriorating effect on the estimated parameters, is not limited to the theory of least squares estimation. Traces of strategies to handle outliers can be found in the literature going back in history as far as over 200 years. An early concept was obviously the *rejection* of observations *too wide off the truth*, Bernoulli (1777) cited in Beckman and Cook (1983). The dispute in the literature, also reviewed by Beckman and Cook (1983), was mainly about the criteria used to decide on acceptance or rejection. Glaisher (1873) seems to be one of the first to publish a procedure that *reduces the influence* of outliers on the results, rather than reject them. This may be the conceptual incipience of the theory of robust estimation.

Glaisher's idea that the observations may come from different populations and their initial weights have to be modified, fits well into Hampel's definition of *robust statistics*, Hampel et al. (1986, p. 7):

> *"Robust statistics, as a collection of related theories, is the statistics of approximate parametric models."*

According to this definition, robust statistics accepts that any model, e.g., of the type GMM (2.8), and any assumptions about the probability distributions involved, e.g., normally distributed errors, is only an approximation. Consequently, the effects of approximation errors are investigated, and methods are developed that are as little as possible affected by these errors.

Robust estimation, as a subset of robust statistics, is concerned with obtaining nearly optimum results in a whole "neighborhood" of the assumed model. This is the reason for the definition of outliers as explained in section 2.2.3. In the context of this thesis, robust estimation is applied to deal with approximation errors of the types eq. (2.35) and eq.(2.36). Note that many authors report percentages of outliers between 0.1% and 10% in real datasets, e.g., citations in Hampel et al. (1986). Caspary (1988b) indicates that 1% outliers may be typically found in geodetic data, and this referred to the pre-GPS era. Considering unmodeled systematic effects in GPS positioning, the percentage of outliers according to (2.35) may be much greater, see sec. 5.4.

Very useful and still up-to-date introductions into robust estimation are provided by Hampel et al. (1986), and Rousseeuw and Leroy (1987). For an outline of the history of robust estimation and a concise review of literature related to outliers and robust estimation until about 1980, see Beckman and Cook (1983). A

short but exceptionally valuable introduction with strong emphasize on geodetic applications is given by Caspary (1988b).

The definition of robust statistics, cited above, is almost too broad a concept, and estimators called *robust* may be found in literature, with actually very different properties. A number of robustness criteria are defined by Hampel et al. (1986). They shall be briefly discussed here, in order to distinguish between robust and non-robust estimators.

2.3.2 The Influence function

The starting point of Hampel's investigation of robustness is the notion of the *influence function* IF, Hampel (1974). An estimate $\hat{\theta}$ of a certain parameter θ, which is related to some observations y, is considered a function of the distribution $F(y)$ of the observations:

$$\hat{\theta} = T(F), \qquad \text{with} \quad y \sim F \tag{2.45}$$

As an example of $T(F)$, consider the estimation of the arithmetic mean $\overline{y}$ of some observations y by computing the first moment of their (assumed) distribution:

$$y \sim F \Rightarrow \hat{\overline{y}} = T(F) := \mu_F$$

with

$$\mu_F = \int_{-\infty}^{\infty} x \cdot f(x)\mathrm{d}x, \qquad \text{and} \quad f(x) = \frac{\mathrm{d}F}{\mathrm{d}x}$$

So, in this example, the functional $T(\bullet)$ is defined as follows:

$$T(\bullet) = \int_{-\infty}^{\infty} x \cdot \left(\frac{\mathrm{d}\bullet}{\mathrm{d}x}\right) \cdot \mathrm{d}x \tag{2.46}$$

Now, let $F(x)$ be an approximation only, and the true distribution of the observations be rather

$$G(x) = (1-\epsilon)F(x) + \epsilon H(x) \tag{2.47}$$

i.e., most of the observations actually belong to the distribution $F(x)$, but a small portion ϵ of the data comes from a *contamination distribution* $H(y)$. The question is, how does the estimate $T(F)$ change, if the true distribution G is considered instead of the approximation F? The answer is given by the influence function, which is defined such that

$$T(G) \approx T(F) + \int \mathrm{IF}_{T,F}(x)g(x)\mathrm{d}x, \qquad \text{with} \quad g(x) = \frac{\mathrm{d}G}{\mathrm{d}x} \tag{2.48}$$

This is accomplished by the pointwise definition of the asymptotic influence function of T at the underlying distribution F:

$$\mathrm{IF}_{T,F}(x) := \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left(T\left[(1-\epsilon) F + \epsilon \Delta_x \right] - T\left(F \right) \right), \quad \forall x \in \mathbb{R} \tag{2.49}$$

where Δ_x is the point mass distribution related to the Dirac Delta function δ by[4]

$$\Delta_x(t) = \int_{-\infty}^{t} \delta(x-u)\mathrm{d}u$$

Let us continue the example from above, in order to demonstrate the use of eqs. (2.49) and (2.48): the IF of the arithmetic mean $\overline{y}$ at a distribution F with first moment μ_F is computed as follows:

$$\begin{aligned} \mathrm{IF}_{\overline{y},F}(x) &= \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\left((1-\epsilon) \int_{-\infty}^{\infty} u f(u)\mathrm{d}u + \epsilon \int_{-\infty}^{\infty} u \delta(x-u)\mathrm{d}u \right) \right. \\ &\qquad \left. - \int_{-\infty}^{\infty} u f(u)\mathrm{d}u \right] \\ &= \lim_{\epsilon \to 0} \left[- \int_{-\infty}^{\infty} u f(u)\mathrm{d}u + x \cdot \int_{-\infty}^{\infty} \delta(x-u)\mathrm{d}u \right] \\ &= x - \mu_F \end{aligned} \tag{2.50}$$

So, if μ_F has been computed as estimate of the arithmentic mean of the observations y, but the true distribution of these observations turns out to be G instead of F, eq. (2.48) may be used to compute the correct estimate from the approximation by means of the influence functions. Substituting $T(F) = \mu_F$ and $\mathrm{IF}_{\overline{y},F}(x)$ as of eq. (2.50) into eq. (2.48), we obtain

$$T(G) \approx \mu_F + \int_{-\infty}^{\infty} (y - \mu_F) g(y)\mathrm{d}y = \int_{-\infty}^{\infty} y g(y)\mathrm{d}y = \mu_G \tag{2.51}$$

which is—in this case—precisely the arithmetic mean of the true distribution G. Note, that with more general estimators, $T(G)$ will only be approximated, as indicated by eq. (2.48).

[4]The symbol δ is used to denote the Dirac Delta function only in this section. It is not to be confused with the outlier, introduced in sec. 2.2.3.

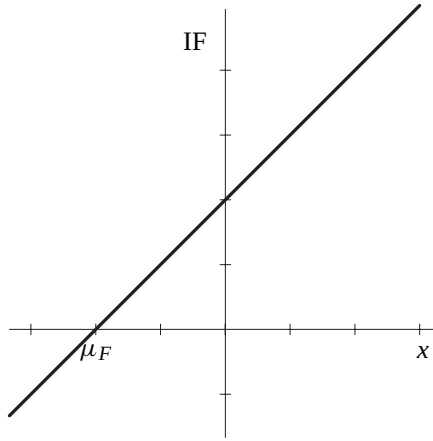

Figure 2.4 Asymptotic influence function $\mathrm{IF}_{\overline{y},F}(x)$ of arithmetic mean at underlying distribution F.

