On the Approximation of the Inverse Error Covariances of High-Resolution Satellite Altimetry Data

Max Yaremchuk  
Naval Research Laboratory

Joseph M. D'Addezio  
University of Southern Mississippi, joseph.daddezio@usm.edu

Gleb Panteleev  
Naval Research Laboratory, gleb@iarc.uaf.edu

Gregg Jacobs  
Naval Research Laboratory

Follow this and additional works at: https://aquila.usm.edu/fac_pubs

Part of the Meteorology Commons

Recommended Citation
Available at: https://aquila.usm.edu/fac_pubs/18142

This Letter to the Editor is brought to you for free and open access by The Aquila Digital Community. It has been accepted for inclusion in Faculty Publications by an authorized administrator of The Aquila Digital Community. For more information, please contact Joshua.Cromwell@usm.edu.
Approximation of the inverse observational error covariances

M. Yaremchuk et al
Naval Research Laboratory, Stennis Space Center, USA
Department of Marine Science, University of Southern Mississippi, USA
Naval Research Laboratory, 1009 Balch Blvd., Stennis Space Center, 39522, MS, USA. E-mail: max.yaremchuk@nrlssc.navy.mil

High-resolution (swath) altimeter missions scheduled to monitor the ocean surface in the near future have observation error covariances (OECs) with slowly decaying off-diagonal elements. This property presents a challenge for the majority of the data assimilation (DA) algorithms which were designed under the assumption of the diagonal OECs being easily inverted. In this note, we present a method of approximating the inverse of a dense OEC by a sparse matrix represented by the polynomial of spatially inhomogeneous differential operators, whose coefficients are optimized to fit the target OEC by minimizing a quadratic cost function. Explicit expressions for the cost function gradient and the Hessian are derived. The method is tested with an OEC model generated by the SWOT simulator.

covariance modelling; data assimilation; observational data analysis; wide swath altimetry

On the approximation of the inverse error covariances of high resolution satellite altimetry data
M. Yaremchuk 1, J.M. D’Addezio 2, G. Panteleev 1 and G. Jacobs 1
Fri Jun 29 09:35:55 2018

1 Introduction

Over the last several decades, representation of the background error covariances by the polynomials of the diffusion operator has been extensively studied in both meteorological and oceanographic DA applications (e.g., [Derber and Rosati(1988)], [Weaver et al(2003)], [Xu(2005)], [Yaremchuk and Smith(2011)], [Yaremchuk et al(2013)]). Among the advantages are

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/qj.3336
the computational efficiency of the approach and its ability to preserve the positive semi-
definite (psd) property of the resulting background error covariance (BEC) estimates. The method proves to be especially useful in heuristic modeling of the climatological (static) BECs which supplement the ensemble-based estimates of the background errors.

In contrast to the BECs, observational error covariances (OECs) are conventionally represented by diagonal matrices under the implicit assumption that observation errors are weakly correlated at spatial scales exceeding the grid step of the numerical models. This assumption, being realistic for most current observational platforms, provides an additional convenience of inexpensive computation of the inverse OECs and of their square roots currently employed by the majority of DA algorithms (e.g., [Cummings(2005)], [Hunt et al(2007)], [Fairbairn et al(2014)]).

Recent developments in high-resolution ("swath") altimetry ([Durand et al(2010)], [Ito et al(2014)], [Ubelmann et al(2015)], [Gaultier et al(2016)]) introduce challenges for data assimilation. Novel issues emerge due to both the data density that may approach model grid scales in the horizontal, and because observational errors at such high resolutions appear to be highly correlated in space [Ruggiero et al(2016)] (hereinafter R16) due to the design of the satellite and sensor. At the same time, recent studies (e.g., [Stewart et al(2013)], [Miyoshi et al(2013)], [Waller et al(2014)]) demonstrate substantial benefits of accounting for spatial correlations of the observation errors even in low-dimensional DA systems. As a consequence, these newly arriving data require special treatment in order to maintain skill and retain the computational efficiency of the DA schemes.

In most of the DA algorithms, the difference between the model sea surface height and the data has to be multiplied by either the inverse of the observation error covariance $R$ or by its inverse square root $R^{-1/2}$, so a computationally efficient representation $R_i$ of these matrices (e.g., by a sparse matrix) is highly desirable. This note contributes to the methodology of approximating $R^{-1}$ (or its square root) using differential operators. Specifically, with the forthcoming Surface Water and Ocean Topography (SWOT) altimeter mission in mind, we postulate that an estimate of $R$ is available on a regular 2-dimensional set of $N = n_x \times n_y$ observation grid points from the simulating software of [Ubelmann et al(2017)].

2 Approximating the inverse covariance

2.1 The matrix parameterization model

In what follows, we present a methodology of parameterizing $R_i$ by a linear combination of sparse matrices with matrix-valued coefficients represented by discretized differential operators. To simplify the notation, the method is illustrated by a particular
example relevant to the SWOT application.

Specifically, we consider $R_i$ of the form

$$R_i = A + \nabla^T B \nabla + \Delta C \Delta \tag{1}$$

where $\nabla$ is the $2N - n_x - n_y \times N$ matrix, representing the first-order approximation of the gradient operator on a 2d grid, $\Delta = \nabla^T \nabla$ is the Laplacian, and $A, B, C$ are sparse control matrices. Their $M$ non-zero elements populating the vector $x \in \mathbb{R}^M$ have to be optimized by minimization of the quadratic cost function, measuring the Frobenius norm $|| \cdot ||_F$ of the respective residual:

$$J = ||R_i R - I||^2_F = \text{tr}[P P^T] \to \min x \tag{2}$$

Here $I$ is the identity matrix and notation $P = R_i R - I$ is introduced. The first variation of (2) has the form

$$\delta J = 2 \text{tr}[\delta PP^T] = 2 \text{tr}[\delta R_i R P^T] \tag{3}$$

(a): map of the sum of two columns of the SWOT covariance matrix $R$ (column positions shown by squares) and its approximation (b) by $R_i^{-1}$. Panel (c) shows the spectra of the SWOT covariance (thin black line) and its approximation. Covariance values are divided by 100 cm$^2$

Taking into account that $\delta R_i = \delta A + \nabla^T \delta B \nabla + \Delta \delta C \Delta$, and introducing the notation $Q = 2PR_i$, transforms $\delta J$ to

$$\delta J = \text{tr}[\delta AQ^T + \delta C \Delta Q^T \Delta] + \text{tr}[\delta B \nabla Q^T \nabla] \tag{4}$$

so that the respective expressions for the gradient are:

$$\frac{\delta J}{\delta A} = Q; \quad \frac{\delta J}{\delta B} = \nabla Q \nabla^T; \quad \frac{\delta J}{\delta C} = \Delta Q \Delta \tag{5}$$

and the system of equations $\delta J/\delta x = 0$ defining the minimum of $J$ takes the form

$$R_i R^2 = R \tag{6}$$
$$\nabla R_i R^2 \nabla^T = \nabla R \nabla^T \tag{7}$$
$$\Delta R_i R^2 \Delta = \Delta R \Delta \tag{8}$$

Equations (6-8) can be rewritten explicitly in terms of the Hessian matrix $H \in \mathbb{R}^{M \times M}$ and the rhs vector $r \in \mathbb{R}^M$:

$$H x = r \tag{9}$$

where $r$ contains non-zero elements in the right-hand sides of (6-8) listed columnwise as in Eq. 20 of the Appendix.

In this note we consider the simplest sparsity pattern for $A, B$ and $C$, assuming that the matrices are diagonal so that their sparsity patterns are the identity matrices of the respective size. In this case, the general expression for the Hessian (see eq. (21) in the Appendix)
simplifies to
\[ H = \begin{bmatrix} R^2 \circ I & R^2 \circ \nabla^T \circ \nabla^T & R^2 \Delta \circ \Delta \\ \nabla R^2 \circ \nabla & \nabla R^2 \circ \nabla \nabla^T & \nabla R^2 \Delta \circ \nabla \Delta \\ \Delta R^2 \circ \Delta & \Delta R^2 \circ \nabla \nabla^T & \Delta R^2 \Delta \circ \Delta^2 \end{bmatrix} \] (10)

where \( \circ \) denotes Hadamard (element-wise) matrix product. The latter relationship could be useful for constructing block-diagonal preconditioners for the iterative solvers of (9), or for direct solution of (9) on the moderate-size \((N < 10^5)\) grids.

