CARI'10

Multigrid methods and data assimilation

Convergence study and first experiments on non-linear equations

Neveu Emilie [*], Debreu Laurent [**], Le Dimet François-Xavier [***]

INRIA MOISE Project-Team and Laboratoire Jean Kuntzmann
Grenoble, France
[*] Emilie.Neveu@imag.fr

[**] Laurent.Debreu@imag.fr,

[***] Francois-Xavier.Le-Dimet@imag.fr

ABSTRACT. In order to limit the computational cost of the variational data assimilation process, we investigate the use of multigrid methods to solve the associated optimal control system. On a linear advection equation, we study the impact of the regularization term and the discretization errors on the efficiency of the coarse grid correction step introduced by the multigrid method. We show that even if for a perfect numerical model the optimal control problem leads to the solution of an elliptic system, discretization errors introduce implicit diffusion that can alter the success of the multigrid methods. Then we test the multigrids configuration and the influence of the algorithmic parameters on a non-linear Burgers equation to show that the algorithm is robust and converges much faster than the monogrid one.

RÉSUMÉ. Afin de limiter le coût de calcul lié aux méthodes variationnelles d'assimilation de données, nous nous intéressons ici à l'utilisation de méthodes multigrilles pour la résolution de systèmes de contrôle optimal. Sur un modèle simple d'advection linéaire, nous étudions l'impact du terme de régularisation du contrôle optimal ainsi que l'impact des erreurs de discrétisation sur l'efficacité de la correction grille grossière introduite par cette méthode. En particulier, nous montrons que pour un modèle numérique parfait, le problème de contrôle optimal est elliptique mais que les erreurs de discrétisation introduisant une diffusion implicite peuvent altérer les performances de la méthode multigrille. Enfin, sur une équation de Burgers, non linéaire, nous étudions l'influence des différents paramètres inhérents aux méthodes multigrilles et montrons que ces méthodes sont robustes et convergent beaucoup plus rapidement que les méthodes monogrilles.

KEYWORDS : variational data assimilation, multigrid methods, optimal control

MOTS-CLÉS : assimilation variationnelle de données, méthodes multigrilles, contrôle optimal

1. Introduction

Data assimilation methods are a way of combining different sources of information: observations and numerical models according to error statistics on these sources. They can be divided into two groups (Bennett, 2002 [1]). First, sequential methods are based on the Kalman filtering approach and make the state vector evolve in time along with its error statistics. Then, variational methods are based on optimal control techniques and minimize the distance between the model trajectory and observations according to a cost function $\mathcal{J}$.

These methods have led to strong improvements in the operational context of weather or ocean forecast. But they still have huge computational costs and have to be simplified for operational purposes.

We will focus on the 4D-variational data assimilation (4D-var), introduced by LeDimet and Talagrand in 1986 [4]. In variational data assimilation, assuming $\mathbf{x}_0$ is the control vector, the necessary condition of optimality is given by the Euler equation $\nabla_{\mathbf{x}_0} J = 0$. In this paper, the multigrid methods are used for solving the resulting system to accelerate the resolution by solving on coarser grids.

In the optimal control framework, several attempts have been made to apply multigrid methods, either for linear or non linear optimization (Ta'asan, 1991 [16]; Nash, 2000 [13]; Nash, 2005 [12]). Nash [12] focuses on the control of the initial condition for a linear advection equation with a specific cost function and discretization scheme that makes the problem really well suited for multigrid methods. In this paper, we will use a cost function typical of data assimilation applications.

We present the multigrid methods in a general case in section 2. In section 3, we apply the multigrids on a data assimilation problem, adapting the convergence properties. Then, in section 4, we define a data assimilation system characterized by a linear advection equation and a cost function associated with a typical regularization term. Additionally, using Fourier analysis, we study how discretization errors in the numerical model can alter the efficiency of the coarse grid correction step. Finally, the section 5 deals with a non-linear Burgers equation. We will compare the multigrid method to the monogrid one and study the influence of the multigrid parameters to test its robustness.

2. Multigrid methods

Multigrid methods have been known since the forties (Southwell, 1935 [14]; 1946 [15]) and have been used for a large class of systems, from linear elliptic ones to Navier Stokes equations (Brandt, 1980 [2]). In this section, basic ideas are reviewed. The next section will discuss the adaptation of the multigrid method to the data assimilation problem.

2.1. A way to correct iterative methods

Let $\hat{\mathbf{x}}_0$ be the solution of the following, potentially non linear, system:

$$A(\hat{\mathbf{x}}_0) = g \quad \text{on } \Omega \qquad [1]$$

Let $\mathbf{x}_0$ be an approximation of $\hat{\mathbf{x}}_0$ and $\delta\mathbf{x}_0 = \hat{\mathbf{x}}_0 - \mathbf{x}_0$ the error.
The residual r, given by $r = g - A(\mathbf{x}_0)$, satisfies the residual equation

$$A(\mathbf{x}_0 + \delta\mathbf{x}_0) - A(\mathbf{x}_0) = r \qquad [2]$$

Commonly, iterative methods are used to solve equation [1]. There exists a whole range of these iterative methods called *smoothers* because they reduce efficiently small scales of the error $\delta\mathbf{x}_0$ but have difficulties in solving its large scale components. Weighted-Jacobi or Gauss-Seidel are two of them. Multigrid methods improve the iterative methods by solving the residual equation [2] on a low resolution grid. Indeed the remaining large scales of the error, when described at a coarser grid resolution, are seen as smaller scales. In the following, for sake of clarity, we restrict ourselves to a two-level multigrid method. We assume we want to solve the system [1] on a fine grid domain Ω^f. The superscript f (resp. c) stands for the fine (resp. coarse) grid. To exchange informations between the two grids, we use restriction operators I_f^c, $\hat{I}_f^c$ and an interpolation or prolongation operator I_c^f. We denote by A^c, A^f and g^c and g^f the discretizations of A and g on the coarse and fine grids.
The multigrid algorithm which solves the residual equation [2] on the coarse grid is called Full Approximation Scheme (FAS) and can be written:

Full Approximation Scheme algorithm:
Loop on k until convergence

1) Apply ν_1 times an iterative method named SMOOTH:

$$\mathbf{x}_0^f = SMOOTH^{\nu_1}(\mathbf{x}_{0,k}^f, A^f, g^f) \qquad [3]$$

2) Coarse grid correction step:

a) Solve exactly or approximately with an iterative method

$$A^c(\mathbf{x}_0^c) - A^c(\hat{I}_f^c\mathbf{x}_0^f) = r^c = I_f^c(g^f - A^f(\mathbf{x}_0^f)) \qquad [4]$$

where r^c is the fine grid residual transfered to the coarse grid: $r^c = I_f^c r^f$.