For estimators T_n based on samples of finite size, rather than on the underlying distribution function F, Hampel et al. (1986) mention a number of approximations to the asymptotic IF, which are mainly based on the idea of manipulating the samples and studying the resulting change of the estimate. As one example, I want to mention Tukey's *sensitivity curve*, which approaches the IF for $n \to \infty$:

$$\mathrm{SC}_{T_n}(y) = n\left[T_n\left(y_1, y_2, \ldots, y_{n-1}; y\right) - T_{n-1}\left(y_1, y_2, \ldots, y_{n-1}\right)\right] \qquad (2.52)$$

This shows, that the influence function may be considered a measure of the normalized effect of an additional observation on the value of the estimate. The IF of the arithmetic mean, computed in the example above, is plotted in fig. 2.4. We can derive from this figure, quite in accordance with experience, that an additional observation exactly equal to the mean μ_F of the other observations, does not change the artihmetic mean. Otherwise, the influence is proportional to the size of the additional observation, and if "$y = \infty$" is added, the arithmetic mean will be infinitely distorted.

An important result in studying the efficiency of estimators is the relation between the variance $\mathrm{D}\{T(F)\}$ of the estimate, and the IF, Hampel et al. (1986, p. 85),

$$\mathrm{D}\{T(F)\} = \int \mathrm{IF}^2_{T(F)}(x)\, f(x)\mathrm{d}x \qquad (2.53)$$

Hampel shows, that an estimator is *best*—in the sense of efficiency—, if the IF is proportional to the log-likelihood derivative corresponding to the distribution F. Although these results are beautiful, and give rise to many of the ideas presented in Hampel et al. (1986), we will not pursue them further, but use the influence function rather as a heuristic tool for the assessment of the robustness properties of an estimator.

The concept of the influence function is not restricted to the estimation of a single parameter using observations from a single distribution. Hampel has extended the theory, briefly presented above, to multivariate problems. Formally, the important results (2.48) and (2.53) remain valid, but of course the visualization is more difficult, since the IF is matrix valued in the general case.

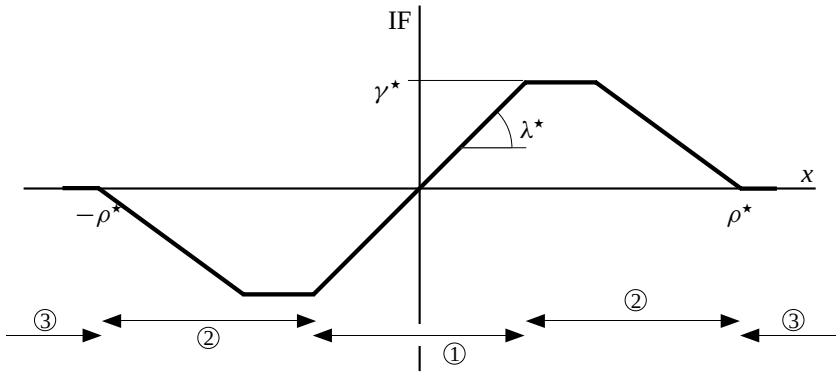

Figure 2.5 Asymptotic influence function and characteristics of a robust estimator

Gross-error sensitivity The *gross-error sensitivity* $\gamma^\star$ measures the worst influence a small amount of contamination of fixed size can have on the estimated value of a parameter. It is defined as the supremum[5] of the IF over all x, where IF exists, see fig. 2.5:

[5]The *supremum* over all elements of a set $\mathbb{M}$ is the minimum upper bound to these elements. As opposed to the maximum, the supremum need not be a member of this set:

$$s = \sup_{x \in \mathbb{M}} \quad \Leftrightarrow \quad s = \min\{\, t \in \mathbb{T} \,|\, t \geq x \;\forall x \in \mathbb{M}\}\,, \quad \text{with} \quad \mathbb{M} \subseteq \mathbb{T}$$

$$\gamma^\star = \sup_x |\,\mathrm{IF}_{T,F}(x)| \tag{2.54}$$

If $\gamma^\star$ is finite, the estimator T is said to be *B-robust*. This indicates, that the maximum bias of T, due to a small contamination of the data, is limited. So, T is robust with respect to *limited bias*. I shall subsequently call *robust*, a B-robust estimator. The comparison of the IF plotted in fig. 2.5 with the one of the arithmetic mean, fig. 2.4, shows that the former is a robust estimator, but the latter is not.

An existing estimator is usually robustified by putting an upper and lower bound to the influence function. We shall see below that this can actually be accomplished easily. However, usually there has to be found a tradeoff between low gross-error sensitivity and high efficiency, since both of them can not be optimized simultaneously.

Local-shift sensitivity The effect of a very small change of an observation, e.g., from x to $x + \epsilon$, on the estimate, may be expressed by the influence of removing the observation x and adding $x + \epsilon$. A standardized measure for this type of change in the observations is the *local-shift sensitivity* $\lambda^\star$, defined as

$$\lambda^\star = \sup_{x,\text{ and } \epsilon \neq 0} \frac{|\,\mathrm{IF}_{T,F}(x+\epsilon) - \mathrm{IF}_{T,F}(x)|}{|\epsilon|} \tag{2.55}$$

Because ϵ may be arbitrarily small, $\lambda^\star$ is the supremum of the slope of the IF, see fig. 2.5. The maximum slope of the IF is therefore a measure of the maximum influence of rounding errors and the likes, on the estimate.

Rejection point If the influence function decreases monotonically outside a certain region, the estimator is a *redescending* estimator, and the estimates are *mainly* based on the consistent observations. This idea may be extended such that the IF is constantly zero in certain regions. If there is a value $\rho^\star$ with

$$\mathrm{IF}_{T,F}(x) = 0, \qquad \forall\, |x| > \rho^\star$$

then this value is called *rejection point*. Observations beyond the rejection point have no influence at all on the estimate, i.e., the value of the estimate is *exclusively* based on the observations which lie within $[-\rho^\star,\ \rho^\star]$. Obviously, such robust estimators incorporate the idea of complete rejection of an observation, which is too far from the bulk of data, see sec. 2.2.3.

Starting with Bernoulli (1777) it has been stated by different authors, that a rejection of observations based on objective or subjective criteria is only acceptable if external evidence, i.e., information not extracted from the observations themselves, indicates that the observations are "bad". Otherwise it is better to treat outlying observations as very improbable, but still possible, samples of the assumed distribution, or as samples which belong to a contamination distribution as of eq. (2.47). Consequently, it is preferable to reduce their influence on the computed estimate rather than reject them, Beckman and Cook (1983). Robust estimators are designed for this purpose.

Breakdown point A strict definition of the *breakdown point* $\epsilon^\star$ may be found in Hampel et al. (1986, p. 97). Loosely speaking, it is the largest fraction of gross errors, that can never carry the estimate over all bounds. Intuitively it is clear that the breakdown point of an estimator will always be smaller than 50%. Actually, a breakdown point of nearly 50% is for example achieved by the sample median, but of course not by the arithmetic mean, for which $\epsilon^\star = 0$.

