

# Some Properties of a Recursive Procedure for High Dimensional Parameter Estimation in Linear Model with Regularization

Hong Son Hoang, Remy Baraille

SHOM/HOM/REC, 42 av Gaspard Coriolis 31057 Toulouse, France

Email: [hhoang@shom.fr](mailto:hhoang@shom.fr)

Received 8 September 2014; revised 6 October 2014; accepted 2 November 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

Theoretical results related to properties of a regularized recursive algorithm for estimation of a high dimensional vector of parameters are presented and proved. The recursive character of the procedure is proposed to overcome the difficulties with high dimension of the observation vector in computation of a statistical regularized estimator. As to deal with high dimension of the vector of unknown parameters, the regularization is introduced by specifying a priori non-negative covariance structure for the vector of estimated parameters. Numerical example with Monte-Carlo simulation for a low-dimensional system as well as the state/parameter estimation in a very high dimensional oceanic model is presented to demonstrate the efficiency of the proposed approach.

## Keywords

Linear Model, Regularization, Recursive Algorithm, Non-Negative Covariance Structure, Eigenvalue Decomposition

---

## 1. Introduction

In [1] a statistical regularized estimator is proposed for an optimal linear estimator of unknown vector in a linear model with arbitrary non-negative covariance structure

$$z = Hx + v \quad (1)$$

where  $z$  is the  $p$ -vector observation,  $H$  is the  $(p \times n)$  observation matrix,  $x$  is the  $n$ -vector of unknown parameters to be estimated,  $v$  is the  $p$ -vector representing the observation error.

It is assumed

**How to cite this paper:** Hoang, H.S. and Baraille, R. (2014) Some Properties of a Recursive Procedure for High Dimensional Parameter Estimation in Linear Model with Regularization. *Open Journal of Statistics*, 4, 921-932.

<http://dx.doi.org/10.4236/ojs.2014.411087>

$$E(v) = 0, \quad E(vv^T) = V \quad (2)$$

$$E(x) = \bar{x}, \quad E(ee^T) = M, \quad E(ev^T) = N, \quad e := x - \bar{x} \quad (3)$$

where  $E(\cdot)$  is the mathematical expectation operator. Throughout this paper let  $p, n$  be any positive integers, the covariance matrix of the joint vector  $\tilde{v} = (e^T, v^T)^T$  may be singular (and hence the model (1)-(3) is called a linear model with arbitrary non-negative covariance structure).

No particular assumption is made regarding the probability density function of  $\tilde{v}$  and  $p, n$  are any positive integers.

In [1] the optimal linear estimator for the unknown vector  $x$  in the model (1)-(3) is defined as

$$\hat{x} = Gz, \quad G = H^+ [I - V_1 \bar{V}_1]^{+T}, \quad \bar{V}_1 = V_1^T (I - H^+ H) \quad (4)$$

where  $V^T$  is the transpose of  $V$ ,  $H^+$  denotes the pseudo-inversion of  $H$ .

As in practice all the matrices  $H, M, N, R$  and the observation vector  $z$  are given only approximately, instead of data set  $D = (H, M, N, R, z)$  we are given their  $\delta$ -approximations

$$D^\delta = (H^\delta, M^\delta, N^\delta, R^\delta, z^\delta) \quad (5)$$

hence the resulting estimate  $\hat{x}^\delta = \hat{x}(D^\delta)$ .

As shown in [1], there are situations when the error  $e := \hat{x}^\delta - \hat{x}$  between the “true” estimate  $\hat{x}$  and its perturbed  $\hat{x}^\delta$  may be very large even for small data error  $\delta$ . The regularization procedure has been proposed in [1] to overcome this difficulty.

In this paper we are interested in obtaining a simple recursive algorithm for computation of  $\hat{x}$  subject to the situation when  $n$  or  $p$  or/and  $n, p$  may be very high.

This problem is very important for many practical applications. As example, consider data assimilation problems in meteorology and oceanography [2]. For typical data set in oceanography, at each assimilation instant we have the observation vector with  $p \approx 10^4 - 10^5$ ,  $n \approx 10^6 - 10^7$ . Under such conditions it is unimaginable to compute the estimate (4) using standard numerical methods. We will show that there exists a simple algorithm to overcome these difficulties by exploiting a recursive character of the algorithm with an appropriate regularization in the form of the priori covariance matrix  $M$  for the vector of unknown parameters  $x$ .

## 2. Simple Recursive Method for Estimating the Vector of Parameters

### 2.1. Problem Statement: Free-Noise Observations

First consider the model (1) and assume that  $v = 0$ . We have then the system of linear equations

$$z = Hx, \quad z \in R^p, \quad x \in R^n \quad (6)$$

and  $z \in R^p$  for the noise-free observations  $z$ .

Suppose that the system (6) is compatible, *i.e.* there exists  $x^0$  such that  $Hx^0 \equiv z$ . In what follows let  $z = (z_1, \dots, z_p)^T$ ,  $H = (h_1^T, \dots, h_p^T)^T$ , *i.e.*  $z_i$  is the  $i^{\text{th}}$  component of  $z$ ,  $h_i$  is the  $i^{\text{th}}$  row-vector of  $H$ . The problem is to obtain a simple recursive procedure to compute a solution of the system (6) when the dimension of  $z$  is very high.

### 2.2. Iterative Procedure

To find a solution to Equation (6), let us introduce the following system of recursive equations

$$x_{i+1} = x_i + K_{i+1} [z_{i+1} - h_{i+1} x_i], \quad x_0 = \bar{x}, \quad i = 0, \dots, p \quad (7a)$$

$$K_{i+1} = M_{i+1} h_{i+1}^T [h_{i+1} M_{i+1} h_{i+1}^T]^{-1} \quad (7b)$$

$$M_{i+1} = M_i - K_{i+1} h_{i+1} M_i, \quad i = 0, 1, \dots, M_0 \text{ is given} \quad (7c)$$

Mention that the system is compatible if  $z \in R[H]$  where  $R[H]$  is a linear space spanned by the columns

of  $H$ . Throughout of the paper, for definiteness, let  $h_i \neq 0$ ,  $i = 1, \dots, p$ .

