A New Hessian Preconditioning Method Applied to Variational Data Assimilation Experiments Using NASA GCM.

Weiyu Yang

Supercomputer Computations Research Institute

Florida State University

I. Michael Navon

Department of Mathematics

and Supercomputer Computations Research Institute

Florida State University

Philippe Courtier

Data Division, Research Department

ECMWF

United Kingdom

July 11, 1995

Revised version submitted to Monthly Weather Review

Abstract

An analysis is provided to show that Courtier's et al.(1994) method for estimating the Hessian preconditioning is not applicable to important categories of cases involving nonlinearity. An extension of the method to cases with higher nonlinearity is proposed in the present paper by designing an algorithm which reduces errors in Hessian estimation induced by lack of validity of the tangent linear approximation. The new preconditioning method was numerically tested in the framework of variational data assimilation experiments using both the NASA Semi-Lagrangian semi-implicit (SLSI) global shallow-water (S-W) equations model and the adiabatic version of NASA/DAO GEOS-1 GCM. Our results show that the new preconditioning method speeds-up convergence rate of minimization when applied to variational data assimilation cases characterized by strong nonlinearity.

Finally we address issues related to computational cost of our new algorithm. These include the optimal determination of the number of random realizations p necessary for Hessian estimation methods. We tested a computationally efficient method which uses a coarser grid point model to estimate Hessian for application to a fine resolution mesh. The tests yielded encouraging results.

1. Introduction

The quality of weather forecasts is highly dependent on the quality of the initial conditions (see e.g. Rabier et al. 1994 for recent results). Good initial conditions that can extend predictability limits may be obtained from four-dimensional variational data assimilation (4-D Var) of meteorological observations. An important feature for a computationally efficient implementation of 4-D Var relates to attaining a fast convergence rate during the early stages of the minimization process, a factor which depends crucially on efficient preconditioning. For example, in the current operational practice at ECMWF (Courtier et al. 1994), the cost of 24 hours of data assimilation is equivalent to the cost of four days of model integration. With the introduction of 3-D Var, this cost could escalate to an equivalent cost of six days of model integration. If 30 iterations of the minimization algorithm are required for attaining a satisfactory convergence in 4-D Var, the CPU time required for a 24 hours 4-D Var is equivalent to the CPU time of 100 days of model integration. Courtier et al.(1994) introduced the incremental approach to reduce this cost by an order of magnitude. Any economy arising from a more efficient minimization can be translated in a relaxation of the simplifying assumptions induced by the incremental method.

The Hessian of the cost function can be used to estimate uncertainty of any model output and can thus determine aspects of the model that are poorly determined by observations (Thacker 1989). Conversely, when some aspects are poorly estimated, the Hessian is ill-conditioned. Convergence properties of the minimization process in 4-D Var are determined by the eigenvalue spectrum of the Hessian and convergence speed is related to the Hessian condition number (the ratio between its maximum and minimum eigenvalues). Geometrically, the iso-surfaces of the cost function consist of multidimensional ellipsoids and the condition number of the Hessian is the ratio between the dominant (maximum and minimum) axis length of this ellipsoid. The larger the condition number, the more the ellipsoid is stretched. An ill-conditioned Hessian corresponds to the case of large distortion of the ellipsoid.

When the Hessian is ill-conditioned, the calculated descent direction may be almost orthogonal to the optimal descent direction, resulting in an extremely slow convergence rate. A way to speed-up the convergence and relax the ill-conditioning of the Hessian is to introduce a preconditioner so that the condition number of the preconditioned Hessian is close to unity for fast convergence of the minimization process. It is clear geometrically that if iso-contours of the cost function are circular, the gradient will always point radially outward from the minimum, and only one descent iteration will be required to attain the minimum. In this case, the Hessian is proportional to the identity matrix and its condition number is unity.

The convergence of conjugate-gradient like methods may be enhanced by preconditioning a symmetric approximation B of the Hessian. Such a preconditioner is a symmetric, positive definite matrix P which is chosen to make the eigenvalues of $P^{-1}B$ cluster around as few distinct eigenvalues as possible. If B were to be positive definite, an ideal preconditioner would be B itself. There is normally a compromise involved between using a good approximation to B with the associated difficulties of finding and storing its factorization and a poor approximation where many conjugate-gradient iterations may be required (Conn et al. 1992). For example, Courtier et al.(1994) used for B a matrix of rank less than 100 in a pattern of size $O(10^6)$.

Choosing a good preconditioner for a given problem is considered to be an art. For the state of the art of preconditioning methods see Axelsson and Barker(1984), Conn et al.(1992), Axelsson(1994), Oppe et al.(1994), Ortega(1987) as well as the papers by Concus et al.(1976), Axelsson (1976) and Axelsson(1985).

To construct a good preconditioner, one should aim at extracting as much information as possible from the Hessian. Without second order Hessian information, one has to construct a preconditioner with empirical parameters. Li et al.(1994) and Zou and Holloway(1995) used rough gradient scaling methods to speed-up convergence rate of the minimization processes. Zupanski(1993a) introduced an empirical parameter to obtain a preconditioning formula and demonstrated that an optimal value of this parameter induced a faster convergence rate. The basic difficulty discussed but not solved in Zupanski's paper is that the Hessian matrix, the best preconditioner, is unknown since its calculation is prohibitively expensive (see Wang et al. 1992, 1995).

Wang et al.(1992) used a second order adjoint model to obtain a Hessian/vector product or a column of the Hessian for a cost comparable with that of a gradient computation. Gauthier(1992) studied the behavior of the covariance matrix of the gradient of a 4-D Var problem (Lorenz model) by considering the observations as random variables. He showed that this matrix is related to the observational error covariance matrix. Rabier and Courtier (1992) used this result to calculate error bars on the solution of their 4-D Var problem. Based on these results (Gauthier 1992; Rabier and Courtier 1992), Courtier et al.(1994) introduced a preconditioning method to estimate the Hessian of linear systems. Since the tangent linear approximation has a wide range of applicability when applied to 3-D or 4-D Var with numerical weather prediction (NWP) models (see Lacarra and Talagrand 1988; Vukicevic 1991; Rabier and Courtier 1992; and Li et al. 1993, 1994), one may at first conclude that Courtier's method may be applied in a straightforward manner in most cases to obtain an accurate Hessian (since the tangent linear approximation is within a valid range in a general sense). In this paper we present a detailed analysis which shows that by applying Courtier's method directly to a nonlinear system, even when the tangent linear approximation is within a valid range in a general sense, the estimated Hessian will be largely erroneous since the Hessian, which is not at the minimum state, is a function of both control variables and observations. Thus we propose here a new algorithm to extend Courtier's method to highly nonlinear variational data assimilation problems related to highly nonlinear physics or to quality control problems(Lorenc and Hammond 1988; Ingleby and Lorenc 1993). Results of numerical preconditioning experiments with NASA SLSI global S-W equations model and the adiabatic version of NASA GEOS-1 C-Grid GCM show our new algorithm to perform well in practice.

The essential difference between the proposed Hessian preconditioning method and empirical parameter preconditioning methods is that we use directly information on the Hessian (the best preconditioner), such as its diagonal elements, in order to build the preconditioner, thus allowing it to be directly applied to any variational data assimilation problem. On the contrary, when empirical parameters are used as preconditioners, they are usually obtained by some trial and error processes.

In the present paper we first developed the tangent linear model (tangent linear model) and adjoint of the adiabatic version of NASA Goddard Earth Observing System-1 (GEOS-1) C-Grid General Circulation Model (GCM) (Yang and Navon 1995a). The nonlinear forward model of the adiabatic version of NASA GEOS-1 C-Grid GCM and its adjoint were used to carry out a series of idealized 4-D Var preconditioning experiments in order to test the accuracy of the estimated Hessian, the impact of a preconditioner consisting of the diagonal elements of the estimated Hessian, as well as the performance of the new proposed Hessian estimation algorithm. All verification checks, such as adjoint check, gradient check, and analyses of validity of tangent linear approximation (for infinitesimal perturbations) were satisfied up to machine precision (Yang and Navon 1995a).

The plan of the paper is as follows. In Section 2 we provide a brief description of the adiabatic version of NASA GEOS-1 C-Grid GCM, its tangent linear and adjoint models as well as various verification checks of their correctness. In Section 3, we first provide a brief description of original Courtier's estimated Hessian preconditioning method (Courtier et al. 1994). We then discuss the issue related to variational assimilation of real observations, and analyze the impact of the validity of the tangent linear approximation in Courtier's Hessian preconditioning method on the accuracy of estimated Hessian, both theoretically and with numerical implementations. In Section 4, a new Hessian estimation algorithm extending Courtier's Hessian preconditioning method to nonlinear cases is proposed, while in Section 5 we provide the numerical results which confirm our analysis. In Section 6, we discuss the impact of the number of random realizations used for estimating the Hessian on the accuracy of the Hessian estimate. In Section 7, we discuss issues related to the computational cost of our new Hessian estimate algorithm and provide numerical results related to computationally efficient implementations. Finally summary and conclusions are presented in Section 8.

2. Tangent linear model and the adjoint of the NASA GEOS-1 C-Grid GCM

2.1. The NASA GEOS-1 C-Grid GCM

In NASA GEOS-1 C-Grid GCM, a σ vertical coordinate is defined by

$$\sigma = \frac{p - p_T}{\pi}$$

where $\pi \equiv p_s - p_T$, p_s is the surface pressure and p_T is a constant prescribed pressure at the top of the model atmosphere. In the current version, p_T=0.

The momentum equations are written in "vector-invariant" form, as in Sadourny(1975) and Arakawa and Lamb(1981), to facilitate derivation of an energy- and enstrophy-conserving differencing scheme. The thermodynamic equation is written in flux form to facilitate derivation of a θ-conserving differencing scheme. The equations involving computation of tendencies for other advected variables such as water vapor and ozone, are also written in flux form. Asselin(1972) time filter and Shapiro filter are used in the dynamic core of NASA GEOS-1 C-Grid GCM.

For further details concerning this GCM, we refer to Suarez et al.(1994) and Takacs et al.(1994).

