Tellus (2003), 55A, 450–454

TELLUS
ISSN 0280–6495

SHORT CONTRIBUTION

A nonlinear preconditioned quasi-Newton method without inversion of a first-guess covariance matrix in variational analyses

By YOSUKE FUJII* and MASAFUMI KAMACHI, *Oceanographic Research Department, Meteorological Research Institute, Nagamine 1-1, Tsukuba 305-0052, Japan*

(Manuscript received 22 April 2002; in final form 28 April 2003)

ABSTRACT

A quasi-Newton method developed for adopting a non-diagonal first-guess covariance matrix in nonlinear variational analyses (i.e. variational analyses adopting a non-quadratic cost function) is presented. As an example, we also show the effect of the nonlinear relationship between temperature and sea-surface height in three-dimensional variational ocean analyses.

1. Introduction

Variational methods are widely applied in objective data analysis of the atmosphere and the ocean. It is necessary for a cost function J to include the constraint of the first-guess (or background) with the inverse of a first-guess error covariance matrix $\mathbf{B}$ in variational analyses (see Courtier, 1997). In that case,

$$J = \frac{1}{2}\mathbf{x}^T\mathbf{B}^{-1}\mathbf{x} + J_{nl}, \tag{1}$$

where the column vector $\mathbf{x}$ is the increment of the state vector, that is, the state vector minus their first-guess. The second term J_{nl} is a function of $\mathbf{x}$ including constraints imposed by observations or physical conditions. It is needed to perform the tough calculation of the inversion $\mathbf{B}^{-1}$ or to find a way avoiding the inversion in order to implement the minimization.

The variable transform of Lorenc (1988), in which the transformed variables are designed to be uncorrelated, is often adopted in meteorology. This method is, however, difficult to adopt in oceanography because of lateral boundaries, the non-homogeneous character of statistics, etc. In oceanography, therefore, the constraint of the first-guess has often been omitted (e.g. Wenzel et al., 2001) or simplified down to a diagonal matrix (e.g. Morrow and DeMey, 1995). In addition, the method of Derber and Rosati (1989) (hereafter DR89) is widely adopted for avoiding the tough inversion in oceanic three-dimensional variational (3DVAR) analyses. The method can, however, not be adopted directly when nonlinear constraints are included in the cost function (i.e. when the cost function is non-quadratic).

We introduce a new nonlinear preconditioned quasi-Newton method based on the method of DR89 in section 2. The method allows us to adopt a non-diagonal first-guess covariance matrix with nonlinear constraints. One example of an application of the method in an oceanic 3DVAR analysis is also shown in section 3. The results are summarized in section 4.

2. Theoretical background

We developed a method for finding the optimal control variables $\mathbf{x}$ that minimize the cost function J in eq. (1). It is noted that our most interesting case in this

*Corresponding author.
e-mail: yfujii@mrijma.go.jp

paper is the one where J_{nl} is not a quadratic function, that is, $\mathbf{g}(\mathbf{x}) = \nabla_{\mathbf{x}} J$ is not linear.

First, we introduce a positive definite symmetric matrix $\mathbf{U}$ satisfying $\mathbf{U}^{-2} = \mathbf{B}$. We then define $\tilde{\mathbf{x}} = \mathbf{U}\mathbf{x}$ and consider J as a function of $\tilde{\mathbf{x}}$. It is noted that preconditioning is performed here in the same manner as Lorenc (1988) by adopting $\mathbf{B}$ as a preconditioner (Golub and Van Loan, 1996). The cost function is rewritten as

$$J = \frac{1}{2}\tilde{\mathbf{x}}^T\tilde{\mathbf{x}} + J_{nl}. \tag{2}$$

We adopt a quasi-Newton method using the BFGS formula with limited memory (Liu and Nocedal, 1989) as follows:

$$\tilde{\mathbf{x}}_k = \tilde{\mathbf{x}}_{k-1} + \alpha_k \tilde{\mathbf{d}}_{k-1}, \tag{3}$$

