Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
STA2014-1256

On a Recursive Procedure for Parameter Estimation in Linear Model with Possibly Singular Covariance

Hong Son Hoang
Research Engineer
SHOM/HOM/REC
France

Remy Baraille
Research Engineer
SHOM/HOM/REC
France

ATINER CONFERENCE PAPER SERIES No: STA2014-1256

An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. The papers published in the series have not been refereed and are published as they were submitted by the author. The series serves two purposes. First, we want to disseminate the information as fast as possible. Second, by doing so, the authors can receive comments useful to revise their papers before they are considered for publication in one of ATINER's books, following our standard procedures of a blind review.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:
Son Hoang, H. and Baraille, R., (2014) **"On a Recursive Procedure for Parameter Estimation in Linear Model with Possibly Singular Covariance",**
Athens: ATINER'S Conference Paper Series, No: **STA2014-1256.**

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209 Email: info@atiner.gr
URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm
Printed in Athens, Greece by the Athens Institute for Education and Research. All rights reserved. Reproduction is allowed for non-commercial purposes if the source is fully acknowledged.
ISSN: **2241-2891**
18/09/2014

On a Recursive Procedure for Parameter Estimation in Linear Model with Possibly Singular Covariance

Hong Son Hoang
Research Engineer
SHOM/HOM/REC
France

Remy Baraille
Research Engineer
SHOM/HOM/REC
France

Abstract

An efficient recursive procedure is proposed for computation of a statistical regularized estimator for the optimal linear estimator in a linear model. Regularization is made by introducing a priori non-negative covariance structure for the vector of estimated parameters.

This procedure is designed to overcome the difficulties related to the very high dimension of the vector of parameters as well as that of the observation vector.

Theoretical results related to properties of the proposed procedure are obtained and proved. Simple numerical example with Monte-Carlo simulation as well as parameter estimation in a very high oceanic model are presented to demonstrate the utility of the proposed approach.

Keywords: Linear model; regularization; recursive algorithm; non-negative covariance structure.

Introduction

Hoang and Baraille (2013) introduce a statistical regularized estimator for an optimal linear estimator of unknown vector in a linear model with arbitrary non-negative covariance structure,

(1.1) $$z = Hx + v$$

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

(1.2) $$E(v) = 0, E(vv^T) = V$$

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

where $E(.)$ 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 Hoang and Baraille (2013) the optimal linear estimator for the unknown vector x in the model (1)-(3) is defined as

(1.4) $$\hat{x} = Gz, G = H^+\left[I - V_1\overline{V}_1\right]^{+,T}, \overline{V}_1 = V_1^T(I - H^+H)$$

where V_1^T is the transpose of V_1, A^+ denotes the pseudo-inversion of A.

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 δ-approximations

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

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

As shown in Hoang and Baraille (2013), there are situations when 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 δ .The regularization procedure has been proposed in Hoang and Baraille (2013) 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 p, n may be very high.

This problem is very important for many practical applications. As example, consider data assimilation problems in meteorology and oceanography (Daley, 1991). For typical data set in oceanography, at each assimilation instant we have the observation vector with $p \approx 10^4 - 10^5$, $n \approx 10^6 - 10^7$. We will show that a simplified algorithm can be designed by regularization of the priori covariance matrix M for the vector of unknown parameters x.

Simple Recursive Method for Estimating the Vector of Parameters

Problem Statement: Free-noise Observations

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

$$z = Hx, z \in R^p, x \in R^n \tag{2.1}$$

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 $z = (z_1,...,z_p)^T$, $H = (h_1^T,...,h_p^T)^T$, i.e. z_i is the i^{th} component of z, h_i is the i^{th} row-vector of H. The problem is to obtain a simple recursive procedure to compute a solution of the system (2.1).

Iterative Procedure

To find a solution to (2.1) let us introduce the following system of recursive equations

$$x_{i+1} = x_i + K_{i+1}\left[z_{i+1} - h_{i+1}x_i\right],\ x_0 = \bar{x}, i = 0,..., p, \tag{2.2a}$$

$$K_{i+1} = M_{i+1}h_{i+1}^T / \left[h_{i+1}M_{i+1}h_{i+1}^T\right], \tag{2.2b}$$

$$M_{i+1} = M_i - K_{i+1}h_{i+1}M_i, i = 0,1,... \ ;\ M_0 \text{ is given.} \tag{2.2c}$$

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

Theorem 2.1. Suppose the system (2.1) is compatible. The for any finite $\bar{x}$ and symmetric positive definitive (SPD) matrix M_0 we have $Hx_p = z$.