2.2. Development of the tangent linear model and its adjoint

To derive the 4-D Var system based on the adiabatic version of NASA GEOS-1 C-grid GCM, we developed the tangent linear model and adjoint of the GEOS-1 GCM. For details related to model derivations, coding methods, as well as the methods and results of the verification of the correctness of these two models, see Yang and Navon(1995a). Comparing with the analysis applied to the adiabatic version of NASA/GLA SLSI GCM (Li et al. 1994), the tangent linear model of the adiabatic version of NASA GEOS-1 C-Grid GCM appears to display a better linearity property than the tangent linear model of NASA/GLA SLSI GCM.

The adjoint model satisfied a verification check up to 13 digits of accuracy using Eq.(2.18) of Navon et al.(1992). A gradient check was then performed to assess the accuracy of the discrete adjoint model. All verifications of the correctness of the tangent linear model and adjoint model were satisfied, thus indicating that the model and its adjoint can be safely used to perform 4-D variational data assimilation experiments.

3. Issues related to use of diagonal of estimated Hessian as preconditioner in 3-D or 4-D Var experiments

3.1. Courtier's method for estimating the Hessian matrix

Courtier et al.(1994) introduced a preconditioning method based on an estimate of the Hessian. They assumed a cost function which includes a background term for a generic problem. Based on the fact that the Hessian matrix is independent of the observations in a linear system, they introduced a bogus observation related to the background term to find the minimum of the cost function. Since in this paper we focus on the impact of the validity of the tangent linear approximation on the accuracy of the Hessian estimate in a nonlinear or weak nonlinear system, we will, for the sake of simplicity, discuss the minimization problem without involving the background term. Then Courtier et al.(1994) method can be applied to our problem as follows:

The cost function measuring misfit between the forecast model solution and the available observations distributed in space and time may be expressed as

$$J(\vec{X}(t_0)) = \frac{1}{2}\sum_{r=0}^{R}\left(B\left(\vec{X}(t_r)\right) - \vec{X}^{obs}(t_r)\right)^T \mathbf{W}(t_r)\left(B\left(\vec{X}(t_r)\right) - \vec{X}^{obs}(t_r)\right) \tag{3.1.1}$$

where B is the observation operator; $\vec{X}(t_r)$ is the vector of model control variables, which consists of the initial conditions in our experiments, at time $t = t_r$; $\vec{X}^{obs}(t_r)$ is the vector of observational data at time $t = t_r$; and $\mathbf{W}(\mathbf{t_r})$ is the inverse of the observation error covariance matrix. For the sake of simplicity we choose R=1, which yields

$$J(\vec{X}(t_0)) = \frac{1}{2}\left(B\left(\vec{X}(t_0)\right) - \vec{X}^{obs}(t_0)\right)^T \mathbf{W}(t_0)\left(B\left(\vec{X}(t_0)\right) - \vec{X}^{obs}(t_0)\right) +$$

$$+\frac{1}{2}\left(B\left(F\left(\vec{X}(t_0)\right)\right) - \vec{X}^{obs}(t_N)\right)^T \mathbf{W}(t_N)\left(B\left(F\left(\vec{X}(t_0)\right)\right) - \vec{X}^{obs}(t_N)\right) \tag{3.1.2}$$

where

$$F = \prod_{n=1}^{N} F_n \tag{3.1.3}$$

is the operator of the model integration from time $t = t_0$ to $t = t_N$. If B and/or BF are linear operators, the value of the Hessian J'' will be independent of the observation values.

At the minimum $\vec{X}_{min}$, the gradient ∇J vanishes. Let us introduce random variables $\vec{\eta}(t_0)$ and $\vec{\eta}(t_N)$, whose expectations are zero and whose covariances are the diagonal elements of $\mathbf{W}^{-1}(t_0)$ and $\mathbf{W}^{-1}(t_N)$, respectively, to the observations.

$$\vec{X}_1^{obs}(t_0) = \vec{X}^{obs}(t_0) + \vec{\eta}(t_0) \tag{3.1.4a}$$

$$\vec{X}_1^{obs}(t_N) = \vec{X}^{obs}(t_N) + \vec{\eta}(t_N) \tag{3.1.4b}$$

Then ∇J (at $\vec{X}_{min}$) is a random variable and we obtain

$$< \nabla J \nabla J^T >= J'' \tag{3.1.5}$$

where $<>$ stands for the mathematical expectation and J'' is the Hessian matrix.

For each realization i of $\vec{X}_1^{obs}(t_0)$ and $\vec{X}_1^{obs}(t_N)$ we can calculate ∇J^i (at $\vec{X}_{min}$) and considering p such realizations, (it follows a Wishart law, see Aitchison and Dunsmore(1975) and Wishart(1952), which is the generalization to the multi-dimensional case of the χ^2 law), we obtain the following approximation of Hessian matrix (see also Rabier and Courtier 1992)

$$H = J'' \simeq J''_p = \frac{1}{p} \sum_{i=1}^{p} \nabla J^i \nabla J^{i\,T} \tag{3.1.6}$$

Analogous methods in the deterministic domain aimed at building approximation of Hessian have been used in some minimization algorithms, e.g., all quasi-Newton methods of the BFGS and DFP variety. Assuming the Hessian is a constant matrix (which corresponds to the cost function being quadratic), quasi-Newton methods use symmetric rank one or rank two updates (of the form uv^T) at every iteration using information gathered as the descent process progresses to update the approximation of the Hessian or the inverse of the Hessian. This process may be called "earn while you learn". As the number of iterations increases, the updated Hessian matrix becomes more accurate and for quadratic functions one can show that the accurate Hessian is obtained in n-steps (see the discussion of the finite step convergence in Luenberger(1984)). The reason for the success of Courtier's (1994) method resides in the fact that mathematically the method translates spread eigenvalues consecutively into the vicinity K_0 of the point 1 ,group by group ,using low-rank transformations for constructing high quality preconditioner.(See Kharcenko and Yeremin(1995) and Eirola and Nevanlinna (1989)).

3.2. Extracting information from the estimated Hessian to build a preconditioner

If we were to have the whole accurate Hessian matrix at our disposal, the new preconditioned Hessian matrix will be the identity matrix **I**, causing the minimization process to attain its minimum in one iteration. However, this is an unattainable goal in 4-D Var with realistic NWP models due to two major limitations on available computer resources. First, the rank of J_p'' is, at most, the number of realizations p, which implies that we cannot directly use J_p'' as a preconditioner but we can extract useful information from J_p'', in this case its diagonal. The second is that due to obvious limitations on computer storage resources, we cannot store the whole Hessian matrix to carry out 4-D Var experiments with the NASA GEOS-1 GCM. Thus in our preconditioning experiments, we focus only on the diagonal elements of the estimated Hessian matrix J_p'' to improve relative scaling and preconditioning of different variables following Thépaut and Moll(1990) and Courtier et al.(1994).

Due to the fact that using the diagonal elements of the real Hessian as the preconditioning matrix will make all diagonal elements of the preconditioned Hessian equal to unity, one may assume that the condition number of the preconditioned Hessian will be greatly improved. This is especially true in the case when the Hessian matrix is strongly diagonally dominant. Besides, Forsythe et al.(1955) have shown that using the diagonal of the Hessian is optimal among all diagonal preconditioning methods.

3.3. Observations

We assume that the cost function measuring misfit between forecast model solution and available observations is of the form given by Eq.(3.1.2). Introducing a bogus observation data $\vec{X}_b^{obs}(t_N) = B\left[F\left(B^*\left(\vec{X}^{obs}(t_0)\right)\right)\right]$, where B^* is an operator from observation space to that of model variables $\vec{X}$ and which assumes the form of the generalized inverse of the operator B satisfying the Moore-Penrose conditions for a unique pseudo-inverse (for details on pseudo-inverse and generalized inverse operators see Nashed 1976; Lawson and Hanson 1974; Golub and Van Loan 1989; Ben-Israel and Greville 1974). Then the term $\left(B\left(F\left(\vec{X}(t_0)\right)\right) - \vec{X}^{obs}(t_N)\right)$ in (3.1.2) may be written as

$$B\left(F\left(\vec{X}(t_0)\right)\right) - \vec{X}^{obs}(t_N) = \left[B\left(F\left(\vec{X}(t_0)\right)\right) - \vec{X}_b^{obs}(t_N)\right] + R^{obs} \qquad (3.3.1)$$

where

$$R^{obs} \equiv \vec{X}_b^{obs}(t_N) - \vec{X}^{obs}(t_N) = B\left[F\left(B^*\left(\vec{X}^{obs}(t_0)\right)\right)\right] - \vec{X}^{obs}(t_N) \qquad (3.3.2).$$

The bogus observation data $\vec{X}_b^{obs}(t_N)$ resembles the so-called "model generated observations". It corresponds to the assumption that both model and observations are perfect and ensures that model outputs are absolutely consistent with the observations.

The second term R^{obs} on the r.h.s. of Eq.(3.3.1) is a term representing the inconsistency between model and observations. It originates from two sources, namely it is either due to deficiencies in dynamics or physics of forecast model and/or is due to inadequate observation data. During the 4-D Var minimization process, if both observational data and model are being kept invariant, R^{obs} will be invariant, i.e., the minimization process will only minimize that part of the cost function related to the first term on the RHS of Eq.(3.3.1). Thus if R^{obs} is large, it would be difficult to reduce the cost function to lower values. To reduce the R^{obs} term, if it originates in deficiencies of dynamics or physics of forecast model, one should strive to improve the model and/or to use the variational continuous assimilation technique proposed by Derber(1989) to correct model solutions throughout the assimilation interval. If the problem is due to inadequate data, additional and more accurate data are required, and sometimes bogus data based on *'a priori'* knowledge may be used to improve the quality of observational data, as discussed by Thacker(1989).