### 2.2 Model reduction

The system of equations (9) may not be well conditioned, so its solution should be sought using a certain parameterization of the original control variables \( x \). In the linear case, such parameterization can be expressed in terms of a projection operator \( \Pi \)

\[ x = \Pi x \] (11)

where \( m \) columns of \( \Pi \) contain the "structure functions", parameterizing spatial variability of \( x \), and \( x \) stands for the vector of the reduced control variables. For instance, to enforce smoothness of the diagonal elements of the control matrices, these functions can be represented by \( m \) smoothest harmonics, the first one being independent on horizontal coordinates. Although in the present study we did not employ any model reduction \((\Pi = I)\), more sophisticated projections (e.g. [Brankart et al, (2009)], R16) can be employed (see Appendix). Using non-trivial structure of \( \Pi \) requires, however, prior information on the spatial variability of the control fields in order to maintain a reasonable balance between the accuracy of the approximation of the target matrix and numerical efficiency.

The reduced normal system \( Hx = r \) is characterized by the reduced Hessian \( H \in \mathbb{R}^{m \times m} \) and the reduced rhs \( r \in \mathbb{R}^m \):

\[ H = \Pi^T H \Pi; \quad r = \Pi^T r \] (12)

Since \( \frac{\delta J}{\delta x} = Hx - r \), the cost function gradient (4) can be conveniently estimated by applying the projection operator to (5):

\[ \frac{\delta J}{\delta x} = \Pi^T \frac{\delta J}{\delta x} \] (13)

Note that since the relationships (5) are valid for arbitrary control matrices, equations (11-13) could be used in optimization algorithms employing gradient information under arbitrary linear constraints on spatial variability of the control matrix elements. In particular, the approach could be useful for maintaining the psd property of \( R_i \) in the process of minimization.

### 3 Numerical testing

#### 3.1 SWOT covariance model
The ansatz (1) for the approximation of the inverse observation error covariance was tested with the target OEC generated by the Jet Propulsion Laboratory’s (JPL) SWOT simulator of [Ubelmann et al(2017)]. The simulator generates realizations of SSH observation error fields based on the latest estimate of the SWOT error budget of [Esteban-Fernandez(2013)]. The error field contains six constituents: Ka-band radar interferometer noise, wet tropospheric error, and errors associated with uncertainties in the estimation of roll, phase, baseline, and timing of the SWOT observational platform. With a reasonable degree of accuracy, the first two error fields can be considered to be uncorrelated. The remaining four error sources are of particular interest because they are highly correlated over large spatial scales: As shown by [Ruggiero et al(2016)], these errors are characterized by typical decorrelation scales of several hundred kilometers along the swath and approximately a hundred across, with the marginal pointwise pdfs being very close to Gaussian.

Same as in Figure 1, but the inverse covariance model is described by eqns. (12) and (14-20).

3.2 Results

In generating the target OEC matrix, we used the SWOT simulator version 2.0.0 with the following parameters: the default cut off wavelength of 40,000 km and a 2 beam wet tropospheric error correction. Additionally, anticipating large decorrelation scales (compared to the projected SWOT resolution of 1-2 km) and the absence of smaller-scale spatial variability in the matrix columns, we elected 40 and 10 km sampling in the along- and across track directions respectively. This selection also decreased the influence of uncorrelated Ka-band noise on the OEC structure. 5,000 random realizations of all error sources summed together were generated by having the simulator recursively sample the same 21 day repeat orbit over a subdomain of the Western Pacific (116 °E-133 °E, 18 °N-34 °N) with a total sampled track length of 2000 km and width of 140 km. Excluding the grid points in the 20 km wide nadir gap, the sampled OEC field dimensions were \( n_x = 14, n_y = 51, (N = n_x n_y = 714) \) with the total number of adjusted degrees of freedom \( 4N - n_x - n_y = 2,791 \) and the number of the independent elements in the target covariance matrix \( N(N + 1)/2 = 255,255 \). The resulting error fields were characterized by approximately Gaussian pointwise pdfs with the average magnitude of the means \( \sim 10^{-4} \) m and the standard deviations varying between of 0.03 m near the nadir gap and 0.31 m at the swath periphery.

Figure 1a shows half the sum of the OEC fields corresponding to the pair of SSH observations located on both edges of the nadir gap in the center of the sampled track. Similar to the results of R16, covariance structures at intermediate scales are barely visible. However, there is a strong anisotropy of the covariance with the typical OEC spatial scales in the along-
and across-track directions differing by an order in magnitude (600 km and 60 km respectively).

Figure 1a and b demonstrate the result of approximating $R$ by the inverse covariance model (1). Due to the limited number of ensemble members a slight asymmetry (of the order of 1%) has been observed in the structure of the mirror rows of $R$. Figure 1b shows that this asymmetry is considerably enhanced in the approximating matrix $R_i^{-1}$ (cf. Fig. 1a, b). The effect is caused by the coarse resolution of the nadir gap which is only 2 grid steps wide, and associated errors in the finit-difference approximation of $R^{-1}$ by the ansatz (1).

Due to the modest dimension of the control space ($N = 2,791$) and low condition number ($\text{cond}(H)=2 \cdot 10^5$) of the Hessian matrix (10), the optimization took a few seconds on a single CPU of a PC using the MatLab sparse system solver. As it is seen, the algorithm provides a reasonably accurate fit to the leading eigenmodes of $R$ (Figure 1c) with the relative error $\text{tr}(R_i^{-1} - R)/\text{tr}(R)$ of 22%.

As a matter of comparison, we performed approximation of $R^{-1}$ in the reduced 5-dimensional space proposed by [Ruggiero et al(2016)], who assigned a fixed spatial variability to the diagonals of the control matrices and minimized (2) by varying five diagonal scaling factors $x$ (see Appendix). In this procedure, we employed the technique of Sections 2.1-2.2, which can be viewed as a generalization of the computational approach of [Ruggiero et al(2016)] who used five-fold expansion of the data space by computing the derivatives of the error fields in SWOT simulator output instead of explicit computation of the Hessian (eq. 19 in Appendix) and its projection on the reduced control space (eq. 12).

Figure 2 shows the results of the reduced space optimization. As it is seen, the reduced method provides a poorer fit to the SWOT spectrum being tested (cf. Fig. 1c and 2c) and a larger error in approximating the columns of the SWOT covariance matrix (cf. Fig. 1b,2b and Fig. 1a). This should be attributed to lesser flexibility of the reduced procedure, as the number of adjusted parameters is approximately $2,791/5 \approx 560$ times smaller compared to the case of full optimization involving solution of eq. (6-8).

4 Summary and discussion

In recent years, there has been an increased interest in inverse OEC modeling due to the high-resolution swath altimetry missions planned in the near future [Durand et al(2010)], [Ito et al(2014)], [Ichikawa l(2014)]. Although this new type of observational platform is characterized by improved accuracy (1-2 cm) and higher spatial resolution (1-2 km), the respective OECs are expected to be highly correlated in space. This property presents a computational challenge for many operational DA algorithms that are based on the diagonal OECs.

This note proposes a methodology of approximating the inverse OECs by a polynomial in differential operators acting on sparse control matrices whose non-zero elements are adjusted to minimize the Frobenius norm of the approximation error. Explicit relationships for the cost
function gradient and the Hessian matrix of the optimization problem have been obtained for control matrices with fixed sparsity patterns. A method of reduction of the optimization problem has been demonstrated for the case of the degenerate Hessian. The proposed approach could be used in realistic data assimilation systems by replacing the code normalizing model-data misfits by observation error variances with the code multiplying the misfits by a sparse matrix retrieved from an estimate of the respective error covariance.

Further developments of the approach can be foreseen in several directions. First, the method does not maintain the positive semi-definite (psd) property of the approximation matrix in the process of optimization. The psd constraint can be imposed in many ways if the method is restricted to the diagonal control matrices. A straightforward way is to constrain all the components of the control vector to be positive in the process of optimization. This approach (combined with the projection technique) was used by [Ruggiero et al(2016)] to ensure the psd property. A somewhat more sophisticated methodology is based on factorizing $R_i = LL^T$ and representing $L$ as a composite of (sparse) control matrices: $L = [L_0 \ \nabla^T L_1 \ ...]^T$ with diagonal controls $L_{i,i} > 0.1 \ldots$. This option, however, destroys the attractive quadratic property of the optimization problem. More general approaches going beyond the fixed sparsity patterns of the controls can also be explored (e.g., [Hsieh et al(2014)]). Our experience with the presented version of the method have shown, however, that optimal $R_i$ was very close to psd with only a few negative eigenvalues that contributed less than 0.1% to the trace of $R_i$.