$$\tilde{\mathbf{g}}_k = \tilde{\mathbf{g}}(\tilde{\mathbf{x}}_k), \tag{4}$$

$$\tilde{\mathbf{H}}_{k,-m} = \gamma_k \mathbf{I}, \tag{5}$$

$$\tilde{\mathbf{H}}_{k,l} = \tilde{\mathbf{V}}_{k+l}^T \tilde{\mathbf{H}}_{k,l-1} \tilde{\mathbf{V}}_{k+l} + \rho_{k+l} \tilde{\mathbf{p}}_{k+l} \tilde{\mathbf{p}}_{k+l}^T \quad (l = -m+1, -m+2, \ldots, 0), \tag{6}$$

$$\tilde{\mathbf{d}}_k = -\tilde{\mathbf{H}}_{k,0} \tilde{\mathbf{g}}_k, \tag{7}$$

where k is an iteration counter, initially set equal to 0, $\tilde{\mathbf{g}}_k = \tilde{\mathbf{g}}(\tilde{\mathbf{x}}_k)$, where $\tilde{\mathbf{g}}(\tilde{\mathbf{x}}) = \nabla_{\tilde{\mathbf{x}}} J$, and $\tilde{\mathbf{d}}_k$ is known as a search direction. The coefficient α_k is determined to minimize $J(\tilde{\mathbf{x}}_{k-1} + \alpha \tilde{\mathbf{d}}_{k-1})$ approximately. The matrix $\tilde{\mathbf{H}}_{k,l}$ is an approximated inverse matrix of the Hessian calculated from the BFGS formula, $\gamma_k = \tilde{\mathbf{y}}_k^T \tilde{\mathbf{p}}_k / \tilde{\mathbf{y}}_k^T \tilde{\mathbf{y}}_k$, $\rho_k = 1/\tilde{\mathbf{y}}_k^T \tilde{\mathbf{p}}_k$, $\tilde{\mathbf{V}}_k = \mathbf{I} - \rho_k \tilde{\mathbf{y}}_k \tilde{\mathbf{p}}_k^T$, $\tilde{\mathbf{p}}_k = \alpha_k \tilde{\mathbf{d}}_{k-1}$, $\tilde{\mathbf{y}}_k = \tilde{\mathbf{g}}_k - \tilde{\mathbf{g}}_{k-1}$, l is the counter for the BFGS formula, and m is the number of old gradient and search direction vectors adopted for the calculation of $\tilde{\mathbf{H}}_{k,0}$. The initial values are $\tilde{\mathbf{x}}_0 = 0$ and $\tilde{\mathbf{d}}_0 = -\tilde{\mathbf{g}}_0$.

Next, we define $\mathbf{d}_k = \mathbf{U}^{-1}\tilde{\mathbf{d}}_k$, $\mathbf{g}_k = \mathbf{g}(\mathbf{x}_k)$, $\mathbf{h}_k = \mathbf{B}\mathbf{g}_k$, $K_k = \mathbf{x}_k^T \mathbf{B}^{-1} \mathbf{x}_k / 2$, $\mathbf{c}_k = \mathbf{B}^{-1}\mathbf{x}_k$, $\mathbf{e}_k = \mathbf{B}^{-1}\mathbf{d}_k$, $\mathbf{p}_k = \alpha \mathbf{d}_{k-1}$, $\mathbf{y}_k = \mathbf{g}_k - \mathbf{g}_{k-1}$, $\mathbf{z}_k = \mathbf{h}_k - \mathbf{h}_{k-1}$, $\mathbf{V}_k = \mathbf{I} - \rho_k \mathbf{y}_k \mathbf{p}_k^T$, and $\mathbf{H}_{k,l} = \mathbf{U}^{-1}\tilde{\mathbf{H}}_{k,l}\mathbf{U}^{-1}$. Then, the iterative calculation is rewritten as follows:

