Conjugate Gradient Algorithm in the Four-Dimensional Variational Data Assimilation System in GRAPES

Yongzhu LIU*, Lin ZHANG, and Zhihua LIAN

National Meteorological Center, *Beijing* 100081

(Received April 8, 2018; in final form August 20, 2018)

ABSTRACT

Minimization algorithms are singular components in four-dimensional variational data assimilation (4DVar). In this paper, the convergence and application of the conjugate gradient algorithm (CGA), which is based on the Lanczos iterative algorithm and the Hessian matrix derived from tangent linear and adjoint models using a non-hydrostatic framework, are investigated in the 4DVar minimization. First, the influence of the Gram-Schmidt orthogonalization of the Lanczos vector on the convergence of the Lanczos algorithm is studied. The results show that the Lanczos algorithm without orthogonalization fails to converge after the ninth iteration in the 4DVar minimization, while the orthogonalized Lanczos algorithm converges stably. Second, the convergence and computational efficiency of the CGA and quasi-Newton method in batch cycling assimilation experiments are compared on the 4DVar platform of the Global/Regional Assimilation and Prediction System (GRAPES). The CGA is 40% more computationally efficient than the quasi-Newton method, although the equivalent analysis results can be obtained by using either the CGA or the quasi-Newton method. Thus, the CGA based on Lanczos iterations is better for solving the optimization problems in the GRAPES 4DVar system.

Key words: numerical weather prediction, Global/Regional Assimilation and Prediction System, four-dimensional variation, conjugate gradient algorithm, Lanczos algorithm

Citation: Liu, Y. Z., L. Zhang, and Z. H. Lian, 2018: Conjugate gradient algorithm in the four-dimensional variational data assimilation system in GRAPES. *J. Meteor. Res.*, **32**(6), 974–984, doi: 10.1007/s13351-018-8053-2.

1. Introduction

The application of high-resolution data assimilation constitutes a mainstream technology for improving numerical weather prediction (NWP) models. Variational data assimilation, which is used to solve analysis problems by minimizing a given cost function, is the best way to estimate model initial conditions by accurately combining observation and background fields (Rabier, 2005; Bannister, 2017). Three-dimensional variational data assimilation (3DVar) was widely used in NWP centers during the twentieth century (Courtier et al., 1998; Rabier et al., 1998; Lorenc et al., 2000). However, 3DVar erroneously assumes that observations acquired at different times are taken at the same time within the assimilation window. To overcome the shortcomings of 3DVar, four-dimensional variational data assimilation (4DVar) seeks an optimal balance between observations scattered through time and space over a finite 4D analysis volume with priori information; consequently, 4DVar is able to closely fit both observations and a priori initial estimates to generate the optimal initial conditions for NWP models (Thépaut et al, 1993; Courtier et al., 1994). For more than a decade, 4DVar has been the most successful data assimilation method for global NWP models; it has been used by many of the main global NWP centers, such as the ECMWF (Rabier et al., 2000), the French national meteorological service Météo-France (Janisková et al., 1999), the Met Office (Rawlins et al., 2007), and the meteorological service of Canada (Laroche et al., 2007). In recent years, some new 4DVar methods for global NWP models have emerged, including the ensemble-based 4DVar technique (Liu and Xiao, 2013) and hybrid 4DVar that adds flow-dependent ensemble covariance to traditional incremental 4DVar, for example, the ensemble data assimilations at ECMWF (Isaksen et al.,

Supported by the China Meteorological Administration Special Public Welfare Research Fund (GYHY201506003).

*Corresponding author: liuyzh@cma.gov.cn.

©The Chinese Meteorological Society and Springer-Verlag Berlin Heidelberg 2018

2010) and the hybrid-4DVar method employed at the Met Office (Clayton et al. 2013； Lorenc et al., 2015).

Variational data assimilation is a solution to large-scale unconstrained optimization problems. The cost function measuring the misfit between the background and the observations is first defined, and the optimal values are then determined by using various large-scale unconstrained minimization algorithms. Variational data assimilation techniques, especially 4DVar approaches based on the tangent linear model and adjoint model, are computationally expensive; thus, the development of a robust and efficient minimization algorithm is crucial (Fisher, 1998; Gürol et al., 2014). Two common minimization algorithms used in 4DVar systems are the conjugate gradient algorithm (CGA; Fisher, 1998) and quasi-Newton methods, including the limited-memory quasi-Newton method (i.e., the limited-memory Broyden–Fletcher–Goldfarb–Shanno, L-BFGS; Liu and Nocedal, 1989) and the truncated Newton method (Nash, 1984). Just as the L-BFGS method attempts to combine the modest storage and computational requirements of CGA methods with the convergence properties of standard quasi-Newton methods, truncated Newton methods attempt to retain the rapid (quadratic) convergence rate of classic Newton methods while making the storage and computational requirements feasible for large sparse matrices (Zou et al., 1993). Zou et al. (1993) compared the L-BFGS method with two truncated Newton methods on several test problems, including problems in meteorology and oceanography; their results confirmed that the L-BFGS seems to be the most efficient approach and is a particularly robust and user-friendly technique. Navon and Legler (1987) compared a number of different CGA and L-BFGS approaches for problems in meteorology and concluded that the L-BFGS is the most adequate for large-scale unconstrained minimization algorithms in meteorology. Furthermore, Fisher (1998) compared different CGA and truncated Newton methods in the ECMWF, and they concluded that the CGA was the most adequate for their 4DVar system. Therefore, the preconditioned CGA is used in the operational 4DVar system of ECMWF (Trémolet, 2007).

The 3DVar operational assimilation system is employed in the Global/Regional Assimilation and Prediction System (GRAPES; Shen et al., 2017) with the L-BFGS minimization algorithm (Xue et al., 2008). The GRAPES dynamical core uses a non-hydrostatic framework with two-time-layer semi-implicit and semi-Lagrangian discretization and employs a latitude–longitude grid with the staggered Arakawa C grid for spatial discretization. A 4DVar system has been developed in the GRAPES to improve its operational prediction quality by using the non-hydrostatic tangent linear and adjoint models, which were developed for the GRAPES global data assimilation system (Liu et al., 2017). The L-BFGS method is currently applied to the GRAPES 4DVar system, but its low convergence rate leads to a low computational efficiency (Zhang and Liu, 2017). In this paper, to select a robust and efficient minimization algorithm for the GRAPES 4DVar system, the convergence of the CGA is thoroughly examined, and the CGA is compared with the L-BFGS method in the GRAPES 4DVar scheme.

This paper is organized as follows. The data and methods are described in Section 2. Section 3 investigates the convergence of the CGA, and some results of the CGA in the GRAPES 4DVar system based on the numerical experiments are presented in Section 4. The conclusions and outlook are presented in Section 5.

2. Data and methods

2.1 *Incremental 4DVar*

Incremental formulation is commonly used in variational data assimilation systems (Courtier et al., 1994; Trémolet, 2007). The incremental scheme offers two main advantages: 1) the tangent linear model and adjoint model can be used with a reduced resolution during minimization, largely reducing the computational cost of 4DVar; 2) the cost function becomes strictly quadratic, and thus, the convergence rate of the minimization can be greatly improved (Fisher, 1998). The incremental formulation scheme includes two components, namely, the inner loop and the outer loop. The outer loop utilizes the initial estimate of the atmospheric state as the initial condition of the forecast model and obtains the model trajectory within the assimilation time windows; this trajectory is then used to calculate the observational increments within the time windows. The purpose of the inner loop is to solve the minimization problem by an iterative algorithm for the variational data assimilation.