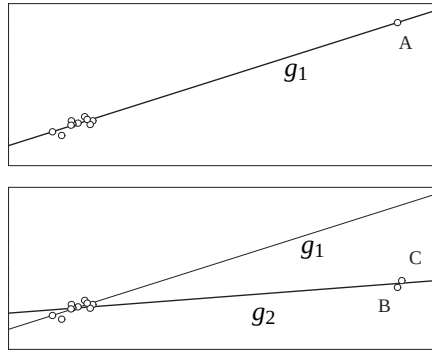

Figure 2.6 Leverage points in linear regression: good leverage point (A) enhances efficiency of estimate (top); outliers in highly influential points, masked by grouping (bottom, points B and C).

With parameter estimation in multivariate models, e.g., coordinate determination using GPS, it is difficult—if not impossible—to obtain a breakdown point different from 0%, in a strict mathematical sense. The reason is, that the geometry, i.e., the matrix $\boldsymbol{A}$ of the GMM, may render individual observations completely uncontrolled. In the sense of Baarda's reliability theory, these are observations with a redundancy number of 0. The maximum influence of such an observation

on the estimate is unbounded. More commonly, in a GMM there will be observations that are poorly controlled—having small, but non-zero (normalized) redundancy numbers. Such observations are called *leverage observations*. The notion hints at points far away from the majority of points in linear regression, which are called leverage points, such as point A, fig. 2.6, top.

Leverage observations are highly influential observations, which heavily distort the estimate, if they are false, but improve the efficiency dramatically, if they are correct. Special attention needs to be paid to these observations, if present. The situation is further complicated by *masking*, Beckman and Cook (1983), which is the situation of a few leverage observations grouped together, see points B and C in fig. 2.6, bottom. Such observations will not stand out by exceptionally low redundancy numbers, and consequently, they are very difficult to detect in a multi-parameter model—where a simple plot of the observations may not be possible. These problems may adversely affect an estimator, which is robust and has a non-zero breakdown point in the one dimensional case, but may have a breakdown point of $\epsilon^\star = 0$ in the extension to a multi-dimensional problem, e.g., the Huber-estimator.

On the other hand, with real world data, infinitely large errors do not occur or are easily detected beforehand, so that an estimator with $\epsilon^\star = 0$ may still be sufficiently robust for many applications.

2.3.3 Definition and realization of robust estimators

M-estimators Estimators may be defined in a variety of ways. Hampel et al. (1986) give a good overview about classes of robust estimators. I shall focus on the *M-estimators*, introduced by Huber (1964). Many robust estimators found in the literature belong to this class or are derived from it. In the field of geodesy, see e.g., Kubik (1982), Caspary and Borutta (1987), Yang (1994), Koch (1996), Koch and Yang (1998), Wicki (1998). Some of the reasons for the popularity of M-estimators may be the conceptual beauty and the ease of realization by means of a well known formalism.

Huber defined M-estimation as a generalization of maximum likelihood estimation, e.g., Koch (1999), by the following minimum problem:

$$\sum_{i=1}^{n} \rho(y_i, \boldsymbol{\xi}) = \min_{\boldsymbol{\xi}}! \tag{2.56}$$

where ρ is the so-called *loss function*, and y_i are the uncorrelated observations. The minimum problem (2.56) is solved by the M-estimate $\hat{\boldsymbol{\xi}}$. If the loss function

is the negative logarithm of the likelihood of y_i, the solution of (2.56) is exactly the maximum likelihood estimate. More generally a variety of estimators, e.g., robust estimators, with different properties may be defined by choosing a suitable loss function

$$\rho : \mathbb{R} \mapsto \mathbb{R}$$

Usually, the argument of the loss function is the corresponding residual divided by the standard deviation of the observation, i.e.,

$$\rho(y_i, \boldsymbol{\xi}) = \rho\left(\frac{e_i}{\sigma_i}\right) = \rho\left(\frac{\boldsymbol{a}_i'\boldsymbol{\xi} - y_i}{\sigma_i}\right) \tag{2.57}$$

where $\boldsymbol{a}_i'$ is the i-th row of the matrix $\boldsymbol{A}$ and $\sigma_i = \sigma\sqrt{\boldsymbol{V}(i,i)}$ from the GMM (2.8). The standardization by σ_i is necessary to provide a scale invariant estimate, Hampel et al. (1986, p. 105).

The solution of the minimum problem (2.56) is obtained by the necessary condition that the first derivation of the sum w.r.t. $\boldsymbol{\xi}$ vanishes. This yields the system of normal equations with one equation for each of the u parameters:

$$\sum_{i=1}^{n} \psi\left(\frac{e_i}{\sigma_i}\right)\frac{1}{\sigma_i}a_{ij} = 0, \qquad \forall\, j \in \{1, 2, \ldots, u\} \tag{2.58}$$

with

$$\psi(x) = \frac{\mathrm{d}\rho}{\mathrm{d}x}, \qquad \text{and} \quad a_{ij} = \boldsymbol{A}(i, j)$$

Note, that the eqs. (2.58) are equally fulfilled, if ψ is multiplied by an arbitrary non-zero factor. Any estimator defined by $\rho(e_i)$ or $\psi(e_i)$ is an M-estimator. If, and only if, $\psi(x)$ is bounded, the corresponding M-estimator is robust, as follows from the important equation

$$\mathrm{IF}_{\psi,F}(x) \quad \propto \quad \psi(x) \tag{2.59}$$

see Hampel et al. (1986, p. 103). So, if a certain loss function ρ is chosen, the robustness properties of the corresponding estimator can immediately be investigated via ψ, the first derivative of ρ. And if, on the other hand, a certain influence function IF* is to be realized, we can immediately name a suitable estimator by $\psi^\star(x) := \mathrm{IF}^\star(x)$. Different robust M-estimators, their influence functions, and their characteristics are extensively discussed in Huber (1981) and Hampel et al. (1986).

An important property of the M-estimates is the fact, that they can be computed using the LS formalism. The least squares estimator, discussed in sec. 2.2, is defined by the loss function

$$\rho(x) := x^2 \tag{2.60}$$

which yields the ψ-function

$$\psi(x) = 2x \tag{2.61}$$

and, with $p_i \propto \sigma_i^{-2}$, the well known normal equations

$$\sum_{i=1}^{n} a_{ij}\, p_i e_i = 0, \qquad \forall\, j \in \{1, 2, \ldots, u\} \tag{2.62}$$

Huber (1981, pp. 183), showed that arbitrary M-estimators need not be computed by directly solving the corresponding normal equations (2.58), but may be realized by iteratively reweighted LS (RLS). The corresponding *equivalent weights* w are a function of the (scaled) residuals and depend on the ψ-function. Comparison of equations (2.62) and (2.58) shows that indeed the solution of (2.58) is obtained through

$$\sum_{i=1}^{n} a_{ij}\, w_i e_i = 0, \qquad \forall\, j \in \{1, 2, \ldots, u\}, \tag{2.63}$$

if

$$w_i = p_i \frac{\psi\,(e_i/\sigma_i)}{e_i/\sigma_i} \tag{2.64}$$

Tukey (1970) calls these types of estimators *W-estimators*. Actually, the idea of reweighting observations based on the size of the residuals, was first published by Glaisher (1873) and Newcomb (1886). It was then not further pursued, until P.J. Huber "set the ball rolling" again in 1964. He reanimated the idea of a contamination distribution, see eq. (2.47), and designed M-estimates in order to assign low weights to the observations originating from the contamination.