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \alpha_k \mathbf{d}_{k-1}, \tag{8}$$

$$K_k = K_{k-1} + \alpha_k \mathbf{d}_{k-1}^T \left(\mathbf{c}_{k-1} + \frac{\alpha_k}{2} \mathbf{e}_{k-1} \right), \tag{9}$$

$$\mathbf{c}_k = \mathbf{c}_{k-1} + \alpha_k \mathbf{e}_{k-1}, \tag{10}$$

$$\mathbf{g}_k = \mathbf{c}_k + \mathbf{g}_{nl}(\mathbf{x}_k), \tag{11}$$

$$\mathbf{h}_k = \mathbf{B}\mathbf{g}_k, \tag{12}$$

$$\mathbf{H}_{k,-m} = \gamma_k \mathbf{B}, \tag{13}$$

$$\mathbf{H}_{k,l} = \mathbf{V}_{k+l}^T \mathbf{H}_{k,l-1} \mathbf{V}_{k+l} + \rho_{k+l} \mathbf{p}_{k+l} \mathbf{p}_{k+l}^T \quad (l = -m+1, -m+2, \ldots, 0), \tag{14}$$

$$\mathbf{d}_k = -\mathbf{H}_{k,0} \mathbf{g}_k, \tag{15}$$

$$\mathbf{e}_k = \mathbf{B}^{-1} \mathbf{d}_k. \tag{16}$$

where $\mathbf{g}_{nl}(\mathbf{x}) = \nabla_{\mathbf{x}} J_{nl}$. The coefficients γ_k and ρ_k are now calculated as $\gamma_k = \mathbf{y}_k^T \mathbf{p}_k / \mathbf{y}_k^T \mathbf{z}_k$ and $\rho_k = 1/\mathbf{y}_k^T \mathbf{p}_k$, respectively. The initial values are $\mathbf{x}_0 = 0$, $K_0 = 0$, $\mathbf{c}_0 = 0$, $\mathbf{d}_0 = -\mathbf{h}_0$, and $\mathbf{e}_0 = -\mathbf{g}_0$. In order to search α_k minimizing $J(\mathbf{x}_{k-1} + \alpha \mathbf{d}_{k-1})$, we can adopt the equations

$$J(\mathbf{x}_{k-1} + \alpha \mathbf{d}_{k-1}) = K_{k-1} + \alpha \mathbf{d}_{k-1}^T \times \left(\mathbf{c}_{k-1} + \frac{\alpha}{2} \mathbf{e}_{k-1} \right) + J_{nl}(\mathbf{x}_{k-1} + \alpha \mathbf{d}_{k-1}), \tag{17}$$

$$\mathbf{g}(\mathbf{x}_{k-1} + \alpha \mathbf{d}_{k-1}) = \mathbf{c}_{k-1} + \alpha \mathbf{e}_{k-1} + \mathbf{g}_{nl}(\mathbf{x}_{k-1} + \alpha \mathbf{d}_{k-1}) \tag{18}$$

as the cost function and its gradient.

It should be noted that $\mathbf{x}_k^T \mathbf{B}^{-1} \mathbf{x}_k / 2$ $(= K_k)$ and $\mathbf{B}_{xk}^{-1}$ $(= \mathbf{c}_k)$, which are the first term of J and its gradient in eq. (1), respectively, are calculated by the recursive equations (9) and (10) without the inversion of $\mathbf{B}$ in the iterative calculation. The calculation of $\mathbf{B}^{-1}$ is required only when updating $\mathbf{e}_k$. We can, however, avoid the calculation of $\mathbf{B}^{-1}$ by implementing the calculation of eqs. (13)–(16) as follows:

$$\mathbf{s}_0 = -\mathbf{h}_k, \quad \mathbf{t}_0 = -\mathbf{g}_k, \tag{19}$$

$$\mathbf{s}_{l-1} = \mathbf{s}_l - \rho_{k+l} \mathbf{t}_l^T \mathbf{p}_{k+l} \mathbf{z}_{k+l} \quad (l = 0, -1, \ldots, -m+1), \tag{20}$$

