ON THE REGULARIZATION PROBLEM IN GRADIOMETRIC DATA ANALYSIS FROM GOCE

Jürgen Kusche[(1)], Roland Klees[(2)]

[(1)] ***DEOS, Delft University of Technology***
Thijsseweg 11, 2629 JA Delft, The Netherlands
j.kusche@citg.tudelft.nl

[(2)] ***DEOS, Delft University of Technology***
Thijsseweg 11, 2629 JA Delft, The Netherlands
r.klees@geo.tudelft.nl

INTRODUCTION

The functional model of the gravity field recovery from GOCE gradiometry observations may be written as a standard Gauss–Markov model

$$y = Ax + e \qquad\qquad D\{y\} = \sigma^2 P^{-1} \ . \tag{1}$$

Here y denotes the vector of gradiometric data, e is the vector of measurement errors with $E\{e\} = 0$, $D\{y\} = \sigma^2 P^{-1}$, and σ^2 is an a–priori variance factor. The design matrix A assigns the observations to the unknown vector x of gravity potential coefficients. Consequently the best linear uniformly unbiased estimate reads

$$\hat{x} = (A^T P A)^{-1} A^T P y \ . \tag{2}$$

Due to the huge number of observations and of unknowns – it is expected to recover about 90.000 potential coefficients from 100 million observations collected during the GOCE mission lifetime – the assembly of the normal equation matrix and right–hand side presents a giant task, from a numerical point of view. With no restrictions imposed on the satellite orbit and the data sampling, the normal matrix $A^T P A$ will be nearly block–structured but, nethertheless, dense. Moreover, the downward continuation process as well as the presence of polar data gaps turnes the problem into an ill–posed one. This property translates into an ill–conditioned normal equation system, and due to the amplification of data errors any solution (2) strongly oscillates and will be physically meaningless. A way out of this problem is regularization. Tikhonov regularization (TR) is the best–investigated regularization method, and has been shown to be numerically equivalent to some other methods such as the least–squares collocation, ridge regression ([10]) and in particulary the so–called Kaula stabilization. We will investigate two recent criteria for the selection of the regularization parameter, which are capable for a large–scale problem like GOCE gradiometry. In particular, we will show that one of these methods outperforms the Kaula method for spectral resolution above $L \approx 150$. The TR version of (2) reads

$$x_\alpha = (A^T P A + \alpha K)^{-1} A^T P y \tag{3}$$

with the regularization or smoothing parameter α and a symmetric positive definite regularization matrix $K = L^T L$. This estimate minimizes the penalized least–squares functional

$$\mathcal{J}_\alpha = (Ax - y)^T P (Ax - y) + \alpha \ x^T K x \ =: \ ||Ax - y||_P^2 + \alpha \ ||x||_K^2 \ . \tag{4}$$

It is well–known that (3) is numerically equivalent to an unbiased least–squares estimate $\hat{x}$, if one introduces the prior information $E\{x\} = 0$, $D\{x\} = (\sigma^2/\alpha) \ K^{-1}$. Identifying the diagonal part of this covariance matrix with the signal variances from a degree-variance model or an existing satellite–derived geopotential model yields the so-called Kaula stabilization, which in often used. Fig. 1 shows the degree–error rms of three different solutions, obtained with different regularization parameters, as well as the signal spectrum from OSU91 used in the simulation. For each individual solution x_α, the degree–error rms is defined by

$$e_l^\alpha = \sqrt{\left(\frac{\sum_{m=0}^{l} (c_{lm} - \bar{c}_{lm}^\alpha)^2 + (s_{lm} - \bar{s}_{lm}^\alpha)^2}{2l+1} \right)}, \tag{5}$$

where c_{lm}, s_{lm} are the "true" OSU91 geopotential coefficients, and $\bar{c}_{lm}^\alpha$, $\bar{s}_{lm}^\alpha$ are the estimated geopotential coefficients (i. e. the elements of x_α). The Kaula matrix was chosen for K, and the three numerical values

were: $\alpha = 1 \cdot 10^{-10}$ (the tiniest value that allowed Cholesky decomposition of the normal matrix), $\alpha = 1$ (i. e. the "nominal" Kaula–stabilized solution), $\alpha = 0.10$ (the value obtained with the RGCV method). It is obvious that some sort of stabilization is necessary in order to avoid strongly distorted solutions. With stabilization a spectral resolution of around $L \approx 270$ is possible. The second fact is that by "tuning" the regularization parameter the solution improves clearly in the high–frequency domain as well as in the low–frequency part (which will partially be covered by SST measurements). However, the optimal regularization parameter with respect to the true solution,

$$\alpha_{\mathrm{mse}} \;=\; \arg\min \frac{1}{u}\,||x_\alpha - x||^2 \;=\; \arg\min \sum_{l=l_{min}}^{l_{max}} (2l+1)(e_l^\alpha)^2 \tag{6}$$

is only accessible in a simulation. In the following we investigate two strategies, which in the mathematical

