Adaptive Observations And Multilevel Optimization
In Data Assimilation

by S. Gratton, M. M. Rincon-Camacho and Ph. L. Toint

Report NAXYS-06-2013   21 May 2013

naxys
Namur Center for Complex   Systems

ENSEEIHT, 2, rue Camichel, 31000 Toulouse, France

CERFACS, 42, avenue Gaspard Coriolis, 31057 Toulouse, France

University of Namur, 61, rue de Bruxelles, B5000 Namur (Belgium)

http://www.unamur.be/sciences/naxys

Adaptive Observations And Multilevel Optimization In Data Assimilation

Serge Gratton,[*] M. Monserrat Rincon-Camacho,[†] and Philippe L. Toint[‡]

June 17, 2013

Abstract

We propose to use a decomposition of large-scale incremental four dimensional (4D-Var) data assimilation problems in order to make their numerical solution more efficient. This decomposition is based on exploiting an adaptive hierarchy of the observations. Starting with a low-cardinality set and the solution of its corresponding optimization problem, observations are adaptively added based on a posteriori error estimates. The particular structure of the sequence of associated linear systems allows the use of a variant of the conjugate gradient algorithm which effectively exploits the fact that the number of observations is smaller than the size of the vector state in the 4D-Var model. The method proposed is justified by deriving the relevant error estimates at different levels of the hierarchy and a practical computational technique is then derived. The new algorithm is tested on a 1D-wave equation and on the Lorenz-96 system, the latter one being of special interest because of its similarity with Numerical Weather Prediction (NWP) systems.

Keywords: Data assimilation, adaptive observations, numerical algorithms, multilevel optimization, a posteriori errors.

1 Introduction

In data assimilation, a substantial body of research has been conducted in the field of adaptive observations. The main idea in this area is to adapt the set of observations used in the model calibration to improve computational efficiency while not sacrificing accuracy. Several techniques have been considered, depending on the practical context of the relevant application. In a Kalman filter framework, Leutbecher (2003b) introduces a Hessian reduced rank estimate whose aim is to predict expected change in the square norm of forecast error due to intermittent modifications of the observation network; its application to the Lorenz96 model (see Lorenz, 1995) is presented in Leutbecher (2003a). Lorenz and Emanuel (1998) present different strategies for weather forecasting which target the locations where the first-guess errors are largest, and verify if the including additional data improve the analyses and forecasts. Berliner, Lu and Snyder (1999) introduce rigorous mathematical criteria to select future data in a dynamical process, based on the uncertainty of the current analysis, the dynamics of error evolution, the form and errors of observations and data assimilation. In Hansen and Smith (2000) the performance of various observation targeting strategies is compared for both direct insertion and ensemble Kalman filter assimilation. The authors conclude that the optimal adaptive observation strategies depend on the combination of model, assimilation scheme and observational network. Daescu and Navon (2004) describe a method based on a periodic update of the adjoint sensitivity field which takes into account the

---

[*]ENSEEIHT, 2, rue Camichel, 31000 Toulouse, France and CERFACS, 42, avenue Gaspard Coriolis, 31057 Toulouse, France. (serge.gratton@enseeiht.fr).

[†]CERFACS, 42, avenue Gaspard Coriolis, 31057 Toulouse, France. (monserratrc@cerfacs.fr).

[‡]Namur Center for Complex Systems, University of Namur, 61, rue de Bruxelles, B-5000 Namur, Belgium. (philippe.toint@unamur.be).

interaction between time-distributed adaptive and routine observations. Cardinali, Pezzulli and Andersson (2004) introduce an influence matrix whose purpose is to identify influential data in a 4D-Var framework . They observe that low-influence data occur in areas with a large number of observations while high-influence data points occur in regions with sparse ones or in dynamically active areas. Finally, Liu, Kalnay, Miyoshi and Cardinali (2009) describe, in a Kalman filter setting, the impact of the self-sensitivity and the cross sensitivity (off-diagonal elements of influence matrix).

Our motivation for the present paper lies in several of the conclusions drawn in this literature. In particular, our work aims at a good numerical exploitation of the conclusion that the spatial distribution of observations may be influential. More specifically, we introduce new techniques for adaptively selecting observations in a 4D-Var setting, based on mathematically sound a posteriori error estimates within a dynamic hierarchy. The technique used for constructing the adaptive set of observations is inspired by adaptive finite elements techniques studied in Section 12.2 (p. 100) of Rincon-Camacho (2011). Related techniques for adaptive multigrid can be found in Brandt (1973) and McCormick (1984).

The paper is structured as follows: Section 2 introduces the 4D-Var vocabulary and the considered hierarchy of observations. Section 3 then covers the associated error estimates and Section 4 presents the new adaptive algorithm. Section 5 discusses its application to a one-dimensional nonlinear wave equation and to the Lorenz96 model. Conclusions and perspectives are discussed in Section 6.

2   The problem and associated observation hierarchy

Consider the nonlinear least-squares problems in 4D-Var data assimilation, whose objective is to find an initial vector state at an initial time denoted as $x = x(t_0) \in \mathbb{R}^n$. The structure of the problem is as follows:

$$\min_{x \in \mathbb{R}^n} \tfrac{1}{2} \|x - x_b\|_{B^{-1}}^2 + \tfrac{1}{2} \sum_{j=0}^{N_t} \|\mathcal{H}_j(x(t_j)) - y_j\|_{R_j^{-1}}^2 \tag{2.1}$$

where the squared norm $\|x\|_M^2$ is induced by the inner product $x^T M x$ for a symmetric positive definite matrix $M \in \mathbb{R}^{l \times l}$ and a vector $x \in \mathbb{R}^l$. Here $x_b \in \mathbb{R}^n$ is the background vector, which is an a priori estimate. The vector $y_j \in \mathbb{R}^{m_j}$ is the vector of observations at time t_j and $\mathcal{H}_j$ is the operator modeling the observation process at the same time. The state vector $x(t_j)$ satisfies the nonlinear model of evolution $x(t_j) = \mathcal{M}_{0 \to j}[x(t_0)]$. The matrix $B \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix representing the background-error covariance and the matrix $R_j \in \mathbb{R}^{m_j \times m_j}$ is also a symmetric positive definite matrix representing the observation-error covariance at time t_j.

The nonlinear least-squares problem (2.1) is solved iteratively, and, at iteration k, the problem is simplified by linearizing the nonlinear observation operator $\mathcal{H}_j(x(t_j))$ at the iterate x_k (which for the moment we denote only by x), leading to the optimization problem

$$\min_{\delta x \in \mathbb{R}^n} \tfrac{1}{2} \|x - x_b + \delta x\|_{B^{-1}}^2 + \tfrac{1}{2} \|H\,\delta x - d\|_{R^{-1}}^2 \tag{2.2}$$

where $R = \mathrm{diag}(R_0, R_1, \ldots, R_{N_t}) \in \mathbb{R}^{m \times m}$, $d \in \mathbb{R}^m$ denotes the concatenated misfits over time $y_j - \mathcal{H}_j(x(t_j))$ and $H \in \mathbb{R}^{m \times n}$ is the concatenated version of the linearized observation process

$$d = \begin{bmatrix} d_0 \\ d_1 \\ \vdots \\ d_{N_t} \end{bmatrix} \in \mathbb{R}^m, \;\; H = \begin{bmatrix} H_0 \\ H_1 M_{0 \to 1} \\ \vdots \\ H_{N_t} M_{0 \to N_t} \end{bmatrix} \in \mathbb{R}^{n \times m}.$$

Here H_j and $M_{0 \to j}$ are a (possibly approximate) linearization of the observation operator $\mathcal{H}_j$ and the model $\mathcal{M}_{0 \to j}$ around $x = x_k(t_0)$ respectively.

Diverse techniques for solving problem (2.1) have been proposed, see for instance Courtier, Thépaut and Hollingsworth (1994) for the so called incremental method which is equivalent to applying a truncated Gauss-Newton iteration to problem (2.1) (see Gratton, Lawless and Nichols, 2007). The general algorithm is the following:

Algorithm 2.1 *Incremental 4D-Var*

1. *Initialize $x_0 = x_b \in \mathbb{R}^n$ and set $k = 0$.*

2. *Compute $x_k(t_j)$ from $x_k(t_0) = x_k$ by running the nonlinear model $\mathcal{M}$ from time t_0 to t_{N_t}.*

3. *Calculate the vectors $d_{k,j} = y_j - \mathcal{H}_j(x_k(t_j))$ for $j = 1, \ldots, N_t$.*

4. *Find an approximate solution δx_k of the minimization problem (2.2), where $x = x_k$, $H = H_k$ and $d = d_k$.*

5. *Update the current solution as $x_{k+1} = x_k + \delta x_k$.*

6. *Set $k := k + 1$. If convergence is not achieved return to Step 2.*

This algorithm is known as the outer loop and the minimization step (Step 4) as the inner loop. Our interest is to reduce the cost of this optimization problem. Termination criteria for the outer and inner loops of this algorithm are discussed in Gratton et al. (2007).