In order to prove Theorem 2.1 we need

Lemma 2.1. The following equalities hold

(2.3) $$h_i M_j = 0, i = 1,..., j$$

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 $h_i M_l = 0, i = 1,..., l$. We have to prove

$$h_i M_{l+1} = 0, i = 1,..., l+1$$

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 $h_i M_l = 0, i = 1,..., l$ it implies $h_i M_l + 1 = 0, i = 1,..., l+1$ (End of Proof).

Lemma 2.2. The following equalities hold

(2.4) $$h_j x_i = z_j, j = 1,..., i$$

Proof. By induction. We have for $i = 1$, $h_1 x_1 = h_1 [x_0 + K_1 (z_1 - h_1 x_0]$. 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 $h_i x_l = z_i, i = 1,..., l$. We have to prove $h_i x_{l+1} = z_i, i = 1,..., l+1$. 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 2.1, $h_i K_{l+1} = h_i M_{l+1} h_{l+1}^T / [M_{l+1} h_{l+1} M_{l+1}] = 0$ as $h_i M_l = 0$, for all $i \le l$. (End of Proof).

Proof of Theorem 2.1.

Theorem follows from Lemma 2.2 since under the conditions of Theorem, from Eq. (9) for $j = p$ it follows $h_j x_p = z_j, j = 1,..., p$ or $Hx_p = z$. (End of Proof).

Corollary 2.1. Suppose the rows of $H_l := [h_1^T,..., h_l^T]$ are linearly independent. Then under conditions of Theorem 2.1,

$$rank[M_l] = n - l .$$

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 $\dim[R(M_1)] \le n-1$. We show now that it is impossible that $\dim[R(M_1)] < n-1$.

Suppose that $\dim[R(M_1)] \le n - n', n' = 2,3,...$ For simplicity, let $n' = 2$. It means that there exist $n-2$ linearly independent vectors $a_1,..., a_{n-2}$ 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 $a_1,..., a_{n-2}, b_1$. 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 $n' = 3,4,..., n$.

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

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

we have $\dim[R(M_1)] = n-1, \dim[\Delta M_{l+1}] = 1$. From Lemma 2.1 it follows $h_i M_{l+1} = 0, i = 1,..., l+1$ 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 is is impossible $\dim[R(M_{l+1})] < n-l-1$ hence $\dim[R(M_{l+1})] = n-l-1$ (End of Proof).

Comment 2.1. By verifying the rank of M_l, Corollary 2.1 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-c) yield the unique solution of the equation $z = Hx$ after n iterations.

Using the result (2.3) in Lemma 2.1, it is easy to see that

Corollary 2.2. Suppose h_{l+1} is linearly dependent on $h_1,...,h_l$. Then in (2.2), $M_{l+1} = M_l$.

Corollary 2.3. Suppose $h_1,...,h_l$ are linearly independent, h_{l+1} is linearly dependent on $h_1,...,h_l$. Then under the conditions of Theorem 2.1, $rank[R(M_{l+1})] = n-l$

Corollary 2.3 follows from the fact that when h_{l+1} is linearly dependent on $h_1,...,h_l$ from Corollary 2.2, $M_{l+1} = M_l$ and hence from Corollary 2.1, $rank[R(M_{l+1})] = n-l$.

Comment 2.2 The last equation for M_i in (2.2) is equivalent to that given in (3.14.1) in Albert (1972) since from Corollary 2.2 it follows automatically $M_{l+1} = M_l$ if h_{l+1} is linearly dependent on $h_1,...,h_l$. Another difference is that in (2.2) the initial M_0 may be any SPD matrix whereas in (3.14.1) of Albert (1972) 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.