*Theorem 1.* Suppose the system (6) is compatible. Then for any finite  $\bar{x}$  and symmetric positive definite (SPD) matrix  $M_0$  we have  $Hx_p = z$ .

In order to prove Theorem 1 we need

*Lemma 1.* The following equalities hold

$$h_i M_j = 0, \quad i = 1, \dots, j \quad (8)$$

*Proof.* By induction. We have for  $j = 1$ ,

$$h_1 M_1 = h_1 M_0 - h_1 K_1 h_1 M_0 = 0$$

since  $h_1 K_1 = h_1 M_0 h_1^T / h_1 M_0 h_1^T = 1$ .

Let the statement be true for some  $1 < l < p$ . We show now that it is true also for  $l+1$ . As the statement is true for  $l$ , it implies  $h_i M_l = 0$ ,  $i = 1, \dots, l$ . We have to prove

$$h_i M_{l+1} = 0, \quad i = 1, \dots, l+1$$

Substituting  $M_{l+1} = M_l - K_{l+1} h_{l+1} M_l$  into  $h_i M_{l+1}$ , taking into account the form of  $K_{l+1}$  one sees that as  $h_i M_l = 0$ ,  $i = 1, \dots, l$  it implies  $h_i M_{l+1} = 0$ ,  $i = 1, \dots, l+1$  (End of Proof).

*Lemma 2.* The following equalities hold

$$h_i x_j = z_i, \quad i = 1, \dots, j \quad (9)$$

*Proof.* By induction. We have for  $j = 1$ ,  $h_1 x_1 = h_1 [x_0 + K_1 (z_1 - h_1 x_0)]$ . As  $h_1 K_1 = 1$ , it is evident that  $h_1 x_1 = z_1$ .

Let the statement be true for some  $1 < l < p$ . We show now that it is true also for  $l+1$ . As the statement is true for  $l$ , it implies  $h_i x_l = z_i$ ,  $i = 1, \dots, l$ . We have to prove  $h_i x_{l+1} = z_i$ ,  $i = 1, \dots, l+1$ . From the definition of  $x_{l+1}$ ,

$$h_i x_{l+1} = h_i [x_l + K_{l+1} (z_{l+1} - h_{l+1} x_l)] = z_i + h_i K_{l+1} (z_{l+1} - h_{l+1} x_l)$$

However from Lemma 1, as  $h_i M_{l+1} = 0$ ,  $h_i K_{l+1} = h_i M_{l+1} h_{l+1}^T [M_{l+1} h_{l+1} M_{l+1}]^+ = 0$  for all  $i \leq l+1$ . (End of Proof).

*Proof of Theorem 1.*

Theorem follows from Lemma 2 since under the conditions of Theorem, from Equation (9) for  $j = p$  it follows  $h_i x_p = z_i$ ,  $i = 1, \dots, p$  or  $Hx_p = z$ . (End of Proof).

*Corollary 1.* Suppose the rows of  $H_1^l := [h_1^T, \dots, h_l^T]^T$  are linearly independent. Then under conditions of Theorem 1,  $\text{rank}[M_l] = n - l$ .

*Proof.* By induction. The fact that for  $l = 1$ ,  $h_1 M_1 = 0$  implies at least the null subspace of  $M_1$  has one nonzero element hence  $\dim[R(M_1)] \leq n - 1$ . We show now that it is impossible that  $\dim[R(M_1)] < n - 1$ .

Suppose that  $\dim[R(M_1)] \leq n - n'$ ,  $n' = 2, 3, \dots$ . For simplicity, let  $n' = 2$ . It means that there exist  $n - 2$  linearly independent vectors  $a_1, \dots, a_{n-2}$  such that any element from the subspace  $R(M_1)$  can be represented on the basis of these  $n - 2$  elements. As to the matrix

$$\Delta M_1 := M_0 h_1^T h_1 M_0 [h_1 M_0 h_1^T]^+$$

its subspace  $R(\Delta M_1)$  has the dimension 1 hence any element from  $R(\Delta M_1)$  can be represented on the basis of some vector  $b_1$ . Thus any element from the subspace  $R(M_1')$  with  $M_1' := M_1 + \Delta M_1$  can be represented as a linear combination of  $n - 1$  elements  $a_1, \dots, a_{n-2}, b_1$ . On the other hand, as  $M_1 + \Delta M_1 = M_0$  is non-singular, any element of  $R(M_0)$  must be represented as a linear combination of  $n$  linearly independent vectors. It contradicts the fact that any element from  $R(M_1')$  could be written as a linear combination of  $n - 1$  linearly independent elements. We conclude that it is impossible  $\dim[R(M_1)] < n - 1$  hence  $\dim[R(M_1)] = n - 1$ .

The same argument is true for  $n' = 3, 4, \dots, n$ .

Suppose now Corollary is true for  $l \geq 1$  and we have to prove that it holds for  $l := l + 1$ . For

$$\Delta M_{l+1} := M_l h_{l+1}^T h_{l+1} M_l \left[ h_{l+1} M_l h_{l+1}^T \right]^+, \quad M_{l+1} = M_l - \Delta M_{l+1}$$

we have  $\dim[R(M_l)] = n-l$ ,  $\dim[\Delta M_{l+1}] = 1$ . From Lemma 1 it follows  $h_i M_{l+1} = 0$ ,  $i = 1, \dots, l+1$  hence the null subspace  $N[M_{l+1}]$  has the dimension  $l+1$  (as all the rows of  $H$  are linearly independent). It follows that the dimension of the subspace  $R[M_{l+1}]$  is at least less or equal to  $n-l-1$ .

