A constrained quadratic programming technique for data-adaptive design of decorrelation filters

L. Roese-Koerner, I. Krasbutter and W.-D. Schuh

Abstract Signals from sensors with high sampling rates are often highly correlated. For the decorrelation of such data, which is often applied for the efficient estimation of parametric data models, discrete filters have proven to be both highly flexible and numerically efficient. Standard filter techniques are, however, often not suitable for eliminating strong local fluctuations or trends present in the noise spectral density. Therefore we propose a constrained least-squares filter design method. The spectral features to be filtered out are specified through inequality constraints regarding the noise spectral density. To solve for the optimal filter parameters under such inequality constraints, we review and apply the Active Set Method, a quadratic programming technique. Results are validated by statistical tests. The proposed filter design algorithm is applied to GOCE gradiometer signals to analyze its numerical behaviour and efficiency for a realistic and complex application.

Keywords Adjustment with inequality constraints · Decorrelation · Filter · Active Set Method

1 Introduction

The most popular method for estimating parameters is ordinary least-squares adjustment, which could easily be extended to handle equality constraints (Koch, 1999, pp. 170-177). However, in order to deal with inequality constraints, a transformation into a quadratic program (Gill et al., 1981, pp 177-186), a linear complementarity problem (Cottle et al., 1992) or a least distance program (Lawson and Hanson, 1974, pp 158-173) is needed. Roese-Koerner (2009) provides a detailed discussion on these methods.

All these problems could be solved by methods of convex optimization (optimization of a convex objective function subject to constraints, which form a convex set). This paper is focused on the Active Set Method (Gill et al., 1981, pp 199-203) - a very stable quadratic programming approach, which is more memory efficient than e.g. Dantzig's Simplex Algorithm for Quadratic Programs (Dantzig, 1998, pp 490-498).

Introducing inequality constraints is helpful in many applications, for example in the design of geodetic networks, to reshape error ellipses (Koch, 1982) or to introduce geometric constraints (Wölle, 1988). Koch (2006) used inequality constraints in the context of semantic integration of GIS data. Applications can be found also in signal processing, as constraints allow for flexible filter design (Fritsch, 1985; Schaffrin, 1981).

In the following, we focus on the design of decorrelation filters. As the spectral density of residuals often contains strong local fluctuations or trends, which need special treatment, design of decorrelation filters in the spectral domain is considered. Therefore we use constraints yielding flexible filter design and an optimal decorrelation. The capability of this approach is shown by decorrelating two different time series. The first one consists of randomly generated colored noise, the second, more sophisticated one, of residuals obtained from adjusted GOCE (Gravity field and steady-state Ocean Circulation Explorer) satellite gravity gradiometry (SGG) data (ESA, 1999).

Lutz Roese-Koerner, Ina Krasbutter, Wolf-Dieter Schuh
e-mail: roese-koerner@geod.uni-bonn.de
Department of Theoretical Geodesy
University of Bonn, Nussallee 17, D-53115 Bonn, Germany

2 Adjustment with inequality constraints

The usual way of estimating the parameter vector $\mathbf{x}$ of an overdetermined linear model

$$\mathbf{A}\mathbf{x} = \boldsymbol{\ell} + \mathbf{v}, \tag{1}$$

with $(n \times m)$ design matrix $\mathbf{A}$, observation vector $\boldsymbol{\ell}$ and residual vector $\mathbf{v}$, is to minimize the (weighted) sum of squared residuals

$$\Phi(\mathbf{v}) = \mathbf{v}^T \boldsymbol{\Sigma}^{-1} \mathbf{v}, \tag{2}$$

where $\boldsymbol{\Sigma}$ is the data covariance matrix. This leads to the well-known least-squares adjustment. As it is often reasonable to restrict the parameters to an interval (e.g. positivity, resource limits, budgets), the objective function (2) may have to be minimized with respect to p linear inequality constraints

$$b_{j,1}x_1 + \ldots + b_{j,m}x_m \leq b_j \quad \Leftrightarrow \quad \mathbf{B}^T\mathbf{x} \leq \mathbf{b}, \tag{3}$$

with constant matrix $\mathbf{B}$, vector $\mathbf{b}$ and $j = 1 \ldots p$. Greater-than-or-equal constraints can be transformed into less-than-or-equal constraints by multiplying the whole equation by minus one. Due to the fact that this problem could not be solved with ordinary least-squares adjustment, we reformulate it as a quadratic program (QP). Once the problem is transformed thusly, there are algorithms capable of dealing with the inequality constraints.

