

Application of the Quasi-Inverse Method to Data Assimilation

EUGENIA KALNAY

Department of Meteorology, University of Maryland, College Park, Maryland

SEON KI PARK*

Cooperative Institute for Mesoscale Meteorological Studies, School of Meteorology, University of Oklahoma, Norman, Oklahoma

ZHAO-XIA PU

NASA Goddard Space Flight Center, Greenbelt, Maryland

JIDONG GAO

Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma

(Manuscript received 25 September 1998, in final form 17 March 1999)

ABSTRACT

Four-dimensional variational data assimilation (4D-Var) seeks to find an optimal initial field that minimizes a cost function defined as the squared distance between model solutions and observations within an assimilation window. For a perfect linear model, Lorenc showed that the 4D-Var forecast at the end of the window coincides with a Kalman filter analysis if two conditions are fulfilled: (a) addition to the cost function of a term that measures the distance to the background at the beginning of the assimilation window, and (b) use of the Kalman filter background error covariance in this term. The standard 4D-Var requires minimization algorithms along with adjoint models to compute gradient information needed for the minimization. In this study, an alternative method is suggested based on the use of the quasi-inverse model that, for certain applications, may help accelerate the solution of problems close to 4D-Var.

The quasi-inverse approach for the forecast sensitivity problem is introduced, and then a closely related variational assimilation problem using the quasi-inverse model is formulated (i.e., the model is integrated backward but changing the sign of the dissipation terms). It is shown that if the cost function has no background term, and has a complete set of observations (as assumed in many classical 4D-Var papers), the new method solves the 4D-Var-minimization problem efficiently, and is in fact equivalent to the Newton algorithm but without having to compute a Hessian. If the background term is included but computed at the end of the interval, allowing the use of observations that are not complete, the minimization can still be carried out very efficiently. In this case, however, the method is much closer to a 3D-Var formulation in which the analysis is attained through a model integration. For this reason, the method is called "inverse 3D-Var" (I3D-Var).

The I3D-Var method was applied to simple models (viscous Burgers' equation and Lorenz model), and it was found that when the background term is ignored and complete fields of noisy observations are available at multiple times, the inverse 3D-Var method minimizes the same cost function as 4D-Var but converges much faster. Tests with the Advanced Regional Prediction System (ARPS) indicate that I3D-Var is about twice as fast as the adjoint Newton method and many times faster than the quasi-Newton LBFGS algorithm, which uses the adjoint model. Potential problems (including the growth of random errors during the integration back in time) and possible applications to preconditioning, and to problems such as storm-scale data assimilation and reanalysis are also discussed.

1. Introduction

Over the last decade many important applications of the backward integration of the adjoint of the linear

tangent model have been introduced in the literature. These include the generation of singular vectors for ensemble prediction (e.g., Molteni et al. 1996), four-dimensional variational data assimilation (e.g., Lewis and Derber 1985; Le Dimet and Talagrand 1986; Courtier et al. 1994), forecast sensitivity to the initial conditions (Rabier et al. 1996; Pu et al. 1997a), and targeted observations (e.g., Rohaly et al. 1998; Pu et al. 1998).

Among advanced methods for data assimilation, four-dimensional variational data assimilation (4D-Var) is the approach that has received the most attention in recent years (e.g., Derber 1989; Courtier et al. 1994; M. Zupanski 1993). A simplified version has been imple-

* Current affiliation: Department of Meteorology, University of Maryland, College Park, Maryland.

Corresponding author address: Prof. Eugenia Kalnay, Department of Meteorology, University of Maryland 3431, Computer and Space Sciences Building, College Park, MD 20742-2425.
E-mail: ekalnay@atmos.umd.edu

mented recently at the European Centre for Medium-Range Weather Forecasts (ECMWF; Bouttier and Rabier 1997; Rabier et al. 1997), and at the time of this writing, the National Centers for Environmental Prediction (NCEP) is testing a 4D-Var system for the Eta Model.

In 4D-Var a cost function is defined as the squared distance between a model integration and the observations in a given assimilation interval. Lorenc (1986, 1988) showed that for a linear perfect model, if (a) a background error term is added to the cost function at the beginning of the assimilation period, and (b) the background error covariance is the same as that used in the Kalman filter (KF) at the initial time, then the 4D-Var analysis at the end of the interval is the same that would be obtained using the KF. This makes 4D-Var attractive, because it is much less expensive than KF (see also Daley 1991; Thepaut et al. 1993).

4D-Var provides initial conditions for a model integration that is close to the observations, but it also has some disadvantages.

- 1) It is difficult to include forecast error covariances in the cost function except at the beginning of the interval, which forces the use of short assimilation intervals in order to maintain the impact of model errors small. It is obvious from Lorenz chaos theory that even with a perfect model, one would not want to perform 4D-Var over, for example, a 2-week data assimilation interval, since the 4D-Var analysis would be given by the state of the model after a 2-week integration, when the predictability has been lost (Pires et al. 1996). Even if the assimilation interval is reduced to a shorter period, such as 6–24 h, the neglect of model errors during the forecast can lead to unrealistic results (Menard and Daley 1996). There have been attempts to include simple evolving model errors (e.g., Derber 1989; D. Zupanski 1997), but much remains to be done in this area.
- 2) 4D-Var has a large computational cost compared to 3D-Var (typically 10–100 or more iterations are required for convergence, equivalent to about 30–300 model integrations per day). ECMWF, for example, has a powerful supercomputer about 25 times faster than a Cray C90, and has been running a model at a horizontal resolution of T213. Nevertheless, ECMWF had to make several simplifying assumptions in their implementation of 4D-Var (such as using a lower horizontal resolution model of T63 and a short assimilation window) in order to reduce the computational cost (Bouttier and Rabier 1997; Rabier et al. 1997).

Recently Wang et al. (1997) suggested the use of backward model integrations in order to accelerate the convergence of 4D-Var (without including background error in the cost function). Pu et al. (1997a) showed that for the problem of forecast sensitivity, closely related

to 4D-Var, a backward integration with the “quasi-inverse” of the tangent linear model (TLM) gave results far superior to those obtained using the adjoint model. The quasi-inverse model is simply the model integrated backward but changing the sign of dissipative terms in order to avoid computational blowup. It can be applied to either the tangent linear or the full nonlinear model, each of which has advantages for different applications. The quasi-inverse linear (QIL) method has been tested successfully at NCEP in several different applications, for example, forecast error sensitivity analysis and data assimilation (Pu et al. 1997a), and adaptive observations (Pu et al. 1998).

Wang et al. (1997) adopted the quasi-inverse approach for their adjoint Newton algorithm (ANA), and applied it to a simplified 4D-Var problem, using simulated data, using the Advanced Regional Prediction System (ARPS; Xue et al. 1995), with impressive results. They assumed the availability of a complete set of observations at the end of the assimilation interval, and showed that the ANA converged in an order of magnitude fewer iterations, and to an error level an order of magnitude smaller than the conventional adjoint approach to solve the same (simplified) 4D-Var problem.

