Meteorol. Atmos. Phys. 50, 3–20 (1992)

Meteorology and Atmospheric Physics

551.509.3

[1] Department of Mathematics and Supercomputer Computations Research Institute, Florida State University, Tallahassee, Florida, U.S.A.
[2] Grenoble, Laboratoire de Modelisation et Calcul, Grenoble Cedex, France
[3] Supercomputer Computations Research Institute, Florida State University, Tallahassee, Florida, U.S.A.

The Second Order Adjoint Analysis: Theory and Applications

Zhi Wang[1], I. M. Navon[1], F. X. Le Dimet[2], and X. Zou[3]

With 17 Figures

Received January 30, 1992
Revised May 26, 1992

Summary

The adjoint method application in variational data assimilation provides a way of obtaining the exact gradient of the cost function J with respect to the control variables. Additional information may be obtained by using second order information. This paper presents a second order adjoint model (SOA) for a shallow-water equation model on a limited-area domain. One integration of such a model yields a value of the Hessian (the matrix of second partial derivatives, $\nabla^2 J$) multiplied by a vector or a column of the Hessian of the cost function with respect to the initial conditions. The SOA model was then used to conduct a sensitivity analysis of the cost function with respect to distributed observations and to study the evolution of the condition number (the ratio of the largest to smallest eigenvalues) of the Hessian during the course of the minimization. The condition number is strongly related to the convergence rate of the minimization. It is proved that the Hessian is positive definite during the process of the minimization, which in turn proves the uniqueness of the optimal solution for the test problem.

Numerical results show that the sensitivity of the response increases with time and that the sensitivity to the geopotential field is larger by an order of magnitude than that to the u and v components of the velocity field. Experiments using data from an ECMWF analysis of the First Global Geophysical Experiment (FGGE) show that the cost function J is more sensitive to observations at points where meteorologically intensive events occur. Using the second order adjoint shows that most changes in the value of the condition number of the Hessian occur during the first few iterations of the minimization and are strongly correlated to major large-scale changes in the reconstructed initial conditions fields.

1. Introduction

The complete description of the initial atmospheric state in a numerical weather prediction method constitutes an important issue. The four-dimensional variational data assimilation (VDA) method offers a promising way to achieve such a description of the atmosphere. It consists of finding the assimilating model solution which minimizes a properly chosen objective function measuring the distance between model solution and available observations distributed in space and time. The control variables are either the initial conditions or the initial conditions plus the boundary conditions. The boundary conditions have to be specified so that the problem is well posed in the sense of Hadamard. In most of the unconstrained minimization algorithms associated with the VDA approach, the gradient of the objective function with respect to the control variables plays an essential role. This gradient is obtained through one direct integration of the model equations followed by a backwards integration in time of the linear adjoint system of the direct model.

VDA was first applied in meteorology by Marchuk (1974) and by Penenko and Obrazstov (1976). Kontarev (1980) further described how to apply the adjoint method to meteorological problems, while Le Dimet (1982) formulated the method in a general mathematical framework related to optimal control of partial differential equations. In the following years, a considerable number of experiments has been carried out on different two-dimensional (2-D) barotropic models by several authors, such as Courtier (1985); Lewis and Derber (1985); Derber (1985); Hoffmann (1986); Le Dimet and Talagrand (1986); Le Dimet and Nouailler (1986); Courtier and Talagrand (1987, 1990); Derber (1987); Talagrand and Courtier (1987); Lorenc (1988a and 1988b); Thacker and Long (1988); Zou et al. (1991). Thépaut and Courtier (1991); Navon et al. (1990, 1992) as well as Chao and Chang (1992) applied the method to 3-D operational NWP models. While major advances have been achieved in the application of the adjoint method, this field of research remains both theoretically and computationally active. Additional research to be carried out includes applications to complicated models such as multilevel primitive equation models related to distributed real data and the inclusion of physical processes in the VDA process.

The SOA model serves to study the evolution of the condition number of the Hessian during the course of the minimization. Two forward integrations of the nonlinear model and the tangent linear model and two backwards integrations in time of the first order adjoint (FOA) model and the SOA system are required to provide the value of Hessian/vector product. This Hessian/vector product is required in truncated Newton-type methods and may be used with the Rayleigh quotient power method to obtain the largest and smallest eigenvalues of the Hessian whose dimension is 1083×1083 for the test problem. The dimension of the Hessian will be more than $10^5 \times 10^5$ for 3-D primitive equations models. If the smallest eigenvalues of the Hessian of the cost function with respect to the control variables are positive at each iteration of the VDA minimization process, then the optimal solution of the VDA is unique. This statement is proven to be true for the shallow water equation model (Section 4.2). The variation of the condition number of the Hessian of the cost function with respect to number of iterations during the minimization process reflects the convergence rate of the minimization. It has been observed (Navon et al., 1992) that large scale changes occur in the process of minimization during the first 30 iterations, while during the ensuing iterations only small scale features are assimilated. This entails that the condition number of the Hessian of the cost function with respect to the initial conditions changes faster at the beginning of the minimization and then remains almost unchanged during the latter iterations. The condition number can also provide information about the error covariance matrix. The rate at which algorithms for computing the best fit to data converge depends on the size of the condition number and the distribution of eigenvalues of the Hessian. The inverse of the Hessian can be identified as the covariance matrix that establishes the accuracy to which the model state is determined by the data; the reciprocals of the Hessian's eigenvalues represent the variance of linear combinations of variables determined by the eigenvectors (Thacker, 1989).

The structure of the paper is as follows: the theory of the SOA is introduced in section 2. In section 3, a detailed derivation of the SOA model of the two-dimensional shallow water equations model is presented. A brief description of the FOA model is provided in Appendix A. Quality control methods for the verification of the correctness of the SOA model are then discussed in Appendix B. Issues concerning uniqueness of the solution and the evolution of the condition number of the Hessian during the course of the minimization as well as related issues of the structure of the reconstructed initial conditions are addressed in section 4. Section 5 is devoted to a sensitivity study of the solution with respect to distributed inaccurate observations. Finally a summary and conclusions are presented in section 6.

2. The SOA Model

*2.1 Theory of the SOA Model**

The forwards and backwards integrations of the nonlinear model and the adjoint model, respec-

* A brief description of the FOA model is provided in Appendix A.

tively, provide the value of the cost function J and its gradient. The following question may then be posed: can we obtain any information about the Hessian (second order derivative matrix) of the cost function with respect to the initial conditions by integrating the adjoint model equations? The calculation of the matrix of the second order derivatives is useful in many instances. For example, a Hessian/vector product is required in the truncated Newton large-scale nonlinear unconstrained optimization algorithm (Nash, 1985). Once the Hessian/vector product is available, the condition number of the Hessian may be obtained. This condition number may then be used to study the convergence rate of VDA. Analysis of the spectrum of the Hessian can provide an in-depth insight into the behavior of the large-scale minimization algorithms (Luenberger, 1984). We will show in this section that one integration of the SOA model yields a Hessian/vector product or a column of the Hessian of the cost function with respect to the initial conditions. Therefore, the SOA model provides an efficient way to compute the Hessian of the cost function by performing N integrations of the SOA model where N is the number of the components of the control variables vector. For a large dimensioned model, obtaining the full Hessian matrix proves to be a computationally prohibitive task beyond the capability of present day computers. The SOA approach will be used to conduct a sensitivity analysis of the observations in section 5 of this paper. We will also study the relative importance of observations distributed at different space and time locations.

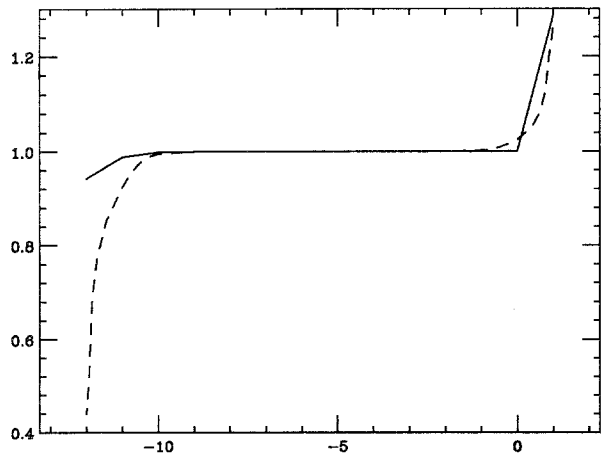


Fig. 1. Verifications of the correctness of the gradient calculation (dash line) and Hessian/vector product calculation (solid line) by FOA and SOA models, respectively

Assume that the model equations can be written as

$$\frac{\partial \boldsymbol{X}}{\partial t} = F(\boldsymbol{X}) \tag{2.1}$$

$$\boldsymbol{X}(t_0) = U \tag{2.2}$$

where $\boldsymbol{X}$ is the state vector (a three-component vector of $(u, v, \phi)^t$ in the shallow-water equations) in a Hilbert space χ whose inner product is denoted by $<,>$, t is the time, t_0 is the initial time, U is the initial condition of $\boldsymbol{X}$ and F is a function of $\boldsymbol{X}$. For any initial condition (2.2), (2.1) has a unique solution, $\boldsymbol{X}(t)$.

Let us define the cost function as

$$J(U) = \frac{1}{2}\int_{t_0}^{T} < W(C\boldsymbol{X} - \boldsymbol{X}^o), C\boldsymbol{X} - \boldsymbol{X}^o > dt \tag{2.3}$$

where W is a weighting matrix often taken to be the inverse of the estimate of the covariance matrix of the observation errors, T is the final time of the assimilation window, the objective function $J(U)$ is the weighted sum of squares of the distance between model solution and available observations distributed in space and time, $\boldsymbol{X}^o$ is an observation vector and the operator C represents the process of interpolating the model solution $\boldsymbol{X}$ to space and time locations where observations are available. The purpose is to find the initial conditions such that the solution of Eq. (2.1) minimizes the cost function $J(U)$ in a least-squares sense. The FOA model as defined by Eqs. (A.11), (A.12) may then be rewritten as

$$-\frac{\partial P}{\partial t} = \left(\frac{\partial F}{\partial \boldsymbol{X}}\right)^* P + C^* W(C\boldsymbol{X} - \boldsymbol{X}^o) \tag{2.4}$$

$$P(T) = 0. \tag{2.5}$$

where P represents the FOA variables vector. The gradient of the cost function with respect to the initial conditions is given by

$$\nabla_U J = P(t_0). \tag{2.6}$$

Let us now consider a perturbation, U', on the initial condition U. The resulting perturbations for the variables $\boldsymbol{X}$, P, $\hat{\boldsymbol{X}}$ and $\hat{P}$ may be obtained from Eqs. (2.1), (2.2), (2.4) and (2.5) as

$$\frac{\partial \hat{\boldsymbol{X}}}{\partial t} = \frac{\partial F}{\partial \boldsymbol{X}}\hat{\boldsymbol{X}} \tag{2.7}$$

$$\hat{X}(0) = U' \tag{2.8}$$

$$-\frac{\partial \hat{P}}{\partial t} = \left(\frac{\partial F}{\partial X}\right)^* \hat{P} + \left[\frac{\partial^2 F}{\partial X^2}\hat{X}\right]^* P + C^* W C \hat{X} \tag{2.9}$$

$$\hat{P}(T) = 0 \tag{2.10}$$

Eqs. (2.7), (2.8) and Eqs. (2.9), (2.10) are called the tangent linear and SOA models, respectively.