2.1 Quadratic program

A quadratic program consists of a quadratic objective function, which is to be minimized with respect to linear (inequality) constraints:

constraints:	$\mathbf{B}^T\mathbf{x} \leq \mathbf{b}$
parameters:	$\mathbf{x} \in \mathbb{R}^m$
objective function:	$\gamma_1 \mathbf{x}^T\mathbf{C}\mathbf{x} + \gamma_2 \mathbf{c}^T\mathbf{x} \ldots \text{Min.}$

$\mathbf{C}$ is a constant, symmetric and positive definite matrix, $\mathbf{c}$ a constant vector, and γ_1 and γ_2 are given scalars.
In order to transform the minimization problem (2) subject to (3) into a quadratic program, solely the objective function has to be reformulated:

$$\begin{aligned}\Phi(\mathbf{x}) &= \mathbf{v}^T\boldsymbol{\Sigma}^{-1}\mathbf{v} \\ &= (\mathbf{A}\mathbf{x} - \boldsymbol{\ell})^T\boldsymbol{\Sigma}^{-1}(\mathbf{A}\mathbf{x} - \boldsymbol{\ell}) \\ &= \mathbf{x}^T\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A}\mathbf{x} - 2\,\mathbf{x}^T\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\ell} + \boldsymbol{\ell}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\ell} \\ &= \gamma_1\mathbf{x}^T\mathbf{C}\mathbf{x} + \gamma_2\mathbf{c}^T\mathbf{x} \ldots \text{Min.}\end{aligned} \tag{4}$$

by using the substitutions

$$\mathbf{C} = 2\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A}, \quad \mathbf{c} = -2\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\ell}, \quad \gamma_1 = \frac{1}{2}, \quad \gamma_2 = 1$$

and neglecting the constant term $\boldsymbol{\ell}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\ell}$, which is irrelevant to the minimization problem. As a method for solving quadratic programs, we will focus on the Active Set Method (Gill et al., 1981, pp 199-203) for the above mentioned reasons (stability and memory efficiency).

2.2 Active Set Method

The idea behind this iterative algorithm is to start from an (arbitrary) feasible point - i.e. a point that satisfies all inequality constraints (3) - and follow the boundary of the feasible set until the minimal objective value is reached. After choosing an initial point $\mathbf{x}^{(0)}$ - e.g. with the "Big-M-Method" (Dantzig and Thapa, 2003, pp 115-116) - a search direction $\mathbf{p}^{(0)}$ and an appropriate step length $q^{(0)}$ are computed (superscript numbers denote the iteration steps). Figure 1 presents a brief overview of the algorithm.
The first step in calculating the search direction is to solve the unconstrained minimization problem (2):

$$\hat{\mathbf{x}} = \left(\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{A}\right)^{-1}\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\ell} = -\mathbf{C}^{-1}\mathbf{c}. \tag{5}$$

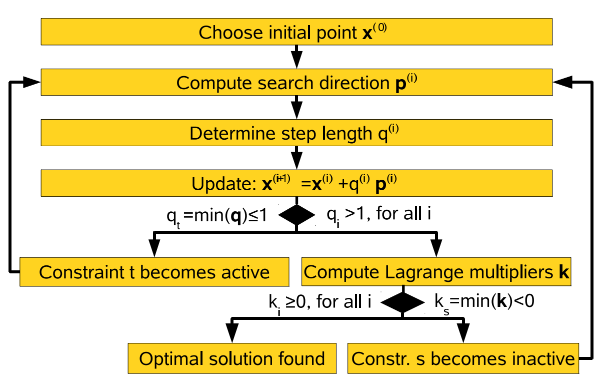


Fig. 1 Flowchart of the Active Set Algorithm.