Extending eqs. (2.63) and (2.64) to a more general notation yields

$$\hat{\boldsymbol{\xi}} = \left(\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\right)^{-1} \boldsymbol{A}'\boldsymbol{W}\boldsymbol{y} \tag{2.65}$$

where

$$\boldsymbol{W} = \boldsymbol{W}(\tilde{\boldsymbol{e}}) \tag{2.66}$$

Since the equivalent weights depend on the residuals, which are initially unknown, the RLS solution is found iteratively, as indicated by fig. 2.7: an initial weight

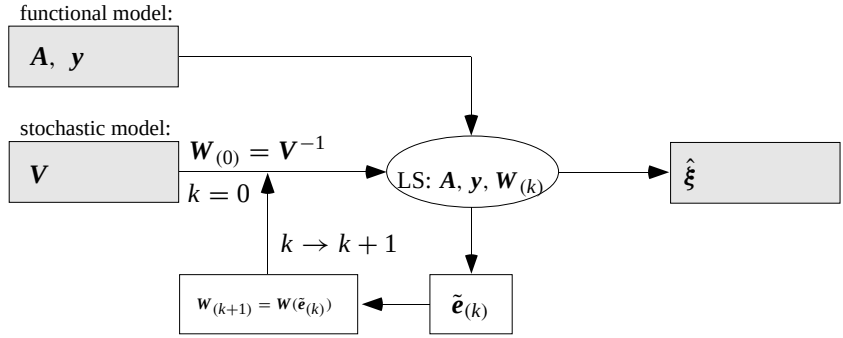

Figure 2.7 Flow chart of reweighted least squares (RLS) using an equivalent weight matrix $\boldsymbol{W}$.

matrix $\boldsymbol{W}_{(0)}$ is computed from the a-priori cofactor matrix of the observations. LS estimation using $\boldsymbol{W}_{(0)}$ and the functional model eq. (2.10), yields the predicted residuals $\tilde{\boldsymbol{e}}_{(0)}$. According to the ψ-function selected, new weights are computed by eq. (2.64) and collected in the updated weight matrix $\boldsymbol{W}_{(1)}$. Using this matrix, the predicted residuals are recomputed. This process of reweighting and LS estimation is repeated until convergence is achieved, i.e., until the maximum change of weights is below a certain threshold. Then the a-posteriori weight matrix $\boldsymbol{W} = \boldsymbol{W}_{(N)}$ is obtained as the weight matrix of the last (N-th) iteration performed, and the robust RLS estimate is computed by LS using this weight matrix.

The benefits of RLS are mainly the use of a simple, well implemented and well understood algorithm to compute the estimates, and the flexibility it offers for modifications, as explained in the next section. On the other hand, the computational burden increases dramatically, if the number of iterations is high. Furthermore, while other algorithms may reveal that the solution is not unique, and indicate disjunct groups of observations matching equally well, RLS may either not converge or yield one of the solutions only, Koch (1996).

Kubik (1982) has given an example of L_1-estimation, i.e., M-estimation with $\rho(x) = |x|$, applied to linear regression, with several possible optimum solutions. RLS converges to one of these solutions only, as opposed to a simplex algorithm, that correctly yields several different regression lines, all of which produce the same minimum sum of absolute residuals.

Usually the a-priori weight matrix should be a good approximation for the majority of the data, and the approximation of the parameters should be close to their true values, in order to facilitate convergence.

Danish Method One of the most remarkable consequences of eqs. (2.63) and (2.64) is, that robust estimators may not only be realized but directly defined by a weight function $w(e_i)$. The relation to the corresponding ψ function is established by eq. (2.64), and ρ and IF can be derived therefrom.

The so-called Danish Method (DM), Krarup et al. (1980), is a well established RLS estimator that has been used in geodetic applications for a long time, e.g. Kubik (1982), Jørgensen et al. (1985), Berber and Hekimoglu (2001). The DM is defined by the equivalent weights

$$w_i = \begin{cases} 1 & \text{for} \quad |\tilde{e}_i| \leq 2\sigma \\ \alpha \exp(-\beta \tilde{e}_i^2) & \text{for} \quad |\tilde{e}_i| > 2\sigma \end{cases} \tag{2.67}$$

which are computed after an initial LS estimation.

In eq. (2.67), σ is the a-priori standard deviation of the independent identically distributed (i.i.d.) observations, and α and β are suitably chosen constants. Fig. 2.8 shows the structure of the influence function of the DM. The gross-error sensitivity is bounded, so actually the DM is robust. Furthermore, the influence of observations outside $\pm 2\sigma$ decreases very rapidly and approaches 0 as an observation tends towards infinity. So the estimator is a redescending estimator, and, although there is no rejection point, observations far from the center of the distribution have virtually no influence on the estimate. Since the IF is discontinuous at $\pm 2\sigma$, the local-shift sensitivity is unbounded, which is one of the criticisms towards application of the DM. $\lambda^\star = \infty$ means, that e.g., rounding errors in the observations may cause large changes in the influence of an observation. On the other hand, this problem is limited to the values $\tilde{e}_i = 2\sigma$ and $\tilde{e}_i = -2\sigma$. Everywhere else, the IF is continuous. If desired, the constants α and β can be determined such that it is continuous everywhere.

It is difficult to attribute a clear statistical meaning to the DM, especially if extended to a multivariate model, Jørgensen et al. (1985). However, as may be seen from fig. 2.8, the DM estimate is mainly determined by the non-outlying observations. According to Kubik (1982), it is computed from the largest consistent subset of observations. In Kubik (1988), this argument is investigated, and the DM is explained as a multivariate extension of the trimmed mean[6].

The DM is very robust w.r.t. few outliers in the observations, Krarup et al. (1980), Caspary and Borutta (1987), Berber and Hekimoglu (2001), and it usually converges quickly, Jørgensen et al. (1985), which is one of the reasons

[6]The α-trimmed mean of a sample of size N is obtained by removing the $\frac{\alpha}{2}N$ largest and the $\frac{\alpha}{2}N$ smallest sample values, and computing the mean of the remaining $(1-\alpha)N$ values.

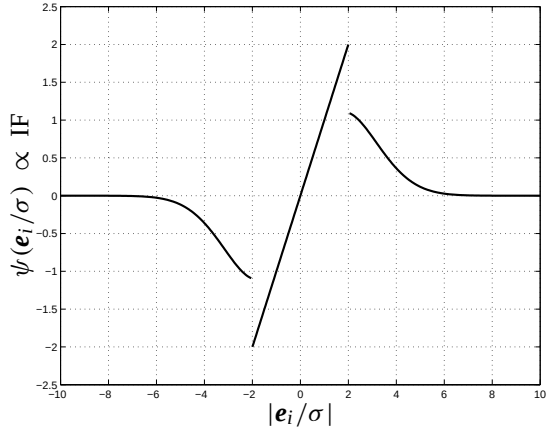

Figure 2.8 Influence function of Danish Method with $\alpha = 1$, $\beta = 0.15$, $\sigma = 1$.

why it is attractive. Furthermore, its estimates are equal to the LS estimates, if the data contain no outliers.