It is most natural to solve the subproblem in Step 4 of Algorithm 2.1 directly in the space of dimension n, the size of x. This technique is referred to as the *primal approach*. At variance, the quadratic optimization problem (2.2) may be rewritten in a space of dimension m, the number of observations. This is known as the *dual approach* and it is especially useful when m is significantly smaller than n, in which case the second term in (2.2) is of relatively low rank compared to the first. A first approach of this type is the Physical-space Statistical Analysis System (PSAS) method (see Courtier, 1997), where the "low rank" observation term in (2.2) is handled by using a Sherman-Morrison formula and applying the standard conjugate-gradient algorithm (see Hestenes and Stiefel, 1952, or Golub and Van Loan, 1996, Section 10.2, p. 520) to the reduced system. This method has the drawback of not maintaining the inherent monotonicity of the conjugate-gradient algorithm in $\mathbb{R}^n$, thereby making any stopping rule of the inner minimization difficult to define (see El Akkroui, Gauthier, Pellerin and Buis, 2008, or Gratton, Gürol and Toint, 2013). Fortunately, a better alternative is available, where the conjugate-gradient algorithm is reformulated in $\mathbb{R}^m$ using the inner product defined by the metric HBH^T in order to guarantee the desired monotonicity properties without incurring additional cost. This technique, known as the Restricted Preconditioned Conjugate Gradient method (RPCG, see Gratton and Tshimanga, 2009 and Gratton et al., 2013), provides an efficient numerical procedure where substantial computing gains are obtained when $m \ll n$. We refer the reader to the cited publications for details.

Our aim in this work is to make the best possible use of this dual technique and to propose an adaptive observations' strategy for solving the data assimilation problem (2.1). Suppose that we have a large set of m observations $\mathcal{O}$ which can be decomposed into a hierarchical collection of sets $\{\mathcal{O}_i\}_{i=0}^r$, each with cardinality m_i, such that

$$\mathcal{O}_i \subset \mathcal{O}_{i+1} \quad \text{and} \quad m_i < m_{i+1} \quad (i = 0, \ldots, r-1)$$

where, by convention, $\mathcal{O}_r = \mathcal{O}$ and $m_r = m$. To each set of observations $\mathcal{O}_i$ we associate a misfit vector $y_i \in \mathbb{R}^{m_i}$ (by selecting the relevant components in the vector y), and the corresponding observation-error covariance matrix R_i. Given $\{\mathcal{O}_i\}_{i=0}^r$, we may therefore consider the hierarchical collection of minimization problems

$$\min_{x \in \mathbb{R}^n} \tfrac{1}{2}\|x - x_b\|_{B^{-1}}^2 + \tfrac{1}{2}\|\mathcal{H}_i(x) - y_i\|_{R_i^{-1}}^2, \quad i = 0, \ldots, r, \tag{2.3}$$

where the vector $\mathcal{H}_i(x)$ denotes the nonlinear observation operator concatenated over time.

Thus our objective is to construct the collection $\{\mathcal{O}_i\}_{i=0}^r$ such that the sequential solution of the problems (2.3) is obtained at significantly lower cost compared to solving (2.2) directly, while at the same time maintaining equivalent accuracy requirements.

3   Mathematical analysis

Our adaptive method is based on the exploitation of a posteriori bounds for the error between the solutions obtained using few observations or many. In particular, we are interested in comparing the accuracy of the solutions of problems (2.3) for $\mathcal{O}_i$ and $\mathcal{O}_{i+1}$. For simplicity of notation, we denote them as $\mathcal{O}_c$ a set with m_c observations and $\mathcal{O}_f$ a set containing m_f observations such that $m_c < m_f$ and $\mathcal{O}_c \subset \mathcal{O}_f$, where the indices c and f stand for 'coarse' and 'fine', respectively. The 'fine' optimization problem (2.3) is given, for the starting point x, by

$$\min_{\delta x_f \in \mathbb{R}^n} \tfrac{1}{2}\|x + \delta x_f - x_b\|_{B^{-1}}^2 + \tfrac{1}{2}\|H_f \delta x - d_f\|_{R_f^{-1}}^2 \tag{3.4}$$

where $d_f = y_f - \mathcal{H}_f(x)$ and H_f is the linearized version of the observation process at x involving the set of observations $\mathcal{O}_f$. We may reformulate (3.4) as a convex quadratic problem with linear equality constraints given by

$$\min_{\delta x_f \in \mathbb{R}^n,\, v_f \in \mathbb{R}^{m_f}} \tfrac{1}{2}\|x + \delta x_f - x_b\|_{B^{-1}}^2 + \tfrac{1}{2}\|v_f\|_{R_f^{-1}}^2, \quad \text{subject to } v_f = H_f \delta x_f - d_f, \tag{3.5}$$

see Gratton et al. (2013). The Lagrange function corresponding to this reformulated problem is given by

$$\mathcal{L}(\delta x_f, v_f, \lambda_f) \overset{\text{def}}{=} \tfrac{1}{2}\|x + \delta x_f - x_b\|_{B^{-1}}^2 + \tfrac{1}{2}\|v_f\|_{R_f^{-1}}^2 - \lambda_f^T(H_f \delta x_f - v_f - d_f).$$

Using this function, the optimality conditions for problem (3.5) can then be expressed as

$$\begin{aligned}
\nabla_{\delta x_f}\mathcal{L}(\delta x_f, v_f, \lambda_f) &= B^{-1}(x + \delta x_f - x_b) - H_f^T \lambda_f = 0, \\
\nabla_{v_f}\mathcal{L}(\delta x_f, v_f, \lambda_f) &= R_f^{-1}(v_f) + \lambda_f = 0, \\
\nabla_{\lambda_f}\mathcal{L}(\delta x_f, v_f, \lambda_f) &= v_f - H_f \delta x_f + d_f = 0,
\end{aligned}$$

which leads to the system

$$\begin{aligned}
H_f^T \lambda_f &= B^{-1}(x + \delta x_f - x_b), \\
-\lambda_f &= R_f^{-1}(H_f \delta x_f - d_f).
\end{aligned} \tag{3.6}$$

Gratton et al. (2013) and Gratton and Tshimanga (2009) show that the solution of this system can be obtained by solving

$$(R_f^{-1} H_f B H_f^T + I_{m_f})\lambda_f = R_f^{-1}(d_f - H_f(x_b - x)), \tag{3.7}$$

for λ_f and then substituting in the second equation of (3.6) for δx_f and v_f.

We now compare the solution $(\delta x_f, \lambda_f)$ to the solution of the coarse level minimization problem

$$\min_{\delta x_c \in \mathbb{R}^n} \tfrac{1}{2}\|x + \delta x_c - x_b\|_{B^{-1}}^2 + \tfrac{1}{2}\|\Gamma_f(H_f \delta x_c - d_f)\|_{R_c^{-1}}^2, \tag{3.8}$$

where Γ_f is the restriction operator $\Gamma_f : \mathbb{R}^{m_f} \to \mathbb{R}^{m_c}$ from the fine observation space to the coarse one. As above, we reformulate (3.8) as a convex quadratic problem with linear equality constraints, and obtain

$$\min_{\delta x_c \in \mathbb{R}^n,\, v_c \in \mathbb{R}^{m_c}} \tfrac{1}{2}\|x + \delta x_c - x_b\|_{B^{-1}}^2 + \tfrac{1}{2}\|v_c\|_{R_c^{-1}}^2, \quad \text{subject to } v_c = \Gamma_f(H_f \delta x_c - d_f), \tag{3.9}$$

whose Lagrangian function is given by

$$\mathcal{L}(\delta x_c, v_c, \lambda_c) \overset{\text{def}}{=} \tfrac{1}{2}\|x + \delta x_c - x_b\|_{B^{-1}}^2 + \tfrac{1}{2}\|v_c\|_{R_c^{-1}}^2 - \lambda_c^T(\Gamma_f H_f \delta x_c - v_c - \Gamma_f d_f).$$

The optimality conditions now become

$$\begin{aligned}
H_f^T \Gamma_f^T \lambda_c &= B^{-1}(x + \delta x_c - x_b), \\
-\lambda_c &= R_c^{-1}\Gamma_f(H_f \delta x_c - d_f)
\end{aligned} \tag{3.10}$$

These conditions may again be solved by computing λ_c such that

$$(R_c^{-1}\Gamma_f H_f B H_f^T \Gamma_f^T + I_{m_c})\lambda_c = R_c^{-1}\Gamma_f(d_f - H_f(x_b - x)). \tag{3.11}$$

Since the dimension of the system (3.11) is smaller than that of the system (3.7), solving problem (3.8) is (often much) cheaper than solving problem (3.4). The RPCG algorithm mentioned in the previous section derives its efficiency by using the formulation (3.11) rather than (3.7), see Gratton et al. (2013) and Gratton and Tshimanga (2009).

