Parametric versus non-parametric methods for optimal weighted averaging of noisy data sets

C. Kotsakis
Department of Geomatics Engineering, University of Calgary
2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4, Email: ckotsaki@ucalgary.ca

I.N. Tziavos
Department of Geodesy and Surveying, Aristotle University of Thessaloniki
University Box 440, GR-54124, Thessaloniki, Greece, Email: tziavos@olimpia.topo.auth.gr

Abstract. Averaging is a simple, yet very effective, technique that can be used for the fusion of repeated data sets and/or multiple estimates of a single parameter vector. With the increasing number of geodetic data sensors and the availability of numerous geodetic signal realizations from various satellite missions, averaging offers an important tool which can significantly reduce the data load while improving the quality of the recovered signal information. A preliminary study of weighted averaging methods for multivariate data ensembles is given in this paper, along with a novel approach for optimal weight determination directly from the available data.

1. Introduction

The problem of *optimal weighted averaging* of data sets that correspond to multiple noisy recordings of a given signal is studied in this paper. Being one of the earliest least-squares applications, the method of averaging provides the simplest and often most efficient technique for noise filtering when several measurements of the same quantity are available. It can also serve as a trend determination and/or removal tool from different random field realizations, as well as a combination strategy in cases where various estimates of an unknown signal (or vector) need to be merged into a single optimal solution. Examples from the use of averaging techniques in geodesy can be found in the analysis and stacking of repeat-track altimetric data (Knudsen, 1993; Sailor, 1994), in the combination procedures employed by IGS to produce precise clock and satellite orbit solutions from the individual submissions of its Analysis Centers (Beutler et al., 1995; Kouba et al., 1995), and in the recovery of topographic gradients from stacked repeat-pass SAR interferograms without phase unwrapping (Sandwell and Sichoix, 2000). We should also mention another important geodetic paradigm whose strong dependence on the proper application of an averaging algorithm has been often overlooked, namely the integrated statistical processing of orthometric, ellipsoidal and geoidal heights for vertical datum and gravimetric geoid refinement studies (Kotsakis, 2003).

Generalizing the traditional estimation problem in statistics, where the expectation of a single random variable is approximated by the sample mean of a set of its observed values, *signal* or *vector averaging* can be used for the analysis and fusion of multi-variate data ensembles. The issue of proper weighting of the different ensembles becomes especially important in this case, since a *single* weight factor may not be able to account for: (i) the noise correlation within each data set, or (ii) the noise cross-correlation between different data sets. This can be very critical in applications where the errors affecting the repeated realizations of a geodetic signal are significantly correlated at various spatial and/or temporal scales (e.g., atmospheric effects). Furthermore, in many cases the statistical characteristics of the data errors are only poorly known, thus introducing additional difficulties for choosing the appropriate data weights.

In this paper, we present a preliminary treatment of the weighted averaging problem for multivariate data ensembles, with the emphasis put on two different approaches for the optimal determination of the ensemble weights. The first approach, which is useful for theoretical or simulation studies, assumes the complete knowledge of the second-order statistics for the data noise. Some examples for the error behaviour of the corresponding optimal averaging estimator are presented to demonstrate the importance of proper weighting when mixing data sets with varying accuracy levels. The second approach is oriented towards more practical situations and it is based on an 'empirical' estimation criterion that allows the determination of

the optimal ensemble weights directly from the available data.

2. Problem formulation

Let us assume that a set of N multivariate noisy observations (ensembles) of a given signal is available. For notational simplicity these ensembles are represented as column vectors $\mathbf{x}_i$ in an m-dimensional *data space* $\Re^m$, where m is the number of data values in each recorded signal realization. Note that unless we are dealing with 1D problems, this formulation requires the choice of a one-to-one mapping between the physical signal description (e.g., the raster scan of a digital image) and the components of the data vectors $\mathbf{x}_i$ (ensemble vectorization).