Optimal Properties of the Solution of (2.2)

Regularized Estimate

Return to Eqs (1.1)-(1.3) and assume that $V=0, N=0$. The optimal estimator is given in the form

(3.1) $$\hat{x} = \bar{x} + MH^T\left[HMH^T\right]^+(z - H\bar{x})$$

where H^+ is the pseudo-inversion of H. Consider first the equation $z_1 = h_1 x$. By the same way the optimal estimator obtained on the basis of the first observation is

(3.2) $$\hat{x}_1 = \bar{x} + Mh_1^T \left[h_1 M h_1^T\right]^+ (z - H\bar{x}),\ M_1 = M_0 - Mh_1^T \left[h_1 M h_1^T\right]^+ h_1 M$$

If we apply Eq. (3.14.1) in [Albert], $M = I$ and instead of $\hat{x}_1$ we have

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

For simplicity, let $\bar{x} = 0$. Comparing (3.2) with (3.3) 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 (3.2) 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 that 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 Hoang and Baraille (2014) 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 $x_i, i = 1,2,...$ are projected onto the space $R[M]$.

Theorem 3.1. Consider the algorithm (2.2). Suppose $\bar{x} \in R[\overline{M}]$. Then all the estimates $x_i, i = 1,2,...$ belong the space spanned by the columns of $\overline{M}$, i.e. $x_i \in R[\overline{M}]$.

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 $x_i \in R[\overline{M}]$, it is sufficient to show that $K_{i+1} z_{i+1} \in R[\overline{M}]$. From $x_i \in R[\overline{M}]$, $x_i = x_{i-1} + K_i \zeta_i$, as $x_{i-1} \in R[\overline{M}]$ it follows that $K_i \zeta_i \in R[\overline{M}]$. But $K_i \zeta_i = M_{i-1} h_i^T \left[h_i M_{i-1} h_i^T\right]^+ \zeta_i$ hence the columns of M_{i-1} must belong to $R[\overline{M}]$. Again from the equation for M_i in (2.2) we have $M_i \in R[\overline{M}]$. It proves $K_{i+1} z_{i+1} \in R[\overline{M}]$
since $M_i \in R[\overline{M}]$ (End of proof).

Theorem 2.2 says that by specifying $M_0 \in R[\overline{M}]$ the algorithm (2.2) will produce the estimates belonging to the subspace $R[\overline{M}]$. Specification of the matrix $\overline{M}$ plays the most important task to produce the estimate with high quality if we are given a priori that the estimates must belong to $R[\overline{M}]$. At the second iteration from Lemma 2.1 it is seen that for $i = 1,2$, $h_i M_2 = 0$ and $h_i M'_2 = 0$. It means that two subspaces $R[M_2]$ and $R[M'_2]$ are orthogonal to null subspace $N[H_2]$ hence $R[M_2] \approx R[M'_2]$.

Comment 3.1

(i) In practice, as $\overline{M}$ is estimated from samples, it is important to ensure that $\bar{x} \in R[\overline{M}]$.

(ii) Theorem 3.1 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 $\bar{x} \in R[\overline{M}], M_0 = \overline{M}$. Thus the algorithm can be considered as that which finds the solution $Hx = z$ under the constraint $\bar{x} \in R[\overline{M}]$. 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 $R[H^T]$.

Minimal Variance Estimate

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

Theorem 3.2. Suppose x is a random variable having the mean $\bar{x}$ and covariance matrix $\overline{M}$. Then x_i generated by the recursive equations (2.2) is an unbiased and minimum variance estimate for x in the class of all unbiased estimates linearly dependent on $\bar{x}$ and $z_1,...,z_i$.

Proof. Introduce for the system (2.1),

$$z_1^i := (z_1,...,z_i)^T, H_1^i := (h_1^T,...,h_1^T)^T \tag{3.4}$$

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

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

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

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

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

(3.5) $$x'_i(B) = \bar{x} + B\left[z_1^i - H_1^i\bar{x}\right]$$

It means that all the estimate in (3.5) is unbiased.

