

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



On Optimal and Simultaneous Stochastic Perturbations with Application to Estimation of High-Dimensional Matrix and Data Assimilation in High-Dimensional Systems

Hong Son Hoang and Remy Baraille

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.77273>

Abstract

This chapter is devoted to different types of optimal perturbations (OP), deterministic, stochastic, OP in an invariant subspace, and simultaneous stochastic perturbations (SSP). The definitions of OPs are given. It will be shown how the OPs are important for the study on the predictability of behavior of system dynamics, generating ensemble forecasts as well as in the design of a stable filter. A variety of algorithm-based SSP methodology for estimation and decomposition of very high-dimensional (Hd) matrices are presented. Numerical experiments will be presented to illustrate the efficiency and benefice of the perturbation technique.

Keywords: predictability, optimal perturbation, invariant subspace, simultaneous stochastic perturbation, dynamical system, filter stability, estimation of high-dimensional matrix

1. Introduction

Study in high-dimensional systems (HdS) today constitutes one of the most important research subjects thanks to the exponential increase of the power and speed of computers: after Moore's law, the number of transistors in a dense integrated circuit doubles approximately every 2 years (see Myhrvold [1]). However, this exponential increase is still far from being sufficient for responding to great demand on computational and memory resources in implementing the

optimal data assimilation algorithms (like Kalman filter (KF) [2], for example) for operational forecasting systems (OFS).

This chapter is devoted to the role of perturbations as an efficient tool for predictability of dynamical system, ensemble forecasting and for overcoming the difficulties in the design of data assimilation algorithms, in particular, of the optimal adaptive filtering for extremely HdS. In [3, 4], Lorenz has studied the problem of predictability of the atmosphere. It is found that the atmosphere is a chaotic system and a predictability limit to numerical forecast is of about 2 weeks. The barrier of predictability has to be overcome in order to increase the time period of a forecast further. The fact that estimates of the current state are inaccurate and that numerical models have inadequacies leads to forecast errors that grow with increasing forecast lead time. Ensemble forecasting aims at quantifying this flow-dependent forecast uncertainty. Today, a medium-range forecast has become a standard product. In the 1990s, the ensemble forecasting (EnF) technique was introduced in operational centers such as the European Centre for Medium-range Weather Forecast (ECMWF) (see Palmer et al. [5]), the NCEP (US National Center for Environmental Prediction) (Toth and Kalnay [6]). It is found that a single forecast can depart rapidly from the real atmosphere. The idea of the ensemble forecasting is to add the perturbations around the control forecast to produce a collection of forecasts that try to better simulate the possible uncertainties in a numerical forecast. The ensemble mean can then act as a nonlinear filter such that its skill is higher than that of individual members in a statistical sense (Toth and Kalnay [6]).

The chapter is organized as follows. Section 2 outlines first the optimal perturbation (OP) theory, on how the OP plays the important role for seeking the most growing direction of prediction error (PE). The predictability theory of the dynamical system as well as a stability of the filtering algorithm all are developed on the basis of OP. The definition of the optimal deterministic perturbation (ODP) and some theoretical results on the ODP are introduced. It is found that the ODP is associated with the right singular vector (SV) of the system dynamics. In Section 3, the two other classes of ODPs are presented: the leading eigenvector (EV) and real Schur vector (SchV) of the system dynamics. Mention that the first EV is the ODS in the eigen invariant subspace (EI-InS) of the system dynamics. As to the leading SchV, it is ODS in the Schur invariant subspace (Sch-InS) which is closely related to the EI-InS in the sense that the subspace of the leading SchVs, generated by the sampling procedure (Sampling-P, Section 3), converges to the EI-InS. In Section 4, we present the other type of OP called as optimal stochastic perturbation (OSP). Mention that the OSP is a natural extension of the ODP which gives insight into understanding of what represents the most growing PE and how one can produce it by stochastically perturbing the initial state. One important class of perturbations (known as simultaneous stochastic perturbation—SSP) is presented in Section 5. It will be shown that the SSP is very efficient for solving optimization problems in high-dimensional (Hd) setting. The different algorithms for estimating, decomposing ... Hd matrices are also presented here. Numerical examples are presented in Section 6 for illustrating the theoretical results and efficiency of the OPs in solving data assimilation problems. The experiment on data assimilation in the Hd ocean model MICOM by the filters constructed on the basis of the Schur ODSs and SSPs is presented in Section 7. The concluding remarks are presented in Section 8.

2. Optimal perturbations: predictability and filter stability

2.1. Stability of filter

The behavior of atmosphere or ocean is recognized as highly sensitive to initial conditions. It means that a small change in an initial condition can alter strongly the trajectory of the system. It is therefore important to be able to know about the directions of rapid growth of the system state. The research on OP is namely aimed at finding the methods to better capture these rapidly growing directions of the system dynamics, to optimize the predictability of the physical process under consideration.

To explain this phenomenon more clearly, consider a standard linear filtering problem

$$x(k+1) = \Phi x(k) + w(k+1), z(k+1) = Hx(k+1) + v(k+1). \quad (1)$$

where $\Phi \in R^{n \times n}$ is the state transition matrix, $H \in R^{p \times n}$ is an observation matrix. Under standard conditions related to the model and observation noises w_k, v_k , the minimum mean squared (MMS) estimate \hat{x}_k can be obtained by the well-known KF [2].

$$\hat{x}(k+1) = \hat{x}(k+1/k) + K\zeta(k+1), \hat{x}(k+1/k) = \Phi\hat{x}(k) \quad (2)$$

where $\zeta(k+1) = z(k+1) - H\hat{x}(k+1/k)$ is the innovation vector, $\hat{x}(k+1)$ is the filtered (or analysis) estimate, $\hat{x}(k+1/k)$ is the one-step ahead prediction for $x(k+1)$. The KF gain K is given by

$$K = MH^T [HMH^T + R]^{-1} \quad (3)$$

From Eq. (2), it can be shown that the transition matrix for the filtered estimate equation is expressed by $L = [I - KH]\Phi$.

For HdS, the KF gain (3) is impossible to compute. In a study by Hoang et al. [7], it is suggested to find the gain with the structure

$$K = P_r K_e \quad (4)$$

with $P_r \in R^{n \times n_e}$ —an operator projecting a vector from the reduced space R^{n_e} to the full system space R^n , $K_e \in R^{n_e \times p}$ is the gain for the reduced filter. One of very important questions arising here is how one can choose a subspace of projection and structure of K_e to make L to be stable? It is found in the work done by Hoang et al. [8] that detectability of the input-output system (1) is sufficient for the existence of a stabilizing gain K and this gain can be constructed with P_r consisting from all unstable EVs (or unstable SVs, SchVs. See Section 3) of the system dynamics.

2.2. Singular value decomposition and optimal perturbations

Consider the singular value decomposition (SVD) of Φ [9],

$$\begin{aligned}\Phi &= UDV^T, D = \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_n], \sigma_1 \geq \sigma_2 \dots \geq \sigma_n, \\ U &= [U_1, U_2], D = \text{block diag} [D_1, D_2], V = [V_1, V_2]\end{aligned}\quad (5)$$

where $U_1, V_1 \in R^{\times n_1}$, $D_{n_1} \in R^{n_1 \times n_1}$, n_1 is the number of all unstable and neutral SVs of Φ . In the future, for simplicity, unless otherwise stated, we say on the set of all unstable SVs as that including all unstable and neutral SVs.

