Correcting the smoothing effect of least-squares collocation with a covariance-adaptive optimal transformation

Christopher Kotsakis

Department of Geodesy and Surveying
Aristotle University of Thessaloniki
University Box 440, GR-54124, Thessaloniki, Greece
Email: kotsaki@topo.auth.gr

Abstract. An optimal modification of the classical LSC prediction method is presented, which removes its inherent smoothing effect while sustaining most of its local prediction accuracy at each computation point. Our 'de-smoothing' approach is based on a covariance-matching constraint that is imposed on a linear modification of the usual LSC solution so that the final predicted field reproduces the spatial variations patterns implied by an adopted covariance (CV) function model. In addition, an optimal criterion is enforced which minimizes the loss in local prediction accuracy (in the mean squared sense) that occurs during the transformation of the original LSC solution to its CV-matching counterpart. The merit and the main theoretical principles of this signal CV-adaptive technique are analytically explained, and a comparative example with the classical LSC prediction method is given.

Keywords. Least-squares collocation, spatial random field, smoothing, covariance matching.

1 Introduction

The prediction of the functional values of a continuous *spatial random field* (SRF), using a set of observed values of the same and/or other SRFs, is a fundamental inverse problem in geosciences. The mathematical model describing such a problem is commonly formulated in terms of the observation equation

$$y_i = L_i(u) + v_i \quad , \quad i = 1, 2, \ldots, n \tag{1}$$

where $u(P)$ denotes the primary random field of interest ($P \in D$, with D being a bounded or unbounded spatial domain) that needs to be determined, at one or more points, using n discrete measurements $\{y_i\}$ which are taken on the same and/or other locations. The symbols $L_i(\cdot)$ correspond to bounded linear or linearized functionals of the unknown field, depending on the physical model that relates the observable quantities with the underlying SRF itself, while the terms $\{v_i\}$ contain the effect of measurement random noise. Typical examples that fall within the realm of the aforementioned SRF prediction scheme include the determination of the disturbing gravity potential on or outside a spherical Earth model using various types of gravity field functionals, the prediction of stationary or non-stationary ocean circulation patterns from satellite altimetry data, the prediction of atmospheric fields (tropospheric, ionospheric) from the tomographic inversion of GPS data, the prediction of crustal deformation fields from geodetic data, etc.

The predominant approach that is generally followed in geodesy for solving such problems is *least-squares collocation* (LSC) which was introduced by Krarup (1969) in a deterministic context as a rigorous approximation method in separable Hilbert spaces with reproducing kernels, and formulated in parallel by Moritz (1970) in a probabilistic setting as an optimal prediction technique for spatially correlated random variables and stochastic processes; see also Sanso (1986), Dermanis (1976).

A critical aspect in LSC is the smoothing effect on the predicted signal values $\hat{u}(P)$, which typically exhibit less spatial variability than the actual field $u(P)$. Consequently, small field values are overestimated and large values are underestimated, thus introducing a likely conditional bias in the final results and possibly creating artifact structures in SRF maps generated through the LSC process. Note that smoothing is an important characteristic which is not solely associated with the LSC method, and it is shared by most interpolation techniques aiming at the unique approximation of a continuous function from a finite number of observed functionals. Its merit is that it guarantees that the recovered field does not produce artificial details not inherent or proven by the actual data, which is certainly a desirable characteristic for an optimal signal interpolator. However, the use of smoothed SRF images or

maps generated by techniques such as LSC provides a shortfall for applications sensitive to the presence of extreme signal values, patterns of field continuity and spatial correlation structure. While founded on *local* optimality criteria that minimize the mean squared error (MSE) at each prediction point, the LSC approach overlooks to some extent a feature of reality that is often important to capture, namely spatial variability. The latter can be considered a *global* field attribute, since it only has meaning in the context of the relationship of all predicted values to one another in space. As a result of the smoothing effect, the ordinary LSC solution does not reproduce either the histogram of the underlying SRF, or the spatial correlation structure as implied by the adopted model of its covariance (CV) function.

In this paper we present an ad-hoc approach that enhances LSC-based field predictions by eliminating their inherent smoothing effect, while preserving most of their local prediction accuracy. Our approach consists of correcting a-posteriori the optimal result obtained from LSC in the inversion of (1), in a way that the corrected field matches the spatial correlation structure implied by the signal CV function that was used to construct the initial LSC solution. Similar predictors have also appeared in the geostatistical literature by constraining the usual unbiased kriging-type solution through a covariance-matching condition, thus yielding new linear SRF predictors that match not only the first moment but the second moment of the primary SRF as well (Aldworth and Cressie 2003, Cressie 1993).

