Computational Mathematics and Modeling, pp. 169-178

Nonlinear Optimization Problems in Atmospheric and Oceanic Sciences

Mu Mu, Duan Wansuo and **Wang Jiafeng**

LASG, Institute of Atmospheric Physics,
Chinese Academy of Sciences, Beijing 100029,China
e-mail: mumu@lasg.iap.ac.cn

ABSTRACT

A number of problems, arising from both theoretical research in atmospheric and oceanic sciences and numerical implementation of weather and climate prediction, can be formulated as nonlinear optimization problems (NOPs). Firstly the predictability problems in numerical prediction of weather and climate are investigated, of which three sub-problems are classified, and then reduced into nonlinear optimization problems. How to solve these NOPs to estimate the prediction errors, the maximal predictable time, and the maximal admissible initial errors is demonstrated by the well-known Lorenz model. Secondly the data assimilation problems in atmospheric and oceanic research are considered. Attentions are particularly paid to the variational data assimilation with " on- off " switch conditions in the model physical processes, which usually yields in non-smooth NOPs. Thirdly the NOPs related to the sensitivity analysis in atmospheric and oceanic studies are also explored. The difficulties both in the theory and in the lack of ripe algorithms are also presented, which leaves future works to both computational mathematicians and scientists in geophysics.

KEY WORDS – weather, climate, optimization, predictability, data assimilation, sensitivity analysis.

1. Introduction

In theoretical research of atmospheric and oceanic sciences and numerical implementation of weather and climate prediction, many problems can be formulated as nonlinear optimization problems (NOPs). This paper is devoted to discussion of the NOPs related to predictability, variational data assimilation and sensitivity analysis.

Numerical prediction of weather and climate consists of solving a set of partial differential equations, which is usually refered as model, with given initial values. But inevitable model deficiencies and initial errors will cause the uncertainties of the forecast results. The studies of these uncertainties have become known as "predictability problems", which can be traced back to Thompson (1957) and the latter Lorenz (1963). Untill now, the predictability problems of numerical weather and climate prediction is still one of the important subjects, e.g., the well-known international research programme on "Climate Variability".

According to above two factors causing the uncertainties of the forecast results, the predictability problems are usually classified into two types: the first kind of predictability, which is related to the initial errors, and the second kind of predictability, which is to the model errors. The definition of the model error varies with authors. In this paper, we adopt the following definition (Talagrand, 1997): If the initial value of the model is the true state, then the difference between the value of the forecast and the true state at the prediction time is called the model error. From this definition, it is easily seen that there are many factors causing model errors. However, in this paper, we only consider the errors of the parameters in the model, which is generally considered to be the main problem in the model.

Obviously, the more accurate the estimate of the initial conditions, the better the quality of the forecasts. Currently, operational numerical weather prediction centers produce initial conditions through a combination of observations and short-range forecasts. This approach is entitled "data assimilation", whose purpose is to determine as accurately as possible the state of the atmospheric (or oceanic) flow by using all the available information (Talagrand, 1997).

Data assimilation consists of two types of approaches: sequential data assimilation and variational data assimilation. In this paper, our attention is paid to four-dimensional variational data assimilation, known as '4D-VAR'. It starts by consructing a cost function with respect to initial condition, which represents the difference between the model solution and the observational data. By seeking the minimum value for the cost function, the optimal initial field is obtained. This problem can be regarded as unconstrained optimization problem mathmetically. However, it is not easy to compute the gradient of the cost function with high dimensional models. In Le Dimet and Talagrand (1986), the adjoint method was proposed, which is an efficient way to catch the gradient. However, to apply adjoint method correctly and effectively in the 4D-VAR with physical processes, there are still a few important problems to be worked out, and the "on-off" switch problem is one of them.

The present study aims to address the above problems. We will, in section 2, derive the three sub-problems of the predictability, and adopt the well-known Lorenz model to show how to realize the idea in solving the three sub-problems. In section 3, the non-smooth NOPs in the variational data assimilation with on-off switch condition are explored. Section 4 is devoted to the NOPs related to the sensitivity analysis. In the last section, the difficulties in theory and in the optimization process are discussed.

2. Predictability problems

2.1 Three problems of predictability

On the basis of practical demands, the predictability problems can be classified into three problems (Mu, et al, 2002).