Suppose the system is *s-detectable* (detectability of all the columns of U_1 or V_1). From the research by Hoang et al. [10], there exists a stabilizing gain K_s (sufficient but not necessary condition) with $P_r = U_1$ (see Eq. (4)) such that the transition matrix $L = (I - K_s H)\Phi$ is stable. It signifies that the filter is stable and the estimation error is bounded. The columns of U_1 , i.e., the left unstable SVs of Φ , serve as a basis for seeking appropriate correction in the filtering algorithm.

On the other hand, in practice, for extreme HdS, one cannot compute all elements of U_1 but only some of its subset $U'_1 \in U_1$. Using U'_1 instead of U_1 cannot guarantee a filter stability. The ensemble forecasting has been proposed as an approach to prevent a possible large error in the forecast and requires a knowledge on the rapidly growing directions of the PE. In this context, the OPs appear to be important which allow to search the rapid growing directions of the PE by model integration of OPs. They (i.e., OPs) are infact the unstable right SVs (RSV).

2.3. Optimal perturbation

Let δx be a given perturbation, representing an error (uncertainty, deterministic, or stochastic) around the true system state x^* , i.e., $\hat{x}_f = x^* + \delta x$, (at some instant k). The prediction of the system state \hat{x}_p can be obtained by forwarding the numerical model (1) on the basis of \hat{x}_f —filtered estimate, i.e., $\hat{x}_p = \Phi \hat{x}_f$. We have then

$$\hat{x}_p = \Phi \hat{x}_f = \Phi[x^* + \delta x] = \Phi x^* + \Phi \delta x, \quad (6)$$

One sees the perturbation δx in the initial system state “grows” into $\Phi \delta x$ which represents uncertainty in the forecast.

In general, the perturbation δx may be any element in the n -dimensional space, R^n , i.e., $\delta x \in R^n$. For $e_f^{(l)} := \delta x_f^{(l)}$ —a sample of the filtered error (FE) e_f , integrating the model by e_f results in $e_p^{(l)} := \Phi \delta x_f^{(l)}$ —a sample for the PE e_p . By generating the ensemble of perturbations $E_f(L) := \{e_f^{(l)}, l = 1, \dots, L\}$ according to the distribution of e_f , one can produce the ensemble of PE samples $E_p(L) := \{e_p^{(l)}, l = 1, \dots, L\}$ and use them to estimate the distribution of the PE e_p . This serves as a basis for the particle filtering [11].

The ensemble-based filtering (EnBF) algorithm is simplified for the standard linear filtering problems with e_f being of zero mean and the error covariance matrix (ECM) P . This technique is aimed to approximate the ECM without solving the matrix Ricatti equation (Ensemble KF—EnKF, [12]). Mention that at the present, one can generate only about $O(100)$ samples at each

assimilation instant. This ensemble size is too small compared to the system dimension. That is why it is important to have a good strategy for selecting the “optimal” samples (perturbations) to better approximate the ECM in the filtering algorithm.

2.3.1. Optimal deterministic perturbation

Introduce

$$\{S(\delta x) : \delta x, \|\delta x\|_2 = \langle \delta x, \delta x \rangle = 1\} \quad (7)$$

where $\|\cdot\|_2$ denotes the Euclidean vector norm (here $\langle \cdot, \cdot \rangle$ denotes the dot product). Let Φ have the SVD (5).

Definition 2.1. The ODP δx^o is the solution of the extremal problem

$$J(\delta x) = \|\Phi \delta x\|_2 \rightarrow \max_{\delta x} \quad (8)$$

under the constraint (7). One can prove

Lemma 2.1. The optimal perturbation in the sense (8) and (7) is

$$\delta x^o = (+/-)v_1$$

where v_1 is the first right SV of Φ .

2.3.2. Subspaces of ODPs

Introduce

$$\Phi_1 := \Phi - \sigma_1 u_1 v_1^T$$

Consider the optimization problem

$$J_1(\delta x) = \|\Phi_1 \delta x\|_2 \rightarrow \max_{\delta x} \quad (9)$$

under the constraint (7). Similar to the proof of Lemma 2.1, one can prove

Lemma 2.2. The optimal perturbation in the sense (9) and (7) is

$$\delta x^o = (+/-)v_2$$

where v_2 is the second right SV of Φ .

By iteration, for

$$\Phi_i := \Phi_{i-1} - \sigma_i u_i v_i^T, i = 1, \dots, n - 1; \Phi_0 = \Phi. \quad (10)$$

applying Lemma 2.2 with slight modifications, one finds that the OPs for $\Phi_i, i = 0, 1, \dots, n - 1$ are $(+/-)v_i, i = 1, 2, \dots, n..$

Theorem 2.1. The optimal perturbation for the matrix Φ_{i-1} , $i = 1, \dots, n$ is $(+/-)v_i$ where v_i is the i th leading right SV of Φ_{i-1} , $i = 1, \dots, n$.

The OP for Φ_{i-1} will be called the i^{th} OP for Φ (or the i^{th} SOP—singular OP).

Comment 2.2. The OPs, presented above, are optimal in the sense of the Euclidean norm $\|\cdot\|_2$. In practice, there is a need to normalize the state vector (using the inverse of the covariance matrix M). The normalization is done by changing $\delta x' = M^{-1/2}\delta x$, and all the results presented above remain valid s.t. the new $\delta x'$,

$$\|\delta x'\|_2 = \langle M^{-1/2}\delta x, M^{-1/2}\delta x \rangle = \langle \delta x, M^{-1}\delta x \rangle := \|\delta x\|_{M^{-1}}$$

The weighted norm $\|\delta x\|_{M^{-1}}$ is known as the Mahanalobis norm.

As to the PE, a normalization is also applied in order to have a possibility to compare different variables like density, temperature, velocity ... In this situation, the norm for $y = \Phi\delta x$ may be seminorm [5].

2.4. Ensemble forecasting

The idea of ensemble forecasting is that instead of performing “deterministic” forecasts, stochastic forecasts should be made: several model forecasts are performed by introducing perturbations in the filtered estimate or in the models.

Since 1994, NCEP (National Centers for Environmental Prediction, USA) has been running 17 global forecasts per day, with the perturbations obtained using the method of breeding growing perturbations. This ensures that the perturbations contain growing dynamical perturbations. The length of the forecasts allows the generation of outlook for the second week. At the ECMWF, the perturbation method is based on the use of SVs, which grow even faster than the bred or Lyapunov vector perturbations. The ECMWF ensemble contains 50 members [13].

3. Perturbations based on leading EVs and SchVs

3.1. Adaptive filter (AF)

The idea underlying the AF is to construct a filter which uses feedback in the form of the PE signal (innovation) to adjust the free parameters in the gain to optimize the filter performance. If in the KF, the optimality is defined as a minimum mean squared error (MMS), in the AF, optimality is understood in the sense of MMS for the prediction output error (innovation). This definition allows to define the optimality of the filter in the realization space, but not in the probability space as done in the KF.