Unfortunately, the DM is not applicable to GPS DD processing, since the phase observations are heterogeneous and correlated—as opposed to the i.i.d. assumption of the DM. Several authors have presented modifications of the DM that account for the heterogeneity of the data, by allowing individual a-priori weights for the observations, e.g. Kubik (1982), Berberan (1992). Most of the texts on robust estimation are, implicitly, on *independent data*, and many of the concepts of robust statistics still lack a well founded extension to dependent observations.

Furthermore, most robust estimators do not take the local redundancies into account, which makes them susceptible to (bad) leverage observations. In Jørgensen et al. (1985), we find a hint on a possible improvement of the DM, using the variance of the residuals instead of σ in eq. (2.67). A similar approach is suggested by Berberan (1992), who uses a *hybrid method* to deal with leverage observations. He combines robust estimation by equivalent weights with data snooping, see sec. 2.2.3. Again, both methods are restricted to uncorrelated observations.

2.3.4 Robust estimation with correlated observations

Correlated observations can easily be transformed into uncorrelated ones by homogenization, e.g., Niemeier (1979): if $\mathrm{D}\{\boldsymbol{e}\} = \sigma^2 \boldsymbol{V}$, and $\boldsymbol{V}$ is non-diagonal,

but of full rank, the linear transformation

$$\boldsymbol{y}^\star = \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{y}, \quad \text{with} \quad \boldsymbol{V}^{-\frac{1}{2}}\boldsymbol{V}\boldsymbol{V}^{-\frac{1}{2}} = \boldsymbol{I} \tag{2.68}$$

yields uncorrelated (and equally distributed) observations $\boldsymbol{y}^\star$. So, this process may also be considered a decorrelation of the observations. However, if $\boldsymbol{V}$ is a fully populated matrix, each decorrelated observation $y^\star$ is a linear combination of *all* original observations, and outliers are spread over all observations by decorrelation. Even a single original outlier affects the majority of the decorrelated data, therefore. No robust estimator can handle this situation. Consequently, decorrelation of the observations and subsequent application of an arbitrary robust estimator is no reasonable approach to robust estimation with correlated observations.

Only few publications on robust estimation address the problem of correlation, e.g., Xu (1989), Yang (1994).

Bivariate robust functions Xu (1989) presents the extension of several well known robust estimators by *bivariate robust functions* $\rho(e_i, e_j)$, which replace the loss function of M-estimators. Xu also suggests an RLS estimator similar to the DM, which can be used with correlated observations. The entries $w_{i,j}$ of the equivalent weight matrix of the observations are computed by *bivariate robust functions* of the following type:

$$w_{i,j} = \begin{cases} w(\tilde{e}_i) & \text{if } i = j, |\tilde{e}_i| > c \\ p_{i,j} & \text{if } i \neq j \end{cases} \tag{2.69}$$

Only the diagonal entries of the fully populated weight matrix are modified according to this scheme. With the restriction $w(e_i) \geq p_{i,i}$ on the weight function, this always yields a positive definite weight matrix. It is clear from Xu's suggestions, that the definition of a robust estimator by an equivalent weight function is actually most flexible. Above all, the *weight matrix* may be defined as a function of the *vector of residuals*, see eq. (2.66), thus allowing for the handling of correlation. However, weight functions of type (2.69) rather *ignore* correlation than *incorporate* it, if no further justification for the modification of only the diagonal entries is provided by the correlation pattern.

IGG-3 scheme The IGG-3 scheme, a weight function proposed by Yang (1994), modifies diagonal and off-diagonal entries of $\boldsymbol{P}$, but yields a (slightly) non-symmetric weight matrix. Yang et al. (2001) propose an improved method, again based on a direct definition of a weight matrix, but one that preserves

symmetry *and* correlation structure. According to this scheme, the correlation coefficient, related to the elements of the cofactor matrix $\overline{\boldsymbol{V}}$ by

$$\rho_{i,j} = \frac{\overline{\boldsymbol{V}}(i,j)}{\sqrt{\overline{\boldsymbol{V}}(i,i)\overline{\boldsymbol{V}}(j,j)}} \qquad \text{with} \quad \overline{\boldsymbol{V}} = \boldsymbol{W}^{-1} \tag{2.70}$$

is equal before and after the modification of the weights. This is a reasonable condition, if the correlations described by the cofactor matrix are of *physical* nature.

Correlation patterns Physical correlation between the observations is independent of the variance of the observations. Assume, for example, that n observations are made at different times, and that they are temporally correlated. The correlation is modeled by a correlation matrix $\boldsymbol{C}$ that contains the individual correlation coefficients $\rho_{i,j} = \boldsymbol{C}(i,j)$. If the n-vector $\boldsymbol{\sigma}$ of the standard deviations of the observations is given, the variance-covariance matrix of the correlated observations follows from

$$\boldsymbol{\Sigma} = \boldsymbol{C} * (\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}') \tag{2.71}$$

and has the structure

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \dots & \rho_{1,n}\sigma_1\sigma_n \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \rho_{1,n}\sigma_1\sigma_n & \dots & & \sigma_n^2 \end{bmatrix} \tag{2.72}$$

With $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{V}$ it follows immediately from (2.72), that

$$\frac{\boldsymbol{V}(i,j)}{\sqrt{\boldsymbol{V}(i,i)\boldsymbol{V}(j,j)}} = \frac{\boldsymbol{\Sigma}(i,j)}{\sqrt{\boldsymbol{\Sigma}(i,i)\boldsymbol{\Sigma}(j,j)}} = \frac{\rho_{i,j}\sigma_i\sigma_j}{\sqrt{\sigma_i^2\sigma_j^2}} = \rho_{i,j}$$

in correspondence with eq. (2.70). This means that the correlation coefficient computed from the corresponding elements of $\boldsymbol{\Sigma}$ or $\boldsymbol{V}$ is $\rho_{i,j}$, no matter how small or how large σ_i and σ_j are.

If the correlations are *purely algebraic*, i.e., introduced by processing *functions* of the original observations, the correlation depends on the individual variances, and it may not be reasonable to impose an invariance condition on the correlation coefficient. This is demonstrated using the $(n-1)$ linearily independent

differences $\overline{y}$ of n uncorrelated observations y. Let

$$\mathrm{D}\{y\} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & & \sigma_n^2 \end{bmatrix}, \qquad \text{and} \quad \overline{\boldsymbol{y}} = \begin{bmatrix} y_1 - y_n \\ \vdots \\ y_{n-1} - y_n \end{bmatrix}$$

i.e.,

$$\overline{\boldsymbol{y}} = \boldsymbol{F}\boldsymbol{y}, \qquad \text{with} \quad \boldsymbol{F} = \begin{bmatrix} 1 & 0 & \dots & & -1 \\ 0 & 1 & & & \vdots \\ \vdots & & \ddots & & \\ 0 & & \dots & 1 & -1 \end{bmatrix}$$

