The University of Reading

Inner loop stopping criteria for incremental four-dimensional variational data assimilation

A.S. Lawless an N.K. Nichols

NUMERICAL ANALYSIS REPORT 5/05

Department of Mathematics

The University of Reading

Whiteknights, PO Box 220

Reading

Berkshire RG6 6AX

Department of Mathematics

Abstract

Incremental four-dimensional variational data assimilation is a method that solves the assimilation problem by minimizing a sequence of approximate 'inner loop' functions. In any implementation of such a scheme a decision must be made as to how accurately to solve each of the inner minimization problems. In this paper we apply theory that we have recently developed to derive a new stopping criterion for the inner loop minimizations, that guarantees convergence of the outer loops. This new criterion is shown to give improved convergence compared to other commonly used inner loop stopping criteria.

Contents

1 Introduction

Four-dimensional variational data assimilation (4D-Var) is a method to estimate the model trajectory most consistent with the available observational data over a particular time window through the minimization of a nonlinear cost function. A major advantage of this technique over previously used asssimilation methods is that the dynamical forecasting model is used as a constraint on the assimilation. Early studies showed how this allows a 4D-Var assimilation to infer information from observations in a dynamically consistent way (for example, Courtier and Talagrand 1987, Thépaut and Courtier 1991, Rabier and Courtier 1992, Thépaut *et al.* 1993). However, in its full, nonlinear formulation 4D-Var is very computationally demanding and so methods of simplification are needed before operational implementation is possible. The most successful of these used currently is the incremental

pd, Rbpteiraationallyal.

appPlyree by ay pp

2 M

where $\mathbf{H}_i$ is the linearization of the observation operator H_i around the state $\mathbf{x}_i^{(k)}$ at time t_i and $\mathbf{L}(t_i, t_0, \mathbf{x}^{(k)})$ is the solution operator of a linear model linearized around the nonlinear model trajectory.

  - Update $\mathbf{x}_0^{(k+1)} = \mathbf{x}_0^{(k)} + \delta\mathbf{x}_0^{(k)}$

- Set analysis $\mathbf{x}^a = \mathbf{x}_0^{(K)}$

For an exact method the linear model $\mathbf{L}(t_i, t_0, \mathbf{x}^{(k)})$ is equal to the linearization of the discrete nonlinear model $\mathcal{S}(t_i, t_0, \mathbf{x}_0)$, but in practice an approximate linearization is often used (for example, Lorenc *et al.* 2000, Mahfouf and Rabier 2000). It was shown by Lawless *et al.* (2005a, 2005b) that incremental 4D-Var with an exact tangent linear model is equivalent to a Gauss-Newton method applied to minimize the nonlinear cost function (1). We now review this equivalence.

2.2 Gauss-Newton algorithm

The Gauss-Newton method is an approximation to a Newton iteration, in which the second order terms of the Hessian are neglected (Dennis and Schnabel, 1996). To illustrate this we consider a general nonlinear least squares problem

$$\min_{\mathbf{x}} \mathcal{J}(\mathbf{x}) = \frac{1}{2} \parallel \mathbf{f}(\mathbf{x}) \parallel_2^2 = \frac{1}{2}\mathbf{f}(\mathbf{x})^{\mathrm{T}}\mathbf{f}(\mathbf{x}), \tag{6}$$

with $\mathbf{x} \in \mathbb{R}^n$, which we assume to have an isolated local minimum $\mathbf{x}^*$. We write

$$\nabla\mathcal{J}(\mathbf{x}) = \mathbf{J}^{\mathrm{T}}\mathbf{f}(\mathbf{x}), \tag{7}$$

$$\nabla^2\mathcal{J}(\mathbf{x}) = \mathbf{J}^{\mathrm{T}}\mathbf{J} + Q(\mathbf{x}), \tag{8}$$

f(x) Tle c t a Gauss-Newto wi peratio h r imml m mig (6 n i ic h f ic g ge

To see the link with incremental 4D-Var, we note that (1) can be written in the form (6) by putting

$$\mathbf{f}(\mathbf{x}_0) = \begin{pmatrix} \mathbf{B}_0^{-1/2}(\mathbf{x}_0 - \mathbf{x}^b) \\ \mathbf{R}_0^{-1/2}(H_0[\mathbf{x}_0] - \mathbf{y}_0^o) \\ \vdots \\ \mathbf{R}_n^{-1/2}(H_n[\mathbf{x}_n] - \mathbf{y}_n^o) \end{pmatrix}. \quad (12)$$

Then, as was shown by Lawless *et al.* (2005a, 2005b), if we apply the Gauss-Newton iteration to the nonlinear cost function (1), the step given by (11) is equivalent to the minimization of the inner loop cost function (3). Thus incremental 4D-Var is equivalent to a Gauss-Newton iteration and we can use the theory for the Gauss-Newton method to analyse the behaviour of incremental 4D-Var.