In the 4DVar incremental formulation, the first-order approximation of the cost function J is written as the control variable $\delta\boldsymbol{x}$ (Courtier et al., 1994; Fisher, 1998; Trémolet, 2007):

$$\begin{aligned} J(\delta\boldsymbol{x}) &= \frac{1}{2}\delta\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}^{-1}\delta\boldsymbol{x} + \frac{1}{2}\sum\nolimits_{i=0}^{n}(\boldsymbol{H}_i\boldsymbol{L}_{0\to i}\delta\boldsymbol{x}-\boldsymbol{d}_i)^{\mathrm{T}}\boldsymbol{R}_i^{-1}(\boldsymbol{H}_i\boldsymbol{L}_{0\to i}\delta\boldsymbol{x}-\boldsymbol{d}_i) \\ &= \frac{1}{2}\delta\boldsymbol{x}^{\mathrm{T}}\left(\boldsymbol{B}^{-1}+\sum\nolimits_{i=0}^{n}\boldsymbol{L}_{i\to 0}^{\mathrm{T}}\boldsymbol{H}_i^{\mathrm{T}}\boldsymbol{R}_i^{-1}\boldsymbol{H}_i\boldsymbol{L}_{0\to i}\right)\delta\boldsymbol{x} - \delta\boldsymbol{x}^{\mathrm{T}}\sum\nolimits_{i=0}^{n}\boldsymbol{L}_{i\to 0}^{\mathrm{T}}\boldsymbol{H}_i^{\mathrm{T}}\boldsymbol{R}_i^{-1}\boldsymbol{d}_i + \frac{1}{2}\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{R}_i^{-1}\boldsymbol{d}_i, \end{aligned} \tag{1}$$

where $\delta\boldsymbol{x}$ is the departure from the background ($\delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_\mathrm{b}$), which will be the analysis increment; $\boldsymbol{x}$ is the model state at time t_0; $\boldsymbol{x}_\mathrm{b}$ is the background state at time t_0; $\boldsymbol{B}$ is the background error covariance matrix; $\boldsymbol{R}_i$ is the observational error covariance matrix at time t_i; $\boldsymbol{H}_i = \partial\mathcal{H}_i/\partial\boldsymbol{x}$ is the linearized observation operator of the nonlinear observation operator $\mathcal{H}_i$ at time t_i; $\boldsymbol{L}_{0\to i} = \partial\mathcal{M}_{0\to i}/\partial\boldsymbol{x}$ is the tangent linear model of the nonlinear model $\mathcal{M}_{0\to i}$ integrated from time t_0 to time t_i; and $\boldsymbol{L}_{i\to 0}^\mathrm{T}$ is the corresponding adjoint operator of $\boldsymbol{L}_{0\to i}$ that constitutes backward integration from time t_i to time t_0; $\boldsymbol{d}_i = \boldsymbol{o}_i - \boldsymbol{H}_i\partial\mathcal{M}_{0\to i}(\boldsymbol{x}_\mathrm{b})$ represents the observational increment at time t_i; $\boldsymbol{o}_i$ is the observation at time t_i. The solution of the adjoint operator can be coded from the corresponding tangent linear model code, and it does not require deriving the adjoint equations analytically (Talagrand and Courtier, 1987).

To solve the minimization problem of Eq. (1), the gradient of the control variable $\delta\boldsymbol{x}$ is calculated with the following equation (Courtier et al., 1994; Fisher, 1998):

$$\nabla J(\delta\boldsymbol{x}) = \left(\boldsymbol{B}^{-1} + \sum_{i=0}^{n}\boldsymbol{L}_{i\to 0}^\mathrm{T}\boldsymbol{H}_i^\mathrm{T}\boldsymbol{R}_i^{-1}\boldsymbol{H}_i\boldsymbol{L}_{0\to i}\right)\delta\boldsymbol{x} - \sum_{i=0}^{n}\boldsymbol{L}_{i\to 0}^\mathrm{T}\boldsymbol{H}_i^\mathrm{T}\boldsymbol{R}_i^{-1}\boldsymbol{d}_i, \tag{2}$$

where the minimization of the cost function can be obtained by minimization algorithms such as the Newton method or CGA. The second partial derivative of $J(\delta\boldsymbol{x})$, the Hessian matrix, is denoted J'' and is calculated as follows (Courtier et al., 1994; Fisher, 1998):

$$J'' = \boldsymbol{B}^{-1} + \sum_{i=0}^{n}\boldsymbol{L}_{i\to 0}^\mathrm{T}\boldsymbol{H}_i^\mathrm{T}\boldsymbol{R}_i^{-1}\boldsymbol{H}_i\boldsymbol{L}_{0\to i}. \tag{3}$$

Thus, the solution of Eq. (2) is equal to the solution of the system of linear equations $J''\delta\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{b} = \sum_{i=0}^{n}\boldsymbol{L}_{i\to 0}^\mathrm{T}\boldsymbol{H}_i^\mathrm{T}\boldsymbol{R}_i^{-1}\boldsymbol{H}_i\boldsymbol{L}_{0\to i}$.

Because $\boldsymbol{B}$ is usually a large sparse matrix and is nearly ill conditioned, it is difficult to solve the minimization problem in Eq. (1). To achieve acceptable convergence rates, it is necessary to perform some transforms and preconditioning for $\boldsymbol{B}$.

In the GRAPES 4DVar system, the basic atmospheric state variables $\boldsymbol{x}$ are the two wind vectors (denoted by $\boldsymbol{u}$ and $\boldsymbol{v}$), the relative humidity $\boldsymbol{q}$, and the non-dimensional pressure as an independent variable (denoted by Π), which is the analysis variable that represents the quality field, instead of potential temperature (denoted by $\boldsymbol{\theta}$). Thus, the analysis increment is $\delta\boldsymbol{x} = (\delta\boldsymbol{u}, \delta\boldsymbol{v}, \delta\Pi)^\mathrm{T}$, which can be transformed into a new vector $\delta\boldsymbol{x}_u = (\delta\psi, \delta\chi_u, \delta\Pi_u)^\mathrm{T}$, $\delta\boldsymbol{x} = \mathcal{P}\delta\boldsymbol{x}_u$, where $\mathcal{P}$ is a physical balance transformation operator (Xue et al., 2008).

Therefore, the background error covariance matrix $\boldsymbol{B}$ can be split into three independent blocked matrices $\boldsymbol{B} = \mathcal{P}\boldsymbol{B}_u\mathcal{P}^{-1}$, thereby reducing the scale of the matrix computation. This method of preconditioning through a change in the variable $(\boldsymbol{B}_u)^{1/2}$ is currently used in the GRAPES 4DVar system. Introducing a new control variable $\boldsymbol{w}$ in the cost function, the preconditioning transform of the variable $\delta\boldsymbol{x}$ is expressed as $\delta\boldsymbol{x} = \mathcal{P}\delta\boldsymbol{x}_u = \mathcal{P}\Sigma_u\mathbb{U}\boldsymbol{w}$, where $\boldsymbol{B}_u = \Sigma_u\mathbb{U}\mathbb{U}\Sigma_u$. Therefore, Eq. (1) can be expressed by using the control variable $\boldsymbol{w}$:

$$J(\boldsymbol{w}) = \frac{1}{2}\boldsymbol{w}^\mathrm{T}\boldsymbol{w} + \frac{1}{2}\sum_{i=0}^{n}(\boldsymbol{H}_i\boldsymbol{L}_{0\to i}\mathcal{P}\Sigma_u\mathbb{U}\boldsymbol{w} - \boldsymbol{d}_i)^\mathrm{T} \cdot \boldsymbol{R}_i^{-1}(\boldsymbol{H}_i\boldsymbol{L}_{0\to i}\mathcal{P}\Sigma_u\mathbb{U}\boldsymbol{w} - \boldsymbol{d}_i). \tag{4}$$

2.2 *The L-BFGS and CGA in GRAPES 4DVar*

The L-BFGS algorithm (Appendix A) in the GRAPES 4DVar system uses the estimation to the inverse Hessian matrix to guide its search through the variable space. For the L-BFGS in the GRAPES 4DVar scheme, the initial Hessian matrix is the identity matrix, and the number of iterations m insomuch that the m previous values s_k and z_k are stored to compute the approximation of the inverse Hessian matrix is 12.

The CGA based on the Lanczos iteration (Appendix B) in GRAPES 4DVar is mainly applied to solve large sparse, symmetric, positive definite linear equations (Paige and Saunders, 1982). With this combination, the orthogonalization of the Lanczos algorithm can sufficiently overcome the instability of the CGA in providing practical solutions to the above equations.