The vector $\mathbf{p}^{*(0)}$ pointing from the initial point $\mathbf{x}^{(0)}$ to the minimum $\hat{\mathbf{x}}$ of the unconstrained problem is given by their difference

$$\mathbf{p}^{*(0)} = \hat{\mathbf{x}} - \mathbf{x}^{(0)}. \tag{6}$$

Next, the so-called Newton direction $\mathbf{p}^{*(0)}$ is projected onto the boundary of the feasible set, i.e. onto the active constraints (cf. Figure 2). A constraint is called "active" if the inequality is exactly satisfied

$$b_{j,1}x_1^{(i)} + b_{j,2}x_2^{(i)} + \ldots + b_{j,m}x_m^{(i)} = b_j. \tag{7}$$

At iteration i all active constraints from $\mathbf{B}$ and $\mathbf{b}$ are summarized in the active set matrix $\mathbf{W}^{(i)}$ and the vector $\mathbf{w}^{(i)}$:

$$\mathbf{W}^{(i)T}\mathbf{x}^{(i)} = \mathbf{w}^{(i)}.$$

Due to the metric, which is defined by the unconstrained least-squares objective function, $\mathbf{W}^{(i)}$ must be rescaled. This can be done by multiplying $\mathbf{W}^{(i)}$ from the left with $\mathbf{C}^{-1}$

$$\overline{\mathbf{W}}^{(i)} = \mathbf{C}^{-1}\mathbf{W}^{(i)}, \quad \text{with} \quad \mathbf{C}^{-1} = (2\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{A})^{-1}$$

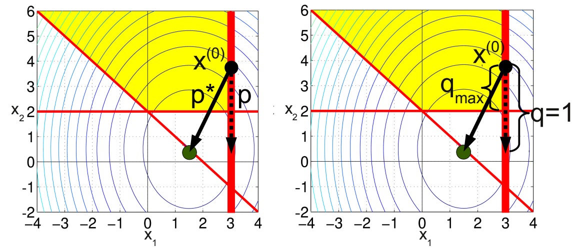
(Roese-Koerner, 2009, p 22). Now the search direction $\mathbf{p}^*$ can be projected into the left null space of the matrix $\overline{\mathbf{W}}^{(i)}$ of active constraints

$$\begin{aligned}\mathbf{p} &= <\mathbf{\Pi}^{\mathbf{C}}_{S^\perp(\overline{\mathbf{W}})}, \mathbf{p}^* >_{\mathbf{C}} \\ &= \mathbf{\Pi}^{\mathbf{C}}_{S^\perp(\overline{\mathbf{W}})}\mathbf{C}\mathbf{p}^* \\ &= (\mathbf{C}^{-1} - \overline{\mathbf{W}}(\overline{\mathbf{W}}^T\mathbf{C}\overline{\mathbf{W}})^{-1}\overline{\mathbf{W}}^T)\mathbf{C}\mathbf{p}^*\end{aligned}$$

(iteration indices neglected). $\mathbf{\Pi}^{\mathbf{C}}_{S^\perp(\overline{\mathbf{W}})}$ denotes the projection matrix under the metric $\mathbf{C}$ (Koch, 1999, pp 64-66). Thereby, no active constraint is violated by a step in direction $\mathbf{p}^{(i)}$. However, this is not true for inactive constraints. Hence the maximum step length $q_{max}^{(i)}$ has to be computed as the distance to the next inactive constraint along direction $\mathbf{p}^{(i)}$. Here, two scenarios are possible (cf. Figure 1).
If $q_{max}^{(i)} \leq 1$, we set $q^{(i)} = q_{max}^{(i)}$. After updating the parameters

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} + q^{(i)}\mathbf{p}^{(i)}, \tag{8}$$

a new constraint will become active and a new iteration is computed, starting with the determination of a new search direction according to (6). Otherwise, if $q_{max} > 1$, it is possible to take a step of optimal length (cf. Figure 2) without violating any constraints, and $q^{(i)} = 1$ is chosen. After computing the update step (8), the objective function (4) is extended to the Lagrange function



(a) Search direction $\mathbf{p}$. (b) Step length q_{max}.