Assuming additive noise in each data set, the following observation equations can now be formed:

$$\mathbf{x}_i = \boldsymbol{\mu} + \mathbf{v}_i \ , \quad i = 1, 2, ..., N \tag{1}$$

where the vector $\boldsymbol{\mu}$ contains the true values of the underlying signal. The random error vectors $\mathbf{v}_i$ have zero mean and they are statistically described in terms of their auto- and cross-covariance matrices

$$\begin{aligned} \mathbf{C}_{ii} &= E\{\mathbf{v}_i \mathbf{v}_i^{\mathrm{T}}\}, \\ \mathbf{C}_{ij} &= E\{\mathbf{v}_i \mathbf{v}_j^{\mathrm{T}}\} \qquad i, j = 1, 2, ..., N \end{aligned} \tag{2}$$

where $E\{\cdot\}$ is the mathematical expectation operator. Note that all vector(-ized) ensembles are assumed to have the same dimensions and they correspond to the same spatial (or temporal) sampling configuration.

A linear estimator of the unknown signal from its repeated realizations has the general form

$$\hat{\boldsymbol{\mu}} = \sum_{i=1}^{N} \mathbf{W}_i \mathbf{x}_i \tag{3}$$

The last formula can easily be identified as a weighted sum of the ensembles $\{\mathbf{x}_i\}$, with $\{\mathbf{W}_i\}$ being appropriate weight matrices that need to be determined in some optimal sense. Due to the limited extent of this paper, we will restrict our study to the case where the weight matrices have the simple diagonal form

$$\mathbf{W}_i = w_i \mathbf{I} \tag{4}$$

where $\mathbf{I}$ denotes the $m \times m$ unit matrix. In this way, the linear estimator of Eq. (3) is simplified to

$$\hat{\boldsymbol{\mu}} = \sum_{i=1}^{N} w_i \mathbf{x}_i \tag{5}$$

where a single weight coefficient w_i is now assigned to each data set. The treatment of the more general case of weighted averaging, according to Eq. (3), will soon be published elsewhere.

Remark: The previous formulation is also suited for problems where the data sets $\mathbf{x}_i$ are not obtained by direct repeated observations of the same field, but they correspond to different a-priori estimates or solutions of an unknown vector quantity. The various ensembles $\{\mathbf{x}_i\}$ then need to be optimally fused into a unique combined solution $\hat{\boldsymbol{\mu}}$ through a weighted averaging procedure according to Eq. (3) or Eq. (5). In such cases, the covariance (CV) matrices given in Eq. (2) represent the accuracy of each individual vector estimate and they are usually determined from separate data adjustment procedures. The cross-CV matrices $\mathbf{C}_{ij}$ should normally vanish if the prior estimates for $\boldsymbol{\mu}$ have been obtained independently from each other. A typical example of this type of problem can be found in the combination strategies implemented by IGS for the determination of precise orbit and clock solutions (Beutler et al., 1995).

3. Rigorous weight determination

The computation of the weight factors in Eq. (5) should be based on a well-defined criterion that optimizes the output signal estimate. In this section, a rigorous *mean square error* (MSE) criterion is adopted for the optimal determination of the unknown weights. Denoting by $\mathbf{e}$ the actual error of the linear estimator in Eq. (5), i.e.

$$\mathbf{e} = \hat{\boldsymbol{\mu}} - \boldsymbol{\mu} \tag{6}$$

the corresponding MSE is a certain functional that quantifies in an average sense the magnitude of the above estimation error. If we use the standard Euclidean vector norm, then it can be shown that the MSE can be expressed by the following form:

$$\begin{aligned} E\{\|\mathbf{e}\|^2\} &= E\{\mathbf{e}^{\mathrm{T}}\mathbf{e}\} \\ &= \mathbf{w}^{\mathrm{T}} \widetilde{\mathbf{C}} \mathbf{w} + \boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{\mu} \left[\left(\sum_{i=1}^{N} w_i \right) - 1 \right]^2 \end{aligned} \tag{7}$$

where the vector $\mathbf{w}$ contains the values of the scalar weights (w_i), and the matrix $\widetilde{\mathbf{C}}$ is defined by the formula

$$\widetilde{\mathbf{C}}:\ [\widetilde{C}_{ij}] = trace\,\mathbf{C}_{ij} \qquad i, j = 1, 2, ..., N \qquad (8)$$