The variance-covariance matrix of the derived observations $\overline{\boldsymbol{y}}$ is then obtained by variance propagation. It reads

$$\boldsymbol{\Sigma}_{\overline{y}} = \begin{bmatrix} (\sigma_1^2 + \sigma_n^2) & \sigma_n^2 & \sigma_n^2 & \dots & \sigma_n^2 \\ \sigma_n^2 & (\sigma_2^2 + \sigma_n^2) & \sigma_n^2 & & \vdots \\ \vdots & & & \ddots & \\ \sigma_n^2 & \sigma_n^2 & \dots & & (\sigma_{n-1}^2 + \sigma_n^2) \end{bmatrix} \tag{2.73}$$

and the correlation coefficient computed from this matrix is

$$\frac{\boldsymbol{\Sigma}_{\overline{y}}(i,j)}{\sqrt{\boldsymbol{\Sigma}_{\overline{y}}(i,i)\boldsymbol{\Sigma}_{\overline{y}}(j,j)}} = \frac{\sigma_n^2}{\sqrt{\sigma_i^2 + \sigma_j^2 + 2\sigma_n^2}}, \qquad i \neq j$$

So, in this case, the correlation depends on the variances, and it changes if at least one of σ_i, σ_j or σ_n changes. Consequently, if under a variance inflation model, the variance of an observation is modified, the correlation must be modified as well. On the other hand, this may be accomplished implicitly by a robust estimator, if, instead of the weights, the *equivalent variances* are computed and the variance-covariance matrix is composed using the law of variance propagation. However, assumptions about the initial variances $\boldsymbol{\sigma}$ will be necessary to compute their inflation as a function of the residuals of the derived, i.e., correlated, observations $\overline{\boldsymbol{y}}$.

If the correlation structure consists of an algebraic *and* a physical component, the variance-covariance matrix of the observations is formally given by

$$\boldsymbol{\Sigma}_{\overline{y}} = \boldsymbol{F}\left(\boldsymbol{C} * \left(\boldsymbol{\sigma} \cdot \boldsymbol{\sigma}'\right)\right)\boldsymbol{F}' \tag{2.74}$$

Basically, this general type of correlations might also be handled with robust estimation by equivalent variances and recomposition of the corresponding dispersion matrix according to eq. (2.74), but it may be very difficult to derive proper inflation factors of the original—only physically correlated—observations. To my knowledge, there are currently no publications treating this problem.

RLSCO The correlation pattern of GPS DD observations, considered in this text, is exactly of type (2.73). A suitable robust estimator for parameter estimation with GPS DD has been presented in Wieser and Brunner (2000) and Wieser and Brunner (2001). It is a reweighted least squares estimator for correlated observations (RLSCO), realized as a modification of the Danish method.

According to the variance inflation model, the variance of outlying observations is increased in order to achieve consistency between the model and the observations. Possibly outlying observations are indicated by the test statistic for outlier detection with correlated data, see eq. (2.43). Furthermore, T_i includes the underlying geometry via $\boldsymbol{V}_{\tilde{\boldsymbol{e}}}$, and the local redundancies are considered by an estimator, if the variance inflation is based on T_i.

The equivalent cofactor matrix for the k-th iteration of RLSCO is determined as follows:

$$\overline{v}_{i,j\,(k+1)} = \begin{cases} v_{i,j} \exp\left(\frac{|T_{i\,(k)}|}{c}\right) & i = j,\ |T_{i\,(k)}| > c \\ v_{i,j} & \text{else} \end{cases} \tag{2.75}$$

and

$$\overline{\boldsymbol{V}}_{(k+1)} = \begin{bmatrix} \overline{v}_{1,1\,(k+1)} & \overline{v}_{1,2\,(k+1)} & \cdots \\ \overline{v}_{1,2\,(k+1)} & \overline{v}_{2,2\,(k+1)} & \\ \vdots & & \end{bmatrix}$$

If the true errors are normally distributed, T_i has a standard normal distribution under the null hypothesis—no outlier in the i-th observation. In this case, c is the quantile of the standard normal distribution at the corresponding level of significance. However, the true distribution of the errors is usually *not* known. So, as with the DM, a fixed threshold value may be used, e.g., $c = 3$.

Fig. 2.9 is a flow chart of the RLSCO procedure. The iterations start with an LS estimation using the inverse of the a-priori cofactor matrix $\boldsymbol{V}$ as weight matrix, eqs. (2.76) and (2.76). In each iteration k the test statistic $T_{i\,(k)}$ is computed for all observations using the predicted residuals of the current iteration, eqs. (2.78) and (2.79). Then the equivalent variances for the next iteration are computed according to eq. (2.75), and another LS estimation is performed.

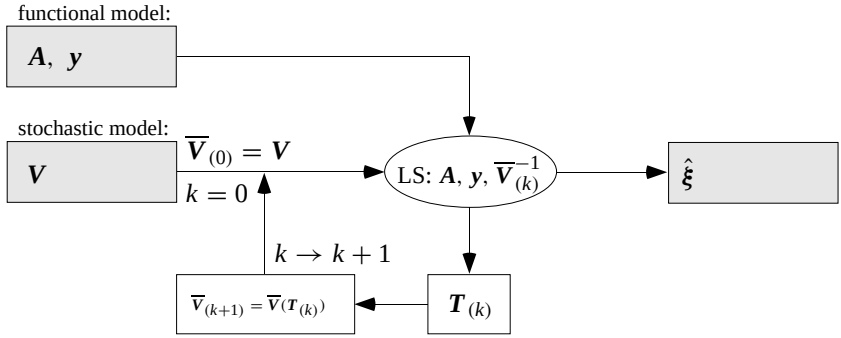

Figure 2.9 Flow chart of RLSCO.

$$\overline{\boldsymbol{V}}_{(0)} = \boldsymbol{V} \tag{2.76}$$

$$\overline{\boldsymbol{P}}_{(k)} = \overline{\boldsymbol{V}}_{(k)}^{-1} \tag{2.77}$$

$$\tilde{\boldsymbol{e}}_{(k)} = \left(\boldsymbol{A} \left(\boldsymbol{A}' \overline{\boldsymbol{P}}_{(k)} \boldsymbol{A} \right)^{-1} \boldsymbol{A}' \overline{\boldsymbol{P}}_{(k)} - \boldsymbol{I} \right) \boldsymbol{y} \tag{2.78}$$

$$T_{i\,(k)} = \frac{\boldsymbol{\eta}'_{(i)} \overline{\boldsymbol{P}}_{(k)} \tilde{\boldsymbol{e}}_{(k)}}{\sigma \sqrt{\boldsymbol{\eta}'_{(i)} \overline{\boldsymbol{P}}_{(k)} \boldsymbol{V}_{\tilde{\boldsymbol{e}}} \overline{\boldsymbol{P}}_{(k)} \boldsymbol{\eta}_{(i)}}} \tag{2.79}$$

The iterations are terminated, once the equivalent cofactors do not change any more significantly, which is usually achieved after less than 10 iterations. The robust estimate is then computed using the last (N-th) equivalent weight matrix:

$$\hat{\boldsymbol{\xi}} = \hat{\boldsymbol{\xi}}_{(N)} = \left(\boldsymbol{A}' \overline{\boldsymbol{P}}_{(N)} \boldsymbol{A} \right)^{-1} \boldsymbol{A}' \overline{\boldsymbol{P}}_{(N)} \boldsymbol{y} \tag{2.80}$$

As can be seen from eq. (2.75), only the diagonal elements of the cofactor matrix are modified during the iterations of RLSCO. This is justified by the special structure of the variance-covariance matrix—and the cofactor matrix—of the GPS DD observations, see sec. 4.3. The structure corresponds to that of eq. (2.73), where y_n is the single differenced reference satellite observation, and y_i with $i \neq n$ refers to the other satellites. The off-diagonal elements depend on the variance of the reference satellite observations only. The diagonal elements, on the other hand, depend on the variances of the observations of

both satellites involved in a double difference observation, see also eq. (2.73). So, if the reference satellite is chosen carefully, we may attribute the necessary variance inflation to the second satellite only. In this case, as can be seen from the structure of the cofactor matrix in eq. (2.73), only the diagonal of the matrix has to be modified, in order to preserve the correct correlation structure. On the other hand, if the reference satellite observations are outliers, the procedure will fail. Sec. 4.4 is devoted to reference satellite selection, and its influence on the processing results. An algorithm for the automatic selection of the reference satellite will be presented in sec. 3.4.