Kalnay and Pu (1998) generalized the Wang et al. (1997) approach by including a background term to the cost function and further simplified the method. The background error term in the cost function allows using incomplete sets of observations, but in order to maintain the efficiency of the method, it is necessary to estimate the background error term at the end of the assimilation interval, rather than at the beginning as in the Lorenc (1986) formulation. We have further generalized the method to allow for the assimilation of data at different times, rather than only at the end of the interval, as in Wang et al. (1997). The results suggest that the quasi-inverse model may be used in data assimilation for accelerating convergence and reducing spinup problems, although problems may arise when tested on comprehensive atmospheric models.

In this paper we first introduce the quasi-inverse approach for the forecast sensitivity problem, and then formulate a closely related variational assimilation problem using the quasi-inverse model (section 2). We show that if the cost function has no background term, and has a complete set of observations (as was assumed in many classical 4D-Var papers), the new method solves the 4D-Var-minimization problem efficiently, and is in fact equivalent to the Newton algorithm but without having to compute a Hessian. If the background term is included but computed at the end of the interval, the minimization can still be carried out very efficiently, but in this case the method is closer to a 3D-Var formulation in which the analysis is attained through a model integration. For this reason, we call the method “inverse 3D-Var” (I3D-Var).

In section 3 we introduce a simple “model” (viscous Burgers’ equation), which includes effects mimicking

the three main components of atmospheric models: advection, large-scale instabilities, and dissipative processes. Using this simple model we show the effects of applying a linear tangent model, its adjoint, the exact inverse, and the quasi-inverse model, and compare one iteration of inverse 3D-Var and adjoint 4D-Var. In section 4 we present preliminary results comparing inverse 3D-Var and the adjoint 4D-Var using Burgers' model and Lorenz's model (Lorenz 1963). We show that when the background term is ignored and complete fields of noisy observations are available at multiple times, the inverse 3D-Var method still minimizes the same cost function as 4D-Var (but much more efficiently). Section 5 discusses several topics related to possible applications of inverse 3D-Var: assimilation of data at multiple time levels, research on a storm-scale model with reversible clouds for storm-forecast initialization, and the problem of random observational errors, which may amplify during the backward integration (Reynolds and Palmer 1998).

2. Formulation of inverse 3D-Var

a. Forecast sensitivity

We introduce inverse 3D-Var by first considering the forecast sensitivity problem posed by Rabier et al. (1996): Find the change in initial conditions $\delta\mathbf{x}_0$ that "optimally" corrects a perceived forecast error at the final time t . In what follows, M is a nonlinear forecast, that is, $\mathbf{x}_t = M(\mathbf{x}_0)$, A is the analysis, $E = M - A$ is the perceived error, L is the linear propagator (linear tangent model integrated forward in time), and L^* is its adjoint with respect to the metric used in the definition of the inner product: $\langle \mathbf{x}, L\mathbf{y} \rangle = \langle L^*\mathbf{x}, \mathbf{y} \rangle$ for any pair of vectors \mathbf{x} , \mathbf{y} . Then $\delta\mathbf{x}_0$ is the solution of

$$M(\mathbf{x}_0 + \delta\mathbf{x}_0) = A \quad \text{or} \quad (1)$$

$$\delta\mathbf{x}_t = L \delta\mathbf{x}_0 \cong M(\mathbf{x}_0) - A = E. \quad (2)$$

1) ADJOINT APPROACH (RABIER ET AL. 1996; PU ET AL. 1997B)

In the standard adjoint approach, we define an error cost function using, for example, an energy norm, that is, defining the inner product such that the norm of a vector \mathbf{x} is given by $\langle \mathbf{x}, \mathbf{x} \rangle = \mathbf{x}^T W^2 \mathbf{x}$, the total energy of a state vector \mathbf{x} . With this inner product, $\langle \mathbf{x}, L\mathbf{y} \rangle = \langle L^*\mathbf{x}, \mathbf{y} \rangle$ defines the adjoint of L with respect to the total energy norm $L^* = W^{-2} L^T W^2$, where L^T is the adjoint of L with respect to the Eulerian norm (transpose of L). The error cost function is then

$$J = \frac{1}{2} \langle M(\mathbf{x}_0) - A, M(\mathbf{x}_0) - A \rangle. \quad (3a)$$

A perturbation $\delta\mathbf{x}_0$ in the initial conditions will lead to a change in the cost function

$$\begin{aligned} \delta J &= \langle M(\mathbf{x}_0) - A, \delta\mathbf{x}_t \rangle = \langle M(\mathbf{x}_0) - A, L \delta\mathbf{x}_0 \rangle \\ &= \langle L^*[M(\mathbf{x}_0) - A], \delta\mathbf{x}_0 \rangle = \langle L^*E, \delta\mathbf{x}_0 \rangle. \end{aligned} \quad (3b)$$

Since by definition $\delta J = \langle \nabla J(\mathbf{x}_0), \delta\mathbf{x}_0 \rangle$, the gradient of the cost function with respect to the initial conditions is given by

$$\nabla J(\mathbf{x}_0) = L^* L \delta\mathbf{x}_0 = L^* E. \quad (4)$$

Equation (4) indicates that to obtain the gradient of the cost function we have to integrate backward with the adjoint model, starting from the perceived error at the final time. The negative gradient gives an "optimal" descent direction that results in the maximum decrease of the cost function for a given size perturbation. As pointed out by Rabier et al. (1996), with this definition of inner product, the gradient of J has the same units as the state vector \mathbf{x} , but it depends on the choice of norm. The adjoint procedure requires in addition an estimation of an appropriate amplitude α , after which the adjoint sensitivity correction becomes

$$\delta\mathbf{x}_0 = \alpha \nabla J(\mathbf{x}_0). \quad (5)$$

This correction can be plugged into (1) and the whole procedure iterated (Pu et al. 1997a).

2) QUASI-INVERSE APPROACH

In the QIL approach we try to solve (2) directly, but in order to do so, we need to have an approximation of the inverse of the TLM:

$$\delta\mathbf{x}_0 = L^{-1} E. \quad (6)$$

The QIL approximation to L^{-1} consists of simply running the TLM backward (changing the sign of Δt , and also changing the sign of the dissipative terms to avoid computational blowup). Pu et al. (1997a) found that this is a rather accurate approximation to the dry-dynamics inverse model. It solves a deterministic problem, so that there is no need to find an optimal amplitude, as required by the adjoint method. Reynolds and Palmer (1998) compared this method with running the exact inverse (using a Runge-Kutta time scheme and no change in sign for dissipation). They found that the presence of dissipation during the backward integration had a beneficial effect of a small reduction in noisiness.

Note that the inverse solution is not "optimal" like the adjoint solution, since there is no constraint on the size of $\delta\mathbf{x}$. However, the inverse approach can be considered to be "perfect": it reaches in a single step the same solution ($\nabla J \cong 0$) that the adjoint approach would reach after many iterations. It should be pointed out that Lorenc (1988) integrated the NCEP nested grid model backward with a change of sign in the physics, but in his experiments he was attempting to approximate the adjoint model, not the inverse model.

b. Inverse 3D-Var

Wang et al. (1997) were trying to solve a 4D-Var problem on what they denote the “estimated Newton descent direction” rather than the descent direction provided by the standard adjoint approach. For this purpose they needed to approximate the inverse TLM, and succeeded by adopting the QIL method. In the experiments they did with simulated data and the adiabatic version of the ARPS model, they got convergence in an order of magnitude fewer iterations with the new method (ANA), and a decrease of the cost function, which was also an order of magnitude better than with the adjoint approach. They assimilated simulated data only at the end of the interval, and only the full model field, so that they did not need a background error term in the cost function.