Let us denote the FOA variable after a perturbation U' on the initial condition U by $P_{U+U'}$, then according to definition

$$P_{U+U'}(t_0) = P(t_0) + \hat{P}(t_0). \tag{2.11}$$

Expanding $\nabla_{U+U'}J$ at U in a Taylor series and only retaining the first order term, results in

$$\nabla_{U+U'}J = \nabla_U J + \nabla^2 J \cdot U' + O(\| U' \|^2). \tag{2.12}$$

From Eq. (2.6), we know that

$$\nabla_{U+U'}J = P_{U+U'}(t_0). \tag{2.13}$$

Combining Eqs. (2.6), (2.11)–(2.13), one obtains

$$\hat{P}(t_0) = \nabla^2 J \cdot U' = HU' \tag{2.14}$$

where $H = \nabla^2 J$ is the second derivative of the cost function with respect to initial conditions.

If we set $U' = e_j$, where e_j is the unit vector with the j-th element equal to 1, the j-th column of the Hessian may be obtained by

$$He_j = \hat{P}(t_0). \tag{2.15}$$

Therefore, theoretically speaking, the full Hessian H can be obtained by M integrations of Eqs. (2.9), (2.10) with $U' = e_i$, $i = 1, \ldots, N$ where N is the number of the components of the control variables vector (the initial conditions $u(t_0)$, $v(t_0)$ and $\phi(t_0)$ in our case).

In summary, the j-th column of the Hessian of the cost function can be obtained by the following procedure:

(a) Integrate the model (2.1), (2.2) and the tangent linear model (2.7), (2.8) forward and store in memory the corresponding sequences of the states X_i and $\hat{X}_i$ $(i = 0, \ldots, M)$;
(b) Integrate the FOA Eqs. (2.4), (2.5) backwards in time and store in memory the sequence of P_i $(i = 0, \ldots, M)$;
(c) Integrate the SOA model (2.9), (2.10) backwards in time. The final value $\hat{P}(t_0)$, yields the j-th column of the Hessian of the cost function.

The verification of the correctness of FOA and SOA models is provided in Appendix B.

2.2 *The Estimate of the Condition Number of the Hessian*

Let us denote the largest and the smallest eigenvalues of the Hessian matrix H and their corresponding eigenvectors by $\lambda_{\max}$, $\lambda_{\min}$, $V_{\max}$ and $V_{\min}$, respectively. Then the condition number of the Hessian is given by

$$\kappa(H) = \frac{\lambda_{\max}}{\lambda_{\min}}. \tag{2.16}$$

Considering the eigenvalue problem $HU = \lambda U$ and assuming that the eigenvalues are ordered in decreasing order with $|\lambda_1| \geqslant |\lambda_2| \geqslant \cdots \geqslant |\lambda_n|$, an arbitrary initial vector X_0 may be expressed as a linear combination of the eigenvectors $\{U_i\}$

$$X_0 = \sum_{i=1}^{n} c_i U_i. \tag{2.17}$$

If λ_i is an eigenvalue corresponding to the i-th eigenvector U_i, the product of m multiplications of the Hessian H with Eq. (2.17) result in,

$$X_m = \sum_{i=1}^{n} c_i \lambda_i^m U_i \tag{2.18}$$

where

$$X_m = H^m X_0.$$

Factoring λ_1^m out, we obtain

$$X_m = \lambda_1^m \sum_{i=1}^{n} c_i \left(\frac{\lambda_i}{\lambda_1}\right)^m U_i. \tag{2.19}$$

Since λ_1 is the largest eigenvalue, the ratio $\left(\frac{\lambda_i}{\lambda_1}\right)^m$ approaches zero as m increases (suppose $\lambda_1 \neq \lambda_2$). Therefore we may write

$$X_m = \lambda_1^m c_1 U_1. \tag{2.20}$$

From (2.20) observe that the largest eigenvalue may then be calculated by

$$\lambda_1 = \frac{j\text{th component of } X_{m+1}}{j\text{th component of } X_m}. \tag{2.21}$$

This technique is called the power method (Strang, 1986). We can normalize the vector X_m by its largest component in absolute value. If we denote

the new scaled iterate to be X'_m, then

$$X_{m+1} = HX'_m \tag{2.22}$$

and the method is called the power method with scaling. It gives us an eigenvector whose largest component is 1.

The main steps in the power method with scaling algorithm are:

(a) Generate a starting vector X_0.
(b) Form a matrix power sequence $X_m = HX_{m-1}$.
(c) Normalize X_m so that its largest component is unity.
(d) Return to step (b) until convergence

$$|X_m - X_{m-1}| \leqq 10^{-6}$$

is satisfied or a prescribed upper limit of the number of iterations has been attained.

The smallest eigenvalue of H may also be computed by applying the shifted iterated power method to the matrix $Z = z \cdot I - H$, where z is the majorant of the spectral radius of H and I the identity matrix.

Since the Hessian H is symmetric, we will employ here the Rayleigh quotient power method which has a better convergence rate.

3. The Derivation of the SOA for the Shallow Water Equations Model

In this section, we consider the application of the SOA model to a two-dimensional limited-area shallow water equations model. Our purpose is to illustrate how to derive the SOA model explicitly.

The shallow water equations model may be written as

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial y} + fv - \frac{\partial \phi}{\partial x} \tag{3.1}$$

$$\frac{\partial u}{\partial t} = -u\frac{\partial v}{\partial x} - v\frac{\partial v}{\partial y} - fu - \frac{\partial \phi}{\partial y} \tag{3.2}$$

$$\frac{\partial \phi}{\partial t} = -\frac{\partial (u\phi)}{\partial x} - \frac{\partial (v\phi)}{\partial y} \tag{3.3}$$

where u, v, ϕ and f are the two components of the horizontal velocity, geopotential field and the Coriolis factor, respectively.

We shall use initial conditions due to Grammeltvedt (1969)

$$h = H_0 + H_1 \tanh \frac{9(y - y_0)}{2D} + H_2 \operatorname{sech} \frac{9(y - y_0)}{D} \sin \frac{2\pi x}{L} \tag{3.4}$$

where $\phi = gh$, $H_0 = 2000$ m, $H_1 = -220$ m, $H_2 = 133$ m, $g = 10\,\mathrm{m\,sec^{-2}}$, $L = 6000$ km, $D = 4400$ km, $f = 10^{-4}\,\mathrm{sec^{-1}}$, $\beta = 1.5 \times 10^{-11}\,\mathrm{sec^{-1}\,m^{-1}}$. Here L is the length of the channel on the β plane, D is the width of the channel and $y_0 = \frac{D}{2}$ is the middle of the channel. The initial velocity fields were derived from the initial height field via the geostrophic relationship, and are given by

$$u = -\frac{g}{f}\frac{\partial h}{\partial y} \tag{3.5}$$

$$v = \frac{g}{f}\frac{\partial h}{\partial x}. \tag{3.6}$$

The time and space increments used in the model were

$$\Delta x = 300\,\mathrm{km}, \quad \Delta y = 220\,\mathrm{km}, \quad \Delta t = 600\,\mathrm{s}. \tag{3.7}$$

which means that there are 21×21 grid point locations in the channel and the number of the components of initial condition vector $(u, v, \phi)^t$ is 1083. Therefore the Hessian of the cost function in our test problem has a dimension of 1083×1083.

The southern and north boundaries are rigid walls where the normal velocity components vanish, and it is assumed that the flow is periodic in the west–east direction with a wavelength equal to the length of the channel.

Let us define

$$\boldsymbol{X} = (u, v, \phi)^T \tag{3.8}$$

$$F = -\begin{pmatrix} u\dfrac{\partial u}{\partial x} + v\dfrac{\partial u}{\partial y} - fv + \dfrac{\partial \phi}{\partial x} \\ u\dfrac{\partial v}{\partial x} + v\dfrac{\partial v}{\partial y} + fu + \dfrac{\partial \phi}{\partial y} \\ \dfrac{\partial (u\phi)}{\partial x} + \dfrac{\partial (v\phi)}{\partial y} \end{pmatrix}. \tag{3.9}$$

Then Eqs. (3.1)–(3.3) assume the form of Eq. (2.1). It is easy to verify that

$$\frac{\partial F}{\partial \boldsymbol{X}} = -\begin{pmatrix} \dfrac{\partial(u(\cdot))}{\partial x} + v\dfrac{\partial(\cdot)}{\partial y} & (\cdot)\dfrac{\partial u}{\partial y} - f(\cdot) & \dfrac{\partial(\cdot)}{\partial x} \\ (\cdot)\dfrac{\partial v}{\partial x} + f(\cdot) & u\dfrac{\partial(\cdot)}{\partial x} + \dfrac{\partial(v(\cdot))}{\partial y} & \dfrac{\partial(\cdot)}{\partial y} \\ \dfrac{\partial(\phi(\cdot))}{\partial x} & \dfrac{\partial(\phi(\cdot))}{\partial y} & \dfrac{\partial(u(\cdot))}{\partial x} + \dfrac{\partial(v(\cdot))}{\partial y} \end{pmatrix} \tag{3.10}$$