In contrast to stochastic simulation schemes which provide multiple equiprobable signal realizations according to some CV model of spatial variability (e.g. Christakos 1992), the methodology presented herein gives a *unique* field estimate that is statistically consistent with a prior model of its spatial CV function. The uniqueness is imposed though an optimal criterion that minimizes the loss in local prediction accuracy (in the MSE sense) which occurs when we transform the LSC solution to match the spatial correlation structure of the underlying SRF.

2 Ordinary Least-Squares Collocation

Denoting by $s_i = L_i(u)$ the signal part in the available data, the system of observation equations in (1) can be written in vector form as

$$\mathbf{y} = \mathbf{s} + \mathbf{v} \tag{2}$$

where **y**, **s** and **v** are random vectors containing the known measurements, and the unknown signal and noise values, respectively, at all observation points $\{P_i\}$. The signal and noise components in (2) are considered uncorrelated with each other (a crucial simplification that is regularly applied in practice), and of known statistical properties in terms of their given expectations and co-variances.

Assuming that the spatial variability of the primary SRF u is described by a known CV function model $C_u(P,Q)$, the elements of the CV matrix of the signal vector **s** are determined according to the CV propagation law (Moritz 1980)

$$\mathbf{C_s}(i,j) = L_i\, L_j\, C_u(P_i, P_j) \tag{3}$$

where L_i and L_j correspond to the functionals associated with the ith and j^{th} observation. In the same way, the cross-CV matrix between the primary field values (at the selected prediction points $\{P_i'\}$) and the observed signal values is obtained as

$$\mathbf{C_{us}}(i,j) = L_j\, C_u(P_i', P_j) \tag{4}$$

The CV matrix $\mathbf{C_v}$ of the data noise is also considered known, based on the availability of an appropriate stochastic model describing the statistical behavior of the zero-mean measurement errors.

An additional postulate on the spatial trend of the primary SRF is often employed as an auxiliary hypothesis for the LSC inversion of (1) or (2). In fact, various LSC prediction algorithms may arise in practice, depending on how we treat the signal detrending problem. For the purpose of this paper and without any essential loss of generality, it will be assumed that we deal only with zero-mean SRFs and signals ($E\{u\} = 0$, $E\{\mathbf{s}\} = \mathbf{0}$).

Based on the previous assumptions, the LSC predictor of the primary SRF, at all selected prediction points, is given by the well known matrix formula

$$\hat{\mathbf{u}} = \mathbf{C_{us}}(\mathbf{C_s} + \mathbf{C_v})^{-1}\mathbf{y} \tag{5}$$

which corresponds to the linear unbiased solution with minimum mean squared prediction error (Moritz 1980, Sanso 1986).

The inherent smoothing effect in LSC can be identified from the CV structure of its optimal result. Applying CV propagation to the predicted field $\hat{\mathbf{u}}$ in (5), we obtain the result

$$\mathbf{C}_{\hat{\mathbf{u}}} = \mathbf{C}_{\mathbf{us}}(\mathbf{C}_{\mathbf{s}} + \mathbf{C}_{\mathbf{v}})^{-1}\mathbf{C}_{\mathbf{us}}^{\mathrm{T}} \tag{6}$$

which generally differs from the CV matrix $\mathbf{C}_{\mathbf{u}}$ of the original SRF at the same set of prediction points, i.e.

$$\mathbf{C}_{\mathbf{u}}(i,j) = C_u(P_i', P_j') \neq \mathbf{C}_{\hat{\mathbf{u}}}(i,j) \tag{7}$$

Moreover, if we consider the vector of the prediction errors $\mathbf{e} = \hat{\mathbf{u}} - \mathbf{u}$, it holds that

$$\mathbf{C}_{\hat{\mathbf{u}}} = \mathbf{C}_{\mathbf{u}} - \mathbf{C}_{\mathbf{e}} \tag{8}$$

where the error CV matrix is given by the equation

$$\mathbf{C}_{\mathbf{e}} = \mathbf{C}_{\mathbf{u}} - \mathbf{C}_{\mathbf{us}}(\mathbf{C}_{\mathbf{s}} + \mathbf{C}_{\mathbf{v}})^{-1}\mathbf{C}_{\mathbf{us}}^{\mathrm{T}} \tag{9}$$

The fundamental relationship in (8) conveys the meaning of the smoothing effect in LSC, which essentially acts as an optimal low-pass filter to the input data. The spatial variability of the LSC prediction errors, in terms of their variances and covariances, is exactly equal to the deficit in spatial variability of the LSC predictor $\hat{\mathbf{u}}$ with respect to the original SRF **u**.