Problem 1 Assume that the initial observation $\vec{u}_0^{obs}$ and the first given value of the parameter $\vec{\mu}^g$ in the numerical model are known. At prediction time T, the maximum allowable prediction error in terms of norm $\|\cdot\|_A$ is

$$\| M_T\left(\vec{u}_0^{obs}, \vec{\mu}^g\right) - \vec{u}_T^t \|_A \le \varepsilon \tag{2.1}$$

where M_T is the propagator, which propagates the state from the initial time to time T, $\vec{u}_T^t$ the true value of the state at time T, and ε is given. So the maximum predictable time T_ε can be formulated into the following nonlinear optimization problem.

$$T_\varepsilon = \max\{\tau \mid \| M_t(\vec{u}_0^{obs}, \vec{\mu}^g) - \vec{u}_t^t \|_A \le \varepsilon, \quad 0 \le t \le \tau\}. \tag{2.2}$$

Since the true value $\vec{u}_t^t$ cannot be obtained exactly, T_ε cannot be derived precisely too. But if we know more information on the errors of initial value and parameters, the useful estimation on T_ε can be derived. For example, if we know the information on the errors of the initial value and the parameter as follows

$$\| \vec{u}_0^t - \vec{u}_0^{obs} \|_A \le \delta_1, \ \| \vec{\mu}^t - \vec{\mu}^g \|_B \le \delta_2. \tag{2.3}$$

Then we investigate the following nonlinear optimization problem

$$T_d = \min_{\vec{u}_0 \in B_{\delta_1}, \vec{\mu} \in B_{\delta_2}} \left\{ T_{\vec{u}_o, \vec{\mu}} \mid T_{\vec{u}_o, \vec{\mu}} = \max \tau, \| M_t(\vec{u}_0, \vec{\mu}) - M_t(\vec{u}_0^{obs}, \vec{\mu}^g) \| \le \varepsilon, 0 \le t \le \tau \right\},$$

where B_{δ_1}, B_{δ_2} are the balls with centers at $\vec{u}_0^{obs}$, $\vec{\mu}^g$ and radius δ_1, δ_2 respectively.

It is not difficult to prove that

$$T_{\varepsilon l} \le T_\varepsilon .$$

Therefore we have the lower bound of the maximum predictable time.

Problem 2 Assume $\vec{u}_0^{obs}$ and $\vec{\mu}^g$ are given, for a fixed prediction time T, the prediction error can be expressed as follows

$$E = \| M_T(\vec{u}_0^{obs}, \vec{\mu}^g) - \vec{u}_T^t \|_{A.}$$

Similar to the above problem, since the true value $\vec{u}_T^t$ cannot be known precisely, it is also impossible to get the exact value of *E*. But we can estimate it by the information on the errors about initial value and parameters as follows.

$$E_u = \max_{\vec{u}_0 \in B_{\delta_1}, \vec{\mu} \in B_{\delta_2}} \| M_T(\vec{u}_0, \vec{\mu}) - M_T(\vec{u}_0^{obs}, \vec{\mu}^g) \|_A .$$

Without much difficulty, we can prove that

$$E \le E_u .$$

In this way, we established the upper bounds on the prediction errors.

Problem 3 Assuming that the initial observation $\vec{u}_0^{obs}$ and the first given value of the parameter $\vec{\mu}^g$ are available. At the prediction time T**,** the allowing maximum prediction error is (2.1). Our purpose is to determine the allowable maximum initial error and the parameter error. More precisely, the purpose is to look for the maximum δ, such that if (2.3) holds with $\delta = \delta_1 + \delta_2$, then (2.1) holds.

This problem can also be reduced to an optimization problem as follows,

$$\delta_{\max} = \max_{\delta} \{ \delta \mid \| \vec{u}_0^t - \vec{u}_0^{obs} \|_A \le \delta_1, \| \vec{\mu}^g - \vec{\mu}^t \|_B \le \delta_2,$$

$$\text{if } \delta_1 + \delta_2 = \delta, \text{ then } \| M_T(\vec{u}_0^{obs}, \vec{\mu}^g) - \vec{u}_T^t \|_A \le \varepsilon \},$$

Following the above idea, we can estimate δ. Investigating the problem

$$\bar{\delta}_{\max} = \max_{\delta} \{ \delta \mid \| M_T(\vec{u}_0^{obs}, \vec{\mu}^g) - M_T(\vec{u}_0, \vec{\mu}) \|_A \le \varepsilon, \vec{u}_0 \in B_{\delta_1}, \vec{\mu} \in B_{\delta_2}, \delta_1 + \delta_2 = \delta \},$$

we can conclude that

$$\bar{\delta}_{\max} \le \delta_{\max} .$$

2.2 Examples — Three problems of Lorenz model

In this sub-section, we study the three problems of the predictability of the Lorenz model as examples. For simplicity, we assume that the model is perfect.