Since the Hessian estimation method is related to the minimum state $\vec{X}_{min}$, we need to find this minimum state. Clearly, if $R^{obs} = 0$, the minimum state is given by $\vec{X}_{min}(t_0) = B^* \left(\vec{X}^{obs}(t_0)\right)$ and the cost function will be zero at this minimum state. However if $R^{obs} \neq 0$, since both two terms of RHS of Eq. (3.1.2) are quadratic, the minimum of the cost function cannot be zero. From

$$\nabla J \left(\vec{X}_{min}(t_0)\right) = 0 \tag{3.3.3}$$

and noting that W are symmetric matrices, we find that the minimum state of the cost function (3.1.2) is given by

$$\begin{aligned}\nabla J \left(\vec{X}_{min}(t_0)\right) &= B'^T \left(\vec{X}_{min}(t_0)\right) \mathbf{W}(t_0) \left(B\left(\vec{X}_{min}(t_0)\right) - \vec{X}^{obs}(t_0)\right) + \\ &+ (BF)'^T \left(\vec{X}_{min}(t_0)\right) \mathbf{W}(t_N) \left[B\left(F\left(\vec{X}_{min}(t_0)\right)\right) - \vec{X}^{obs}(t_N)\right] = 0\end{aligned} \tag{3.3.4}$$

where "$'$" denotes the first order derivative. For a linear system whose Hessian is invertible, Eq. (3.3.4) may be written as

$$\vec{X}_{min}(t_0) = H^{-1}\left[B'^T\mathbf{W}(t_0)\vec{X}^{obs}(t_0) + (BF)'^T\mathbf{W}(t_N)\vec{X}^{obs}(t_N)\right] \qquad (3.3.5)$$

where H^{-1} is the inverse of the Hessian H. To obtain the minimum state $\vec{X}_{min}(t_0)$ when $R^{obs} \neq 0$, one should use an iterative method. Moreover, in a nonlinear system, Eq. (3.3.3) does not constitute a sufficient condition for finding the global minimum.

Thus we conclude that $R^{obs} \neq 0$ will render our analysis of Hessian preconditioning more complicated. In this paper we use model generated data as "observations" for all the numerical experiments. Issues related to use of real observations will be analyzed in a forthcoming paper.

3.4. Validity of the tangent linear approximation

To apply Courtier's method, we must pay attention to two issues. The first issue is that this method is based on the assumption that B and (BF) are linear operators, which brings us to consider the issue of what happens if B and (BF) are either nonlinear or weakly nonlinear operators. As the tangent linear approximation has a wide range of applicability when applied to 3-D or 4-D Var with NWP models (Lacarra and Talagrand 1988; Vukicevic 1991; Rabier and Courtier 1992; and Li et al. 1993, 1994), one may conclude intuitively that Courtier's method may be applied in a straightforward manner in most cases to obtain an accurate Hessian estimate (since the tangent linear approximation is within a valid range in a general sense). However, our detailed analysis reveals that in many situations, if an accurate estimated Hessian is required, the restriction on the validity of the tangent linear approximation using Courtier's method will be too stringent for it to be applicable to nonlinear cases of interest. The second issue relates to minimizing the computational cost of the method, which consists in choosing an optimal number p of perturbed gradients of the cost function and using economical Hessian estimation methods. These issues will be addressed in Sections 6 and 7, respectively.

From the cost function (3.1.2), noting that W are symmetric matrices, we have

$$
\begin{aligned}
H = & J'' \left(\vec{X}(t_0)\right) = B''^T \left(\vec{X}(t_0)\right) \mathbf{W}(t_0) \left(B\left(\vec{X}(t_0)\right) - \vec{X}^{obs}(t_0)\right) + \\
& + B'^T \left(\vec{X}(t_0)\right) \mathbf{W}(t_0) B' \left(\vec{X}(t_0)\right) + \\
& + (BF)'^T \left(\vec{X}(t_0)\right) \mathbf{W}(t_N)(BF)' \left(\vec{X}(t_0)\right) + \\
& + (BF)''^T \left(\vec{X}(t_0)\right) \mathbf{W}(t_N) \left[B\left(F\left(\vec{X}(t_0)\right)\right) - \vec{X}^{obs}(t_N)\right]
\end{aligned}
\tag{3.4.1}
$$

where "$''$" denotes the second derivative. During the minimization process, if either B and/or (BF) are nonlinear operators, the control variable $\vec{X}$ is not at the minimum point and the Hessian will depend both on values of the control variable $\vec{X}$ and on the observations $\vec{X}^{obs}$, i.e., the real Hessian cannot to be obtained with Courtier's method. To apply Eq. (3.1.6) to obtain an accurate estimated Hessian, we must require that the variational data assimilation cases satisfy conditions where the tangent linear approximation is valid. Since all the computational steps in Courtier's method are related to $\vec{X}_{min}(t_0)$, we choose this minimum state as the basic state for assessing the validity of the tangent linear approximation in this paper.

The term $\left(B\left(F\left(\vec{X}(t_0)\right)\right) - \vec{X}^{obs}(t_N)\right)$ in (3.1.2) may be written by omitting R^{obs} term as:

$$
B\left(F\left(\vec{X}(t_0)\right)\right) - \vec{X}^{obs}(t_N) \simeq \left[B\left(F\left(\vec{X}(t_0)\right)\right) - B\left(F\left(B^*\left(\vec{X}^{obs}(t_0)\right)\right)\right)\right] \tag{3.4.2}
$$

and we can then rewrite the cost function with "model generated observations" as

$$
\begin{aligned}
J(\vec{X}(t_0)) = & \frac{1}{2}\left(B\left(\vec{X}(t_0)\right) - \vec{X}^{obs}(t_0)\right)^T \mathbf{W}(t_0)\left(B\left(\vec{X}(t_0)\right) - \vec{X}^{obs}(t_0)\right) + \\
& + \frac{1}{2}\left(B\left(F\left(\vec{X}(t_0)\right)\right) - B\left(F\left(B^*\left(\vec{X}^{obs}(t_0)\right)\right)\right)\right)^T \\
& \mathbf{W}(t_N)\left(B\left(F\left(\vec{X}(t_0)\right)\right) - B\left(F\left(B^*\left(\vec{X}^{obs}(t_0)\right)\right)\right)\right)
\end{aligned}
\tag{3.4.3}
$$

Since

$$B\left(\vec{X}(t_0)\right) - \vec{X}^{obs}(t_0) = B\left(\vec{X}(t_0)\right) - B\left(B^*\left(\vec{X}^{obs}(t_0)\right)\right)$$
$$= L_1\left(\vec{X}(t_0) - B^*\left(\vec{X}^{obs}(t_0)\right)\right) + O\left(\left(\vec{X}(t_0) - B^*\left(\vec{X}^{obs}(t_0)\right)\right)^2\right) \quad (3.4.5)$$

and

$$B\left(F\left(\vec{X}(t_0)\right)\right) - B\left(F\left(B^*\left(\vec{X}^{obs}(t_0)\right)\right)\right)$$
$$= L_2\left(\vec{X}(t_0) - B^*\left(\vec{X}^{obs}(t_0)\right)\right) + O\left(\left(\vec{X}(t_0) - B^*\left(\vec{X}^{obs}(t_0)\right)\right)^2\right) \quad (3.4.6)$$

where L_1 is the tangent linear operator of B, and L_2 is the tangent linear operator of (BF). Define

$$J_L(\vec{X}(t_0)) = \frac{1}{2}\left(L_1\left(\vec{X}(t_0) - B^*\left(\vec{X}^{obs}(t_0)\right)\right)\right)^T \mathbf{W}(t_0)$$
$$\left(L_1\left(\vec{X}(t_0) - B^*\left(\vec{X}^{obs}(t_0)\right)\right)\right) + \frac{1}{2}\left(L_2\left(\vec{X}(t_0) - B^*\left(\vec{X}^{obs}(t_0)\right)\right)\right)^T$$
$$\mathbf{W}(t_N)\left(L_2\left(\vec{X}(t_0) - B^*\left(\vec{X}^{obs}(t_0)\right)\right)\right) \quad (3.4.6)$$

When the difference between $\vec{X}(t_0)$ and $\left(B^*\left(\vec{X}^{obs}(t_0)\right)\right)$ is small enough, we obtain

$$J_L(\vec{X}(t_0)) \simeq J(\vec{X}(t_0)) \quad (3.4.7)$$

and thus the Hessian matrices of $J(\vec{X}(t_0))$ and $J_L(\vec{X}(t_0))$ are nearly equal. We found that only by satisfying the condition expressed by Eq. (3.4.7), can Eqs. (3.1.5) and (3.1.6) be used to yield an accurate Hessian estimate.

In order to investigate the impact of validity of tangent linear approximation, we first analyze a simple example (Appendix A). We find that even when the nonlinearity of operator F_n (in Eq. (3.1.3)) is very weak and the validity of tangent linear approximation is within an acceptable range, the estimated Hessian still displays a large error. Since operators in realistic NWP models possess much stronger nonlinear properties than the operator F_n chosen in this simple example, our analysis reveals that an accurate Hessian matrix cannot be estimated with Courtier's method applied directly to 3-D or 4-D Var using a realistic NWP model. Thus in the next section we propose a new algorithm to solve this problem and obtain satisfactory results in preconditioning variational data assimilation experiments where nonlinearity is an issue.

The numerical results of preconditioning variational data assimilation experiments with either NASA SLSI global S-W model (Figures omitted) and the NASA GEOS-1 GCM (Yang and Navon 1995b) show clearly that validity of the tangent linear approximation strongly impacts the convergence rate of minimization when the preconditioning Hessian was estimated by Eq. (3.1.6). In cases involving Hessian preconditioning experiments, where the validity of the tangent linear approximation is not satisfactory, the resulting convergence rate of the minimization processes is much slower than in cases characterized by a good validity of the tangent linear approximation. This fact reveals that with higher validity of the tangent linear approximation, the Hessian estimate will be more accurate.

It should be noted that in practical 4-D Var, attention is paid primarily to the convergence rate of the minimization process and much less to the accuracy of the preconditioner. In some instances the convergence rate of the minimization process is found to be satisfactory, although the preconditioning estimated Hessian is inaccurate. This is due to the fact that some unconstrained minimization packages are endowed with their own preconditioning schemes which compensate to some extent effects of this error. Such is the case with L-BFGS or E04DGF minimization algorithms (See Liu and Nocedal 1989, Gill and Murray 1979, Gill et al. 1981). In such instances one does not have to find a more accurate Hessian. However if the error of the preconditioning Hessian estimate is too large to yield a satisfactory convergence rate of the large scale unconstrained minimization algorithm in variational data assimilation, one can beneficially apply our new Hessian estimation algorithm to obtain a more accurate Hessian estimate.