b) Interpolate the correction and deduce the approximate solution on Ω^f, which may be expressed by:

$$\mathbf{x}^f_{0,\text{ after CGC}} = \mathbf{x}_0^f + I_c^f(\mathbf{x}_0^c - \hat{I}_f^c\mathbf{x}_0^f) \qquad [5]$$

after CGC meaning after Coarse Grid Correction

3) Apply ν_2 times the iterative method:

$$\mathbf{x}_{0,k+1}^f = SMOOTH^{\nu_2}(\mathbf{x}^f_{0,\text{ after CGC}}, A^f, g^f) \qquad [6]$$

End Loop, $k = k + 1$

SMOOTH is an iterative method suited for non linear systems. The residual equation [4] can also be solved by a multigrid method if more than 2 grids are used, making the algorithm recursive.
In the following, the residual equation [4] will be written as:

$$A^c_{\text{FAS}}(\mathbf{x}_0^c) = I_f^c g^f \qquad [7]$$

with

$$A^c_{\text{FAS}}(\mathbf{x}_0^c) = A^c(\mathbf{x}_0^c) + I_f^c A^f(\mathbf{x}_0^f) - A^c(\hat{I}_f^c\mathbf{x}_0^f) \qquad [8]$$

2.2. Convergence properties

The algorithm converges to $\hat{\mathbf{x}}_0^f$ depending on the transfer operators, the coarse grid system and the properties of the smoothing method (see Hackbusch, 1985 [8]).

Linear case

– *Ellipticity:* For linear cases where $A = \mathbf{A}$; if the operator $\mathbf{A}$ is elliptic, it is possible to find iterative methods that, if applied on the fine grid, will efficiently remove small scales in the error $\delta\mathbf{x}_0^f$ (see Trottenberg, 2001 [19]). However if the ellipticity insures that the multigrids will be more efficient than monogrids, it is not a necessary condition for convergence. Even if with weaker ellipticity properties, the multigrid methods may still be efficient.

– *Smoothing property:* the smoothing method needs to satisfy :

$$\exists\, \alpha, \|\mathbf{A}^f SMOOTH^\nu\| \leq \eta(\nu)(\Delta x^f)^{-\alpha}, \quad \forall \nu \geq 1$$

with $\lim_{\nu\to\infty} \eta(\nu) = 0$, and Δx^f, the fine grid space step.

– *Approximation property:* the coarse grid operators and transfer operators need to satisfy:

$$\|(\mathbf{A}^f)^{-1} - I_c^f(\mathbf{A}^c)^{-1} I_f^c\| \leq C(\Delta x^f)^\alpha, \quad C > 0 \qquad [9]$$

The satisfaction of the approximation property puts constraints on the interpolation/restriction orders and requires consistency of the discretizations of $\mathbf{A}^f$ and $\mathbf{A}^c$ with $\mathbf{A}$. This last condition can be insured if $\mathbf{A}^c$ is built using the Galerkin relation (Trottenberg, 2001 [19]):

$$\mathbf{A}^c = I_f^c \mathbf{A}^f I_c^f$$

However, in general, such a Galerkin operator is difficult to derive if the operator $\mathbf{A}^f$ is itself complex and additionally for nonlinear operators it would in practice require the explicit evaluation of this operator at high resolution and thus would be too costly. We will rather attempt to approach it.
The efficiency of the coarse grid correction step itself is given by the amplification factor

$$K_{\text{mgrid}} = \mathbf{I} - I_c^f(\mathbf{A}^c)^{-1} I_f^c \mathbf{A}^f \qquad [10]$$

that have to be small for large scales. This factor is relevant when the coarse grid problem is solved exactly. For practical applications, the coarse grid problem does not have to be solved exactly for the multigrid iterations to converge. We will verify this assumption further.

Non linear case

For non linear cases, the convergence study made by Yavneh et al., 2006 [20] emphasizes the importance for the coarse grid operator $A^c(\mathbf{x}_0^c)$ to be close enough to the fine grid one $A^f(\mathbf{x}_0^f)$ but it is not a sufficient condition for convergence (see Lewis and Nash, 2005 [12] and Ta'asan, 1997[17] for optimization-based convergence studies).

3. Multigrid methods for variational data assimilation

In a variational data assimilation method, the aim is to minimize a given cost function $\mathcal{J}(\mathbf{x}_0)$ playing with the control vector $\mathbf{x}_0$. The cost function measures the distance between a set of observations $\mathbf{y}^o$ (available for times t_n) and the model as follows

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2}\sum_n \|H_{t_n}(\mathbf{x}_0) - \mathbf{y}^o_{t_n}\|^2_{\mathbf{R}^{-1}_{t_n}} + \frac{1}{2}\|\mathbf{x}_0 - \mathbf{x}^b\|^2_{\mathbf{B}^{-1}} \qquad [11]$$

The first term measures the misfit to data while the last term is a regularization one, $\mathbf{x}^b$ being the background or current estimate of the initial state.
H_{t_n} is an operator that includes both the model trajectory $M_{0,n}(.)$ from time 0 to time n and the observation operator that goes from the model space to the observations space. $\mathbf{B}$ is the background error covariance matrix associated to $\mathbf{x}^b$ and $\mathbf{R}_{t_n}$ the observational error covariance matrix.

Let $\hat{\mathbf{x}}_0$ be a minimum of the cost function : $\mathcal{J}(\hat{\mathbf{x}}_0) = \min_{\mathbf{x}_0} \mathcal{J}(\mathbf{x}_0)$.
Then, a necessary condition for $\hat{\mathbf{x}}_0$ is to satisfy the Euler equation :

$$\nabla_{\mathbf{x}_0}\mathcal{J}(\hat{\mathbf{x}}_0) = 0 \qquad [12]$$

Variational data assimilation (Le Dimet, 1986 [4]; Ide, 1997 [10]; Bennett, 1992 [1]) is based on a minimization procedure, that requires the knowledge of the gradient of J which itself requires integration of both direct and adjoint models. As one integration of an operational model is already of a high cost of resolution, solving data assimilation control optimal system can then become very expensive. In addition, the non-linearities and the complexities of physical phenomena make the non-quadradic cost function difficult to minimize. The resulting cost of the data assimilation process can easily be 10 times higher than the model cost, depending on the algorithmic configuration.
Our objective is to alleviate the cost of 4D-Var calculation by solving the Euler equation [12] using a multigrid method. Nash [13], and Ta'asan [17] have studied some optimal control problems solved by multigrid methods. There, we look more specifically at optimal control problems that come from a variational data assimilation problem. They differ by the shape of the cost function.
In order to relate this problem to the one introduced in the previous section, we write the Euler equation [12] under the form $A(\mathbf{x}_0) = g$ where $\nabla\mathcal{J}(\mathbf{x}_0) = A(\mathbf{x}_0) - g$
The Full Approximation Scheme consists in applying the multigrid algorithm to solve the non linear system $\nabla\mathcal{J}^f(\mathbf{x}^f_0) = 0$. The residual equation solved on coarse grid is inspired by [7] and [8]:

$$\nabla\mathcal{J}^c_{\text{FAS}}(\mathbf{x}^c_0) = \nabla\mathcal{J}^c(\mathbf{x}^c_0) + I^c_f\nabla\mathcal{J}^f(\mathbf{x}^f_0) - \nabla\mathcal{J}^c(\hat{I}^c_f\mathbf{x}^f_0) = 0 \qquad [13]$$