By the same way as proved for  $l=1$  one can show that it is impossible  $\dim[R(M_{l+1})] < n-l-1$  hence  $\dim[R(M_{l+1})] = n-l-1$  (End of Proof).

*Comment 1.* By verifying the rank of  $M_l$ , Corollary 1 allows us to check if the computer code is correct. In particular if  $H$  is non-singular, at the end of the iterative procedure the matrix  $M_n$  should be zero. The recursive Equations (7a)-(7c) then yield the unique solution of the equation  $z = Hx$  after  $n$  iterations.

Using the result (8) in Lemma 1, it is easy to see that:

*Corollary 2.* Suppose  $h_{l+1}$  is linearly dependent on  $h_1, \dots, h_l$ . Then in (7),  $M_{l+1} = M_l$ .

*Corollary 3.* Suppose  $h_1, \dots, h_l$  are linearly independent,  $h_{l+1}$  is linearly dependent on  $h_1, \dots, h_l$ . Then under the conditions of Theorem 1,  $\text{rank}[R(M_{l+1})] = n-l$ .

Corollary 3 follows from the fact that when  $h_{l+1}$  is linearly dependent on  $h_1, \dots, h_l$  from Corollary 2,  $M_{l+1} = M_l$  and hence by Corollary 1,  $\text{rank}[R(M_{l+1})] = n-l$ .

*Comment 2.* Equation (7c) for  $M_l$  is similar to that given in (3.14.1) in [3] since from Corollary 2 it follows automatically  $M_{l+1} = M_l$  if  $h_{l+1}$  is linearly dependent on  $h_1, \dots, h_l$ . Their difference is that in (7c) the initial  $M_0$  may be any SPD matrix whereas in (3.14.1) of [3] it is assumed  $M_0 = I$ .

In the next section we shall show that the fact  $M_0$  may be any SPD matrix is important to obtain the optimal in mean squared estimator for  $x$ .

### 3. Optimal Properties of the Solution of (7)

#### 3.1. Regularized Estimate

Theorem 1 says only that Equations (7a)-(7c) give a solution to (1). We are going now to study the question on whether a solution of Equations (7a)-(7c) is optimal, and if yes, in what sense? Return to Equations (1)-(3) and assume that  $V = 0$ ,  $N = 0$ . The optimal estimator (4) is then given in the form

$$\hat{x} = \bar{x} + MH^+ [HMH^+]^+ (z - H\bar{x}) \quad (10)$$

where  $H^+$  is the pseudo-inversion of  $H$ . Consider first the equation  $z_1 = h_1 x$ . In this case, (10) yields the optimal estimator (obtained on the basis of the first observation)

$$\hat{x}_1 = \bar{x} + Mh_1^T [h_1 Mh_1^T]^+ (z_1 - h_1 \bar{x}) \quad (11a)$$

and one can prove also that the mean square error for  $\hat{x}_1$  is equal to

$$M_1 = M_0 - Mh_1^T [h_1 Mh_1^T]^+ h_1 M \quad (11b)$$

If we apply Equation (3.14.1) in [3],  $M = I$  then instead of  $\hat{x}_1$  we have

$$\hat{x}'_1 = h_1^T / (h_1 h_1^T) z_1 = h_1^+ z_1, \quad M'_1 = I - h_1^T / (h_1 h_1^T) h_1 \quad (12)$$

For simplicity, let  $\bar{x} = 0$ . Comparing Equation (12) with Equation (11) shows that if  $\hat{x}'_1$  is the orthogonal projection of  $x$  onto the subspace spanned by  $h_1^T$ , the estimate  $\hat{x}_1$  belongs to the subspace spanned by  $M$ . Thus the algorithm (11) takes into account the fact that we known a priori  $x$  belongs to the space  $R[M]$ . This fact is very important when the number of observations  $p$  is much less than the number of the estimated parameters  $n$  as it happens in oceanic data assimilation: today usually  $p = 10^4 - 10^5$ ,  $n = 10^6 - 10^7$ .

In [4] a similar question has been studied which concerns the choice of adequate structure for the Error Covariance Matrix (ECM)  $M$ .

We prove now a more strong result saying that all the estimates  $x_i, i = 1, 2, \dots$  are projected onto the space  $R[M]$ .

*Theorem 2.* Consider the algorithm (7). Suppose  $\bar{x} \in R[\bar{M}]$ . Then all the estimates  $x_i, i = 1, 2, \dots$  belong to

the space spanned by the columns of  $\bar{M}$ , i.e.  $x_i \in R[\bar{M}]$ .

*Proof.* For  $i=1$  the statement is evident as shown above.

Suppose the statement is true for some  $i$ . We will show that it is true also for  $i := i+1$ .

Really as  $x_i \in R[\bar{M}]$ , it is sufficient to show that  $K_{i+1}\zeta_{i+1} \in R[\bar{M}]$ ,  $\zeta_{i+1} := z_{i+1} - h_{i+1}x_i$ . From  $x_i \in R[\bar{M}]$ ,  $x_i = x_{i-1} + K_i\zeta_i$ , as  $x_{i-1} \in R[\bar{M}]$  it follows that  $K_i\zeta_i \in R[\bar{M}]$ . But  $K_i\zeta_i = M_i h_i^T [h_i M_i h_i^T]^+ \zeta_i$  hence the columns of  $M_i$  must belong to  $R[\bar{M}]$ . Again from the equation for  $M_{i+1}$  in Equation (7) we have  $M_{i+1} \in R[\bar{M}]$ . It proves  $K_{i+1}\zeta_{i+1} \in R[\bar{M}]$  since  $M_i \in R[\bar{M}]$  (End of proof).