4. An new algorithm for obtaining a more accurate Hessian estimate

4.1. The basic algorithm

To extend Courtier's Hessian preconditioning method (Courtier et al. 1994) to cases characterized by higher nonlinearity, we propose to apply a new algorithm to 4-D Var experiments to reduce the error in estimated Hessian induced by lack of validity of the tangent linear approximation. The basic idea(illustrated in a schematic diagram Fig. 1) consists in dividing the distance L between $\left(B\left(\vec{X}^{initial}(t_0)\right)\right)$ and $\vec{X}^{obs}(t_0)$ into n $(n > 1)$ parts resulting in the fact that the distance L/n in each part will be much shorter for large enough n. We also introduce intermediate bogus "observations" $\vec{X}_k^{obs}(t_0)$, which are related to the original observations, located within the part closer to $\left(B\left(\vec{X}\right)\right)$. Thus the difference between $\left(B\left(\vec{X}\right)\right)$ and intermediate "observations" $\vec{X}_k^{obs}(t_0)$ can be made small enough for Eq. (3.4.7) to be satisfied, and we can use Eqs. (3.1.5)-(3.1.6) to obtain an accurate Hessian estimate. First, we provide a basic description of our new algorithm.

Let the distance between initial conditions and the initial time observations be denoted by

$$\Delta\vec{X} = B\left(\vec{X}^{initial}(t_0)\right) - \vec{X}^{obs}(t_0) \tag{4.1.1}$$

choose $n \geq 1$, let $k = n - 1$ and

$$\frac{1}{n}\|\Delta\vec{X}\| \leq \varepsilon \tag{4.1.2}$$

where ε is a critical value, (i.e. the value of the perturbation $\frac{1}{n}\|\Delta\vec{X}\|$ which acting upon $B\left(\vec{X}_{min}(t_0)\right)$ yields a value smaller than the prescribed threshold value ε).

(i) substitute $\vec{X}_k^{obs}(t_0) = \left(\vec{X}^{obs}(t_0) + \frac{k}{n}\Delta\vec{X}\right)$ for $\vec{X}^{obs}(t_0)$ in the cost function (3.4.3), and define an intermediate cost function J_k,

$$J_k(\vec{X}_k(t_0)) = \frac{1}{2}\left(B\left(\vec{X}_k(t_0)\right) - (\vec{X}^{obs}(t_0) + \frac{k}{n}\Delta\vec{X})\right)^T$$
$$\mathbf{W}(t_0)\left(B\left(\vec{X}_k(t_0)\right) - (\vec{X}^{obs}(t_0) + \frac{k}{n}\Delta\vec{X})\right) +$$
$$+\frac{1}{2}\left(B\left(F(\vec{X}_k(t_0))\right) - B\left(F\left(B^*\left(\vec{X}^{obs}(t_0) + \frac{k}{n}\Delta\vec{X}\right)\right)\right)\right)^T$$
$$\mathbf{W}(t_N)\left(B\left(F(\vec{X}_k(t_0))\right) - B\left(F\left(B^*\left(\vec{X}^{obs}(t_0) + \frac{k}{n}\Delta\vec{X}\right)\right)\right)\right) \quad (4.1.3)$$

where $\vec{X}_k(t_0)$ are the intermediate initial conditions at stage k; $F\left(\vec{X}_k(t_0)\right)$ is the model operator integrated from initial time t_0 to t_N starting from $\vec{X}_k(t_0)$; and $\left(B\left(F\left(B^*\left(\vec{X}^{obs}(t_0) + \frac{k}{n}\Delta\vec{X}\right)\right)\right)\right)$ is the model operator starting from intermediate observations $\vec{X}_k^{obs}(t_0)$ to generate $\vec{X}_k^{obs}(t_N)$.

(ii) calculate the diagonal elements of the estimated Hessian J_k using Eq. (3.1.6), and use them as the preconditioner.

(iii) perform a preconditioned minimization iteration process to minimize J_k.

(iv) if $k = 0$, stop the computation process, otherwise, update k by $k - 1$, and go to (i).

Applying this algorithm to the simple scalar variable example in Appendix A, we obtained a clear illustration of the effects of our proposed algorithms on the accuracy of the estimated Hessian (Fig. 2), where the relative error of estimated Hessian is defined by

$$R_H = \sqrt{\frac{1}{n}\sum_{k=0}^{n-1}\left(\frac{J''_{k\,ESTIMATE} - J''_{k\,REAL}}{J''_{k\,REAL}}\right)^2} \quad (4.1.4)$$

.

The solid line in Fig. 2 stands for R_H corresponding to the Hessian being calculated at $\vec{X}_{min}$; the dashed line stands for R_H corresponding to the Hessian being estimated by the r.h.s. of Eq.(3.1.6), with p=30 (short dashed line) and p=3000 (long dashed line), respectively. From the results of this test we conclude that our algorithm given by steps (i) - (iv) sizably reduces the error of estimating the Hessian matrix. Additionally, we draw the following conclusions:

a) By applying Courtier's method directly to nonlinear cases, one cannot obtain an accurate Hessian matrix no matter how large a value of p is chosen.

b) We found that for a chosen parameter p, R_H will be close to a fixed value $R_{H\,critical}$ when n exceeds a threshold value $n_{critical}$. In other words, if p is fixed, no matter how accurate the tangent linear approximation is, one cannot render the estimated Hessian matrix more accurate than a threshold relative error $R_{H\,critical}$. The larger p is, the smaller $R_{H\,critical}$ becomes. For a fixed p, there is an optimal value of n, which allows us to reduce computational costs while obtaining an accurate estimate of the Hessian matrix. In Fig. 2, $n_{critical}$ for p=30 is 5, while for p=3000 $n_{critical}$ is about 80, respectively, yielding corresponding values for $R_{H\,critical}$ which are 1.6×10^{-1} and 1.6×10^{-2}, respectively.

This is just a basic analysis which is was derived from a scalar case where the Hessian matrix consists of just one entry. In practice, these restrictions turn out to be more stringent if we just use the diagonal elements of an estimated Hessian as a preconditioner.

The chosen value of n depends on the value of $\Delta\vec{X}$. In practice, n may be chosen to assume values between 2-6. If $n = 1$, our new algorithm is equivalent to the original Courtier's method.

4.2. A modified algorithm for Hessian estimation

From the discussion in last subsection, we concluded that our new algorithm can sizably reduce the error of estimated Hessian induced by limitation on the validity of the tangent linear approximation in Courtier's method. However in practical 4-D Var experiments, two problematic issues emerged, related to unsatisfactory results obtained due to minimizing an intermediate cost function J_k.

The first issue is that, since J_k is just an intermediate cost function, and our final goal is to minimize J in (3.4.3), a satisfactory retrieved initial conditions vector which minimizes J_k cannot yield a large reduction in the normalized cost J in Eq. (3.4.3), especially when a larger value of n is chosen. To verify this point, using the No. 4 4-D Var experiment (to be described in the next paragraph) in Yang and Navon(1995b), we carried out three preconditioning 4-D Var experiments with different values of n.

In these test experiments, we used the 4-D Var system of the adiabatic version of NASA GEOS-1 GCM with January 1, 1985 ECMWF 00Z data as the initial time observation data. The surface pressure field of the observational data is displayed in Fig. 3. The first guess initial condition is a randomly perturbed initial condition from the observation data, i.e.,

$$\vec{X}^{initial} = \mathbf{B}\vec{X}^{obs}(t_0) \tag{4.2.1}$$

where $\mathbf{B}$ is a randomly perturbed coefficient diagonal matrix whose diagonal elements are given by

$$b_{i,i} = C_1 + C_2 \times D_i,\ i = 1, 2, ..., n \tag{4.2.2}$$

with C_1=0.75 and C_2=0.5. The length of the data assimilation window is 6 hours and the time-step length is 5 minutes. The weighting matrix $\mathbf{W}$ used in these 4-D Var preconditioning experiments is a constant diagonal matrix whose diagonal components are $\mathbf{W}_u = \mathbf{W}_v = 10^{-3}\mathbf{I}\ \ s^2m^{-2}$, $\mathbf{W}_T = 10^{-2}\mathbf{I}\ \ K^{-2}$ and $\mathbf{W}_{Ps}$=$10^{-3}\mathbf{I}\ \ hPa^{-2}$, respectively. The value of the parameter p used in (3.1.6) was chosen as p=60. The L-BFGS minimization algorithm (Liu and Nocedal 1989) was used in these experiments with an update formula parameter m=5. This 4-D Var case is characterized by a severe lack of validity of the tangent linear approximation and as a consequence the convergence rate of the unconstrained minimization process cannot result in an efficient speed-up using Courtier's method.

We then conducted experiments to test our new basic algorithm in steps of $k = n - 1$ using different values of n. We chose n=2, 4, and 8. The results are presented in Fig. 4. We calculated the cost function given by J in Eq. (3.4.3) using the same retrieved control variable fields as used when calculating J_k. The solid lines correspond to variations of J_k while the dashed lines provide variations of J, both versus the number of minimization iterations.

From Fig. 4, we see that when a larger parameter n is used, the minimization of J_k proceeds much faster than for a smaller value of n. This confirms experimentally that our basic idea is correct, i.e., as n increases, the distance between intermediate observations and the intermediate first guess initial conditions is shortened, thus the validity of the tangent linear approximation is increased, which in turn yields a more accurate estimate of the Hessian with better ability to speed-up convergence of the minimization process.

However while J_k is more efficiently minimized by using a larger n, the corresponding change in the cost function J becomes smaller. This is due to the fact that J_k is just an intermediate cost function, thus, as n increases, the difference between J_k and J will increase. For instance, for n=8, even though the cost function J_k is minimized to zero, the retrieved fields are reduced to only about 76% of normalized original cost function J. Since our goal is to minimize J, we must modify our basic algorithm to achieve this goal.