In this subsection we generalize their approach by including both data and background in the cost function, with appropriate error covariances. To maintain the ability to solve the minimization efficiently, however, the background term is estimated at the end of the interval, rather than at the beginning as in Lorenc (1986). Our derivation is also considerably simpler than the ANA method, and our method does not use line minimization, as in Wang et al. (1997). As a result, our method is about twice as fast as the ANA approach.

Assume (for the moment) that data \mathbf{y}^o is available at the end of the assimilation interval t , with $\delta\mathbf{x} = \mathbf{x}^a - \mathbf{x}^b$, $\delta\mathbf{y} = \mathbf{y}^o - \mathbf{H}(\mathbf{x}^b)$. Here \mathbf{x}^a and \mathbf{x}^b are the analysis and first guess, respectively, and \mathbf{H} is the “forward observation operator,” which converts the model first guess into first guess observations (Ide et al. 1997). The cost function that we minimize is the 3D-Var cost function at the end of the interval. It is given by the distance to the background of the forecast at the end of the time interval (weighted by the inverse of the forecast error covariance B), plus the distance to the observations (weighted by the inverse of the observational error covariance R), also at the end of the interval:

$$J = \frac{1}{2}(L \delta\mathbf{x})^T B^{-1} (L \delta\mathbf{x}) + \frac{1}{2}(HL \delta\mathbf{x} - \delta\mathbf{y})^T R^{-1} (HL \delta\mathbf{x} - \delta\mathbf{y}). \quad (7)$$

Here $\delta\mathbf{x}$ (the control variable) is the difference between the analysis and the background (at the present iteration) at the beginning of the assimilation window, L and L^T are, as before, the TLM and its adjoint, and H is the tangent linear version of the forward observation operator \mathbf{H} . If we take the gradient of J with respect to the initial change $\delta\mathbf{x}$, we obtain

$$\nabla J = L^T [B^{-1} L \delta\mathbf{x} + H^T R^{-1} (HL \delta\mathbf{x} - \delta\mathbf{y})]. \quad (8)$$

From this equation we see that the gradient of the cost function is given by the backward adjoint integration of the rhs terms in (8). In the adjoint 4D-Var, the

gradient information is needed in an iterative minimization algorithm (such as quasi-Newton, conjugate gradient), which is used to minimize the cost function. The iterative process can be expressed simply in the form

$$\mathbf{x}_{k+1} = \mathbf{x}_k + a_k \mathbf{p}_k, \quad (9)$$

where, for iteration number k , the vector \mathbf{p}_k represents a search direction, and the positive scalar a_k is the step length. All minimization algorithms require the computation of the search direction, which is a function of ∇J . For example, $\mathbf{p}_k = -\nabla J$ for the steepest descent algorithm, $\mathbf{Q}\mathbf{p}_k = -\nabla J$ for the Newton method where \mathbf{Q} is a Hessian matrix, and $\mathbf{S}\mathbf{p}_k = -\nabla J$ for the quasi-Newton method where \mathbf{S} is an approximate Hessian matrix. These algorithms require many iterations until ∇J becomes very small and the minimum of J is reached. For some algorithms (e.g., LBFGS), each iteration requires a few corrections (or function calls) to compute the approximate Hessian, so that the number of direct and adjoint integrations required by 4D-Var can be significantly larger than the number of iterations.

In the inverse 3D-Var, however, we seek to obtain directly the “perfect solution,” that is, the $\delta\mathbf{x}$ that makes $\nabla J = 0$ for small $\delta\mathbf{x}$. From (8) we can eliminate the adjoint operator, and obtain the “perfect” solution given by

$$L \delta\mathbf{x} = (B^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} \delta\mathbf{y}. \quad (10)$$

Since we have a good approximation of L^{-1} at hand (the quasi-inverse model obtained by integrating the tangent linear model backward, but changing the sign of frictional terms), we can apply it and obtain

$$\delta\mathbf{x} = L^{-1} (B^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} \delta\mathbf{y}. \quad (11)$$

This can be interpreted as starting from the 3D-Var analysis increment at the end of the interval and integrating backward with the TLM or an approximation of it. If we do not include the forecast error covariance term B^{-1} , (11) reduces to the ANA algorithm of Wang et al. (1997) except that we do not need to run a minimization algorithm though a few quasi-inverse iterations are needed due to the discrepancy between the full nonlinear model and the linear model. We have tested the inverse 3D-Var with the ARPS model and found that for this reason, the inverse 3D-Var is computationally about twice as fast as the Wang et al. (1997) ANA scheme.

c. Equivalence of inverse 3D-Var and the Newton minimization algorithm

It is easy to prove that if (i) the forward model is linear, and (ii) the quasi-inverse tangent approximates the true inverse tangent linear model, then the inverse 3D-Var approach is equivalent to solving the minimization problem (at each time level) using the ideal Newton iterative method (e.g., Gill et al. 1981). Suppose that we are seeking the minimum of a cost function at $\mathbf{x} +$

$\delta\mathbf{x}$, and our present estimate of the solution is \mathbf{x} . Then by Taylor expansion,

$$\nabla J(\mathbf{x} + \delta\mathbf{x}) = \nabla J(\mathbf{x}) + \nabla^2 J(\mathbf{x}) \delta\mathbf{x} = 0. \quad (12)$$

Here $\nabla^2 J(\mathbf{x})$ is the Hessian matrix $\nabla^2 J_{ij} = \partial^2 J / \partial x_i \partial x_j$. The Newton algorithm, which has quadratic rate of convergence, solves the rhs part of equation (12): $\nabla J(\mathbf{x}) + \nabla^2 J(\mathbf{x}) \delta\mathbf{x} = 0$. Therefore the Newton iteration is given by

$$\delta\mathbf{x} = -[\nabla^2 J(\mathbf{x})]^{-1} \nabla J(\mathbf{x}). \quad (13)$$

Using the full-Newton algorithm (13) is extremely expensive because it requires the computation of both the gradient and the inverse of the Hessian. Navon and Legler (1987) reviewed various alternatives to the full-Newton algorithm for meteorological application (e.g., quasi-Newton, limited-memory quasi-Newton, and truncated Newton methods). With the truncated Newton method the Hessian-vector product $[\nabla^2 J(\mathbf{x}) \delta\mathbf{x}]$ in (12) is obtained approximately by the difference of gradients, while with the adjoint truncated Newton method (Wang et al. 1995) it is obtained exactly by solving the second-order adjoint. Although both methods can reduce the computing cost of the full-Newton iterations and show more efficient performance than quasi-Newton methods (Wang et al. 1995), they still require many iterations and are computationally expensive.

For the specific cost function (7), the Hessian is given by

$$\nabla^2 J = L^T(B^{-1} + H^T R^{-1} H)L. \quad (14)$$

Therefore the first iteration with the Newton descent algorithm is

$$\begin{aligned} \delta\mathbf{x}_1 &= [L^T(B^{-1} + H^T R^{-1} H)L]^{-1} L^T H^T R^{-1} \delta\mathbf{y} \\ &= L^{-1}(B^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} \delta\mathbf{y}, \end{aligned} \quad (15)$$

which is identical to the inverse 3D-Var solution (11).