Consider the minimization problem

$$J(B) = traceE(ee^T) \to \arg\min_B, e := x - x'_i(B)$$

We have

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

$$\overline{M} - \overline{M}H_1^{i,T}B^T - BH_1^i\overline{M} + BH_1^i\overline{M}H_1^{i,T}B^T$$

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

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

from which follows one of the solutions

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

If now instead of (3.4) we consider the system

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

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

(3.7a) $$\hat{x}_i(\alpha) = \bar{x} + K(\alpha)\left[z_1^i - H_1^i\bar{x}\right]$$

(3.7b) $$K(\alpha) = \overline{M}H_1^{i,T}\left[H_1^i\overline{M}H_1^{i,T} + \alpha I\right]^{-1}$$

Using the properties of the pseudo-inverse, one can prove that

(3.8) $$\lim_{\alpha \to 0} K(\alpha) = B_0$$

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 Hoang and Baraille (2011a) for the case of uncorrelated sequence $\{v_i\}$, one can show that

(3.9a) $$\hat{x}_{i+1}(\alpha) = \hat{x}_i(\alpha) + k_{i+1}(\alpha)\left[z_{i+1} - h_{i+1}\hat{x}_i(\alpha)\right],$$

(3.9b) $$k_{i+1}(\alpha) = M_{i+1}(\alpha)h_{i+1}^T / \left[h_{i+1}M_{i+1}(\alpha)h_{i+1}^T + \alpha\right],$$

(3.9c) $$M_{i+1}(\alpha) = M_i(\alpha) - k_{i+1}(\alpha)h_{i+1}M_i(\alpha)$$

(3.9d) $$M_0(\alpha) = \overline{M}$$

Letting $\alpha \to 0$ one comes to $\hat{x}_i(\alpha) \to x_i$ in (2.2) (End of proof).

Noisy Observations

The algorithm (3.9) thus yields the unbiased minimal variance (UMV) estimates for x in the situation when α represents the observation noise variance. We want to stress that these algorithms produce the UMV estimates only if $\overline{M} = M$ where M is the true covariance of the error $e := x - \bar{x}$ before arriving $z_i, i = 1,2,...$

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, of orders of $10^6 - 10^7$ (Daley, 1991). This happens because x is a collection of several variables defined in the three dimensional grid. It is therefore impossible to evaluate the matrices M_i in (2.2) and (3.9) with the number of elements $10^{12} - 10^{14}$. One of possible ways is to seek some leading eigenvectors of M and use them to approximate M. More precisely, let us have the following eigen-decomposition for M (Golub and Van Loan, 1996)

(3.10) $$M = UDU^T$$

In (3.10) the columns of U are the eigenvectors of M and D is diagonal with the elements $\lambda_1 \geq \lambda_2 \geq ...\lambda_n$ at the diagonal – the eigen-values of M. Let $U = [U_1, U_2]$, $D = diag[D_1, D_2]$. If we put in the algorithms (2.2) or (3.9) $\overline{M} = U_1 D_1 U_1^T$, then the algorithm (2.2), 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 $n \times m, m \le n$. Let $\overline{M}(m) = U_1(m)D_1(m)U_1^T(m)$.

Main Theoretical Results

Theorem 3.3

Consider two algorithms of the type (2.2) subject to two matrices

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

where the columns of the columns of $U_i(m_i)$ consist of m_i leading eigenvectors
of M, $m_1 \le m_2$. Then the following inequalities hold

$$E\left\|e(m_2)\right\|^2 \le E\left\|e(m_1)\right\|^2, m_1 \le m_2, e(m_i) = \hat{x}(m_i) - x, i = 1,2 \tag{3.11}$$

$\hat{x}(m_i)$ is the estimate produced by the algorithm (2.2) subject to $\overline{M} := \overline{M}(m_i)$, 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 $\bar{x} = 0$.

$$x = UD^{1/2}y = Ly, L := UD^{1/2}, M = UDU^T \tag{3.12}$$

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

Let y^l is a sample of y. Theorem 3.2 states that for all $x^l := Ly^l$, the algorithm (2.2) 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;

$\overline{M}$ - the initialized ECM in the algorithm (2.2).

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

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

There are the following different cases

1. $\overline{M} = M$: By Theorem 3.2 the algorithm (2.2) will produce the estimates of minimal variance for all x^l; This is true also if (2.2) is applied to $x^l(m)$.