The cost function could also be defined by $J = \|I - R_i RR_i\|_F^2$ to directly retrieve a sparse approximation to $R_i^{-1/2}$ that may be more useful in the DA applications. Our numerical experiments with this formulation have shown that one has to pay more attention to initialization of the control variables, because starting the quasi-newtonian descent from $x = 0$ proved inefficient for several simulated classes of OECs. In contrast, the considered quadratic/diagonal formulation (1) performed well and never encountered convergence/conditioning problems for the same classes of OECs.

Elaboration of an efficient reduction scheme also remains an important issue. [Ruggiero et al (2016)] have shown that certain OECs can be efficiently approximated with just a few parameters, if an appropriate projection method is elected. In particular, useful information on the structure of $\Pi$ could be retrieved from the structure of the diagonal cells of the Hessian matrices. An alternative way of regularizing the problem is to augment (2) with the terms which penalize high-frequency variations of the control variables. However, the respective low-pass filter should be designed with caution, as the high-frequency variations of the inverse matrix elements (partly simulated by the differential operators) are a key component of the optimized matrix.

We believe that further studies of the matrix approximation methodologies in application to the class of psd matrices with slowly varying spatial structure has good prospects in the future development of DA techniques in geophysical applications and may benefit more
This study was supported by Office of Naval Research projects (Program Elements 0646352N, 0602435N). J. D’Addezio was supported by the Naval Research Laboratory Cooperative Agreement BAA-N00173-03-13-01 awarded to the University of Southern Mississippi. Helpful discussions with Prof. C. Beattie are acknowledged.

5 Appendix

[Ruggiero et al(2016)] proposed a simplified method of estimating the inverse of $R$ through the adjustment of only five free parameters. In the notation of section 2, their scheme involves a combination of the inverse covariance model containing four diagonal matrices

$$R_i = A + \nabla^T \nabla + \Delta_x \Delta_x + \Delta_y \Delta_y$$

with a projection scheme, which assigns a certain spatial structure to the matrices $A, B, C$ and $D$. In (14) the operators $\Delta_x$ and $\Delta_y$ stand for the cross- and along-track constituents of the Laplacian: $\Delta = \Delta_x + \Delta_y$.

Specifically, the projection adopted in R16 is defined by

$$A = \alpha_0 (R \circ I)^{-1}$$

$$B = \alpha_{1c} (\partial_x^2 R \circ \partial_x^T \circ I)^{-1}$$

$$C = \alpha_{2c} (\Delta_x \Delta_x \circ I)^{-1}$$

$$D = \alpha_{2a} (\Delta_y \Delta_y \circ I)^{-1}$$

with the control vector $x = [\alpha_0 \alpha_{1c} \alpha_{1a} \alpha_{2c} \alpha_{2a}]^T$ and the projection operator $\Pi$ represented by $5N - n_x - n_y \times 5$ block-diagonal matrix containing the inverse matrices in right-hand sides of eq. (15-18) on the diagonal.

The system of equations $Hx = r$ is

$$H = \begin{bmatrix} R \circ I & R^2 \circ I & R^2 \circ \nabla \circ \nabla^T & R^2 \circ \nabla \circ \Delta_x \circ \nabla^T & R^2 \circ \Delta_y \circ \nabla \circ \nabla^T & R^2 \circ \Delta_y \circ \Delta_y \circ \nabla \circ \nabla^T \\ \nabla R \circ \nabla & \nabla R \circ \nabla \circ \nabla^T & \nabla \Delta_x \circ \Delta_x \circ \nabla \circ \nabla^T & \nabla \Delta_x \circ \Delta_x \circ \Delta_x \circ \nabla \circ \nabla^T & \nabla \Delta_y \circ \nabla \circ \Delta_y \circ \nabla \circ \nabla^T & \nabla \Delta_y \circ \nabla \circ \Delta_y \circ \Delta_y \circ \nabla \circ \nabla^T \\ \Delta_x \circ \Delta_x & \Delta_x \circ \nabla \circ \nabla \circ \Delta_x \circ \nabla \circ \nabla^T & \Delta_x \circ \Delta_x \circ \nabla \circ \nabla \circ \Delta_x \circ \nabla \circ \nabla^T & \Delta_x \circ \Delta_x \circ \nabla \circ \nabla \circ \Delta_x \circ \nabla \circ \nabla^T & \Delta_x \circ \Delta_x \circ \Delta_x \circ \nabla \circ \nabla \circ \nabla^T & \Delta_x \circ \Delta_x \circ \Delta_x \circ \Delta_x \circ \nabla \circ \nabla^T \\ \Delta_y \circ \Delta_y & \Delta_y \circ \nabla \circ \nabla \circ \Delta_y \circ \nabla \circ \nabla^T & \Delta_y \circ \Delta_y \circ \nabla \circ \nabla \circ \Delta_y \circ \nabla \circ \nabla^T & \Delta_y \circ \Delta_y \circ \nabla \circ \nabla \circ \Delta_y \circ \nabla \circ \nabla^T & \Delta_y \circ \Delta_y \circ \Delta_y \circ \nabla \circ \nabla \circ \nabla^T & \Delta_y \circ \Delta_y \circ \Delta_y \circ \Delta_y \circ \nabla \circ \nabla^T \end{bmatrix}$$

$$r = \begin{bmatrix} R \circ I \\ \nabla R \circ \nabla \circ \nabla^T \circ I \\ \Delta_x \circ \Delta_x \circ \nabla \circ \nabla^T \circ I \\ \Delta_y \circ \Delta_y \circ \nabla \circ \nabla^T \circ I \\ \Delta_x \circ \Delta_x \circ \Delta_x \circ \nabla \circ \nabla^T \circ I \end{bmatrix}$$

It is noteworthy, that the ansatz (14) produced a degenerate Hessian (19) which did not allow us to compare the results of full optimizations with the inverse OEC models (1) and (14). Reduction of the control space regularized the problem, but resulted in a relatively poor fit to
the spectrum of the tested covariance (cf. Fig. 1c and Fig. 2c).

As a final note, we present the general expression for the Hessian matrix associated with the column-vectorized form of (6-8). Defining the sparsity pattern $S$ of a matrix $A$ by replacing non-zero elements of $A$ with ones, and adopting the notation $S_{AB} = \text{vec}(S \otimes A) \otimes \text{vec}(S \otimes B^T)$ for mutual Kronecker products of the sparsity patterns, the Hessian is given by

$$H = \begin{bmatrix}
(R^2 \otimes I) \circ S_{AA} & (R^2 \Delta \otimes \Delta) \circ S_{AB} & (R^2 \Delta \otimes \Delta) \circ S_{AC} \\
(R^2 \Delta \otimes \Delta) \circ S_{BA} & (R^2 \Delta \otimes \Delta) \circ S_{BB} & (R^2 \Delta \otimes \Delta) \circ S_{BC} \\
(\Delta R^2 \otimes \Delta) \circ S_{CA} & (\Delta R^2 \otimes \Delta) \circ S_{CB} & (\Delta R^2 \otimes \Delta) \circ S_{CC}
\end{bmatrix}.$$  \tag{21}

References


This article is protected by copyright. All rights reserved.


Approximation of the inverse observational error covariances

Naval Research Laboratory, Stennis Space Center, USA
Department of Marine Science, University of Southern Mississippi, USA
Naval Research Laboratory, 1009 Balch Blvd., Stennis Space Center, 39522, MS, USA. Email: max.yaremchuk@nrlssc.navy.mil

High-resolution (swath) altimeter missions scheduled to monitor the ocean surface in the near future have observation error covariances (OECs) with slowly decaying off-diagonal elements. This property presents a challenge for the majority of the data assimilation (DA) algorithms which were designed under the assumption of the diagonal OECs being easily inverted. In this note, we present a method of approximating the inverse of a dense OEC by a sparse matrix represented by the polynomial of spatially inhomogeneous differential operators, whose coefficients are optimized to fit the target OEC by minimizing a quadratic cost function. Explicit expressions for the cost function gradient and the Hessian are derived. The method is tested with an OEC model generated by the SWOT simulator.

covariance modelling; data assimilation; observational data analysis; wide swath altimetry

On the approximation of the inverse error covariances of high resolution satellite altimetry data

M. Yaremchuk 1, J.M. D’Addezio 2, G. Panteleev 1 and G. Jacobs 1
Mon Jun 25 12:03:01 2018

1 Introduction

Over the last several decades, representation of the background error covariances by the polynomials of the diffusion operator has been extensively studied in both meteorological and oceanographic DA applications (e.g., [Derber and Rosati(1988)], [Weaver et al(2003)], [Xu(2005)], [Yaremchuk and Smith(2011)], [Yaremchuk et al(2013)]). Among the advantages are the computational efficiency of the approach and its ability to preserve the positive semidefinite (psd) property of the resulting background error covariance (BEC) estimates. The method proves to be especially useful in heuristic modeling of the climatological (static) BECs which supplement the ensemble-based estimates of the background errors.