3 Optimal "De-Smoothing" of the LSC solution

Our objective is to develop a correction algorithm that can be applied to the optimal field prediction obtained from LSC for the purpose of removing its inherent smoothing effect, while sustaining most of its local prediction accuracy. In general terms, we seek a "de-smoothing" transformation to act upon the LSC predictor, $\hat{\mathbf{u}}' = \Re(\hat{\mathbf{u}})$, such that the CV structure of the primary SRF is recovered. This means that the transformation $\Re(\cdot)$ should guarantee that

$$\mathbf{C}_{\hat{\mathbf{u}}'} = \mathbf{C}_{\mathbf{u}} \tag{10}$$

where $\mathbf{C}_{\mathbf{u}}$ is the CV matrix formed through the CV function $C_u(P, Q)$ of the primary SRF; see (7).

In addition, the prediction errors $\mathbf{e}' = \hat{\mathbf{u}}' - \mathbf{u}$ associated with the field predictor $\hat{\mathbf{u}}'$ should remain small in some sense, so that the new solution can provide not only a CV-adaptive representation for the SRF variation patterns, but also locally accurate predicted values on the basis of the given data. For this purpose, the formulation of the operator $\Re(\cdot)$ should additionally incorporate some kind of optimality principle by minimizing, for example, the trace of the new error CV matrix $\mathbf{C}_{\mathbf{e}'}$.

Let us now introduce a straightforward linear approach to modify the LSC predictor $\hat{\mathbf{u}}$, through

$$\hat{\mathbf{u}}' = \mathbf{R}\,\hat{\mathbf{u}} \tag{11}$$

where **R** is a square filtering matrix that needs to be determined according to some optimal criteria for the new predictor $\hat{\mathbf{u}}'$, including its CV-matching property given in (10).

The predicted field obtained from (11) should reproduce the CV structure of the primary SRF, in the sense that $\mathbf{C}_{\hat{\mathbf{u}}'} = \mathbf{C}_{\mathbf{u}}$ for the given spatial distribution of all prediction points $\{P_i'\}$. Hence, the filtering matrix **R** has to satisfy the constraint

$$\mathbf{R}\mathbf{C}_{\hat{\mathbf{u}}}\mathbf{R}^{\mathrm{T}} = \mathbf{C}_{\mathbf{u}} \tag{12}$$

where $\mathbf{C}_{\mathbf{u}}$ and $\mathbf{C}_{\hat{\mathbf{u}}}$ correspond to the CV matrices of the primary and the LSC-predicted SRFs.

The assessment of the prediction accuracy of the new solution $\hat{\mathbf{u}}'$ can be made through its error CV matrix

$$\mathbf{C}_{\mathbf{e}'} = E\{(\hat{\mathbf{u}}' - \mathbf{u})(\hat{\mathbf{u}}' - \mathbf{u})^{\mathrm{T}}\} \tag{13}$$

which, taking (11) into account, yields

$$\mathbf{C}_{\mathbf{e}'} = \mathbf{R}\mathbf{C}_{\hat{\mathbf{u}}}\mathbf{R}^{\mathrm{T}} + \mathbf{C}_{\mathbf{u}} - \mathbf{R}\mathbf{C}_{\hat{\mathbf{u}}\mathbf{u}} - \mathbf{C}_{\mathbf{u}\hat{\mathbf{u}}}\mathbf{R}^{\mathrm{T}} \tag{14}$$