Note, that the cofactor matrix $\boldsymbol{V}_{\tilde{\boldsymbol{e}}}$ of the residuals is not updated during the iterations, see eq. (2.79). This is necessary for several reasons, although it deprives T_i of its strict statistical significance, after the first iteration. If $\boldsymbol{V}_{\tilde{\boldsymbol{e}}}$ were correctly computed after each iteration, RLSCO would not converge: a high value of T_i, say $T_{i,(k)} > c$, requires variance inflation. If the observation is actually an outlier, the reduction of its influence will yield a higher absolute residual, and perhaps even a higher value of T_i than before. If this happens, the variance will be further inflated. Once, the inflation is sufficient, i.e., the a-priori variance of the observation is low enough to explain its discrepancy w.r.t. the other observations by chance, T_i will not indicate an outlier anymore, if $\boldsymbol{V}_{\tilde{\boldsymbol{e}}}$ is correctly recomputed after each iteration. So, there will be an iteration $l > k$, when $T_{i,(l)} < c$. According to eq. 2.75, the equivalent cofactor would consequently be reset to its inital value, and $T_{i,(l+1)} > c$ would immediately result, i.e., RLSCO would not converge, if there is at least one observation with a high value of T.

A possible work-around would be, to update the cofactor matrix of the previous iteration rather than the a-priroi cofactor matrix, i.e., replace $v_{i,j}$ by $\overline{v}_{i,j(k)}$ in eq. (2.75). So, a low value of T_i in an iteration would just mean that the variance need not be *further* inflated. However, with this procedure, an observation which was subject to variance inflation once, can never obtain a low variance again—even if it is actually a good observation. In view of these problems, T_i is computed as of eq. (2.79), using a constant matrix $\boldsymbol{V}_{\tilde{\boldsymbol{e}}}$.

Wicki (1998), who proposes a generalized M-estimator (BIBER) for use with geodetic networks, takes a similar decision. The BIBER estimator is based on the normalized residuals (of uncorrelated observations). BIBER may be realized by different iterated algorithms, all of which require that the normalization be performed by constant values, which do not change during the iterations, (p. 102, ibid). With BIBER, these normalization constants are the initial standard deviations of the predicted residuals.

Under the assumption of a variance-inflation model, see sec. 2.2.3, the residuals of outlying observations are still distributed with expectation $\mathbf{0}$, but with an increased variance. RLSCO inflates the variances of these observations accordingly. So, the a-posteriori variance factor and the cofactor matrix of the parameters may be estimated by

$$\hat{\sigma}^2 = \frac{\tilde{\boldsymbol{e}}' \overline{\boldsymbol{P}}_{(N)} \tilde{\boldsymbol{e}}}{f}, \qquad \text{with} \quad f = n - u \tag{2.81}$$

$$\boldsymbol{V}_{\hat{\boldsymbol{\xi}}} = \left(\boldsymbol{A}' \overline{\boldsymbol{P}}_{(N)} \boldsymbol{A} \right)^{-1} \tag{2.82}$$

where f is the degree of freedom.

The (inflated) a-priori variances $\overline{\sigma}_y^2$ of the individual observations can be computed from the equivalent cofactor matrix of the final iteration by

$$\overline{\sigma}_{y_i}^2 = \sigma^2 \overline{\boldsymbol{V}}_{(N)}(i,i) \tag{2.83}$$

This equation can also be used to track how RLSCO modifies the individual a-priori standard deviations of the observations during its iterations. Fig. 2.10 shows this for the example of a single epoch of L1 GPS DD phase observations, and tab. 2.1 lists the corresponding T_i values during the iterations. The initial variances are computed using the SIGMA-ϵ variance model, see sec. 5.5.1. The standard deviation is a few mm for each of the seven DD observations, see fig. 2.10 (iteration 1). Two of the observations yield high values of T_i: PRN2 and PRN7, see table 2.1. Correspondingly, the variances of these observations are inflated by RLSCO. However, after iteration 2, only the observation involving PRN7 still stands out by a high value of T—actually, T is even higher than before. So, the variance of the PRN7 observation is inflated further ($\overline{\sigma}_y \approx 40$ mm), while the variance of the PRN2 observation is reset to its low initial value, because the corresponding value of T is below the threshold $c = 3$. The situation does not change much after the next iteration. The maximum change of the cofactors is below 0.1% after the fourth iteration, and RLSCO terminates.

The following characteristics of RLSCO can be seen from this example:

- The variances of all suspicious observations are inflated simultaneously, which helps in quickly reducing the influence of outlying observations (that need not correspond to the largest values of T). Note that with traditional outlier detection techniques, and some robust estimation procedures, e.g., Wicki (1998), only the observation with the largest value of the test statistic is considered for rejection or variance inflation during each iteration step, see sec. 2.2.3.

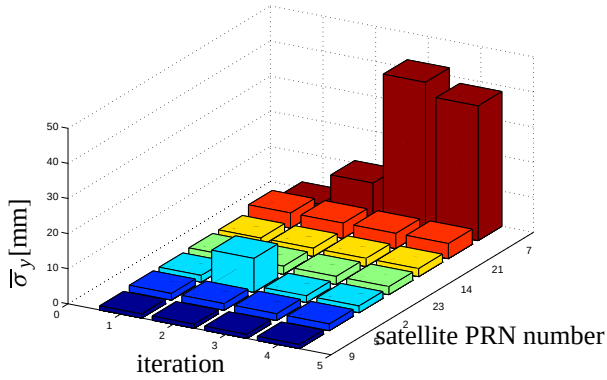

Figure 2.10 A-priori standard deviation $\overline{\sigma}_y$ of GPS DD observations, as assigned by the variance function during iterations of RLSCO.

- The inflated variances may be reduced or reset to their initial values in subsequent iteration steps. Observations erroneously considered outliers, may therefore be added to the group of consistent observations again. This backward testing is not routinely applied to conventional outlier detection techniques, see sec. 2.2.3, but is essential for maintaining good efficiency.

From table 2.1 the success of RLSCO in the example is also evident from the estimated variance factor (last column). While its initial value of 0.011 clearly indicates an erroneous model, the final value matches the a-priori variance factor $\sigma = 0.002$.