2.3 The truncated algorithm

In practice the inner loop cost function of incremental 4D-Var is minimized using an iterative procedure, such as a conjugate gradient method. Such minimization methods employ a stopping criterion to determine when the solution has been found to sufficient accuracy. Hence the minimum of the inner cost function is not found exactly, but only to within the degree of accuracy determined by the stopping criterion. In the context of the Gauss-Newton algorithm this can be considered as a method in which the step (9) of the algorithm is solved inexactly. Thus we obtain the truncated Gauss-Newton (TGN) algorithm

$$\text{Solve for } \delta\mathbf{x}^{(k)}: \quad (\mathbf{J}(\mathbf{x}^{(k)})^{\mathrm{T}}\mathbf{J}(\mathbf{x}^{(k)}))\delta\mathbf{x}^{(k)} = -[\mathbf{J}(\mathbf{x}^{(k)})^{\mathrm{T}}\mathbf{f}(\mathbf{x}^{(k)}) + \mathbf{r}^{(k)}], \quad (13)$$

$$\text{Update:} \quad \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta\mathbf{x}^{(k)}, \quad (14)$$

where $\mathbf{r}^{(k)}$ is the residual arising from the premature termination of the inner loop minimization. Convergence of the TGN algorithm is guaranteed by the following theorem, first stated in Lawless *et al.* (2005b) and proved in Gratton *et al.* (2004):

Theorem 1 *Assume that $\hat{\beta} < 1$ and that on each iteration the Gauss-Newton method is truncated with*

$$\| \mathbf{r}^{(k)} \|_2 \leq \beta_k \| \mathbf{J}(\mathbf{x}^{(k)})^T \mathbf{f}(\mathbf{x}^{(k)}) \|_2, \quad (15)$$

where

$$\beta_k \leq \frac{\hat{\beta} - \| (\mathbf{J}^T(\mathbf{x}^{(k)})\mathbf{J}(\mathbf{x}^{(k)}))^{-1}Q(\mathbf{x}^{(k)}) \|_2}{1 + \| (\mathbf{J}^T(\mathbf{x}^{(k)})\mathbf{J}(\mathbf{x}^{(k)}))^{-1}Q(\mathbf{x}^{(k)}) \|_2}. \tag{16}$$

Then there exists $\eta \quad 0$ *such that, if* $\| \mathbf{x}_0 - \mathbf{x}^* \|_2 \leq \eta$*, the truncated Gauss-Newton iteration (TGN) converges to the solution* $\mathbf{x}^*$ *of the nonlinear least squares problem (6).*

Proof: See Gratton *et al.* (2004). □

In the next section we demonstrate how this theorem may be used to determine a suitable stopping criterion for the inner loop of incremental 4D-Var.

Defining a suitable stopping criterion

As stated in the introduction, it is important to choose carefully the stopping criterion for the inner loop minimization of incremental 4D-Var. If too few inner iterations are performed, then the outer loop iterations may not converge to the solution of the original problem. Although this may not appear to be a problem if we are not running many outer loops, it does mean that with too few inner iterations the system may diverge and the final analysis may be further from the truth than the first guess. Hence even in the case where only a few outer loops are performed, adequate minimization of the inner loop is necessary. However, we also wish to avoid performing too many iterations on the inner loop. If the current outer iterate $\mathbf{x}^{(k)}$ is far from the true solution, then iteration of the inner minimization to too high an accuracy may result in extra computational work which does not lead to increased accuracy in the outer loops. We discuss some common ways in which the inner loop minimization is stopped and then use the theory of Section 2 to propose a new stopping criterion.

3.1 Choice of criterion

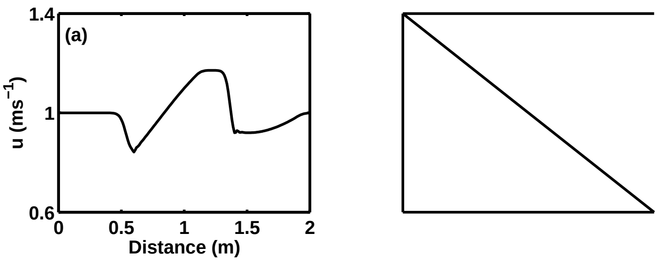
Since we know that the gradient of a function is zero at its minimum, a natural method for stopping the inner iteration is to stop when the norm of the gradient of the inner loop cost function falls below a specified tolerance. If we use ϵ to denote a user-set tolerance and use subscript m to denote the iteration count of the inner loop, then we can write this criterion as