After obtaining the solution $(\delta x_c, \lambda_c)$, our objective is now to compute the difference between the (still unknown) λ_f and $\Pi_c \lambda_c$ in order to identify the set of observations where this difference is large. Here Π_c is the prolongation operator given by

$$\Pi_c \stackrel{\text{def}}{=} \sigma_f \Gamma_f^T \tag{3.12}$$

for some $\sigma_f > 0$. Our intention is then to construct the "fine" problem from the "coarse" one by adding to the coarse the observations that are singled out by this comparison.

We start by computing the difference between λ_f and $\Pi_c\lambda_c$ in an appropriate norm. Using (3.6) and (3.10), we define

$$
\begin{aligned}
E_1 &\stackrel{\text{def}}{=} \|\lambda_f - \Pi_c\lambda_c\|_{R_f}^2 \\
&= \langle R_f(\lambda_f - \Pi_c\lambda_c), -R_f^{-1}(H_f \delta x_f - d_f) - \Pi_c\lambda_c\rangle \\
&= \langle \lambda_f - \Pi_c\lambda_c, -H_f \delta x_f + d_f - R_f \Pi_c\lambda_c\rangle,
\end{aligned}
$$

and

$$
\begin{aligned}
E_2 &\stackrel{\text{def}}{=} \|\lambda_f - \Pi_c\lambda_c\|_{H_f B H_f^T}^2 \\
&= \|H_f^T \lambda_f - H_f^T \Pi_c\lambda_c\|_B^2 \\
&= \langle B(H_f^T \lambda_f - H_f^T \Pi_c\lambda_c), B^{-1}(x + \delta x_f - x_b) - H_c^T \Pi_c\lambda_c\rangle \\
&= \langle \lambda_f - \Pi_c\lambda_c, H_f(x - x_b) + H_f \delta x_f - H_f B H_f^T \Pi_c\lambda_c\rangle.
\end{aligned}
$$

By adding errors E_1 and E_2 we obtain that

$$
\begin{aligned}
E_1 + E_2 &= \langle \lambda_f - \Pi_c\lambda_c, -H_f \delta x_f + d_f - R_f \Pi_c\lambda_c\rangle \\
&\quad + \langle \lambda_f - \Pi_c\lambda_c, H_f(x - x_b) + H_f \delta x_f - H_f B H_f^T \Pi_c\lambda_c\rangle \\
&\quad + \langle \lambda_f - \Pi_c\lambda_c, H_f \delta x_c - H_f \delta x_c\rangle \\
&= \langle \lambda_f - \Pi_c\lambda_c, d_f - R_f \Pi_c\lambda_c - H_f \delta x_c\rangle \\
&\quad + \langle \lambda_f - \Pi_c\lambda_c, H_f(x - x_b + \delta x_c) - H_f B H_f^T \Pi_c\lambda_c\rangle.
\end{aligned}
$$

As a consequence, we deduce that

$$
\begin{aligned}
E_1 + E_2 &= \langle R_f^{1/2}(\lambda_f - \Pi_c\lambda_c), R_f^{-1/2}(d_f - H_f \delta x_c - R_f \Pi_c\lambda_c)\rangle \\
&\quad + \langle H_f B H_f^T(\lambda_f - \Pi_c\lambda_c), (H_f B H_f^T)^{-1}(H_f(x - x_b + \delta x_c) - H_f B H_f^T \Pi_c\lambda_c)\rangle \\
&= \langle R_f^{1/2}(\lambda_f - \Pi_c\lambda_c), R_f^{-1/2}(d_f - H_f \delta x_c - R_f \Pi_c\lambda_c)\rangle \\
&\quad + \langle B^{1/2}H_f^T(\lambda_f - \Pi_c\lambda_c), B^{1/2}H_f^T(H_f B H_f^T)^{-1}(H_f(x - x_b + \delta x_c) - H_f B H_f^T \Pi_c\lambda_c)\rangle.
\end{aligned}
$$

Thus, we obtain that

$$
\begin{aligned}
E_1 + E_2 &\leq \|\lambda_f - \Pi_c\lambda_c\|_{R_f} \|d_f - H_f \delta x_c - R_f \Pi_c\lambda_c\|_{R_f^{-1}} \\
&\quad + \|H_f^T \lambda_f - H_f^T \Pi_c\lambda_c\|_B \|H_f^T(H_f B H_f^T)^{-1}(H_f(x - x_b + \delta x_c) - H_f B H_f^T \Pi_c\lambda_c)\|_B \\
&\leq \|\lambda_f - \Pi_c\lambda_c\|_{R_f} \|d_f - H_f \delta x_c - R_f \Pi_c\lambda_c\|_{R_f^{-1}} \\
&\quad + \|\lambda_f - \Pi_c\lambda_c\|_{H_f B H_f^T} \|H_f(x - x_b + \delta x_c) - H_f B H_f^T \Pi_c\lambda_c\|_{(H_f B H_f^T)^{-1}}.
\end{aligned}
$$

The Cauchy-Schwarz inequality then yields that

$$
\begin{aligned}
E_1 + E_2 \quad &\le \quad \left[\|\lambda_f - \Pi_c\lambda_c\|_{R_f}^2 + \|\lambda_f - \Pi_c\lambda_c\|_{H_f B H_f^T}^2 \right]^{1/2} \times \\
&\qquad \left[\|d_f - H_f\delta x_c - R_f\Pi_c\lambda_c\|_{R_f^{-1}}^2 + \|H_f(x - x_b + \delta x_c) - H_f B H_f^T\Pi_c\lambda_c\|_{(H_f B H_f^T)^{-1}}^2 \right]^{1/2} \\
&= \quad [\, E_1 + E_2 \,]^{1/2} \times \\
&\qquad \left[\|d_f - H_f\delta x_c - R_f\Pi_c\lambda_c\|_{R_f^{-1}}^2 + \|H_f(x - x_b + \delta x_c) - H_f B H_f^T\Pi_c\lambda_c\|_{(H_f B H_f^T)^{-1}}^2 \right]^{1/2} .
\end{aligned}
$$

This inequality leads to the a posteriori error bound

$$
E_1 + E_2 \le \|d_f - H_f\delta x_c - R_f\Pi_c\lambda_c\|_{R_f^{-1}}^2 + \|H_f(x - x_b + \delta x_c) - H_f B H_f^T\Pi_c\lambda_c\|_{(H_f B H_f^T)^{-1}}^2. \tag{3.13}
$$

The question then arises of how to compute the second term in the right hand side of this inequality, i.e.

$$
\begin{aligned}
&\|H_f(x - x_b + \delta x_c) - H_f B H_f^T\Pi_c\lambda_c\|_{(H_f B H_f^T)^{-1}}^2 \\
&\quad = \langle H_f(x - x_b + \delta x_c) - H_f B H_f^T\Pi_c\lambda_c, (H_f B H_f^T)^{-1}\left(H_f(x - x_b + \delta x_c) - H_f B H_f^T\Pi_c\lambda_c \right)\rangle \\
&\quad = \langle H_f(x - x_b + \delta x_c) - H_f B H_f^T\Pi_c\lambda_c, (H_f B H_f^T)^{-1}H_f(x - x_b + \delta x_c) - \Pi_c\lambda_c\rangle.
\end{aligned} \tag{3.14}
$$