The optimal gain thus can be determined from solving the optimization problem by adjusting all elements of the filter gain. There are two major difficulties:

- i. **Instability:** As the filter gain is estimated stochastically during the optimization process, the filter may become unstable due to the stochastic character of the filter gain.
- ii. **Reduction of tuning parameters:** For extreme HdS, the number of elements in the filter gain is still very high. Reduction of the number of tuning gain elements is necessary.

3.2. Leading EVs and SchVs as optimal perturbations

Interest on stability of the AF arises soon after the AF has been introduced. The study on the filter stability shows that it is possible to provide a filter stability when the system is *detectable* [8]. For the different parameterized stabilizing gain structures based on a subspace of unstable and neutral EVs, see [8]. As the EVs may be complex and their computation is unstable (Lanczos [14]), in [8], it is proved that one can also ensure a stability of the filter if the space of projection is constructed from a set of unstable and neutral SchVs of the system dynamics. The unstable and neutral real SchVs are referred to as SchVs associated with the unstable and neutral eigenvalues of the system dynamics. The advantage of the real SchVs is that they are real, orthonormal, and their computation is stable. Moreover, the algorithm for estimating dominant SchVs is simple which is based on the power iteration procedure (Sampling-P, see [15]). As to the unstable SVs, although they are real and orthonormal, their computation requires an adjoint operator (the transpose matrix Φ^T). Construction of adjoint code (AC) is a time-consuming and tedious process. Approximating leading SVs without (AC) can be done on the basis of Algorithms 5.2.

3.3. EVs as optimal perturbations in the invariant subspace

Let Φ be diagonalizable. Introduce the set

$$\{EV_1(x, \lambda) : x \in C^n, \|x\|_2 = 1, \lambda \in C^1 : \Phi x = \lambda x\}. \tag{11}$$

The subspace of $x \in C^n$ satisfying $\Phi x = \lambda x$ for some $\lambda \in C^1$ is known as an invariant subspace of Φ : the matrix Φx acts on to stretch the vector x but conserves the direction of x . Consider the optimization problem

$$J(\delta x) = \|\Phi \delta x\|_2 \rightarrow \max_{\delta x, (\delta x, \lambda) \in EV_1(\delta x, \lambda)} \tag{12}$$

It is seen that the optimal solution is the first EV $x_{ei}(1)$ of Φ with the largest magnitude equal to $|\lambda_1|$. We will call λ_1 a first optimal EV perturbation (denoted as EI-OP).

For a symmetric matrix, the EI-OP coincides with the SOP. The EI-OP is not unique.

By solving the optimization problem (8) s.t.

$$\{EV_2(x, \lambda) : x \in C^n, \|x\|_2 = 1 : \Phi x = \lambda x, \lambda \in C^1, |\lambda| < |\lambda_1|\}. \tag{13}$$

one finds the second EI-OP $x_{ei}(2)$. In a similar way, by defining

$$\{EV_i(x, \lambda) : x \in \mathbb{C}^n, \|x\|_2 = 1 : \Phi x = \lambda x, \lambda \in \mathbb{C}^1, |\lambda| < |\lambda_{i-1}|\}. \quad (14)$$

for $i = 1, 2, \dots, n-1$, we obtain a sequence of EI-OPs $x_{ei}(i)$, $i = 1, 2, \dots, n$. The first n_e EI-OPs are unstable SVs.

In general, for a defective case (not diagonalizable), Φ does not have n linearly independent EVs and the independent generalized EVs can serve as “optimal” perturbations to construct a subspace of projection in the AF.

To summarize, let the EV decomposition be

$$X_{ei} J X_{ei}^{-1} = \Phi \quad (15)$$

where J is a matrix of Jordan canonical form, X_{ei}^{-1} is the matrix inverse of X_{ei} (see Golub and Van Loan [9]). The columns of X_{ei} are the EVs of Φ , J is a block diagonal with the diagonal blocks of 1 or 2 dimensions. The rank k decomposition is $X_{ei,1} J_1 \tilde{X}_{ei,1}$ where

$$\begin{aligned} EV_k &:= X_{ei,1} J_1 \tilde{X}_{ei,1}, \\ X_{ei} &= [X_{ei,1}, X_{ei,2}], J = \text{block diag}[J_1, J_2], \\ \tilde{X}_{ei} &:= X_{ei}^{-1} = [\tilde{X}_{ei,1}^T, \tilde{X}_{ei,2}^T], \end{aligned} \quad (16)$$

with $X_{ei,1} \in \mathbb{R}^{n \times k}$, $X_{ei,2} \in \mathbb{R}^{n \times (n-k)}$. Multiplying the right of EV_k by $X_{ei,1}$ yields $X_{ei,1} J_1$, i.e., we obtain the k largest (in modulus) perturbations in the eigen (invariant) space of Φ . The perturbations being the column vectors of $X_{ei,1}$ (i.e., the k first EVs of Φ) are the first k OPs of Φ in the eigen-invariant subspace (EI-InS).

3.4. Dominant SchVs as OPs in the Schur invariant subspace

The study of Hoang et al. [8] shows that the subspace of projection of the stable filter can be constructed on the basis of all unstable EVs or SchVs of Φ .

Compared to the EVs, the approach based on real Schur decomposition is of preference in practice since the SchVs are real and orthonormal. Moreover, there exists a simple, power iterative algorithm for approaching the set of real leading SchVs. According to Theorem 7.3.1 in Golub and Van Loan [9], the subspace $R[X_{s,1}]$ spanned by the n_u leading SchVs converges to the unique invariant subspace $D_{n_u}(\Phi)$ (called a *dominant* invariant subspace) associated with the eigenvalues $\lambda_1, \dots, \lambda_{n_u}$ if $|\lambda_{n_u}| > |\lambda_{n_u+1}|$. In this sense, we consider the leading SchVs as OPs (denoted as Sch-OP) in the Schur invariant subspace (Sch-InS).

4. Optimal stochastic perturbation (OSP)

In Section 2, the perturbation δx is deterministic (see Definition 2.1). In practice, it happens that δx is of stochastic nature. For example, the priori information on the FE is an zero mean

random vector (RV) with the ECM P . The question arising here is how one determine the OP in such situation and how to find it.

We will consider now δx as an element of the Hilbert space \mathcal{H} of RVs. This space \mathcal{H} is a complete normed linear vector space equipped with the inner product

$$\langle x, y \rangle_H = E \langle x, y \rangle \tag{17}$$

where $E(\cdot)$ denotes the mathematical expectation. The norm in \mathcal{H} is defined as

$$\|x\|_H = \sqrt{E \langle x, x \rangle} \tag{18}$$

All elements of \mathcal{H} are of finite variance and for simplicity, we assume they all have zero mean value.

Introduce the set of RVS

$$S_s(\delta x) := \{ \delta x : \|\delta x\|_H = 1 \} \tag{19}$$

Definition 4.1. The optimal stochastic perturbation (OSP) δx^o is the solution of the extremal problem

$$J(\delta x) = \|\Phi \delta x\|_H \rightarrow \max_{\delta x} \tag{20}$$

under the constraint (19). One can prove

Lemma 4.1. For $\delta x \in S_s(\delta x)$, there exists $\delta y = (\delta y_1, \dots, \delta y_n)^T \in S_s(\delta x)$ such that $\delta x = \sum_{k=1}^n v_k \delta y_k$.