By defining

$$L = I^c_f\nabla\mathcal{J}^f(\mathbf{x}^f_0) - \nabla\mathcal{J}^c(\hat{I}^c_f\mathbf{x}^f_0) \qquad [14]$$

we have to minimize on the coarse grid the new cost function $\mathcal{J}^c_{\text{FAS}}$:

$$\mathcal{J}^c_{\text{FAS}}(\mathbf{x}^c_0) = \mathcal{J}^c(\mathbf{x}^c_0) + < L, \mathbf{x}^c_0 > \qquad [15]$$

The iterate $\mathbf{x}^f_0$ is transposed on coarse grid with the restriction operator $\hat{I}^c_f$ that may be different from the one transposing the residual $-\nabla\mathcal{J}^f[\mathbf{x}^f_0]$. Next, we will simply use the

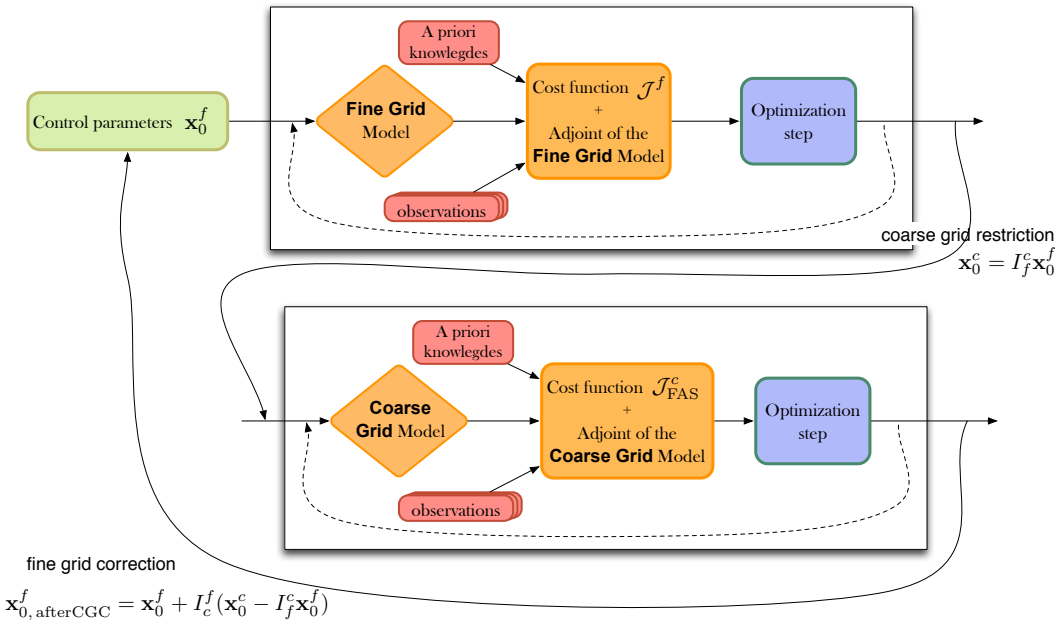

Figure 1. *The 4D-Var FAS Algorithm .*
same restriction operator I_f^c. Figure 1 shows the schematic of this algorithm.

Convergence Properties in the linear case:

– *Ellipticity:* In data assimilation process, as shown Figure 1, the iterative method used is an optimization procedure as conjugate gradient algorithm or Quasi-Newton methods. The work of Gratton (2004, [7]) indicates that even if they are not strictly smoothers, they are likely to act as smoothing operators.

– *Approximation property:* The coarse grid correction step gives us a convergence condition, as in equation [9]. The coarse grid operator $\mathbf{A}^c$ needs to approximate $\mathbf{A}^f$. In our variational data assimilation problem, the operator $\mathbf{A}$ corresponds to the Hessian $\mathcal{H}$ of the cost function $\mathcal{J}$. It can be expressed using the cost function [11]:

$$\mathcal{H} = \sum_i (\mathbf{H}_{t_n})^T \, \mathbf{R}^{-1} \mathbf{H}_{t_n} + \mathbf{B}^{-1} \qquad [16]$$

The following conditions are thus needed to satisfy the approximation property:

$$\begin{cases} \|\mathbf{H}^c_{t_n} - \mathbf{H}^f_{t_n} I^f_c\| \leq \epsilon, \forall n \\ \|(\mathbf{B}^c)^{-1} - I^c_f (\mathbf{B}^f)^{-1} I^f_c\| \leq \epsilon \\ (I^f_c)^T = c\, I^c_f \quad \text{(operators are ajdoints)} \end{cases} \qquad [17]$$

4. Convergence study on a linear advection equation

In this section, we study the application of the previously described method on a linear advection equation. First, the continuous and discrete problems are introduced, then we proceed to a Fourier analysis of the convergence properties of the coarse grid correction step.

4.1. Model and cost function

We use an elementary advection equation on the one-dimensional domain Ω

$$\left.\begin{array}{l} \partial_t \mathbf{x} + c\,\partial_x \mathbf{x} = 0 \quad \text{with } c > 0 \\ \mathbf{x}(x, t=0) = \mathbf{x}_0(x) \end{array}\right\} \qquad [18]$$

with periodic boundary conditions.
We suppose that the observations data set $\mathbf{y}^o$ is available continuously (the observation operators $\mathbf{H}_{t_n}$ are all equal to identity) and that the observational error covariance matrix $\mathbf{R}$ is a diagonal matrix with a constant variance equal to σ_o^2 on fine and coarse grids.
The cost function [11] makes use of the background error covariance matrix $\mathbf{B}$. In typical applications, $\mathbf{B}$ is representative of errors correlated with a Gaussian shape function. In that case, the regularization term can be approximated using spatial derivatives (see Bennett, 2002 [1]), σ_b^2 being the error variance.
The continuous cost function is given by

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2\sigma_o^2}\int_0^T \|\mathbf{x}(x,t) - \mathbf{y}^o\|^2 + \frac{\beta}{2\sigma_b^2}\|(\mathbf{I} - \frac{\sigma_b^2}{4}\triangle)\left(\mathbf{x}_0 - \mathbf{x}^b\right)\|^2 \qquad [19]$$

where β is a positive constant that adapts the weight of the regularization term.
Using the continuous solution of equation [18], $\mathbf{x}(x,t) = \mathbf{x}_0(x - ct)$, the expression of the Hessian of $\mathcal{J}$ can be derived :

$$\mathcal{H} = \frac{T}{\sigma_o^2}\left(1 + \gamma\left(\mathbf{I} + \frac{\sigma_b^2}{2}\frac{d^2}{dx^2} + \frac{\sigma_b^4}{16}\frac{d^4}{dx^4}\right)\right), \quad \text{where } \gamma = \frac{\beta\sigma_o^2}{T\sigma_b^2} \qquad [20]$$

$\widetilde{\mathcal{H}}(k)$ the Fourier symbol of the Hessian, defined by $\mathcal{H}(e^{ikx}) = \widetilde{\mathcal{H}}(k)e^{ikx}$, is given by the expression:

$$\widetilde{\mathcal{H}}(k) = \frac{T}{\sigma_o^2}\left[1 + \gamma\left(1 + \frac{\sigma_b^2}{2}k^2 + \frac{\sigma_b^4}{16}k^4\right)\right] \qquad [21]$$

With $\gamma = 0$, that is to say if the cost function does not include the regularization term, then the Hessian does not depend on k. The optimization method will have a similar behavior at all scales so that the multigrid method will loose its theoretical efficiency.
However if the regularization term is added, γ is stricly positive and in that case, an elliptic operator is obtained.