(C1) Absolute norm of gradient

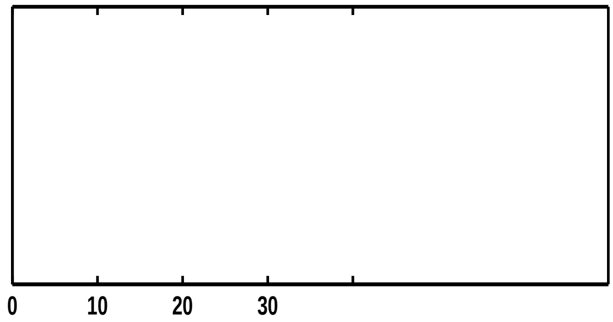
$$\| \nabla \tilde{\mathcal{J}}^{(k)}_{(m)} \|_2 < \epsilon. \tag{17}$$

An alternative stopping criterion from optimization theory is to stop when the relative change in the inner loop cost function itself is less than a given tolerance (Gill *et al.*

A comparison of this with (13) reveals that $\mathbf{r}^{(k)}_{(M)} = \nabla$ ˜


(a)
1.4
1
0.6
u (ms⁻¹)
0
0.5
1
1.5
2
Distance (m)

We first consider the relative computational costs of the three experiments, noting that for this system, with the inner loop at the same spatial resolution as the outer loop, the inner and outer iterations are of comparable cost. We see from Figure 2 that the experiment using criterion (C1) takes the most iterations to converge. After the first few iterations of each inner loop the function value remains almost constant, while the gradient norm continues to decrease. Thus many inner iterations are performed which have little eff


0
10
20
30

with the least amount of work.

5.3 Premature termination of outer loops

In practical data assimilation the outer loops are not usually run to complete convergence, so it is important to understand the effect of the different inner loop stopping criteria where the outer loops are stopped prematurely. For example, if limited computing resources are available, then it may be necessary to stop the assimilation after a given amount of computation. If the inner and outer iterations are of similar cost, this corresponds to stopping after a fixed number of iterations. From Figures 2 and 3 we see that after the first few iterations the assimilation using stopping criterion (C3) always gives a lower value of the cost function than the assimilations using the other criteria. It is possible in such a case that the inner-loop gradient norm will be lowest in the assimilation using stopping criterion (C1), for example if the experiment with imperfect observations is stopped after 10 iterations. However, when the solution to the inner problem is added onto the linearization state to give a new outer iterate, the gradient of the nonlinear cost function may be very different. In fact, we see that if we choose to stop the assimilation after a fixed number of outer loops (as is often done in practice), then not only is the cost function highest for the experiment using criterion (C1), but the gradient norm is also highest for this experiment.

5.4 Summary

From the experiments in this section the new stopping criterion (C3) is seen to be the most useful of the three convergence criteria proposed. It avoids the performance of excessive inner iterations, while providing faster convergence of the outer loop cost function and gradient. This conclusion seems to hold whether the outer loops are converged or whether they are stopped after a fixed amount of computational work. Hence the numerical experiments support the conclusion from the Gauss-Newton theory that a relative gradient stopping criterion is a more natural way of stopping the inner loop in incremental 4D-Var. However, the usefulness of this criterion depends on a good choice of the tolerance. We consider this question in