The adjoint of an operator L, L^*, is defined by the relationship

$$\langle \boldsymbol{LX}, \boldsymbol{Y} \rangle = \langle \boldsymbol{X}, \boldsymbol{L^*Y} \rangle \tag{3.11}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product

$$\langle \cdot, \cdot \rangle = \iint_D \cdots dD \tag{3.12}$$

where D is the spatial domain. Using the definition (3.12), the adjoint of (3.11) can be derived as

$$\left[\frac{\partial F}{\partial \boldsymbol{X}}\right]^* = -\begin{pmatrix} -u\dfrac{\partial(\cdot)}{\partial x} - \dfrac{\partial(v(\cdot))}{\partial y} & (\cdot)\dfrac{\partial v}{\partial x} + f(\cdot) & -\phi\dfrac{\partial(\cdot)}{\partial x} \\ (\cdot)\dfrac{\partial u}{\partial y} - f(\cdot) & -v\dfrac{\partial(\cdot)}{\partial y} - \dfrac{\partial(u(\cdot))}{\partial x} & -\phi\dfrac{\partial(\cdot)}{\partial y} \\ -\dfrac{\partial(\cdot)}{\partial x} & -\dfrac{\partial(\cdot)}{\partial y} & -u\dfrac{\partial(\cdot)}{\partial x} - v\dfrac{\partial(\cdot)}{\partial y} \end{pmatrix} \tag{3.13}$$

Therefore the first order adjoint model with the forcing terms may be written as

$$-\frac{\partial u^*}{\partial t} = -\left(-u\frac{\partial u^*}{\partial x} - \frac{\partial(vu^*)}{\partial y} + v^*\frac{\partial v}{\partial x} + fv^* - \phi\frac{\partial \phi^*}{\partial x}\right) + W_u(u - u^o) \tag{3.14}$$

$$-\frac{\partial v^*}{\partial t} = -\left(u^*\frac{\partial u}{\partial y} - fu^* - v\frac{\partial v^*}{\partial y} - \frac{\partial(uv^*)}{\partial x} - \phi\frac{\partial \phi^*}{\partial y}\right) + W_v(v - v^o) \tag{3.15}$$

$$-\frac{\partial \phi^*}{\partial t} = -\left(-\frac{\partial u^*}{\partial x} - \frac{\partial v^*}{\partial y} - u\frac{\partial \phi^*}{\partial x} - v\frac{\partial \phi^*}{\partial y}\right) + W_\phi(\phi - \phi^o) \tag{3.16}$$

with final conditions

$$u(T) = 0, \quad v(T) = 0, \quad \phi(T) = 0 \tag{3.17}$$

where $P = (u^*, v^*, \phi^*)^t$ is the first order adjoint variable. W_u, W_v, W_ϕ are weighting factors which are taken to be the inverse of estimates of the statistical root-mean-square observational errors on geopotential and wind components, respectively. In our test problem, values of $W_\phi = 10^{-4}\,\mathrm{m}^{-4}\,\mathrm{s}^4$ and $W_u = W_v = 10^{-2}\,\mathrm{m}^{-2}\,\mathrm{s}^2$ are used.

Now let us consider a perturbation, U', on the initial condition for $\boldsymbol{X}, \boldsymbol{X}(t_0)$. The resulting corresponding perturbations for variables $\boldsymbol{X}$ and P, $\hat{\boldsymbol{X}} = (\hat{u}, \hat{v}, \hat{\phi})^t$ and $\bar{P} = (\bar{u}, \bar{v}, \bar{\phi})^t$, are obtained from Eqs. (3.1)–(3.3) and (3.14)–(3.17) as

$$\frac{\partial \hat{u}}{\partial t} = -\left(\frac{\partial(u\hat{u})}{\partial x} + v\frac{\partial \hat{u}}{\partial y} + \hat{v}\frac{\partial u}{\partial y} - f\hat{v} + \frac{\partial \hat{\phi}}{\partial x}\right) \tag{3.18}$$

$$\frac{\partial \hat{v}}{\partial t} = -\left(\hat{u}\frac{\partial v}{\partial x} + f\hat{u} + u\frac{\partial \hat{v}}{\partial x} + \frac{\partial(v\hat{v})}{\partial y} + \frac{\partial \hat{\phi}}{\partial y}\right) \tag{3.19}$$

$$\frac{\partial \hat{\phi}}{\partial t} = -\left(\frac{\partial(\phi\hat{u})}{\partial x} + \frac{\partial(\phi\hat{v})}{\partial y} + \frac{\partial(u\hat{\phi})}{\partial x} + \frac{\partial(v\hat{\phi})}{\partial y}\right) \tag{3.20}$$

with zero initial conditions, and

$$-\frac{\partial \bar{u}}{\partial t} = -\left(-u\frac{\partial \bar{u}}{\partial x} - \frac{\partial(v\bar{u})}{\partial y} + \bar{v}\frac{\partial v}{\partial x} + \phi\frac{\partial \bar{\phi}}{\partial x}\right.$$

$$-\hat{u}\frac{\partial u^*}{\partial x}-\frac{\partial(\hat{v}u^*)}{\partial y}+v^*\frac{\partial\hat{v}}{\partial x}-\hat{\phi}\frac{\partial\phi^*}{\partial x}\Big)+W_u\hat{u} \tag{3.21}$$

$$-\frac{\partial\bar{v}}{\partial t}=-\Big(\bar{u}\frac{\partial u}{\partial y}-v\frac{\partial\bar{v}}{\partial y}-\frac{\partial(u\bar{v})}{\partial x}-\phi\frac{\partial\bar{\phi}}{\partial y}$$
$$+u^*\frac{\partial\hat{u}}{\partial y}-\hat{v}\frac{\partial v^*}{\partial y}-\frac{\partial(\hat{u}v^*)}{\partial x}-\hat{\phi}\frac{\partial\phi^*}{\partial y}\Big)+W_v\hat{v} \tag{3.22}$$

$$-\frac{\partial\bar{\phi}}{\partial t}=-\Big(-\frac{\partial\bar{u}}{\partial x}-\frac{\partial\bar{v}}{\partial y}-u\frac{\partial\bar{\phi}}{\partial x}-v\frac{\partial\bar{\phi}}{\partial y}$$
$$-\hat{u}\frac{\partial\phi^*}{\partial x}-\hat{v}\frac{\partial\phi^*}{\partial y}\Big)+W_\phi\hat{\phi} \tag{3.23}$$

with final condition

$$\bar{u}(T)=0,\quad \bar{v}(T)=0,\quad \bar{\phi}(T)=0. \tag{3.24}$$

Therefore

$$\hat{P}(t_0)=(\bar{u}(t_0),\bar{v}(t_0),\bar{\phi}(t_0))^t=HU' \tag{3.25}$$

where H is the Hessian of the cost function with respect to the initial conditions. Equation (3.25) gives the Hessian/vector product. If we choose U' to be the unit vector e_j where the j-th component is unity and all its other components are zeros, then the corresponding column H_j of the Hessian H will be obtained after one integration of the SOA backwards in time.

4. Second Order Adjoint Information

4.1 Calculation of the Hessian/Vector Product

There are two practical ways to calculate the Hessian/vector product at a point $\boldsymbol{X}$ associated with VDA. One way consists in using a finite-difference method while the other way is by using the SOA method. The finite-difference approach assumes the following form

$$f(\alpha)=VJ(\boldsymbol{X}+\alpha\boldsymbol{Y})-VJ(\boldsymbol{X})=\alpha H\boldsymbol{Y}+O(\alpha^2) \tag{4.1}$$

where $\boldsymbol{Y}$ is a random perturbation vector and H is the Hessian of the cost function. A second way to obtain Hessian/vector product is to integrate the SOA equations model backwards in time. According to Eq. (2.14), we also have

$$f(\alpha)=\alpha H\boldsymbol{Y}. \tag{4.2}$$

The computational cost required to obtain the Hessian/vector product is approximately the same for both methods. The SOA approach requires us to integrate the original nonlinear model and its tangent linear model forward in time once and integrate the FOA model and the SOA model backwards in time once. The finite difference approach requires the integration of the original nonlinear model forward in time twice and the FOA model backwards in time twice. The computational costs for integrating the tangent linear model forward in time, the FOA model backwards in time or the SOA model backwards in time once are comparable. However, the SOA model method gives an accurate value of the Hessian/vector product while the finite-difference method yields only an approximated value, which can be a very poor estimate when the value α is not properly chosen. Figures 2–4 present a comparison between the first 50 components of Hessian/vector products at the optimal point obtained by using both the SOA and finite-difference approaches for various scalars α varying from 10, 3 to 0.01. It is clearly seen that the Hessian/vector product obtained by using finite-difference approach converges to that obtained by SOA as the scalar α decreases. With the SOA approach an accurate result can be obtained with a relatively large perturbation ($\alpha=10$), while the finite-difference approach is very sensitive to the magnitude of perturbations. When the perturbations are large, say for $\alpha=10$, the finite-differencing yields no meaningful results (Fig. 2). When the perturbations are small, the finite-difference

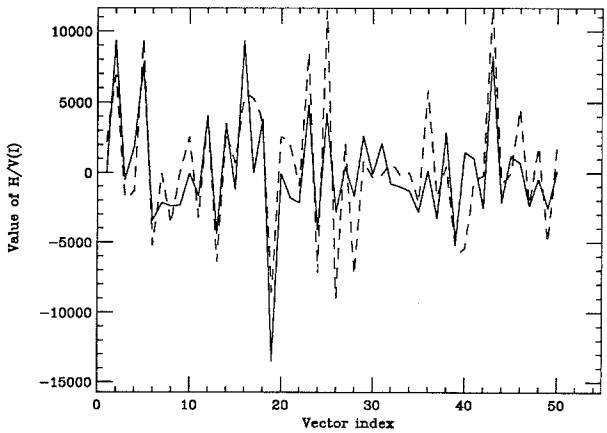


Fig. 2. The first 50 components of the Hessian/vector products at the optimal solution obtained by the finite-difference approach (dash line), and SOA method (solid line) when $\alpha=10$