But the first equation of (3.10) multiplied by $H_f B$ gives that

$$
H_f B H_f^T\Gamma_f^T\lambda_c = H_f(x - x_b + \delta x_c), \tag{3.15}
$$

and hence that

$$
\Gamma_f^T\lambda_c = (H_f B H_f^T)^{-1}H_f(x - x_b + \delta x_c). \tag{3.16}
$$

Therefore

$$
\begin{aligned}
&\|H_f(x - x_b + \delta x_c) - H_f B H_f^T\Pi_c\lambda_c\|_{(H_f B H_f^T)^{-1}}^2 \\
&\quad = \langle H_f(x - x_b + \delta x_c) - H_f B H_f^T\Pi_c\lambda_c, \Gamma_f^T\lambda_c - \Pi_c\lambda_c\rangle \\
&\quad = (\tfrac{1}{\sigma_f} - 1)\langle H_f(x - x_b + \delta x_c) - H_f B H_f^T\Pi_c\lambda_c, \Pi_c\lambda_c\rangle,
\end{aligned} \tag{3.17}
$$

because of (3.12). We note also that

$$
\|\lambda_f - \Pi_c\lambda_c\|_{R_f + H_f B H_f^T}^2 = \|\lambda_f - \Pi_c\lambda_c\|_{R_f}^2 + \|\lambda_f - \Pi_c\lambda_c\|_{H_f B H_f^T}^2 = E_1 + E_2. \tag{3.18}
$$

As a consequence, the left-hand side of this inequality (the sought error estimate) can be bounded above using inequality (3.13), giving the following a posteriori error.

Theorem 3.1 *Let δx_f be the solution to the problem (3.5) and λ_f the corresponding Lagrange multiplier to the constraint in (3.5) such that the couple $(\delta x_f, \lambda_f)$ satisfies (3.6). Analogously, let δx_c be the solution to (3.9) and λ_c the corresponding Lagrange multiplier such that $(\delta x_c, \lambda_c)$ satisfies (3.10). Then the a posteriori error bound satisfies the inequality*

$$
\begin{aligned}
\|\lambda_f - \Pi_c\lambda_c\|_{R_f + H_f B H_f^T}^2 \quad &\le \quad \|d_f - H_f\delta x_c - R_f\Pi_c\lambda_c\|_{R_f^{-1}}^2 \\
&\quad + (\tfrac{1}{\sigma_f} - 1)\langle H_f(x - x_b + \delta x_c) - H_f B H_f^T\Pi_c\lambda_c, \Pi_c\lambda_c\rangle.
\end{aligned} \tag{3.19}
$$

Note that the bound (3.19) does not involve the computation of λ_f or δx_f.

The use of an a posteriori error bound depending on a primal-dual problem in adaptive finite elements was studied in Rincon-Camacho (2011).

In the next section we describe how to make use of the bound (3.19) in order to algorithmically construct the 'fine' problem from the 'coarse' one.

4 A new adaptive algorithm

Starting from a small set of observations $\mathcal{O}_c$, our goal is to add only significant observations to produce $\mathcal{O}_f$ so that the a posteriori error (3.19) is reduced. Our strategy is to define an auxiliary set of *potential* fine observations $\tilde{\mathcal{O}}_f$ from which the observations in $\mathcal{O}_f$ are selected. However, describing our strategy (and algorithm) requires additional assumptions on the hierarchy of (potential) observations. More specifically, we complete our assumptions as follows.

- The observations correspond to localizations in some underlying continuous measurable "observation space".

- The coarse observation set partitions the observation space in a finite number of *cells* $\{c_j\}_{j=1}^{p_c}$ of measures $\{w_j\}_{j=1}^{p_c}$.

- The auxiliary set $\tilde{\mathcal{O}}_f$ is constructed by considering all observations in $\mathcal{O}_c$ with the addition of a single additional potential observation point in the interior of each cell. The cell is said to be associated with this additional potential observation.

- There exists a prolongation operator $\tilde{\Pi}_c$ from $\mathcal{O}_c$ to $\tilde{\mathcal{O}}_f$ such that, for each potential observation o_j in $\tilde{\mathcal{O}}_f \setminus \mathcal{O}_c$, $\tilde{\Pi}_c$ defines the value of this observation only in terms of the observations of the associated cell c_j. As expected, we define $\tilde{\Pi}_c = \sigma_f \tilde{\Gamma}_f^T$.

We illustrate these assumptions by an example: suppose that the observation space is the plane and the coarse observation set $\mathcal{O}_c$ is the rectangular 'grid' shown in Figure 4.1 (a): the cells are elementary rectangles in this grid, whose measure is given by their surface. We may then define $\tilde{\mathcal{O}}_f$ as the grid shown in Figure 4.1 (b), which we obtained by locally adding a new potential observation in the center of each rectangle and four additional ones on its boundary (effectively doubling the mesh in every direction). The observations in $\mathcal{O}_f$ (as shown in Figure 4.1 (c)) can then be extracted from $\tilde{\mathcal{O}}_f$.

(a) Coarse observation set $\mathcal{O}_c$ (b) Auxiliary observation set $\tilde{\mathcal{O}}_f$
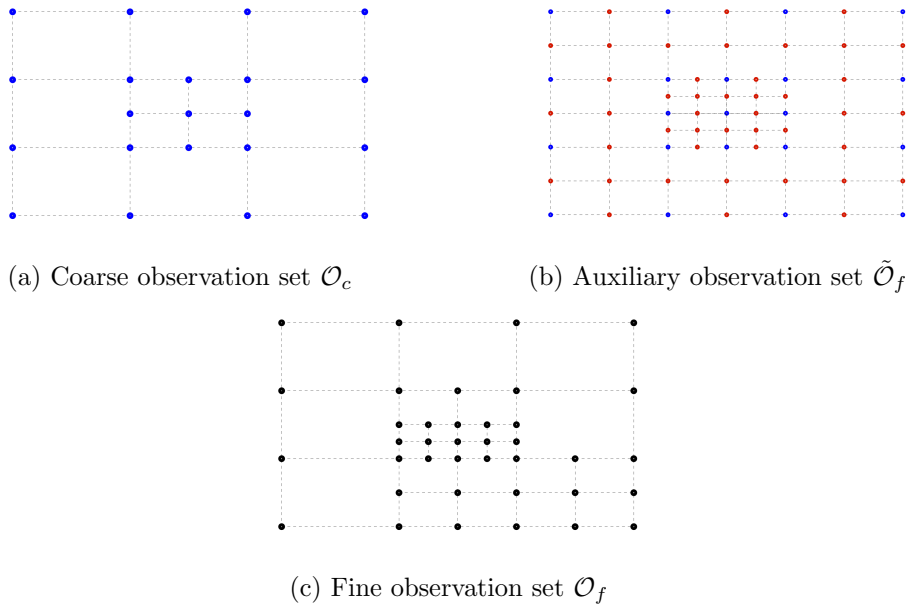
(c) Fine observation set $\mathcal{O}_f$

Figure 4.1: Auxiliary observations

In this example, the restriction operator $\tilde{\Gamma}_f$ can be defined as the usual full weighting operator associated with bilinear interpolation prolongations (which justifies the introduction of the

boundary points). This full-weighting restriction operator is given, on every grid node, by the stencil

$$\frac{1}{16} \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \tag{4.20}$$

In order to achieve our goal to select 'important' observations from $\tilde{\mathcal{O}}_f$, we need to compute localized error indicators (3.19). We define them as

$$\eta_j \stackrel{\text{def}}{=} w_j \langle (\tilde{d}_f - \tilde{H}_f \delta x_c - \tilde{R}_f \tilde{\Pi}_c \lambda_c)|_j, \tilde{R}_f^{-1}(\tilde{d}_f - \tilde{H}_f \delta x_c - \tilde{R}_f \tilde{\Pi}_c \lambda_c)|_j \rangle,$$

$$\nu_j \stackrel{\text{def}}{=} (\tfrac{1}{\sigma_f} - 1) w_j \langle (\tilde{H}_f(x - x_b + \delta x_c) - \tilde{H}_f B \tilde{H}_f^T \tilde{\Pi}_c \lambda_c)|_j, \tilde{\Pi}_c \lambda_c|_j \rangle,$$

where $\tilde{d}_f, \tilde{H}_f, \tilde{R}_f$ are constructed from the set of observations $\tilde{\mathcal{O}}_f$ and where the symbol $|_j$ denotes the restriction of the associated quantity to the cell c_j. Note that the η's correspond to the first term on the right-hand side of (3.19) while the ν's correspond the the second term.