Using (8) and the following relation that is always valid for the LSC predictor $\hat{\mathbf{u}}$ (assuming that there is zero correlation between the observed signals **s** and the measurement noise **v**)

$$\mathbf{C}_{\hat{\mathbf{u}}\mathbf{u}} = \mathbf{C}_{\hat{\mathbf{u}}} \tag{15}$$

the new error CV matrix can be finally expressed as

$$\mathbf{C}_{\mathbf{e}'} = \mathbf{C}_{\mathbf{e}} + (\mathbf{I} - \mathbf{R})\mathbf{C}_{\hat{\mathbf{u}}}(\mathbf{I} - \mathbf{R})^{\mathrm{T}} \tag{16}$$

where $\mathbf{C}_{\mathbf{e}}$ is the error CV matrix of the LSC solution. Evidently, the prediction accuracy of the modified solution $\hat{\mathbf{u}}'$ will always be worse than the prediction accuracy of the original LSC solution $\hat{\mathbf{u}}$, regardless of the form of the filtering matrix **R**. This is expected since LSC provides the best (in the

MSE sense) unbiased linear predictor from the available measurements, which cannot be further improved by additional linear operations. Nevertheless, our aim is to determine an optimal filtering matrix that satisfies the CV-matching constraint in (12), while minimizing the loss of the MSE prediction accuracy in the recovered SRF, in the sense that

$$trace(\mathbf{C}_{\mathbf{e}'} - \mathbf{C}_{\mathbf{e}}) = trace(\delta\mathbf{C}_{\mathbf{e}'}) = \text{minimum} \qquad (17)$$

where $\delta\mathbf{C}_{\mathbf{e}'} = (\mathbf{I}-\mathbf{R})\mathbf{C}_{\hat{\mathbf{u}}}(\mathbf{I}-\mathbf{R})^{\mathrm{T}}$ represents the part of the error CV matrix of the new predictor $\hat{\mathbf{u}}'$ which depends on the choice of the filtering matrix.

The determination of the filtering matrix **R** that (i) satisfies the CV-matching constraint (12), and (ii) minimizes the loss in the MSE prediction accuracy of the predictor $\hat{\mathbf{u}}'$ according to (17), is analytically described in Kotsakis (2007). Due to space limits, we will only present the final result herein, without going into any technical details regarding its mathematical proof. The optimal filtering is

$$\mathbf{R} = \mathbf{C}_{\mathbf{u}}^{1/2}(\mathbf{C}_{\mathbf{u}}^{1/2}\mathbf{C}_{\hat{\mathbf{u}}}\mathbf{C}_{\mathbf{u}}^{1/2})^{-1/2}\mathbf{C}_{\mathbf{u}}^{1/2} \qquad (18)$$

or equivalently (see Appendix)

$$\mathbf{R} = \mathbf{C}_{\hat{\mathbf{u}}}^{-1/2}(\mathbf{C}_{\hat{\mathbf{u}}}^{1/2}\mathbf{C}_{\mathbf{u}}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2})^{-1/2}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2} \qquad (19)$$

Note that the above result was originally derived in Eldar (2001) under a completely different context than the one discussed in this paper, focusing on applications such as matched-filter detection, quantum signal processing and signal whitening.

4 Numerical Test

A numerical example is presented in this section to demonstrate the performance of the CV-adaptive predictor $\hat{\mathbf{u}}'$, in comparison with the classical LSC predictor $\hat{\mathbf{u}}$. The particular test refers to a standard noise filtering problem for a set of simulated gridded gravity anomaly data. The image shown in Fig. 1(a) is the actual realization of a free-air gravity anomaly field which has been simulated within an 50×50 km^2 area and with a uniform sampling resolution of 2 km, according to the following model of the spatial CV function

$$C_u(P,Q) = \frac{C_o}{1+(r_{PQ}/a)^2} \qquad (20)$$

where $C_o = 220\,\text{mgal}^2$, r_{PQ} is the planar distance between points *P* and *Q*, and the parameter *a* is selected such that the correlation length of the gravity anomaly field is equal to 7 km.

The noisy data grid is shown in Fig. 1(b), with the underlying noise level being equal to ±15 mgals. Note that the additive random errors have been simulated as a set of uncorrelated random variables, thus enforcing a white noise assumption for the gridded data. In Fig. 1(d) we see the filtered signal as obtained from the classical LSC algorithm (i.e. Wiener filtering), whereas in Fig. 1(c) is shown the result obtained from the CV-adaptive solution $\hat{\mathbf{u}}'$. It is seen that, although LSC provides in principle the most accurate (in the MSE sense) filtered signal, the result obtained from the CV-adaptive predictor clearly looks more similar to the original SRF that is depicted in Fig. 1(a). The emulation of the spatial variability of the primary SRF by the CV-adaptive solution $\hat{\mathbf{u}}'$, in contrast to the smoothed representation obtained by the LSC predictor $\hat{\mathbf{u}}$, can also be seen in the histograms plotted in Fig. 2, as well as in the signal statistics listed in Table 1.