A second issue is related to the fact that we used the L-BFGS minimization algorithm in our 4-D Var experiments. This algorithm requires information from the last m iterations to update the formula generating approximation to the Hessian matrix. In our experiments, we chose m=5. If the number of total iterations is less than m, the L-BFGS minimization algorithm cannot be used efficiently. On the other hand, due to computational efficiency considerations, it is not desirable to carry out 4-D Var experiments requiring a large number of minimization iterations. As a rule the total number of iterations should be a small integer value, for instance a total of 8 minimization iterations is typical. Thus, in each step of our basic preconditioning algorithm the total number of iterations will be divided into n (for instance $n = 4$) steps. If we do not store information from the previous minimization iterations with J_k for the latter J_{k-1} minimization iterations to be used in the L-BFGS algorithm, the L-BFGS minimization algorithm will turn out to be inefficient, due to the small number of iterations carried out at each step of the algorithm. However if we save information from previous minimization iterations with J_k for the latter J_{k-1} minimization iterations, the values of cost function and its gradient will experience sudden jumps since we minimize different cost functions, causing either failure or non-smooth performance of the L-BFGS algorithm.

To remedy the negative impact of above mentioned two issues requires a logical modification of our basic algorithm. This consists in applying the minimization iteration process directly towards minimizing the original cost function J in our modified algorithm, in Step (iii) rather than minimizing the intermediate cost function J_k. This is the only change introduced in the basic algorithm presented in Subsection 4.1. For reference, we provide a detailed description of the modified algorithm in Appendix B.

Since we did not make any assumption concerning the nature of the nonlinear operators B and F in (i)-(iv), our proposed preconditioning algorithm may be theoretically applied to any nonlinear 3-D or 4-D Var system.

5. Numerical results of preconditioning variational data assimilation experiments

We carried out preconditioning variational data assimilation experiments using either the NASA SLSI global S-W equations model or the adiabatic version of NASA/DAO GEOS-1 GCM to test our newly proposed preconditioning algorithm. The numerical results show that this algorithm does indeed possess the ability to speed-up the convergence rate of minimization processes due to the increased accuracy of the estimated Hessian.

5.1. Experiments using the NASA SLSI global S-W equations model.

We first tested our preconditioning algorithm by using the variational data assimilation system of NASA SLSI global S-W equations model. For a detailed description of the model see Bates et al.(1990).

The observation used consists of a 12-h preparatory integration output of 500 hPa fields of January 15, 1979, 00Z ECMWF data(for the geopotential field see Fig. 6). The model resolution is $(\Delta\theta, \Delta\lambda)$=$(7.5^o, 7.5^o)$, respectively. The length of the assimilation window is 3 hours and the time-step length is 1 hour. The first guess initial condition is taken to be a randomly perturbed initial condition from the observation data using Eqs. (4.2.1)-(4.2.2) with C_1=0.75 and C_2=0.5. The total number of minimization iterations is 12, and we chose n=2 and n=6 to carry out two preconditioning experiments. The value of the parameter p of the number of realizations used in (3.1.6) was chosen as p=120.

The variation of the normalized cost function versus the number of iterations is displayed in Fig. 5. The geopotential field difference between the first guess initial condition and the assimilation retrieved results is shown in Figs. 7. It is clear that our new preconditioning algorithm performs effectively. This case is characterized by strong randomly perturbed initial conditions and as such does not satisfy conditions for the tangent linear approximation to be valid. Thus using Courtier's method cannot obtain an accurate Hessian estimation which is a prerequisite for effectively accelerating the convergence rate of the minimization process. Using our new algorithm results in an impressive speed-up of the minimization convergence rate, which reveals that the error in the estimation of the Hessian is sizably reduced.

5.2. Experiments using the adiabatic version of the NASA GEOS-1 GCM.

We carried out two 4-D Var preconditioning experiments similar to the case used in Subsection 4.2, taking the total number of minimization iterations to be 8, and we chose n=2 and n=4 for the two experiments, respectively.

The variation of the normalized cost function versus the number of iterations is displayed in Fig. 8 and the surface pressure field difference between the first guess initial condition and the assimilation retrieved results is shown in Fig. 9. From these two figures we conclude that our new algorithm does indeed possess the ability to speed up the convergence rate of the minimization algorithm. As n increases, the validity of the tangent linear approximation increases, resulting in an increased accuracy of the estimated Hessian and a better ability to improve convergence rate of the minimization algorithm. The value of the cost function attained after 8 minimization iterations with the original Courtier's preconditioning method required just 3 iterations (in the n=4 case) and 4 iterations (in the n=2 case) when the new preconditioning algorithm was applied. Other results of 4-D Var preconditioning with different first guess initial conditions also confirm that our new algorithm achieves a significant speed-up in the convergence rate of minimization algorithm (figures omitted). In Fig. 8, we observe that the convergence rate of the minimization is slower in the first few iterations. This is due to the fact that at the beginning of the minimization process, the L-BFGS minimization algorithm was not provided with enough information to allow it to use its own preconditioning scheme.

A study of the reduction in the condition number of the estimated Hessian also shows the new algorithm to perform very effectively. We only know the diagonal elements of estimated Hessian at $\vec{X}_{min}$ (denoted here by $H_{m,d}$, the subscript m denoting the minimum state $\vec{X}_{min}$ while d denotes the diagonal matrix) in our 4-D Var preconditioning experiments. Since in a nonlinear system, the Hessian matrix of cost function depends upon the control variables, as the minimization process proceeds, and retrieved control variable fields get closer to observations in a least-squares sense, H will approach H_m. In our new algorithm, at different stages $\frac{k}{n}$, as k decreases, the minimum state $\vec{X}_{k,m}$ gets closer to $\vec{X}_m = \vec{X}^{obs}(t_0)$. Thus we may use the condition numbers of the intermediate Hessians $H_{k,m,d}$ at different stages $\frac{k}{n}$ of minimization as indicators for estimating the reduction in the condition number of the Hessian as the preconditioned minimization algorithm progresses.

In Table 1, $K_{k,m,d}$ is the condition number of non-preconditioned Hessian $H_{k,m,d}$ corresponding to stage $\frac{k}{n}$, and $K_{p,\ k,m,d}$ is the condition number of the preconditioned Hessian $H_{p,\ k,m,d}$

$$H_{p,\ k,m,d} = H_{m,d}^{-\frac{1}{2}} H_{k,m,d} H_{m,d}^{-\frac{1}{2}} \tag{4.3.1}$$

where $H_{m,d}$ is the Hessian preconditioner estimated with the original Courtier's method.

From Table 1, we see that after preconditioning with Courtier's method, the condition number of $H_{p,\ k,m,d}$ remains still quite large, especially at the beginning of the 4-D Var process. However using our new algorithm, the condition number of the preconditioned $H_{k,m,d}$ will equal unity. This reveals the reason of the efficiency of our new algorithm.

In this paper, all analyses are based on the "model generated observations" assumption. Extension to cases involving model systematic error or involving real observational data requires the inclusion of additional terms in cost function (3.1.1) and redefines the minimum point of the new cost function (e.g., by Eq. (3.3.4)), implying that general conclusions previously derived may still apply.

6. Determination of the parameter p

Another aspect to be considered is how to minimize the computational cost of the Hessian estimation method, consisting here in choosing an optimal number p of realizations of perturbed gradients of the cost function and using economical Hessian estimation methods. In this section we focus the discussion on how to determine the value of the parameter p in Eq. (3.1.6).

First let us consider this issue for linear systems. To obtain a basic insight, we analyzed a simple example, i.e., let $\vec{X}$ in Eq. (3.4.3) be a scalar variable, $B = 1$ and let $F(x) = Ax$, where A is a constant coefficient. We calculated the relationship between p and the accuracy of the estimated Hessian (Fig. 10). The results obtained show that for estimating the Hessian r.h.s. term of (3.1.6) with a relative error of less than 1% requires p to assume a value of about 3000, and for a relative error of less than 10% the parameter p is required to satisfy $30 < p < 2000$. Thus in linear systems, p may be chosen between 30 and 60 as in Courtier et al.(1994). Since in a linear system, the Hessian has quadratic form (see the second term of RHS of Eq. (3.4.1)), this basic estimation is valid only for simple linear systems.

In nonlinear systems, both $R_{H\,critical}$ and $n_{critical}$ depend on chosen p(see Subsection 4.1). Comparing Figs. 1 and 7, one finds that the relative error of the estimated Hessian in a nonlinear system is larger than that in a linear system. So if the same accuracy of the estimated Hessian is required, one has to use a larger value of p. Fig. 11 depicts variation of the average variance of diagonal elements of the estimated Hessian versus p using the adiabatic version of the NASA GEOS-1 ARIES GCM and its adjoint. This figure shows that the average variance of Hessian will tend to a low stable level when p exceeds values of between 90-120. This value may be used as a reference for choosing p.

In operational 3-D or 4-D Var applications, if the convergence criterion of minimization algorithm is met, we may choose a lower value of p, such as 30-60 to reduce computational costs. In Section 5, a value of p=60 was chosen for the 4-D Var experiments with NASA GEOS-1 GCM; For the experiments with the shallow water equations model, since the computational cost is not as important a consideration, a larger value of p, i.e., p=120, was chosen.

7. Reducing the computational cost of Hessian estimation

Since using Eq. (3.1.6) to estimate the Hessian requires running the adjoint model p times, we need to pay attention to the critical issue of computational cost. In fact, for some strong nonlinear cases, if one does not apply our new algorithm, the convergence rate of the minimization processes will be extremely slow due to ill-conditioning of the corresponding Hessian matrices. Thus the computational costs exceed the cost required for applying our algorithm to accurately estimate the Hessian.

In Appendix C we provide such an example, a variational data assimilation case which does not apply our new algorithm and display an extremely slow convergence rate.

We ran the experiments in Appendix C on the CRAY YMP Supercomputer. The CPU time for carrying out 500 minimization iterations(the number of function calls is 559) in the variational data assimilation experiment without preconditioning required 232.988 S; for carrying out 500 VDA minimization iterations(the number of function calls is 728) using the original Courtier's method preconditioning required a CPU of 350.164 S; however when we ran the n=2 case with the new algorithm and 12 minimization iterations(the number of function calls is 29), the CPU time required is 95.523 S while for the n=6 case(the number of function calls is 18) it is 255.687 S. All CPU times quoted in above experiments include the CPU time required for estimating the Hessian matrix. From this analysis, we may conclude that, using our new Hessian estimate algorithm, one may achieve significant computational cost savings while obtaining more satisfactory minimization results. It should be noted that when using our new algorithm, most of the CPU time is spent on estimating the Hessian matrix which requires running the adjoint model $n \times p$ times. For practical implementation of our new algorithm, this computational burden issue should be paid attention to.