In order to decide which cells will be chosen to include a new interior potential observation, we use the bulk criterion (also known as Dörfler marking, see, for instance, Dörfler, 1996, Morin, Nochetto and Siebert, 2000, Logg, Mardal and Wells, 2012). For some constants $\theta_1, \theta_2 \in (0, 1)$, we construct the sets $\mathcal{S}_\eta$ and $\mathcal{S}_\nu$ such that

$$\theta_1 \left(\sum_{j=1}^p \eta_j \right) \leq \sum_{k \in \mathcal{S}_\eta} \eta_k \qquad \text{and} \qquad \theta_2 \left(\sum_{j=1}^p \nu_j \right) \leq \sum_{k \in \mathcal{S}_\nu} \nu_k. \tag{4.21}$$

In practice this construction is carried out by progressively constructing each set using a greedy heuristic which includes first the non-included cell with maximal indicator value.

Once $\mathcal{S}_\eta$ and $\mathcal{S}_\nu$ are constructed, we decide that a cell k of $\mathcal{O}_c$ is 'refined' if $k \in \mathcal{S}_\eta$ or $k \in \mathcal{S}_\nu$, meaning that the observations associated with the corresponding cell in $\tilde{\mathcal{O}}_f$ are added to the set $\mathcal{O}_c$ to construct the new set O_f. More formally,

$$\mathcal{O}_f \stackrel{\text{def}}{=} \mathcal{O}_c \cup \left(\bigcup_{k \in \mathcal{S}_\eta \cup \mathcal{S}_\nu} o_k \right). \tag{4.22}$$

Thus, starting with a small set of observations $\mathcal{O}_0$, we progressively add observations using the method just described, resulting in the following algorithm.

Algorithm 4.1 *An algorithm using adaptive observations*

1. *Set $i = 0$, initialize x and the coarse observation set $\mathcal{O}_0$.*

2. *Given the set of observations $\mathcal{O}_i$, construct the auxiliary set $\tilde{\mathcal{O}}_{i+1}$ such that the conditions described at the beginning of this section hold.*

3. *Find the solution $(\delta x_i, \lambda_i)$ to the problem*

$$\min_{\delta x_i \in \mathbb{R}^n} \tfrac{1}{2} \|x_i + \delta x_i - x_b\|_{B^{-1}}^2 + \tfrac{1}{2} \|\tilde{\Gamma}_{i+1}(\tilde{H}_{i+1} \delta x_i - \tilde{d}_{i+1})\|_{R_i^{-1}}^2, \tag{4.23}$$

   by approximately solving the system

$$(R_i^{-1} \tilde{\Gamma}_i \tilde{H}_{i+1} B \tilde{H}_{i+1}^T \tilde{\Gamma}_{i+1}^T + I_{m_i}) \lambda_i = R_i^{-1} \tilde{\Gamma}_{i+1} (\tilde{d}_{i+1} - \tilde{H}_{i+1}(x_b - x_i))$$

   using RPCG and then setting $\delta x_i = x_b - x_i + B \tilde{H}_{i+1}^T \tilde{\Gamma}_i^T \lambda_i$.

4. *For each cell c_j of observation set $\mathcal{O}_i$ compute the error indicators*

$$\eta_j = w_j \langle (\tilde{d}_{i+1} - \tilde{H}_{i+1} \delta x_i - \tilde{R}_{i+1} \tilde{\Pi}_i \lambda_i)|_j, \tilde{R}_{i+1}^{-1}(\tilde{d}_{i+1} - \tilde{H}_{i+1} \delta x_i - \tilde{R}_{i+1} \tilde{\Pi}_i \lambda_i)|_j \rangle,$$

$$\nu_j = (\tfrac{1}{\sigma_{i+1}} - 1) w_j \langle (\tilde{H}_{i+1}(x_i - x_b + \delta x_i) - \tilde{H}_{i+1} B \tilde{H}_{i+1}^T \tilde{\Pi}_i \lambda_i)|_j, \tilde{\Pi}_i \lambda_i|_j \rangle.$$

5. *Build the sets $\mathcal{S}_\eta$ and $\mathcal{S}_\nu$ such that*

$$\theta_1 \left(\sum_{j=1}^{p_i} \eta_j \right) \leq \sum_{k \in \mathcal{S}_\eta} \eta_k, \qquad and \qquad \theta_2 \left(\sum_{j=1}^{p_i} \nu_j \right) \leq \sum_{k \in \mathcal{S}_\nu} \eta_k.$$

using the bulk criterion.

6. *Construct the set O_{i+1} as*

$$O_{i+1} := O_i \cup \left(\bigcup_{k \in \mathcal{S}_\eta \cup \mathcal{S}_\nu} o_k \right)$$

.

7. *Update $x \leftarrow x + \delta x_i$.*

8. *Increment i and return to Step 2.*

We note that the computation of $(\delta x_i, \lambda_i)$ in Step 3 corresponds to applying the RCPG algorithm to (4.23), thereby making this computation essentially dependent on m, the number of observations, which the algorithm maintains as small as necessary by design. We also note that the bulk of the remaining work is in the computation of the error indicators in Step 4, as this requires the product of $\tilde{H}_{i+1}$ with δx_i. Observe that, because of (3.12), $\tilde{H}_{i+1}^T \tilde{\Pi}_i \lambda_i = \sigma_{i+1} \tilde{H}_{i+1}^T \tilde{\Gamma}_{i+1}^T \lambda_i$ and that both this quantity and the product $\tilde{H}_{i+1}(x_i - x_b)$ are already available as by-products of Step 3.

Similar methods for constructing adaptive grids are the multi-level adaptive technique (MLAT) studied in Brandt (1973) and the fast adaptive composite grid (FAC) presented in McCormick (1984).

5   Numerical experiments

5.1   Two test problems

This section is devoted to showing the performance of our new adaptive algorithm on two test cases. The first one is a one-dimensional wave equation system, which we refer to as the 1D-Wave model from now on. The dynamics on this model are governed by the following nonlinear wave equations:

$$\begin{aligned}
&\frac{\partial^2}{\partial t^2} u(z,t) - \frac{\partial^2}{\partial z^2} u(z,t) + f(u) = 0, \\
&u(0,t) = u(1,t) = 0, \\
&u(z,0) = u_0(z), \ \frac{\partial}{\partial t} u(z,0) = 0, \\
&0 \leq t \leq T, \ \ 0 \leq z \leq 1,
\end{aligned} \qquad (5.24)$$

where we have chosen $f(u) = \mu e^{\eta u}$. In this case we look for the initial function $u_0(z)$, which corresponds to x in the data assimilation problem (2.1). We illustrate in Figure 5.2 (a) the initial state vector u_0 ($x = u_0$) and the evolution of the system in Figure 5.2 (b) (view from the top) where the space domain corresponds to the horizontal axis and the time domain to the vertical axis.

Our second example is the model referred as Lorenz96 presented in Lorenz (1995). The variable $\bar{u}$ is a vector of N-equally spaced entries around a circle of constant latitude, i.e. $\bar{u}(t) = (u_1(t), u_2(t), \ldots, u_N(t))$. The N-dimensional system is determined by the following N equations

$$\frac{du_j}{dt} = \frac{1}{\kappa}(-u_{j-2}u_{j-1} + u_{j-1}u_{j+1} - u_j + F), \ \ j = 1, \ldots N, \qquad (5.25)$$

(a) Initial $x_0 := u_0(z)$
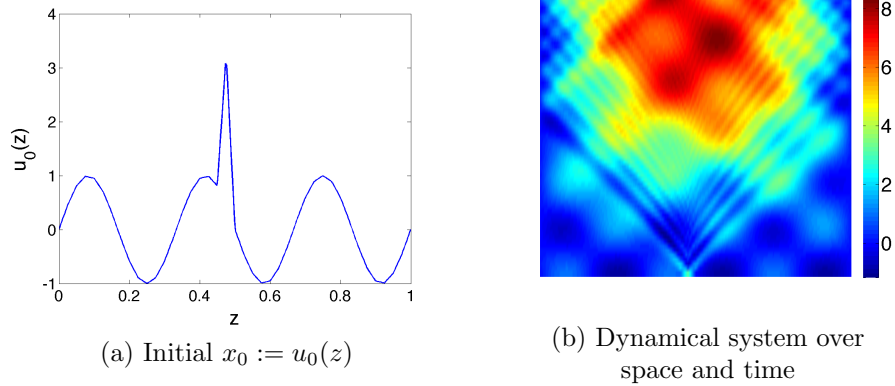
(b) Dynamical system over
space and time