The inverse 3D-Var algorithm solves exactly the same problem but takes advantage of the fact that the lhs of (12) $\nabla J(\mathbf{x} + \delta\mathbf{x}) = 0$ can be solved directly [cf. Eqs. (8) and (11)]. Therefore the inverse 3D-Var iteration (11) is identical to the Newton algorithm iteration (assuming the quasi-inverse approximates the true inverse), but it is not necessary to compute the Hessian or the gradient, just to integrate the linear tangent model backward.

The results of Wang et al. (1997) and Pu et al. (1997b) support considerable optimism for this method. For a quadratic function, the Newton algorithm (and the equivalent inverse 3D-Var) converges in a single iteration. Since the cost functions used in 4D-Var are close to quadratic functions, inverse 3D-Var can be considered equivalent to perfect preconditioning of the simplified 4D-Var problem.

d. Multiple time levels of data

If there are data at different time levels we can choose to bring the data increments to the same initial time

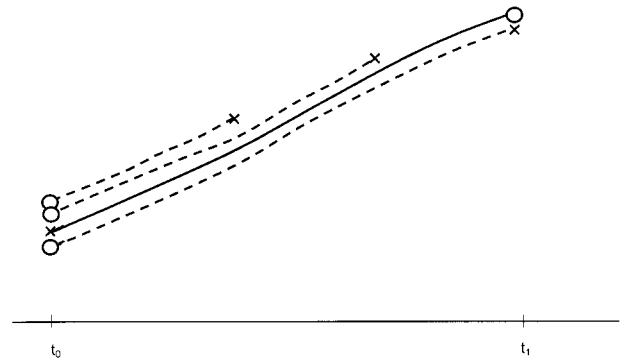


FIG. 1. Schematic of inverse 3D-Var approach when observations are available at several time levels in an assimilation window ($t_1 - t_0$), in which the observational increments are brought back to the beginning of the assimilation window. Solid line implies forward integration with the nonlinear model or the TLM, while dashed lines are backward integrations with the quasi-inverse nonlinear model or the quasi-inverse TLM. Analysis data, which can be a full analysis or an observation increment, are represented as \times . Model states are shown as \circ .

level (as shown schematically in Fig. 1) so that the increments corresponding to the different data can be averaged, with weights that may depend on the time level or the type of data. For applications in which “knowing the future” is allowed, such as reanalysis, the observational increments could be brought to the center of an interval, and used for the final analysis. In section 4 we show that, in a simple nonlinear model with complete data, when increments are brought to the same initial time, we solve a separate minimization for each time level, but that in fact (at least for this model) the 3D-Var minimizes the same multiple-level cost function as the simplified 4D-Var problem.

3. Burgers’ equation example

a. Simple TLM, adjoint, and inverse model formulation

Consider the simplest example of a nonlinear model with advection and diffusion, based on Burgers’ equation

$$\frac{\partial \bar{u}}{\partial t} = -\bar{u} \frac{\partial \bar{u}}{\partial x} + \nu \frac{\partial^2 \bar{u}}{\partial x^2}, \quad (16)$$

where $\bar{u} = u + \delta u$ and ν is a diffusion coefficient, and we assume that the basic flow $u(x)$ is a slowly varying function of x , and neglect its time changes.

The linear perturbation model is then

$$\frac{\partial \delta u}{\partial t} = -u \frac{\partial \delta u}{\partial x} - \delta u \frac{du}{dx} + \nu \frac{\partial^2 \delta u}{\partial x^2}. \quad (17)$$

Assume

$$\delta u = A(t) e^{ik(x-ut)}.$$

Then

$$\frac{dA}{dt} - ikuA = -ikuA - A \frac{du}{dx} - \nu k^2 A \quad \text{or}$$

$$A(t) = A_0 e^{-[(du/dx) + \nu k^2]t}.$$

We can interpret the first term in the exponent as an instability associated with the large-scale flow (perturbations grow where there is convergence, $du/dx < 0$), whereas the second term represents small-scale dissipative processes. The imaginary exponent represents the effects of large-scale advection.

So the solution is

$$\delta u(t) = e^{-ikut} e^{-(du/dx)t} e^{-\nu k^2 t} \delta u(0), \quad (18)$$

which can be interpreted as

final perturbation

= (large-scale advection)(large-scale instability)

× (diffusion)initial perturbation.

The TLM or propagator between time = 0 and time = t , is then

$$L = e^{-(du/dx + \nu k^2)t - ikut}. \quad (19)$$

The adjoint model is obtained by taking the complex conjugate of the transpose:

$$L^* = e^{-(du/dx + \nu k^2)t + ikut}. \quad (20)$$

The exact inverse linear tangent model is obtained by integrating backward in time (changing the sign of time):

$$L^{-1} = e^{(du/dx + \nu k^2)t + ikut} \quad (21)$$

and the approximate inverse (quasi-inverse) QTLM (Pu et al. 1997a) is obtained by integrating backward in time except for changing the sign of the diffusion terms:

$$L' = e^{(du/dx - \nu k^2)t + ikut}. \quad (22)$$

If we integrate forward with the linear model followed by a backward integration with the exact inverse, we get the exact initial conditions:

$$L^{-1}L = I. \quad (23)$$

If we integrate forward with the linear model followed by a backward integration with the adjoint, the unstable modes grow both during the forward and the backward integration, and damping also occurs twice:

$$L^*L = e^{-2(du/dx + \nu k^2)t}, \quad (24)$$

whereas if we follow the forward integration with a backward integration with the quasi-inverse QTLM, we get

$$L'Le^{-2\nu k^2 t}, \quad (25)$$

that is, we get the exact initial condition except smoothed twice by the diffusive terms.

b. Application of 4D-Var and inverse 3D-Var to Burgers' equation

Let us assume that $\mathbf{H} = I$ (observations are made in the model variable space). We assume

$$B = \alpha U^2 I, \quad R = U^2, \quad (26)$$

where U^2 is the observational error variance, and αU^2 is the background error variance.

Then, from (10), the inverse 3D-Var analysis is given by

$$L \delta u_0 = \frac{\alpha}{1 + \alpha} \delta u^{\text{obs}}, \quad (27)$$

where $\alpha \ll 1$ corresponds to small background errors (good forecast) and $\alpha \gg 1$ to a poor forecast (relative to the observations). If we neglect diffusion, the inverse 3D-Var solution is

$$\delta u_0 = e^{(du/dx)t + ikut} \frac{\alpha}{1 + \alpha} \delta u^{\text{obs}}. \quad (28)$$

The observational increments are appropriately weighted and moved back in time (advected and decreased where the flow is unstable). Note that in areas of large-scale decay ($du/dx > 0$) the initial increments will be larger than at the final time. This is not of concern, because when integrated forward, they will decay to their proper observed values.

If using the method of steepest descent, the first iteration of the regular 4D-Var, on the other hand, is

$$\delta u_1 = a \nabla J_0 = a \exp[\theta(du/dx)t + ikut] \delta u^{\text{obs}}, \quad (29)$$

where a is an appropriately chosen amplitude. Like the forecast sensitivity problem, it moves the observational increment backward in time, but enhancing the growing modes during the adjoint integration. It should be noted that when the 4D-Var is iterated, eventually it should converge to the same solution (28).