One may try several economical alternative estimation methods for reducing the computational cost for estimating Hessian with our algorithm, such as using a lower value of p. In practical applications, as suggested by Courtier et al.(1994), the Hessian estimate could be done off-line. This is based on the fact that the main features of the observational network data are fairly stable from day to day. Thus a Hessian estimate from observational data on a certain day may be used effectively as a preconditioner for the next several days. Another way to reduce the computational cost, suggested by Courtier et al.(1994) is that of using a coarser grid point model to estimate the Hessian, then using it on a finer mesh as a preconditioner. Besides, we may use the adiabatic version of the adjoint model to estimate the Hessian for 3-D or 4-D Var while using a full-physics forecast model, like Zupanski(1993b). Using these approaches, one can obtain a faster convergence rate of 3-D or 4-D Var minimization without incurring high computational cost. In the present paper, we implement a method using a coarser grid point model to estimate the Hessian, thus alleviating to a large extent the computational cost attached to the implementation of this new algorithm.

7.1. Use of a coarser grid point model to estimate Hessian in the new algorithm.

We carried out a series of experiments similar to those in Subsection 5.1 but using a coarser grid point model to estimate the Hessian. The model resolution for estimating Hessian matrix is $(\Delta\theta, \Delta\lambda)$=$(15^o, 15^o)$, respectively. Thus the allowed time-step length is extended to 1.5 hours. To further reduce computational cost, a smaller parameter p was chosen, namely p=30. After obtaining the coarse grid point estimated Hessian, we interpolated it to a fine mesh of grid points. Then we used the interpolated Hessian as the preconditioner to rerun the variational data assimilation cases in Subsection 5.1 with a finer mesh resolution (see also Zhu et al. 1994 for a similar approach).

The results are shown in Fig. 12. We find that the efficiency of the Hessian estimated from the coarse grid point model is just slightly worse than that of the original estimated Hessian, while the computational cost is sizably reduced. As the number of grid points and the value of p decrease, they allow us to carry out the task on the CRAY YMP supercomputer to a higher level of vectorization and compiler optimization. The CPU times used for estimating Hessian with the original model are 82.264 S(n=2) and 246.664 S(n=6), respectively, while with the coarse mesh model they are reduced to just 4.049 S(n=2) and 12.013 S(n=6), respectively. This is a very encouraging result. Some additional test experiments related to the off-line Hessian estimate and use of adiabatic version of adjoint model to estimate the Hessian for 3-D or 4-D Var with full-physics forecast model will be presented in a forthcoming paper. One should note that the speed-up obtained by using the coarse grid estimated Hessian exceeds the ratio of computational cost reduction expected. We hypothesize that this is due to better capabilities of the optimizing complier induced by the reduced array size.

8. Summary and conclusions

In this paper, we presented the development of the tangent linear model and adjoint of adiabatic version of NASA GEOS-1 C-Grid GCM. We then analyzed Courtier's preconditioning method using estimated Hessian as well as a newly proposed Hessian estimation algorithm extending original Courtier's method to some 3-D and 4-D Var problems characterized by stronger nonlinear properties (such as we may be confronting in highly nonlinear physics packages in numerical weather prediction).

Since analysis of observations in Subsection 3.3 showed that 3-D or 4-D Var preconditioning research involving real observation data will be more difficult, we assumed in this paper that the model is good enough and the error in data is small enough to allow use of model generated data as "observations" in all 4-D Var experiments. Issues involving use of real observational data will be analyzed in a forthcoming paper.

The impact of validity of the tangent linear approximation on accuracy of the estimated Hessian in the preconditioning method proposed by Courtier et al.(1994) was first analyzed for a simple analytic case. We then analyzed the results of preconditioning experiments using the variational data assimilation systems of both NASA SLSI global S-W equations model and the adiabatic version of NASA GEOS-1 C-Grid GCM. The results show that the validity of tangent linear approximation strongly impacts upon the accuracy of estimated Hessian. We found that in many instances if an accurate Hessian estimate is required, restriction on the validity of the tangent linear approximation with Courtier's method will be so stringent as to impose a modification of Courtier's method to fit cases of interest. (nonlinearity in 3-D or 4-D Var)

We proposed an extension of Courtier's estimated Hessian preconditioning method to cases with higher nonlinearity by designing a new algorithm reducing error in the Hessian estimation induced by lack of validity of the tangent linear approximation. We applied our new algorithm to variational data assimilation cases with stronger nonlinearity using both NASA SLSI global S-W equations model and the adiabatic version of NASA GEOS-1 C-Grid GCM. The results show that our new algorithm speeds-up convergence rate of the minimization algorithm better than the original Courtier's method. Since we did not make any restrictive assumptions when deriving this algorithm, it has the potential of being applicable to the preconditioning of any nonlinear 3-D or 4-D Var system of interest. Some more complex data-model tests of this new algorithm will be carried out in future research.

We found that by applying Courtier's method directly to nonlinear cases, one cannot obtain an accurate estimated Hessian matrix, no matter how large a value of p was used. We also concluded that if the parameter p in Courtier's method is fixed, no matter how accurate the tangent linear approximation is, one cannot reduce the error of the estimated Hessian matrix beyond a threshold value.

We also addressed the issue of how to determine the value of the parameter p in the method used for estimating the Hessian matrix. We found that the relative error in estimating the Hessian matrix for a nonlinear system is larger than that of a linear system if the same value of p is used. In linear systems, the value of p may be chosen between 30 and 60 like in Courtier et al.(1994). In nonlinear systems, in order to obtain a more accurate Hessian, one may attempt to choose p between 90 to 120. If the convergence criterion of the minimization algorithm is satisfied, one may relax this requirement and choose a lower value of p, such as 30-60, for the sake of computational efficiency.

Finally, we discuss issues related to computational cost. We found out that for some strong nonlinear data assimilation cases, one should use some very effective preconditioning methods such as provided by our new algorithm, otherwise the convergence rate will be intolerably slow. A new strategy was tested using a coarser grid point model to estimate the Hessian and applying the resulting coarse mesh estimated Hessian as the preconditioner for variational data assimilation using a fine mesh of grid points model for reducing computational cost. The results obtained are encouraging and point to major computational cost savings.

Acknowledgments: The authors would like to thank L. L. Takacs and T. Wang for their very kind help. They provided all computer codes and documents of NASA GEOS-1 C-Grid GCM and all necessary data to run the model as well as a wealth of helpful advice for this study. We wish to thank Dr. M. Zupanski of NMC for some critical and beneficial suggestions. The incisive remarks and constructive criticism of two anonymous reviewers, which contributed to improvement in the presentation of this paper is hereby acknowledged. This work has been supported by NASA grant NAG-5-1660.

Appendix A

Investigation of the impact of validity of tangent linear approximation on the accuracy of estimated Hessian: A simple example

We provide a simple example whose analysis shows that the impact of validity of tangent linear approximation on the accuracy of estimated Hessian using Courtier's preconditioning method is very serious.

Let $\vec{X}$ be a scalar variable and $B = 1$, let F_n in Eq. (3.1.3) be given by

$$F_n = x^{1+\alpha} \tag{A.1}$$

with

$$\alpha << 1 \tag{A.2}$$

Since

$$x^{1+\alpha} = x \ + \ x(\ln x)\alpha \ + \ O(\alpha^2) \tag{A.3}$$

we have

$$F_n \simeq x(1+\beta) \tag{A.4}$$

where $\beta = \alpha \ln x$. Since we wish F_n to possess good linear properties, we choose $x = x_0 + \Delta x$ with x_0=100 and $|\Delta x| << x_0 \ln x_0$, $\alpha = 0.0048251 << 1$. Using a Taylor expansion, we obtain

$$\beta \simeq \alpha \ln x_0 = const \tag{3.4.12}$$

Hence F_n is a very weak nonlinear function. Fig. 13 shows that the relative error RF of F_n

$$RF = \frac{x^{1+\alpha} - x(1+\alpha \ln x_0)}{x(1+\alpha \ln x_0)} \tag{A.5}$$

is very small. We choose N=144, which corresponds to a 12 hour data assimilation window with a 5 minutes integration time step, mimicking a typical atmospheric numerical prediction model (e.g., such as in NASA GEOS-1 C-Grid GCM). The initial conditions and observations were chosen to yield a relative error of the tangent linear approximation given by

$$\|\mathbf{D}\|/\|\mathbf{L}\| = \frac{\left\| F\left(\vec{X}(t_0)\right) - F\left(\vec{X}^{obs}(t_0)\right) - L_2\left(\vec{X}(t_0) - \vec{X}^{obs}(t_0)\right)\right\|}{\left\| L_2\left(\vec{X}(t_0) - \vec{X}^{obs}(t_0)\right)\right\|} = 15\% \quad (A.6)$$

The value of the ratio $\|\mathbf{D}\|/\|\mathbf{L}\|$ is within the range of a good degree of validity of tangent linear approximation (Rabier and Courtier 1992; and Li et al. 1993, 1994).

Yet even for this weak nonlinear operator F_n and with satisfactory validity of tangent linear approximation, we find that the estimated Hessian has a large relative error. The relative error of the estimated Hessian is given by

$$R_H = \frac{|J''_{ESTIMATE} - J''_{REAL}|}{|J''_{REAL}|} \quad (A.7)$$

where J''_{REAL} is the real analytical value of Hessian, $J''_{ESTIMATE}$ is estimated Hessian. We analyzed three different types of $J''_{ESTIMATE}$. The first type is analytical value of Hessian at the minimum point $\left(\vec{X}(t_0) = \vec{X}^{obs}(t_0)\right)$, while the others are estimated Hessians using Eq. (3.1.6) with p=30 and p=3000, respectively. We obtained values of R_H given by 158%, 183% and 157%, respectively. Since operators in realistic NWP models possess much stronger nonlinear properties than the operator F_n chosen in our example, our analysis reveals that an accurate Hessian matrix cannot be estimated with Courtier's method applied directly to 3-D or 4-D Var using a realistic nonlinear NWP model.