If we further impose the condition that the result of Eq. (5) should lead to an *unbiased* signal estimate, then we are faced with the following constrained optimization problem:

$$\mathbf{w}^{\mathrm{T}}\widetilde{\mathbf{C}}\mathbf{w} = minimum\,,\ \text{subject to}\ \sum_{i=1}^{N} w_i = 1 \qquad (9)$$

Using the method of Lagrange multipliers, a unique weight solution can be obtained as follows:

$$\hat{\mathbf{w}} = \frac{1}{\mathbf{s}^{\mathrm{T}}\widetilde{\mathbf{C}}^{-1}\mathbf{s}}\widetilde{\mathbf{C}}^{-1}\mathbf{s} \qquad (10)$$

where $\mathbf{s}$ is an auxiliary vector whose elements are all ones. The minimum MSE value that corresponds to the above optimal weight vector is

$$MSE_{\min} = (\mathbf{s}^{\mathrm{T}}\widetilde{\mathbf{C}}^{-1}\mathbf{s})^{-1} \qquad (11)$$

For comparison, it is interesting to give also the value that the MSE takes when a *uniform* weighting scheme is followed (i.e. $w_i = 1/N$). In such a case we have

$$MSE\ (\text{for equal weights}) = (\mathbf{s}^{\mathrm{T}}\widetilde{\mathbf{C}}\mathbf{s})/N^2 \qquad (12)$$

Note that the last expression is generally sub-optimal and it decays much slower, for increasing sample size N, than the optimal error value given in Eq. (11); see also Sect. 4.

In the special case where the matrix $\widetilde{\mathbf{C}}$ is <u>diagonal</u> (i.e. the noise cross-correlation between different data sets is zero, $\mathbf{C}_{ij} = \mathbf{0}$), the optimal weights associated with the ensembles $\mathbf{x}_i$ are given by the formula

$$\hat{w}_i = \left[\frac{1}{\sum_{k=1}^{N}\left(1/trace\,\mathbf{C}_{kk}\right)}\right]\frac{1}{trace\,\mathbf{C}_{ii}} \qquad (13)$$

whereas the resultant minimum MSE value is

$$MSE_{\min} = \frac{1}{\sum_{k=1}^{N}\left(1/trace\,\mathbf{C}_{kk}\right)} \qquad (14)$$

Again, for comparison purposes, the MSE value obtained by using equal weight coefficients (in uncorrelated data ensembles) is given below

$$MSE\ (\text{for equal weights}) = \frac{\sum_{k=1}^{N} trace\,\mathbf{C}_{kk}}{N^2} \qquad (15)$$

Remark: The MSE of the weighted average $\hat{\boldsymbol{\mu}}$, defined by Eq. (7), is insensitive to the noise correlation that may exist *within* each data set $\mathbf{x}_i$. That is because the off-diagonal elements of the CV matrices $\mathbf{C}_{ii}$, which contain the statistical information for the noise correlation within every ensemble, do not enter into the definition of the matrix $\widetilde{\mathbf{C}}$ according to Eq. (8). As a result, an averaging scheme that employs only scalar weighting coefficients may not be suitable for data sets that have highly correlated errors, since the individual weights obtained by the MSE minimization in Eq. (9) cannot reflect this type of information.

The *cross-ensemble* noise correlation, on the other hand, is not totally ignored in the optimal weights given in Eq. (10). The MSE definition encompasses <u>the traces of all cross-CV matrices</u> $\mathbf{C}_{ij}$, and thus the weight optimization according to Eq. (9) takes into account the noise correlation at every data point between different ensembles. Obviously, any statistical information on the noise cross-correlation between different ensembles <u>and</u> different data points is excluded from the optimal weights of Eq. (10).