$$\mathbf{t}_{l-1} = \mathbf{t}_l - \rho_{k+l} \mathbf{t}_l^T \mathbf{p}_{k+l} \mathbf{y}_{k+l} \quad (l = 0, -1, \ldots, -m+1), \tag{21}$$

$$\mathbf{s}'_m = \gamma_k \mathbf{s}_m, \quad \mathbf{t}'_m = \gamma_k \mathbf{t}_m, \tag{22}$$

$$\mathbf{s}'_l = \mathbf{s}'_{l-1} + \rho_{k+l} \left(\mathbf{t}_l^T \mathbf{p}_{k+l} - \mathbf{s}'^T_{l-1} \mathbf{y}_{k+l} \right) \mathbf{p}_{k+l} \quad (l = -m+1, -m+2, \ldots, 0), \tag{23}$$

$$\mathbf{t}'_l = \mathbf{t}'_{l-1} + \rho_{k+l} \left(\mathbf{t}_l^T \mathbf{p}_{k+l} - \mathbf{s}'^T_{l-1} \mathbf{y}_{k+l} \right) \mathbf{q}_{k+l} \quad (l = -m+1, -m+2, \ldots, 0), \tag{24}$$

$$\mathbf{d}_k = \mathbf{s}'_0, \quad \mathbf{e}_k = \mathbf{t}'_0, \tag{25}$$

where $\mathbf{q}_k = \mathbf{e}_k - \mathbf{e}_{k-1}$. The calculation of $\mathbf{B}^{-1}$ is completely avoided in this way. It is also noted that the linearity of the gradient $\mathbf{g}$ is not assumed in this derivation. This method can be adopted for nonlinear variational analyses.

3. Example

We examined the effects of the nonlinear relationship between temperature and sea surface height (SSH) in an oceanic 3DVAR analysis. The method of DR89 is adopted for avoiding the inverse calculation in oceanography. The nonlinear relation has often been linearized in those analyses, because the method is available only for a quadratic function. The nonlinearity, however, has an effect that needs to be accounted for, and can be handled with the new quasi-Newton method.

The analyses were performed in the ocean east of Japan, 30–45°N and 140–150°E. A vertical section along 144°E of the analyses at 15 July 1999 is shown with a field observed by the Kofu-maru, the observation vessel of Japan Meteorological Agency (JMA), in the period 16–23 July 1999. SSH observed by TOPEX/POSEIDON (T/P) altimetry (Kuragano and Shibata, 1997) and gridded data of sea-surface temperature (SST) produced by JMA were employed. The analysis method is based on the idea of Fujii and Kamachi (2003). The vertical coupled temperature–salinity empirical orthogonal function (EOF) modes are adopted as control variables. The cost function is as follows:

$$J = \frac{1}{2}\mathbf{x}^T\mathbf{B}^{-1}\mathbf{x} + \frac{1}{2}[H(\mathbf{x}) - \mathbf{x}_o]^T\mathbf{R}^{-1}[H(\mathbf{x}) - \mathbf{x}_o], \tag{26}$$

where $\mathbf{x}_o$ is observed values, $\mathbf{R}$ is an observation error covariance matrix, assumed to be diagonal, and H is an observation function transforming the control variables to the equivalent values of observation. The first-guess error covariance matrix $\mathbf{B}$ is not diagonal because of the horizontal correlation. An approximated Gaussian function (decorrelation scale is 120 km) is adopted for $\mathbf{B}$. Because the SSH observation is employed, the calculation of sea surface dynamic height (SDH) from the grid point values of temperature and salinity is included in H. SDH is calculated by the following nonlinear equation

$$h = -\frac{1}{\rho_s}\int_0^{z_m} \rho'(T, S, p)\,\mathrm{d}z, \tag{27}$$

where h is the difference of SDH from a reference state, ρ_s is the surface density, z_m is the reference depth, ρ' is the difference of the density from the reference state, z denotes the vertical coordinate, and T, S and p denote temperature, salinity and pressure, respectively. The density was calculated according to the nonlinear relation between T, S, p and density by UNESCO (1981). We chose the state of 0 °C and 35 psu as the reference state, and 1000 m as the reference depth. We tried two analyses. One is adopting the exact nonlinear expression of H. The other is adopting the tangent linear operator of H.