Fig. 2 Search direction $\mathbf{p}$ and step length q_{max} of the Active Set Algorithm. In Figure 2(a) $\mathbf{p}^*$ denotes the vector to the unconstrained minimum (green point), the red lines symbolize constraints ($x_1 \leq 3, x_2 \geq 2, x_1 + x_2 \geq 2$) and the feasible set is colored yellow. The active (i.e. exactly satisfied) constraint is highlighted. In Figure 2(b) the optimal step length $q = 1$ is illustrated, which could not be chosen, because it violates an inactive constraint.

$$\Phi^*(\mathbf{x}, \mathbf{k}) = \frac{1}{2}\mathbf{x}^T\mathbf{C}\mathbf{x} + \mathbf{c}^T\mathbf{x} + \mathbf{k}^T(\mathbf{B}^T\mathbf{x} - \mathbf{b}), \tag{9}$$

as described in Wölle (1988). He proved, that the solution is optimal if and only if all Lagrange multipliers $\mathbf{k}^{(i+1)}$ related to active constraints are positive and that $\mathbf{p}^{*(i+1)}$ is in the column space of $\overline{\mathbf{W}}^{(i+1)}$ and can therefore be expressed as the linear combination

$$\mathbf{p}^{*(i+1)} = \overline{\mathbf{W}}^{(i+1)}\mathbf{k}^{(i+1)}, \tag{10}$$

with the Lagrange multipliers as weighting factors. Hence $\mathbf{k}^{(i+1)}$ could be computed by solving (10). If all Lagrange multipliers linked with active constraints are positive, the optimal solution is found. Otherwise all constraints with negative Lagrange multipliers are deactivated (Gill et al., 1981, p 170), i.e. removed from the active set $\mathbf{W}$, and the algorithm is started again.

3 Applications in filter design

With constrained quadratic programming techniques more flexibility is gained in many applications. Fritsch (1985) and Schaffrin (1981), for instance, applied Lemke's algorithm to design filters in the frequency domain using the transfer function. Thereby, one can control the spectral behavior of the filters by using inequality constraints. In the following we will use the power spectral density (PSD) to estimate decorrelation filters.

3.1 Design of simple decorrelation filters

The aim is to decorrelate data with colored noise characteristics by a decorrelation filter to obtain white noise. An algebraically simple approach is to estimate the coefficients of a non-recursive filter, such as of a symmetric moving-average filter:

$$y_i = \sum_{k=-N}^{N} \beta_k u_{i-k}, \qquad \text{with } \beta_{-k} = \beta_k, \qquad (11)$$

where β_k are the unknown filter coefficients, N is the order of the filter, $\mathbf{u}$ the input and $\mathbf{y}$ the output sequence. The filter coefficients may then be estimated by constrained least-squares adjustment using the PSD of the residuals as observations $\boldsymbol{\ell}$. The representation of the filter equation (11) in the frequency domain is the square of its transfer function (Schuh, 2003)

$$S(\boldsymbol{\beta},\omega) = (\beta_0 + 2\beta_1 \cos(\omega) + \ldots + 2\beta_k \cos(k\omega))^2,$$

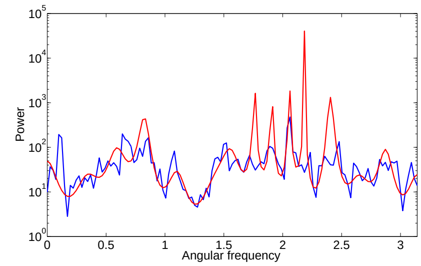
with angular frequency ω. In a numerical simulation the Active Set Method is used to decorrelate a randomly generated colored noise sequence, solving the quadratic program

$$\boxed{\begin{aligned}&\textbf{constraints: } \tilde{\ell}_j = \frac{1}{S(\boldsymbol{\beta},\omega_j)} \geq \ell_j, \quad j = 1,\ldots,p\\ &\textbf{parameters: } \boldsymbol{\beta} \in \mathrm{I\!R}^m\\ &\textbf{obj. function: } \sum_{i=1}^{n} \left(\tilde{\ell}_i - \ell_i\right)^2 \ldots \text{Min.}\end{aligned}}$$