Theorem 2 says that by specifying  $M_0 \in R[\bar{M}]$  the algorithm (7) will produce the estimates belonging to the subspace  $R[\bar{M}]$ . Specification of the priori matrix  $\bar{M}$  plays the most important task if we want the algorithm (7) to produce the estimate with high quality.

*Comment 3*

1) If  $\bar{x}$  does not belong to  $R[\bar{M}]$ , by considering  $e := x - \bar{x}$  in place of  $x$  Theorem 3 remains valid for the algorithm (7) written for  $e$ . In this case at the first iteration, as  $e := x - \bar{x}$  represents the priori error for  $x$ , it is natural for the algorithm to seek the correction (i.e., the estimate for the error  $e := x - \bar{x}$ ) belonging to the space  $R[\bar{M}]$ . In what follows, unless otherwise stated, for simplicity we assume  $\bar{x} = 0$ .

2) Theorem 2 says that there is a possibility to regularize the estimate when the number of observations is less than the number of estimated parameters by choosing  $M_0 = \bar{M}$ . Thus the algorithm can be considered as that which finds the solution  $Hx = z$  under the constraint  $\bar{x} \in R[\bar{M}]$ . In the procedure in Albert (1972) putting  $M_0 = I$  means that there is no constraint on  $x$  hence the best way to do is to project  $x$  orthogonally onto subspace of  $R[H^T]$ .

### 3.2. Minimal Variance Estimate

Suppose  $x$  is a random variable having the mean  $\bar{x}$  and covariance matrix  $\bar{M}$ . We have then the following result.

*Theorem 3.* Suppose  $x$  is a random variable having the mean  $\bar{x}$  and the covariance matrix  $\bar{M}$ . Then  $x_i$  generated by the recursive Equation (7) is an unbiased and minimum variance estimate for  $x$  in the class of all unbiased estimates linearly dependent on  $\bar{x}$  and  $z_1, \dots, z_i$ .

*Proof.* Introduce for the system (6),

$$z_1^i := (z_1, \dots, z_i)^T, \quad H_1^i := (h_1^T, \dots, h_i^T)^T \quad (13)$$

and the class of all estimates  $x_i'$  linearly dependent on  $\bar{x}$  and  $z_1, \dots, z_i$ .

$$x_i'(A, B) = A\bar{x} + Bz_1^i$$

The condition for unbiasedness of  $x_i'(A, B)$  is

$$E[x_i'(A, B)] = E(A\bar{x} + Bz_1^i) = \bar{x}$$

or

$$(A + BH_1^i)\bar{x} = \bar{x}, \quad \forall \bar{x} \in R^n$$

from which follows  $A = I - BH_1^i$ . Substituting this relation into  $x_i'(A, B) = A\bar{x} + Bz_1^i$  leads to

$$x_i'(B) = \bar{x} + B[z_1^i - H_1^i\bar{x}] \quad (14)$$

It means that all the estimate in Equation (14) is unbiased.

Consider the minimization problem

$$J(B) = \text{trace}E(ee^T) \rightarrow \text{argmin}_B, e := x - x_i'(B)$$

We have

$$E(ee^T) = E\left[x - \bar{x} - B(z_1^i - H_1^i \bar{x})\right]\left[x - \bar{x} - B(z_1^i - H_1^i \bar{x})\right]^T = \bar{M} - \bar{M}H_1^{i,T}B^T - BH_1^i\bar{M} + BH_1^i\bar{M}H_1^{i,T}B^T$$

Taking the derivative of  $J(B)$  with respect to  $B$  implies the following equation for finding  $B$ ,

$$BH_1^i\bar{M}H_1^{i,T} = \bar{M}H_1^{i,T}$$

from which follows one of the solutions

$$B_0 = \bar{M}H_1^{i,T}\left[H_1^i\bar{M}H_1^{i,T}\right]^+$$

If now instead of Equation (13) we consider the system

$$z_1^i = H_1^i x + v_1^i, \quad E(v_1^i v_1^{i,T}) = \alpha I, \quad E\left[v_1^i (x - \bar{x})^T\right] = 0 \quad (15)$$

and repeat the same proof, one can show that the unbiased minimum variance estimate for  $x$  is given by

$$\hat{x}_i(\alpha) = \bar{x} + K(\alpha)\left[z_1^i - H_1^i \bar{x}\right] \quad (16a)$$

$$K(\alpha) = \bar{M}H_1^{i,T}\left[H_1^i\bar{M}H_1^{i,T} + \alpha I\right]^{-1} \quad (16b)$$

Using the properties of the pseudo-inverse (Theorem 3.8 [3]), one can prove that

$$\lim_{\alpha \rightarrow 0} K(\alpha) = B_0 \quad (17)$$

Thus for a very small  $\alpha > 0$  the estimate  $\hat{x}_i(\alpha)$  is unbiased minimum variance which can be made as close as possible to  $x_i^*(B_0)$ .

On the other hand, applying Lemma 1 in [5] for the case of uncorrelated sequence  $\{v_i\}$ , one can show that

$$\hat{x}_{i+1}(\alpha) = \hat{x}_i(\alpha) + k_{i+1}(\alpha)\left[z_{i+1} - h_{i+1}\hat{x}_i(\alpha)\right], \quad (18a)$$

$$k_{i+1}(\alpha) = M_{i+1}(\alpha)h_{i+1}^T / \left[h_{i+1}M_{i+1}(\alpha)h_{i+1}^T + \alpha\right], \quad (18b)$$

$$M_{i+1}(\alpha) = M_i(\alpha) - k_{i+1}(\alpha)h_{i+1}M_i(\alpha) \quad (18c)$$

$$M_0(\alpha) = \bar{M} \quad (18d)$$

Letting  $\alpha \rightarrow 0$  one comes to  $\hat{x}_i(\alpha) \rightarrow x_i$  in (7) (End of proof).