4. Numerical experiments

We have performed preliminary experiments with the NCEP global model (Pu et al. 1997a), and with two simple models, viscous Burgers' equation and the Lorenz (1963) model.

a. The NCEP global model

The application to the global NCEP model was a forecast sensitivity approach. The 24-h forecast error was estimated from the 24-h analysis, and the difference between the analysis and the forecast was integrated backward, using the TLM of the NCEP global model, with the sign of surface friction and horizontal diffusion changed. The results were very encouraging, indicating that the correction of the forecast was considerably better than using the adjoint sensitivity approach, even

when the latter was iterated five times (Pu et al. 1997a). Several important points should be noted:

- 1) Ideally, the TLM integrated backward should have a reversible formulation of the Hamiltonian (energy conserving) dynamics. In practice, the NCEP model has only approximately reversible dynamics (e.g., the Robert time filter is slightly diffusive). This will introduce additional diffusion during the backward integration; this subject is further discussed in the next subsection.
- 2) A known nonlinear solution was integrated backward and again forward over 24 h. The error in reproducing the full nonlinear perturbation with the TLM and the quasi-inverse TLM was about 10% in both total and kinetic energy throughout the model atmosphere, except near the surface, where the effect of changing the sign of friction is most important and where the error reached about 25%.
- 3) The amplitude of the quasi-inverse sensitivity was much larger than the adjoint sensitivity. This is because the adjoint sensitivity focuses only on the fastest growing modes [cf. Eq. (24)], whereas the quasi-inverse sensitivity includes both growing and decaying modes (and the latter grow during the backward integration). This may result in unwanted noise growth, and needs to be handled carefully.

b. Burgers' equation

We performed some numerical tests with viscous Burgers' equation (16). It should be noted that the model had been originally programmed using the Lax scheme (e.g., Anderson et al. 1984), which is highly diffusive and far from reversible (S. K. Park, personal communication 1998). Such a scheme would not be appropriate for a method that requires approximating the inverse of the model by running it backward. This was easily solved by rewriting the model with a leapfrog scheme for advection (with a forward first time step) and DuFort-Frankel for diffusion (e.g., Anderson et al. 1984). Therefore the numerical scheme was fully reversible except for the first forward time step. This allowed us to compare the effects of using the exact inverse linear model (IL) and the QIL as long as the Reynolds number was large enough (i.e., low dissipation) for the exact inverse to remain computationally stable.

The results were excellent. Figure 2a shows the cost function for a case in which the first guess included errors of 50%, and the observations were exact. As in Wang et al. (1997), the observation field at the end of the assimilation interval was complete, and the cost function did not include a background term. Unless noted otherwise, the results described below are computed with the QIL. The 4D-Var minimization was performed using the LBFGS algorithm (Liu and Nocedal 1989), based on a limited-memory quasi-Newton method. It should be noted that the parameter for checking direc-

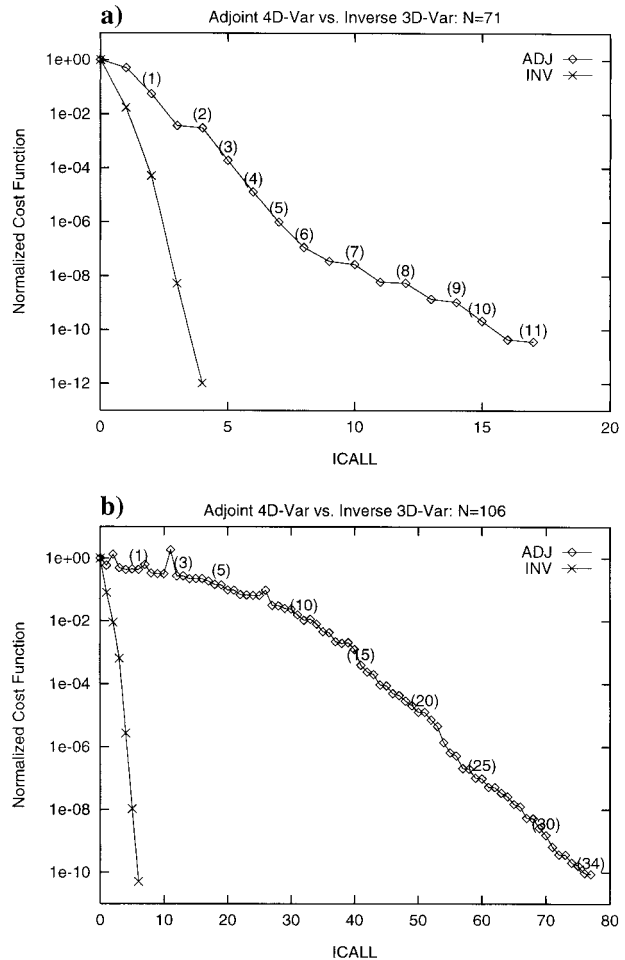


FIG. 2. Convergence rate of cost functions for two methods (adjoint 4D-Var vs inverse 3D-Var) as a function of CPU time, which is proportional to ICALL (number of calls for the nonlinear model and adjoint or inverse model). For the adjoint 4D-Var, the LBFGS algorithm is employed, in which one iteration requires more than one ICALL. The number of iterations is described in the parentheses. Reynolds number is 1000, and the initial error magnitude is 50% for u . The assimilation window length is (a) 71 and (b) 106.

tional derivative condition (GTOL) in the LBFGS algorithm has to be chosen appropriately for each problem; after some experimentation, we chose it to be $GTOL = 0.1$, which we found to be optimal for the 4D-Var in our case. For an assimilation window of 71 time steps (Fig. 2a), the inverse 3D-Var converged to less than 10^{-12} of its initial value in 4 iterations, whereas the adjoint algorithm required 11 iterations (and 17 computations of the gradient, each one involving a forward and backward integration) in order to converge to 10^{-10} . For longer assimilation windows, the advantages of inverse 3D-Var became more apparent. For example, when the assimilation window was extended to 106 time steps (Fig. 2b), inverse 3D-Var converged to 10^{-10} in only five iterations. The 4D-Var, on the other hand, converged to the same value in 34 iterations, and almost 80 computations of the gradient (each one involving a

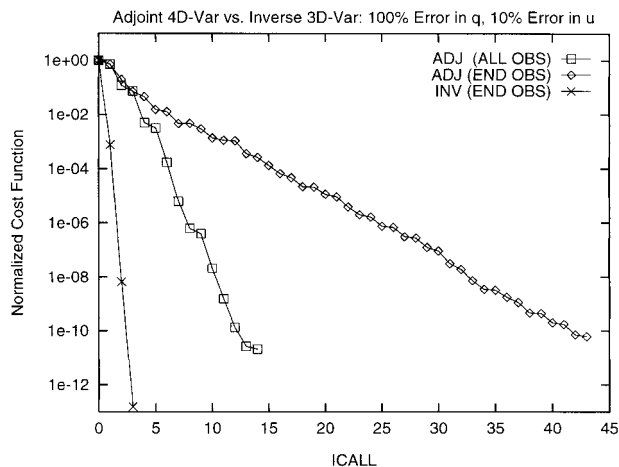


FIG. 3. Same as in Fig. 2 except that the assimilation is 101, and that, for the adjoint 4D-Var, the case with observations only at the end of the assimilation window (END OBS) is compared with the case with observations at all time steps (ALL OBS).

forward and backward integration of the model or its adjoint). Choices of smaller GTOL resulted in a lack of convergence.