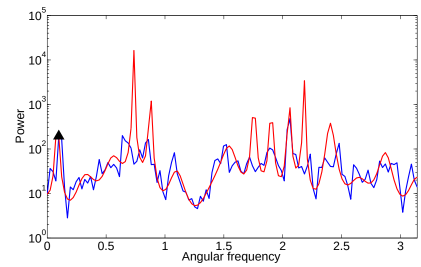
where p inverse, adjusted observations $\tilde{\ell}_j$ have to be greater than or equal to the original observations ℓ_j. Figure 3 shows the values $\boldsymbol{\ell}$ of the data PSD (blue) and the inverse values of the estimated filter $\tilde{\boldsymbol{\ell}}$ (red) with $N = 50$. The filter illustrated in Figure 3(a) is determined with unconstrained least-squares adjustment. The second filter, shown in Figure 3(b), has a constraint at angular frequency $\omega = 0.1$ Hz, which is badly denoised in the unconstrained case.

As described in Krasbutter (2009), the estimated decorrelation filters may be verified by statistical tests. One possible choice is the test of autocorrelation (Schlittgen and Streitberg, 2001, pp 243-246). Results of the test are illustrated in Figure 4. The red lines are boundaries of the confidence interval for type-I-error $\alpha = 0.05$, i.e. for white noise, 95 % of the autocorrelation coefficients must lie within this interval.

Using the unconstrained filter, the boundaries were violated by 26.0 % of the autocorrelation coefficients of the adjusted observations (see Figure 4). Introducing only one constraint improves the decorrelation process significantly; then only 5.6 % of the autocorrelation coefficients are outside the confidence interval. The five percent limit could be reached by applying more constraints.

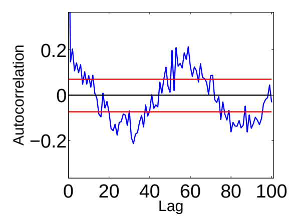


(a) No constraint.

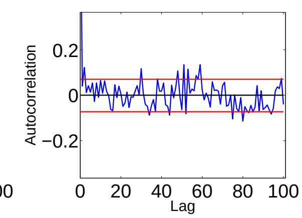


(b) Constraint at $\omega = 0.1$ Hz.

Fig. 3 Power spectral density of colored noise (blue) and the estimated reciprocal filter (red). The constraint is indicated by a black triangle.



(a) No constraint.



(b) Constraint at $\omega = 0.1$ Hz.

Fig. 4 Test of autocorrelation. For white noise, 95 % of the autocorrelation coefficients (blue) of different lags have to be in the confidence interval (red lines).

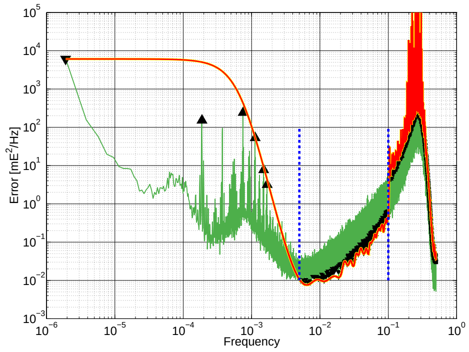


Fig. 5 PSD of simulated GOCE residuals (zz-component, green) and estimated inverse filter (red). Constraints are indicated by black triangles. The dotted vertical lines denote the measurement bandwidth between 0.005 Hz and 0.1 Hz.