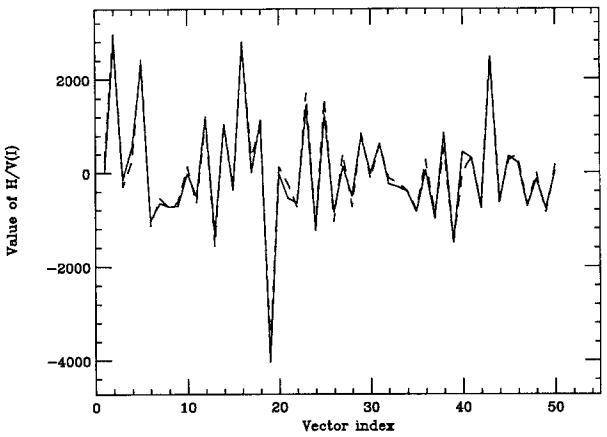


Fig. 3. Same as Fig. 2 except $\alpha = 3$

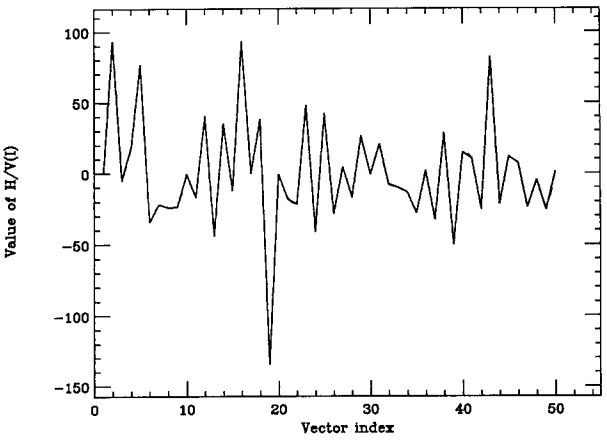


Fig. 4. Same as Fig. 2 except $\alpha = 0.01$

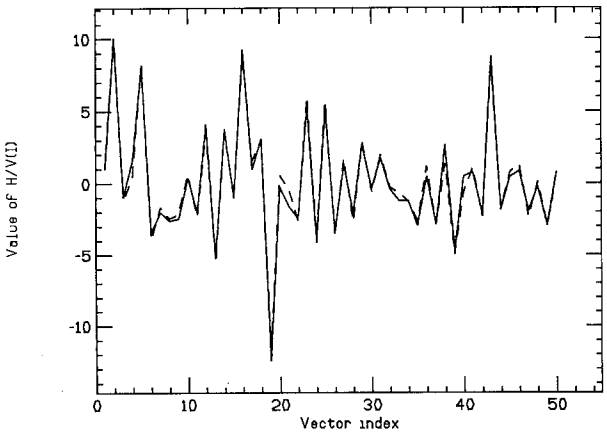


Fig. 5. Same as Fig. 4 except at the initial guess point

approach might involve a subtraction of nearly equal numbers which results in the cancellation of significant digits and the results thus obtained are an inaccurate estimate of the Hessian/vector product. This is the case when the Hessian/vector product is estimated at the initial guess point with $\alpha = 0.01$ (Fig. 5). Therefore it is much more advantageous to use the SOA approach than to use the finite-difference approach.

The calculation of a Hessian/vector product is required in many occurrences. For instance, Nash's (1984) Truncated Newton method requires the values of Hessian/vector products. It may also be used to carry out eigenvalue calculations and sensitivity analysis.

4.2 The Uniqueness of the Solution

An important issue in VDA is to determine whether the solution is unique. If more than one local minimum exists, then the solution of the minimization process may possibly change depending on different initial guesses.

There are two different but complementary ways to characterize the solution to unconstrained optimization problems. In the local approach, one examines the relation of a given point to that of its neighbors. The conclusion is that at an unconstrained relative minimum point of a smooth cost function, the gradient of the cost function vanishes and the Hessian is positive semidefinite; and conversely, if at a point the gradient vanishes and the Hessian is positive definite, that point is a relative minimum point. This characterization has a natural extension to the global approach where convexity ensures that if the gradient vanishes at a point, that point is a global minimum point.

The Hessian (the matrix of second order derivatives of the cost function with respect to the control variables) is the generalization to E^n of the concept of the curvature of the function, and correspondingly, positive definiteness of the Hessian is the generalization of positive curvature. We sometimes refer to a function as being locally strictly convex if its Hessian is positive definite in the region. In these terms we see that the second order sufficiency result requires that the function be locally strictly convex at the point X^*.

A simple experiment was conducted to find out about the uniqueness of the cost function with respect to the initial conditions using the shallow-water equation model. The experiment is devised as follows: the model-generated values starting

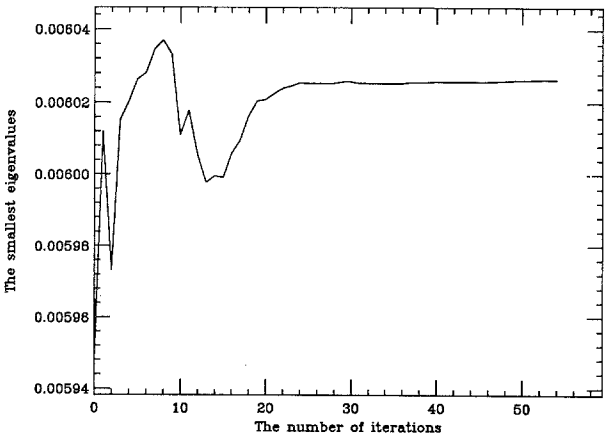


Fig. 6. Variation of the smallest eigenvalue of the Hessian of the cost function with the number of iterations

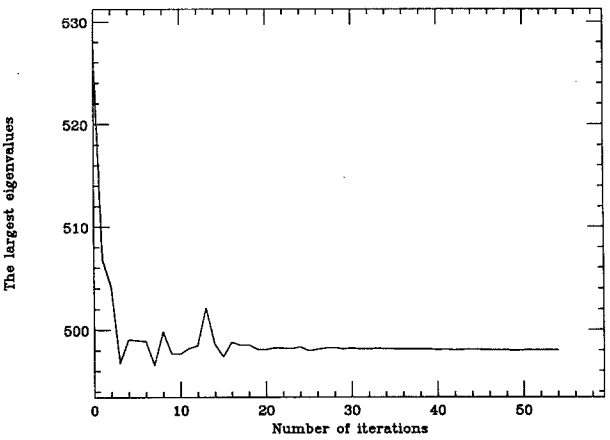


Fig. 7. Variation of the largest eigenvalue of the Hessian of the cost function with the number of iterations

from the initial condition of Grammeltvedt (Eq. (3.4)) are used as observations, the initial guess is a randomly perturbed Grammeltvedt initial condition, and the length of the assimilation is 10 hours. We know exactly what the solution is, and the value of the cost function at the minimum must be zero. All the random perturbations used in this paper are from a uniform distribution. The limited memory quasi-Newton large-scale unconstrained minimization method of Liu and Nocedal (1989) is used for all experiments in this paper.

The symmetric versions of the power and shifted power methods are used to obtain the largest and smallest eigenvalues of the Hessian at each iteration. The results are shown in Figs. 6 and 7. The smallest eigenvalues at each iteration of the minimization process are small positive numbers. The positiveness of the smallest eigenvalues implies the positive definiteness of the Hessian, which in turn proves the uniqueness of the optimal solution.

4.3 Convergence Analysis

The largest and smallest eigenvalues and the condition numbers are considered here. The purpose of this study is to provide an in-depth diagnosis of the convergence of the VDA applied to a meteorological problem. The various scale changes of different field reconstructions with the number of minimization iterations of VDA has attracted the attention of several researchers (Navon et al., 1992). In this research work we will attempt to provide an explanation of this phenomenon based on the evolution of the condition number of the Hessian of the cost function with respect to control variables (Thacker, 1989). It has been observed that in VDA, large scale changes occur in the first few iterations and small scale changes occur during the latter iterations in the process of the minimization of the cost function.

The same experiment as described in section 4.2 was conducted again this time to follow the quality of the reconstructed initial conditions at different stages of the minimization process. Figures 8–10 show the perturbed geopotential

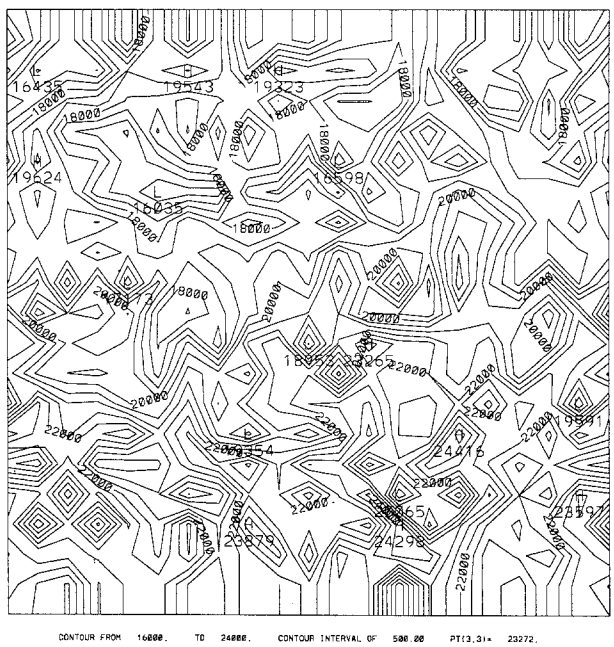


Fig. 8. Distribution of the randomly perturbed geopotential field

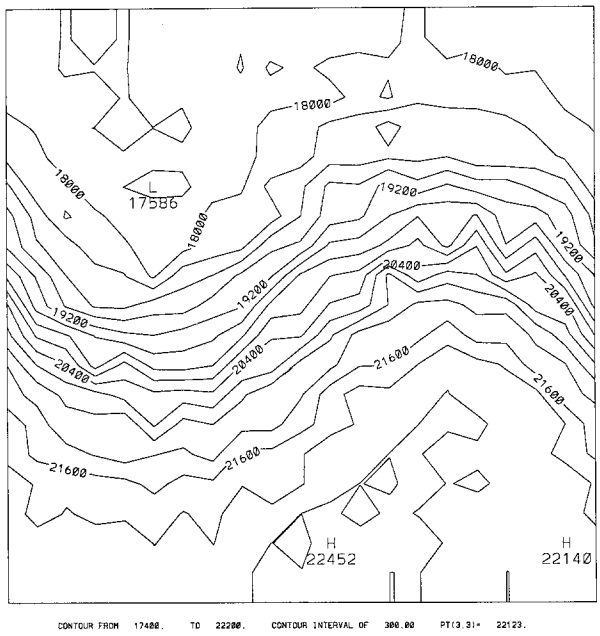


Fig. 9. Reconstructed geopotential field after 6 iterations of minimization

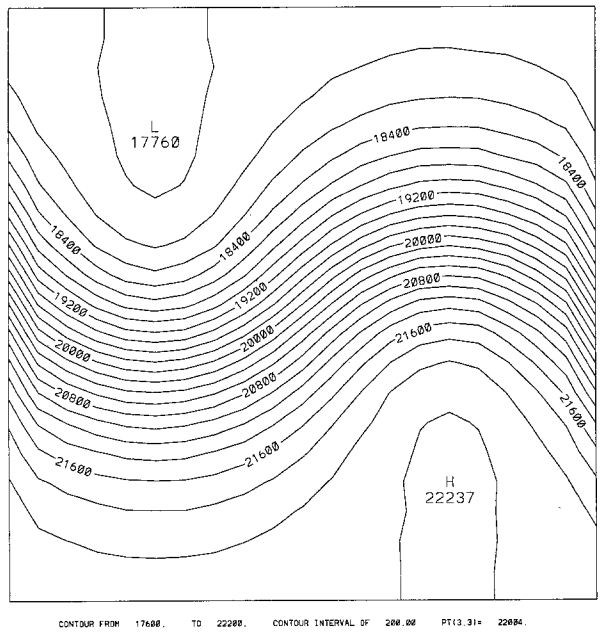


Fig. 10. Reconstructed geopotential field after 25 iterations of minimization

field and the reconstructed geopotential fields after 6 and 25 iterations, respectively. It can be clearly seen that most of the large scale reconstructions occur within the first 25 iterations of the minimization process. The geopotential field reconstructed after 25 iterations is very similar to the one reconstructed after 54 iterations at which stage the prescribed convergence criteria