The horizontal grid spacing is 0.25° with 19 vertical levels (0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000 m). The monthly climatology of Kuragano and Kamachi (1997) is adopted as the first guess. Observations from the World Ocean Database 1998 (Conkright et al., 1998) (WOD98) are used for calculating statistics. At first, we made a covariance matrix $\mathbf{T}_v$ from the difference between observation and the climatology, and divided it into the first-guess error covariance matrix $\mathbf{B}_v$ and the observation error covariance matrix $\mathbf{R}_v$ as $\mathbf{B}_v = (1 - \eta)\,\mathbf{T}_v$ and $\mathbf{R}_v = \eta\,\mathbf{T}_v$, because the difference between the observation and the first-guess is equal to the sum of first-guess and observation errors. Measurement and representativeness errors are included in the observation error. The representativeness error is necessary because the grid points are too sparse to represent the true state fully in analysis (Cohn, 1997). The parameter η is the ratio of the inexpressible variability to the total variability, and is set equal to 0.2. The EOF modes were calculated from $\mathbf{B}_v$ as the vertical eigen modes of it. The first eight modes are applied. The prior observation error for SSH is set equal to 5 cm. That for SST is picked up from $\mathbf{R}'_v$, where $\mathbf{R}'_v$ is the sum of $\mathbf{R}_v$ and a matrix expressing variability of higher modes not adopted in the analyses. This value includes both the measurement and the representativeness errors.

The results are shown in Fig. 1 to be compared with the observation field in Fig. 2. Especially in warm water, the estimation with the nonlinear observation operator (the nonlinear estimation) is colder than the one with the linear observation operator (the linear estimation), and has smaller errors. The root mean square error of the nonlinear estimation (2.06 °C) is smaller than that of the linear estimation (2.26 °C). This is

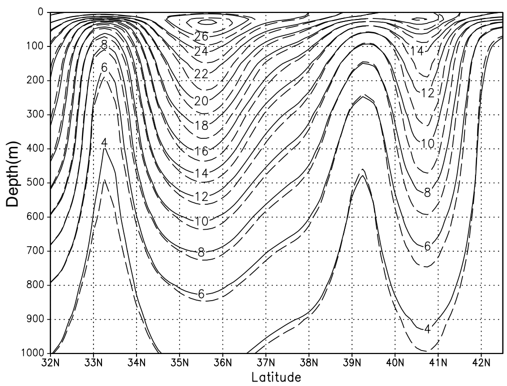

Fig. 1. The analyzed temperature (in °C) fields of the 144°E section. Solid (dashed) lines show the estimation with the nonlinear (linear) operator.

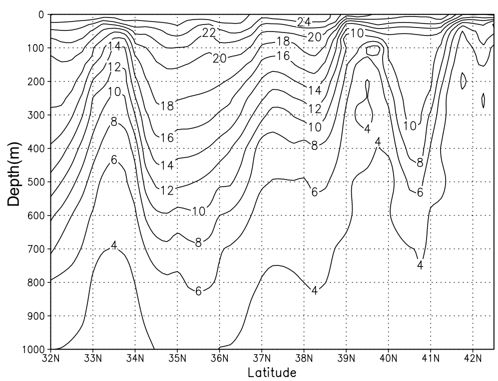

Fig. 2. The observed temperature (in °C) field of the 144°E section.

because the nonlinear operator expresses SDH more precisely than the linear operator. In this region, the first EOF mode represents whether the whole water column is warm or cold. The higher the temperature of a water column is, the more sensitive SDH is to the variation of temperature, which is not expressed by the linear operator (see also Fujii and Kamachi, 2003). The linear estimation therefore overestimates (underestimates) a warm (cold) deviation. It is noted that the linear estimation is more than 1 °C worse in representing the maximum than the nonlinear estimation. This effect is too large to be neglected.