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

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

2.2. 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

$$E\left\|e(n)\right\|^2 \rfloor \le E\left\|e(m)\right\|^2 \rfloor, m \le n, e(m) = \hat{x}(m) - x$$

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

3.1. $\overline{M} = M(m_1)$:

(i) $x^l \in R[M(m_1)]$ :: the estimates are of minimal variance
(ii) $x^l \notin R[M(m_1)]$: the estimates are not of minimal variance;

3.2. $\overline{M} = M(m_2)$: The algorithm (2.2) 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

(3.14) $$E\left\|e(m_2)\right\|^2 \rfloor \le E\left\|e(m_1)\right\|^2 \rfloor, m_1 \le m_2, e(m) = \hat{x}(m) - x$$

Simplified Algorithm

Theorem 3.4

Consider the algorithm (2.2) subject to
$\overline{M} := M(m), M(m) = U(m)D(m)U^T(m), m \le n$.
Then this algorithm can be rewritten in the form

(3.15a) $$x_{i+1} = U(m)x_e(i+1), x_e(i+1) \in R^m,$$

(3.15b) $$x_e(i+1) = x_e(i) + K_e(i+1)[z_{i+1} - h_e(i+1)x_e(i)],$$

$x_0 = \bar{x}, i = 0,1,..., p$

(3.15c) $$K_e(i+1) = M_e(i+1)h_e^T(i+1)\left[h_e(i+1)M_e(i+1)h_e^T(i+1)\right]^+,$$

(3.15d) $$h_e(i+1) = h_{i+1}U(m), M_e(0) = D(m)$$

It is seen that in the algorithm (3.15a-d), 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 maybe is very high but there is only some leading directions (leading eigenvectors of M) which are very important to be captured. Thus the algorithm (15a-d) is quite adapted for solving such problems if the initial covariance M can be constructed from physical considerations or numerical model, and next to decompose it to obtain an approximated decomposition

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

Mention that the version (3.15a-d) is very closed to that studied in Hoang *et al.* (2001) for ensuring a stability of the filter.

Numerical Example

Consider the system (1.1) subject to the covariance M ,

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

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

(4.2) $$H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Numerical computation of eigen-decomposition of $M = UDU^T$ yields

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

(4.4) $$D = diag(1.25, 0.64, 0.162)$$

Three algorithms of the type (3.15) are applied subject to three covariance matrices $\overline{M} = M(m) = U(m)D(m)U^T(m)$. They are denoted as ALG(m).

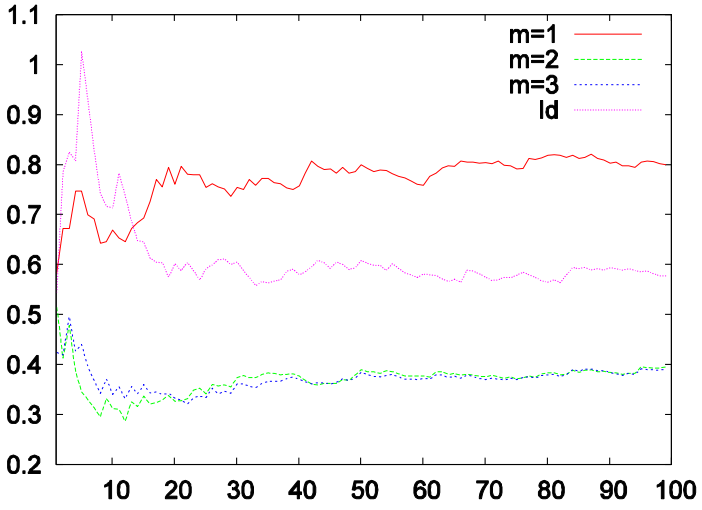
In Figure 4.1 we show the numerical results obtained from the Monte-Carlo simulation.

There are 100 samples simulating the true x which are generated by a random generator

distributed according to the normal distribution $N(./\bar{x}, M)$. The curves in Figure 4.1 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).

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

Figure 4.1. *Performance (rms) of the Algorithm (3.15) Subject to Different Projection Subspaces*



As seen from Figure 4.1, 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 is a simple way to improve the quality of the estimate by appropriately introducing it into the algorithm 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, simple 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 λ_3 is too small. That explains why ALG(2) and ALG(3) have produced almost the same results.

Experiment with Oceanic MICOM Model

MICOM Model

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

The MICOM model used in this experiment is exactly as that presented in Hoang and Baraille (2011b). We recall only that the model configuration is a domain situated in the North Atlantic from $30°N$ to $60°N$ and $80°W$ to $44°W$. The grid spacing is about $0.2°$ in longitude and in latitude, requiring the horizontal mesh $i = 1,...,140$; $j = 1,...,180$. The distance between two points $\Delta x = x_{i+1} - x_i \approx 20km$, $\Delta y = y_{i+1} - y_i \approx 20km$. 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^{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 $i_o = 10,20,...,140$; $j_o = 10,20,...,180$ 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.