$$\| \nabla J(X_k) \| \leqslant 10^{-14} \times \max \{1, \| X_k \| \}$$

is satisfied. This clearly indicates that the VDA achieves most of the large scale reconstructions during the first 25 iterations and that in the latter part of the minimization process only small scale features are being assimilated. In this case by stopping the minimization process prior to the cost function satisfying the preset convergence criteria, the expensive computational cost of the VDA process could be cut by more than a half, while satisfactory results may still be obtained.

This in turn is related to the evolution of the largest and smallest eigenvalues and thus to the change in the condition number of the Hessian with iterations (Figs. 6, 7 and 11). From these figures we observe:

(a) The smallest eigenvalues are positive at each iteration and remain approximately the same except for rather small changes during the first few iterations (Fig. 6).
(b) The largest eigenvalues decrease quickly during the first few iterations of the minimization process, then change only slightly for the next 15 iterations and remain approximately the same during the latter minimization stages until the convergence criteria is attained (Fig. 7).
(c) The condition numbers of the Hessian/vector product at different steps of the minimization vary in a way similar to that of the evolution of the largest eigenvalues during the minimization process and their magnitude is about 83,000 which is very large (Fig. 11).

We conclude that most changes in the condition numbers occur during the early stage of the VDA minimization process. This explains why large scale reconstructions occur during the first 30 iterations of the minimization process.

The large condition numbers in the initial stage of the minimization imply that the contour lines of the cost function $J(=\text{constant})$ are strongly elongated in the parameter space (Lions, 1971; Fletcher, 1987; Gill et al., 1981; Luenberger et al., 1984), which explains the slow convergence rate of the VDA process. The above experiment was carried out without adding either a penalty or a smoothing term. The addition of such a penalty

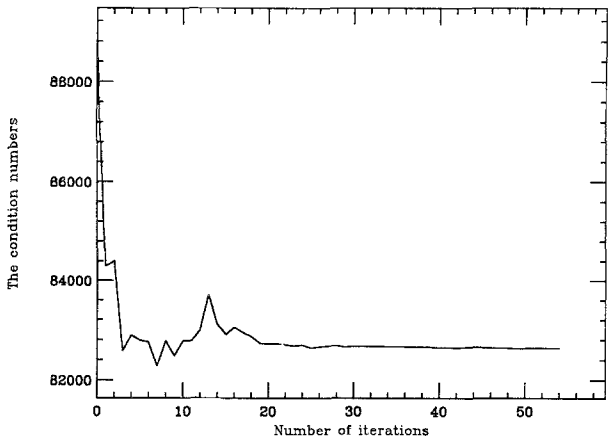


Fig. 11. Variation of the condition numbers of the Hessian of the cost function with respect to the number of iterations

term, which is positive definite and quadratic with respect to the initial conditions, will definitely increase the convexity of the cost function. Thus the addition of an adequate quadratic penalty term adding additional information to the cost function changes the condition number of the Hessian and speeds up the convergence of the VDA process.

5. Sensitivity Analysis for Observations

5.1 Sensitivity Theory for Observation

The cost function is also a function of the observations. Different observations will result in different solutions. Due to the errors inherent in the heterogeneous observations, it is important to obtain the sensitivities of the cost function to the changes in the observations which quantify the extent to which the perturbations in the observations correspond to the perturbation in the solution. If the sensitivities are large, then the model will possess a large uncertainty with respect to changes in the observations.

Conventional evaluation of the sensitivities with respect to model parameters is carried out by changing the values of model parameters and recalculating each model solution for every parameter. Such a calculation is prohibitive for models with a large number of parameters since it requires an exceedingly large amount of computing time. The adjoint sensitivity method (Cacuci, 1981; Hall and Cacuci, 1982, 1983; Sykes et al., 1985) proved to be an efficient method for carrying out sensitivity analysis. The objective of the sensitivity analysis considered here is to estimate changes in the cost function, J, arising from changes in observations which are distributed in space and time. This will illustrate the relative importance of observations at different time and space locations.

Due to the equivalent position of the state vector and the observation vector in Eq. (A.1), the cost function can be viewed as depending on both of them, namely

$$J = J(X - X^o). \tag{5.1}$$

As such, the following identities can be proved using the chain rule:

$$\frac{\partial J}{\partial X} = -\frac{\partial J}{\partial X^o} \tag{5.2}$$

$$\frac{\partial^2 J}{\partial X^2} = \frac{\partial^2 J}{\partial X^{o^2}}. \tag{5.3}$$

These two equations are used in the following sensitivity analysis.

Let us denote a change in the observations by δX^o. If this change is small, then we may expand the cost function J around X^o as

$$\begin{aligned} &J(X^o(t_n) + \delta X^o(t_n)) \\ &\quad = J(X^o(t_n)) + \frac{\partial J(X^o(t_n))}{\partial X^o(t_n)} \delta X^o(t_n) \\ &\qquad + \frac{1}{2} \delta X^o(t_n)^t \frac{\partial^2 J(X^o(t_n))}{\partial X^o(t_n)^2} \delta X^o(t_n) \\ &\qquad + O(\| \delta X^o(t_n) \|^3). \end{aligned} \tag{5.4}$$

According to the identities given by Eqs. (5.2) and (5.3), Eq. (5.4) can be written as

$$\begin{aligned} &J(X^o(t_n) + \delta X^o(t_n)) \\ &\quad = J(X^o(t_n)) - \frac{\partial J(X^o(t_n))}{\partial X(t_n)} \delta X^o \\ &\qquad + \frac{1}{2} \delta X^o(t_n)^t \frac{\partial^2 J(X^o(t_n))}{\partial X(t_n)^2} \delta X^o \\ &\qquad + O(\| \delta X^o(t_n) \|^3) \end{aligned} \tag{5.5}$$

where t_n denotes the time, $t_n = t_0 + n\Delta t$ and Δt is given by Eq. (3.7). Since the first order term in Eq. (5.5) dominates, we obtain from Eq. (5.5)

$$\begin{aligned} J' &= J(X^o(t_n) + \delta X^o(t_n)) - J(X^o(t_n)) \\ &= -\frac{\partial J(X^o(t_n))}{\partial X(t_n)} \delta X^o + O(\| \delta X^o(t_n) \|^2). \end{aligned} \tag{5.6}$$

This equation describes changes in the cost function resulting from changes in the observation at time t_n.

If the gradient of the cost function with respect to the state vector $X(t_n)$ is zero, then we obtain

$$J' = J(X^o(t_n) + \delta X^o(t_n)) - J(X^o(t_n))$$
$$= \frac{1}{2}\delta X^o(t_n)^t \frac{\partial^2 J(X^o(t_n))}{\partial X(t_n)^2}\delta X^o + O(\|\delta X^o(t_n)\|^3) \tag{5.7}$$

where the second derivative of J with respect to the observations is the Hessian of J with respect to the state variable at time t_n. Equation (5.7) describes the changes in the cost function resulting from a change in the observation at time t_n.

Let us now calculate the first derivative of the cost function J with respect to $X(t_n)$. The variation of the cost function J in Eq. (A.6) can be written as

$$\delta J = \sum_{i=0}^{M} \langle W(X_i - X_i^o), \hat{X}_i \rangle$$
$$= \sum_{i=0}^{n-1} \langle W(X_i - X_i^o), \hat{X}_i \rangle + \langle W(X_n - X_n^o), \hat{X}_n \rangle$$
$$+ \sum_{i=n+1}^{M} \langle W(X_i - X_i^o), \hat{X}_i \rangle$$
$$= \delta J_1 + \langle W(X_n - X_n^o), \hat{X}_n \rangle + \delta J_2 \tag{5.8}$$

where $\hat{X}_i$ is the resulting perturbation on X_i and is defined by the tangent linear Eqs. (2.7), (2.8). It can be shown from Appendix A that

$$\hat{X}_i = \prod_{j=1}^{i-1} \left[I + \Delta t \left(\frac{\partial F}{\partial X} \right)_j \right] \hat{X}_n \tag{5.9}$$

for $i > n$, and

$$\hat{X}_n = \prod_{j=i}^{n-1} \left[I + \Delta t \left(\frac{\partial F}{\partial X} \right)_j \right] \hat{X}_i \tag{5.10}$$

for $n > i$. Using the definition of the adjoint operators, Eq. (5.10) yields

$$\hat{X}_i = \left\{ \prod_{j=i}^{n-1} \left[I + \Delta t \left(\frac{\partial F}{\partial X} \right)_j \right] \right\}^* \hat{X}_n \tag{5.11}$$

for n being larger than i. Now we can write J_1, J_2 corresponding to Eqs. (5.9) and (5.11) respectively as

$$\delta J_1 = \sum_{i=0}^{n-1} \left\langle W(X_i - X_i^o), \left\{ \prod_{j=i}^{n-1} \left[I + \Delta t \left(\frac{\partial F}{\partial X} \right)_j \right] \right\}^* \hat{X}_n \right\rangle$$

$$\delta J_2 = \sum_{i=n+1}^{M} \left\langle W(X_i - X_i^o), \prod_{j=n}^{i-1} \left[I + \Delta t \left(\frac{\partial F}{\partial X} \right)_j \right] \hat{X}_n \right\rangle$$