In contrast to the BECs, observational error covariances (OECs) are conventionally

This article is protected by copyright. All rights reserved.
represented by diagonal matrices under the implicit assumption that observation errors are weakly correlated at spatial scales exceeding the grid step of the numerical models. This assumption, being realistic for most current observational platforms, provides an additional convenience of inexpensive computation of the inverse OECs and of their square roots currently employed by the majority of DA algorithms (e.g., [Cummings(2005)], [Hunt et al(2007)], [Fairbain et al(2014)]).

Recent developments in high-resolution ("swath") altimetry ([Durand et al(2010)], [Ito et al(2014)], [Ubelmann et al(2015)], [Gaultier et al(2016)]) introduce challenges for data assimilation. Novel issues emerge due to both the data density that may approach model grid scales in the horizontal, and because observational errors at such high resolutions appear to be highly correlated in space [Ruggiero et al(2016)] (hereinafter R16) due to the design of the satellite and sensor. At the same time, recent studies (e.g., [Stewart et al(2013)], [Miyoshi et al(2013)], [Waller et al (2014)]) demonstrate substantial benefits of accounting for spatial correlations of the observation errors even in low-dimensional DA systems. As a consequence, these newly arriving data require special treatment in order to maintain skill and retain the computational efficiency of the DA schemes.

In most of the DA algorithms, the difference between the model sea surface height and the data has to be multiplied by either the inverse of the observation error covariance \( R \) or by its inverse square root \( R^{-1/2} \), so a computationally efficient representation \( R_i \) of these matrices (e.g., by a sparse matrix) is highly desirable. This note contributes to the methodology of approximating \( R^{-1} \) (or its square root) using differential operators. Specifically, with the forthcoming Surface Water and Ocean Topography (SWOT) altimeter mission in mind, we postulate that an estimate of \( R \) is available on a regular 2-dimensional set of \( N = n_x \times n_y \) observation grid points from the simulating software of [Ubelmann et al(2017)].

## 2 Approximating the inverse covariance

### 2.1 The matrix parameterization model

In what follows, we present a methodology of parameterizing \( R_i \) by a linear combination of sparse matrices with matrix-valued coefficients represented by discretized differential operators. To simplify the notation, the method is illustrated by a particular example relevant to the SWOT application.

Specifically, we consider \( R_i \) of the form

\[
R_i = A + \nabla^T B \nabla + \Delta C \Delta
\]

where \( \nabla \) is the \( 2N - n_x - n_y \times N \) matrix, representing the first-order approximation of the gradient operator on a 2d grid, \( \Delta = \nabla^T \nabla \) is the Laplacian, and \( A, B, C \) are sparse control matrices. Their \( M \) non-zero elements populating the vector \( x \in \mathbb{R}^M \) have to be optimized by minimization of the quadratic cost function, measuring the Frobenius norm \( \| \cdot \|_F \) of the respective residual:

\[
I = \| R_i R - I \|_F^2 = \text{tr}[P P^T] \to \min x
\]
Here $I$ is the identity matrix and notation $P = R_i R - I$ is introduced. The first variation of (2) has the form
\[ \delta J = 2 \text{tr}[\delta P P^T] = 2 \text{tr}[\delta R_i R P^T] \] (3)

Taking into account that $\delta R_i = \delta A + \nabla^T \delta B \nabla + \Delta \delta C \Delta$ and introducing the notation $Q = 2P R$, transforms $\delta J$ to
\[ \delta J = \text{tr}[\delta A Q^T + \delta C \Delta Q^T \Delta] + \text{tr}[\delta B \nabla Q^T \nabla^T] \] (4)

so that the respective expressions for the gradient are:
\[ \frac{\delta J}{\delta A} = Q; \quad \frac{\delta J}{\delta B} = \nabla Q \nabla^T; \quad \frac{\delta J}{\delta C} = \Delta Q \Delta \] (5)

and the system of equations $\delta J / \delta x = 0$ defining the minimum of $J$ takes the form
\[ R_i R^2 = R \] (6)
\[ \nabla R_i R^2 \nabla^T = \nabla R \nabla^T \] (7)
\[ \Delta R_i R^2 \Delta = \Delta R \Delta \] (8)

Equations (6-8) can be rewritten explicitly in terms of the Hessian matrix $H \in \mathbb{R}^{M \times M}$ and the rhs vector $r \in \mathbb{R}^M$
\[ H x = r \] (9)

where $r$ contains non-zero elements in the right-hand sides of (6-8) listed columnwise as in Eq. 20 of the Appendix.

In this note we consider the simplest sparsity pattern for $A$, $B$ and $C$, assuming that the matrices are diagonal so that their sparsity patterns are the identity matrices of the respective size. In this case, the general expression for the Hessian (see eq. (21) in the Appendix) simplifies to
where $\circ$ denotes Hadamard (element-wise) matrix product. The latter relationship could be useful for constructing block-diagonal preconditioners for the iterative solvers of (9), or for direct solution of (9) on the moderate-size ($N < 10^4$) grids.

2.2 Model reduction

The system of equations (9) may not be well conditioned, so its solution should be sought using a certain parameterization of the original control variables $x$. In the linear case, such parameterization can be expressed in terms of a projection operator $\Pi$,

$$x = \Pi x$$

(11)

where $m$ columns of $\Pi$ contain the "structure functions", parameterizing spatial variability of $x$, and $x$ stands for the vector of the reduced control variables. For instance, to enforce smoothness of the diagonal elements of the control matrices, these functions can be represented by $m$ smoothest harmonics, the first one being independent on horizontal coordinates. Although in the present study we did not employ any model reduction ($\Pi = I$), more sophisticated projections (e.g. [Brankart et al, (2009)], R16) can be employed (see Appendix). Using non-trivial structure of $\Pi$ requires, however, prior information on the spatial variability of the control fields in order to maintain a reasonable balance between the accuracy of the approximation of the target matrix and numerical efficiency.

The reduced normal system $Hx = r$ is characterized by the reduced Hessian $H \in \mathbb{R}^{m \times m}$ and the reduced rhs $r \in \mathbb{R}^m$:

$$H = \Pi^T \Pi; \quad r = \Pi^T r$$

(12)

Since $\delta J/\delta x = Hx - r$, the cost function gradient (4) can be conveniently estimated by applying the projection operator to (5):

$$\frac{\delta J}{\delta x} = \Pi^T \frac{\delta J}{\delta x}$$

(13)

Note that since the relationships (5) are valid for arbitrary control matrices, equations (11-13) could be used in optimization algorithms employing gradient information under arbitrary linear constraints on spatial variability of the control matrix elements. In particular, the approach could be useful for maintaining the psd property of $R_{\delta x}$ in the process of minimization.

3 Numerical testing

3.1 SWOT covariance model

The ansatz (1) for the approximation of the inverse observation error covariance was tested with the target OEC generated by the Jet Propulsion Laboratory’s (JPL) SWOT simulator of [Ubelmann et al(2017)]. The simulator generates realizations of SSH observation error fields based on the latest estimate of the SWOT error budget of [Esteban-Fernandez(2013)]. The error field contains six constituents: Ka-band radar interferometer noise, wet tropospheric error, and
errors associated with uncertainties in the estimation of roll, phase, baseline, and timing of the SWOT observational platform. With a reasonable degree of accuracy, the first two error fields can be considered to be uncorrelated. The remaining four error sources are of particular interest because they are highly correlated over large spatial scales: As shown by [Ruggiero et al (2016)], these errors are characterized by typical decorrelation scales of several hundred kilometers along the swath and approximately a hundred across, with the marginal pointwise pdfs being very close to Gaussian.

Same as in Figure 1, but the inverse covariance model is described by eqns. (12) and (14-20).