Lemma 4.2. The optimal perturbation in the sense (20) and (19) is

$$\delta x^o = \psi v_1$$

where ψ is a RV with zero mean and unit variance, v_1 is the first right SV of Φ .

Comment 4.1. Comparing the ODP with OSP shows that if the ODS is the first right SV (defined up to the sign), the OSP is an ensemble of vectors lying in the subspace of the first right SV with the lengths being the samples of the RV of zero mean and unit variance.

Introduce

$$\Phi_1 := \Phi - \sigma_1 u_1 v_1^T$$

and consider the objective function

$$J(\delta x) = \|\Phi_1 \delta x\|_H \rightarrow \max_{\delta x} \tag{21}$$

Lemma 4.3. The optimal perturbation in the sense (21) and (19) is $\delta x^o = \psi v_2$, where ψ is an RV with zero mean and unit variance, v_2 is the first right SV of Φ .

By iteration, for

$$\Phi_i := \Phi_{i-1} - \sigma_i u_i v_i^T, i = 1, \dots, n - 1; \Phi_0 = \Phi. \quad (22)$$

applying Lemma 4.3 with slight modifications, one finds that the OSP for $\Phi_k, k = 0, 1, \dots, n - 1$ are $\psi v_k, k = 1, 2, \dots, n$, where ψ is an RV with zero mean and unit variance, v_k is the k^{th} right SV of Φ .

Theorem 4.1. The optimal perturbation for the matrix $\Phi_{i-1}, i = 1, \dots, n$ is ψv_i , where ψ is an RV with zero mean and unit variance, v_k is the k^{th} right SV of Φ .

5. Simultaneous stochastic perturbations (SSP)

In [16], Spall proposes a simultaneous perturbation stochastic approximation (SPSA) algorithm for finding optimal unknown parameters by minimizing some objective function. The main feature of the simultaneous perturbation gradient approximation (SPGA) resides in the way to approximate the gradient vector (in average): a sample gradient vector is estimated by perturbing simultaneously all components of the unknown vector in a stochastic way. This method requires only two or three measurements of the objective function, regardless of the dimension of the vector of unknown parameters. In a study by Hoang and Baraille [17], the idea of the SPGA is described in detail, with a wide variety of applications in engineering domains. The application to estimation of ECM in the filtering problem is given in the work done by Hoang and Baraille [18]. In the research by Hoang and Baraille [19], a simple algorithm for estimating the elements of an unknown matrix as well as the way to decompose the estimated matrix into a product of two matrices, under the condition that only the matrix-vector product is accessible, has been proposed.

5.1. Theoretical background of SPGA: gradient approximation

The component-wise perturbation is a method for numerical computation of the cost function with respect to the vector of unknown parameters. It is based on the idea to perturb separately each component of the vector of parameters. For very HdS, this technique is impossible to implement. An alternative to the component-wise perturbation is the SSP approach.

Let

$$DJ(\theta_0) = \left[\frac{\partial J(\theta_0)}{\partial \theta_1}, \dots, \frac{\partial J(\theta_0)}{\partial \theta_{n_\theta}} \right]^T$$

denote the gradient of $J(\theta)$ computed at $\theta = \theta_0$. Suppose $\Delta_j, j = 1, \dots, n$ are RVs independent identically distributed (i.i.d) according to the Bernoulli law which assumes two values +1 or -1 with equal probabilities 1/2. It implies that

$$E(\Delta_j) = 0, E(\Delta_j)^2 = 1, E(\Delta_j^{-1}) = 0, E(\Delta_j^{-1})^2 = 1, j = 1, 2, \dots, n \quad (23)$$

Suppose $J(\theta)$ is infinitely differentiable at $\theta = \theta_0$. Using a Taylor series expansion,

$$\Delta J := J(\theta_0 + \bar{\delta}\theta) - J(\theta_0) = \bar{\delta}\theta^T DJ(\theta_0) + (1/2)\bar{\delta}\theta^T D^2J(\theta_0)\bar{\delta}\theta + \dots \quad (24)$$

where $D^2J(\theta_0)$ is the Hessian matrix computed at $\theta := \theta_0$. For the choice

$$\bar{\delta}\theta := (\delta\theta_1, \dots, \delta\theta_{n_\theta})^T = c\bar{\Delta}, \bar{\Delta} = (\Delta_1, \Delta_2, \dots, \Delta_{n_\theta})^T, \quad (25)$$

c is a small positive value, from Eq. (24)

$$\Delta J(\theta_0) = c\bar{\Delta}^T DJ(\theta_0) + (c^2/2)\bar{\Delta}^T D^2J(\theta_0)\bar{\Delta} + \dots$$

Dividing both sides of the last equality by $\delta\theta_k = c\Delta_k$ implies

$$\begin{aligned} \Delta J(\theta_0)/\delta\theta_k &= DJ(\theta_0)^T \bar{\Delta}^k + (c/2)\bar{\Delta}^{k,T} D^2J(\theta_0)\bar{\Delta} + \dots \\ \bar{\Delta}^k &:= (\Delta_1\Delta_k^{-1}, \dots, 1, \dots, \Delta_n\Delta_k^{-1})^T \end{aligned} \quad (26)$$

Taking the mathematical expectation for both sides of the last equation yields

$$\begin{aligned} E[\Delta J(\theta_0)\delta\theta_k^{-1}] &= \\ DJ(\theta_0)^T E(\bar{\Delta}^k) + (c/2)E[\bar{\Delta}^T D^2J(\theta_0)\bar{\Delta}^k] + \dots \end{aligned} \quad (27)$$

One sees that from the assumptions on $\bar{\Delta}$, $E(\bar{\Delta}^k) = (0, \dots, 1, \dots, 0)^T$ it follows $DJ(u_0)^T E(\bar{\Delta}^k) = \partial J(\theta_0)/\partial\theta_k$. Moreover, as all the moments of the Bernoulli variables Δ_i and Δ_i^{-1} are finite, $E[\bar{\Delta}^T D^2J(\theta_0)\bar{\Delta}^k] = 0$ since there exists a finite $D^2J(\theta_0)$, one concludes that

$$E[\Delta J(\theta_0)\delta\theta_k^{-1}] = \partial J(\theta_0)/\partial\theta_k + O(c^2) \quad (28)$$

The result expressed by Eq. (28) constitutes a basis for approximating the gradient vector by simultaneous perturbation. The left of Eq. (28) can be easily approximated by noticing that for an ensemble of L i.i.d samples $[\bar{\Delta}^{(1)}, \dots, \bar{\Delta}^{(L)}]$, we can generate the corresponding ensemble of L i.i.d sample estimates for the gradient vector at $\theta = \theta_0$,

$$DJ^{(l)}(\theta_0) = \left[\frac{\Delta J^{(l)}(\theta_0)}{c\Delta_1^{(l)}}, \dots, \frac{\Delta J^{(l)}(\theta_0)}{c\Delta_n^{(l)}} \right]^T, l = 1, 2, \dots, L \quad (29)$$

where $\Delta_k^{(l)}$ is the k^{th} component of the l^{th} sample $\bar{\Delta}^{(l)}$. The left of Eq. (28) is then well approximated by averaging L sample gradients in Eq. (29),

$$E[\Delta J(\theta_0)/\delta\theta_k] \approx (1/L) \sum_{l=1}^L \eta_k^{(l)}, \eta_k^{(l)} := \frac{\Delta J^{(l)}(\theta_0)}{c\Delta_k^{(l)}}, \quad (30)$$