Different Filters

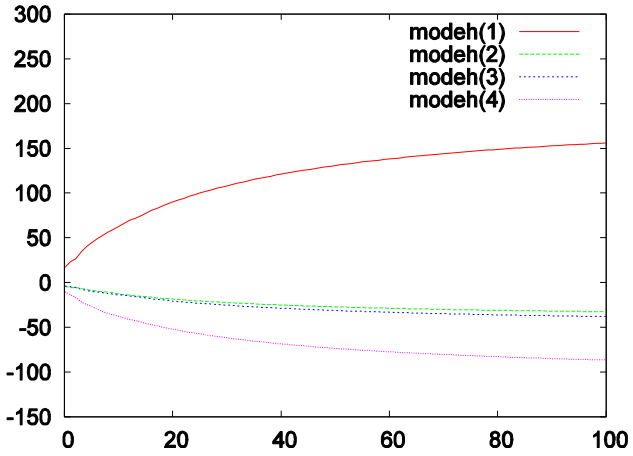
The different filters will be implemented to solve this assimilation problem. It is well known that determining the filter gain is one of the most important tasks in the design of a filter. As for the considered problem it is impossible to apply the standard Kalman filter (Kalman, 1960) since in the 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 (3.1) 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 (Hoang and Baraille, 2011b). Parallelly two other filters, one is the Cooper-Haines filter (CHF) (Cooper and Haines, 1996) and another is an EnOI (Ensemble based Optimal Interpolation) filter (Greenslade and Young, 2005) 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 Hoang and Baraille, 2011b). 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 Hoang and Baraille (2011b) (and it will be denoted as En(PEF)). As for the EnOI, the ensemble of background errors samples (the term used in Greenslade and Young (2005)) and will be denoted by En(EnOI)) will be used. The elements of En(EnOI) are constructed according to the method in Greenslade and Young (2005). 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 Hoang and Baraille (2014), using the hypothesis on separable vertical-horizontal structure for the ECM, we represent $M = M_v M_h$, 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 $M = M_v M_h$, the gain filter can be represented in the form

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

Figure 5.1. *Vertical Gain Coefficients Obtained during Application of the Samp-Proc for Layer Thickness Correction*



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 (Spall, 2000). The data matrix M_d is obtained from samples of the leading real Schur vectors as described in Hoang and Baraille (2011b).

Figure 5.1 shows the estimated vertical coefficients $k_{v_l}, l = 1,2,3,4$ 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

(5.1a) $$K_v^{pef} = [144.59, -29.53, -34.44, -80.12]^T \otimes I_p,$$

(5.1b) $$K_v^{enoi} = [34.04, -7.53, -3.3, -22.21]^T$$

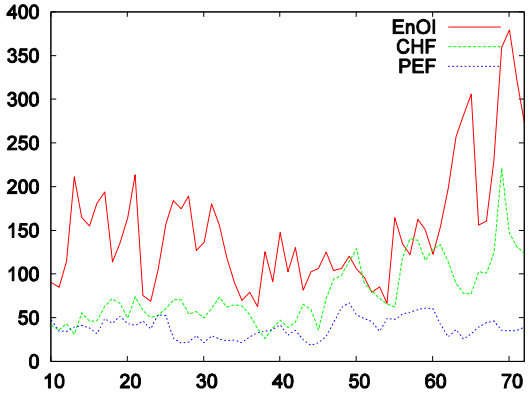
The reason for the choice $t = 72$ is that in practice the ensemble En(EnOI) has only a limited number of samples and for the comparison purpose we want to use two ensembles of the same number of samples. We remark that all the gain coefficients in two filters are of identical sign but the elements of K_v^{enoi} are of much less magnitudes than that of K_v^{pef}. Two gains in (5.1a-b) will be used in the two filters PEF and EnOI to assimilate the observations.