### 3.2 Results

In generating the target OEC matrix, we used the SWOT simulator version 2.0.0 with the following parameters: the default cut off wavelength of 40,000 km and a 2 beam wet tropospheric error correction. Additionally, anticipating large decorrelation scales (compared to the projected SWOT resolution of 1-2 km) and the absence of smaller-scale spatial variability in the matrix columns, we elected 40 and 10 km sampling in the along- and across track directions respectively. This selection also decreased the influence of uncorrelated Ka-band noise on the OEC structure. 5,000 random realizations of all error sources summed together were generated by having the simulator recursively sample the same 21 day repeat orbit over a subdomain of the Western Pacific (116 °E-133 °E, 18 °N-34 °N) with a total sampled track length of 2000 km and width of 140 km. Excluding the grid points in the 20 km wide nadir gap, the sampled OEC field dimensions were \( n_x = 14, n_y = 51 \), \( N = n_x n_y = 714 \) with the total number of adjusted degrees of freedom \( 4N - n_x - n_y = 2,791 \) and the number of the independent elements in the target covariance matrix \( N(N + 1)/2 = 255,255 \). The resulting error fields were characterized by approximately Gaussian pointwise pdfs with the average magnitude of the means \( \sim 10^{-4} \) m and the standard deviations varying between of 0.03 m near the nadir gap and 0.31 m at the swath periphery.

Figure 1a shows half the sum of the OEC fields corresponding to the pair of SSH observations located on both edges of the nadir gap in the center of the sampled track. Similar to the results of R16, covariance structures at intermediate scales are barely visible. However,
there is a strong anisotropy of the covariance with the typical OEC spatial scales in the along- 
and across-track directions differing by an order in magnitude (600 km and 60 km respectively).

Figure 1a and b demonstrate the result of approximating $R$ by the inverse covariance 
model (1). Due to the limited number of ensemble members a slight asymmetry (of the order of 
1%) has been observed in the structure of the mirror rows of $R$. Figure 1b shows that this 
asymmetry is considerably enhanced in the approximating matrix $R_i^{-1}$ (cf. Fig. 1a, b). The effect 
is caused by the coarse resolution of the nadir gap which is only 2 grid steps wide, and 
associated errors in the finite-difference approximation of $R^{-1}$ by the ansatz (1).

Due to the modest dimension of the control space ($N = 2,791$) and low condition 
number ($\text{cond}(H)=2 \cdot 10^4$) of the Hessian matrix (10), the optimization took a few seconds on a 
single CPU of a PC using the MatLab sparse system solver. As it is seen, the algorithm provides a 
reasonably accurate fit to the leading eigenmodes of $R$ (Figure 1c) with the relative error 
$\text{tr}(R_i^{-1} - R)/\text{tr}(R)$ of 22%.

As a matter of comparison, we performed approximation of $R_i^{-1}$ in the reduced 5-
dimensional space proposed by [Ruggiero et al(2016)], who assigned a fixed spatial variability to 
the diagonals of the control matrices and minimized (2) by varying five diagonal scaling factors 
$x$ (see Appendix). In this procedure, we employed the technique of Sections 2.1-2.2, which can 
be viewed as a generalization of the computational approach of [Ruggiero et al(2016)] who 
used five-fold expansion of the data space by computing the derivatives of the error fields in 
SWOT simulator output instead of explicit computation of the Hessian (eq. 19 in Appendix) and 
its projection on the reduced control space (eq. 12).

Figure 2 shows the results of the reduced space optimization. As it is seen, the reduced 
method provides a poorer fit to the SWOT spectrum being tested (cf. Fig. 1c and 2c) and a 
larger error in approximating the columns of the SWOT covariance matrix (cf. Fig. 1b,2b and Fig. 
1a). This should be attributed to lesser flexibility of the reduced procedure, as the number of 
adjusted parameters is approximately $2,791/5 \approx 560$ times smaller compared to the case of full 
optimization involving solution of eq. (6-8).

4 Summary and discussion

In recent years, there has been an increased interest in inverse OEC modeling due to the 
high-resolution swath altimetry missions planned in the near future [Durand et al(2010)], [Ito et 
al(2014)], [Ichikawa l(2014)]. Although this new type of observational platform is characterized 
by improved accuracy (1-2 cm) and higher spatial resolution (1-2 km), the respective OECs are 
expected to be highly correlated in space. This property presents a computational challenge for 
many operational DA algorithms that are based on the diagonal OECs.

This note proposes a methodology of approximating the inverse OECs by a polynomial in 
differential operators acting on sparse control matrices whose non-zero elements are adjusted 
to minimize the Frobenius norm of the approximation error. Explicit relationships for the cost 
function gradient and the Hessian matrix of the optimization problem have been obtained for 
control matrices with fixed sparsity patterns. A method of reduction of the optimization 
problem has been demonstrated for the case of the degenerate Hessian. The proposed 
approach could be used in realistic data assimilation systems by replacing the code normalizing
model-data misfits by observation error variances with the code multiplying the misfits by a sparse matrix retrieved from an estimate of the respective error covariance.

Further developments of the approach can be foreseen in several directions. First, the method does not maintain the positive semi-definite (psd) property of the approximation matrix in the process of optimization. The psd constraint can be imposed in many ways if the method is restricted to the diagonal control matrices. A straightforward way is to constrain all the components of the control vector to be positive in the process of optimization. This approach (combined with the projection technique) was used by [Ruggiero et al(2016)] to ensure the psd property. A somewhat more sophisticated methodology is based on factorizing $R_i = LL^T$ and representing $L$ as a composite of (sparse) control matrices: $L = [L_0 \quad \nabla^T \nabla_1 \ldots]^T$ with diagonal controls $L_i, i = 0, 1, \ldots$. This option, however, destroys the attractive quadratic property of the optimization problem. More general approaches going beyond the fixed sparsity patterns of the controls can also be explored (e.g., [Hsieh et al(2014)]). Our experience with the presented version of the method have shown, however, that optimal $R_i$ was very close to psd with only a few negative eigenvalues that contributed less than 0.1% to the trace of $R_i$.

The cost function could also be defined by $J = ||1 - R_i R_i||_F^2$ to directly retrieve a sparse approximation to $R^{-1/2}$ that may be more useful in the DA applications. Our numerical experiments with this formulation have shown that one has to pay more attention to initialization of the control variables, because starting the quasi-newtonian descent from $x = 0$ proved inefficient for several simulated classes of OECs. In contrast, the considered quadratic/diagonal formulation (1) performed well and never encountered convergence/conditioning problems for the same classes of OECs.

Elaboration of an efficient reduction scheme also remains an important issue. [Ruggiero et al(2016)] have shown that certain OECs can be efficiently approximated with just a few parameters, if an appropriate projection method is elected. In particular, useful information on the structure of $\Pi$ could be retrieved from the structure of the diagonal cells of the Hessian matrices. An alternative way of regularizing the problem is to augment (2) with the terms which penalize high-frequency variations of the control variables. However, the respective low-pass filter should be designed with caution, as the high-frequency variations of the inverse matrix elements (partly simulated by the differential operators) are a key component of the optimized matrix.

We believe that further studies of the matrix approximation methodologies in application to the class of psd matrices with slowly varying spatial structure has good prospects in the future development of DA techniques in geophysical applications and may benefit more general areas such as the search for efficient preconditioners.

This study was supported by Office of Naval Research projects (Program Elements 0646352N, 0602435N). J. D’Addezio was supported by the Naval Research Laboratory Cooperative Agreement BAA-N00173-03-13-01 awarded to the University of Southern Mississippi. Helpful discussions with Prof. C. Beattie are acknowledged.

5 Appendix

[Ruggiero et al(2016)] proposed a simplified method of estimating the inverse of $R$...
through the adjustment of only five free parameters. In the notation of section 2, their scheme involves a combination of the inverse covariance model containing four diagonal matrices

$$R_i = A + \nabla^T B \nabla + \Delta_x C \Delta_x + \Delta_y D \Delta_y$$  \hspace{1cm} (14)

with a projection scheme, which assigns a certain spatial structure to the matrices $A, B, C$ and $D$. In (14) the operators $\Delta_x$ and $\Delta_y$ stand for the cross- and along-track constituents of the Laplacian: $\Delta = \Delta_x + \Delta_y$.