The Hessian matrix J'' in Eq. (3) is a sparse, real, symmetric, positive definite matrix that can be computed by using $\boldsymbol{B}$, $\boldsymbol{R}$, $\boldsymbol{H}$, $\boldsymbol{L}$, and $\boldsymbol{L}^\mathrm{T}$. The convergence efficiency of the inner loop minimization of 4DVar is largely determined by the shape of the Hessian matrix, and the computational efficiency largely depends on that of the tangent linear model $\boldsymbol{L}$ and the adjoint model $\boldsymbol{L}^\mathrm{T}$ as well as the number of iterations in the minimization. Therefore, this approach effectively improves the computational efficiency of the 4DVar minimization by choosing an efficient iterative minimization algorithm.

2.3 *Orthogonalization of the CGA*

Rounding errors greatly affect the behavior of the Lanczos iteration for a practical minimization problem (Paige, 1970). For a 4DVar system in particular, there are often some computational errors from the tangent model and adjoint model of the Hessian matrix J'' as well as rounding errors from the iterations. These errors lead to a quick loss of orthogonality in the Lanczos vectors $\boldsymbol{q}_k$ in addition to the problem of "ghost" eigenvalues during the

Lanczos iterations. Moreover, there are multiple eigenvalues of $\boldsymbol{T}_k$ that correspond to simple eigenvalues of the Hessian matrix J''; this results in additional iterations and convergence failure. Thus, the application of the Lanczos algorithm can easily cause numerical instabilities in the solutions of large symmetric matrices. However, this issue can be overcome by conducting Gram-Schmidt orthogonalization on the Lanczos vectors (Paige, 1970), which is conducted primarily by three methods as follows:

(1) Full orthogonalization (Paige, 1970). This process conducts Gram-Schmidt orthogonalization to make the Lanczos vector $\boldsymbol{q}_{k+1}$ orthogonal to all of the previously computed Lanczos vectors. In detail, the Gram-Schmidt orthogonalization is applied to the residual vector $\boldsymbol{r}_{k+1}$ derived from the third step of the Lanczos algorithm [Eq. (B4)] and the Lanczos vector groups $(\boldsymbol{q}_{1,\cdots},\boldsymbol{q}_k)$, i.e., $\boldsymbol{r}_{k+1} = \boldsymbol{r}_{k+1} - \sum_{i=1}^{k} \langle \boldsymbol{r}_{k+1}, \boldsymbol{q}_i \rangle \boldsymbol{q}_i$. Thus, the Lanczos vector $\boldsymbol{q}_{k+1}$ will be orthogonal to the previously computed Lanczos vectors $(\boldsymbol{q}_{1,\cdots},\boldsymbol{q}_k)$.

(2) Partial orthogonalization (Simon, 1984). Consequently, instead of orthogonalizing $\boldsymbol{q}_{k+1}$ against all the previously computed Lanczos vectors, the same effect can be achieved by orthogonalizing $\boldsymbol{q}_{k+1}$ against the previously computed Lanczos vectors that are not orthogonal to $\boldsymbol{q}_{k+1}$. The detailed steps are similar to those in the full orthogonalization method. However, this method reduces the number of orthogonalized inner products and therefore improves the computational efficiency.

(3) Selective orthogonalization (Parlett and Scott, 1979). The method is similar to partial orthogonalization but orthogonalizing $\boldsymbol{q}_{k+1}$ against the much smaller set of converged eigenvectors of the Hessian matrix J''. This method can avoid some calculations of repeated eigenvalues, reduce the additional Lanczos iterations, and improve the computational efficiency. However, extra space is needed to store the eigenvectors.

The CGA has been successfully applied in the 4DVar system of ECMWF (Fisher, 1998; Trémolet, 2007). However, there are many differences between the GRAPES and ECMWF 4DVar systems. First, the ECMWF tangent linear model and adjoint model use a hydrostatic framework with spectral and reduced grids, while those in GRAPES employ a non-hydrostatic framework with a latitude–longitude grid. Especially in polar regions, the denser grid distribution of GRAPES adds a gradient sensitivity computed by the adjoint model, leading to an increase in the condition number of the Hessian matrix J'', thereby affecting the convergence rate. Second, the state variables of assimilation and the tangent linear model variables are the same as those in the ECMWF 4DVar system. However, there is a variable physical transform between the tangent linear model and the assimilation system (see Section 2.1).

3. Data and experiment

To further analyze the effectiveness of the CGA in a practical 4DVar system, we conduct one cycling assimilation experiment for a month. The time ranges from 0900 UTC 1 June to 0900 UTC 1 July 2016. The data used for the assimilation include conventional Global Telecommunication System (GTS) observations, including temperature, wind and relative humidity data derived from sounding, pressure data from ships and the Synoptic Ocean Prediction (SYNOP) experiment, and wind data from pilot readings, in addition to data from satellite-based platforms, such as the NOAA-15 Advanced Microwave Sounding Unit-A (AMSUA), NOAA-18 AMSUA, NOAA-19 AMSUA, MetOp-A AMSUA, MetOp-B AMSUA, National Polar-orbiting Partnership (NPP) Advanced Technology Microwave Sounder (ATMS) AMSUA, SeaWinds scatterometer, and Global Navigation Satellite System (GNSS) radio occultations. Satellite observations compose approximately 70% of the total observations. The assimilation window is 6 h, and the observational interval is 30 min. The horizontal resolution of the outer loop is 0.5°, and the model time step is 450 s. The horizontal resolution of the inner loop is 1.5°, and the model time step is 900 s. The number of vertical levels is 60, and the maximum number of iterations is 70 in the 4DVar minimization.

The following linearized physical processes are used in this 4DVar experiment: two dry linearized physical processes (vertical diffusion and subgrid-scale orographic effects) to improve the representation of perturbed fields in the tangent linear model (Liu et al., 2017), and two newly developed moist linearized physical parameterizations consisting of deep cumulus convection based on a new simplified Arakawa-Shubert scheme (Han and Pan, 2006) and the large-scale cloud and precipitation scheme described in Tompkins and Janisková (2004). The experimental environment is based on the high-performance computer (Sugon PI) at the China Meteorological Administration. In total, 256 CPU cores are used in these experiments. Two configurations of 4DVar experiments are tested:

(1) CGA experiments, in which the CGA is used for minimization in the 4DVar system; and

(2) L-BFGS experiments, in which the L-BFGS is

used for minimization in the 4DVar.

4. Results of the CGA in 4DVar

We perform numerical experiments on the GRAPES 4DVar system to investigate the convergence of the CGA therein. The experimental configuration is the same as that in the batch experiments in Section 3, which begin at 0900 UTC 1 June 2016, and the number of iterations is 120 in the minimization. Then, the numerical stability of the Lanczos algorithm in the 4DVar is tested against the four orthogonalization schemes described in Section 2.3: full orthogonalization, partial orthogonalization, selective orthogonalization, and without orthogonalization.

4.1 *Orthogonalization analysis of the CGA*

The convergence of the gradient norm $\|\nabla J(\delta \boldsymbol{x})\|$ [Eq. (2)] with the non-orthogonalized Lanczos vector $\boldsymbol{q}_k$ is shown in Fig. 1. The gradient norm fails to converge starting at the 9th iteration, which is partly the result of computational errors. As the iteration continues, the reduced orthogonality of the Lanczos vector $\boldsymbol{q}_k$ gives rise to a higher gradient norm. However, the convergence of the gradient norm is much better after performing full orthogonalization on the Lanczos vector $\boldsymbol{q}_k$ (blue dashed line in Fig. 1). In addition, the results of the first nine iterations are the same as those without orthogonalization. This outcome indicates that the orthogonalization on Lanczos vector $\boldsymbol{q}_k$ does not change the iteration results of the Lanczos algorithm when the effect of the computational errors is small, while the orthogonalization on the Lanczos vector $\boldsymbol{q}_k$ can effectively eliminate the effects of computational errors, leading to the stable convergence of the Lanczos algorithm when the computational errors become larger. Further, the results of partial orthogonalization (red dotted line in Fig. 1) on the Lanczos vector $\boldsymbol{q}_k$ are the same as those of full orthogonalization, and selective orthogonalization also produces the same results as full orthogonalization.