4. Error analysis for the weighted average: an example

A simple example is presented in this section to demonstrate the importance of proper weighting in data averaging procedures. A series of 40 different ensembles, all corresponding to the same gravity signal profile, is considered. The ensembles are assumed to be affected by additive stationary noise whose standard deviation varies among them according to the simulated values shown in Figure 1.

Assuming also that the data noise is uncorrelated between the different ensembles, we can use Eq. (14) to plot the cumulative MSE of the optimal weighted average $\hat{\boldsymbol{\mu}}$, as the number of the processed ensembles (N) increases from 2 up to 40. In this case, we will have that

$$trace\,\mathbf{C}_{kk} = m\sigma_k^2 \qquad (16)$$

where σ_k is the noise standard deviation of the k-th ensemble (see Figure 1), and m denotes the constant

number of data points within each ensemble. In a similar way, Eq. (15) can be used to determine the cumulative MSE of the unweighted (or, more precisely, uniformly weighted) signal average $\hat{\boldsymbol{\mu}}$ as a function of the number of the averaged ensembles.

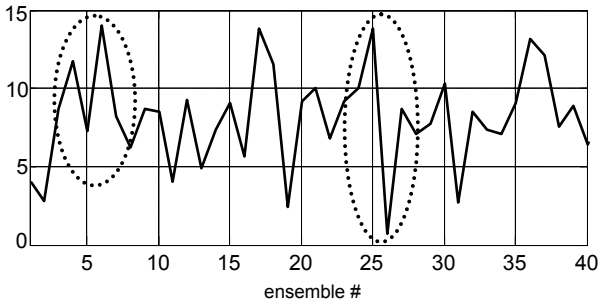

Figure 1. Simulated values (in μgal) for the noise standard deviation in the various gravity signal ensembles.

The results for both cases are shown in Figure 2, which clearly reveals the problems of uniform weighting for data sets with uneven accuracy levels. It is interesting to point out the two regions that are circled in the plots of Figure 2. The first (upper left) region shows that the accuracy of the unweighted average can actually worsen whilst more data sets are used for its computation! As seen in Figure 1, the noise level in the first two data sets is much smaller than in the next few ensembles. This affects the MSE of the unweighted average $\hat{\boldsymbol{\mu}}$ accordingly when the more noisy ensembles enter into the averaging procedure. On the other hand, if the varying noise level in the different ensembles is properly taken into account through an optimal weighting scheme (see Sect. 3), then the accuracy of the averaged estimate will be consistently improving as more data sets are processed. Nonetheless, the MSE of the unweighted average seems to 'stabilize' and starting to decrease constantly when more precise data sets enter into the averaging procedure, although its decay rate is still slower than in the optimal weighting case.

In the second circled region shown in Figure 2, we notice the drastic MSE reduction in the optimal weighted average when a very accurate data set is used to update $\hat{\boldsymbol{\mu}}$ (see also the corresponding circled area in Figure 1 where the decrease in the noise level between the 25th and the 26th ensemble is indicated). In the case of the uniformly weighted average (dotted line) no such considerable improvement occurs, since the equal weighting of all data sets cannot offer any benefit from the use of a significantly more accurate signal ensemble.

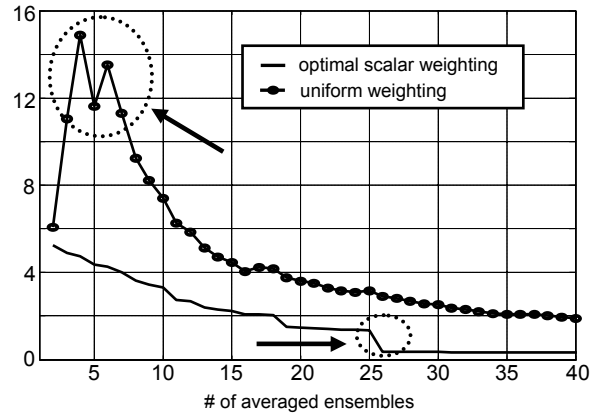

Figure 2. Cumulative MSE (in μgal^2) for the weighted signal average using: (i) optimal weights, and (ii) equal weights.