Specifically, the projection adopted in R16 is defined by

$$A = \alpha_0 (R \circ I)^{-1}$$  \hspace{1cm} (15)  

$$B = \alpha_{1c} (\partial_x R \circ \partial_x^T \circ I)^{-1}$$  \hspace{1cm} (16)  

$$C = \alpha_{2c} (\Delta_x R \Delta_x \circ I)^{-1}$$  \hspace{1cm} (17)  

$$D = \alpha_{2a} (\Delta_y R \Delta_y \circ I)^{-1}$$  \hspace{1cm} (18)

with the control vector $x = [\alpha_0 \alpha_{1c} \alpha_{1a} \alpha_{2c} \alpha_{2a}]^T$ and the projection operator $\Pi$ represented by $5N - n_x - n_y \times 5$ block-diagonal matrix containing the inverse matrices in right-hand sides of eq. (15-18) on the diagonal.

The system of equations $Hx = r$ is

$$H = \begin{bmatrix}
  R^2 \circ I & R^2 \nabla \circ \nabla^T & R^2 \Delta_x \circ \Delta_x & R^2 \Delta_y \circ \Delta_y \\
  \nabla R^2 \circ \nabla & \nabla R^2 \circ \nabla^T & \nabla R^2 \Delta_x \circ \Delta_x & \nabla R^2 \Delta_y \circ \Delta_y \\
  \Delta_x R^2 \circ \Delta_x & \Delta_x R^2 \nabla \circ \Delta_x \nabla^T & \Delta_x R^2 \Delta_x \circ \Delta_x^2 & \Delta_x R^2 \Delta_y \circ \Delta_{xy} \\
  \Delta_y R^2 \circ \Delta_y & \Delta_y R^2 \nabla \circ \Delta_y \nabla^T & \Delta_y R^2 \Delta_x \circ \Delta_{xy} & \Delta_y R^2 \Delta_y \circ \Delta_y^2
\end{bmatrix}$$  \hspace{1cm} (19)

$$r = \begin{bmatrix}
  R \\
  \nabla R \nabla \circ I \\
  \Delta_x R \Delta_x \circ I \\
  \Delta_y R \Delta_y \circ I
\end{bmatrix}$$  \hspace{1cm} (20)

It is noteworthy, that the ansatz (14) produced a degenerate Hessian (19) which did not allow us to compare the results of full optimizations with the inverse OEC models (1) and (14). Reduction of the control space regularized the problem, but resulted in a relatively poor fit to the spectrum of the tested covariance (cf. Fig. 1c and Fig. 2c).

As a final note, we present the general expression for the Hessian matrix associated with the column-vectorized form of (6-8). Defining the sparsity pattern $S$ of a matrix $A$ by replacing non-zero elements of $A$ with ones, and adopting the notation $S_{AB} = \text{vec}(S \ A) \otimes \text{vec}(S \ B^T)$ for mutual Kronecker products of the sparsity patterns, the Hessian is given by

$$H = \begin{bmatrix}
  (R^2 \otimes I) \circ S_{AA} & (R^2 \nabla \otimes \nabla^T) \circ S_{AB} & (R^2 \Delta \otimes \Delta) \circ S_{AC} \\
  (\nabla R^2 \otimes \nabla) \circ S_{BA} & (\nabla R^2 \nabla \otimes \nabla^T) \circ S_{BB} & (\nabla R^2 \Delta \otimes \Delta) \circ S_{BC} \\
  (\Delta R^2 \otimes \Delta) \circ S_{CA} & (\Delta R^2 \nabla \otimes \Delta^T) \circ S_{CB} & (\Delta R^2 \Delta \otimes \Delta^2) \circ S_{CC}
\end{bmatrix}$$  \hspace{1cm} (21)

References

This article is protected by copyright. All rights reserved.


[Ichikawa l(2014)] Ichikawa, K., 2014: Satellite altimeters in the early 21st Century,


On the approximation of the inverse error covariances of high resolution satellite altimetry data

M. Yaremchuk$^1$, J.M. D’Addezio$^2$, G. Pantelev$^1$ and G. Jacobs$^1$

$^1$Naval Research Laboratory, Stennis Space Center, USA

$^2$Department of Marine Science, University of Southern Mississippi, USA

*Correspondence to: Naval Research Laboratory,1009 Balch Blvd., Stennis Space Center, 39522, MS, USA. E-mail: max.yaremchuk@nrlssc.navy.mil

High-resolution (swath) altimeter missions scheduled to monitor the ocean surface in the near future have observation error covariances (OECs) with slowly decaying off-diagonal elements. This property presents a challenge for the majority of the data assimilation (DA) algorithms which were designed under the assumption of the diagonal OECs being easily inverted. In this note, we present a method of approximating the inverse of a dense OEC by a sparse matrix represented by the polynomial of spatially inhomogeneous differential operators, whose coefficients are optimized to fit the target OEC by minimizing a quadratic cost function. Explicit expressions for the cost function gradient and the Hessian are derived. The method is tested with an OEC model generated by the SWOT simulator.

Key Words: covariance modelling; data assimilation; observational data analysis; wide swath altimetry

Received …

1. Introduction

Over the last several decades, representation of the background error covariances by the polynomials of the diffusion operator has been extensively studied in both meteorological and oceanographic DA applications (e.g., Derber and Rosati (1988), Weaver et al (2003), Xu (2005), Yaremchuk and Smith (2011), Yaremchuk et al (2013)). Among the advantages are the computational efficiency of the approach and its ability to preserve the positive semi-definite (psd) property of the resulting background error covariance (BEC) estimates. The method proves to be especially useful in heuristic modeling of the climatological (static) BECs which supplement the ensemble-based estimates of the background errors.
In contrast to the BECs, observational error covariances (OECs) are conventionally represented by diagonal matrices under the implicit assumption that observation errors are weakly correlated at spatial scales exceeding the grid step of the numerical models. This assumption, being realistic for most current observational platforms, provides an additional convenience of inexpensive computation of the inverse OECs and of their square roots currently employed by the majority of DA algorithms (e.g., Cummings (2005), Hunt et al (2007), Fairbairn et al (2014)).

Recent developments in high-resolution (“swath”) altimetry (Durand et al (2010), Ito et al (2014), Ubelmann et al (2015), Gautier et al (2016)) introduce challenges for data assimilation. Novel issues emerge due to both the data density that may approach model grid scales in the horizontal, and because observational errors at such high resolutions appear to be highly correlated in space Ruggiero et al (2016) (hereinafter R16) due to the design of the satellite and sensor. At the same time, recent studies (e.g., Stewart et al (2013), Miyoshi et al (2013), Waller et al (2014)) demonstrate substantial benefits of accounting for spatial correlations of the observation errors even in low-dimensional DA systems. As a consequence, these newly arriving data require special treatment in order to maintain skill and retain the computational efficiency of the DA schemes.

In most of the DA algorithms, the difference between the model sea surface height and the data has to be multiplied by either the inverse of the observation error covariance $R$ or by its inverse square root $R^{-1/2}$, so a computationally efficient representation $R_i$ of these matrices (e.g., by a sparse matrix) is highly desirable. This note contributes to the methodology of approximating $R^{-1}$ (or its square root) using differential operators. Specifically, with the forthcoming Surface Water and Ocean Topography (SWOT) altimeter mission in mind, we postulate that an estimate of $R$ is available on a regular 2-dimensional set of $N = n_x \times n_y$ observation grid points from the simulating software of Ubelmann et al (2017).

## 2. Approximating the inverse covariance

### 2.1. The matrix parameterization model

In what follows, we present a methodology of parameterizing $R_i$ by a linear combination of sparse matrices with matrix-valued coefficients represented by discretized differential operators. To simplify the notation, the method is illustrated by a particular example relevant to the SWOT application.

Specifically, we consider $R_i$ of the form

\[ R_i = A + \nabla^T B \nabla + \Delta C \Delta \quad (1) \]

where $\nabla$ is the $2N - n_x - n_y \times N$ matrix, representing the first-order approximation of the gradient operator on a 2d grid, $\Delta = \nabla^T \nabla$ is the Laplacian, and $A, B, C$ are sparse control matrices. Their $M$ non-zero elements populating the vector $x \in \mathbb{R}^M$ have to be optimized by minimization of the quadratic cost function, measuring the Frobenius norm $|| \cdot ||_F$ of the respective residual:

\[ J = ||R_iR - I||^2_F = \text{tr} \left[ PP^T \right] \rightarrow \min_x \quad (2) \]

Here $I$ is the identity matrix and notation $P = R_iR - I$ is introduced. The first variation of (2) has the form

\[ \delta J = 2 \text{tr} \left[ \delta PP^T \right] = 2 \text{tr} \left[ \delta R_iRP^T \right] \quad (3) \]
Figure 1. (a): map of the sum of two columns of the SWOT covariance matrix \( R \) (column positions shown by squares) and its approximation (b) by \( R^{-1} \). Panel (c) shows the spectra of the SWOT covariance (thin black line) and its approximation. Covariance values are divided by 100 cm\(^2\).