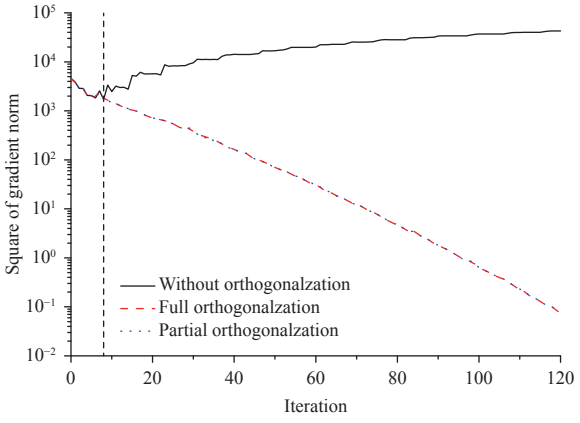

Fig. 1. Convergence of the conjugate gradient norm as a function of the number of iterations for a 4DVar cost function. The vertical axis is the square of the gradient norm, and it denotes the difference between the control vector at a given iteration and its 120th iteration (black solid line: without orthogonalization; red dashed line: full orthogonalization; blue dotted line: partial orthogonalization; vertical dotted line shows that the gradient norm fails to converge starting at the 9th iteration without orthogonalization).

The eigenvalue distribution of the Hessian matrix J'' under different orthogonalization methods is illustrated in Fig. 2. The eigenvalue distribution without orthogonalization on the Lanczos vector $\boldsymbol{q}_k$ is indicated by the solid line (Fig. 2). There are 53 convergent eigenvalues in total (circles on the solid line in Fig. 2); many repeated eigenvalues are associated with redundant iterations due to the loss of orthogonality of the Lanczos vector $\boldsymbol{q}_k$ during the iterations. This Lanczos algorithm is numerically unstable in the 4DVar minimization. The red dashed line in Fig. 2 shows the eigenvalue distribution with full orthogonalization performed on the Lanczos vector $\boldsymbol{q}_k$. Moreover, the number of convergent eigenvalues (triangles on the dashed line in Fig. 2) is 53, but these eigenvalues are no longer repeated. This result implies that the Lanczos algorithm is stable in the 4DVar minimization after conducting full orthogonalization on the Lanczos vector $\boldsymbol{q}_k$. The number of convergent eigenvalues (blue dotted line in Fig. 2) with partial orthogonalization applied to the Lanczos vector $\boldsymbol{q}_k$ is 49, which is 4 fewer than that with full orthogonalization. However, the eigenvalue distribution with partial orthogonalization is generally similar to that with full orthogonalization. Similarly, the eigenvalue distribution with selective orthogonalization is also generally similar to that with full orthogonal-

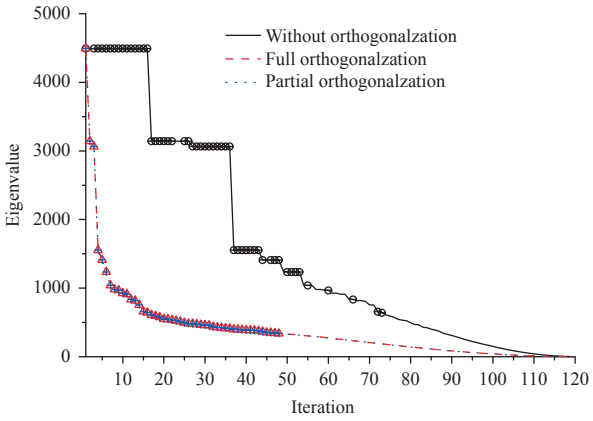

Fig. 2. Eigenvalue distribution of the Hessian matrix of a 4DVar minimization with different schemes of the orthogonalization of Lanczos vectors against the number of iterations (black solid line: without orthogonalization; red dashed line: full orthogonalization; blue dotted line: partial orthogonalization).

ization. Therefore, the Lanczos algorithm is stable in the 4DVar minimization using full orthogonalization, partial orthogonalization, or selective orthogonalization.

4.2 *Convergence analysis of the CGA*

In the 4DVar minimization, the convergence rate of the CGA depends on the eigenvalue distribution of the Hessian matrix J'' and the condition number κ (the ratio of the maximum eigenvalue to the minimum eigenvalue). The convergence estimation of the CGA, namely, the conjugated error $\left(\boldsymbol{e}_j = \delta\boldsymbol{x} - \delta\boldsymbol{x}_j\right)$, is based on the norm of the Hessian matrix, and it satisfies the following (Paige, 1970; Fisher, 1998):

$$\left\|\boldsymbol{e}_j\right\|_{J''}^2 = \left(\delta\boldsymbol{x} - \delta\boldsymbol{x}_j\right)^{\mathrm{T}} J'' \left(\delta\boldsymbol{x} - \delta\boldsymbol{x}_j\right) \leqslant 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^j \left\|\boldsymbol{e}_0\right\|_{J''}^2 . \quad (5)$$

Here, $\delta\boldsymbol{x}$ is the solution of the 4DVar minimization in Eq. (1) (the value of which is the estimated solution of the last iteration of the CGA), while $\delta\boldsymbol{x}_j$ is the estimated solution at the jth iteration of CGA. According to Eq. (B8), the CGA should converge better than the linear algorithm. Moreover, the convergence can be improved by the pre-optimization step of reducing the condition number.

To explore the convergence of the CGA in the GRAPES 4DVar system, we conduct an assimilation experiment (beginning at 0900 UTC 1 June 2016) with 120 iterations of 4DVar minimization. The maximum (minimum) eigenvalue of the Hessian matrix J'' estimated by the CGA in the 4DVar minimization is 4492.1 (1.03). Per Eq. (5), the convergence rate $\left(\sqrt{\kappa}-1\right)/\left(\sqrt{\kappa}+1\right)$ estimated by the condition number is 0.970, and the upper bound on the convergence rate is expressed by the solid line in Fig. 3. This result implies that the convergence of the Hessian matrix is unsatisfactory. However, in a practical calculation of the 4DVar minimization based on the CGA, the Hessian norm of the true iteration error $\left\|\boldsymbol{e}_j\right\|_{J''}^2$ decreases in magnitude from 10^3 to 10^{-2} after 120 iterations. The descent rate is clearly quicker than the convergence rate estimated by the condition number, which constitutes superlinear convergence. The above results are consistent with those based on the Integrated Forecasting System of ECMWF (Fisher, 1998).

In short, the Lanczos algorithm is numerically more stable in the 4DVar minimization if the Lanczos vector is orthogonal during the Lanczos iterations. Thus, performing Gram-Schmidt orthogonalization on the Lanczos vector is an effective way to ensure the numerical stability in the Lanczos algorithm. In this way, the convergence rate of 4DVar minimization is also improved. Considering the short computational time for orthogonalization in the whole 4DVar system, we exploit the Lanczos algorithm with full orthogonalization in the GRAPES 4DVar system to guarantee the orthogonality of the Lanczos vector.

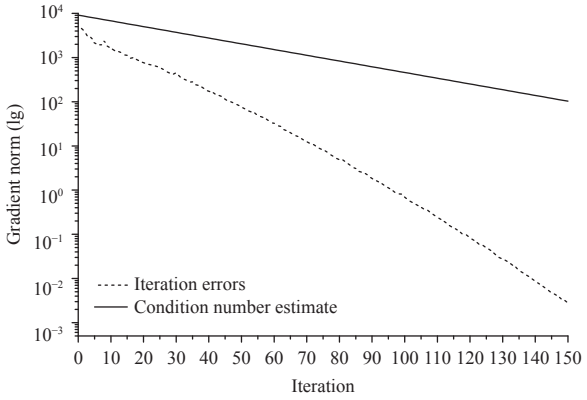

Fig. 3. Convergence of the CGA as a function of the number iterations for a 4DVar cost function. The dashed line is the square of the Hessian norm of the difference between the control vector and the value of the last iteration. The solid line is the upper bound of the convergence rate defined by Eq. (5).

5. Results of numerical experiments in 4DVar

5.1 *Comparison of convergence with the L-BFGS*

The CGA and quasi-Newton method are both quadratically convergent in theory and have the same quadratic termination property. However, these methods behave quite differently in practical applications, especially when applied to certain problems such as solving the minimization of 4DVar. To better compare the convergences of these two methods in 4DVar minimization problems, both cycling assimilation experiments begin on 10 June 2016. Recalculations are performed in the 4DVar experiments with the CGA using the background from the L-BFGS tests. The ratio of the root mean square of the gradient norm to its initial value is regarded as the convergence criterion (set as 0.03) during the minimization iterations. The maximum number of iterations is 70.

