Computers & Fluids 46 (2011) 252–256

Contents lists available at ScienceDirect

Computers & Fluids

journal homepage: www.elsevier.com/locate/compfluid

Conditioning and preconditioning of the variational data assimilation problem

S.A. Haben [1], A.S. Lawless [2], N.K. Nichols [*,2]

Department of Mathematics, University of Reading, PO Box 220, Reading RG6 6AX, UK

ARTICLE INFO

Article history:
Received 26 April 2010
Received in revised form 22 August 2010
Accepted 24 November 2010
Available online 30 November 2010

Keywords:
Variational data assimilation
Error covariances
Least-squares optimization
Condition number
Hessian

ABSTRACT

Numerical weather prediction (NWP) centres use numerical models of the atmospheric flow to forecast future weather states from an estimate of the current state. Variational data assimilation (VAR) is used commonly to determine an optimal state estimate that miminizes the errors between observations of the dynamical system and model predictions of the flow. The rate of convergence of the VAR scheme and the sensitivity of the solution to errors in the data are dependent on the condition number of the Hessian of the variational least-squares objective function. The traditional formulation of VAR is ill-conditioned and hence leads to slow convergence and an inaccurate solution. In practice, operational NWP centres precondition the system via a control variable transform to reduce the condition number of the Hessian. In this paper we investigate the conditioning of VAR for a single, periodic, spatially-distributed state variable. We present theoretical bounds on the condition number of the original and preconditioned Hessians and hence demonstrate the improvement produced by the preconditioning. We also investigate theoretically the effect of observation position and error variance on the preconditioned system and show that the problem becomes more ill-conditioned with increasingly dense and accurate observations. Finally, we confirm the theoretical results in an operational setting by giving experimental results from the Met Office variational system.

1. Introduction

Variational data assimilation (VAR) is popularly used in numerical weather and ocean forecasting to combine observations with a model forecast in order to produce a 'best' estimate of the current state of the system and enable accurate prediction of future states. The estimate minimizes a weighted nonlinear least-squares measure of the error between the model forecast and the available observations and is found using an iterative optimization algorithm. Under certain statistical assumptions the solution to the variational data assimilation problem, known as the *analysis*, yields the *maximum a posteriori* Bayesian estimate of the state of the system [7].

In practice an incremental version of VAR is implemented in many operational centres, including the Met Office [11] and the European Centre for Medium-Range Weather Forecasting (ECMWF) [10]. This method solves a sequence of linear least-squares approximations to the nonlinear least-squares problem and is equivalent to an approximate Gauss–Newton method for determining the analysis [6]. Each approximate linear least-squares problem is solved using an 'inner' gradient iteration method, such as the conjugate gradient method, and the linearization state is then updated in an 'outer' iteration loop.

The rate of convergence of the inner loop of the VAR iteration scheme and the sensitivity of the analysis to perturbations in the data of the problem are proportional to the condition number, that is, the ratio of the largest to the smallest eigenvalue, of the Hessian of the linear least-squares objective function [3]. Experimental results indicate that in operational systems the Hessian is ill-conditioned, with undesirable features [8]. Operationally the system is preconditioned by transforming the state variables to new variables where the errors are assumed to be approximately uncorrelated. Preconditioning reduces the sensitivity of the problem to be solved and hence enables a more accurate analysis to be computed [3]. Experimental comparisons have demonstrated that operationally the preconditioning significantly improves the speed and accuracy of the assimilation scheme [2,8].

A variety of explanations are offered in the literature for the ill-conditioning of the VAR problem and for the benefits of preconditioning in the operational system [9,1,12]. In this paper we examine the conditioning and preconditioning of the variational assimilation method theoretically. We give expressions for bounds on the conditioning of the Hessian of the problem in the case of a single, periodic, spatially-distributed system parameter. It is

* Corresponding author. Tel.: +44 0 118 378 8988; fax: +44 0 118 931 3423.
E-mail address: n.k.nichols@reading.ac.uk (N.K. Nichols).

[1] The work of this author was supported by an EPSRC CASE Award with the Met Office. The author is also grateful for support from the European Science Foundation OPTPDE Network on Optimization with PDE Constraints.

[2] The work of these authors was supported in part by the NERC National Centre for Earth Observation.