Taking into account that \( \delta R_i = \delta A + \nabla^T \delta B \nabla + \Delta \delta C \Delta \), and introducing the notation \( Q = 2PR \), transforms \( \delta J \) to

\[
\delta J = \text{tr} \left[ \delta A Q^T + \delta C \Delta Q^T \Delta \right] + \text{tr} \left[ \delta B \nabla Q^T \nabla \nabla^T \Delta \right].
\]  

(4)

so that the respective expressions for the gradient are:

\[
\frac{\delta J}{\delta A} = Q; \quad \frac{\delta J}{\delta B} = \nabla Q \nabla^T; \quad \frac{\delta J}{\delta C} = \Delta Q \Delta.
\]  

(5)

and the system of equations \( \delta J/\delta x = 0 \) defining the minimum of \( J \) takes the form

\[
R, R^2 = R
\]

(6)

\[
\nabla R, R^2 \nabla^T = \nabla R \nabla^T
\]

(7)

\[
\Delta R, R^2 \Delta = \Delta R \Delta
\]

(8)

Equations (6-8) can be rewritten explicitly in terms of the Hessian matrix \( H \in \mathbb{R}^{M \times M} \) and the rhs vector \( r \in \mathbb{R}^M \)

\[
Hx = r
\]

(9)

where \( r \) contains non-zero elements in the right-hand sides of (6-8) listed columnwise as in Eq. 20 of the Appendix.

In this note we consider the simplest sparsity pattern for \( A, B \) and \( C \), assuming that the matrices are diagonal so that their sparsity patterns are the identity matrices of the respective size. In this case, the general expression for the Hessian (see eq. (21) in the Appendix) simplifies to

\[
H = \begin{bmatrix}
R^2 \circ I & R^2 \nabla^T \circ \nabla & R^2 \Delta \circ \Delta \\
\nabla R^2 \circ \nabla & \nabla R^2 \nabla^T \circ \nabla \nabla^T & \nabla R^2 \Delta \circ \Delta \\
\Delta R^2 \circ \Delta & \Delta R^2 \nabla^T \circ \Delta \nabla^T & \Delta R^2 \Delta \circ \Delta^2
\end{bmatrix},
\]

(10)
where \( \odot \) denotes Hadamard (element-wise) matrix product. The latter relationship could be useful for constructing block-diagonal preconditioners for the iterative solvers of (9), or for direct solution of (9) on the moderate-size \((N < 10^4)\) grids.

### 2.2. Model reduction

The system of equations (9) may not be well conditioned, so its solution should be sought using a certain parameterization of the original control variables \( \mathbf{x} \). In the linear case, such parameterization can be expressed in terms of a projection operator \( \Pi \)

\[
\mathbf{x} = \Pi \tilde{\mathbf{x}}
\]  

(11)

where \( m \) columns of \( \Pi \) contain the "structure functions", parameterizing spatial variability of \( \mathbf{x} \), and \( \tilde{\mathbf{x}} \) stands for the vector of the reduced control variables. For instance, to enforce smoothness of the diagonal elements of the control matrices, these functions can be represented by \( m \) smoothest harmonics, the first one being independent on horizontal coordinates. Although in the present study we did not employ any model reduction \((\Pi = I)\), more sophisticated projections \((e.g.\text{Brankart et al., 2009}, R16)\) can be employed (see Appendix). Using non-trivial structure of \( \Pi \) requires, however, prior information on the spatial variability of the control fields in order to maintain a reasonable balance between the accuracy of the approximation of the target matrix and numerical efficiency.

The reduced normal system \( \tilde{H} \tilde{\mathbf{x}} = \tilde{\mathbf{r}} \) is characterized by the reduced Hessian \( \tilde{H} \in \mathbb{R}^{m \times m} \) and the reduced rhs \( \tilde{\mathbf{r}} \in \mathbb{R}^{m} \):

\[
\tilde{H} = \Pi^T H \Pi; \quad \tilde{\mathbf{r}} = \Pi^T \mathbf{r}
\]  

(12)

Since \( \delta J / \delta \mathbf{x} = H \mathbf{x} - \mathbf{r} \), the cost function gradient (4) can be conveniently estimated by applying the projection operator to (5):

\[
\frac{\delta J}{\delta \tilde{\mathbf{x}}} = \Pi^T \frac{\delta J}{\delta \mathbf{x}}
\]  

(13)

Note that since the relationships (5) are valid for arbitrary control matrices, equations (11-13) could be used in optimization algorithms employing gradient information under arbitrary linear constraints on spatial variability of the control matrix elements. In particular, the approach could be useful for maintaining the psd property of \( \mathbf{R} \) in the process of minimization.

### 3. Numerical testing

#### 3.1. SWOT covariance model

The ansatz (1) for the approximation of the inverse observation error covariance was tested with the target OEC generated by the Jet Propulsion Laboratory’s (JPL) SWOT simulator of Ubelmann et al (2017). The simulator generates realizations of SSH observation error fields based on the latest estimate of the SWOT error budget of Esteban-Fernandez (2013). The error field contains six constituents: Ka-band radar interferometer noise, wet tropospheric error, and errors associated with uncertainties in the estimation of roll, phase, baseline, and timing of the SWOT observational platform. With a reasonable degree of accuracy, the first two error fields can be considered to be uncorrelated. The remaining four error sources are of particular interest because they are highly correlated over large spatial scales: As shown by Ruggiero et al (2016), these errors are characterized by typical decorrelation scales of several hundred kilometers along the swath and approximately a hundred across, with the marginal pointwise pdfs being very close to Gaussian.
3.2. Results

In generating the target OEC matrix, we used the SWOT simulator version 2.0.0 with the following parameters: the default cut-off wavelength of 40,000 km and a 2 beam wet tropospheric error correction. Additionally, anticipating large decorrelation scales (compared to the projected SWOT resolution of 1-2 km) and the absence of smaller-scale spatial variability in the matrix columns, we elected 40 and 10 km sampling in the along- and across track directions respectively. This selection also decreased the influence of uncorrelated Ka-band noise on the OEC structure. 5,000 random realizations of all error sources summed together were generated by having the simulator recursively sample the same 21 day repeat orbit over a subdomain of the Western Pacific (116°E-133°E, 18°N-34°N) with a total sampled track length of 2000 km and width of 140 km. Excluding the grid points in the 20 km wide nadir gap, the sampled OEC field dimensions were \( n_x = 14 \), \( n_y = 51 \), \( N = n_x n_y = 714 \) with the total number of adjusted degrees of freedom \( 4N - n_x - n_y = 2,791 \) and the number of the independent elements in the target covariance matrix \( N(N + 1)/2 = 255,255 \). The resulting error fields were characterized by approximately Gaussian pointwise pdfs with the average magnitude of the means \( \sim 10^{-4} \) m and the standard deviations varying between 0.03 m near the nadir gap and 0.31 m at the swath periphery.

Figure 1a shows half the sum of the OEC fields corresponding to the pair of SSH observations located on both edges of the nadir gap in the center of the sampled track. Similar to the results of R16, covariance structures at intermediate scales are barely visible. However, there is a strong anisotropy of the covariance with the typical OEC spatial scales in the along- and across-track directions differing by an order in magnitude (600 km and 60 km respectively).

Figure 1a and b demonstrate the result of approximating \( \mathbf{R} \) by the inverse covariance model (1). Due to the limited number of ensemble members a slight assymetry (of the order of 1%) has been observed in the structure of the mirror rows of \( \mathbf{R} \). Figure 1b shows that this assymetry is considerably enhanced in the approximating matrix \( \mathbf{R}^{-1} \) (cf. Fig. 1a, b). The effect is caused by the coarse resolution of the nadir gap which is only 2 grid steps wide, and associated errors in the finite-difference approximation of \( \mathbf{R}^{-1} \) by the ansatz (1).

Due to the modest dimension of the control space \( (N = 2,791) \) and low condition number \( \text{cond}(\mathbf{H})=2\cdot10^{14} \) of the Hessian matrix (10), the optimization took a few seconds on a single CPU of a PC using the MatLab sparse system solver. As it is seen, the algorithm provides a reasonably accurate fit to the leading eigenmodes of \( \mathbf{R} \) (Figure 1c) with the relative error \( \text{tr}(\mathbf{R}^{-1} - \mathbf{R})/\text{tr}(\mathbf{R}) \) of 22%.