In the first dozen 4DVar minimization iterations in the four assimilation experiments, the gradient norms of the CGA and L-BFGS experiments both decrease with some oscillations, and a smaller amplitude is observed for the CGA experiments (Fig. 4). Then, the oscillation of the gradient norm for the CGA experiments becomes small, and these experiments satisfy the convergence criterion after approximately 40 iterations. However, the oscillation of the gradient norm for the L-BFGS experiments is still large, and the experiments do not converge after 70 iterations. In addition, the gradient norm in the L-BFGS experiments descends very slowly in the later stages, and the minimum is similar to that after approximately 40 it-

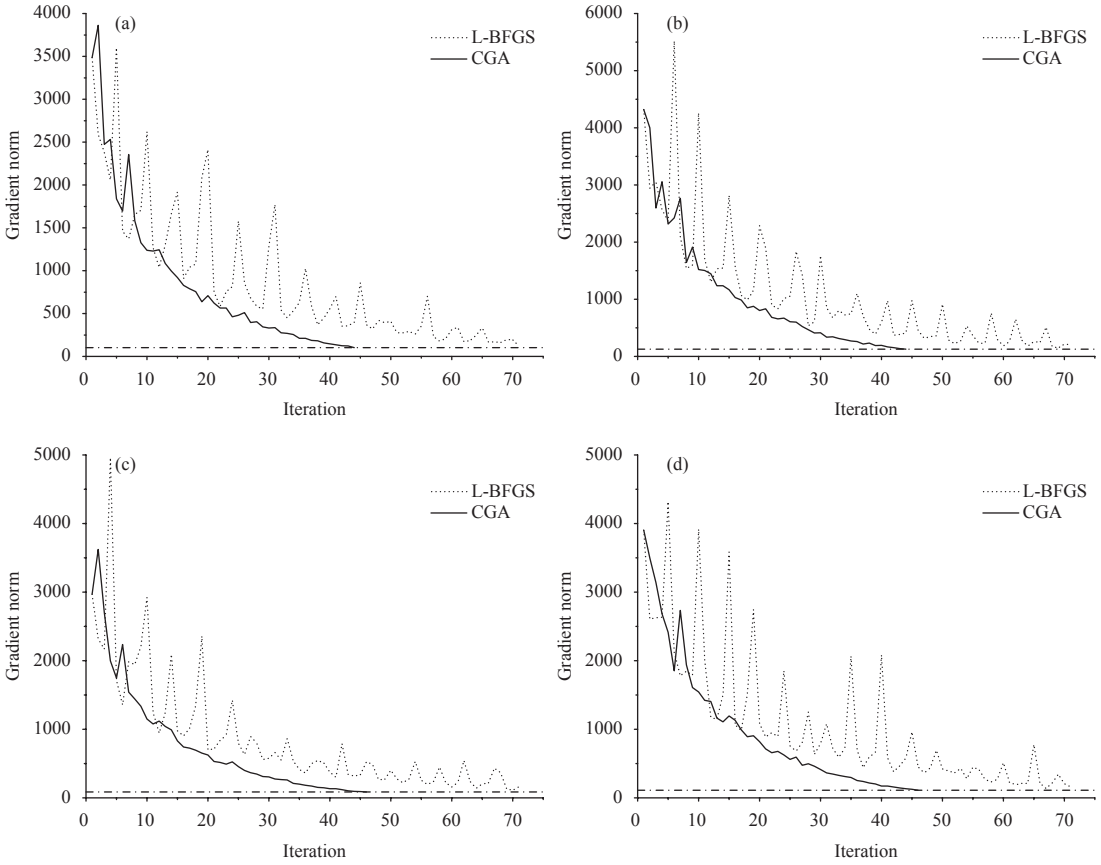

Fig. 4. Convergence of the Hessian norm for a 4DVar cost function under the GRAPES global 4DVar assimilation system starting at (a) 0300 UTC 10 June 2016, (b) 0900 UTC 10 June 2016, (c)1500 UTC 10 June 2016, and (d) 2100 UTC 10 June 2016 (solid line: CGA experiment; dotted line: L-BFGS experiment).

erations in the CGA experiments (dotted line in Fig. 4). The above results indicate that the convergence of the CGA is better than that of the L-BFGS in the 4DVar minimization. Moreover, the computational cost of the CGA is lower. Therefore, the CGA is preferable.

To better compare the convergences of the cost functions between the two sets of experiments, normalization is applied to the cost function to calculate the ratio of all the cost functions to the initial cost function within the iterations. The convergences of the cost functions for both the CGA experiments and the L-BFGS experiments on 20 June 2016 in the 4DVar cyclical assimilations are illustrated in Fig. 5. In the first twenty iterations, the descent rate of the cost functions of the CGA experiments are faster than those of the L-BFGS experiments. The convergence in the CGA experiments after 40 iterations is similar to that in the L-BFGS experiments after 70 iterations. Thus, the convergence rate of the CGA is much faster than that of the L-BFGS in the 4DVar minimization.

5.2 *Computational efficiency*

In the 4DVar minimization, the term $J''\boldsymbol{q}_k$ should be calculated with the iteration of the tangent and adjoint models. Therefore, the computational cost, which is determined by the number of iterations, is very high. Hence, improving the convergence rate and reducing the number of iterations is an effective way to improve the computational efficiency of the 4DVar minimization.

The numbers of iterations and the calculation times for the 4DVar minimization in the 121 cyclical assimilation tests for both experimental sets are plotted in Fig. 6. The CGA experiments satisfy the requirement for convergence within a maximum of 70 iterations in the minimization. The average number of iterations to reach convergence is 37, and the average minimization time in these CGA experiments is 861 s. However, most of the L-BFGS experiments do not meet this convergence requirement within the maximum number of iterations because some of the cost functions in the iterations do not decrease when using the L-BFGS minimization, and thus, it is necessary to choose another descent direction for the L-BFGS experiments, which requires additional calculations. Thus, the number of iterations for some L-BFGS experiments is greater than the maximum number of iter-