Table 2.1 Values of T of DD GPS phase observations, and a-posteriori variance during iterations (It) of RLSCO ($\sigma = 0.002$).

	test statistic T							
It	PRN 9	PRN 5	PRN 2	PRN 23	PRN 14	PRN 21	PRN 7	$\hat{\sigma}$
1	0.09	1.93	9.77	2.81	1.80	0.71	11.23	0.011
2	0.61	0.14	0.85	1.22	1.92	0.99	19.53	0.003
3	0.52	0.32	0.20	1.13	1.77	1.09	18.96	0.002
4	0.52	0.32	0.19	1.13	1.77	1.09	18.95	0.002

RLSCO is a robust estimator, similar to the Danish Method, but applicable to heterogeneous and correlated observations. Its influence function depends on the geometry of the estimation problem, since the equivalent weights are computed using the cofactor matrix of the residuals. Its influence function can only be plotted beforehand, if the matrix $\boldsymbol{A}$ and the initial variances are given. However, its shape is similar to the one shown in fig. 2.8, with a discontinuity at $|T| = c$. This discontinuity does not cause any practical problems.

In the case of mathematically correlated GPS DD phase observations with a single, well chosen reference satellite per epoch, RLSCO preserves the correct correlation structure. The procedure converges rapidly, and it is able to identify multiple outliers. Furthermore, it takes the redundancy contribution of the individual observations into account when identifying outliers. It is therefore robust w.r.t. outliers in weakly controlled observations as well. If no outlying observations are contained in the data, the RLSCO results are equal to the LS results.

RLSCO-2 If the cofactor matrix $\boldsymbol{V}$ of the observations contains physical correlations, these correlations should be preserved, see page 38. In this case, RLSCO has to be slightly modified, following the idea of Yang et al. (2001).

First, the correlation matrix $\boldsymbol{C}$ is computed from the initial cofactor matrix $\boldsymbol{V}$:

$$c_{i,j} = \frac{\boldsymbol{V}(i,j)}{\sqrt{\boldsymbol{V}(i,i)\boldsymbol{V}(j,j)}}, \quad \text{and} \quad \boldsymbol{C} = \begin{bmatrix} 1 & c_{1,2} & c_{1,3} & \cdots \\ c_{1,2} & 1 & c_{2,3} & \\ c_{1,3} & c_{2,3} & 1 & \\ \vdots & & & \ddots \end{bmatrix} \tag{2.84}$$

Then the equivalent weight matrix of the first iteration is computed as before, using eqs. (2.76) and (2.76). An LS estimation yields the initial residuals and T_i values. The cofactor matrix for the next iteration $(k+1)$ is computed in a two-step procedure: the diagonal elements of $\overline{\boldsymbol{V}}_{(k+1)}$ are calculated as before:

$$\overline{v}_{i,i\,(k+1)} = \begin{cases} v_{i,i} \exp\left(\frac{|T_{i\,(k)}|}{c}\right) & |T_{i\,(k)}| > c \\ v_{i,i} & \text{else} \end{cases} \tag{2.85}$$

but the total cofactor matrix is computed from the correlation matrix and the diagonal elements:

$$\overline{\boldsymbol{V}}_{(k+1)} = \boldsymbol{C} * \left(\begin{bmatrix} \sqrt{\overline{v}_{1,1}} \\ \sqrt{\overline{v}_{2,2}} \\ \vdots \\ \sqrt{\overline{v}_{n,n}} \end{bmatrix} \cdot \begin{bmatrix} \sqrt{\overline{v}_{1,1}} & \sqrt{\overline{v}_{2,2}} & \dots & \sqrt{\overline{v}_{n,n}} \end{bmatrix} \right) \tag{2.86}$$

This is the equivalent cofactor matrix of the $(k+1)$-th iteration of RLSCO-2. The iterated process of LS estimation and recomputation of $\overline{\boldsymbol{V}}$ is performed as described with RLSCO, see fig. 2.9.

Currently, physical correlations are usually neglected when processing GPS observations. So, RLSCO will be used as robust estimator in this thesis. However, RLSCO-2 may be more suitable for other applications, where physical correlations are known and introduced into the stochastic model used for parameter estimation.

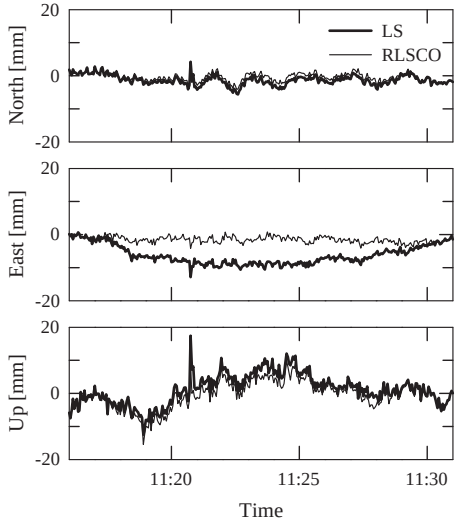

Figure 2.11 GPS phase data from a static session: processing results using robust estimator (RLSCO) and least squares (LS).

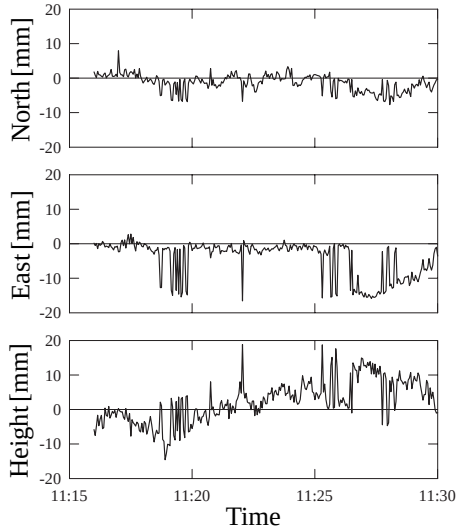

Figure 2.12 RLSCO epoch-by-epoch processing results of L1 GPS DD phase observations of a static baseline: deviation from ground-truth.

2.3.5 Examples and practical considerations

Fig. 2.11 shows the LS and RLSCO results of processing 15 minutes of GPS data from a static baseline of length 250 m. The LS result is the epoch-by-epoch solution based on the SIGMA-ϵ variance model (sec. 5.5.1). The RLSCO time series is computed on an epoch-by-epoch basis, using the equivalent a-priori variances from static RLSCO processing of the whole session. According to sec. 2.2.2, the static session result is approximately the average of the individual epoch results. The plot shows the deviations from ground truth. It reveals, that RLSCO succeeds in removing the bias of 10 mm in the east component. This is achieved by variance inflation of 291 observations out of 2104 total DD observations. The rms of the east component, as computed from the time series is reduced from 2.9 mm (LS) to 1.0 mm by RLSCO.

On the other hand, robust estimation is hardly applicable on the single epoch level, since the redundancy is too low then. This is demonstrated by fig. 2.12, that shows a time series of independent[7] epoch-by-epoch solutions using the same

[7]The epochs are tied together by the ambiguities that have been fixed in advance.

data as in fig. 2.11. Now the poor redundancy causes the estimator to fail in many epochs. It may not sufficiently inflate the variance of the outlying observations, and in some epochs, it even inflates the variance of good observations. This indicates clearly, that RLSCO is not applicable to the processing of single epochs of GPS data.

However, the suitability of RLSCO for processing short static GPS sessions is demonstrated in Wieser and Brunner (2001). The application of RLSCO to several datasets is discussed in chapter 7. In section 5.6, the RLSCO procedure is extended by a fuzzy system, that incorporates external information in order to facilitate applicability on the single epoch level.