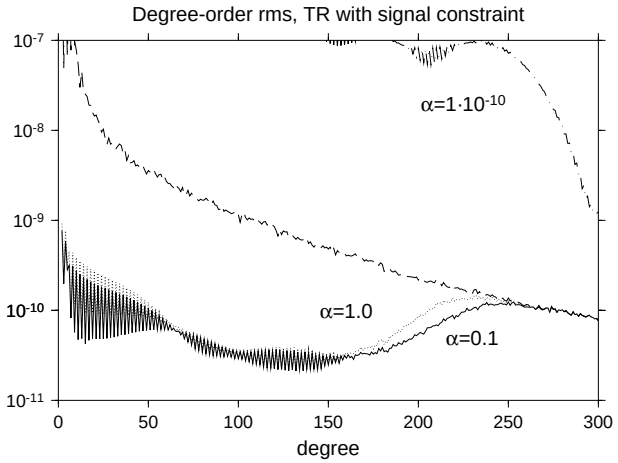

Figure 1: Degree–error rms. The solutions are computed with almost no regularization (dash–dotted), Kaula stabilization (dotted), and TR with RGCV parameter choice (solid)

literature are referred to as heuristic a–posteriori methods: The L–curve criterion and a randomized version of generalized cross–validation. A–posteriori means that one does not rely on prior smoothness assumptions concerning the unknown solution but uses the actual observations to extract this information. Heuristic means that it is not necessary to know exactly the variance factor σ^2 of the data. Non–heuristic methods like the so–called discrepancy principle are known to be weak against inexact knowledge of the data error norm; and we argue that this could be a clear disadvantage concerning GOCE spaceborn data because 1) the instrument performance has never been validated in orbit but only under laboratory conditions, and 2) the actual observations will also be contaminated with an aliasing signal due to the unmodelled high frequencies of the true gravitational field.

REGULARIZATION AND PARAMETER CHOICE METHODS

Parameter choice by the L–curve criterion

The Tikhonov L–curve is a parametric plot of the residual norm versus the norm of the solution for all valid values of the regularization parameter. It is a conveniant graphical tool for displaying the trade–off between the size of the regularized solution and its fit to the given data, as the regularization parameter varies. It is common practice to plot the curve on a log–log scale, since only in this style the appearance of the curve is independent with regard to the choice of units. Choosing the same norms that are penalized in the least–squares functional (4), the graph of the L–curve is given by

$$\Big(\xi(\alpha), \eta(\alpha)\Big) = (\log||Ax_\alpha - y||_P, \log||x_\alpha||_K) \;\; . \tag{7}$$

For typical ill–posed problems, however, the L–curve in fact exhibits a distinct "L–shape" with a "corner point", where a certain balance in the trade-off is present. This situation can be observed very clearly in our GOCE

simulation, see figure 3 left. Following [4] the location of the L–curve corner may be used in order to approximate the optimal regularization parameter. This will be called the "L–curve criterion" in the remainder of this paper. The exact definition of the corner point is not a trivial problem and different methods have been proposed in the literature. An obvious option is to maximize the curvature of the graph of (7) (see [4]),

$$\alpha_{\mathrm{lc(curv)}} \;=\; \arg\max \frac{\xi'\eta'' - \xi''\eta'}{\left((\xi')^2 + (\eta')^2\right)^{3/2}} \;, \tag{8}$$

where ξ', η', ξ'' and η'' denote derivatives of ξ and η with respect to α. It is common to replace the derivatives in (8) by differences. An alternative definition is the following: A point is considered the corner of the L–curve, if the curve is concave in its neighbourhood and the tangent at the point has a slope of -1. It can been shown that this definition is equivalent to the formulation

$$\alpha_{\mathrm{lc(tang)}} \;=\; \arg\min \, ||Ax_\alpha - y||_P \, ||x_\alpha||_K, \tag{9}$$

which is easier implementable than (8). One might justify the tangent definition (9) by the simple observation that the corner point typically connects a nearly vertical "leg" of the curve with a nearly horizontal "leg". Although the L–curve criterion cannot hide a certain lack in mathematical foundation, many researchers investigated its application numerically with good results. In the context of gravity field determination from airborne/satellite data, the L–curve criterion has been investigated by [1] and [9]. We mention that the criterion proposed by [6] follows a similar line of reasoning and seems therefore related to the L–curve.