Table 1. Statistics of the actual (simulated) signal $\mathbf{u}$, the LSC-filtered signal $\hat{\mathbf{u}}$ and the CV-matching filtered signal $\hat{\mathbf{u}}'$ (all values in mgals).

	Max	Min	Mean	σ
Actual grid values	45.33	-42.88	-0.04	±14.96
LSC solution	27.42	-29.10	0.74	±10.29
CV-matching solution	41.47	-42.48	0.73	±14.64

5 Conclusions

Due to its inherent smoothing effect, the LSC prediction algorithm does not reproduce the spatial correlation structure implied by the CV function of the primary SRF that needs to be recovered from its observed functionals. The method presented in this paper offers an alternative approach for optimal SRF prediction, which preserves the signal's spatial variability as dictated by its known CV function.

Evidently, the rationale of the proposed technique relies on the knowledge of the *true* CV function of the underlying SRF, an assumption which is also inbuilt in the theoretical development of the classical LSC method (Moritz 1980). In practice, an *empirical* signal CV function is often first estimated from a given and possibly noisy data record, and then used in the implementation of the LSC proce-

dure for the (sub-optimal) recovery of the primary SRF at a set of prediction points. For such cases, it is reasonable to question whether it would be meaningful to let the spatial variability of the LSC-predicted field to be adapted to an empirical CV function by following the CV-matching approach presented in this paper. A more reasonable methodology would be to additionally incorporate a *variance component estimation* approach, in a way that the final predicted field becomes adapted to an "improved" model of spatial variability (i.e. with respect to the one imposed by the empirical CV function). In contrast to the standard CV-matching constraint introduced in (10), we can impose in this case the alternative CV-tuning constraint

$$\mathbf{C}_{\hat{\mathbf{u}}'} = \sigma^2 \, \mathbf{Q}_{\mathbf{u}} \tag{21}$$

where the known CV matrix $\mathbf{Q}_{\mathbf{u}}$ is formed through the empirically determined signal CV function, and σ^2 is an unknown variance factor which controls the consistency between the empirical and the true CV function for the underlying unknown signal.

Tackling the above problem along with the study of one-step CV-matching linear predictors (see Schaffrin 1997, 2002), instead of the two-step constructive approach that was presented herein, may be an interesting subject for future investigation.

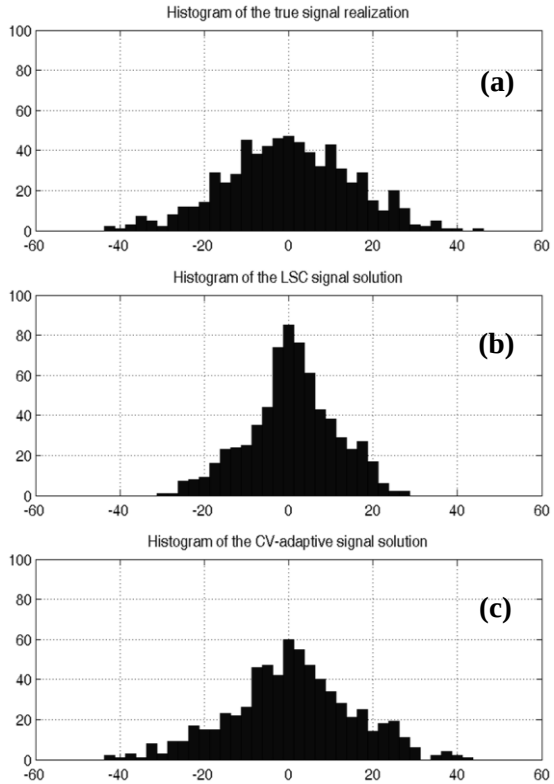

Figure 2. Histograms of the actual (simulated) gravity anomaly signal (a), the LSC-filtered signal (b), and the CV-matching filtered signal (c).