Introduce the notations

$$\bar{m}_k := E[\Delta J(\theta_0)/\delta\theta_k], m_k(L) := (1/L) \sum_{l=1}^L \eta_k^{(l)}, e_k := m_k(L) - \bar{m}_k. \quad (31)$$

Theorem 1 (Hoang and Baraille [19]) states that the estimate $m(L) := (m_1(L), \dots, m_{n_\theta}(L))^T$ converges to the gradient vector $DJ(\theta_0)$ as $L \rightarrow \infty$ and $c \rightarrow 0$ with the order $O(1/L)$ where

$$m(L) := (1/L) \sum_{l=1}^L \eta^{(l)}.$$

5.2. Algorithm for estimation of an unknown matrix

Let $\bar{\Delta} := (\Delta_1, \dots, \Delta_n)^T$, $\Delta_i, i = 1, \dots, n$ be Bernoulli independent and identically distributed (i.i.d.) variables assuming two values ± 1 with equal probability $1/2$. Introduce $[\bar{\Delta}]^{-1} := (1/\Delta_1, \dots, 1/\Delta_n)^T$, $\bar{\Delta}_c := c\bar{\Delta}$, $c > 0$ is a small positive value.

Algorithm 5.1. Suppose it is possible to compute the product $\Phi x = b(x)$ for a given x . At the beginning let $l = 1$. Let the value u be assigned to the vector x , i.e., $x := u$, L be a (large) fixed integer number.

Step 1. Generate $\bar{\Delta}^{(l)}$ whose components are l^{th} samples of the Bernoulli i.i.d. variables assuming two values $+/- 1$ with equal probabilities $1/2$;

Step 2. Compute $\delta b^{(l)} = \Phi(u + \bar{\Delta}_c^{(l)}) - \Phi u$, $\bar{\Delta}_c^{(l)} = c\bar{\Delta}^{(l)}$.

Step 3. Compute $g_i^{(l)} = \delta b_i^{(l)} [\bar{\Delta}_c^{(l)}]^{-1}$, δb_i is the i^{th} component of δb , $g_i^{(l)}$ is the column vector consisting of derivative of $b_i(u)$ w.r.t. u , $i = 1, \dots, m$.

Step 4. Go to Step 1 if $l < L$. Otherwise, go to Step 5.

Step 5. Compute

$$\hat{g}_i = \frac{1}{L} \sum_{l=1}^L g_i^{(l)}, i = 1, \dots, m, \hat{\Phi}(L) := D_x b = [\hat{g}_1, \dots, \hat{g}_m]^T.$$

5.3. Operations with Φ and its transpose

Algorithm 5.1 allows to store $\hat{\Phi}(L)$ as composed from the two ensembles of vectors elements:

$$\begin{aligned} En_L(\delta x) &:= [\delta x^{(1)}, \dots, \delta x^{(L)}], \delta x^{(l)} = c\bar{\Delta}^{(l)}, \\ En_L(\delta b) &:= [\delta b^{(1)}, \dots, \delta b^{(L)}], \delta b^{(l)} = (\delta b_1^{(l)}, \dots, \delta b_m^{(l)})^T, l = 1, \dots, L. \end{aligned} \quad (32)$$

The product $z = \widehat{\Phi}(L)y, y \in R^n$, can be performed as $z_i = \sum_{k=1}^n \widehat{\Phi}_{ik} y_k = \sum_{k=1}^n \left[\frac{1}{L} \sum_{l=1}^L \frac{\delta b_l^{(l)}}{\delta x_k^{(l)}} \right] y_k$, or in a more compact form

$$z = \frac{1}{L} \sum_{l=1}^L \alpha_l \delta b^{(l)}, \alpha_l := \sum_{k=1}^n \frac{y_k}{\delta x_k^{(l)}}. \tag{33}$$

Eq. (33) allows to perform $z = \widehat{\Phi}(L)y$ with $L(m + 2n) + 1$ elementary operations.

Similarly, computation of z_i of $z = \widehat{\Phi}^T(L)y, y \in R^m$ is performed as

$$z_i = \frac{1}{L} \sum_{l=1}^L \frac{1}{\delta x_i^{(l)}} \sum_{k=1}^m \delta b_k^{(l)} y_k, i = 1, \dots, n \tag{34}$$

5.4. Estimation of decomposition of Φ

Let Φ be a matrix of dimensions $(m \times n)$. For definiteness, let $m \leq n$ with $\text{rank}(\Phi) = m$. We want to find the best approximation for Φ among members of the class of matrices

$$\Phi_e = AB^T, A \in R^{m \times r}, B \in R^{n \times r}. \tag{35}$$

under the constraint

Condition (C) A, B are matrices of dimension $m \times r, r \times n, r \leq m, \text{rank}(AB^T) = r$.

Under the condition (C), the optimization problem is formulated as

$$J(A, B) = \|\Phi - \Phi_e\|_F^2 = \|\Phi - AB^T\|_F^2 \rightarrow \min_{(A,B)}, \tag{36}$$

where $\|\cdot\|_F$ denotes the Frobenius matrix norm. Consider Φ and let $U\Sigma V^T$ be SVD of Φ (5). Let $\tilde{\Phi} = \Phi + \Delta\Phi, \tilde{\Phi} = \tilde{U}\tilde{\Sigma}\tilde{V}^T$ and $\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \dots \geq \tilde{\sigma}_m, \tilde{\sigma}_k$ be the k^{th} singular value of $\tilde{\Phi}$. Then, we have

$$J(A_o, B_o) = \sum_{k=r+1}^m \sigma_k^2 \tag{37}$$

where $A_o B_o^T$ is a solution to the problem (36) s.t Condition (C) (Theorem 3.1 of Hoang and Baraille [19]).

Theorem 3.1. Hoang and Baraille [19] implies that $\Phi_e^o := A_o B_o^T$ is equal to the matrix formed by truncating the SVD of Φ to its first r SVs and singular values. It allows to avoid storing elements of the estimate $\widehat{\Phi}(L)$ of Φ (their number is of order $O(10^{m \times n})$).

5.4.1. Decomposition algorithms

Let the elements of Φ (or $\widehat{\Phi}$) be available (may be in algorithmic form). By perturbing stochastically simultaneously all the elements of A and B , one can write out the iterative algorithm for estimating the elements of A and B . For more detail, see Hoang and Baraille [19].

5.4.2. Iterative decomposition algorithm

Another way to decompose the matrix Φ is to solve iteratively the following optimization problems

Algorithm 5.2

At the beginning let $i = 1$.

Step 1. For $i = 1$, solve the minimization problem

$$J_1 = \|\Phi^1 - ab^T\|_F^2 \rightarrow \min_{a,b}, \quad a \in R^m, b \in R^n. \\ \Phi^1 := \Phi, \quad \text{rank}(ab^T) = 1$$

Its solution is denoted as $\hat{a}(i), \hat{b}(i)$.