Parameter choice by generalized cross–validation

Generalized cross–validation (GCV) as a method for choosing the regularization parameter has been originally proposed by [3]. The philosophy of cross–validation is based on the leave–out–one idea, which is well–known in geodetic testing theory: Omitting the k–th observation y_k, the respective leave–out–one solution vector $x_\alpha^{[k]}$ can be used to predict the "missing" observation. A good regularization parameter will produce solutions where the misfit $(Ax_\alpha^{[k]})_k - y_k$ is low in average over all possible y_k. The generalized cross–validation parameter α_{gcv}

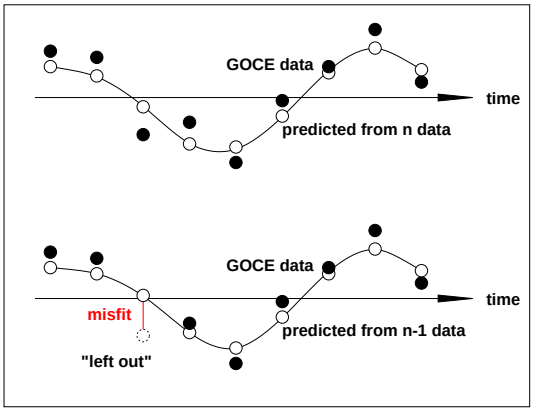

Figure 2: Leave–out–one idea: GOCE true observations (•) and predicted ones (∘)

is given by

$$\alpha_{\mathrm{gcv}} \;=\; \arg\min \frac{n \, ||Ax_\alpha - y||^2}{\left(\mathrm{trace}(I - Q^\alpha)\right)^2} \tag{10}$$

where Q^α is the influence matrix defined by $Ax_\alpha = Q^\alpha y$. α_{gcv} estimates the minimizer of the unknown predictive mean square error

$$\alpha_{\mathrm{pmse}} \;=\; \arg\min \frac{1}{n} \, ||Ax_\alpha - Ax||^2 \;. \tag{11}$$

For problems with coloured noise like GOCE, the functionals (10), (11) have to be decorrelated using the weight matrix P. Inserting further (3) for the definition of Q^α, the GCV parameter can be re–written in the form

$$\alpha_{\mathrm{gcv}} \;=\; \arg\min \frac{n \, ||Ax_\alpha - y||_P^2}{(n - u + \alpha\mathcal{T}_\alpha)^2}, \qquad \mathcal{T}_\alpha = \mathrm{trace}\left(L(A^T P A + \alpha K)^{-1} L^T\right) \;. \tag{12}$$

For a large–scale problem like GOCE the last equation is better suited than (10) since here we have to determine the trace of a $u \times u$ matrix, whereas the influence matrix Q^α in (10) is of $n \times n$ dimension. However, with 90.000 unknowns to be determined from GOCE gradiometry the computation of the trace–term $\mathcal{T}_\alpha$, which involves the inverse normal matrix, seems prohibitive. A way out is the use of trace–estimators ([2], [5]), and the resulting algorithm is called Monte Carlo GCV. Trace–estimation goes as follows: Let z be a random vector with $E\{z\} = 0$, $D\{z\} = I$; then

$$\overline{\mathcal{T}}_\alpha = z^T L(A^T PA + \alpha K)^{-1} L^T z \tag{13}$$

is an unbiased estimator of $\mathcal{T}_\alpha$, where the variance of this estimator depends on the probability density function of z. In practice, $\overline{\mathcal{T}}_\alpha$ is computed in three steps: 1) a random vector z is generated, 2) one solves the auxiliary system $(A^T PA + \alpha K)q_\alpha = L^T z$, and finally 3) one computes $\overline{\mathcal{T}}_\alpha = z^T L q_\alpha$. Numerical benefit can be drawn from the fact that the auxiliary system involves only an additional right–hand side vector $L^T z$ but the same normal matrix than the "main" system in (3). The dispersion of the trace–estimate may be further reduced by averaging single estimates obtained with different realizations of the random vector, but for large problems it is usually sufficient to employ only one realization of z. Monte Carlo GCV with only one realization is also called randomized cross–validation (RGCV). GCV methods are occasionally reported to perform weak in problems with coloured noise; this means, when no decorrelation takes place in (12) at all. We adressed this question in our numerical study, since decorrelation of the residual vector will be a time–consuming task in gradiometric data analysis.