0045-7930/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compfluid.2010.11.025

assumed that the errors in the initial background states (model forecast) have a Gaussian correlation structure, although other forms of the error correlation structure can be analysed using the same theory. We consider three questions: (i) how does the condition number of the Hessian depend on the length-scale in the correlation structures? (ii) how does the variance of the observation errors affect the conditioning of the Hessian? and (iii) how does the distance between observations, or density of the observations, affect the conditioning of the Hessian?

In the next section we introduce the incremental variational assimilation method. In Section 3 we give bounds on the conditioning of the problem and examine our three questions. In Section 4 we present experimental results obtained using the Met Office Unified Model supporting the theory and in Section 5 we summarize the conclusions. In this paper we present results only for the 3D-variational method, but our techniques can be extended to the 4D-variational scheme and will be discussed in a subsequent paper.

2. Variational data assimilation

The aim of the variational assimilation problem is to find an optimal estimate for the initial state of the system $\mathbf{x}_0$ (the *analysis*) at time t_0 given a *prior* estimate $\mathbf{x}_0^b$, (the *background*) and observations $\mathbf{y}_i$ at times t_i, subject to the nonlinear forecast model given by

$$\mathbf{x}_i = \mathscr{M}(t_i, t_{i-1}, \mathbf{x}_{i-1}), \tag{1}$$

$$\mathbf{y}_i = \mathscr{H}_i(\mathbf{x}_i) + \boldsymbol{\delta}_i, \tag{2}$$

for $i = 0, \ldots, n$. Here $\mathscr{M}$ and $\mathscr{H}_i$ denote the evolution and observation operators of the system. The errors $(\mathbf{x}_0 - \mathbf{x}_0^b)$ in the background and the errors $\boldsymbol{\delta}_i$ in the observations are assumed to be random with mean zero and covariance matrices $\mathbf{B}$ and $\mathbf{R}_i$, respectively. The assimilation problem is then to minimize, with respect to $\mathbf{x}_0$, the objective function

$$\mathscr{J}(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_0^b) + \frac{1}{2}\sum_{i=0}^{n}(\mathscr{H}_i(\mathbf{x}_i) - \mathbf{y}_i)^T \mathbf{R}_i^{-1}(\mathscr{H}_i(\mathbf{x}_i) - \mathbf{y}_i), \tag{3}$$

subject to the model forecast Eqs. (1) and (2). If observations are given at several points t_i, $i = 0, 1, \ldots, n$ over a time window $[t_0, t_n]$ with $n > 0$, the assimilation scheme is known as the four-dimensional variational method (4DVar). If observations are given only at the initial time with $n = 0$, then the optimization problem reduces to the three-dimensional data assimilation problem (3DVar).

In practice, to improve the computational efficiency of the variational assimilation procedure, a sequence of linear least-squares approximations to the nonlinear least-squares problem (3) is solved. Given the current estimate of the analysis $\mathbf{x}_0$, the nonlinear objective function is linearized about the corresponding model trajectory $\mathbf{x}_i$, $i = 1, \ldots, n$, satisfying the nonlinear forecast model. An increment $\delta\mathbf{x}_0$ to the current estimate of the analysis is then calculated by minimizing the linearized least-squares objective function subject to the linearized model equations. The linear least-squares minimization problem is solved in an inner loop by a gradient iteration method. The current estimate of the analysis is then updated with the computed increment and the process is repeated in the outer loop of the algorithm. This data assimilation scheme is known as incremental variational assimilation.

In each outer loop the incremental scheme minimizes, with respect to $\delta\mathbf{x}_0$, the current linearized least-squares objective function, which may be written as

$$\begin{aligned}\tilde{\mathscr{J}}[\delta\mathbf{x}_0] = {} & \frac{1}{2}[\delta\mathbf{x}_0 - (\mathbf{x}_0^b - \mathbf{x}_0)]^T \mathbf{B}^{-1}[\delta\mathbf{x}_0 - (\mathbf{x}_0^b - \mathbf{x}_0)] \\ & + \frac{1}{2}(\widehat{\mathbf{H}}\delta\mathbf{x}_0 - \hat{\mathbf{d}})^T \widehat{\mathbf{R}}^{-1}(\widehat{\mathbf{H}}\delta\mathbf{x}_0 - \hat{\mathbf{d}}),\end{aligned} \tag{4}$$