The effect of the noise correlation (between different signal ensembles) on the weighted average $\hat{\boldsymbol{\mu}}$

The different gravity signal ensembles are again presumed to be affected by additive stationary noise whose standard deviation varies according to the values of Figure 1. In addition, it is assumed that a pointwise cross-ensemble correlation of the data noise exists, for all pairs of the available ensembles, with a constant correlation coefficient value of ρ= 0.7. This is a purely theoretical scenario, which nevertheless allows us to easily simulate a fully populated matrix $\tilde{\mathbf{C}}$ that is needed for a rigorous error analysis of the weighted average $\hat{\boldsymbol{\mu}}$ according to Eqs. (11) and (12). In this case, the off-diagonal elements of $\tilde{\mathbf{C}}$ will be given by the simple formula

$$[\tilde{C}_{ij}] = trace\,\mathbf{C}_{ij} = m\rho\sigma_i\sigma_j \qquad (17)$$

where σ_i and σ_j are the noise standard deviations of the *i*-th and *j*-th ensemble, respectively. Three different schemes for computing the weighted average $\hat{\boldsymbol{\mu}}$ are now considered, namely using: (i) the optimal weights according to Eq. (10), (ii) equal weights, and (iii) the weights obtained from Eq. (13) which take into account only the different noise levels in the various data sets (i.e. they neglect the cross-ensemble noise correlation).

As before, the cumulative MSEs of the corresponding signal averages are computed and plotted against the number of the averaged ensembles (*N*). The results are shown in Figure 3, where we can see the 'explosive' behavior of the resulting MSE for the uniformly weighted average.

Even in the case where the weights account for the uneven noise levels among the various data sets (dashed line), the corresponding weighted average still seems to be unstable as the number of processed ensembles increases. This is due to the neglected noise cross-correlation between the different ensembles, which in practice can cause serious problems in the fusion of repeated geodetic sets (e.g., sea surface topography from single/multiple altimetry missions, gravity and gradiometry maps from single/multiple satellite missions, etc.).

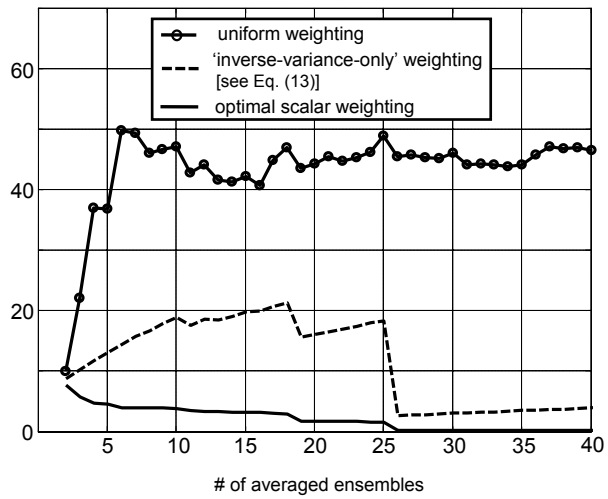

Figure 3. Cumulative MSE (in μgal^2) for the weighted signal average in the case of correlated ensembles ($\tilde{\mathbf{C}}$ is a fully populated matrix).

Note that Figures 2 and 3 show normalized MSE values, since the results of Eqs. (11), (12), (14) and (15) have been divided by the number of data points (m) within every ensemble.

5. Empirical weight determination

The weight optimization procedure that was described in Sect. 3 requires the knowledge of the noise CV matrices for all available data sets. A more realistic approach for the determination of the ensemble weights should consider that the noise statistics are often unknown in practice (or, at least, partially known). This applies also for the case where the CV matrices in Eq. (2) represent not the statistical variability of actual measurement noise, but the accuracy of some a-priori vector estimates $\mathbf{x}_i$ which need to be merged into a single solution $\hat{\boldsymbol{\mu}}$.