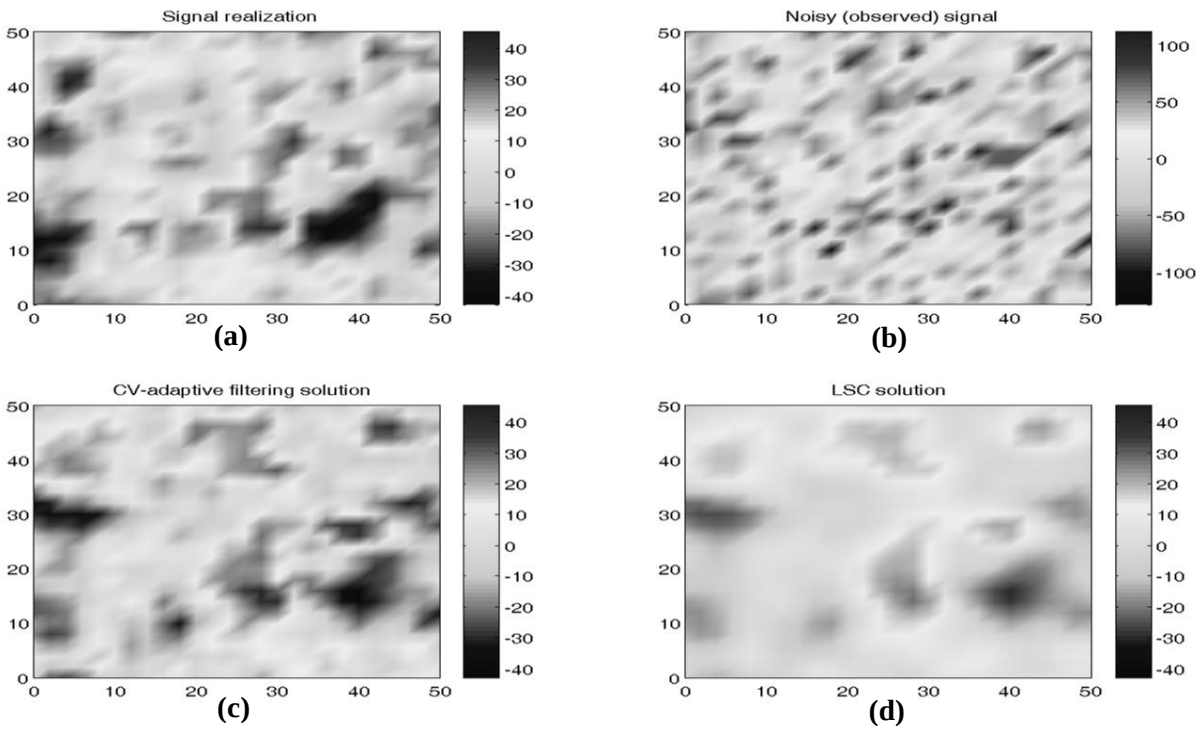

Figure 1. Plots of the actual (simulated) gravity anomaly signal (a), the noisy observed signal (b), the CV-matching filtered signal (c), and the LSC-filtered signal (d).

Acknowledgements. The author would like to acknowledge the valuable suggestions and comments by the two reviewers, as well as the constructive criticism provided by the responsible editor Prof. Burkhard Schaffrin.

References

Aldworth J, Cressie N (2003) Prediction of nonlinear spatial functionals. J Stat Plan Inf, 112: 3-41.

Christakos G (1992) Random Field Models for Earth Sciences. Academic Press, New York, NY.

Cressie N (1993) Aggregation in geostatistical problems. In A. Soares (ed) *Geostatistics Troia 1992*, vol 1, pp. 25-36, Kluwer, Dordrecht.

Dermanis A (1976) Probabilistic and deterministic aspects of linear estimation in geodesy. Rep No. 244, Dept. of Geodetic Science, Ohio State University, Columbus, Ohio.

Eldar YC (2001) Quantum signal processing. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA.

Krarup T (1969) A contribution to the mathematical foundation of physical geodesy. Danish Geodetic Institute, Report no. 47, Copenhagen.

Kotsakis C (2007) Least-squares collocation with covariance-matching constraints. J Geod, DOI 10.1007/s00190-007-0133-5.

Moritz H (1970) Least-squares estimation in physical geodesy. Deutsche Geodätische Kommission, Reihe A, Heft 69, Muenchen.

Moritz H (1980) Advanced Physical Geodesy. Herbert Wichmann Verlag, Karlsruhe.

Sanso F (1986) Statistical Methods in Physical Geodesy. In (H. Suenkel, ed.): Mathematical and Numerical Techniques in Physical Geodesy, Lecture Notes in Earth Sciences, vol 7, pp. 49-156, Springer Verlag, Berlin.