subject to the linearized model equations

$$\delta\mathbf{x}_i = \mathbf{M}(t_i, t_{i-1})\delta\mathbf{x}_{i-1}, \tag{5}$$

where

$$\widehat{\mathbf{H}} = \left[\mathbf{H}_0^T, (\mathbf{H}_1\mathbf{M}(t_1, t_0))^T, \ldots, (\mathbf{H}_n\mathbf{M}(t_n, t_0))^T\right]^T,$$

$$\hat{\mathbf{d}}^T = \left[\mathbf{d}_0^T, \mathbf{d}_1^T, \ldots, \mathbf{d}_n^T\right], \quad \text{with} \quad \mathbf{d}_i = \mathbf{y}_i - \mathscr{H}_i(\mathbf{x}_i).$$

The matrices $\mathbf{M}(t_i, t_0)$ and $\mathbf{H}_i$ are linearizations of the evolution and observation operators $\mathscr{M}(t_i, t_0, \mathbf{x}_0)$ and $\mathscr{H}_i(\mathbf{x}_i)$ about the current estimated state trajectory $\mathbf{x}_i$, $i = 0, \ldots, n$ and $\widehat{\mathbf{R}}$ is a block diagonal matrix with diagonal blocks equal to $\mathbf{R}_i$.

The minimizer of (4) is also the solution to $\nabla\tilde{\mathscr{J}} = 0$, which may be written explicitly as the linear system

$$(\mathbf{B}^{-1} + \widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}})\delta\mathbf{x}_0 = \mathbf{B}^{-1}(\mathbf{x}_0^b - \mathbf{x}_0) + \widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\hat{\mathbf{d}}. \tag{6}$$

Iterative gradient methods are used to solve (4), or equivalently (6). The gradients are found by an adjoint procedure.

3. Conditioning of the assimilation problem

A measure of the accuracy and efficiency with which the data assimilation problem can be solved is given by the *condition number* of the Hessian matrix

$$\mathbf{A} = (\mathbf{B}^{-1} + \widehat{\mathbf{H}}^T\widehat{\mathbf{R}}^{-1}\widehat{\mathbf{H}}) \tag{7}$$

of the linearized objective function (4). Our aim here is to present explicit bounds on the condition number of $\mathbf{A}$ and investigate its properties in terms of the background and observation error covariance matrices $\mathbf{B}$ and $\widehat{\mathbf{R}}$.

The condition number of the Hessian, which is a square, symmetric, positive definite matrix, is defined in the L_2-norm by

$$\kappa(\mathbf{A}) = ||\mathbf{A}||_2||\mathbf{A}^{-1}||_2 \equiv \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})}, \tag{8}$$

where $\lambda(\mathbf{A})$ denotes an eigenvalue of the matrix. The condition number measures the sensitivity of the solution to the linearized least-squares problem (4), or equivalently the solution to the gradient Eq. (6), to perturbations in the data of the problem. If the condition number of the Hessian, $\kappa(\mathbf{A})$, is very large, the problem is 'ill-conditioned' and, even for small perturbations to the system, the relative error in the solution may be extremely large. For the gradient methods that are commonly used to solve the problem, such as the conjugate gradient method, the rate of convergence then may also be very slow.

Here we consider specifically the conditioning of the 3DVar linearized least-squares problem. In this case observations are given at only one point in time and $\widehat{\mathbf{H}} = \mathbf{H} \equiv \mathbf{H}_0$. We consider in theory the case of a single periodic system parameter with background error variance σ_b^2 and uncorrelated observation errors with variance σ_o^2.

3.1. Conditioning of the background error covariance matrix

We write the background error covariance in the form $\mathbf{B} = \sigma_b^2\mathbf{C}$, where $\mathbf{C}$ denotes the correlation structure of the background errors. The condition number $\kappa(\mathbf{B})$ then equals the condition number $\kappa(\mathbf{C})$. We assume that the correlation structure is homogeneous, where the correlations depend only on distance between states and not position. Under these conditions the correlation matrices used commonly in practice have a circulant structure [4], which we exploit to obtain our theoretical bounds. For example, the Gaussian, Markov and SOAR correlation matrices have this structure, as do those based on Laplacian smoothing. A circulant matrix