TEST SETUP AND SIMULATIONS

The performance of the two regularization parameter choice rules has been investigated using simulated data. First, a circular orbit 258 km above the Earth's surface was generated. The orbit has an inclination of 96.6°; it is almost repeat with 961 orbit revolutions during 59.8 nodal days. The time for one revolution is about 5380 seconds. The difference between OSU91a and GRS80 defines the 'true' disturbing potential to be recovered. Along the orbit second radial derivatives of the disturbing potential were generated at known positions with a sampling rate of 5 seconds, which gives about one million observations. The normal equations were computed up to degree and order 300. The observations were corrupted by coloured noise, generated from a power spectral density function with a flat spectrum of 9 $\mathrm{mE}^2/\mathrm{Hz}$ between 0.005 Hz and 0.1 Hz and a $1/\mathrm{f}^2$ behaviour between $3.7 \cdot 10^{-4}$ Hz and 0.005 Hz, see [8]. The time-wise approach was followed, i.e. the measurements are considered as an (uninterrupted) time series along the satellite orbit, and the unknown disturbing potential is expressed in orbital elements instead of the usual representation in an Earth-fixed coordinate system using spherical coordinates. It is known that the time-wise approach yields a strictly block-diagonal normal matrix provided that the orbit is circular and exactly repeat, the data period is an integer multiple of the repeat period, and the data flow is uninterrupted. The generated orbit and the generated time series of gravity gradients fulfil these requirements up to a slight non-closure of the orbit, which strictly spoken yields a dense normal matrix with a strong block-diagonal dominance. This slight deviation from a strictly block-diagonal normal matrix was neglected in the simulations, which causes small disturbancies in the estimated disturbing potential coefficients. These small disturbancies could have been removed by iteration (e.g. [7]), but this has not been done, since the effects on the solution were not significant.

RESULTS AND DISCUSSION

Regularization with signal constraints

In the following the question will be addressed whether the optimal parameter α_{mse} can be approached by the parameter choice strategies discussed before, and what the effect of choosing different constraints in (4) would be. The results are summarized in table 1. First we apply TR, where the matrix K is diagonal with elements $10^{10} l^4$, which is the inverse of the well–known Kaula rule. The L–curves (7), using unfiltered gradiometric residuals as well as decorrelated residuals, are presented in figure 3 in double–logarithmic style. Both curves exhibit a distinct shape: A nearly vertical part for smooth solutions, where the residual norm increases dramatically, and a nearly horizontal leg for unstabilized, "rough" solutions with the residuals levelling on a low plateau. ¿From the figure it is clear that the corner points are located somewhere left to the optimal parameter $\alpha_{\mathrm{mse}} = 0.1$. Thus, solutions computed with α_{lc} will be "oversmooth". Indeed we find $\alpha_{\mathrm{lc(tang)}} = 3.0$ and $\alpha_{\mathrm{lc(curv)}} = 2.0$ using filtered residuals and $\alpha_{\mathrm{lc(tang)}} = 4.0$ and $\alpha_{\mathrm{lc(curv)}} = 3.0$ with unfiltered residuals. Finally we mention that it is

choice of K	α_{mse}	L-curve		RGCV	
		$\|\cdot\|_2$	$\|\cdot\|_P$	$\|\cdot\|_2$	$\|\cdot\|_P$
$K \sim l^4$	0.1	3.0	2.0	0.5	0.1
$K \sim l^6$	$3 \cdot 10^{-6}$	$1 \cdot 10^{-4}$	$8 \cdot 10^{-5}$	$6 \cdot 10^{-6}$	$3 \cdot 10^{-6}$
$K \sim l^8$	$5 \cdot 10^{-11}$	$7 \cdot 10^{-9}$	$4 \cdot 10^{-9}$	$9 \cdot 10^{-11}$	$5 \cdot 10^{-11}$

Table 1: Regularization parameter using L-curve criterion and randomized generalized cross-validation (RGCV) for various choices of the regularization matrix K. $\|\cdot\|_2$: neglecting correlations, $\|\cdot\|_P$: taking correlations into account by filtering the residuals

absolutely necessary to measure the size of the solution in the same $||\cdot||_K$–norm that is used in (4), i. e. here by weighting the estimated geopotential coefficients according to Kaula's law. Otherwise the respective curves are wiggly with multiple "corners", and cannot enable a unique interpretation. Next we apply randomized GCV.