### 3.3. Noisy Observations

The algorithm (18a)-(18d) thus yields an unbiased minimal variance (UMV) estimates for  $x$  in the situation when  $\alpha$  represents the observation noise variance. We want to stress that these algorithms produce the UMV estimates only if  $\bar{M} = M$  where  $M$  is the true covariance of the error  $e := x - \bar{x}$  before arriving  $z_i$ ,  $i = 1, 2, \dots$ .

### 3.4. Very High Dimension of $x$ : Simplified Algorithms

In the field of data assimilation in meteorology and oceanography usually the state vector  $x$  is of very high dimension, the orders of  $10^6 - 10^7$  [2]. This happens because  $x$  is a collection of several variables defined in the three dimensional grid. If the algorithm (7) allows to overcome the difficulties with the high dimension of the observation vector  $z$  (each iteration involves one component of  $z$ ), due to high dimension of  $x$ , it is impossible to handle Equation (7c) to evaluate the matrices  $M_i$  (with the number of elements  $10^{12} - 10^{14}$ ). This section is devoted to the question on how one can overcome such difficulties.

Let us consider the eigen-decomposition for  $M$  [6]

$$M = UDU^T \quad (19)$$

In (19) the columns of  $U$  are the eigenvectors of  $M$  and  $D$  is diagonal with the elements  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  at the diagonal—the eigen-values of  $M$ . Let  $U = [U_1, U_2]$ ,  $D = \text{diag}[D_1, D_2]$ . If we put in the algorithms (7) or (18)  $\bar{M} = U_1 D_1 U_1^T$ , then the algorithm (7), for example, will yield the best estimate for  $x$

projected in the subspace  $R(U_1)$ . Let  $U_1 = U_1(m)$  has the dimension  $n \times m$ ,  $m \leq n$ . Let

$$\bar{M}(m) = U_1(m)D_1(m)U_1^T(m).$$

### 3.4.1. Main Theoretical Results

*Theorem 4*

Consider two algorithms of the type (7) subject to two matrices

$$\bar{M} = \bar{M}_i(m_i) = U_i(m_i)D_i(m_i)U_i^T(m_i), \quad i = 1, 2$$

where the columns of  $U_i(m_i)$  consist of  $m_i$  leading eigenvectors of  $M$ ,  $m_i \leq m_2$ . Then the following inequalities hold

$$E[\|e(m_2)\|^2] \leq E[\|e(m_1)\|^2], \quad m_1 \leq m_2, \quad e(m_i) = \hat{x}(m_i) - x, \quad i = 1, 2 \quad (20)$$

$\hat{x}(m_i)$  is the estimate produced by the algorithm (7) subject to  $\bar{M} := \bar{M}(m_i)$ , where the strict inequality takes place if  $\lambda_{m_2} > 0$ .

*Proof*

Write the representation of  $x$  in the terms of decomposition of  $M$  on the basis of its eigenvectors (for simplicity, let  $\bar{x} = 0$ ),

$$x = UD^{1/2}y = Ly, \quad L := UD^{1/2}, \quad M = UDU^T \quad (21)$$

where  $y$  is of zero mean and has the covariance matrix  $I$ .

Let  $y^l$  is a sample of  $y$ . Theorem 3 states that for all  $x^l := Ly^l$ , the algorithm (7) will yield the estimate with the minimal variance.

In what follows we introduce the notation:

$M = UDU^T$  —the true ECM of  $x$ ;

$M(m) = U(m)D(m)U^T(m)$  —a truncated covariance coming from  $M$ ;

$\bar{M}$  —the initialized ECM in the algorithm (7a)-(7c).

The samples  $x^l$  of  $x$  are coming from a variable having zero mean and covariance  $M$ .

$x^l(m)$  —sample coming from a variable having zero mean and covariance  $M(m)$ .

There are the following different cases

1)  $\bar{M} = M$ : By Theorem 3 the algorithm (7) will produce the estimates of minimal variance for all  $x^l$ ; This is true also if Equation (7) is applied to  $x^l(m)$ .

2)  $\bar{M} = M(m)$ ,  $m < n$ ,  $\lambda_{m+1} > 0$ :

a) For samples belonging to  $R[U(m)]$ : The estimates will be of minimal variance.

b) For samples belonging to  $R^n/R[U(m)]$  (*i.e.* belonging to  $R^n$  but not to  $R[U(m)]$ ): The estimates will not be of minimal variance.

Thus in the mean sense

$$E[\|e(n)\|^2] \leq E[\|e(m)\|^2], \quad m \leq n, \quad e(m) = \hat{x}(m) - x \quad (22)$$

3) Consider two initializations  $\bar{M} = M(m_1)$  and  $\bar{M} = M(m_2)$ ,  $m_1 \leq m_2 \leq n$ ,  $\lambda_{m_1+1} > 0$ . In the same way we have

a)  $\bar{M} = M(m_1)$ :

i)  $x^l \in R[M(m_1)]$ : the estimates are of minimal variance;

ii)  $x^l \in R^n/R[M(m_1)]$ : the estimates are not of minimal variance.

b)  $\bar{M} = M(m_2)$ : The algorithm (7) will produce the estimates

i) of minimal variance for  $x^l \in R[M(m_1)]$ ;

ii) of minimal variance for  $x^l \in R[M(m_2)]/R[M(m_1)]$ ;

iii) not of minimal variance for  $x^l \in R[M(n)]/R[M(m_2)]$ .

Thus in the mean sense

$$E\left[\|e(m_2)\|^2\right] \leq E\left[\|e(m_1)\|^2\right], \quad m_1 \leq m_2, \quad e(m_i) = \hat{x}(m_i) - x, \quad i = 1, 2 \quad (23)$$

### 3.4.2. Simplified Algorithm

*Theorem 5*

Consider the algorithm (7) subject to  $\bar{M} := M(m)$ ,  $M(m) = U(m)D(m)U^T(m)$ ,  $m \leq n$ .