4. Summary

We have introduced a quasi-Newton method developed for the application of a non-diagonal first-guess covariance matrix in nonlinear variational analyses. We showed that the nonlinear relation between the temperature profile and the sea-surface height must be taken into account in an oceanic 3DVAR analysis. The method is useful, especially for oceanographers who want to consider correlations among control variables in variational analyses. The method is to be applied in the new global ocean assimilation system of the Meteorological Research Institute in JMA.

5. Acknowledgments

The observed vertical temperature and salinity profiles along 144°E and the gridded SST data were kindly provided by the Japan Meteorological Agency. WOD98 and the T/P altimeter data were kindly provided by the Japan Oceanographic Data Center and Dr. T. Kuragano, respectively. We thank three reviewers for fruitful discussions. Part of this study was supported by Category 7 of the MEXT RR2002 Project for Sustainable Coexistence of Humanity, Nature and the Earth.

REFERENCES

Cohn, S. E. 1997. An introduction to estimation theory. *J. Meteorol. Soc. Jpn.* **75**, 257–288.

Conkright, M. E., Levitus, S., O'Brien, T., Boyer, T. P., Stephens, C., Johnson, D., Stathoplos, L., Baranova, O., Antonov, J., Gelfeld, R., Burney, J., Rochester, J. and Forgy, C. 1998. World Ocean Database 1998 CD-ROM Data Set Documentation Version 1.2. National Oceanographic Data Center Internal Report 14. NOAA, Washington D.C., 37 pp.

Courtier, P. 1997. Dual formulation of four-dimensional variational assimilation. *Q. J. R. Meteorol. Soc.* **123**, 2449–2461.

Derber, J. and Rosati, A. 1989. A global oceanic data assimilation system. *J. Phys. Oceanogr.* **19**, 1333–1347.

Fujii, Y. and Kamachi, M. 2003. A reconstruction of observed profiles in the sea east of Japan using vertical coupled temperature–salinity EOF modes. *J. Oceanogr.* **59**, 173–186.

Golub, G. H. and Van Loan, C. F. 1996. *Matrix computations*, 3rd edn. The Johns Hopkins University Press, Baltimore, USA, 694 pp.

Kuragano, T. and Shibata, A. 1997. Sea surface dynamic height of the Pacific Ocean derived from

TOPEX/POSEIDON altimeter data: Calculation method and accuracy. *J. Oceanogr.* **53**, 585–599.

Kuragano, T. and Kamachi, M. 1997. Investigation of temperature and salinity fields statistically derived from TOPEX/POSEIDON altimeter datay. In: *Proceedings of the Symposium Monitoring the Ocean in the 2000s. An integrated approach*. Biarritz, France, Session 4, p. 29.

Liu, D. C. and Nocedal, J. 1989. On the limited memory BFGS method for large scale optimization. *Math. Program.* **45**, 503–528.

Lorenc, A. C. 1988. Optimal nonlinear objective analysis. *Q. J. R. Meteorol. Soc.* **114**, 205–240.

Morrow, R. and DeMey, P. 1995. Adjoint assimilation of altimetric, surface drifter, and hydrographic data in a quasi-geostrophic model of Azores current. *J. Geophys. Res.* **100**, 25 007–25 025.

UNESCO 1981. Tenth report of the joint panel on oceanographic tables and standard, Sidney, B. C., September 1980. *UNESCO technical papers in marine science*, vol. 36, 25 pp.

Wenzel, M., Schröter, J. and Olbers, D. 2001. The annual cycle of the global ocean circulation as determined by 4D VAR data assimilation. *Progr. Oceanogr.* **48**, 73–119.