Substituting J_1, J_2 into Eq. (5.8) and using basic concepts of adjoint operators, we obtain the following expression

$$\delta J = \sum_{i=0}^{n-1} \left\langle \prod_{j=i}^{n-1} \left[I + \Delta t \left(\frac{\partial F}{\partial X} \right)_j \right] W(X_i - X_i^o), \hat{X}_n \right\rangle$$
$$+ \langle W(X_n - X_n^o), \hat{X}_n \rangle$$
$$+ \sum_{i=n+1}^{M} \left\langle \left\{ \prod_{j=n}^{i-1} \left[I + \Delta t \left(\frac{\partial F}{\partial X} \right)_j \right] \right\}^*$$
$$\times W(X_i - X_i^o), \hat{X}_n \right\rangle. \tag{5.12}$$

We know however that

$$\delta J = \langle \nabla_{X(t_n)} J, \hat{X}_n \rangle. \tag{5.13}$$

Equating Eqs. (5.12) and (5.13), we obtain the gradient of the cost function with respect to $X(t_n)$ as

$$\nabla_{X(t_n)} J = \sum_{i=0}^{n-1} \prod_{j=i}^{n-1} \left[I + \Delta t \left(\frac{\partial F}{\partial X} \right)_j \right] W(X_i - X_i^o)$$
$$+ W(X_n - X_n^o)$$
$$+ \sum_{i=n+1}^{M} \left\{ \prod_{j=n}^{i-1} \left[I + \Delta t \left(\frac{\partial F}{\partial X} \right)_j \right] \right\}^*$$
$$\times W(X_i - X_i^o). \tag{5.14}$$

In summary, the perturbation in the cost function resulting from a perturbation in the observation at the time t_n may be obtained by performing the following operations

(a) Generate a perturbation on the observation at time t_n;
(b) Calculate the gradient of the cost function with respect to state variable $X(t_n)$, which is the sum of the results of integrating Eq. (5.15) and Eq. (5.16) plus the middle term in Eq. (5.14),
  (1) Starting from $P_M = W(X_M - X_M^o)$, integrate the "forced" adjoint equation

$$P_i = \left[I + \Delta t \left(\frac{\partial F}{\partial X} \right)_i^* \right] P_{i+1} + W(X_i - X_i^o) \tag{5.15}$$

backwards in time from t_M to t_n. The final result P_n is the sum of the last two terms in Eq. (5.14),

(2) Starting from $\hat{X}_0 = W(X_0 - X_0^o)$, integrate the "forced" equation

$$\hat{X}_i = \left[I + \Delta t \left(\frac{\partial F}{\partial X} \right)_{i-1} \right] \hat{X}_{i-1} + W(X_i - X_i^o) \tag{5.16}$$

forward in time from t_0 to t_n. The final result $\hat{X}_n$ is the sum of the first two terms in Eq. (5.14).

(c) Use Eq. (5.6) to obtain the corresponding perturbation in the cost function resulting from the perturbation in the observations at time t_n.

It is worthwhile noting that we do not need to integrate the FOA equations repeatedly to obtain the gradient of the cost function with respect to X_{t_n}. We only need to integrate the FOA equations backwards in time starting from t_M to t_n and store the FOA variable at each iteration in memory, then integrate the forced linear equation forward in time from t_0 to t_n and store the results at each iteration in memory. The sum is the gradient of the cost function with respect to the state variables at time t_n plus $W(X_n - X_n^o)$ for $n = 0, 1, \ldots, M$. Once these gradients are calculated, they need not be recalculated. They can be used repeatedly to calculate the perturbations in the cost function for different perturbations in the observations.

5.2 Numerical Results from Model Generated Data

A sensitivity study was conducted by using the same model as that described in section 3. First we choose a point (x_{15}, y_{10}) in the assimilation window where $x_{15} = x_0 + 15\Delta x$ and $y_{10} = y_0 + 10\Delta y$. Suppose a 1% perturbation in the observations occurs only at this point for the two components of the wind velocity field and the geopotential field. The variation of the resulting perturbations in the cost function as a function of the number of time steps is displayed in Fig. 12 (solid line). The results indicate that the changes in the cost function with respect to the changes in the observations at a fixed point are different for different times in the assimilation window. If perturbations are imposed firstly only on the u-wind component, then only on the v-wind component and then only on the geopotential field ϕ, the corresponding perturbations in the cost function

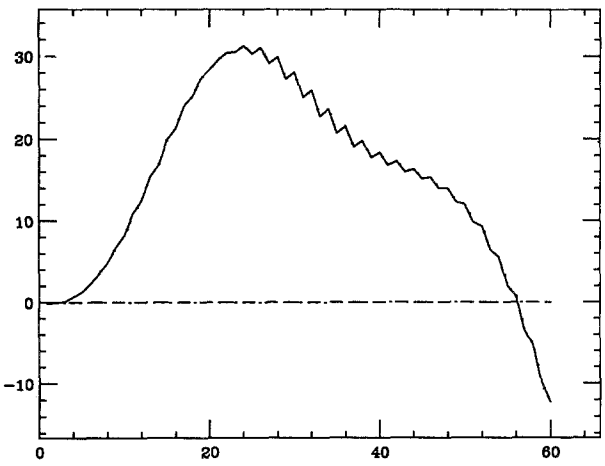


Fig. 12. Time variation of the sensitivities of cost function J to 1% observational error in the v-wind component (dashed line), the u-wind component (dotted line which coincides with the dashed line), the geopotential field ϕ (dashed-dotted line which coincides with the solid line) and in all the three fields (solid line) at point (x_{15}, y_{10})

exhibit different variations with time as shown in Fig. 12 by the dotted line, dashed line and dash-dot line, respectively. This figure indicates also that perturbations in the observed geopotential field have more impact on the cost function than those in the observed velocity field. The changes in the cost function arising from changes in the u-wind component and v-wind component observations are close to zero at all times. Similar experiments conducted at different grid points yield similar results.

To study the importance of observations at different space locations, three different points are chosen. They are located at (x_5, y_{15}), (x_{10}, y_{10}), (x_{15}, y_5), respectively, representing low, middle and high points in the isoline values of the geopotential field. From Fig. 13 we observe that the changes in the observations occurring at the end of the assimilation period result in larger changes in the magnitudes of the cost function than corresponding changes in the observations occurring at the beginning of the assimilation window. This means that recent events have more impact on the cost function that older events.

Finally, we study the impact of the perturbations on all the observational data. The results are displayed in Fig. 14. This figure clearly indicates that perturbations of the observations at the end of the assimilation window have a larger impact on the sensitivity of the cost function with respect to the observations.

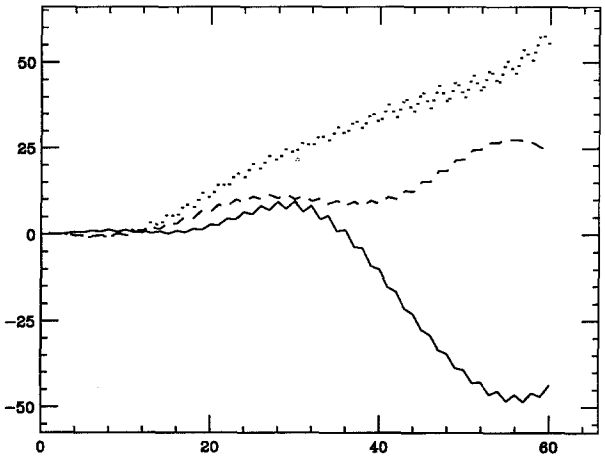


Fig. 13. Time variation of the sensitivities of cost function J to 1% observational error at points (x_{10}, y_{10}) (solid line) (x_5, y_{15}) (dotted line), and (x_{15}, y_5) (dash line) in the wind and the geopotential fields

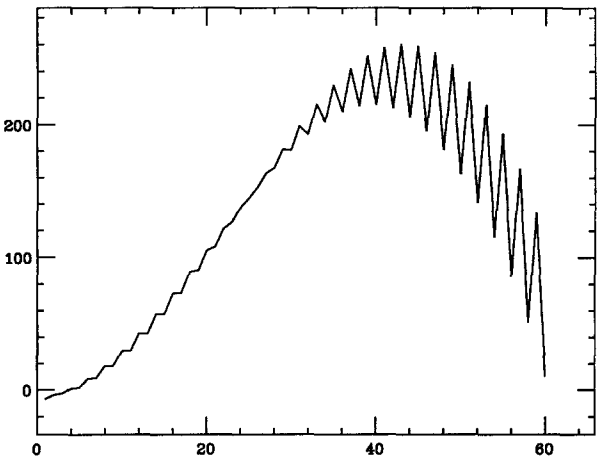


Fig. 14. Time variation of the sensitivities of cost function J to 1% observational error on all grid points in the wind and the geopotential fields

5.3 *Numerical Results Using Real Analysis of FGGE Data*

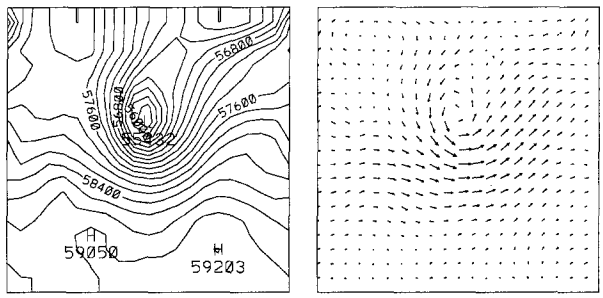
To examine the sensitivity of the cost function with respect to real analyses, we employed a set of FGGE data of height and horizontal wind fields at 500 mb level at 0 and 18UTC, May 26, 1979. The data are equally spaced with $\Delta\lambda = \Delta\phi = 1.875°$. Using the formula

$$\phi(J) = \frac{y}{\alpha} + \phi_0 = \frac{-2200 + (J-1)*220}{\alpha} + \phi_0$$

$$\lambda(I) = \frac{x}{\alpha \cos\phi(J)} + \lambda_0 \qquad (5.8)$$

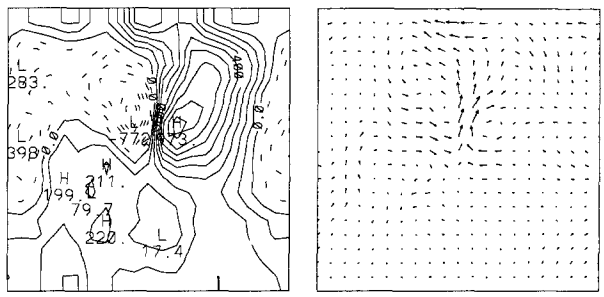
we obtain a correspondence between points on the sphere and grid points located on a limited area on a β-plane approximation at $(32°, 130°)$, which approximately represents the center of the zonal jet. Using a cubic interpolation we obtained the height and horizontal wind data on the grid points. Then we carried out another cubic interpolation near the left boundary in order to impose a periodic boundary condition in the x-direction. Near the top and bottom boundaries we used a linear interpolation to impose solid boundary conditions. The fields thus obtained are shown in Fig. 15.