3.2 GOCE decorrelation filters

Decorrelation filters are also used for GOCE data analysis as the observations of the gravity gradiometer are highly correlated (Schuh, 1996). As described in Schuh (2003) especially symmetric moving-average filters allow for an efficient computation of the parameters within iterative approaches.

The PSD of simulated, highpass-filtered gradiometer residuals (zz-component) and the inverse of the estimated filter are presented in Figure 5. The gradiometer is very sensitive in the bandwidth from 0.005 Hz to 0.1 Hz. Due to that fact, the signal outside the measurement bandwidth contains less information and should therefore be eliminated, while the signal inside should be preserved. This goal could be reached using inequality constraints as illustrated in Figure 5.

4 Conclusion

Inequality-constrained least-squares problems can easily be transformed into quadratic programs and be solved with the Active Set Method. Concerning the design of decorrelation filters, constrained quadratic programming techniques were demonstrated to lead to improved flexibility (cf. 3.2) and better rewhitening (cf. 3.1). One drawback of quadratic programming algorithms is that they yield no stochastic information about the estimated parameters. One possibility to overcome this, is to combine them with Monte Carlo methods in the future.

Acknowledgments: This work was financially supported by the BMBF Geotechnologien program REAL GOCE and the ESA contract No. 18308/04/NL/MM.

References

Cottle, R.W., Pang, J.-S. and Stone, R.E. (1992). The Linear Complementarity Problem. *Computer Science and Scientific Computing.* Academic Press Inc., San Diego

Dantzig, G.B. (1998). Linear Programming and Extensions. Princeton University Press, New Jersey

Dantzig, G.B. and Thapa, M.N. (2003). Linear Programming 2: Theory and Extensions. Springer Verlag, Berlin

ESA (1999). The four candidate Earth explorer core missions - gravity field and steady-state ocean circulation mission. ESA Report SP-1233(1), Granada

Fritsch, D. (1985). Some additional informations on the capacity of the linear complementarity algorithm. In: Grafarend, E.W. and F. Sansò (ed.), *Optimization and Design of Geodetic Networks*, Springer, Berlin, pp 169-184

Gill, P., Murray, W. and Wright, M. (1981). Practical Optimization. Academic Press, San Diego

Koch, K.-R. (1982). Optimization of the configuration of geodetic networks. *Proceedings of the international symposium on geodetic networks and computations*, Munich

Koch, K.-R. (1999). Parameter Estimation and Hypothesis Testing in Linear Models. Springer Verlag, Berlin

Koch, A. (2006). Semantic integration of two-dimensional GIS data and digital terrain models. *PhD thesis (in german)* DGK series C, No. 601, University of Hannover

Lawson, C. and Hanson, R. (1974). Solving Least Squares Problems. Prentice-Hall, London

Krasbutter, I. (2009). Decorrelation and validation of GOCE residuals. *Diploma thesis (in german)*, Institute of Geodesy and Geoinformation, University of Bonn

Roese-Koerner, L. (2009). Quadratic programming with inequality constraints. *Diploma thesis (in german)*, Institute of Geodesy and Geoinformation, University of Bonn

Schaffrin, B. (1981). Parameter estimation with inequality constraints. *Allgemeine Vermessungs-Nachrichten (in german)*, 88, pp 227-238

Schlittgen, R. and B. H. J. Streitberg (2001). Time series analysis. 9th edition *(in german)*, Oldenbourg Verlag, Munich

Schuh, W.-D. (1996). Tailored numerical solution strategies for the global determination of the earth's gravity field. *Mitteilungen der Geodätischen Institute der TU, No.81*, Graz

Schuh, W.-D. (2003). The processing of band-limited measurements; filtering techniques in the least squares context and in the presence of data gaps. In: Beutler, G. and Drinkwater, M. and Rummel, R. and von Steiger, R. (ed.), *Earth Gravity Field from Space - From Sensors to Earth Sciences, Space Science Reviews*, 108, pp 67-78

Wölle, N.A. (1988). Graphical estimation through Quadratic Programming with linear equalities and inequalities. *Diploma thesis (in german)*, Technical University of Graz