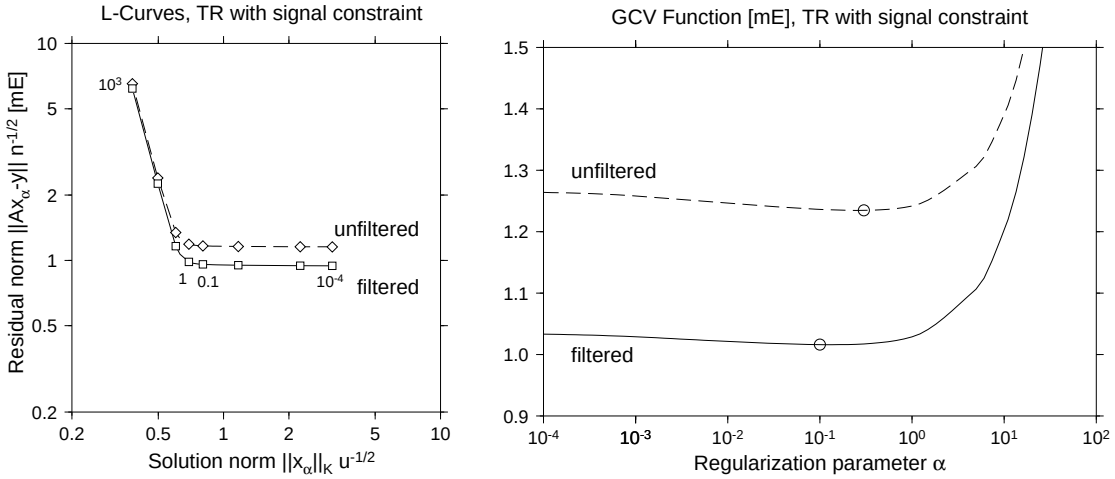

Figure 3: TR parameter choice criteria: L–curve (left), RGCV (right). Solid lines apply to filtered residuals, dashed lines to unfiltered residuals

Figure 3 (right) shows on a semi–logarithmic scale the GCV function

$$GCV(\alpha) = \sqrt{n}\ \frac{||Ax_\alpha - y||}{n - u + \alpha \overline{\mathcal{T}}_\alpha} \tag{14}$$

for a wide range of regularization parameters, and for $||\cdot|| = ||\cdot||_2$ (unfiltered residuals) as well as $||\cdot|| = ||\cdot||_P$ (filtered residuals). Units are mE. Both curves are relatively flat in the domain of reasonable α's, but the minima α_{gcv} are indeed very close to $\alpha_{\text{mse}} = 0.1$. With α_{gcv} determined from filtered residuals the estimated geopotential coefficients $\bar{c}^\alpha_{lm}$, $\bar{s}^\alpha_{lm}$ nearly coincide with the best–possible regularized solution! The overall results are summarized in figure 4 (left), which shows the degree–error rms of the estimated potential coefficients according to the parameter choice rules discussed before. Another possibility is regularization with first or second (radial) derivative constraint, i. e. the matrix K is diagonal with elements $10^{10}(l+1)^2 l^4$ or $10^{10}(l+1)^2(l+2)^2 l^4$. High degree coefficients are then more constrained than in the previous experiments. The results are given in table 1. These values are generally smaller than those of the previous experiment, since the entries of K are now orders of magnitude larger. The L–curve solutions are nethertheless oversmooth, especially if the tangent definition (9) is applied. Randomized GCV again leads to parameters which are very close to the optimal α_{mse}. The overall results of this section are summarized in figure 4 (right), which shows the degree–error rms for geopotential coefficient solutions according to the parameter choice rules discussed before.