Appendix B

The new Hessian estimation algorithm

Our new Hessian estimation algorithm, which can be applied to 3-D Var or 4-D Var problems characterized by higher nonlinearity, is as follows:

Let the distance between initial conditions and the initial time observations be denoted by

$$\Delta\vec{X} = B\left(\vec{X}^{initial}(t_0)\right) - \vec{X}^{obs}(t_0) \qquad (B.1)$$

choose $n \geq 1$, let $k = n - 1$ and

$$\frac{1}{n}\|\Delta\vec{X}\| \leq \varepsilon \qquad (B.2)$$

where ε is a critical value, (i.e. the value of the perturbation $\frac{1}{n}\|\Delta\vec{X}\|$ which acting upon $B\left(\vec{X}_{min}(t_0)\right)$ yields a value smaller than the prescribed threshold value ε).

(i) substitute $\vec{X}_k^{obs}(t_0) = \left(\vec{X}^{obs}(t_0) + \frac{k}{n}\Delta\vec{X}\right)$ for $\vec{X}^{obs}(t_0)$ in the cost function (3.4.3), and define an intermediate cost function J_k,

$$\begin{aligned} J_k(\vec{X}_k(t_0)) = \frac{1}{2}&\left(B\left(\vec{X}_k(t_0)\right) - (\vec{X}^{obs}(t_0) + \frac{k}{n}\Delta\vec{X})\right)^T \\ &\mathbf{W}(t_0)\left(B\left(\vec{X}_k(t_0)\right) - (\vec{X}^{obs}(t_0) + \frac{k}{n}\Delta\vec{X})\right) + \\ &+\frac{1}{2}\left(B\left(F(\vec{X}_k(t_0))\right) - B\left(F\left(B^*\left(\vec{X}^{obs}(t_0) + \frac{k}{n}\Delta\vec{X}\right)\right)\right)\right)^T \\ &\mathbf{W}(t_N)\left(B\left(F(\vec{X}_k(t_0))\right) - B\left(F\left(B^*\left(\vec{X}^{obs}(t_0) + \frac{k}{n}\Delta\vec{X}\right)\right)\right)\right) \end{aligned} \qquad (B.3)$$

where $\vec{X}_k(t_0)$ are the intermediate initial conditions at stage k; $F\left(\vec{X}_k(t_0)\right)$ is the model operator integrated from initial time t_0 to t_N starting from $\vec{X}_k(t_0)$; and $\left(B\left(F\left(B^*\left(\vec{X}^{obs}(t_0) + \frac{k}{n}\Delta\vec{X}\right)\right)\right)\right)$ is the model operator starting from intermediate observations $\vec{X}_k^{obs}(t_0)$ to generate $\vec{X}_k^{obs}(t_N)$.

(ii) calculate the diagonal elements of the estimated Hessian J_k using Eq. (3.1.6), and use them as the preconditioner.

(iii) perform the preconditioning minimization iteration process to minimize the original cost function J in (3.4.3).

(iv) if $k = 0$, stop the computation process, otherwise, update k by $k - 1$, and go to (i).

Appendix C

An example illustrating an extremely slow convergence rate.

We rerun the data assimilation experiment in Subsection 5.1 without applying our new algorithm and with a large number of minimization iterations. The result is provided in Fig. 14. We can clearly see that when our new algorithm is not applied, the variational data assimilation minimization process displays an extremely slow convergence rate. After 500 iterations, the normalized cost function is reduced to 74%(without preconditioning) and 32%(using original Courtier's method); yet using our new algorithm, only 4 minimization iterations were sufficient to reduce the normalized cost function to 31.5% (n=2) and 13%(n=6) of its original value, respectively(Fig. 5). This example shows that, for some strong nonlinear data assimilation cases, one must use some very effective preconditioning methods such as provided by our new algorithm.

References

Aitchison, J., and I. R. Dunsmore, 1975: *Statistical prediction analysis*, Cambridge University Press, New York, 273 pp.

Arakawa, A., and V. R. Lamb, 1981: A potential enstrophy and energy conserving scheme for the shallow water equations, *Mon. Wea. Rev.*, **109**, 18-36.

Asselin, R., 1972: Frequency filter for time integrations, *Mon. Wea. Rev.*, **100**, 487-490.

Axelsson, O., 1976: A class of iterative methods for finite element equations, *Comput. Methods Appl. Mech. Eng.*, **9**, 123-137.

Axelsson, O., 1985: A survey of vectorizable preconditioning methods for large scale finite element matrix problems, *BIT*, **25**, 166-187.

Axelsson, O., and V. A. Barker, 1984: *Finite Element Solution of Boundary Value Problems, Theory and Computation*, Academic Press, 432 pp.

Axelsson, O., 1994: *Iterative Solution Methods*, Cambridge University Press, 654 pp.

Bates, J. R., F. H. M. Semazzi, R. W. Higgins, and R. W. Barros, 1990: Integration of the shallow water equations on the sphere using a vector semi-Lagrangian scheme with a multigrid solver, *Mon. Wea. Rev.*, **118**, 1615-1627.

Ben-Israel, A., and T. N. E. Greville, 1974: *Generalized inverses: theory and applications*, Wiley, New York, 395 pp.

Brown, J. A., and K. Campana, 1978: An economical time-differencing system for numerical weather prediction, *Mon. Wea. Rev.*, **106**, 1125-1136.

Concus, P., G. Golub, and D. O'Leary, 1976: *A generalized conjugate gradient method for the numerical solution of elliptic partial differential equations, in Sparse Matrix Computations*, J. Bunch and D. Rose (eds.), Academic Press, New York, pp. 309-322.

Conn, A. R., N. I. M. Gould, and Ph. L. Toint, 1992: *LANCELOT - A Fortran Package for Large-Scale Nonlinear Optimization (Release A)*, Springer-Verlag, 330 pp.

Courtier, P., 1987: *Application du contrôle optimal á la prévision numerique en Météorologie.* These de doctorat de l'universite Paris 6 (Reading 1987).

Courtier, P., and O. Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint equation - Part I. Numerical results. *Q. J. R. Meteorol. Soc.*, **113**, 1329-1347.

Courtier, P., J. -N. Thepaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach, *Q. J. R. Meteorol. Soc.*, **120**, 1367-1387.

Derber, J.C., 1989: A variational continuous assimilation technique.*Mon. Wea. Rev.*, **117**, 2437-2446.

Eirola T and O. Nevanlinna,1989: Accelerating with rank-one updates. *Lin. Alg. Appl.* ,**121**, 511-520.

Forsythe, G. E., and E. G. Strauss, 1955: On best conditioned matrices, *Proc. Ameri. Math. Soc.*, **6**, 340-345.

Fox-Rabinovitz, M., H. M. Helfand, A. Hou, L. L. Takacs, and A. Molod, 1991: Numerical experiments on forecasting, climate simulation and data assimilation with the new 17 layer GLA GCM, *Proceedings of the AMS Ninth Conference on Numerical Weather Prediction*, 21-25 October 1991, Denver, CO, 506-509.

Gates, W. L., 1992: AMIP: The Atmospheric Model Intercomparison Project, *Bull. Ameri. Meteor. Soc.*, **73**, 1962-1970.

Gauthier, P., 1992: Chaos and quadri-dimensional data assimilation: a study based on the Lorenz model, *Tellus*, **44A**, 2-17.

Gill, P. E., and W. Murray, 1979: Conjugate-gradient methods for large-scale nonlinear optimization, *Technical Report SOL 79-15*, Systems Optimization Laboratory, Department of Operations Research, Stanford University, 61pp.

Gill, P. E., W. Murray, and M. H. Wright, 1981: *Practical optimization*, Academic Press, Inc. New-York, 401pp.

Golub, G. H., C. F. Van Loan, 1989: *Matrix computations*, 2nd Edition, The Johns Hopkins University Press, Baltimore and London, 642 pp.

Helfand, H. M., M. Fox-Rabinovitz, L. Takacs, and A. Molod, 1991: Simulation of the planetary boundary layer and turbulence in the GLA GCM, *Proceedings of the AMS Ninth Conference on Numerical Weather Prediction*, 21-25 October 1991, Denver, CO, 514-517.

Ingleby, N. B., and A. C. Lorenc, 1993: Bayesian quality control using multivariate normal distributions, *Q. J. R. Meteorol. Soc.*, **119**, 1195-1225.

Kalnay, E. M. Kanamitsu, J. Pfaendtner, J. Sela, M. Suarez, J. Stackpole, J. Tuccillo, L. Umscheid, and D. Williamson, 1989: Rules for the interchange of physical parameterizations, *Bull. Ameri. Meteor. Soc.*, **70**, 620-622.

Kharcenko S.A. and A.Yu. Yeremin, 1995: Translation based preconditioners for the GMRES(k) Method.*Num. Lin. Algebra with Applications*, **2(1)**, 51-77.

Lacarra, J.-F., and O. Talagrand, 1988: Short-range evolution of small perturbations in a barotropic model, *Tellus*, **40A**, 81-95.

Lawson, C. L., and R. J. Hanson, 1974: *Solving least squares problems*, Prentice-Hall, Englewood Cliffs, N. J., 340 pp.

Li, Y., I. M. Navon, P. Courtier, and P. Gauthier, 1993: Variational data assimilation with a semi-Lagrangian semi-implicit global shallow water equation model and its adjoint, *Mon. Wea. Rev.*, **121**, 1759-1769.

Li, Y., I. M. Navon, W. Yang, X. Zou, J. R. Bates, S. Moorthi, and R. W. Higgins, 1994: Four-Dimensional Variational Data Assimilation Experiments with a Multilevel semi-Lagrangian semi-Implicit GCM. *Mon. Wea. Rev.*, **122**, 5, 966-983.

Liu, D. C., and J., Nocedal, 1989: On the limited memory BFGS method for large scale optimization, *Mathematical Programming*, **45**, 503-528.

Lorenc, A. C., and O. Hammon, 1988: Objective quality control of observations using Bayesian methods: Theory and a practical implementation, *Q. J. R. Meteorol. Soc.*, **114**, 515-543.