Figure 3 shows the performance of two methods for the same case as in Fig. 2 but with an assimilation window of 101 time steps and different number of observations. The inverse 3D-Var converges to 10^{-12} of its original value after three iterations, but 4D-Var with data at the end of the interval requires 44 equivalent model integrations to converge to 10^{-10} . If we provide 4D-Var with complete observations for every time step, it converges to 10^{-10} of the original cost function in 12 time integrations. Many other experiments including observational and background errors were performed with uniformly good results. Some of the conclusions from these experiments (S. K. Park 1998, personal communication) are:

- 1) Results from inverse 3D-Var are very good in essentially every case. In general, for large Reynolds number (low dissipation), the QIL converges slightly faster than the exact IL. For small Reynolds number, the exact IL becomes unstable, but the QIL still converges fast.
- 2) We also tested the results of having multiple time levels in the observations. We included random errors in the observations with maximum amplitude of 10% of the total range. We followed the approach of Fig. 1, that is, we brought the observational increments (innovations) from different observational times backward to the same initial time level. Performing an average of the simultaneous increments gives very good results, and improves forecasts beyond the assimilation window roughly like $(n)^{-1/2}$, where n is the number of time levels in the observations. Additional iterations are performed integrating the nonlinear model from the updated initial

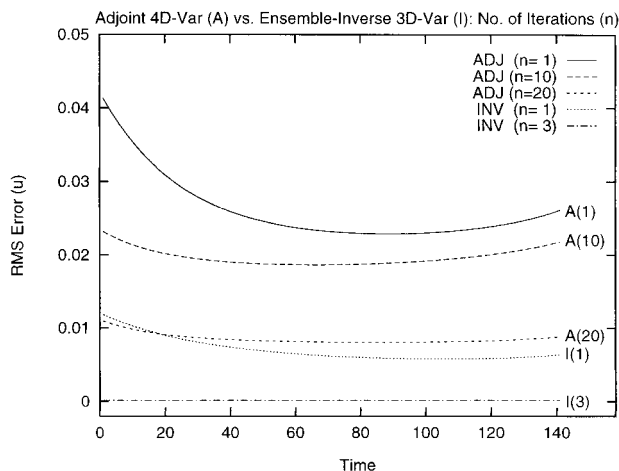


FIG. 4. Rms errors of u between true fields and forecast fields from various initial conditions generated by adjoint 4D-Var and inverse 3D-Var methods. Each method has been compared for different iteration numbers where initial conditions are generated using observations at six time levels ($t = 0, 21, 41, 61, 81,$ and 101). The inverse 3D-Var has been performed by running quasi-inverse TLM starting from each observation time (except $t = 0$) and those ensemble initial conditions are averaged including the observation fields at $t = 0$. Reynolds number is 1000, and initial error magnitude is 10% for u . Each observation includes random errors with maximum magnitude of 10%.

conditions, and integrating backward the observational increments. The results of the forecasts with one iteration of inverse 3D-Var were comparable to those of 20 iterations of 4D-Var, whereas three iterations of inverse 3D-Var resulted in much better forecasts (Fig. 4).

- 3) It is important to note that in this experiment, the inverse 3D-Var approach in practice minimizes the same total cost function as the variational approach, even though it is only guaranteed to minimize one observation level at a time (Table 1).

The variational approach gives a slightly better minimization after four iterations, but this does not translate into a more accurate forecast (Fig. 4, verified against truth), since there is slight overfitting of noisy observations.

TABLE 1. Variations in the cost function computed from different initial conditions generated by the same procedure as in Fig. 4 except that each variational assimilation (adjoint and ensemble inverse) has stopped at iteration numbers of 1–5.

Iteration	Adjoint	Ensemble inverse
0	41.8588	41.8588
1	22.5407	1.7420
2	2.1480	0.9197
3	2.0967	0.8881
4	0.8558	0.8787
5	0.8666	0.8746

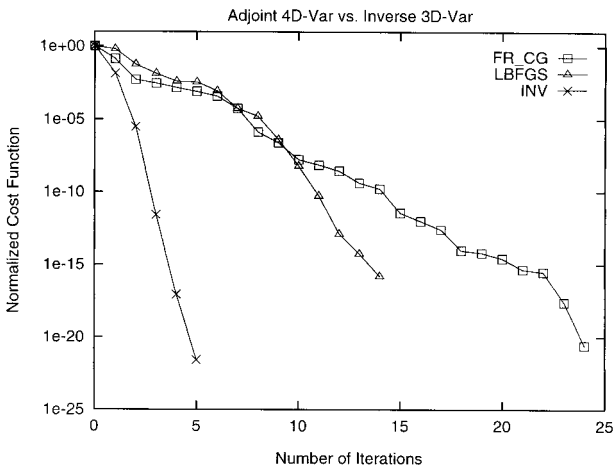


FIG. 5. Convergence rate of cost functions for adjoint 4D-Var and inverse 3D-Var (INV) using the Lorenz model. For adjoint 4D-Var, two methods are compared: Fletcher–Reeves conjugate gradient method (FR_CG) and the limited memory quasi-Newton method (LBFGS).

c. Experiments with Lorenz 3-variable model

Inverse 3D-Var was also tested and compared with regular (adjoint) 4D-Var using Lorenz (1963) 3-variable model. Figure 5 shows the evolution of the cost function using two different library minimization algorithms for the 4D-Var approach [limited-memory quasi-Newton or LBFGS, and Fletcher–Reeves conjugate gradient or FR_CG—see Navon and Legler (1987) for details] and the inverse 3D-Var. The results are similar to those obtained with Burgers’ equation: inverse 3D-Var reduces the cost function to 10^{-10} in three iterations and to 10^{-22} in five iterations (Fig. 5). The conjugate gradient and quasi-Newton methods converge to 10^{-14} in about 20 and 14 iterations, respectively (where each iteration includes several forward nonlinear and backward adjoint integrations). Figure 6 shows the cost function in the (X_0, Y_0) space and the descent approach of the three algorithms. The fact that inverse 3D-Var is equivalent to a Newton algorithm is apparent by the directness of its convergence: both the descent direction and the amplitude of the step are optimal. Several additional experiments with random errors in the initial conditions, multiple levels of observations, and observations of subset of variables and their combinations have also given uniformly excellent results (J. Gao 1998, personal communication).

5. Discussion

In the following discussion we first consider the relationship between the traditional 6-h cycle for 3D-Var, 4D-Var, and inverse 3D-Var (Cohn 1997; Courtier 1997). In 3D-Var (or previously in optimal interpolation) the observations were lumped together within a ± 3 -h window, and were assumed to have been taken at

the center of the interval. For example, an analysis at 1200 UTC included observations taken between 0900 UTC and 1500 UTC, but assumed to be observed at 1200 UTC. For observations made, for example, at 1000 UTC, this introduces two (relatively small) errors: the innovations (observations minus background) are computed with respect to a forecast at the wrong time, and the innovations are applied at the wrong time (1200 UTC instead of 1000 UTC). The first error can be easily corrected within 3D-Var: the background can be computed at the time of the observation, rather than at the center of the window. This correction is currently done at NCEP. However, only 4D-Var corrects the second error, 4D-Var (as implemented at ECMWF) computes the initial conditions valid at 0900 UTC that best fit the data at their correct times throughout the 0900–1500 UTC interval. It minimizes a cost function that includes distance to the background at 0900 UTC, plus the distance to the observations at their correct time (binned into 1-h intervals). The 4D-Var “analysis” at 1200 UTC is defined as the 3-h forecast from the optimal initial conditions at 0900 UTC. Because the minimization required about 80 iterations before reaching a satisfactory level, ECMWF has used a T63 model for the minimization, whereas the forecast model has T213 resolution.