In this section, an alternative optimization method is presented which results in ensemble weights that can be directly computed from the available data. It can thus be characterized as a *non-parametric* method (as opposed to the parametric approach of Sect. 3), since any a-priori assumptions for the statistical properties of the various data sets are completely avoided.

Data weighting based on a maximum-SNR criterion

The criterion that will be now used for the optimal determination of the ensemble weights is based on the maximization of an 'empirical' signal-to-noise ratio (SNR) expression. In general, the SNR that corresponds to the averaging estimator of Eq. (5) has the form

$$SNR = \frac{E_s}{E_n} = \frac{\hat{\boldsymbol{\mu}}^{\mathrm{T}}\hat{\boldsymbol{\mu}}}{E_n} \tag{18}$$

where E_s and E_n denote the Euclidean norms of the estimated signal and the data noise, respectively.

Two different methods can be followed to quantify the data noise norm. In particular, we have

Case I

$$\mathbf{e}_i = \mathbf{x}_i - \hat{\boldsymbol{\mu}}\ , \quad i = 1, 2, \ldots, N \tag{19}$$

$$E_n = \sum_i w_i \left\| \mathbf{e}_i \right\|^2 = \sum_i w_i \left(\mathbf{x}_i - \hat{\boldsymbol{\mu}}\right)^{\mathrm{T}} \left(\mathbf{x}_i - \hat{\boldsymbol{\mu}}\right) \tag{20}$$

or an alternative approach,

Case II

$$\mathbf{e}_i^{(w)} = w_i \mathbf{x}_i - \hat{\boldsymbol{\mu}}\ , \quad i = 1, 2, \ldots, N \tag{21}$$

$$E_n = \sum_i \left\| \mathbf{e}_i^{(w)} \right\|^2 = \sum_i \left(w_i \mathbf{x}_i - \hat{\boldsymbol{\mu}}\right)^{\mathrm{T}} \left(w_i \mathbf{x}_i - \hat{\boldsymbol{\mu}}\right) \tag{22}$$

The previous approaches are considered 'empirical' in the sense that they both quantify the data noise not with respect to the true unknown signal $\boldsymbol{\mu}$, but with reference to its optimal estimate $\hat{\boldsymbol{\mu}}$. Note that *Case I* provides a straightforward generalization of the weighted least-squares error norm for the case of multivariate measurements $\{\mathbf{x}_i\}$.

Based on Eqs. (19) and (20), the SNR of the weighted signal average for *Case I* is

$$SNR_{(\mathrm{I})} = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{R}\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{D}\mathbf{s} - \mathbf{w}^{\mathrm{T}}\mathbf{R}\mathbf{w}} \tag{23}$$

where $\mathbf{R}$ is a symmetric Grammian-type matrix that consists of the Euclidean inner products $(\mathbf{x}_i, \mathbf{x}_j)$ between all possible pairs of data sets

$$\mathbf{R}: \ [R_{ij}] = \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j \qquad i, j = 1, 2, ..., N \tag{24}$$

and $\mathbf{D}$ is a diagonal matrix obtained from $\mathbf{R}$ by setting to zero its off-diagonal elements

$$\mathbf{D}: \ [D_{ii}] = \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_i \qquad i = 1, 2, ..., N \tag{25}$$

The maximization of the SNR expression in Eq. (23) yields a unique solution that is given by the closed formula

$$\underline{\textit{Case I}}: \qquad \hat{\mathbf{w}} = \frac{1}{\mathbf{s}^{\mathrm{T}} \mathbf{R}^{-1} \mathbf{D} \mathbf{s}} \mathbf{R}^{-1} \mathbf{D} \mathbf{s} \tag{26}$$

As it can be seen from the last equation, the optimal weight vector can now be computed directly from the available data, without any other knowledge of their statistical properties.