Then this algorithm can be rewritten in the form

$$x_{i+1} = U(m)x_e(i+1), \quad x_e(i+1) \in R^m, \quad (24a)$$

$$x_e(i+1) = x_e(i) + K_e(i+1)[z_{i+1} - h_e(i+1)x_e(i)], \quad x_0 = \bar{x}, \quad i = 0, 1, \dots, p \quad (24b)$$

$$K_e(i+1) = M_e(i+1)h_e^T(i+1)[h_e(i+1)M_e(i+1)h_e^T(i+1)]^+, \quad (24c)$$

$$h_e(i+1) = h_{i+1}U(m), \quad M_e(0) = D(m) \quad (24d)$$

It is seen that in the algorithm (24a)-(24d), the estimate  $x_e(i) \in R^m$  belongs to the linear space of dimension  $m$ . In data assimilation, it happens that the dimension of  $x$  may be very high but there is only some leading directions (leading eigenvectors of  $M$  representing the directions of most rapid growth of estimation error) to be captured. Thus the algorithm (24a)-(24d) is quite adapted for solving such problems: first the initial covariance  $M$  is constructed from physical considerations or numerical model, and next to decompose it (numerically) to obtain an approximated decomposition

$$M(m) = U(m)D(m)U^T(m), \quad m \ll n$$

Mention that the version (24a)-(24d) is very closed to that studied in [7] for ensuring a stability of the filter.

## 4. Numerical Example

Consider the system (1) subject to the covariance  $M$ ,

$$M = \begin{bmatrix} 0.85 & 0.525 & 0 \\ 0.525 & 0.5625 & 0 \\ 0 & 0 & 0.64 \end{bmatrix} \quad (25)$$

Here we assume that the 1st and 3rd components of  $x$  is observed, *i.e.*

$$z = Hx, \quad H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (26)$$

Numerical computation of eigendecomposition of  $M = UDU^T$  yields

$$U = [u_1, u_2, u_3] = \begin{bmatrix} 0.795 & -6.E-16 & 0.606 \\ 0.6066 & 7.8E-16 & -0.795 \\ -1.4E-45 & 0.9999 & 4E-15 \end{bmatrix} \quad (27)$$

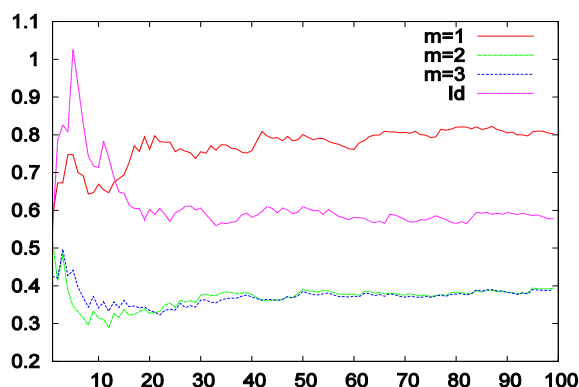
$$D = \text{diag}(1.25, 0.64, 0.162) \quad (28)$$

The algorithm (24a)-(24d) is applied subject to three covariance matrices  $\bar{M} = M(m) = U(m)D(m)U^T(m)$ ,  $m = 1, 2, 3$ . They are denoted as ALG( $m$ ).

In **Figure 1** we show the numerical results obtained from the Monte-Carlo simulation.

There are 100 samples simulating the true  $x$  which are generated by a random generator distributed according to the normal distribution  $N(\cdot/\bar{x}, M)$ . The curves in **Figure 1** represent rms of the estimation error  $e = x - \hat{x}$  obtained by different algorithms. Here the curves  $m = 1$ ,  $m = 2$ ,  $m = 3$  correspond to the three algorithms ALG(1), ALG(2), ALG(3).





**Figure 1.** Performance (rms) of the algorithm (24) subject to different projection subspaces.

The curve  $Id$  denotes the 4th algorithm ALG (4) which is run subject to  $\bar{M} = I$ —identity matrix. This is equivalent to the orthogonal projection (using the pseudo-inversion of  $H$ ) of  $z$  into the subspace  $R[H^T]$ .

As seen from **Figure 1**, the estimation error is highest in ALG (1). There is practically no difference between ALG (2) and ALG (3) which are capable of decreasing considerably the estimation error (50%) compared to ALG (1). As to the ALG (4), its performance is situated between ALG (1) and ALG (2). This experiment confirms the theoretical results and demonstrates that if we are given a good priori information on the estimated parameters, there exists a simple way to improve the quality of the estimate by appropriately introducing the priori information in the form of the regularization matrix  $M$ .

The results produced by ALG (1) and ALG (4) show also that when the priori information is insufficiently rich, the algorithm naturally produces the estimates of poor quality. In such situation, simply applying orthogonal projection can yield a better result. For the present example, the reason is that using the 2nd mode  $u_2$  allows to capture the important information contained in the second observation  $v_2$ . Ignoring it (as does ALG (1)) is equivalent to ignoring the second observation  $z_2$ . As to the third mode  $u_3$ , it has a weak impact on the estimation since the corresponding eigenvalue  $\lambda_3$  is small. That explains why ALG (2) and ALG (3) have produced almost the same results.

## 5. Experiment with Oceanic MICOM Model

### 5.1. MICOM Model

In this section we will show importance of the regularization factor in the form of the priori covariance  $M$  in the design of a filter for systems of very high dimension.