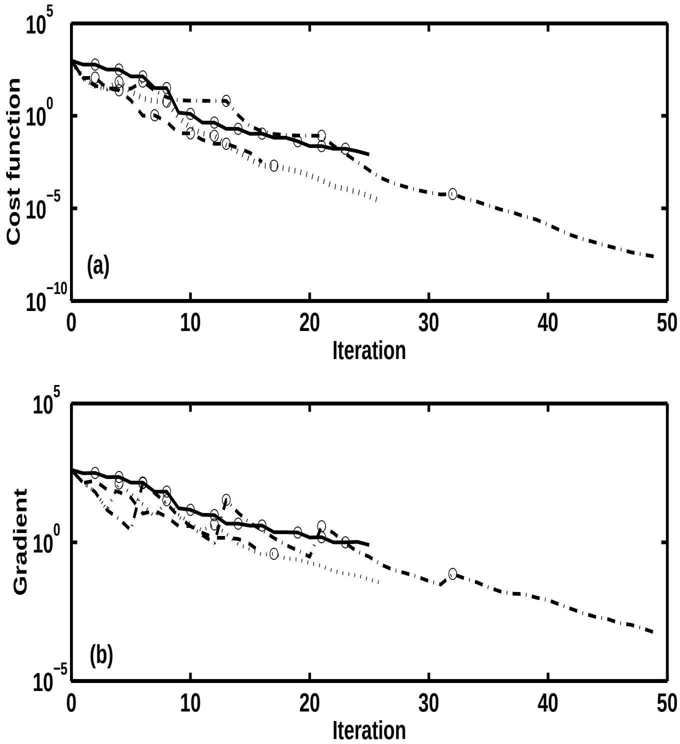
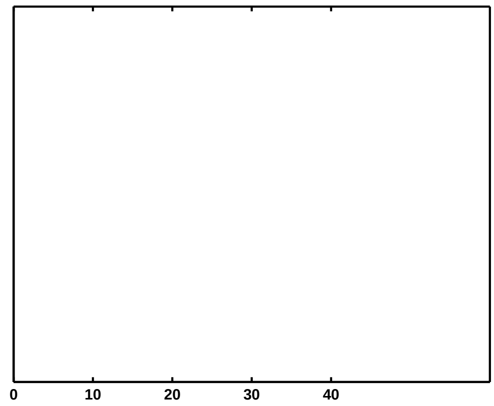
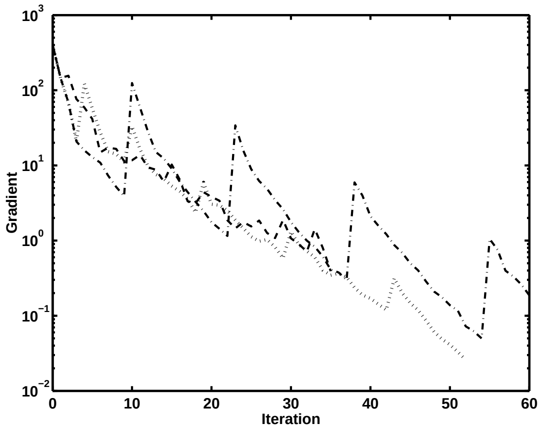
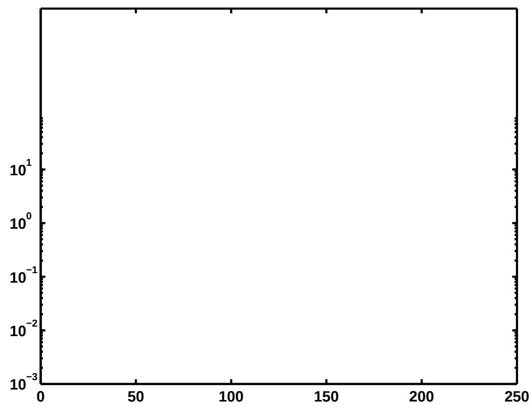


Figure 4: Convergence of (a) cost function and (b) gradient for experiment with

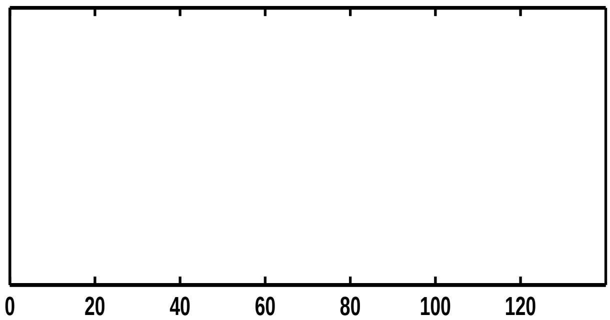

0
10
20
30
40



Gradient
10^3
10^2
10^1
10^0
10^{-1}
10^{-2}
0
10
20
30
40
50
60
Iteration



10^1
10^0
10^−1
10^−2
10^−3
0
50
100
150
200
250


0
20
40
60
80
100
120

then the final solution may not be close to the minimum of the nonlinear cost function. However, if the inner minimization is solved to too great an accuracy, much computational effort is wasted for very little gain. In this study we have proposed

forecast experiments with a four-dimensional variational assimilation system. *Q. J. R. Meteorol. Soc.*, **124,** 1861–1887.

Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A., (2000). The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.*, **126,** 1143–1170.

Thépaut, J.N. and Courtier, ., (1991). Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. *Q. J. R. Meteorol. Soc.*, **117,** 1225–1254.

Thépaut, J.N., Hoffman, R.N. and Courtier, ., (1993). Interactions of dynamics and observations in a four-dimensional variational assimilation. *Mon. Wea. Rev.*, **121,** 3393–3414.

Weaver, A.T., Vialard, J. and Anderson, D.L.T., (2003). Three and four-dimensional variational assimilation with a general circulation model of the Tropical acific Ocean. art I: Formulation, internal diagnostics, and consistency checks. *Mon. Wea. Rev.*, [illegible]