Performance of the parameter choice strategies

The choice of the regularization parameter in GOCE gradiometric data analysis is an important issue, and

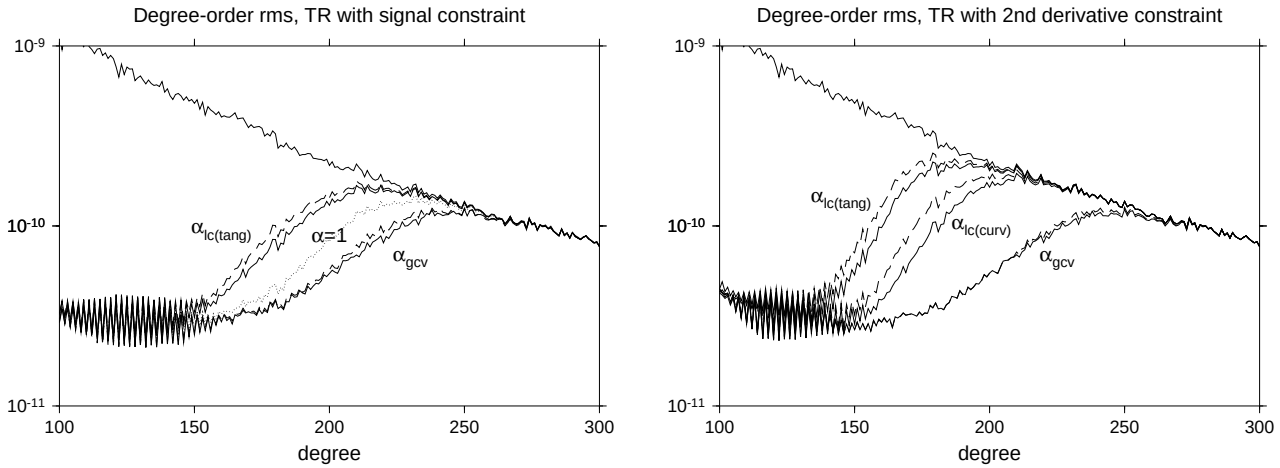

Figure 4: Degree–error rms. TR–Solutions for signal constraint (left) and 2nd derivative constraint (right), according to L–curve and RGCV. Solid lines apply to decorrelated residuals, dashed lines to unfiltered residuals.

considerable improvements with regard to the Kaula method are possible when applying cross–validatory methods. We found that the L–curve method gives oversmooth solutions and should therefore be applied with care. Different corner definitions lead to different solutions and we recommend to adhere to the curvature–based definition and not to the tangent definition. We found that randomized cross–validation (RGCV) outperforms the L–curve criterion by far and, when combined with an appropriate decorrelation, gives the best–possible solution. Moreover, the RGCV solution is much better than the Kaula-stabilized solution. The price to be paid is roughly the work of one additional solution of the normal equation system for a different right–hand side. Filtering the residuals is generally helpful for finding a good regularization parameter. But even without any decorrelation the RGCV solution is very close to the best–possible solution. The L–curve criterion seems to a larger extend weak against unmodelled coloured noise.

REFERENCES

[1] J. Bouman, "Quality of Regularization Methods", *DEOS Report* No. 98.2, Delft University Press

[2] D. A. Girard, "A fast 'Monte–Carlo Cross–Validation' procedure for large least squares problems with noisy data", *Num Math* 56: 1–23, 1989

[3] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross–validation as a method for choosing a good ridge parameter", *Technometrics* 21: 215–223, 1979

[4] P. C. Hansen, D. P. O'Leary, "The use of the L–curve in the regularization of discrete ill–posed problems", *SIAM J Sci Comput*, 14: 1487–1503, 1993

[5] M. F. Hutchinson, "A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines", *Comm Stat, Sim and Comp* 19: 433–450, 1990

[6] K. H. Ilk, "Regularization for High Resolution Gravity Field Recovery by future Satellite Techniques", in G. Angerer G et al (eds) *Inverse Problems: Principles and Applications in Geophysics, Technology, and Medicine.* Akademie Verlag, pp 189–214, 1993

[7] R. Klees, R. Koop, P. Visser, and J. van den IJssel, "Efficient gravity field recovery from GOCE gravity gradient observations" *J Geod* 74: 561–571, 2000

[8] R. Koop R, P. Visser, J. van den IJssel, R. Klees, "Detailed scientific Data Processing Approach", in H. Sünkel *From Eötvös to Milligal.* ESA/ESTEC Contract No. 13392/98/NL/GD, 2000

[9] J. Kusche, K. H. Ilk, "The Polar Gap Problem", in H. Sünkel *From Eötvös to Milligal.* ESA/ESTEC Contract No. 13392/98/NL/GD, 2000

[10] P. Xu, R. Rummel, "Generalized ridge regression with applications in determination of potential fields", *manuscripta geodaetica*, 20: 8–20, 1994