The numerical model used in this experiment is the MICOM (Miami Isopycnic Coordinate Ocean Model) which is exactly as that presented in [8]. We recall only that the model configuration is a domain situated in the North Atlantic from  $30^\circ$  N to  $60^\circ$  N and  $80^\circ$  W to  $44^\circ$  W. The grid spacing is about  $0.2^\circ$  in longitude and in latitude, requiring the horizontal mesh  $i = 1, \dots, 140$ ;  $j = 1, \dots, 180$ . The distance between two points  $\Delta x = x_{i+1} - x_i \approx 20$  km,  $\Delta y = y_{i+1} - y_i \approx 20$  km. The number of layers in the model  $n_v = 4$ . We note that the state of the model  $x := (h, u, v)$  where  $h = h(i, j, l)$  is the thickness of the  $l^{\text{th}}$  layer,  $u = u(i, j, l)$ ,  $v = v(i, j, l)$  are two velocity components. The “true” ocean is simulated by running the model from “climatology” during two years. Each ten days the sea-surface height (SSH) are stored at the grid points  $i_o = 10, 20, \dots, 140$ ;  $j_o = 10, 20, \dots, 180$  which are considered as observations in the assimilation experiment. The sequence of true states will be available and allows us to compute the estimation errors. Thus the observation operator  $H$  is constant at all assimilation instants.

The assimilation experiment consists of using the SSH to correct the model solution, which is initialized by some arbitrarily chosen state resulting from the control run.

### 5.2. Different Filters

The different filters are implemented to estimate the oceanic circulation. It is well known that determining the

filter gain is one of the most important tasks in the design of a filter. As for the considered problem it is impossible to apply the standard Kalman filter [9] since in the present experiment,  $n = O(10^6)$  and the number of elements in the ECM is of order  $O(n^2)$ . At each assimilation instant, the estimate for the system state in all filters is computed in the form (10) with the corresponding ECM  $M$ . As the number of observations  $p = 252$  is largely inferior to the dimension of the system state, the choice of  $M$  as a regularization factor has a great impact on the quality of the produced estimates. In this assimilation experiment the following filters will be employed. First the Prediction Error Filter (PEF) whose ECM is obtained on the basis of leading real Schur vectors [8]. Parallely two other filters, one is the Cooper-Haines filter (CHF) [10] and another is an EnOI (Ensemble based Optimal Interpolation) filter [11] will be used. Mention that the ECM in the CHF is obtained on the basis of the principle of a vertical rearrangement of water parcels (see also [8]). The method conserves the water masses and maintains geostrophy. The main difference between PEF and EnOI is lying in the way to generate the ensembles of Prediction Error (PE) samples. In the PEF, the ensemble of PE samples is generated using the sampling procedure described in [8] (and it will be denoted as En(PEF)). As for the EnOI, the ensemble of background errors samples (the term used in [8]) and will be denoted by En(EnOI) will be used. The elements of En(EnOI) are constructed according to the method in [11]. It consists of using 2-year mean of true states as the background field and the error samples are calculated as differences between individual 10-day true states during this period and the background.

According to Corollary 4.1 in [4], using the hypothesis on separable vertical-horizontal structure for the ECM, we represent  $M = M_v \otimes M_h$  where  $M_v$ ,  $M_h$  are the ECM of vertical and horizontal variables respectively. In the case of sea-surface height observations, from the representation  $M = M_v \otimes M_h$ , the gain filter can be represented in the form

$$K = K_v \otimes K_h, \quad K_v = [k_1, k_2, k_3, k_4]^T$$

where  $\otimes$  denotes the Kronecker product. The gain  $K_v$  allows the correction available at the surface to propagate into all vertical subsurface layers. As to  $K_h$ , it represents an operator of (horizontal) optimal interpolation which interpolates the observations over all horizontal grid points at the surface. Mention that the elements of  $M_v$  and the correlation length parameter in  $M_h$  are estimated by minimizing the mean distance between the data matrix  $M_d$  and  $M$  using a simultaneous perturbation stochastic approximation algorithm [12]. The data matrix  $M_d$  is obtained from samples of the leading real Schur vectors as described in [8].

**Figure 2** shows the estimated vertical coefficients  $k_l$ ,  $l = 1, 2, 3, 4$  obtained on the basis of the ECM spanned by the elements of En(PEF). It is seen that the estimates converge quite quickly. The estimated vertical gain coefficients  $K_v = [k_1, k_2, k_3, k_4]^T$  computed on the basis of the ECM from two ensembles En(PEF), En(EnOI) at the iteration  $t = 72$  are

$$K_v^{\text{pef}} = [144.59, -29.53, -34.44, -80.12]^T \quad (29a)$$

$$K_v^{\text{enoi}} = [34.04, -7.53, -3.3, -22.21]^T \quad (29b)$$

We remark that all the gain coefficients in two filters are of identical sign but the elements of  $K_v^{\text{enoi}}$  are of much less magnitudes than that of  $K_v^{\text{pef}}$ . It means that the EnOI will make less correction (compared to the PEF) to the forecast estimate. Two gains in (29a), (29b) will be used in the two filters PEF and EnOI to assimilate the observations.

The vertical gain coefficients for the CHF are taken from [8] and are equal to

$$K_v^{\text{chf}} = [185.97, 0, 0, -184.97]^T \quad (30)$$

### 5.3. Numerical Results

In **Figure 3** we show the instantaneous variances of the SSH innovation produced by three filters EnOI, CHF and PEF. It is seen that initialized by the same initial state, if the innovation variances in EnOI, CHF have a tendency to increase, this error remains stable for the PEF during all assimilation period. At the end of assimilation, the PE in the CHF is more than two times greater than that produced by the PEF. The EnOI has produced poor estimates, with error about two times greater than the CHF has done.

For the velocity estimates, the same tendency is observed as seen from **Figure 4** for the surface velocity PE

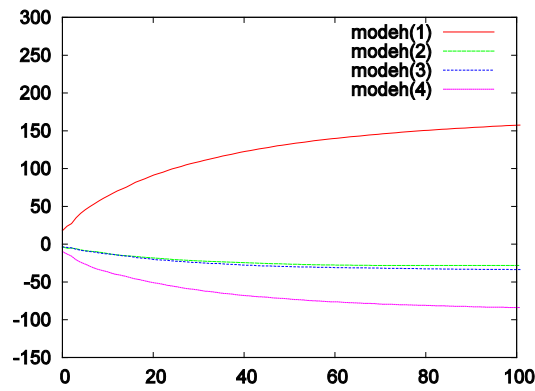


Figure 2. Vertical gain coefficients obtained during application of the Sampling Procedure for layer thickness correction during iterations.