Equation [18] is discretized using a finite difference method based on a uniform grid with time step Δt and space step Δx, using an Euler upwind scheme.
We note $\mathbf{x}_n^j$ the approximation of the value of $\mathbf{x}(x,t)$ at $x = j\Delta x$ and $t = n\Delta t$.
Defining $\lambda = c\Delta t/\Delta x$ the Courant number, the numerical scheme writes:

$$\mathbf{x}_{n+1}^j - \mathbf{x}_n^j = -\lambda\left(\mathbf{x}_n^j - \mathbf{x}_n^{j-1}\right) \qquad [22]$$

The laplacian $\triangle$ of the background term is discretized using a second order centered scheme.

The properties of the Euler upwind scheme are well known. It is first order accurate in space and time and is conditionally stable under the constraint $0 \leq \lambda \leq 1$. By a Taylor expansion, it can be proved that the numerical solution produced by scheme [22] is a first order approximation in time and space of the advection equation [18] but is a second order approximation of the following advection-diffusion equation (see 2010, [5], p. 109):

$$\partial_t \mathbf{x} + c\,\partial_x \mathbf{x} = \epsilon \partial_{xx} \mathbf{x}, \text{ with } \epsilon = \frac{c}{2}\left(\Delta x - c\,\Delta t\right) \qquad [23]$$

Note that for the particular case of $\lambda = 1$, $\epsilon = 0$ and the numerical model actually leads to the exact solution.
If $\lambda \neq 1$, ϵ is positive due to the stability constraint $0 \leq \lambda \leq 1$ and the additional term on the right hand side corresponds to a diffusion term.
Using the same cost function with the above advection-diffusion equation changes the expression of the Hessian, which in Fourier space gives:

$$\widetilde{\mathcal{H}}(k) = \frac{1 - e^{-2\epsilon k^2 T}}{\sigma_o^2 2\epsilon k^2} + \frac{\beta}{\sigma_b^2}\left(1 + \frac{\sigma_b^2}{2}k^2 + \frac{\sigma_b^4}{16}k^4\right) \qquad [24]$$

Its Taylor expansion at order 2 is:

$$\widetilde{\mathcal{H}}(k) = \frac{T}{\sigma_o^2}\left(1 - \epsilon T k^2 + \gamma\left(1 + \frac{\sigma_b^2}{2}k^2\right)\right) + O(k^4) \qquad [25]$$

This expression is also valid for the large scales of the discrete solution given by [22]. It shows that at large scales there is a balance between the artificial numerical diffusion and the regularization term. We note γ_{lim}, the γ-coefficient that cancels the second order term:

$$\gamma_{\text{lim}} = \frac{2\epsilon T}{\sigma_b^2} \qquad [26]$$

If $\gamma \geq \gamma_{\text{lim}}$, the Hessian is truly elliptic (i.e. $|\widetilde{\mathcal{H}}(k)| \geq \alpha k^2, \alpha > 0$).

4.2. Discrete analysis

4.2.1. Hessian ellipticity

We studied above the behavior of the Hessian for large scales ($k\Delta x \ll 1$). To complete this study, we numerically compute the inverse of the discrete Hessian according to the mode k, $|\widetilde{\mathcal{H}^{-1}}(k)|$. The two parameters T and L are given by $T = 1$ and $L = 4$. Others parameters are fixed so that $\gamma_{\text{lim}} = 1$. Values of $|\widetilde{\mathcal{H}^{-1}}(k)|$ shown here are the true discrete values, the expression [24] being only valid at large scales $k\Delta x \ll 1$.
Figure 2 and Figure 3 show the influence of γ on the modulus of the inverse of the fine grid Hessian for $\lambda = 0.9$. When the regularization term is omitted ($\gamma = 0$, see Figure 2), the numerical diffusion makes the model less controllable. Even if no multigrid method is used, a regularization term $\gamma > 0$ is required to make the fine grid optimization easier. In Figure 3, we omit the value $\gamma = 0$ to better visualize the results for $\gamma > 0$. With $\gamma < \gamma_{\text{lim}}$, at medium scales there is a competition between the artificial diffusion induced by the

discretization and the regularization term as found on equation [25]. When $\gamma \geq \gamma_{\text{lim}}$, the Hessian becomes elliptic.

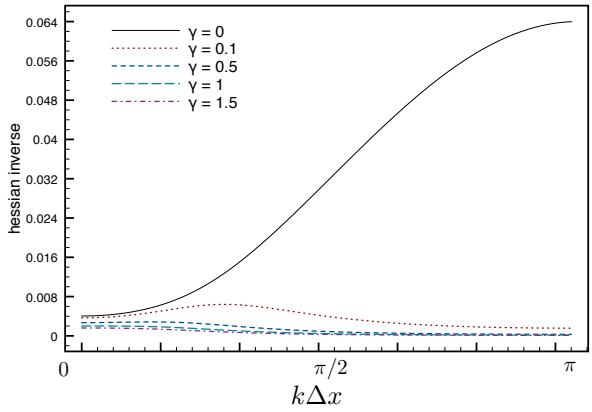

Figure 2. *The inverse of the Hessian, $|\widetilde{\mathcal{H}^{-1}}(k)|$, according to $k\Delta x$ for different values of γ.*

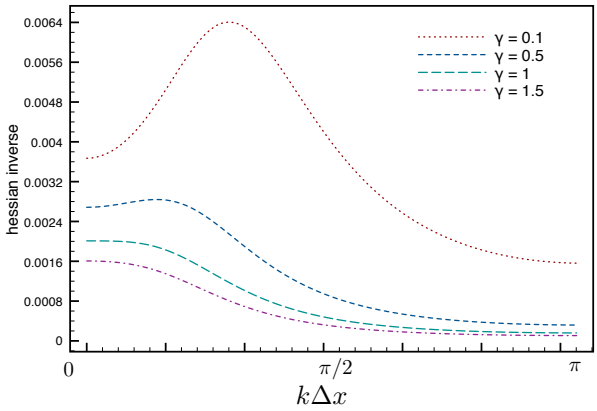

Figure 3. *The inverse of the Hessian, $|\widetilde{\mathcal{H}^{-1}}(k)|$, according to $k\Delta x$ for different values of γ, when omitting the value $\gamma = 0$. The Hessian is elliptic when $\gamma \geq \gamma_{lim} = 1$.*