Figure 5.2: Nonlinear 1D Wave equation

where F and κ are constants independent of j. To form a cyclic chain, we set $u_N = u_0$, $u_{-1} = u_{N-1}$ and $u_{N+1} = u_1$. This system is known to have a chaotic behaviour over time depending on the parameters N, F and κ (see for instance Karimi and Paul, 2009) For a given set of parameters N, F and κ for which a stable behavior is observed, i.e. where data assimilation can be performed, we consider then the following dynamical system:

$$\frac{du_{j+\theta}}{dt} = \frac{1}{\kappa}(-u_{j+\theta-2}u_{j+\theta-1} + u_{j+\theta-1}u_{j+\theta+1} - u_{j+\theta} + F), \quad j = 1, \ldots N, \quad \theta = 1, \ldots, \Theta, \quad (5.26)$$

where θ and Θ are integers. Thus, the new size of the vector $\bar{u}(t)$ is $N \times \Theta$, which may be specified as large as needed in our numerical experiments. The dynamical system is plotted for $N = 40$, $F = 8$, $\kappa = 120$ and $\Theta = 10$, using the coordinate graph described in Figure 5.3 (a). For an initial state $x = \bar{u}(0)$ as that shown in Figure 5.3 (b) the system develops over time as described in Figure 5.3 (c) (view from the top). As we observe that the system becomes chaotic after a certain time, we consider a reduced window of assimilation plotted in Figure 5.3 (d).

5.2   Results

We now provide numerical results using Algorithm 4.1, which uses the RPCG algorithm in Step 3, as was mentioned in Section 2. When tuning the accuracy parameter for stopping the inner iterations, we noticed that it was suitable to choose a value in the middle of the possible range. In the case of the 1DWave equation the range of the parameter is more or less defined by $[10^{-1}, 10^{-4}]$, thus we choose 10^{-2}, while the range is approximately $[10^{-1}, 10^{-6}]$ in the case of the Lorenz96 model, and we chose the value 10^{-4} in our experiment.

For the 1DWave problem, we depict the background vector in a black dashed line together with the true solution as a red line, in Figure 6.4 (a). The result given by our algorithm is plotted as a blue line in Figure 6.4 (b). The background vector and the true solution are plotted in Figure 6.5 (a) and the result given by our algorithm is presented in Figure 6.5 (b) for the Lorenz96 model.

In order to observe the adaptive nature of the algorithm, for an intermediate iteration i, we display two consecutive sets of observations $\mathcal{O}_i$ and $\mathcal{O}_{i+1}$ in Figures 6.6 (a)-(b) and 6.7 (a)-(b) for the 1DWave and the Lorenz96 respectively. In order to appreciate the impact of the local error indicators defined in Step 4, we also illustrate, in Figures 6.6 (e) and 6.7 (e), the local behaviour of the error between the prolongation of the current λ_i to the set $\mathcal{O}_{i+1}$ and the true $\tilde{\lambda}_{i+1}$, as given by

$$\epsilon_j = \left\langle (\tilde{\lambda}_{i+1} - \Pi_i\lambda_i)|_j, [(\tilde{R}_{i+1} + \tilde{H}_{i+1}B\tilde{H}_{i+1}^T)(\tilde{\lambda}_{i+1} - \Pi_i\lambda_i)]|_j \right\rangle,$$

together with that of the local error indicators themselves (η_j and ν_j are displayed Figures 6.6 (c)-(d) and 6.7 (c)-(d), respectively).

(a) Coordinate system

(b) Initial $x_0 := (u_1(0), u_2(0), \ldots, u_N(0))$

(c) Dynamical system
over space and time
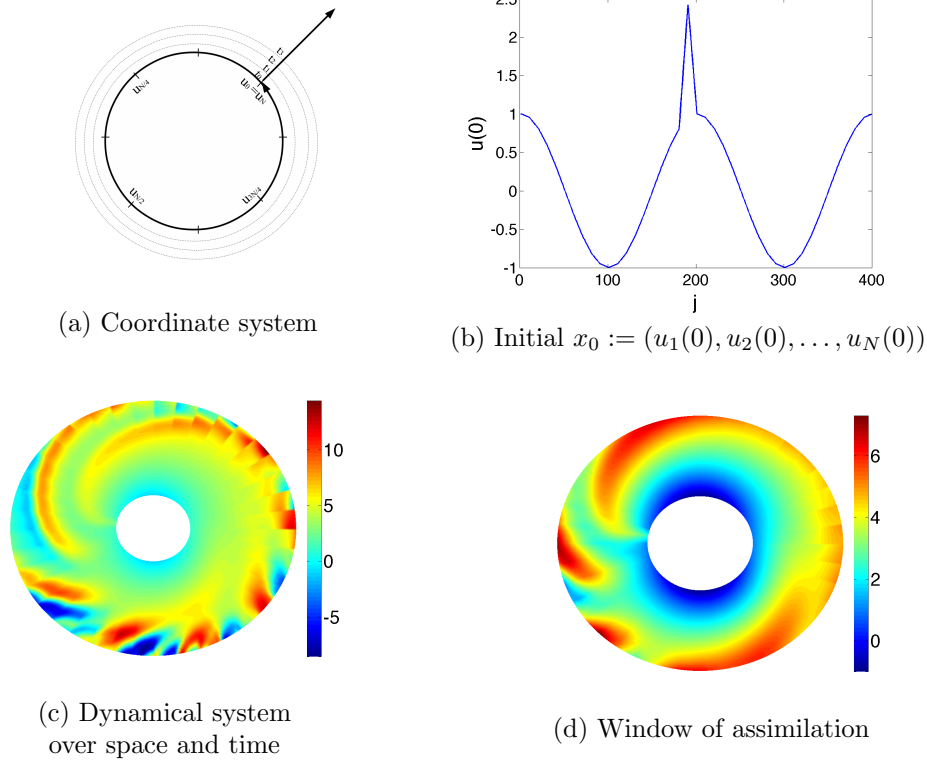

(d) Window of assimilation

Figure 5.3: Lorenz96 problem

In the case of the 1DWave equation we observe in Figure 5.2 (a) that the peak in the middle of the signal produces a dynamical reaction over space and time, shown in part (b) of the same figure. The errors η_j and ν_j in Figure 6.6 (c) and (d) are larger in the regions of strong dynamical activity, whose identification is clear in Figure 6.6 (a)-(b) which shows the evolution of the observation sets. In Figure 6.6 (e), we also observe that the regions where the difference ϵ_j is large coincide with those where η_j and ν_j are also large in Figure 6.6 (c)-(d).

In the case of the Lorenz96 model where the initial signal has also a peak in the middle (Figure 5.3 (b)), the high dynamics are on the left part of the space and time graph (Figure 5.3 (d)). In this case the quantities η_j and ν_j also provide correct indication of where more observations are needed, see Figure 6.7 (c)-(d). We also note that the distance ϵ_j in Figure 6.7 (c) is consistent with η_j and ν_j. Indeed, the sets of observations in Figure 6.7 (a)-(b) show that the observations concentrate on the left part of the graph where the highly dynamic nature of the model is most evident.

In Figures 6.8 and 6.9, we compare the performance of the new algorithm with the simple use of uniform observations and a benchmark method where a pre-established hierarchy of uniform distributed and progressively denser observations sets is used in Algorithm 4.1 (skipping Steps 4-6). This last method bypasses the new adaptive features of the method and its associated computing cost. For this comparison, we define, for a given iterate x resulting from Step 7 of Algorithm 4.1, the cost function as the value of the function in (2.3) when $i = r$, i.e. when all the possible observations are used. We then plot the evolution of this cost function (in logarithmic scale) against the number of observations used and the associated flop (floating-point operations) counts for the three algorithms. (The curves for the uniform case do not correspond to an algorithm, but show the accuracy and computational costs associated with directly solving the problem on a uniform grid for each specific size.)

We observe in both cases that the new method achieves a smaller error (plotted in red) than

that obtained by using uniformly distributed observations (plotted in blue) or the benchmark (plotted in black), which is explained by the fact that most observations used by the adaptive algorithm are in regions of space where their contribution to accuracy is largest.

Computational experience not reported here also indicates that the effect of the choice of starting value of x does not affect significantly the hierarchy of observation sets beyond the fact that observations concentrate in regions of high dynamical activity. We also found that $[0.4, 0.7]$ appears to be an adequate range for the parameters θ_1 and θ_2 in (4.21), yielding a satisfactory rate of inclusion of new observations at each iteration.