Step 2. For $i < r$, put $i := i + 1$ and solve the problem

$$J_{i+1} = \|\Phi^i - ab^T\|_F^2 \rightarrow \min_{a,b}, \quad a \in R^m, b \in R^n. \quad \Phi^i := \Phi - \sum_{k=1}^{i-1} \hat{a}(k)\hat{b}^T(k), \\ \text{rank}(ab^T) = 1$$

Step 3. If $i = r$, compute

$$\hat{\Phi} = \hat{A}(r)\hat{B}^T(r), \quad \hat{A}(r) = [\hat{a}(1), \dots, \hat{a}(r)], \quad \hat{B}(r) = [\hat{b}(1), \dots, \hat{b}(r)].$$

and stop. Otherwise, go to *Step 2*.

From Theorem 3.2 of Hoang and Baraille [19], the couple $\hat{A}(r), \hat{B}(r)$ is a solution for the problem (36)(C).

6. Numerical example

Consider the matrix $\Phi \in R^{2 \times 2}$

$$\phi_{11} = 5, \phi_{12} = 7, \phi_{21} = -2, \phi_{22} = -4.$$

The singular values and the right SVs for Φ are displayed in **Table 1** which are obtained by solving the classical equations for eigenvalues of $\Phi^T\Phi$.

First, we apply Algorithm 5.1 to estimate the matrix Φ . **Figure 1** shows the estimates produced by Algorithm 5.1. It is seen that the estimates are converging quickly to the true elements of Φ .

Next Algorithm 5.2 (Iterative Decomposition Algorithm) has been applied to estimate the decomposition of the matrix Φ . After each i^{th} iteration, the algorithm yields $\hat{b}^{(i)}, \hat{c}^{(i)}$.

The different OPs are shown in **Table 2** where $x_{sv}^r(i)$, $x_{ei}(i)$, and $x_{sch}(i)$ are the theoretical SV-POs, EI-POs, Sch-POs. The vectors $\hat{x}_{sch}(i)$ are the components of X_i computed by *Algorithm 3.1*

Element	ϕ_{ij}	$\hat{\phi}_{ij}$	$\hat{\phi}_{ij^{(1)}}$	$\hat{\phi}_{ij^{(2)}}$
(1,1)	5.00	4.677	4.914	-0.237
(1,2)	7.00	7.07	7.662	-0.592
(2,1)	-2.00	-2.041	-1.08	-0.962
(2,2)	-4.00	-4.081	-1.683	-2.398

Table 1. Estimates of Φ obtained by Algorithm 5.2.

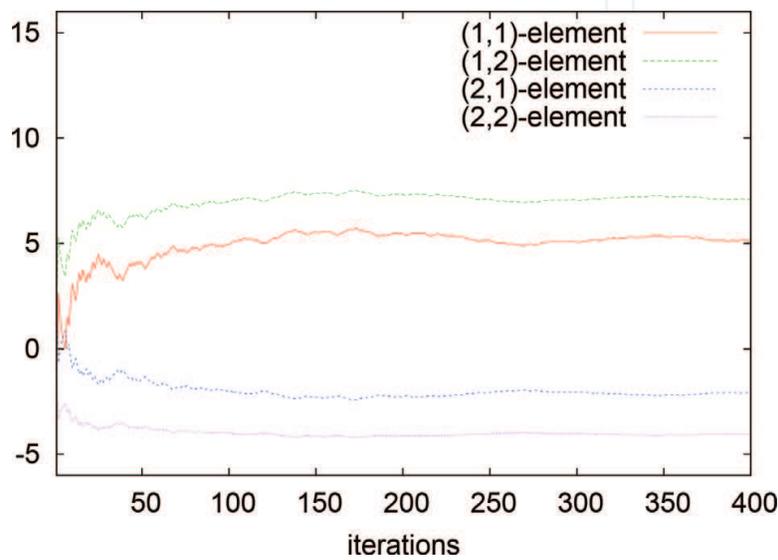


Figure 1. Estimates of elements of the matrix Φ .

Perturbations	Vector	Predictor	Amplification
$x_{sv}^r(1)$	$(0.554, 0.833)^T$	$(8.597, -4.438)^T$	9.676
$x_{sv}^r(2)$	$(0.833, -0.554)^T$	$(0.284, 0.551)^T$	0.62
$x_{ei}(1)$	$(0.962, -0.275)^T$	$(2.885, -0.824)^T$	3
$x_{ei}(2)$	$(0.707, -0.707)^T$	$(-1.414, 1.414)^T$	2
$x_{sch}(1)$	$(0.707, 0.707)^T$	$(8.485, -4.243)^T$	9.487
$x_{sch}(2)$	$(0.707, -0.707)^T$	$(-1.414, 1.414)^T$	2
$\hat{x}_{sch}(1)$	$(0.962, -0.275)^T$	$(-8.1, 4.4)^T$	9.22
$\hat{x}_{sch}(2)$	$(-0.275, -0.962)^T$	$(2.885, -0.824)^T$	3
$\hat{c}_n^{(1)}$	$(0.54, 0.842)^T$	$(8.592, -4.447)^T$	9.674
$\hat{c}_n^{(2)}$	$(0.372, 0.928)^T$	$(8.358, -4.457)^T$	9.472

Table 2. Different OPs.

(Sampling-P). As to $\widehat{c}_n^{(i)}$, they are the results of normalization (with the unit Euclidean norm) of $\widehat{c}^{(i)}$.

Looking at the first OPs, one sees that $x_{sv}^r(1)$, $x_{sch}(1)$, $\widehat{x}_{sch}(1)$, and $\widehat{c}_n^{(1)}$ produce almost the same amplification. The first $x_{ei}(1)$ has the amplification three times less than those of $x_{sv}^r(1)$, $x_{sch}(1)$. The second $x_{sch}(2)$ is much less optimal than $x_{sv}^r(2)$ and $\widehat{c}_n^{(2)}$. By comparing $x_{sv}^r(i)$ with $\widehat{c}_n^{(i)}$ for $i = 1, 2$, one concludes that the obtained results justify the correctness of Theorem 3.1 of Hoang and Baraille [19]. Mention that only $\widehat{x}_{sch}(i)$ and $\widehat{c}_n^{(i)}$ can be calculated for HdS.

In **Table 1**, we show the results obtained by Algorithm 5.2 after two consecutive iterations (matrix estimation in R^1 subspace). The elements of the true $\Phi = [\phi_{ij}]$ are displayed in the second column, whereas their estimates—in the third column,

$$\widehat{\Phi} = \widehat{\Phi}^{(1)} + \widehat{\Phi}^{(2)} = \sum_{i=1}^2 \widehat{b}^{(i)} \widehat{c}^{(i),T} \widehat{\Phi}^{(i)} := [\widehat{\phi}_{ij}^{(i)}] = \widehat{b}^{(i)} \widehat{c}^{(i),T}$$

The estimates, resulting from the first iteration, are the elements of $\widehat{\Phi}^{(1)}$ (**Table 1**, column 4). After the first iteration, $\Phi^{(2)} := \Phi - \widehat{\Phi}^{(1)} = b^{(2)} c^{(2),T}$ and the optimization yields the estimates $\widehat{\phi}_{ij}^{(2)}$ displayed in the column 5. From the columns 4–5, one sees that the first iteration allows to well estimate the two biggest elements $\Phi_{11} = 5$, $\Phi_{12} = 7$. In the similar way, the second iteration captures the two biggest elements of $\Phi^{(2)}$.