4.2.2. The coarse grid correction

Assuming we solve exactly the linear coarse grid system [13] , the impact of the coarse grid correction can be studied using the convergence factor [10], with $\mathbf{A} = \mathcal{H}$:

$$K_{\text{4dvar-mgrid}} = \mathbf{I} - I_c^f (\mathcal{H}^c)^{-1} I_f^c \mathcal{H}^f \qquad [27]$$

For the multigrid process to be efficient, this factor should be small at large scales.
We use two different grids, Ω^f, a fine grid of resolution $\Delta x^f = 0.1$ in space and $\Delta t^f =$

$\frac{\lambda \Delta x^f}{c}$ in time, and a coarse one Ω^c of resolution Δx^c in space and Δt^c in time. We assume here that the spatial and temporal refinement factors are equal to 2. Thus the value of the Courant number $\lambda = 0.9$ is the same for the two grids, but numerical diffusion is twice higher on the coarse grid. The other parameters have identical values on fine and coarse grids.

The chosen restriction operator corresponds to a full-weighting operator while the interpolation operator is linear. Their stencils are given by :

$$I_f^c = \frac{1}{2} \left[\begin{array}{ccc} \frac{1}{2} & 1 & \frac{1}{2} \end{array} \right]_f \quad I_c^f = \left[\begin{array}{c} \frac{1}{2} \\ 1 \\ \frac{1}{2} \end{array} \right]_c \qquad [28]$$

In Figure 4, we plot the symbol of the convergence factor $K_{\text{4dvar-mgrid}}$ of equation [27] for different values of γ.

We observe, for both curves, a local maximum at large scales ($k\Delta x \in [0.2; 0.8]$). This

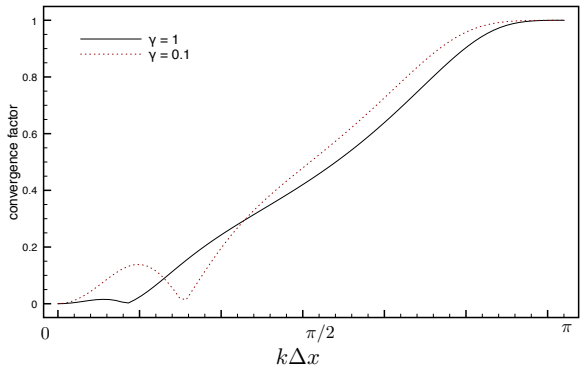

Figure 4. *The convergence factor [27] for one coarse grid correction according to $k\Delta x$, for different values of γ.*

extremum is certainly due to the fact that the regularization term, which reduces the non-ellipticity effect, is not effective for the large scales. Even with the presence of this local extremum, the convergence factor for $k \leq \frac{\pi}{2}$ is smaller than for $k > \frac{\pi}{2}$, which insures us that large scales are better resolved than small ones.

To compare to the monogrid case, Figure 5 represents in Fourier space the convergence factor of the error after ten high resolution weighted-Jacobi iterations. One iteration of a weighted-Jacobi is defined as:

$$\mathbf{x}_0^f = \left[\mathbf{I} - \omega \mathbf{D}^{-1} \mathbf{A}^f\right] \mathbf{x}_0^f + \omega \mathbf{D}^{-1} \mathbf{g}^f \quad \omega \in [0, 1] \qquad [29]$$

with $\mathbf{A} = \mathcal{H}$, $\mathbf{D}$ being the diagonal of $\mathcal{H}^f$ and $\mathbf{g}^f = \mathbf{0}$. We use here $\omega = 2/3$ and for the expression of the Hessian $\gamma = 1$ and $\lambda = 0.9$. This experiment is named Monogrid.

We compare it to the convergence factor of the error after one cycle of the FAS multigrid scheme using the same total number of fine grid iterations ($\nu_1 = \nu_2 = 5$ on the FAS algorithm). It is used the same relaxation scheme and the same parameters on fine grid as the Monogrid experiment. For this experiment, the convergence factor of the error is a combination of the coarse grid convergence factor [27] with the convergence factor already calculated by the Monogrid experiment. This experiment is named Multigrid and is

also shown on Figure 5.
Comparing both curves on Figure 5, the Multigrid method removes much better the large

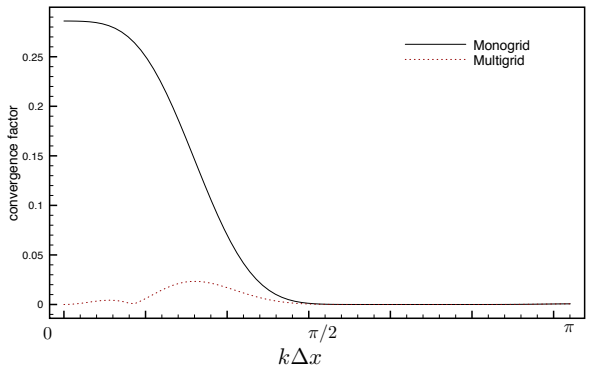

Figure 5. *The convergence factor of the Monogrid and the Multigrid experiments (both experiments use the same relaxation method on fine grid).*

scales than monogrid does, which emphasizes the coarse grid correction efficiency.

The study on the linear advection equation has been made from a theoretical point of view with a detailed discrete analysis of the Hessian and of the coarse grid correction step. We have seen that the background term of the cost function adds ellipticity to the system and that the coarse grid correction step proves to be useful in reducing large scales components of the error. In the next section, we interest us in a numerical study of the behavior of the multigrids. We also add non-linearities to test the robustness of the algorithm.

5. Experiments on a Burgers equation

We apply the FAS multigrid algorithm, presented in section 3 on a 4D-Var non linear optimal system based on a 1-D Burgers equation. We implement a monogrid high resolution method and compare it to the multigrid method. We also test different parameters of the multigrid algorithms to assess its robustness.