The vertical gain coefficients for the CHF are taken from Hoang and Baraille (2011b) and are equal to

(5.2) $$K_v^{chf} = [185.97, 0, 0, -184.97]^T$$

Numerical Results

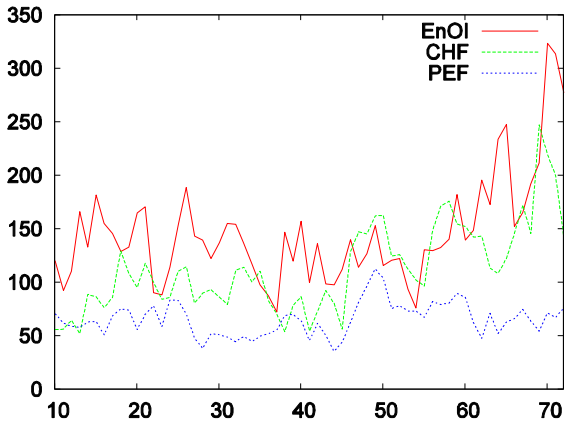
Figure 5.2. *Performance Comparison of EnOI, CHF and PEF: Variance of SSH Innovation Resulting from the Filters EnOI, CHF and PEF*



In Figure 5.2 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 very poor estimates, with error about two times greater than the CHF has done.

Figure 5.3. *The Prediction Error Variance of the* u *Velocity Component at the Surface* (cm/s) *Resulting from the EnOI, CHF and PEF*



For the velocity estimates, the same tendency is observed as seen from Figure 5.3 for the surface velocity PE errors. These results prove that regularization of the estimate on the basis of the ECM spanned by the members of En(PEF) allows to 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.

Conclusions

We have presented 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 problem studied here is emphasized on the situation with a priori non-negative covariance structure of estimated parameters. Initialization of the procedure with this covariance is proved to be important in regularization of the estimate when the number of observations is much less than the dimension of the vector of estimated parameters. That initialization plays also an important role in reducing the number of estimated parameters by using a principal component analysis which is very useful for the state estimation in very high dimensional systems. The efficiency of the proposed recursive procedure has been demonstrated by some numerical experiments, with small and very high dimension of the vector of estimated parameters.

References

A. Albert. *Regression and the Moore-Penrose Pseudo-Inverse*, Academy Press, NY and London, 1972.

R. Daley. *Atmospheric data analysis*. Cambridge University Press, New York, 1991.

A. Doucet, S. Godsill, C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10 (3): pages 197-208. doi:10.1023/A:1008935410038, 2000.

M. Cooper and K. Haines. Altimetric assimilation with water property conservation. *J. Geophys. Res*., 101(C1), pages 1059-1077, 1996.

G.H. Golub, and C.F. van Loan. *Matrix Computations, 3rd edition*, Johns Hopkins University Press, 1996.

D.J.M. Greenslade and I.R. Young. The Impact of Altimeter Sampling Patterns on Estimates of Background Errors in a Global Wave Model. *J. of Atmos. and Ocean Tech*., December, pages 1895-1917, 2005.

H. S. Hoang, R. Baraille. Approximate Approach to Linear Filtering Problem with Correlated Noise. *Engineering and Technology*, Vol:5, pages 11-23, 2011 (a).

H.S. Hoang and R. Baraille. Prediction error sampling procedure based on dominant Schur decomposition. Application to state estimation in high dimensional oceanic model. *Appl. Math. Comp*., v.218, no.7, pages 3689-3709, 2011 (b).

H.S. Hoang, R. Baraille. A Regularized Estimator For Linear Regression Model With Possibly Singular Covariance. *IEEE Trans. Automat. Contr*. 58(1): pages 236-241, 2013.

H.S. Hoang H.S. and R. Baraille. A Low Cost Filter Design for State and Parameter Estimation in Very High Dimensional Systems. *Proceedings of the 19th IFAC Congress*, Cape Town, August 24-29, pages 3156-3161, 2014.

H.S. Hoang, R. Baraille and O. Talagrand . On the design of a stable adaptive filter for high dimensional systems. *Automatica*, Issue 37, pages 341-359, 2001.

R.E. Kalman. A New Approach to Linear Filtering and Prediction Problems, *Transactions of the ASME - Journal of Basic Engineering*, Vol. 82: pages 35-45, 1960.

J.C. Spall. Adaptive Stochastic Approximation by the Simultaneous Perturbation Method, *IEEE Transactions on Automatic Control*, vol. 45, pages 1839 -1853, 2000.