7. Assimilation in high-dimensional ocean model MICOM

7.1. Ocean model MICOM

To see the impact of optimal SchVs in the design of filtering algorithm for HdS, in this section, we present the results of the experiment on the Hd ocean model MICOM (Miami Isopycnal Ocean Model). This numerical experiment is identical to that described in Hoang and Baraille [15]. The model configuration is a domain situated in the North Atlantic from 30°N to 60°N and 80°W to 44°W; for the exact model domain and some main features of the ocean current produced by the model, see and Baraille [15]. The system state $x = (h, u, v)$ where $h = h(i, j, k)$ is a layer thickness and $u = u(i, j, k)$, $v = v(i, j, k)$ are two velocity components. Mention that after discretization, the dimension of the system state is $n = 302400$. The observations available at each assimilation instant are the sea surface height (SSH) with dimension $p = 221$.

7.1.1. Data matrix based on dominant Sch-Ops

The filter is a reduced-order filter (ROF) with the variable h as a reduced state and u, v are calculated on the basis of the geostrophy hypothesis. To obtain the gain in the ROF, first the *Algorithm 3.1* has been implemented to generate an ensemble of dominant SchVs (totally 72 SchVs, denoted as $En(SCH)$). The sample ECM $M^d(SCH)$ is computed on the basis of the

$En(SCH)$. Due to rank deficiency, the sample $M^d(SCH)$ is considered only as a data matrix. The optimization procedure is applied to minimize the distance between the data matrix $M^d(SCH)$ and the structured parametrized ECM $M(SCH) = M_v \otimes M_h$ which is written in the form of the Schur product of two matrices $M(SCH) = M_v(\theta) \otimes M_h(\theta)$. Here, M_v is the vertical ECM, M_h is the horizontal ECM [18]), (θ) is a vector of unknown parameters. Mention that the hypothesis on separability of the vertical and horizontal variables in the ECM is not new in the meteorology [20]. The gain is computed according to Eq. (3) with $R = \alpha I$, $\alpha > 0$ is a small positive value. The ROF is denoted as PEF (SSP).

7.1.2. Data matrix based on SSP approach

The second data matrix $M^d(SSP)$ is obtained by perturbing the system state according to the SSP method. The SSP samples are simulated in the way similar to that described above for generating $En(SCH)$, with the difference that the perturbation components $\delta h^{(l)}(i, j, k)$ are the i. d. random Bernoulli variables assuming two values ± 1 with the equal probability 1/2. The same optimization procedure has been applied to estimate $M(SSP)$. The obtained ROF is denoted as PEF (SSP).

Figure 2 shows the evolution of estimates for the gain coefficients $k(1)$ computed from the estimated coefficients of \hat{c}_{kl} of $M^d(SCH)$ and $M^d(SSP)$ on the basis of $En(SCH)$ (curve "schur") and $En(SSP)$ (curve "random"), during model integration. It is seen that two coefficients are evolved in nearly the same manner, of nearly the same magnitude as that of $k(1)$ in the CHF (Cooper-Haines filter, Cooper and Haines [21]). Mention that the CHF is a filter widely used in the oceanic data assimilation, which projects the PE of the surface height data by lifting or lowering of water columns.

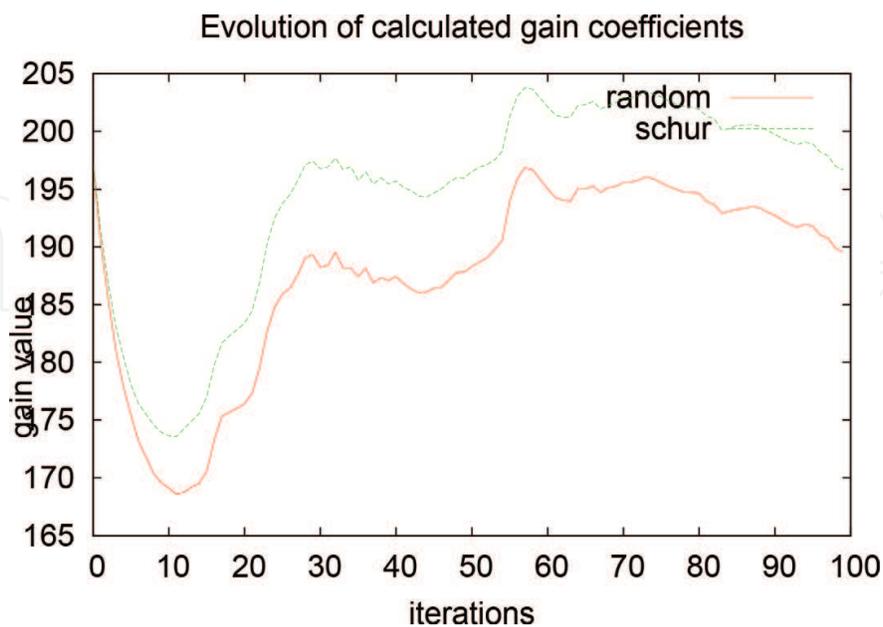


Figure 2. Evolution of estimates for the gain coefficients $k(1)$ computed from \hat{c}_{kl} on the basis of $En(SCH)$ (curve "Schur") and $En(SSP)$ (curve "random"), during model integration. It is seen that two coefficients are evolved in nearly the same manner, of nearly the same magnitude as that of $k(1)$ in the CHF. The same picture is obtained for other c_k , $k = 2, 3, 4$.

The same pictures are obtained for the estimates $\hat{c}_k, k = 2, 3, 4$. Mention that in the CHF, $c_2 = c_3 = 0$.

7.2. Performance of different filters

In **Table 3**, the performances of the three filters are displayed. The errors are the averaged (spatially and temporally) rms of PE for the SSH and for the two velocity components u and v .

The results in **Table 3** show that two filters PEF (SCH) and PEF (SSP) are practically of the same performance, and their estimates are much better compared to those of the CHF, with a slightly better performance for the PEF (SSP). We note that as the PEF (SCH) is constructed on the basis of an ensemble of samples tending to the first dominant SchV, its performance must be theoretically better than that of the PEF (SSP). The slightly better performance of PEF (SSP) (compared to that of PEF (SCH)) may be explained by the fact that the best theoretical performance of PEF (SCH) can be obtained only if the model is linear, stationary, and the number of PE samples in $En(SCH)$ at each iteration must be large enough. The ensemble size of $En(SCH)$ in the present experiment is too small compared with the dimension of the MICOM model.

rms	CHF (cm)	PEF (SCH) (cm/s)	PEF (RAN) (cm/s)
ssh(fcst)	7.39	5.09	4.95
u(fcst)	7.59	5.36	5.29
v(fcst)	7.72	5.73	5.64

Table 3. rms of PE for *ssh*, and u, v velocity components.