Inverse 3D-Var offers some additional flexibility: if observations are complete, it allows transporting all the innovations from 0900 to 1500 UTC to the desired time (1200 UTC) essentially exactly (Fig. 1). The innovations at 1200 UTC can then be analyzed into a 3D-Var that includes different background weights depending on the length of the forecast. In general, however, the observations are not complete, and a background error covariance needs to be introduced into the cost function. In that case, the inverse 3D-Var analysis at the end of the assimilation interval is equivalent to 3D-Var, but the analysis is reached through a model integration, which can be advantageous in reducing problems of spin up. When knowledge of “future” observations is available (as in reanalysis), and the goal is to optimize the analysis (rather than to improve the forecast), the inverse 3D-Var can also be used, as suggested by the forecast sensitivity applications. In addition, it may be possible to use the inverse 3D-Var as a first iteration in the complete 4D-Var problem, thus acting as a kind of preconditioner.

We have seen that inverse 3D-Var has several potential advantages, including accuracy, efficiency, and flexibility, and these have been apparent in the simple model experiments. It also has some potentially serious disadvantages, but we believe they can be overcome with further development and experimentation:

- 1) Growth of noise that projects on decaying modes during the backward integration. This is a very serious problem, but it need not be a “show-stopper” since those errors will decay again during the next forward integration. The results of Pu et al. (1997a) for integrations 24 h and longer, and those obtained

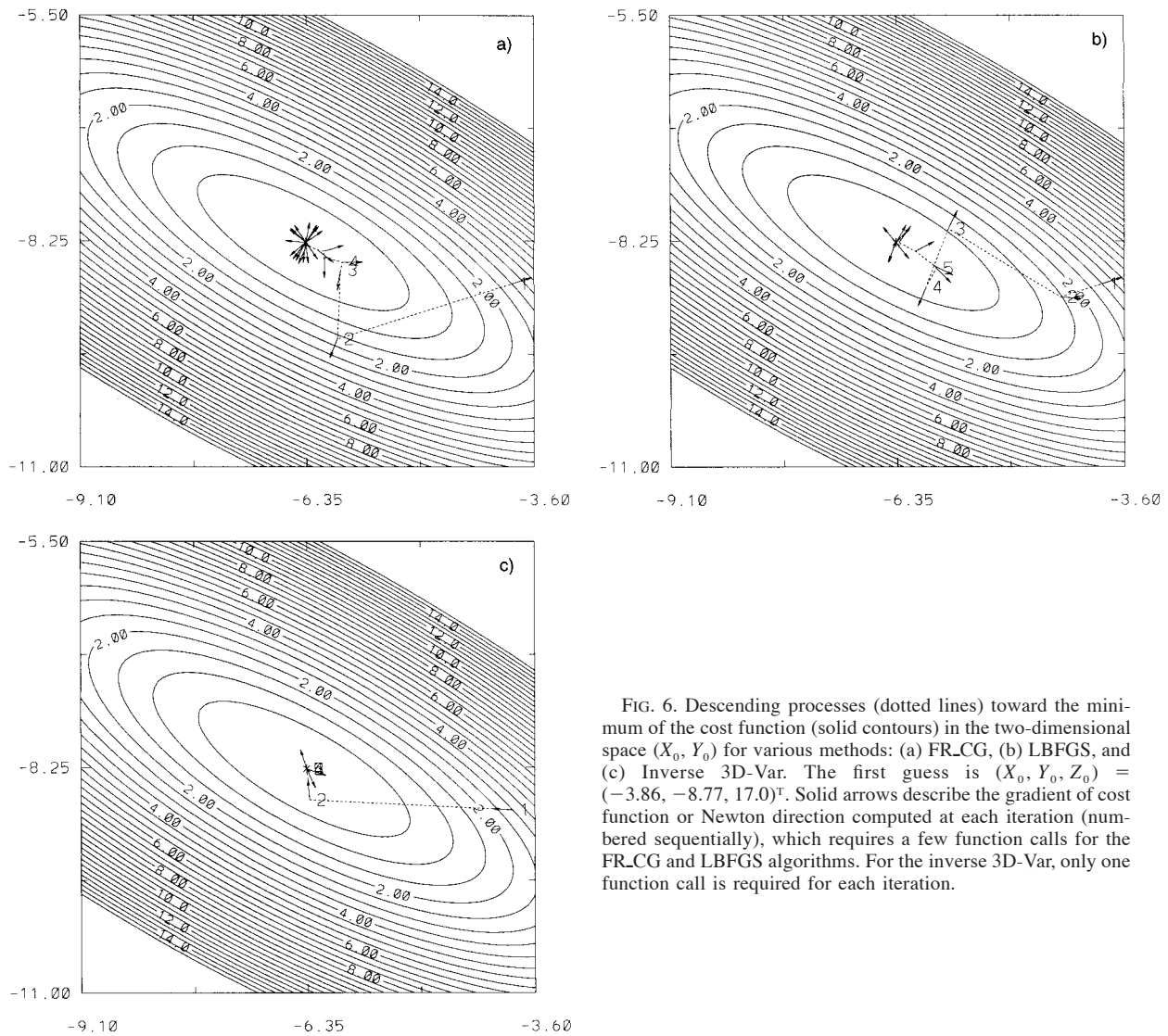


FIG. 6. Descending processes (dotted lines) toward the minimum of the cost function (solid contours) in the two-dimensional space (X_0, Y_0) for various methods: (a) FR-CG, (b) LBFGS, and (c) Inverse 3D-Var. The first guess is $(X_0, Y_0, Z_0) = (-3.86, -8.77, 17.0)^T$. Solid arrows describe the gradient of cost function or Newton direction computed at each iteration (numbered sequentially), which requires a few function calls for the FR-CG and LBFGS algorithms. For the inverse 3D-Var, only one function call is required for each iteration.

here with the Burgers' equation and observational noise are quite encouraging in this respect. It should be noted that the results obtained by Reynolds and Palmer (1998) when studying this problem are in full agreement with those of Pu et al. (1998) and of the present paper. They found that analysis uncertainties grew during the backward integration, but that during the forward integration they decayed again. As a result the forecast error reduction achieved at the end of the interval using the quasi-inverse was equivalent to that derived using the pseudo-inverse method with 60–90 singular vectors, but it was obtained at a computational cost several orders of magnitude smaller.

2) Physical processes are generally not parameterized in a reversible form in atmospheric models. This is also serious, but to some extent it can be overcome.

For example, moist convective processes can be simplified and parameterized in a reversible manner through the first hour of model integration (Jerry Straka 1998, personal communication). We are testing this idea with the ARPS model, where we plan to use a reversible parameterization of convection to “phase correct” the background field when, for example, the model predicts a squall line shifted in space and time from the observations.