5.1. Model and cost function

We use this non linear inviscid Burgers equation:

$$\partial_t \mathbf{x} + \frac{1}{2}\partial_x \mathbf{x}^2 = 0 \quad \text{on } \Omega = [0,T] \times [0,L] \qquad [30]$$

with periodic boundary conditions.
Discretization is based on the staggered Lax-Friedrichs scheme (see 2007, [18], p. 112) with a space step Δx and a time step Δt. This discretized scheme is conditionally stable and is known to generate high numerical diffusion. We note $\mathbf{x}_n^j$ the value of $\mathbf{x}$ at point

$\mathbf{x}(j\Delta x, n\Delta t)$, $\mathbf{x}^j_{n+\frac{1}{2}}$ being an intermediate value calculated on fictious time $n+\frac{1}{2}$ on a staggered grid:

$$\left\{ \begin{array}{rcl} \mathbf{x}^j_{n+\frac{1}{2}} & = & \frac{1}{2}\left(\mathbf{x}^j_n + \mathbf{x}^{j+1}_n\right) - \frac{\Delta t}{2\Delta x}\left(\frac{1}{2}(\mathbf{x}^{j+1}_n)^2 - \frac{1}{2}(\mathbf{x}^j_n)^2\right) \\ \mathbf{x}^j_{n+1} & = & \frac{1}{2}\left(\mathbf{x}^{j-1}_{n+\frac{1}{2}} + \mathbf{x}^j_{n+\frac{1}{2}}\right) - \frac{\Delta t}{2\Delta x}\left(\frac{1}{2}(\mathbf{x}^j_{n+\frac{1}{2}})^2 - \frac{1}{2}(\mathbf{x}^{j-1}_{n+\frac{1}{2}})^2\right) \end{array} \right. \quad [31]$$

The scheme can be written:

$$\mathbf{x}_n = M_{n-1,n}(\mathbf{x}_{n-1}) = M_{0,n}(\mathbf{x}_0) \quad [32]$$

where $M_{0,n} = M_{n-1,n} \circ M_{n-2,n-1} \circ ... \circ M_{0,1}$.
Observations are generated by an integration of the model starting from a known initial state $\mathbf{x}^t_0$, and adding a Gaussian noise perturbation.
We use the same representation of the background term as with the linear model (see function [19]), that gives us the discrete cost function below:

$$\mathcal{J}(\mathbf{x}_0) = \frac{1}{2\sigma_o^2}\sum_{n=0}^{N-1}\|\mathbf{x}_n - \mathbf{y}^o_n\|^2 + \frac{\beta}{2\sigma_b^2}\|(\mathbf{I} - \frac{\sigma_b^2}{4}\triangle)(\mathbf{x}_0 - \mathbf{x}^b)\|^2 \quad [33]$$

5.2. Grids and configuration

We integrate the model on the grid $\Omega = [0, L] \times [0, T]$ with $L = 1$ and $T = 0.512$. The true initial state used to generate the observations is:

$$\mathbf{x}^t_0(x) = \sin(2\pi x/L) \quad [34]$$

We use a two-grids configuration. The fine and coarse grid models use the same discretization scheme [31] but with spatial and temporal refinement factors equal to 2 : $\Delta x^f = 0.0025$, $\Delta t^f = 0.001$ and $\Delta x^c = 0.005, \Delta t^c = 0.002$. Here our numerical problem has a relatively low computation cost so that we limit ourselves to a two grid algorithm. In practical applications with very large systems, the use of more than two grids can of course be more advantageous.
Considering observations are available each n_x space steps and each n_t time steps:

$$\mathbf{y}^o_n = M_{0,n}(\mathbf{x}^t_0) + \mathcal{G}(0, \sigma_o^2) \quad [35]$$

$\mathcal{G}(0, \sigma_o^2)$ being a Gaussian perturbation of variance σ_o^2 and mean 0 as we modeled in the cost functions [19] and [33].
We use a density of observations with $n^f_t = 32$ and $n^f_x = 16$. On coarse grid, these values become $n^c_t = n^f_t/2$ and $n^c_x = n^f_x/2$. The other parameters have identical values on fine and coarse grids: $\sigma_o^2 = 0.02$, $\sigma_b^2 = 0.2$. The chosen restriction operator corresponds to a full-weighting operator while the interpolation operator is linear. We recall that their stencils are given by equation [28].
The fine grid background state $\mathbf{x}^{bf}$ which is also the initial guess for all algorithms is defined as :

$$\mathbf{x}^{bf}(x) = 0.9\sin(2\pi x/L + \pi) + 0.05\sin(5 \times 2\pi x/L) \quad [36]$$

Constructing the initial guess by adding high frequencies and a dephasing of π to the true initial state is a way of adding high frequencies in the initial error.

The coarse grid background is equal to the restriction of the fine grid one: $\mathbf{x}^{b^c} = I_f^c \mathbf{x}^{b^f}$. In the linear case, we found the cost function to be fully elliptic for $\gamma \geq \gamma_{\text{lim}}$. Here, to find an optimal solution $\hat{\mathbf{x}}_0$ not too close to the background state, we take $\gamma = 0.01$.

As a reference experiment, we implement the non linear 4D-VAR on fine grid, using a limited memory Quasi-Newton optimization procedure M1QN3 (Gilbert and Lemaréchal, 1989 [6]). For each algorithm that we test, we use the same M1QN3 optimization at high and low resolution as the iterative method. The resulting optimal state is shown on Figure 6 and its evolution in time on Figure 7. Due to non-linearities, a shock is produced at about $t = 0.2s$.

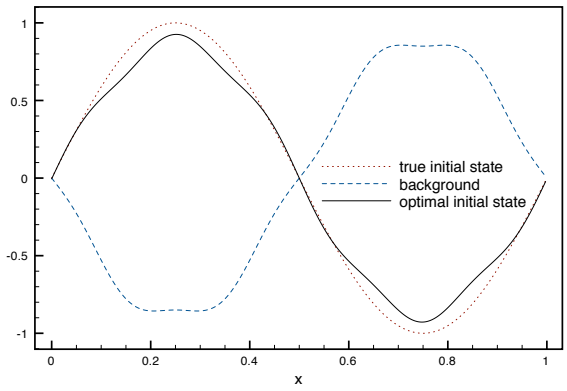

Figure 6. *The true $\mathbf{x}_0^t$, optimal $\hat{\mathbf{x}}_0$ and background $\mathbf{x}^b$ initial states according to x.*

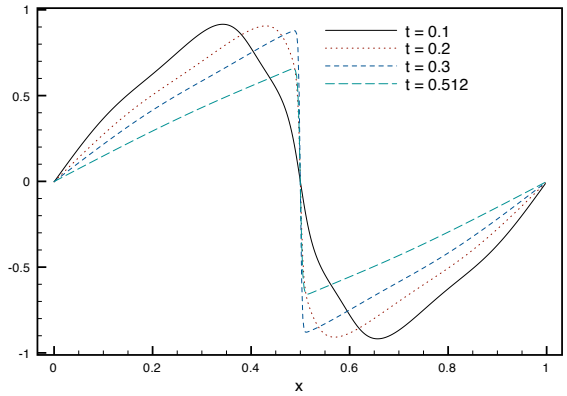

Figure 7. *The evolution in time of the initial optimal state $\hat{\mathbf{x}}_0$ according to x, $M_{0,t}(\hat{\mathbf{x}}_0)$, for different values of time t.*

5.3. Coarse grid approximation

Here we take interest in the errors introduced by the coarse grid approximation. To do so, we compute (see Figure 8) the error between fine and coarse grid models introducing the CGM error (Coarse Grid Model error):

$$CGM(t) = \frac{\|M^f_{0,t}(\mathbf{x}^f_0) - I^f_c M^c_{0,t}(I^c_f \mathbf{x}^f_0)\|}{\|M^f_{0,t}(\mathbf{x}^f_0)\|}(t) \qquad [37]$$

$\mathbf{x}^f_0$ being here the estimated solution computed after a first iteration on fine grid.