Both results on accuracy and computational costs are therefore highly encouraging.

6 Conclusions and perspectives

A method which identifies the influential data on a 4D-Var data assimilation problem is proposed, as well as an algorithm that exploits this identification to improve on computing efficiency. The cpu-time gains are obtained for two reasons, the first being that the available number of observations is used very effectively, and the second the fact that the cost of the subproblem solution is signficantly reduced by the use of dual-space conjugate-gradient techniques like RPCG. Numerical experience has been presented on two nonlinear test problems, and the results are so far encouraging.

Further refinements of the algorithm could be considered, such as the use of adaptive preconditioners in the subproblem solver, and continued experience with the method is of course desirable to assess its true potential. Extensions of these ideas in other domains are also possible: we think in particular of data assimilation problems in frameworks where each state of the system is itself an image on which adaptive reconstruction techniques could be applied.
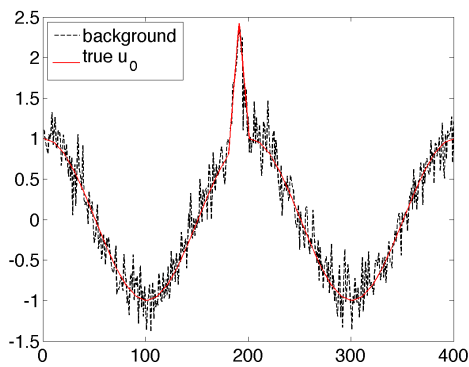
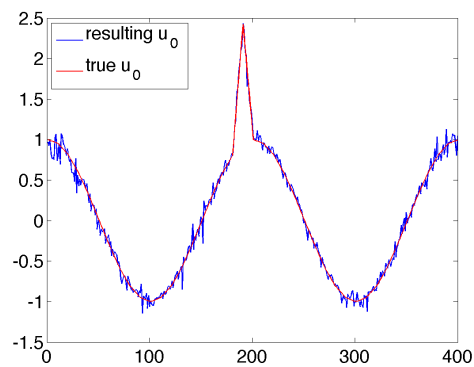(a) Background vector
and true initial $u(0)$
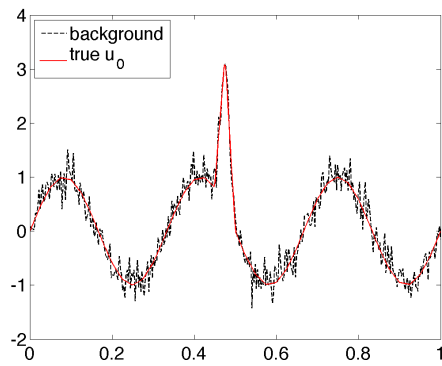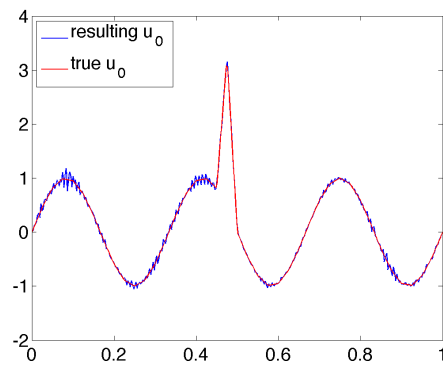
(b) True $u(0)$
and algorithm solution

Figure 6.4: Results: Nonlinear 1D Wave equation

(a) Background vector
and true initial $u(0)$

(b) True $u(0)$
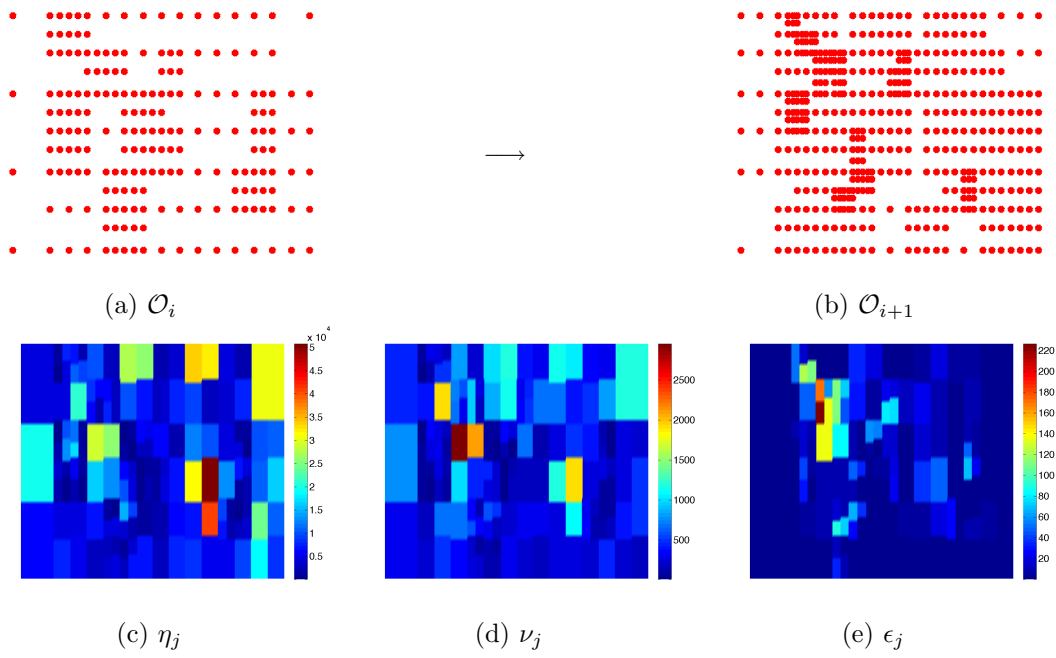and algorithm solution

Figure 6.5: Results: Lorenz96

(a) $\mathcal{O}_i$ (b) $\mathcal{O}_{i+1}$

(c) η_j (d) ν_j (e) ϵ_j

Figure 6.6: Observations set and adaptive errors

(a) $\mathcal{O}_i$ (b) $\mathcal{O}_{i+1}$

(c) η_j (d) ν_j (e) ϵ_j
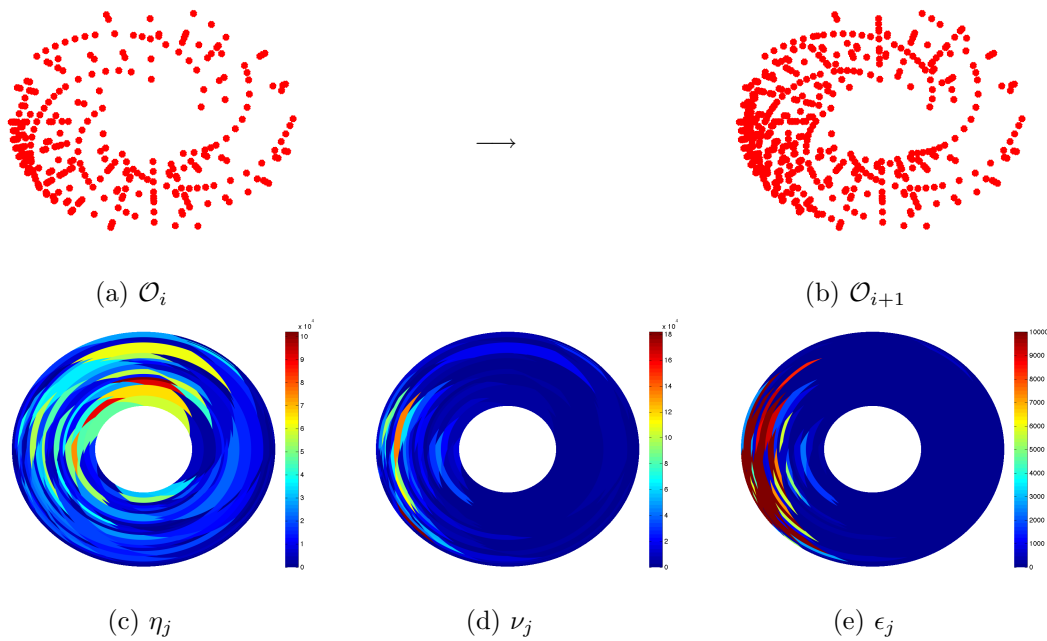
Figure 6.7: Observations set and adaptive errors