If the second formulation for the data noise norm E_n is used (*Case II*), then the corresponding SNR of the weighted signal average takes the form

$$SNR_{(\mathrm{II})} = \frac{\mathbf{w}^{\mathrm{T}} \mathbf{R} \mathbf{w}}{\mathbf{w}^{\mathrm{T}} \mathbf{D} \mathbf{w} - \mathbf{w}^{\mathrm{T}} \mathbf{R} \mathbf{w}} \tag{27}$$

which is slightly different from the previous SNR expression in Eq. (23). Its maximization leads to a *generalized eigenvalue problem*, as follows:

$$\underline{\textit{Case II}}: \qquad \mathbf{R} \mathbf{w} = \lambda \mathbf{D} \mathbf{w} \tag{28}$$

In this case, the optimal weight vector $\hat{\mathbf{w}}$ is just the eigenvector corresponding to the maximum eigenvalue λ_{max} in Eq. (28); more mathematical details and proofs will be given in an upcoming journal publication. Note that generalized eigenvalue problems can be easily converted to standard eigenvalue problems (i.e. $\mathbf{A} \mathbf{x} = \lambda \mathbf{x}$) through simple linear transformations; see Hairville (1997, pp. 562-564). Most high-end scientific computing packages, such as Matlab™, contain also built-in routines for solving directly this type of algebraic problems.

Remark: Both of the 'empirically optimal' weight solutions described above will produce unbiased signal estimates when used in the averaging estimator of Eq. (5). That is because the sum of their values $\{w_i\}$ is, or it can be easily made after a simple global rescaling, equal to one. Note that this does not exclude cases where the optimal weight solution contains a mixture of both positive and negative values, as long as their total sum is still equal to one! Actually, such cases will arise often when the data noise shows significant cross-correlation among the different signal ensembles.

6. Conclusions

The practice of averaging multiple realizations of a given signal, either in the form of repeatedly observed ensembles or as a-priori estimates obtained from different individual procedures, is a valuable tool in geodetic data analysis. In this paper we have briefly presented a general treatment of the weighted averaging problem for multivariate data sets. Our main focus has been on: (i) the optimal determination of the ensemble weights (with and without the knowledge of the data noise statistics), and (ii) the error behaviour of the signal average for different weighting schemes. In terms of future work we plan to present a more detailed version of the study outlined herein that will include all of the omitted mathematical details, as well as more sophisticated weighting schemes for data averaging such as the use of fully populated weight matrices $\mathbf{W}_i$ and the simultaneous bias estimation/removal from the data sets $\{\mathbf{x}_i\}$.

References

Beutler G, Kouba J, Springer T (1995) Combining the orbits of the IGS Analysis Centers. *Bulletin Géodésique*, 69:200-222.

Hairville DA (1997) *Matrix Algebra from a Statistician's Perspective*. Springer-Verlag, New York.

Knudsen P (1993) Altimetry for geodesy and oceanography. In *Lecture notes for the NKG-Autumn School 1992* (J. Kakkuri, ed.): "Geodesy and Geophysics", Report of the Finnish Geodetic Institute, No. 115, pp. 87-130, Helsinki.

Kouba J, Mireault Y, Lahaye F (1995) IGS Orbit/Clock Combination and Evaluation (Appendix I of the Analysis Coordinator Report), *International GPS Service for Geodynamics (IGS) 1994 Annual Report*, pp. 70-94.

Kotsakis C (2003) Some remarks on the integrated statistical processing of GPS, levelling and geoid data (in preparation).

Sailor RV (1994) Signal Processing Techniques. In *Geoid and its Geophysical Interpretations* (P. Vanicek and N.T. Christou, eds.), pp. 147-186, CRC Press.

Sandwell DT, Sichoix L (2000) Topographic phase recovery from stacked ERS interferometry and a low resolution digital elevation model. *Journal of Geophysical Research*, vol. 105, no. B12, pp. 28,211-28,222.