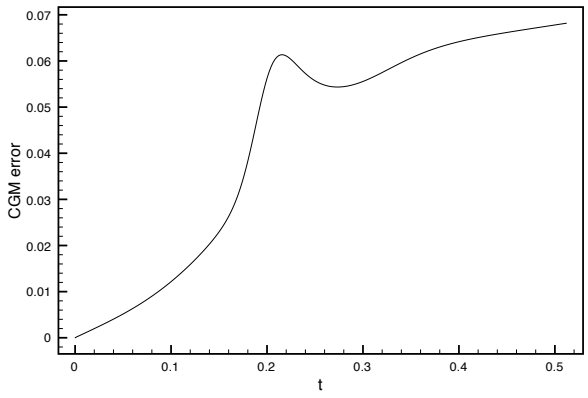

Figure 8. *Coarse grid approximation errors [37] according to the time integration* t.

We observe that the error is increasing until the time $t = 0.2$, where the shock is produced. After that, the error slowly evolves. Note that at this time the solution corresponds to a stationary shock. Non-linearities have an impact on the approximation property but the error remains bounded.

5.4. Critical parameters

We now study the influence of the different parameters involved in the multigrid algorithm: for each cycle, let ν_1 be the number of fine grid iterations before the coarse grid step, and ν_2 be the number of fine grid iterations after the coarse grid step (see the FAS algorithm on section 2.1). For simplicity, we will impose ν_2 to be equal to ν_1 and refer to them as ν: $\nu_1 = \nu_2 = \nu$. So the total number of fine grid iterations per cycle will be equal to $\nu_1 + \nu_2 = 2\nu$.
In order to obtained good properties of convergence, the coarse grid solution has to be solved accurately enough. This accuracy is dictated by the stopping criteria of the M1QN3 procedure:

$$\epsilon = \frac{\|\nabla \mathcal{J}_k\|}{\|\nabla \mathcal{J}_{\text{initial}}\|} \qquad [38]$$

We want to check how much satisfying this condition is important for the method to converge and so we will investigate the behavior of the algorithm for different values of ϵ.

To compare the results, we use as diagnostic variable the root mean square error $RMS(\mathbf{x}_k, \mathbf{x}^t)$ that evaluates the distance between the k^{th} estimated solution $\mathbf{x}_k$ on fine grid and the true one $\mathbf{x}^t$:

$$RMS(\mathbf{x}_k, \mathbf{x}^t) = \sqrt{\frac{1}{N_x.N} \sum_{n=0}^{N-1} \sum_{j=0}^{N_x-1} (\mathbf{x}_n^j - (\mathbf{x}_n^j)^t)^2} \qquad [39]$$

All the following plots show the decrease of this error according to the computational time. This enables the comparisons between the different algorithms by taking into account the coarse grid resolution cost.
We use different values of ν (the number of M1QN3 iterations per cycle on the fine grids): 1, 10, and 100 on Figure 9. For these experiments, the coarse grid problem is solved with high accuracy, using the stopping critera of the M1QN3 procedure (defined equation [38]) equal to $\epsilon = 0.001$.
FAS-100 denotes the use of $\nu = 100$, the same logic holds for FAS-10, and FAS-1. More iterations are done on the fine grids, less are needed on the coarse grid to converge and this globally leads to similar computational times.

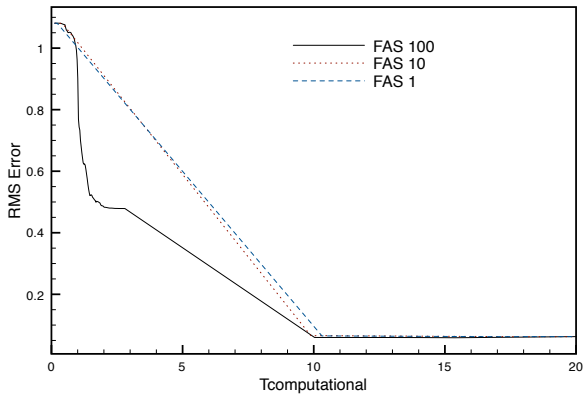

Figure 9. *The decrease of the fine grid RMS error [39] for different numbers of fine grid iterations as a function of the computing time.*

For next experiments, ν will be fixed to 1 which means that only two fine grid iterations are done per multigrid cycle. Note that the cost of one coarse grid correction step is not negligible in our experiment. But in a 2D case with a refinement factor of 3 in time and space, the ratio between high and coarse grid resolution model would equal 27. So that the number of fine grid iterations has to be the smallest possible. According to this argument, the multigrid algorithm could be initialized by starting the minimization on the coarse grid first, which is called the Full-Multigrid configuration. With complex systems this can have a strong beneficial impact on the computational time. However in this paper we focus on the benefit of the coarse grid correction on the fine grid estimation.

We now change the number of coarse grid iterations by modifying the stopping critera ϵ of the M1QN3 procedure : 0.001, 0.1 and 0.7, see Figure 10.

We observe that going until convergence on the coarse grid (small ϵ) is not the choice that

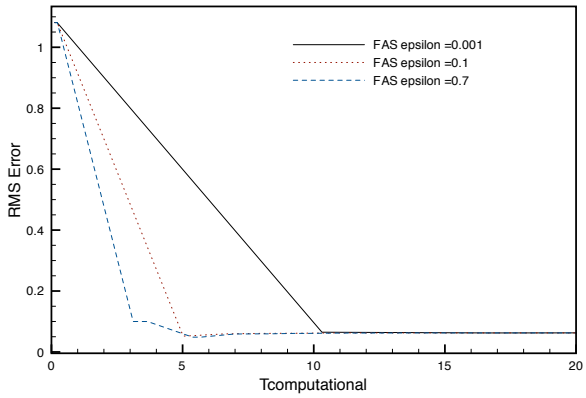

Figure 10. *The decrease of the fine grid RMS error [39] for different stopping critera ϵ of the coarse grid minimization as a function of the computing time.*

minimizes the total cost. For the first cycle, the number of coarse grid iterations is about, 1600 for $\epsilon = 0.001$, 700 with $\epsilon = 0.1$ and 450 with $\epsilon = 0.7$. The choice $\epsilon = 0.1$ is optimal for this experiment. We will then fix ϵ to 0.1.

5.5. Monogrid vs. Multigrid

Using the previous parameters for FAS algorithm ($\nu = 1$ and $\epsilon = 0.1$), we now compare the RMS error [39] decrease for the monogrid algorithm with the multigrid one on Figure 11 and Figure 12. The global behavior shows, on Figure 11, that the multigrids method is about 40 times faster. Moreover, as we noted above, on Figure 12 we see that the FAS algorithm almost reaches convergence after the first coarse grid correction.