(a) Cost function versus number of observations   (b) Cost function versus flops
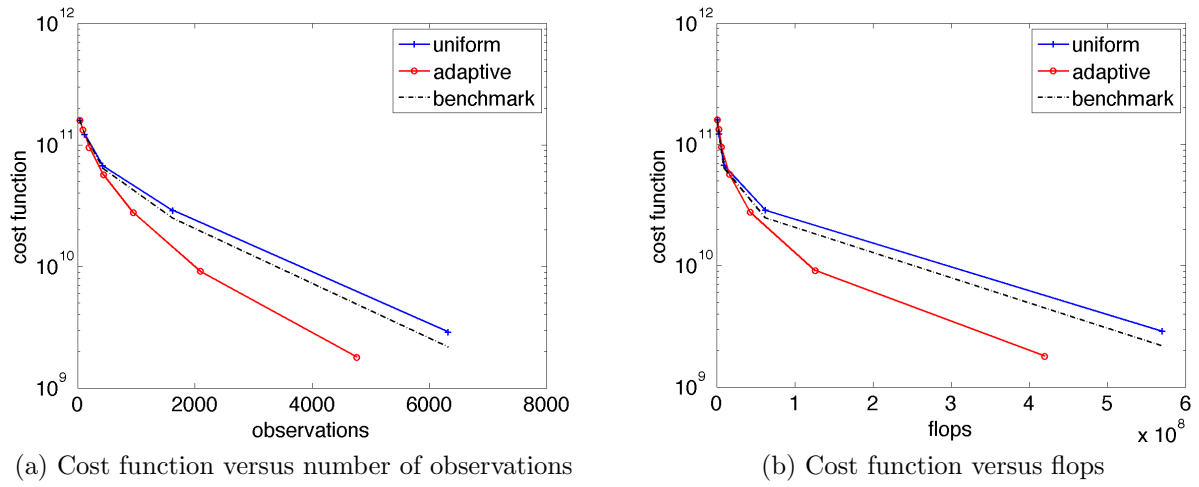
Figure 6.8: Performance of the algorithm on the nonlinear wave equation

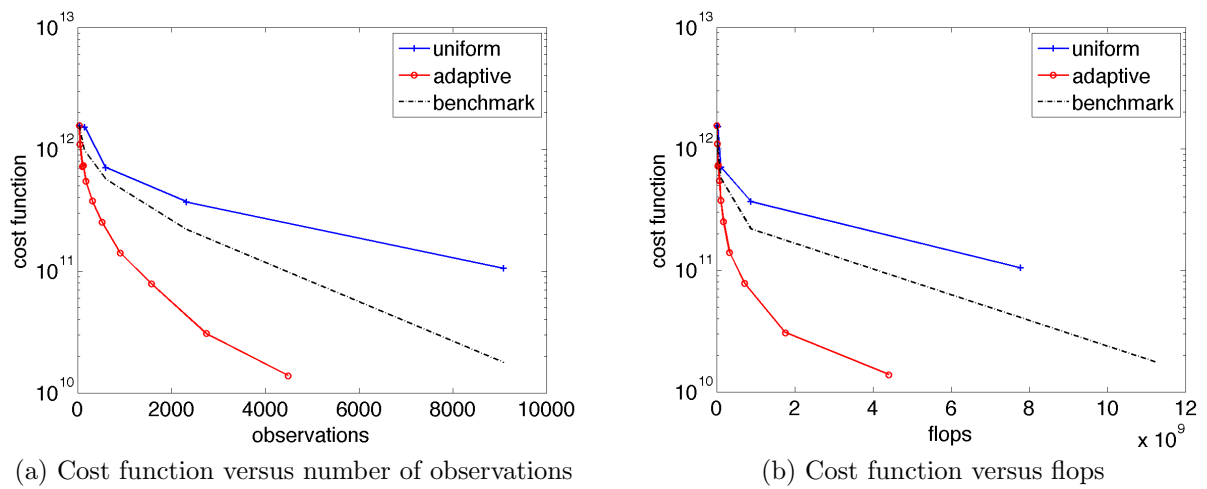(a) Cost function versus number of observations   (b) Cost function versus flops

Figure 6.9: Performance of the algorithm on the Lorenz96 problem

References

L.M. Berliner, Z.Q. Lu, and C. Snyder. Statistical design for adaptive weather observations. *Journal of the Atmospheric Sciences*, **56**(15), 2536–2552, 1999.

A. Brandt. Multi-level adaptive technique (MLAT) for fast numerical solution to boundary value problems. *in* 'Proceedings of the Third International Conference on Numerical Methods in Fluid Mechanics', pp. 82–89, Heidelberg, Berlin, New York, 1973. Springer Verlag.

C. Cardinali, S. Pezzulli, and E. Andersson. Influence-matrix diagnostic of a data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **130**(603), 2767–2786, 2004.

Ph. Courtier. Dual formulation of four-dimensional variational assimilation. *Quarterly Journal of the Royal Meteorological Society*, **123**, 2449–2461, 1997.

Ph. Courtier, J.-N. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4D-Var using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, **120**, 1367–1388, 1994.

D.N. Daescu and I.M. Navon. Adaptive observations in the context of 4D-Var data assimilation. *Meteorology and Atmospheric Physics*, **85**(4), 205–226, 2004.

W. Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM Journal on Numerical Analysis*, **33**(3), 1106–1124, 1996.

A. El Akkroui, P. Gauthier, S. Pellerin, and S. Buis. Intercomparison of the primal and dual formulations of variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **134**(633), 1015–1025, 2008.

G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins University Press, Baltimore, third edn, 1996.

S. Gratton and J. Tshimanga. An observation-space formulation of variational assimilation using a restricted preconditioned conjugate-gradient algorithm. *Quarterly Journal of the Royal Meteorological Society*, **135**, 1573–1585, 2009.

S. Gratton, S. Gürol, and Ph. L. Toint. Preconditioning and globalizing conjugate gradients in dual space for quadratically penalized nonlinear-least squares problems. *Computational Optimization and Applications*, **54**(1), 1–25, 2013.

S. Gratton, A. Lawless, and N. K. Nichols. Approximate Gauss-Newton methods for nonlinear least-squares problems. *SIAM Journal on Optimization*, **18**, 106–132, 2007.

J.A. Hansen and L.A. Smith. The role of operational constraints in selecting supplementary observations. *Journal of the Atmospheric Sciences*, **57**(17), 2859–2871, 2000.

M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of the National Bureau of Standards*, **49**, 409–436, 1952.

A. Karimi and M.R. Paul. Extensive chaos in the Lorenz-96 model. *Arxiv preprint arXiv:0906.3496*, 2009.

M. Leutbecher. Adaptive observations, the Hessian metric and singular vectors. *in* 'Proc. ECMWF Seminar on Recent developments in data assimilation for atmosphere and ocean, Reading, UK', pp. 8–12, 2003*a*.

M. Leutbecher. A reduced rank estimate of forecast error variance changes due to intermittent modifications of the observing network. *Journal of the Atmospheric Sciences*, **60**(5), 729–742, 2003*b*.

J. Liu, E. Kalnay, T. Miyoshi, and C. Cardinali. Analysis sensitivity calculation in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, **135**(644), 1842–1851, 2009.

A. Logg, K.A. Mardal, and G.N. Wells. *Automated solution of differential equations by the finite element method*, Vol. 84 of *Lecture Notes in Computational Science and Engineering*. Springer Verlag, Heidelberg, Berlin, New York, 2012.

E.N. Lorenz and K.A. Emanuel. Optimal sites for supplementary weather observations: Simulation with a small model. *Journal of the Atmospheric Sciences*, **55**(3), 399–414, 1998.

E. N. Lorenz. Predictability: a problem partly solved. *in* 'Seminar on Predictability', Vol. 1, pp. 1–18, Reading, UK, 1995. ECMWF.

S. T. McCormick. Fast adaptive composite grid (FAC) methods: Theory for the variational case. *in* 'Defect correction methods', pp. 115–121. Springer Verlag, Heidelberg, Berlin, New York, 1984.

P. Morin, R.H. Nochetto, and K.G. Siebert. Data oscillation and convergence of adaptive FEM. *SIAM Journal on Numerical Analysis*, **38**(2), 466–488, 2000.

M. M. Rincon-Camacho. *Adaptive methods for Total Variation based image restoration.* PhD thesis, University of Graz, Graz, Austria, 2011.