As a matter of comparison, we performed approximation of \( \mathbf{R}^{-1} \) in the reduced 5-dimensional space proposed by Ruggiero et al. (2016), who assigned a fixed spatial variability to the diagonals of the control matrices and minimized (2) by varying five diagonal scaling factors \( \tilde{x} \) (see Appendix). In this procedure, we employed the technique of Sections 2.1-2.2, which can be viewed as a...
generalization of the computational approach of Ruggiero et al (2016) who used five-fold expansion of the data space by computing the derivatives of the error fields in SWOT simulator output instead of explicit computation of the Hessian (eq. 19 in Appendix) and its projection on the reduced control space (eq. 12).

Figure 2 shows the results of the reduced space optimization. As it is seen, the reduced method provides a poorer fit to the SWOT spectrum being tested (cf. Fig. 1c and 2c) and a larger error in approximating the columns of the SWOT covariance matrix (cf. Fig. 1b,2b and Fig. 1a). This should be attributed to lesser flexibility of the reduced procedure, as the number of adjusted parameters is approximately 2,791/5≈560 times smaller compared to the case of full optimization involving solution of eq. (6-8).

4. Summary and discussion

In recent years, there has been an increased interest in inverse OEC modeling due to the high-resolution swath altimetry missions planned in the near future Durand et al (2010), Ito et al (2014), Ichikawa l (2014). Although this new type of observational platform is characterized by improved accuracy (1-2 cm) and higher spatial resolution (1-2 km), the respective OECs are expected to be highly correlated in space. This property presents a computational challenge for many operational DA algorithms that are based on the diagonal OECs.

This note proposes a methodology of approximating the inverse OECs by a polynomial in differential operators acting on sparse control matrices whose non-zero elements are adjusted to minimize the Frobenius norm of the approximation error. Explicit relationships for the cost function gradient and the Hessian matrix of the optimization problem have been obtained for control matrices with fixed sparsity patterns. A method of reduction of the optimization problem has been demonstrated for the case of the degenerate Hessian. The proposed approach could be used in realistic data assimilation systems by replacing the code normalizing model-data misfits by observation error variances with the code multiplying the misfits by a sparse matrix retrieved from an estimate of the respective error covariance.

Further developments of the approach can be foreseen in several directions. First, the method does not maintain the positive semi-definite (psd) property of the approximation matrix in the process of optimization. The psd constraint can be imposed in many ways if the method is restricted to the diagonal control matrices. A straightforward way is to constrain all the components of the control vector to be positive in the process of optimization. This approach (combined with the projection technique) was used by Ruggiero et al (2016) to ensure the psd property. A somewhat more sophisticated methodology is based on factorizing \( R_i = L i L^T \) and representing \( L \) as a composite of (sparse) control matrices: \( L = [L_0 \quad \nabla^T L_1 \ ...]^T \) with diagonal controls \( L_i, i = 0, 1, ... \). This option, however, destroys the attractive quadratic property of the optimization problem. More general approaches going beyond the fixed sparsity patterns of the controls can also be explored (e.g., Hsieh et al (2014)). Our experience with the presented version of the method have shown, however, that optimal \( R_i \) was very close to psd with only a few negative eigenvalues that contributed less than 0.1% to the trace of \( R_i \).

The cost function could also be defined by \( J = || I - R_i L_i R_i ||^2_F \) to directly retrieve a sparse approximation to \( R^{-1/2} \) that may be more useful in the DA applications. Our numerical experiments with this formulation have shown that one has to pay more attention to initialization of the control variables, because starting the quasi-newtonian descent from \( x = 0 \) proved inefficient for several simulated classes of OECs. In contrast, the considered quadratic/diagonal formulation (1) performed well and never encountered convergence/conditioning problems for the same classes of OECs.

Elaboration of an efficient reduction scheme also remains an important issue. Ruggiero et al (2016) have shown that certain OECs can be efficiently approximated with just a few parameters, if an appropriate projection method is elected. In particular, useful information on the structure of \( \Pi \) could be retrieved from the structure of the diagonal cells of the Hessian matrices. An alternative way of regularizing the problem is to augment (2) with the terms which penalize high-frequency variations of the control variables. However, the respective
low-pass filter should be designed with caution, as the high-frequency variations of the inverse matrix elements (partly simulated by the differential operators) are a key component of the optimized matrix.

We believe that further studies of the matrix approximation methodologies in application to the class of psd matrices with slowly varying spatial structure has good prospects in the future development of DA techniques in geophysical applications and may benefit more general areas such as the search for efficient preconditioners.

Acknowledgements

This study was supported by Office of Naval Research projects (Program Elements 0646352N, 0602435N). J. D’Addezio was supported by the Naval Research Laboratory Cooperative Agreement BAA-N00173-03-13-01 awarded to the University of Southern Mississippi. Helpful discussions with Prof. C. Beattie are acknowledged.

5. Appendix

Ruggiero et al (2016) proposed a simplified method of estimating the inverse of $R$ through the adjustment of only five free parameters. In the notation of section 2, their scheme involves a combination of the inverse covariance model containing four diagonal matrices

$$R_i = A + \nabla^T B \nabla + \Delta_x C \Delta_x + \Delta_y D \Delta_y$$

(14)

with a projection scheme, which assigns a certain spatial structure to the matrices $A, B, C$ and $D$. In (14) the operators $\Delta_x$ and $\Delta_y$ stand for the cross- and along-track constituents of the Laplacian: $\Delta = \Delta_x + \Delta_y$.

Specifically, the projection adopted in R16 is defined by

$$A = \alpha_0 (R \circ I)^{-1}$$

$$B = \alpha_{1c} (\partial_x R \circ I)^{-1}$$

$$C = \alpha_{2c} (\Delta_x R \circ I)^{-1}$$

$$D = \alpha_{2a} (\Delta_y R \circ I)^{-1}$$

(15-18)

with the control vector $\mathbf{x} = [\alpha_0 \alpha_{1c} \alpha_{1a} \alpha_{2c} \alpha_{2a}]$ and the projection operator $\Pi$ represented by $5N - n_x - n_y \times 5$ block-diagonal matrix containing the inverse matrices in right-hand sides of eq. (15-18) on the diagonal.

The system of equations $Hx = r$ is

$$H = \begin{bmatrix}
R^2 \circ I & R^2 \nabla^T \circ \nabla^T & R^2 \Delta_x \circ \Delta_x & R^2 \Delta_y \circ \Delta_y \\
\nabla R^2 \circ \nabla & \nabla R^2 \nabla^T \circ \nabla^T & \nabla R^2 \Delta_x \circ \Delta_x & \nabla R^2 \Delta_y \circ \Delta_y \\
\Delta_x R^2 \circ \Delta_x & \Delta_x R^2 \nabla^T \circ \Delta_x \nabla^T & \Delta_x R^2 \Delta_x \circ \Delta_x^2 & \Delta_x R^2 \Delta_y \circ \Delta_y \\
\Delta_y R^2 \circ \Delta_y & \Delta_y R^2 \nabla^T \circ \Delta_y \nabla^T & \Delta_y R^2 \Delta_x \circ \Delta_y & \Delta_y R^2 \Delta_y \circ \Delta_y \\
\end{bmatrix}$$

(19)
It is noteworthy, that the ansatz (14) produced a degenerate Hessian (19) which did not allow us to compare the results of full optimizations with the inverse OEC models (1) and (14). Reduction of the control space regularized the problem, but resulted in a relatively poor fit to the spectrum of the tested covariance (cf. Fig. 1c and Fig. 2c).

As a final note, we present the general expression for the Hessian matrix associated with the column-vectorized form of (6-8). Defining the sparsity pattern $S_A$ of a matrix $A$ by replacing non-zero elements of $A$ with ones, and adopting the notation $S_{AB} = \text{vec}(S_A) \otimes \text{vec}(S_B^T)$ for mutual Kronecker products of the sparsity patterns, the Hessian is given by

$$H = \begin{bmatrix} R \otimes I & \nabla R \nabla^T \otimes I & \Delta_x R \Delta_x \otimes I \\ \nabla R^2 \otimes \nabla & \nabla R^2 \nabla^T \otimes \nabla^T & \nabla R^2 \Delta \otimes \Delta \\ \Delta_x R \Delta_x \otimes I & \Delta_x R \Delta_y \otimes I & \Delta_y R \Delta_y \otimes I \end{bmatrix}$$

References


Approximation of the inverse observational error covariances