The geopotential and wind fields at time 0 UTC were used to produce the model-generated observations. The minimization started from geopotential and wind fields distribution at time 18 UTC. The difference between these two fields is shown in Fig. 16. Having the model-generated observa-



a b

Fig. 15. Distribution of (a) the geopotential and (b) the wind fields for the FGGE data at 0 UTC 05/26, 1979 on the 500 mb. The contour intervals are 200 m^2/s^2 and the magnitude of maximum velocity vector is 0.311E + 02 m/s



a b

Fig. 16. Distribution of the difference fields of the geopotential (a) and the wind (b) fields at 18 UTC and 0 UTC on 500 mb between 18 UTC and 0 UTC times. The contour intervals are 100 m^2/s^2 and the magnitude of maximum velocity vector is 0.210E + 02 m/s

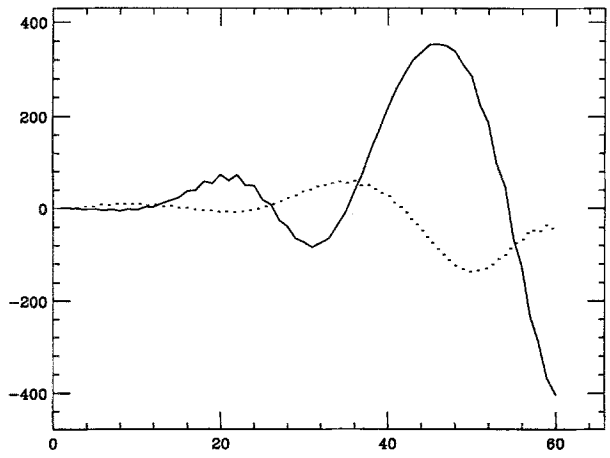


Fig. 17. Time variation of the sensitivities of the cost function J to 1% observational error in the wind and the geopotential fields at grid points (x_{10}, y_{10}) (solid line) and (x_7, y_5) (dotted line)

tions, the minimization should be able to reduce the value of the cost function as well as the norm of its gradient, and the reconstructed differences should be zero. This turns out to be the case.

From Fig. 15, we note that meteorologically-intensive events occur at the center area of the limited-area domain while fewer events occur at the corners of the limited-area domain. We chose two points (x_{10}, y_{10}) and (x_7, y_5), which are located in the center and in the bottom left corner of the limited area domain, respectively. We then introduced a 1% perturbation in the geopotential and wind fields at these two points. The variations of sensitivities of the cost function with the number of time steps in the assimilation window are displayed in Fig. 17, the solid line corresponding to sensitivity at the point (x_{10}, y_{10}) and the dotted line to sensitivity at the point (x_7, y_5). Clearly the sensitivity of the cost function with respect to observations at point (x_{10}, y_{10}) is larger than that at the point (x_7, y_5). This confirms that the cost function is more sensitive to observations at points where intensive events occur. It also means that the accuracy of observations at locations where intensive events occur has more impact on the quality of the VDA retrieval.

6. Summary and Conclusions

In this paper, a SOA model was developed, providing second order information. The coding work involving in obtaining the SOA model was rather modest once the FOA model has been developed.

One integration of the SOA model yields an accurate value of a column of the Hessian of the cost function provided the perturbation vector is a unit vector with one component being unity and the remainder being zeros. Numerical results show that the use of the SOA approach to obtain the Hessian/vector product is much more advantageous than the corresponding finite-difference approach, since the latter yields only an approximated value of the Hessian/vector product which may be a very poor estimate. The numerical cost of using the SOA approach is roughly the same as that of using the finite-difference approach. This application of the SOA model is crucial in the implementation of the large-scale truncated Newton method, which was proved to be a very efficient for large-scale unconstrained minimization (Zou et al., 1991).

Another application of the SOA model is in the calculation of eigenvalues and eigenvectors of the Hessian. There are several iterative methods such as the power method, Rayleigh quotient or the Lanczos method (Strang, 1986), which require only the information of the Hessian-vector product to calculate several eigenvalues and eigenvectors. Such a calculation using the power method is presented in this paper and reveals that most changes of the largest eigenvalue occur during the first few iterations of the minimization procedure, which might explain why most of large-scale features are reconstructed earlier than the small scale features in the VDA retrieval solution during minimization (Navon et al., 1992) and the positivity of the smallest eigenvalues of the Hessians of the cost function during the minimization process indicates the uniqueness of the optimal solution.

We also examined the sensitivity of the cost function to observational errors using a two dimensional limited-area shallow water equation model. We found that the sensitivity depends on the time when the errors occur, the specific field containing the errors, and the spatial location where the errors occur. The cost function is more sensitive to the observational errors occurring at the end of the assimilation window, to errors in the geopotential field, and to errors at these grid point locations where intensive events occur.

Sensitivity analysis using balanced perturbations will be reported in a future paper where we will pay special attention to the spatial scale of the perturbations. Further research on the issue of calculating the inverse Hessian multiplied by a vector is currently under consideration, the latter being of crucial importance for developing a new efficient large-scale minimization algorithm.

Appendix A

Brief Description of the FOA

The theory and application of the FOA model is discussed by several authors, e.g. Talagrand and Courtier (1987) and Navon et al. (1990). In order to provide a comprehensive description of the SOA model, we briefly summarize the theory of the FOA model.

The distance function, which measures the distance between the model solution and the available observations distributed in time and space, is defined in discrete form as

$$J = \frac{1}{2}\sum_{i=0}^{M} \langle W(\boldsymbol{X}_i - \boldsymbol{X}_i^o), (\boldsymbol{X}_i - \boldsymbol{X}_i^o)\rangle \tag{A.1}$$

where $\boldsymbol{X}_i$ and $\boldsymbol{X}_i^o$ are the model solution and observation at i-th time level, respectively, and W is the weighting function which can be taken as the inverse of the estimate of the statistical root-mean-square observation errors (see remarks in body of text).

Now consider a perturbation, U', on the initial condition U, Eqs. (2.1), (2.2) become

$$\frac{\partial(\boldsymbol{X}+\hat{X})}{\partial t} = F(\boldsymbol{X}+\hat{X}) \tag{A.2}$$

$$\boldsymbol{X}(t_0) + \hat{X}(t_0) = U + U' \tag{A.3}$$

where $\hat{X}$ is the resulting perturbation of the variable $\boldsymbol{X}$. Expanding (A.2) at $\boldsymbol{X}$ and retaining only the first order term, one obtains

$$\frac{\partial \hat{X}}{\partial t} = \frac{\partial F}{\partial \boldsymbol{X}}\hat{X}, \tag{A.4}$$

$$\hat{X}(t_0) = U'. \tag{A.5}$$

Equations (A.4), (A.5) are defined as the tangent linear equations of Eqs. (2.1), (2.2).

The variation of the distance function J due to the perturbation U' is

$$\delta J = \sum_{i=0}^{M} \langle W(\boldsymbol{X}_i - \boldsymbol{X}_i^o), \hat{X}_i\rangle. \tag{A.6}$$

Using the Euler time differencing scheme for example one obtains from (A.4)

$$\begin{aligned}\hat{X}_{i+1} &= \hat{X}_i + \Delta t\left[\frac{\partial F}{\partial \boldsymbol{X}}\right]_i \hat{X}_i \\ &= \left[I + \Delta t\left(\frac{\partial F}{\partial \boldsymbol{X}}\right)_i\right]\hat{X}_i \\ &= \prod_{j=0}^{i}\left[I + \Delta t\left(\frac{\partial F}{\partial \boldsymbol{X}}\right)_j\right]\hat{X}_0 \end{aligned} \tag{A.7}$$

where Δt is the constant time step, I is the unit matrix operator, $\left(\frac{\partial F}{\partial \boldsymbol{X}}\right)_j$ represents the j-th row of the matrix $\frac{\partial F}{\partial \boldsymbol{X}}$, and $\prod_{j=0}^{i}$ denotes the product of $i+1$ factors.

Substituting (A.7) into (A.6) and using basic concepts of adjoint operators, we obtain the following expression

$$\begin{aligned}\delta J = \sum_{i=1}^{M} \left\langle \left\{\prod_{j=0}^{i-1}\left[I + \Delta t\left(\frac{\partial F}{\partial \boldsymbol{X}}\right)_j\right]\right\}^* W(\boldsymbol{X}_i - \boldsymbol{X}_i^o), \hat{X}_0 \right\rangle \\ + \langle W(\boldsymbol{X}_0 - \boldsymbol{X}_0^o), \hat{X}_0\rangle \end{aligned} \tag{A.8}$$

where $(\ )^*$ denotes the adjoint of $(\ )$. On the other hand, we have

$$\delta J = \langle \nabla_U J, \hat{X}_0\rangle. \tag{A.9}$$

Equating Eqs. (A.8) and (A.9), one obtains the gradient of the cost function with respect to the initial conditions as

$$\nabla_U J = \sum_{i=0}^{M}\left\{\prod_{j=0}^{i-1}\left[I + \Delta t\left(\frac{\partial F}{\partial \boldsymbol{X}}\right)_j\right]\right\}^* W(\boldsymbol{X}_i - \boldsymbol{X}_i^o). \tag{A.10}$$

The i-th term in (A.10) can be obtained by a backwards integration of the following adjoint equation

$$-\frac{\partial P}{\partial t} = \left(\frac{\partial F}{\partial \boldsymbol{X}}\right)^* P \tag{A.11}$$

from the i-th time step to the initial step, starting from

$$P_i = W(\boldsymbol{X}_i - \boldsymbol{X}_i^o) \tag{A.12}$$

where P represents the adjoint variables corresponding to $\hat{X}$. It appears that M integrations of the adjoint model, starting from different time steps $t_M, t_{M-1}, \ldots, t_1$, are required to obtain the gradient $\nabla_U J$. However, since the adjoint model (A.11) is linear, only one integration from t_M to t_0 of the adjoint equation is required to calculate the gradient of the cost function with respect to the initial conditions.