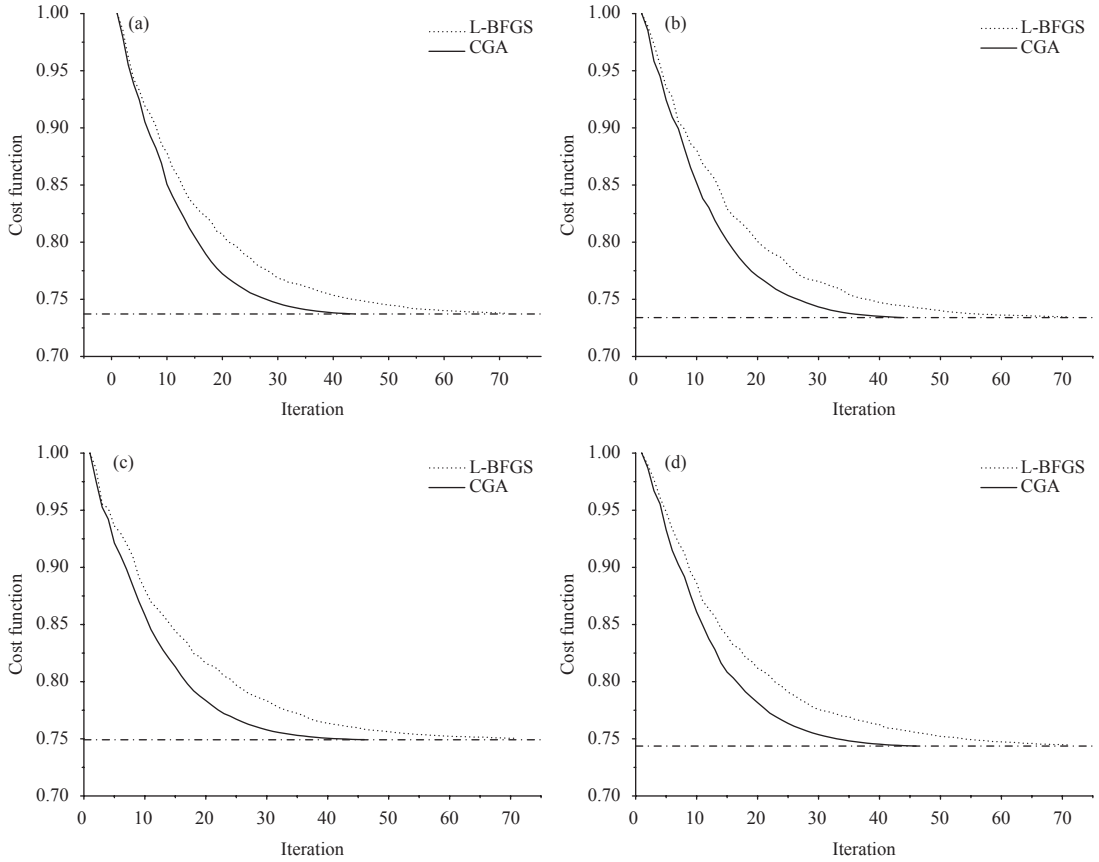

Fig. 5. Convergence of the 4DVar cost function under the GRAPES global 4DVar assimilation system starting at (a) 0300 UTC 20 June 2016, (b) 0900 UTC 20 June 2016, (c) 1500 UTC 20 June 2016, and (d) 2100 UTC 20 June 2016 (solid line: CGA experiment; dotted line: L-BFGS experiment; horizontal dashed line: final convergence rate of the cost function in the CGA experiment).

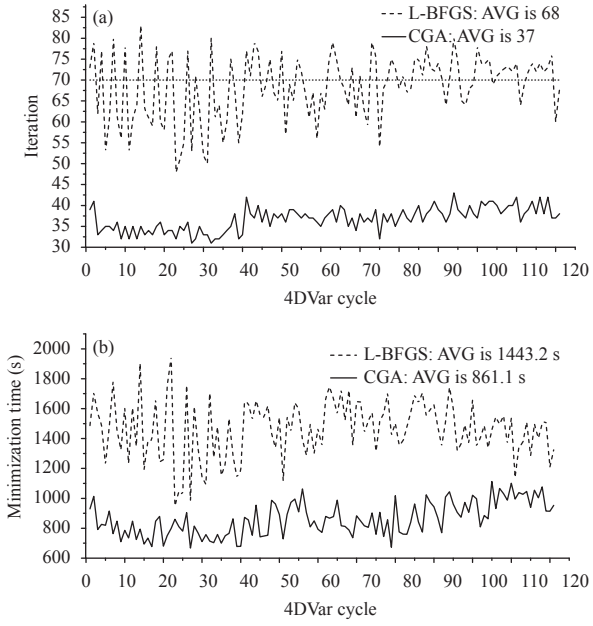

Fig. 6. (a) Number and (b) time of iterations in the 4DVar minimization in the 121 cyclical assimilation tests for the two experimental sets (solid line: CGA experiment; dashed line: L-BFGS experiment; AVG: the average number or time of iterations over one month)

ations (70). Furthermore, the average number of iterations in these L-BFGS experiments is 68, and the average minimization time in these L-BFGS experiments is 1443 s. The average number of iterations in the CGA experiments is 32 less than that in the L-BFGS experiments, representing a 45% improvement in the computational efficiency. Hence, the CGA can largely improve the computational efficiency without affecting the convergence in the GRAPES 4DVar system.

5.3 *The assimilation and forecasting results*

To compare the assimilation results of the batch experiments more reasonably, we exploit the 21-day (from 10 to 30 June 2016) results of the cycling assimilation tests to avoid the influence of the initial field. A statistical analysis of the batch background and analysis deviations from the radiosonde temperature observations for the L-BFGS experiment (shown in black) and the CGA experiment (shown in red) is shown in Fig. 7. The two experiments show very similar standard deviations (Fig. 7a) and biases (Fig. 7b) of the background fields and analysis fields at all levels. In addition, the statistics of the other types of observations between the two experiments are

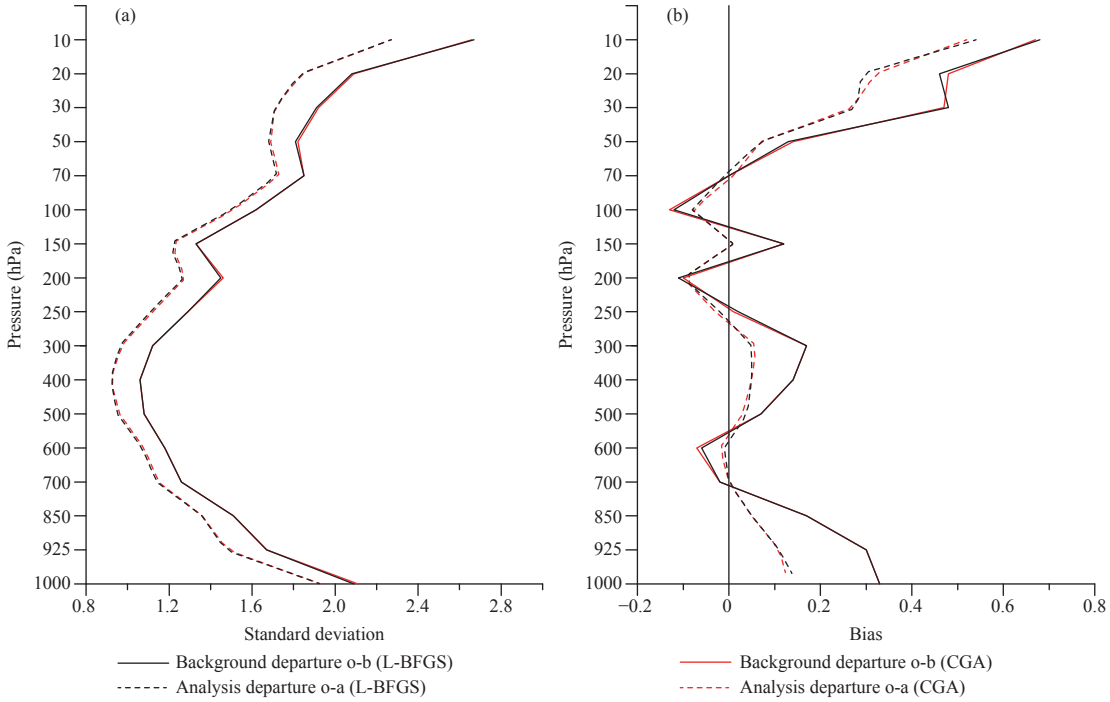

Fig. 7. (a) Standard deviations and (b) biases of background and analysis fields from radiosonde temperature observations for the L-BFGS experiment (black) and CGA experiment (red) (solid line: background departure o-b; dotted line: analysis departure o-a).

also similar. Therefore, this shows that the estimated solutions [Eq. (1)] for these two minimization algorithms are both reasonable when they reach the same convergence rate in the 4DVar minimization (Fig. 6). This result validates the potential of using these two iterative methods in the GRAPES 4DVar minimization.

6. Conclusions and discussion

A CGA based on the Lanczos iteration is investigated in this paper that considers latitudinal and longitudinal grid characteristics under a non-hydrostatic framework in the GRAPES tangent linear and adjoint models. This approach solves the convergence problem through orthogonalization in the Lanczos iteration. The CGA produces equivalent analysis results with far fewer iterations and a higher computational efficiency than the L-BFGS in the batch experiments on the 4DVar system. This conclusion for the GRAPES 4DVar system is consistent with that for the ECMWF 4DVar system. However, the denser grid distribution of GRAPES adds a gradient sensitivity computed by the adjoint model, leading to an increase in the condition number of the Hessian matrix, thereby affecting the convergence rate. However, this issue can be addressed by orthogonalization. Thus, the CGA is more suitable for the operational development of the GRAPES 4DVar system, indicating that the CGA is more suitable for minimization problems such as those in the 4DVar system.

To further improve the convergence of the 4DVar minimization problem, we need to explore the preconditioned CGA based on the eigenvector of the low-resolution minimization, perform additional outer loop updates in the framework of incremental analysis, and ultimately improve the 4DVar analysis technique.