3) The basic hydrodynamics of a model may not be written in a reversible fashion. If the numerical discretization of the hydrodynamics is excessively dissipative, this may require some rewriting, as discussed in section 4. Slightly dissipative schemes, when integrated backward will also fall within the “QIL” approach: they will dissipate both forward and backward.

- 4) The truly dissipative processes in the atmospheric model are not reversible, or may be reversible only for short intervals. If dissipation is a major factor during an assimilation window, it will not be well represented by inverse 3D-Var. On the other hand, both our results and those of Reynolds and Palmer (1998), suggest that in most cases, the quasi-inverse approximation slightly improves the results.
- 5) Finally, it has to be demonstrated with more complex systems that in nonlinear integrations this method will provide an improvement upon what can be attained with 3D-Var alone.

We are currently planning to test the inverse 3D-Var approach on the ARPS model, by combining it with a 3D-Var analysis including Doppler radar observations of radial velocities and reflectivities, using a linear tangent model with simplified reversible physics. If successful, we will also attempt to apply this method to the NCEP global model, where it could be applied, for example, in the second phase of the global Reanalysis project (Kalnay et al. 1996), where “future” data is available during the assimilation.

Acknowledgments. This work was started at NCEP’s Environmental Modeling Center. We would like to express our special gratitude to David Zhi Wang and his co-authors, for introducing similar ideas on the acceleration of 4D-Var, and Jerry Straka for developing a reversible parameterization of convection. Discussions with Kelvin Droegemeier, Keith Brewster, Alan Shapiro, and Brian Fiedler at the University of Oklahoma; Mark Iredell, Wan-Shu Wu, Zoltan Toth, M. Zupanski, and D. Zupanski at NCEP; R. Menard and S. Cohn at NASA DAO; and Carolyn Reynolds and Tim Palmer were also helpful. We are especially grateful to Olivier Talagrand, Andrew Lorenc, Francois Bouttier, Istvan Szunyogh, Jim Purser, John Derber, I. Michael Navon, and David Parrish for their insightful comments on an early version of the paper.

The support and encouragement of Peter Lamb (CIMMS), Kelvin Droegemeier (CAPS), and Fred Carr (School of Meteorology), and Ron McPherson and Stephen Lord (NCEP) are gratefully acknowledged. This research was supported by DOC NOAA under Grant NA67J0160 through CIMMS and NSF under Grant ATM91-20009 through CAPS.

REFERENCES

- Anderson, D. A., J. C. Tannehill, and R. H. Pletcher, 1984: *Computational Fluid Mechanics and Heat Transfer*. Hemisphere Publishing, 599 pp.
- Bouttier, F., and F. Rabier, 1997: The operational implementation of 4D-Var. *ECMWF Newsl.*, **78**, 2–5.
- Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan*, **75**, 257–288.
- Courtier, P., 1997: Variational methods. *J. Meteor. Soc. Japan*, **75**, 211–218.
- , J.-N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1388.
- Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press, 457 pp.
- Derber, J. C., 1989: A variational continuous assimilation technique. *Mon. Wea. Rev.*, **117**, 2437–2446.
- Gill, P. E., W. Murray, and M. H. Wright, 1981: *Practical Optimization*. Academic Press, 401 pp.
- Ide, K., P. Courtier, M. Ghil, and A. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential and variational. *J. Meteor. Soc. Japan*, **75**, 181–197.
- Kalnay, E., and Z.-X. Pu, 1998: Application of the quasi-inverse method to accelerate 4D-Var. Preprints, *12th Conf. on Numerical Weather Prediction*, Phoenix, AZ, Amer. Meteor. Soc., 41–42.
- , and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Le Dimet, F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97–110.
- Lewis, J. M., and J. C. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37A**, 309–322.
- Liu, D. C., and J. Nocedal, 1989: On the limited memory BFGS method for large scale minimization. *Math. Program.*, **45**, 503–528.
- Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194.
- , 1988: A practical approximation to optimal four-dimensional objective analysis. *Mon. Wea. Rev.*, **116**, 730–745.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.
- Menard, R., and R. Daley, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus*, **48A**, 221–237.
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF Ensemble Prediction System: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
- Navon, I. M., and D. M. Legler, 1987: Conjugate-gradient methods for large-scale minimization in meteorology. *Mon. Wea. Rev.*, **115**, 1479–1502.
- Pires, C., R. Vautard, and O. Talagrand, 1996: On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus*, **48A**, 96–121.
- Pu, Z.-X., E. Kalnay, J. Sela, and I. Szunyogh, 1997a: Sensitivity of forecast errors to initial conditions with a quasi-inverse linear model. *Mon. Wea. Rev.*, **125**, 2479–2503.
- , —, J. Derber, and J. Sela, 1997b: An inexpensive technique for using past forecast errors to improve future forecast skill. *Quart. J. Roy. Meteor. Soc.*, **123**, 1035–1054.
- , S. J. Lord, and E. Kalnay, 1998: Forecast sensitivity with drop-windsonde data and targeted observations. *Tellus*, **50A**, 391–410.
- Rabier, F., E. Klinker, P. Courtier, and A. Hollingsworth, 1996: Sensitivity of forecast errors to initial conditions. *Quart. J. Roy. Meteor. Soc.*, **122**, 121–150.
- , and Coauthors, 1997: The ECMWF operational implementation of four-dimensional variational assimilation. ECMWF Research Department Tech. Memo. 240, 62 pp. [Available from ECMWF, Shinfield Park, Reading RG2 9AX, United Kingdom.]
- Reynolds, C., and T. N. Palmer, 1998: Decaying singular vectors and their impact on analysis and forecast corrections. *J. Atmos. Sci.*, **55**, 3005–3023.
- Rohaly, G. D., R. H. Langland, and R. Gelaro, 1998: Identifying regions where the forecast of tropical cyclone tracks is most sensitive to initial condition uncertainty using adjoint methods. Preprints, *12th Conf. on Numerical Weather Prediction*, Phoenix, AZ, Amer. Meteor. Soc., 337–340.
- Thepaut, J.-N., R. N. Hoffman, and P. Courtier, 1993: Interactions of dynamics and observations in a 4-dimensional variational assimilation. *Mon. Wea. Rev.*, **121**, 3393–3414.
- Wang, Z., I. M. Navon, X. Zou, and F. X. Le Dimet, 1995: A truncated

- Newton optimization algorithm in meteorology applications with analytic Hessian/vector products. *Comput. Optim. Appl.*, **4**, 241–262.
- , K. K. Droegemeier, L. White, and I. M. Navon, 1997: Application of a new adjoint Newton algorithm to the 3D ARPS storm-scale model using simulated data. *Mon. Wea. Rev.*, **125**, 2460–2478.
- Xue, M., K. K. Droegemeier, V. Wong, A. Shapiro, and K. Brewster, 1995: ARPS version 4.0 users' guide. Center for Analysis and Prediction of Storms, University of Oklahoma, 380 pp. [Available from Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, OK 73019.]
- Zupanski, D., 1997: A general weak constraint applicable to operational 4DVAR data assimilation systems. *Mon. Wea. Rev.*, **125**, 2274–2292.
- Zupanski, M., 1993: Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Wea. Rev.*, **121**, 2396–2408.