Schaffrin B (1997) On suboptimal geodetic network fusion. Poster presented at the IAG Scientific Assembly, September 3-9, Rio de Janeiro, Brazil.

Schaffrin B (2002) Reproducing estimates via least-squares: An optimal alternative to Helmert transformation. In: Grafarend EW, Krumm FW, Schwarze VS (eds) *Geodesy – The Challenge of the 3rd Millennium*, pp. 387-392, Springer, Berlin, Heidelberg, New York.

Appendix

In this appendix, we will establish the equivalency between the two forms of the optimal filtering matrix **R** that were given in (18) and (19), respectively.

Starting from (18) and using the following matrix identity (which is easy to verify for all invertible matrices **S** and **T**)

$$(\mathbf{ST})^{-1/2} = \mathbf{S}(\mathbf{TS})^{-1/2}\mathbf{S}^{-1} \tag{A1}$$

we have

$$\begin{aligned}
\mathbf{R} &= \mathbf{C}_{\mathbf{u}}^{1/2}(\mathbf{C}_{\mathbf{u}}^{1/2}\mathbf{C}_{\hat{\mathbf{u}}}\mathbf{C}_{\mathbf{u}}^{1/2})^{-1/2}\mathbf{C}_{\mathbf{u}}^{1/2} \\
&= \mathbf{C}_{\mathbf{u}}^{1/2}(\underbrace{\mathbf{C}_{\mathbf{u}}^{1/2}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2}}_{\mathbf{S}}\underbrace{\mathbf{C}_{\hat{\mathbf{u}}}^{1/2}\mathbf{C}_{\mathbf{u}}^{1/2}}_{\mathbf{T}})^{-1/2}\mathbf{C}_{\mathbf{u}}^{1/2} \\
&= \mathbf{C}_{\mathbf{u}}^{1/2}\underbrace{(\mathbf{C}_{\mathbf{u}}^{1/2}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2})}_{\mathbf{S}}\underbrace{(\mathbf{C}_{\hat{\mathbf{u}}}^{1/2}\mathbf{C}_{\mathbf{u}}^{1/2}}_{\mathbf{T}}\underbrace{\mathbf{C}_{\mathbf{u}}^{1/2}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2})}_{\mathbf{S}}{}^{-1/2}\times \\
&\quad \times\underbrace{(\mathbf{C}_{\mathbf{u}}^{1/2}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2})}_{\mathbf{S}}{}^{-1}\mathbf{C}_{\mathbf{u}}^{1/2} \\
&= \mathbf{C}_{\mathbf{u}}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2}(\mathbf{C}_{\hat{\mathbf{u}}}^{1/2}\mathbf{C}_{\mathbf{u}}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2})^{-1/2}\mathbf{C}_{\hat{\mathbf{u}}}^{-1/2}\mathbf{C}_{\mathbf{u}}^{-1/2}\mathbf{C}_{\mathbf{u}}^{1/2} \\
&= \mathbf{C}_{\hat{\mathbf{u}}}^{-1/2}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2}\mathbf{C}_{\mathbf{u}}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2}(\mathbf{C}_{\hat{\mathbf{u}}}^{1/2}\mathbf{C}_{\mathbf{u}}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2})^{-1/2}\mathbf{C}_{\hat{\mathbf{u}}}^{-1/2} \\
&= \mathbf{C}_{\hat{\mathbf{u}}}^{-1/2}(\mathbf{C}_{\hat{\mathbf{u}}}^{1/2}\mathbf{C}_{\mathbf{u}}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2})(\mathbf{C}_{\hat{\mathbf{u}}}^{1/2}\mathbf{C}_{\mathbf{u}}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2})^{-1/2}\mathbf{C}_{\hat{\mathbf{u}}}^{-1/2} \\
&= \mathbf{C}_{\hat{\mathbf{u}}}^{-1/2}(\mathbf{C}_{\hat{\mathbf{u}}}^{1/2}\mathbf{C}_{\mathbf{u}}\mathbf{C}_{\hat{\mathbf{u}}}^{1/2})^{1/2}\mathbf{C}_{\hat{\mathbf{u}}}^{-1/2}
\end{aligned} \tag{A2}$$

The last expression in the above equation is identical to the matrix form given in (19), and thus the equivalency between (18) and (19) has been established.