In summary, the gradient of the cost function with respect to the initial condition U can be obtained by the following procedure:

(a) Integrate the model from t_0 to t_M from initial condition (2.2) and store in memory the corresponding sequence of the model states $\boldsymbol{X}_i$ $(i = 0, 1, \ldots, M)$;

(b) Starting from $P_M = W(\boldsymbol{X}_M - \boldsymbol{X}_M^o)$, integrate the "forced" adjoint equation (A.11) backwards in time from t_M to t_0 with a forcing term $W(\boldsymbol{X}_i - \boldsymbol{X}_i^o)$ being added to the right-hand-side of (A.11) at the i-th time step when an observation is encountered. The final result P_0 is the value of gradient of the cost function with respect to the initial condition.

It is worth noting that

(i) When the observations do not coincide with the model grid points, the model solution should be interpolated to the observations, i.e., $C\boldsymbol{X} - \boldsymbol{X}^o$ should be used instead of $\boldsymbol{X} - \boldsymbol{X}^o$ in the cost function definition, where the operator C represents the

process of interpolating the model solution to space and time locations where observations are available.

(ii) We note that the numerical cost of the adjoint model computation is about the same as the cost of one integration of the tangent linear model, the latter involving a computational cost of between 1 and 2 integrations of the nonlinear model.

Appendix B

The Verification of the Correctness of FOA and SOA

It is very important to verify the correctness of the FOA and SOA codes. A Taylor expansion in the direction of Y leads to

$$J(X+\alpha Y) = J(X) + \alpha \frac{\partial J(X)}{\partial X} Y + \frac{1}{2}\alpha^2 Y^t \frac{\partial^2 J(X)}{\partial X^2} Y + O(\alpha^3) \tag{B.1}$$

where α is a small scalar, Y is a random perturbation vector which can be generated by using the randomizer on the Cray-YMP computer and Y^t denotes the transpose of the vector Y. Equation (B.1) can be used to define two functions of α

$$\psi(\alpha) = \frac{J(X+\alpha Y) - J(X)}{\alpha \dfrac{\partial J(X)}{\partial X} Y} \tag{B.2}$$

and

$$\phi(\alpha) = \frac{J(X+\alpha Y) - J(X) - \alpha \dfrac{\partial J(X)}{\partial X} Y}{\dfrac{1}{2}\alpha^2 Y^t \dfrac{\partial^2 J(X)}{\partial X^2} Y} \tag{B.3}$$

then for small α we have

$$\psi(\alpha) = 1 + O(\alpha) \tag{B.4}$$

$$\phi(\alpha) = 1 + O(\alpha). \tag{B.5}$$

For values of α which are small but not very close to the machine zero, one should except a value of $\psi(\alpha)$ or $\phi(\alpha)$ approaching 1 linearly for a wide range of magnitudes of α.

The experiment was performed using a limited area 2-D shallow water equation model. The results are shown in Fig. 1. It is clearly seen that for values of α between 10^0–10^{-11}, unit values for $\psi(\alpha)$ and $\phi(\alpha)$ are obtained. The correctness of the gradient of the cost function and the correctness of the Hessian/vector product have therefore been verified.

Acknowledgments

The authors gratefully acknowledge the support of Dr. Pamela Stephens, Director of GARP Division of the Atmospheric Sciences at NSF. This research was funded by NSF Grant ATM-9102851. Additional support was provided by the Supercomputer Computations Research Institute at Florida State University, which is partially funded by the Department of Energy through contract No. DE-FC0583-ER250000. Finally, the authors would like to acknowledge the thorough work of both reviewers, whose in-depth remarks have rendered the paper clearer and made it more legible, and in particular the perceptive remarks of Prof. Stephen Cohn.

References

Cacuci, Dan G., 1981: Sensitivity theory for nonlinear systems. II. Extensions to additional classes of responses. *J. Math. Phy.*, **22**(12), 2803–2812.

Courtier, P., Talagrand, O., 1990: Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. *Tellus*, **42A**, 531–549.

Chao, W. C., Chang, L., 1992: Development of a 4-dimensional analysis system using the adjoint method at GLA. Part 1: Dynamics. *Mon. Wea. Rev.*, **120**, 1661–1673.

Courtier, P., 1985: Experiments in data assimilation using the adjoint model technique. Workshop on High-Resolution Analysis ECMWF (UK) June 1985.

Courtier, P., Talagrand, O., 1987: Variational assimilation of meteorological observations with the adjoint equations Part 2. Numerical results. *Quart. J. Roy. Meteor. Soc.*, **113**, 1329–1347.

Derber, J. C., 1985: The variational four dimensional assimilation of analysis using filtered models as constraints. Ph.D. Thesis, Univ. of Wisconsin-Madison, 141 pp.

Derber, J. C., 1987: Variational four dimensional analysis using the quasigeostrophic constraint. *Mon. Wea. Rev.*, **115**, 998–1008.

Fletcher, R., 1987: *Practical Methods of Optimization, Vol. 1, 2nd ed.*, Chichester, New York: John Wiley, 436 pp.

Gill, P. E., Murray, W., Wright, M. H., 1981: *Practical Optimization.* London, Academic Press, 401 pp.

Grammeltvedt, A., 1969: A survey of finite-difference schemes for the primitive equations for a barotropic fluid. *Mon. Wea. Rev.*, **97**, 387–404.

Hall, M. C. G., Cacuci, D. G., Schlesinger, M. E., 1982: Sensitivity analysis of a radiative–convective model by the adjoint method. *J. Atmos. Sci.*, **39**, 2038–2050.

Hall, M. C. G., Cacuci, D. G., 1983: Physical interpretation of the adjoint functions for sensitivity analysis of atmospheric models. *J. Atmos. Sci.*, **40**, 2537–2546.

Hoffmann, R. N., 1986: A four dimensional analysis exactly satisfying equations of motion. *Mon. Wea. Rev.*, **114**, 388–397.

Kontarev, G., 1980: The adjoint equation technique applied to meteorological problem. *EC MWF Tech. Rep.*, Reading, U.K., 21, 1–21 pp.

Le Dimet, F. X., 1982: A general formalism of variational analysis. CIMMS Report, Norman, OK 73091, 22, 1–34 pp.

Le Dimet, F. X., Talagrand, O., 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97–110.

Le Dimet, F. X., Nouailler, A., 1986: In: Sasaki, Y. K. (ed.) *Assimilation of Dynamics in Geosciences.* Amsterdam: Elsevier, 1986, 181–198.

Lewis, J. M., Derber, J. C., 1985: The use of adjoint equations

to solve a variational adjustment problem with advective constraints. *Tellus,* **37A**, 309–322.

Liu, D. C., Jorge Nocedal, 1989: On the limited memory BFGS method for large scale minimization. *Mathematical Programming,* **45**, 503–528.

Lions, J. L., 1971: *Optimal Control of Systems Governed by Partial Differential Equations* (Translated by Mitter, S. K.). Berlin, Heidelberg: Springer, 396 pp

Lorenc, A. C., 1988a: Optimal nonlinear objective analysis. *Quart. J. Roy. Meteor. Soc.,* **114,** 205–240.

Lorenc, A. C., 1988b: A practical approximation to optimal four dimensional objective analysis. *Mon. Wea. Rev.,* **116**, 730–745.

Luenberger, D. G., 1984: *Linear and Nonlinear Programming,* 2nd ed. Reading, MA: Addison-Wesley, 491 pp.

Marchuk, G. I., 1974: *Numerical Solution of the Problem of Dynamics of Atmosphere and Ocean* (in Russian). Leningrad: Gidrometeoizdat, 303 pp.

Nash, S. G., 1985: Preconditioning of truncated-Newton methods. *SIAM J. Sci. Stat. Comput.,* **6**(3), 599–616.

Navon, I. M., Zou, X. L., Johnson, K., Derber, J., Sela, J., 1990: The adjoint of NMC spectral model. Tech. Report, FSU, Tallahassee, FSU-SCRI-REPORT-90-22, 71 pp.

Navon, I. M., Zou, X. L., Derber, J., Sela, J., 1992: Variational data assimilation with an adiabatic version of the NMC spectral model. *Mon. Wea. Rev.,* **120**, 1433–1466.

Penenko, V., Obraztsov, N. N., 1976: A variational initialization method for the fields of the meteorological elements. *Meteorol. Gidrol.* (English translation), **11**, 1–11.

Strang, G. 1986: *Introduction to Applied Mathematics.* Wellesley: Wellesley-Cambridge Press, 758 pp.

Sykes, J. F., Wilson, J. L., R. W. Andrews, 1985: Sensitivity analysis for steady state groundwater flow using adjoint operators. *Water Resources Research,* **21**(3), 359–371.

Talagrand, O., Courtier, P., 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation-Part 1. Theory. *Quart. J. Roy. Meteor. Soc.,* **113**, 1311–1328.

Thacker, W. C., Long, R. B., 1988: Fitting dynamics to data. *J. Geophys. Res.,* **93**, 1227–1240.

Thacker, W. C., 1989: The Role of Hessian Matrix in Fitting models to Measurements. *J. Geophys. Res.,* **94**, 6177–6196.

Thépaut J. N., Courtier, 1991: Four-dimensional variational data assimilation using the adjoint of multilevel primitive equation model. *Quart. J. Roy. Meteor. Soc.,* **117**, 1225–1254.

Zou, X. L., Navon, I. M., Berger, M., Paul K. H. Phua, Schlick, T., Le Dimet, F. X., 1992: Numerical experience with limited-Memory Quasi-Newton methods and Truncated Newton methods. *SIAM J. Optimization* (in press).

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

Zou, X. L., Navon, I. M., Sela, J., 1992: Control of gravity oscillations in variational data assimilation. *Mon. Wea. Rev.* (in press).

Authors' addresses: Zhi Wang and I. M., Navon, Department of Mathematics and Supercomputer Computations, Research Institute, Florida State University, Tallahassee, FL 32306, U.S.A.; F. X. Le Dimet, Grenoble, Laboratoire de Modelisation et Calcul, B.P. 53X, F-38041 Grenoble Cedex, France and X. Zou, Supercomputer Computations Research Institute, Florida State University, Tallahassee, FL 32306, U.S.A.