Appendix A: A description of the L-BFGS

The L-BFGS algorithm is an optimization algorithm in the family of quasi-Newton methods that employs a limited amount of computer memory. The algorithm starts with an initial estimate $\delta \boldsymbol{x}_0$, and the initial gradient of the cost function is $\boldsymbol{g}_0 = \nabla J(\delta \boldsymbol{x}_0)$. A positive definite initial approximation of the inverse Hessian matrix is defined as $\boldsymbol{E}_0$ (which may be an identity matrix). Thus, the L-BFGS algorithm has the following basic structure for minimizing $J(\delta \boldsymbol{x})$ (Liu and Nocedal, 1989) for $k = 0, 1, \cdots$:

Step 1. Compute the search direction $d_k = -\boldsymbol{E}_k \boldsymbol{g}_k$, and set $\delta \boldsymbol{x}_{k+1} = \delta \boldsymbol{x}_k + \alpha_k d_k$, where α_k is the step size obtained by a safeguarded procedure, $\boldsymbol{E}_k$ is the approximation of the inverse Hessian matrix, and $\boldsymbol{g}_k = \nabla J(\delta \boldsymbol{x}_k)$.

Step 2. Set $\boldsymbol{s}_k = \delta \boldsymbol{x}_k - \delta \boldsymbol{x}_{k+1}$ and $\boldsymbol{z}_k = \boldsymbol{g}_k - \boldsymbol{g}_{k+1}$. To reduce the memory usage in the algorithm, $\boldsymbol{E}_{k+1}$ is generally updated by the previous m iterations $\boldsymbol{s}_k, \boldsymbol{s}_{k-1}, \cdots, \boldsymbol{s}_{k-m}$ and $\boldsymbol{z}_k, \boldsymbol{z}_{k-1}, \cdots, \boldsymbol{z}_{k-m}$: $\boldsymbol{E}_{k+1} = \left(\mathbf{I} - \rho_k \boldsymbol{s}_k \boldsymbol{z}_k^{\mathrm{T}}\right) \boldsymbol{E}_k \left(\mathbf{I} - \rho_k \boldsymbol{z}_k \boldsymbol{s}_k^{\mathrm{T}}\right) + \rho_k \boldsymbol{s}_k \boldsymbol{s}_k^{\mathrm{T}}$, where $\rho_k = \left(\boldsymbol{z}_k^{\mathrm{T}} \boldsymbol{s}_k\right)^{-1}$.

Step 3. Generate a new search direction $d_{k+1} = -\boldsymbol{E}_{k+1}\boldsymbol{g}_{k+1}$, and then go to step 1.

The L-BFGS algorithm attempts to combine modest storage and computational requirements for minimizing $J(\delta\boldsymbol{x})$. Therefore, with a lower number of iterations ($k < m$), $\boldsymbol{E}_{k+1}$ can capture only insufficient information of the Hessian matrix, and thus, the convergence efficiency of the L-BFGS algorithm is affected.

Appendix B: A description of the CGA based on Lanczos iterations

The CGA searches in the direction of conjugated base vectors over the Krylov subspace $\mathcal{K}(J'', \boldsymbol{r}_0)$ and derives the minimum of the target function (Fletcher and Reeves, 1964). The equation $\boldsymbol{r}_0 = \boldsymbol{b} - J''\delta\boldsymbol{x}_0$ is the initial residual. The Lanczos approach converts the large sparse symmetric matrix into a symmetric tridiagonal matrix by an orthogonal similarity transform (Paige, 1970).

The Lanczos approach is applied on the Hessian matrix to iterate and generate a tridiagonal matrix $\boldsymbol{T}$ and an orthogonal matrix $\boldsymbol{Q}$ satisfying the relation $\boldsymbol{T} = \boldsymbol{Q}^{\mathrm{T}}J''\boldsymbol{Q}$ (Golub and Van Loan, 1996). After k steps of Lanczos iterations, we generate a matrix $\boldsymbol{Q}_k = [\boldsymbol{q}_1, \cdots, \boldsymbol{q}_k]$ with orthonormal columns and a tridiagonal matrix

$$\boldsymbol{T}_k = \begin{bmatrix} \alpha_1 & \beta_1 & 0 & 0 \\ \beta_1 & \alpha_2 & \ddots & 0 \\ 0 & \ddots & \ddots & \beta_{k-1} \\ 0 & 0 & \beta_{k-1} & \alpha_k \end{bmatrix}.$$

Equating the columns in $J''\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{T}$, we conclude the following

$$J''\boldsymbol{q}_k = \beta_{k-1}\boldsymbol{q}_{k-1} + \alpha_k\boldsymbol{q}_k + \beta_k\boldsymbol{q}_{k+1}. \tag{B1}$$

The CGA based on Lanczos iterations starts with an initial estimate $\delta\boldsymbol{x}_0$; the initial gradient of the cost function is $\boldsymbol{g}_0 = \nabla J(\delta\boldsymbol{x}_0)$, and the initial descent direction is $\boldsymbol{d}_0 = -\boldsymbol{g}_0$, while the initial Lanczos vector is $\boldsymbol{q}_1 = \boldsymbol{d}_0/\|\boldsymbol{d}_0\|_2$, $\beta_0 = 0, \boldsymbol{q}_0 = 0$. We have the following basic structure for minimizing $J(\delta\boldsymbol{x})$ (Goluband and Van Loan, 1996) for $k = 0, 1, \cdots$:

Step 1. Calculate the multiplication of the matrix and vector at the kth step:

$$\boldsymbol{g}_k = J''\boldsymbol{q}_k, \tag{B2}$$

where the largest calculation lies over the whole iterative algorithm because the Hessian matrix is computed using the tangent linear model $\boldsymbol{L}$ and the adjoint model $\boldsymbol{L}^{\mathrm{T}}$, the calculations of which are very large and time consuming.

Step 2. Estimate the kth diagonal element of the matrix $\boldsymbol{T}$:

$$\alpha_k = \langle \boldsymbol{g}_k, \boldsymbol{q}_k \rangle. \tag{B3}$$

Here, the notation $\langle \cdots, \cdots \rangle$ stands for the inner product.

Step 3. Calculate the residual vector:

$$\boldsymbol{r}_k = \boldsymbol{g}_k - \alpha_k\boldsymbol{q}_k - \beta_{k-1}\boldsymbol{q}_{k-1}. \tag{B4}$$

Step 4. Calculate the $(k+1)^{\text{th}}$ secondary diagonal element of the matrix $\boldsymbol{T}$:

$$\beta_k = \|\boldsymbol{r}_k\|_2, \tag{B5}$$

Step 5. Determine the Lanczos vector $\boldsymbol{q}_{k+1}$ for the next iteration:

$$\boldsymbol{q}_{k+1} = \boldsymbol{r}_k/\beta_k, \tag{B6}$$

which is equivalent to the normalization of the residual vector $\boldsymbol{r}_k$.

Equation (B1) may be written in matrix form as follows:

$$J''\boldsymbol{Q}_k = \boldsymbol{Q}_k\boldsymbol{T}_k + \boldsymbol{r}_k(\mu_k)^{\mathrm{T}}, \tag{B7}$$

where $(\mu_k)^{\mathrm{T}} = (0, \cdots, 0, 1)$. Then, in terms of the quadratic linear equation $\boldsymbol{T}_k\delta\boldsymbol{y}_k = \boldsymbol{Q}_k{}^{\mathrm{T}}\boldsymbol{b}$, which consists of the Lanczos matrix $\boldsymbol{T}_k$, the solution $\delta\boldsymbol{y}_k$ of the kth step is calculated. Further, the kth approximate solution of the minimization [Eq. (1)] is estimated and associated with the Lanczos vectors $(\boldsymbol{q}_1, \cdots, \boldsymbol{q}_k)$.

$$\delta\boldsymbol{x}_k = \delta\boldsymbol{x}_0 + \sum_{i=1}^{k} \langle \boldsymbol{q}_i, \delta\boldsymbol{y}_k \rangle. \tag{B8}$$

In addition, the eigenvalues and eigenvectors of the Lanczos matrix $\boldsymbol{T}_k$ can be estimated during the iteration of the Lanczos approach. The eigenvectors of $\boldsymbol{T}_k$, when pre-multiplied by $\boldsymbol{Q}_k$, approximate the eigenvectors of the Hessian matrix J''. We can use the eigenvectors of the Hessian matrix to estimate the covariance matrix of the analysis errors because the error matrix is equal to the inversion of the Hessian matrix in the variational assimilation (Fisher, 1998). This relation can be used to precondition the CGA and improve the convergence rate.