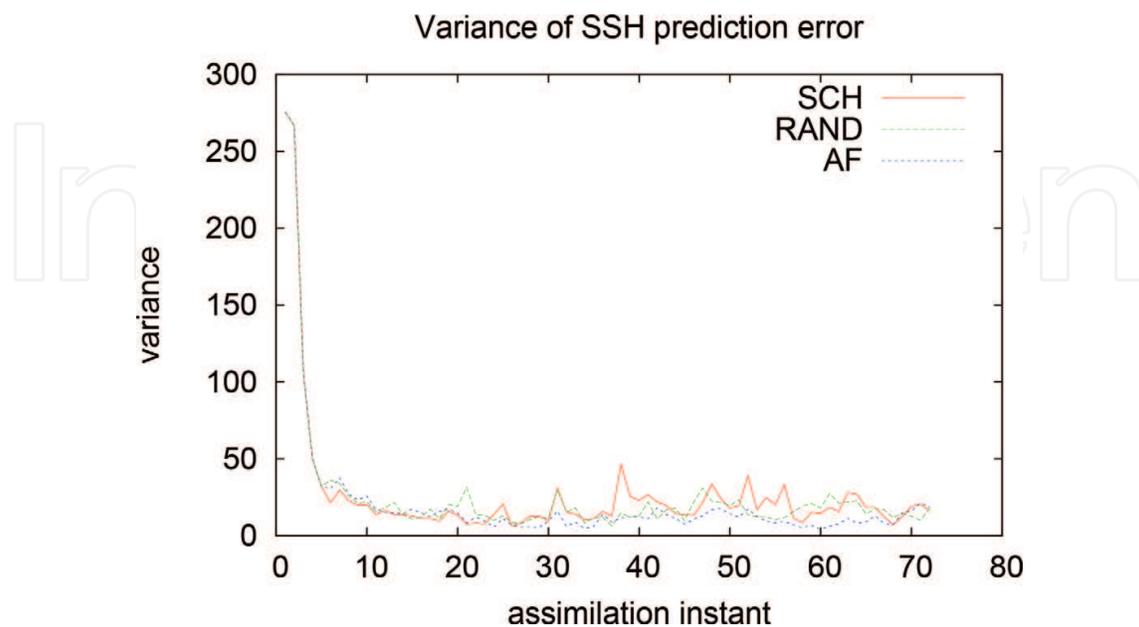


Figure 3. Variance of PE resulting from three filters PEF (SCH), PEF (SSP), and AF. It is seen that the AF yields much better performance compared to the two other nonadaptive filters PEF (SCH) and PEF (SSP).

To illustrate the efficiency of adaptation, in **Figure 3**, we show the cost functions (variances of innovation) resulting from the three filters PEF (SCH), PEF (SSP), and AF (i.e., APEF based on PEF (SCH); the same performance is observed for the AF based on PEF (SSP)). Undoubtedly, the adaptation allows to improve considerably the performances of nonadaptive filters.

8. Conclusion remarks

We have presented in this chapter the different types of OPs—deterministic, stochastic, or optimal, the invariant subspaces of the system dynamics. The ODPs and OSPs play an important role in the study on the predictability of the system dynamics as well as in construction of optimal OFS for environmental geophysical systems.

One other class of perturbation known as SSP is found to be a very efficient tool for solving optimization and estimation problems, especially with Hd matrices and in computing the optimal perturbations.

The numerical experiments presented in this chapter confirm the important role of the different types of OPs in the numerical study of Hd assimilation systems.

Author details

Hong Son Hoang* and Remy Baraille

*Address all correspondence to: hhoang@shom.fr

SHOM/HOM/REC, Toulouse, France

References

- [1] Myhrvold N. Moore's Law Corollary: Pixel Power. New York Times. June 7, 2006 [Retrieved: 2011-11-27]
- [2] Kalman RE. A new approach to linear filtering and prediction problems. Transactions of the ASME-Journal of Basic Engineering. 1960;**82**:35-45
- [3] Lorenz EN. Deterministic non-periodic flow. Journal of the Atmospheric Sciences. 1963;**20**: 130-141
- [4] Lorenz EN. Atmospheric predictability experiments with a large numerical model. Tellus. 1982;**34**:505-513
- [5] Palmer TN, Barkmeijer J, Buizza R, Petroliagis T. The ECMWF ensemble prediction system. Quarterly Journal of the Royal Meteorological Society. 1997;**4**(4):301-304

- [6] Toth Z, Kalnay E. Ensemble forecasting at NMC: The generation of perturbations. *Bulletin of the American Meteorological Society*. 1993;**74**:2317-2330
- [7] Hoang HS, De Mey P, Talagrand O, Baraille R. A new reduced-order adaptive filter for state estimation in high dimensional systems. *Automatica*. 1997;**33**:1475-1498
- [8] Hoang HS, Talagrand O, Baraille R. On the design of a stable adaptive filter for high dimensional systems. *Automatica*. 2001;**37**:341-359
- [9] Golub GH, Van Loan CF. *Matrix Computations*. 2nd ed. Baltimore-London: Johns Hopkin Press, 1993
- [10] Hoang HS, Baraille R, Talagrand O. On the stability of a reduced-order filter based on dominant singular value decomposition of the systems dynamics. *Automatica*. 2009; **45**(10):2400-2405. DOI: 10.1016/j.automatica.2009.06.032
- [11] Doucet A, Johansen AM. A tutorial on particle filtering and smoothing: fifteen years later (PDF). Technical report. Department of Statistics, University of British Columbia; 2008
- [12] Evensen G. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*. 2003;**53**:343-367
- [13] Kalnay E. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press; 2002
- [14] Lanczos C. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *Journal of Research of the National Bureau of Standards*. 1950;**45**:255-282
- [15] Hoang HS, Baraille R. Prediction error sampling procedure based on dominant Schur decomposition. Application to state estimation in high dimensional oceanic model. *Applied Mathematics and Computation*. 2011;**218**(7):3689-3709. DOI: 10.1016/j.amc.2011.09.012
- [16] Spall JC. Accelerated second-order stochastic optimization using only function measurements. In: *Proceedings of 36th IEEE Conference on Decision and Control*; 1997. pp. 1417-1424
- [17] Hoang HS, Baraille R. Stochastic simultaneous perturbation as powerful method for state and parameter estimation in high dimensional systems. In: Baswell AR, editor. *Advances in Mathematics Research*. Vol. 20. New-York: Nova Science Publishers; 2015. pp. 117-148
- [18] Hoang HS, Baraille R. On the efficient low cost procedure for estimation of high-dimensional prediction error covariance matrices. *Automatica*. 2017;**83**:317-330
- [19] Hoang HS, Baraille R. A simple numerical method based simultaneous stochastic perturbation for estimation of high dimensional matrices. *Multidimensional Systems and Signal Processing*. 2018. DOI: 10.1007/s11045-018-0551-y
- [20] Daley R. The effect of serially correlated observation and model error on atmospheric data assimilation. *Monthly Weather Review*. 1992;**120**:165-177
- [21] Cooper M, Haines K. Altimetric assimilation with water property conservation. *Journal of Geophysical Research*. 1996;**101**:1059-1077