Luenberger, D. G., 1984: *Linear and Nonlinear Programming*, Second Edition, Addison-Wesley Publishing Company, 491 pp.

Nashed, M. Z., 1976: *Generalized inverses and applications: Proceedings of an advanced seminar*, Mathematics Research Center, University of Wisconsin - Madison, No. 32, Academic Press, New York, 1054 pp.

Navon, I. M., X. Zou, J. Derber, and J. Sela, 1992: Variational data assimilation with an adiabatic version of the NMC spectral model, *Mon. Wea. Rev.*, **120**, 1433-1446.

Navon, I. M., X. Zou, M. Berger, P.K.H. Phua, T.Schlick and F.X.Le Dimet, 1992a: Numerical experience with Limited Memory Quasi-Newton and truncated Newton Methods, *Optimization Techniques and Applications*, Vol 1, Edited by K. H. Phua et al. World Scientific Publishing Co., 1232 pp., 33-48.

Navon, I. M., X. Zou, M. Berger, P.K.H. Phua, T.Schlick and F.X.Le Dimet, 1992b: Testing for Reliability and Robustness of Optimization Codes for Large Scale Optimization Problems, *Optimization Techniques and Applications*, Vol 1, Edited by K. H. Phua et al. World Scientific Publishing Co., 1232 pp., 445-480.

NAG Fortran Library Manual, Mark 15, Volume 4, *E04-Minimizing or Maximizing a Function*, NAG Ltd, Wilkinson House, Jordan Hill Road, OXFORD, June 1991.

Nemhauser, G. L., A. H. G. Rinnooy Kan, M. J. Todd, 1989: *Optimization*, Handbooks in Operations Research and Management Science, Volume 1, 709 pp., 5-6.

Oppe, T. C., W. H. Holter, and I.M.Navon, 1994: Parallelizable PCG Methods for the CRAY-YMP and the TMV CM-2, In *Proceedings of the Second International Colloquium on Numerical Analysis*, D. Bainov and V. Covachev (eds), VSP International Science Publishers, Zeist, The Netherlands, 155-164, 198 pp.

Ortega, J. M., 1987: *Introduction to Parallel and Vector Solution of Linear Systems, Frontiers of Computer Science*, Plenum Press, 305 pp.

Rabier, F., and P. Courtier, 1992: Four-dimensional assimilation in the presence of baroclinic instability, *Q. J. R. Meteorol. Soc.*, **118**, 649-672.

Rabier, F., Klinker, P. Courtier, B. Strauss, and M. Herveou, 1994: Sensitivity of forecast error to initial conditions using the adjoint technique, *Proceedings of Tenth Conference on Numerical Weather Prediction*, July 18-22, 1994, Portland, Oregon pp. 463-465.

Sadourny, R., 1975: The dynamics of finite difference models of the shallow water equations, *J. Atmos. Sci.*, **32**, 680-689.

Schubert, S. D., J. Pfaendtner, and R. Rood, 1993: An assimilated data set for Earth Science applications, *Bull. Ameri. Meteor. Soc.*, **74**, 2331-2342.

Schubert, S., M. Suarez, C. K. Park, and S. Moorthi, 1993: GCM simulations of interseasonal variability in the Pacific/North American region, *J. Atmos. Sci.*, **50**, 1991-2007.

Suarez, M. J., and L. L. Takacs, 1994: *Documentation of the ARIES/GEOS Dynamical Core*, Version 2, May 19, 1994.

Takacs, L. L., A. Molod, and T. Wang, 1994: *Goddard Earth Observing System (GEOS) General Circulation Model (GCM)*, Version 1, May, 17, 1994.

Takano, K., and M. G. Wurtele, 1982: *A fourth-order energy and potential enstrophy conserving difference scheme*, Air Force Geophysics Laboratory Report **AFGL-TR-82-0205, (NTIS AD-A126626)**, AFGL Hanscom AFB, Massachusetts, 01731, 85pp.

Thacker, W. C., 1989: The role of the Hessian matrix in fitting models to measurements, *J. Geophys. Res.*, **94**, 6177-6196.

Thépaut, J.-N., and P. Moll, 1990: Variational inversion of simulated TOVS radiances using the adjoint technique, *Q. J. R. Meteorol. Soc.*, **116**, 1425-1448.

Vukićević, T., 1991: Nonlinear and linear evolution of initial forecast errors, *Mon. Wea. Rev.*, **119**, 1602-1611.

Wang, Z., I. M. Navon, F. X. Le Dimet, and X. Zou, 1992: The second order adjoint analysis, theory and applications, *Meteorology and Atmospheric Physics*, **50**, 3-20.

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. *Computational Optimization and Applications*, **4**, 241-262.

Wishart, J., 1952: Moment coefficients of the k-statistics in samples from a finite population, *Biometrika*, **39**, 1-13.

Yang, W., and I. M. Navon, 1995a: *Documentation of the tangent linear model and its adjoint of the adiabatic version of the NASA GEOS-1 C-grid GCM (version 5.2)*, Technical Report, **FSU-SCRI-95T-52**, Supercomputer Computations Research Institute, Florida State University, 60 pp., available from SCRI, Florida State University, Tallahassee, FL 32308-4052, also to appear as a NASA Technical Memorandum.

Yang, W., and I. M. Navon, 1995b: *A New Hessian Preconditioning Method Applied to Variational Data Assimilation Experiments using Adiabatic Version of NASA/GEOS-1 GCM*, Technical Report, **FSU-SCRI-95-03**, Supercomputer Computations Research Institute, Florida State University, 43pp, available from SCRI, Florida State University, Tallahassee, FL 32308-4052.

Young, D. M., 1971: *Iterative solution of large linear systems, Computer Science and Applied Mathematics*, A series of Monographs and Textbooks, Academic Press, 570pp, 42, 214-215.

Zhu, Keyun, I. M. Navon, and X. Zou, 1994: Variational data assimilation with a variable resolution finite-element shallow-water equations model, *Mon. Wea. Rev.*, **122**, 946-965.

Zou, J., and G. Holloway, 1995: Improving steady-state fit of dynamics to data using adjoint equation for gradient preconditioning, *Mon. Wea. Rev.*, **123**, 199-211.

Zou, X., I. M. Navon, M. Berger, P. K. Phua, and T. Schlick and F. X. Le Dimet, 1993: Numerical experience with limited-memory quasi-Newton methods for large-scale unconstrained nonlinear minimization, *SIAM Journal on Optimization*, **3**, 3, 582-608.

Zou, X., I. M. Navon, and F. X. Le Dimet, 1992: Incomplete observations and control of gravity waves in variational data assimilation, *Tellus*, **44A**, 273-296.

Zupanski, M., 1993a: A preconditioning algorithm for large scale minimization problems, *Tellus*, **45A**, 478-492.

Zupanski, M., 1993b: Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment, *Mon. Wea. Rev.*, **121**, 2396-2408.

Captions

Fig. 1. Schematic diagram illustrating the new Hessian estimation algorithm.

Fig. 2. Variation of the logarithm of relative error of Hessian R_H with the index n of the intermediate cost functions.

a) n from 1 to 200.

b) n from 1 to 2000.

Fig. 3. Surface pressure filed of January 1, 1985 ECMWF 00Z data.

Fig. 4. Variation of normalized cost function J/J_0 versus the number of minimization iterations at the $k = n - 1$ stage.

Solid line: intermediate cost function J_k.

Dashed line: original cost function J.

a) n=2.

b) n=4.

c) n=8.

Fig. 5. Variation of the normalized cost function J/J_0 versus the number of minimization iterations using the NASA SLSI global S-W equations model.

Solid line: Without preconditioning.

Dashed line: preconditioning with Courtier's method to estimate Hessian.

Dotted line: preconditioning with new Hessian estimation algorithm, n=2.

Short-long-dashed line: preconditioning with new Hessian estimation algorithm, n=6.

Fig 6. Geopotential field used as the observation for carrying out variational data assimilation experiments using the NASA SLSI global S-W equations model.

Fig. 7. The difference of the geopotential fileds between the assimilation retrieved results and the first guess initial condition(the contour interval is 100 m).

a) Without preconditioning.

b) Preconditioning with Courtier's method to estimate Hessian.

c) Preconditioning with new Hessian estimation algorithm, n=2.

d) Preconditioning with new Hessian estimation algorithm, n=6.

Fig. 8. Variation of the normalized cost function J/J_0 versus the number of minimization iterations using the 4-D Var system of the adiabatic version of NASA GEOS-1 GCM.

Solid line: Without preconditioning.

Dashed line: preconditioning with Courtier's method to estimate Hessian.

Short-dashed line: preconditioning with new Hessian estimation algorithm, n=2.

Short-long-dashed line: preconditioning with new Hessian estimation algorithm, n=4.

Fig. 9. Difference of the surface pressure fields between the assimilation retrieved results and the first guess initial condition(the contour interval is 5 hPa).

a) Without preconditioning.

b) Preconditioning with Courtier's method to estimate Hessian.

c) Preconditioning with new Hessian estimation algorithm, n=2.

d) Preconditioning with new Hessian estimation algorithm, n=4.

Fig. 10. Variation of the ratio $(Hessian_{estimate}/Hessian_{real})$ with number of realizations p in a linear system.

Fig. 11. Variation of the average variance of the estimated Hessian with number of realizations p for the NASA/DAO GEOS-1 GCM.

Fig. 12. Variation of the normalized cost function J/J_0 with the number of minimization iterations using the NASA SLSI global S-W equations model.

Solid line: preconditioning with coarse grid estimated Hessian and new algorithm, n=2.

Dotted line: preconditioning with standard grid estimated Hessian and new algorithm, n=2.

Dashed line: preconditioning with coarse grid estimated Hessian and new algorithm, n=6.

Long-short-dashed line: preconditioning with standard grid estimated Hessian and new algorithm, n=6.

Fig. 13. Variation of the relative error RF of F_n with respect to x.

Fig. 14. Variation of the normalized cost function J/J_0 with the number of minimization iterations using the NASA SLSI global S-W equations model(for large number of iterations).

Solid line: Without preconditioning.

Dashed line: preconditioning with Courtier's method to estimate Hessian.