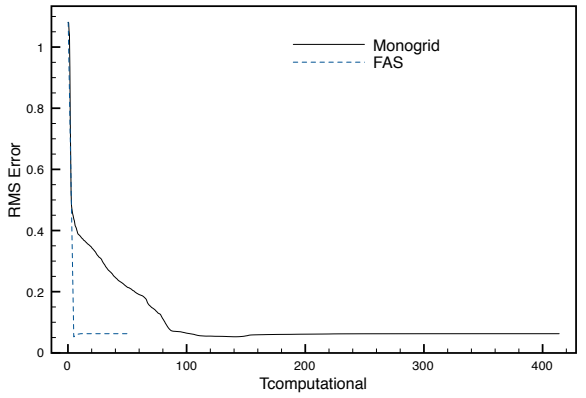

Figure 11. *The decrease of the fine grid RMS error [39] for the monogrid and the multigrid algorithms as a function of the computing time.*

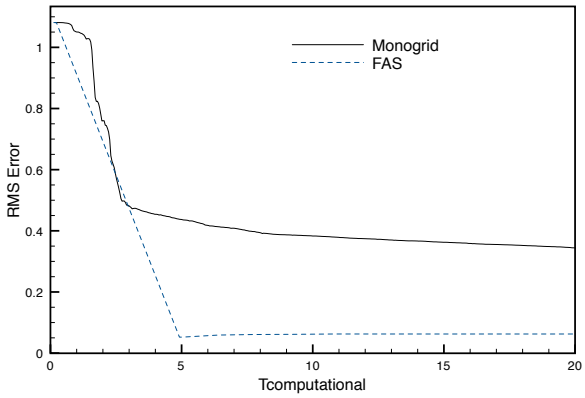

Figure 12. *The decrease of the fine grid RMS error [39] for the monogrid and the multigrid algorithms as a function of the computing time, showing only the beginning of the computation.*

6. Conclusion

We focused on the solution of variational data assimilation problems (4D-var), a large and complex optimal system characterized by the definition of a specific cost function that measures the distance between the model and the observations data set. The multigrid methods are a way of solving a system by using grids at coarser resolutions. Following Nash [12], the idea was to adapt the multigrid methods to data assimilation and more specifically to study its efficiency in the purpose of getting lower computational costs.
On a linear model, we proceeded to a detailed study of the efficiency of the coarse grid correction according to the parameters of the cost function and emphasized the importance of defining a well suited cost function.
More precisely, we showed that the discretization errors affect the ellipticity of the discrete Hessian while the regularization term could counterbalance this negative impact.
For non linear models, multigrid methods can be adapted using the Full-Approximation Scheme (FAS). We applied this algorithm to solve a 4D-Var optimal system based on a Burgers equation and compared the multigrid solution to the monogrid one. The multigrid algorithm proved its robustness to non-linearities and to the definition of the different parameters such as the number of fine and coarse grid iterations. Moreover the multigrid method reached convergence 40 times faster than the monogrid method does. The multigrid methods have shown they can be efficient in reducing the cost of solving the data assimilation optimal systems.
The objectives are now to do further developments in order to be able to apply such techniques to more complex geophysical models.

7. References

[1] ANDREW F. BENNETT. “Inverse Modeling of the Ocean and Atmosphere”, *Cambridge University Press*, 2002.

[2] BRANDT, A. “Multi-level adaptive computations in fluid dynamics.” *AIAA J.*, vol. 18: 1165–1172, 1980.

[3] A. BRANDT. " Multigrid techniques 1984 Guide with applications to fluid dynamics", *Gesellschaft fuer Mathematik und Datenverarbeitung m.b.H. Bonn (GMD), St. Augustin (Germany)*, 1984.

[4] FRANÇOIS-XAVIER LE DIMET AND O. TALAGRAND. " Variational algorithms for analysis and assimilation of meteorological observations: Therotical aspects." *Tellus*, vol. 38(A):97–110, 1986.

[5] DURAN, D. " Numerical Methods for Fluid Dynamics: With Applications to Geophysics" *Springer; 2nd Edition*, vol. 2010.

[6] GILBERT, J-C. AND C. LEMARÉCHAL "Some numerical experiments with variable storage quasi-Newton algorithms." *Mathematical Programming*, vol. B25: 407–435, 1989.

[7] GRATTON, S. AND LAWLESS, A. S. AND NICHOLS, N. K. "Approximate Gauss-Newton methods for non-linear least square problems." *SIAM J. Optimization*, vol. 18: 106–132, 2007.

[8] HACKBUSCH, W. "Multi-grid Methods and Applications." *Springer Series in Computational Mathematics*, vol. 1985.

[9] HEMKER, P.W. "A note on defect correction processes with an approximate inverse of deficient rank." *Appl. Math. Comp.*, vol. 8: 137–139, 1982.

[10] IDE, KAYO AND COURTIER, PHILIPPE AND GHIL, MICHAEL AND LORENC, ANDREW C. "Unified notation for data assimilation: Operational, and Variational." *J. Met. Soc. of Japan*, vol. 75(1B):181-189, 1997.

[11] MCCORMICK, S. F. "An algebraic interpretation of multigrid methods." *SIAM J. Num. Anal.*, vol. 19: 548–560, 1982.

[12] ROBERT MICHAEL LEWIS AND STEPHEN G. NASH. " Model problems for the multigrid optimization of systems governed by differential equations." *SIAM J. SCI. COMPUT.*, vol. 26(6):1811–1837, 2005.

[13] STEPHEN G. NASH. " A multigrid approach to discretized optimization problems." *Journal of Optimization Methods and Software*, vol. 14:99–116, 2000.

[14] SOUTHWELL, R. V "Relaxation Methods in Theoretical Physics. " *Oxford University Press, Technical Report*, vol. 1946.

[15] SOUTHWELL, R. V "Stress-Calculation in Frameworks by the Method of "Systematic Relaxation of Constraints. " *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, vol. 151: 56–95, 1935.

[16] SHLOMO. TA'ASAN. " One shot methods for optimal control of distributed parameters systems i: Finite dimensional control." *Technical Report No 91-2, ICASE Report*, 1991.

[17] SHLOMO. TA'ASAN. " Multigrid one-shot methods and Design Strategy. " *Carnegie Mellon University*, 1997.

[18] TRANGENSTEIN, J., " Numerical solution of hyperbolic partial differential equations. " *Cambridge University Press*, 2007.

[19] ULRICH TROTTENBERG, CORNELIS W. OOSTERLEE, AND ANTON SCHÜLLER. " Multigrid " *Academic Press*, 2001.

[20] YAVNEH, IRAD AND DARDYK, GREGORY "A multilevel nonlinear method " *SIAM J. SCI. COMPUT.*, vol. 28: 24–46, 2006.