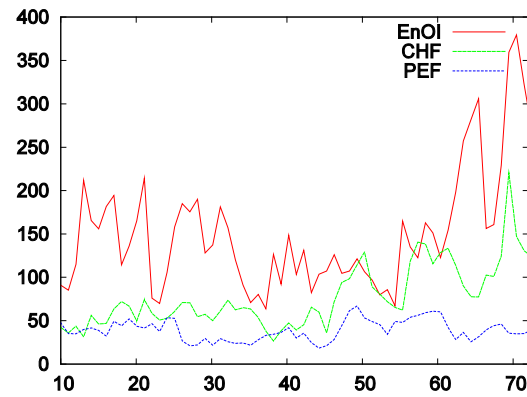


Figure 3. Performance comparison of EnOI, CHF and PEF: Variance of SSH innovation resulting from the filters EnOI, CHF and PEF.

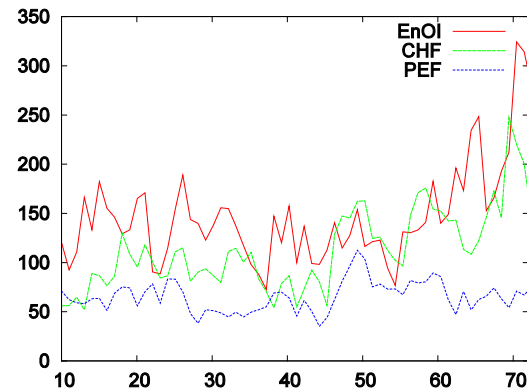


Figure 4. The prediction error variance of the  $u$  velocity component at the surface (cm/s) resulting from the EnOI, CHF and PEF.

errors. These results prove that the choice of ECM as a regularization factor on the basis of members of the En(PEF) allows to much better approach the true system state compared to that based on the samples taken from En(EnOI) or to that constructed on the basis of the physical consideration as in the CHF. The reason is that the members of En(PEF) by construction [8] are samples of the directions along which the prediction error increases

most rapidly. In other words, the correction in the PEF is designed to capture the principal important components in the decomposition of the covariance of the prediction error.

## 6. Conclusion

We have presented some properties of an efficient recursive procedure for computation of a statistical regularized estimator for the optimal linear estimator in a linear model with arbitrary non-negative covariance structure. The main objective of this paper is to obtain an algorithm which allows overcoming the difficulties concerned with high dimensions of the observation vector as well as that of the estimated vector of parameters. As it was seen, the recursive nature of the proposed algorithm allows dealing with high dimension of the observation vector. By initialization of the associated matrix equation by a low rank approximation covariance which accounts for only first leading components of the eigenvalue decomposition of the priori covariance matrix, the proposed algorithm permits to reduce greatly the number of estimated parameters in the algorithm. The efficiency of the proposed recursive procedure has been demonstrated by numerical experiments, with the systems of small and very high dimension.

## References

- [1] Hoang, H.S. and Baraille, R. (2013) A Regularized Estimator for Linear Regression Model with Possibly Singular Covariance. *IEEE Transactions on Automatic Control*, **58**, 236-241. <http://dx.doi.org/10.1109/TAC.2012.2203552>
- [2] Daley, R. (1991) Atmospheric Data Analysis. Cambridge University Press, New York.
- [3] Albert, A. (1972) Regression and the Moore-Penrose Pseudo-Inverse. Academy Press, New York.
- [4] Hoang, H.S. and Baraille, R. (2014) A Low Cost Filter Design for State and Parameter Estimation in Very High Dimensional Systems. *Proceedings of the 19th IFAC Congress*, Cape Town, 24-29 August 2014, 3156-3161.
- [5] Hoang, H.S. and Baraille, R. (2011) Approximate Approach to Linear Filtering Problem with Correlated Noise. *Engineering and Technology*, **5**, 11-23.
- [6] Golub, G.H. and van Loan, C.F. (1996) Matrix Computations. 3rd Edition, Johns Hopkins University Press, Baltimore.
- [7] Hoang, H.S., Baraille, R. and Talagrand, O. (2001) On the Design of a Stable Adaptive Filter for High Dimensional Systems. *Automatica*, **37**, 341-359. [http://dx.doi.org/10.1016/S0005-1098\(00\)00175-8](http://dx.doi.org/10.1016/S0005-1098(00)00175-8)
- [8] Hoang, H.S. and Baraille, R. (2011) Prediction Error Sampling Procedure Based on Dominant Schur Decomposition. Application to State Estimation in High Dimensional Oceanic Model. *Applied Mathematics and Computation*, **218**, 3689-3709. <http://dx.doi.org/10.1016/j.amc.2011.09.012>
- [9] Kalman, R.E. (1960) A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, **82**, 35-45. <http://dx.doi.org/10.1115/1.3662552>
- [10] Cooper, M. and Haines, K. (1996) Altimetric Assimilation with Water Property Conservation. *Journal of Geophysical Research*, **101**, 1059-1077. <http://dx.doi.org/10.1029/95JC02902>
- [11] Greenslade, D.J.M. and Young, I.R. (2005) The Impact of Altimeter Sampling Patterns on Estimates of Background Errors in a Global Wave Model. *Journal of Atmospheric and Oceanic Technology*, 1895-1917.
- [12] Spall, J.C. (2000) Adaptive Stochastic Approximation by the Simultaneous Perturbation Method. *IEEE Transactions on Automatic Control*, **45**, 1839-1853. <http://dx.doi.org/10.1109/TAC.2000.880982>