Acknowledgments. The authors thank the editor and two anonymous reviewers for their valuable comments and suggestions in improving this manuscript.

REFERENCES

Bannister, R. N., 2017: A review of operational methods of variational and ensemble-variational data assimilation. *Quart. J. Roy. Meteor. Soc.*, **143**, 607–633, doi: 10.1002/qj.2982.

Clayton, A. M., A. C. Lorenc, and D. M. Barker, 2013: Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Quart. J. Roy. Meteor. Soc.*, **139**, 1445–1461, doi: 10.1002/qj.2054.

Courtier, P., J. N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387, doi: 10.1002/qj.49712051912.

Courtier, P., E. Andersson, W. Heckley, et al., 1998: The ECM-

WF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quart. J. Roy. Meteor. Soc.*, **124**, 1783–1807, doi: 10.1002/qj.49712455002.

Fisher, M., 1998: Minimization algorithms for variational data assimilation. Annual Seminar on Recent Developments in Numerical Methods for Atmospheric Modelling, Shinfield Park, Reading, 7–11 September 1998, ECMWF, 364–385.

Fletcher, R., and C. M. Reeves, 1964: Function minimization by conjugate gradients. *The Computer Journal*, **7**, 149–154, doi: 10.1093/comjnl/7.2.149.

Golub, G. H., and C. F. Van Loan, 1996: Matrix computations. *Mathematical Gazette*, **47**, 392–396.

Gürol, S., A. T. Weaver, A. M. Moore, et al., 2014: **B**-preconditioned minimization algorithms for variational data assimilation with the dual formulation. *Quart. J. Roy. Meteor. Soc.*, **140**, 539–556, doi: 10.1002/qj.2150.

Han, J., and H. L. Pan, 2006: Sensitivity of hurricane intensity forecast to convective momentum transport parameterization. *Mon. Wea. Rev.*, **134**, 664–674, doi: 10.1175/MWR3090.1.

Isaksen, L., M. Bonavita, R. Buizza, et al., 2010: Ensemble of Data Assimilations at ECMWF. ECMWF Technical Memoranda No. 636, Shinfield Park, Reading, 1–48.

Janisková, M., J. N. Thépaut, and J. F. Geleyn, 1999: Simplified and regular physical parameterizations for incremental four-dimensional variational assimilation. *Mon. Wea. Rev.*, **127**, 26–45, doi: 10.1175/1520-0493(1999)127<0026:SARPPF>2.0.CO;2.

Laroche, S., P. Gauthier, M. Tanguay, et al., 2007: Impact of the different components of 4DVAR on the global forecast system of the meteorological service of Canada. *Mon. Wea. Rev.*, **135**, 2355–2364, doi: 10.1175/MWR3408.1.

Liu, C. S., and Q. N. Xiao, 2013: An ensemble-based four-dimensional variational data assimilation scheme. Part III: Antarctic applications with advanced research WRF using real data. *Mon. Wea. Rev.*, **141**, 2721–2739, doi: 10.1175/MWR-D-12-00130.1.

Liu, D. C., and J. Nocedal, 1989: On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, **45**, 503–528, doi: 10.1007/BF01589116.

Liu, Y. Z., L. Zhang, and Z. Y. Jin, 2017: The optimization of GRAPES global tangent linear model and adjoint model. *J. Appl. Meteor. Sci.*, **28**, 62–71, doi: 10.11898/1001-7313.20170106. (in Chinese)

Lorenc, A. C., S. P. Ballard, R. S. Bell, et al., 2000: The Met. Office global three-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **126**, 2991–3012, doi: 10.1002/qj.49712657002.

Lorenc, A. C., N. E. Bowler, A. M. Clayton, et al., 2015: Comparison of Hybrid-4DEnVar and Hybrid-4DVar data assimilation methods for global NWP. *Mon. Wea. Rev.*, **143**, 212–229, doi: 10.1175/MWR-D-14-00195.1.

Nash, S. G., 1984: Newton-type minimization via the Lanczos method. *SIAM J. Numer. Anal.*, **21**, 770–788, doi: 10.1137/0721052.

Navon, I. M., and D. M. Legler, 1987: Conjugate-gradient methods for large-scale minimization in meteorology. *Mon. Wea. Rev.*, **115**, 1479–1502, doi: 10.1175/1520-0493(1987)115<1479:CGMFLS>2.0.CO;2.

Paige, C. C., 1970: Practical use of the symmetric Lanczos process with re-orthogonalization. *BIT Numerical Mathematics*, **10**, 183–195, doi: 10.1007/BF01936866.

Paige, C. C., and M. A. Saunders, 1982: LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, **8**, 43–71, doi: 10.1145/355984.355989.

Parlett, B. N., and D. S. Scott, 1979: The Lanczos algorithm with selective orthogonalization. *Mathematics of Computation*, **33**, 217–238, doi: 10.1090/S0025-5718-1979-0514820-3.

Rabier, F., 2005: Overview of global data assimilation developments in numerical weather-prediction centres. *Quart. J. Roy. Meteor. Soc.*, **131**, 3215–3233, doi: 10.1256/qj.05.129.

Rabier, F., A. McNally, E. Andersson, et al., 1998: The ECMWF implementation of three-dimensional variational assimilation (3D-Var). II: Structure functions. *Quart. J. Roy. Meteor. Soc.*, **124**, 1809–1829, doi: 10.1002/qj.49712455003.

Rabier, F., H. Järvinen, E. Klinker, et al., 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143–1170, doi: 10.1002/qj.49712656415.

Rawlins, F., S. P. Ballard, K. Bovis, et al., 2007: The Met Office global four-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **133**, 347–362, doi: 10.1002/qj.32.

Shen, X. S., Y. Su, J. L. Hu, et al., 2017: Development and operation transformation of GRAPES global middle range forecast system. *J. Appl. Meteor. Sci.*, **28**, 1–10, doi: 10.11898/1001-7313.20170101. (in Chinese)

Simon, H. D., 1984: The Lanczos algorithm with partial reorthogonalization. *Mathematics of Computation*, **42**, 115–142, doi: 10.1090/S0025-5718-1984-0725988-X.

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Quart. J. Roy. Meteor. Soc.*, **113**, 1311–1328, doi: 10.1002/qj.49711347812.

Thépaut, J. N., R. N. Hoffman, and P. Courtier, 1993: Interactions of dynamics and observations in a four-dimensional variational assimilation. *Mon. Wea. Rev.*, **121**, 3393–3414, doi: 10.1175/1520-0493(1993)121<3393:IODAOI>2.0.CO;2.

Tompkins, A. M., and M. Janisková, 2004: A cloud scheme for data assimilation: Description and initial tests. *Quart. J. Roy. Meteor.*, **130**, 2495–2517, doi: 10.1256/qj.03.162.

Trémolet, Y., 2007: Incremental 4D-Var convergence study. *Tellus A: Dynamic Meteorology and Oceanography*, **59**, 706–718, doi: 10.1111/j.1600-0870.2007.00271.x.

Xue, J. S., S. Y. Zhuang, G. F. Zhu, et al., 2008: Scientific design and preliminary results of three-dimensional variational data assimilation system of GRAPES. *Chinese Science Bulletin*, **53**, 3446–3457, doi: 10.1007/s11434-008-0416-0.

Zhang, L., and Y. Z. Liu, 2017: The preconditioning of minimization algorithm in GRAPES global four-dimensional variational data assimilation system. *J. Appl. Meteor. Sci.*, **28**, 168–176, doi: 10.11898/1001-7313.20170204. (in Chinese)

Zou, X., I. M. Navon, M. Berger, et al., 1993: Numerical experience with limited-memory Quasi-Newton and Truncated Newton methods. *SIAM J. Optim.*, **3**, 582–608, doi: 10.1137/0803029.

Tech & Copy Editor: Hongqun ZHANG