Lorenz's model consists of a set of ordinary differential equations:

$$\begin{cases} \dfrac{dx}{dt} = -\sigma x + \sigma y, \\ \dfrac{dy}{dt} = -xz + rx - y, \\ \dfrac{dz}{dt} = xy - bz, \\ (x, y, z)|_{t=0} = (x_0, y_0, z_0), \end{cases}$$

where $\sigma = 10,\ r = 28,\ b = 8/3$ (Lorenz, 1965) are the parameters?

It is easy to find out that Lorenz's model has three stationary points:

$$\begin{cases} O : (x, y, z) = (0,0,0), \\ C_1 : (x, y, z) = (-\sqrt{b(r-1)}, -\sqrt{b(r-1)}, r-1), \\ C_2 : (x, y, z) = (\sqrt{b(r-1)}, \sqrt{b(r-1)}, r-1). \end{cases}$$

We choose these three stationary points as initial observations to study the three problems. It should be pointed out that other points could be adopted as the initial observations too. The system is integrated by middle point scheme with time step $dt = 0.01$.

2.2.1 Problem 1

Let M_t be the propagator of the Lorenz model, O^* the initial observation, $\vec{X} = (x_0, y_0, z_0)$, and $\vec{u}_0 = O^* + \vec{X}$. With these notations, the lower bound of the maximum predictable time is

$$T_{\varepsilon l} = \min_{\|\vec{X}\| \le \delta} \{ T_{\vec{X}} \mid T_{\vec{X}} = \max \tau, \| M_t(O^* + \vec{X}) - M_t(O^*) \| \le \varepsilon, 0 \le t \le \tau \}.$$

In this paper, two norms $\|\vec{X}\|_2 = \sqrt{x_0^2 + y_0^2 + z_0^2}$ and $\|\vec{X}\|_\infty = \max\{|x_0|, |y_0|, |z_0|\}$ are employed to measure the errors. For different initial observational error bounds $\|\vec{X}\|_2 \le \delta$ and $\|\vec{X}\|_\infty \le \delta$, we obtain the lower bound of the maximum predictable time for the initialobservations O and C_1. Because C_1 and C_2 are z-axis symmetrical, the lower bound of the maximum predictable time should be equivalent. For simiplicity, we only show the results of O and C_1.

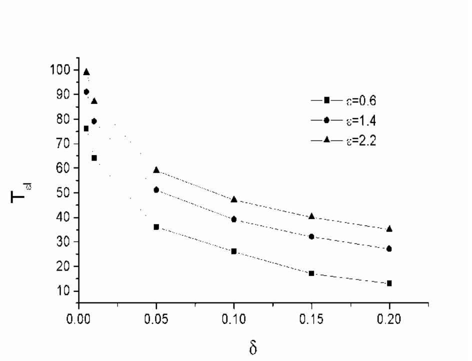

Fig 1. $T_{\varepsilon l}$ *for* O *with norm* $\|\cdot\|_\infty$

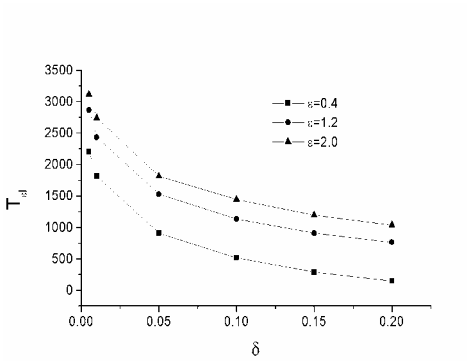

Fig 2. $T_{\varepsilon l}$ *for* C_1 *with norm* $\|\cdot\|_\infty$

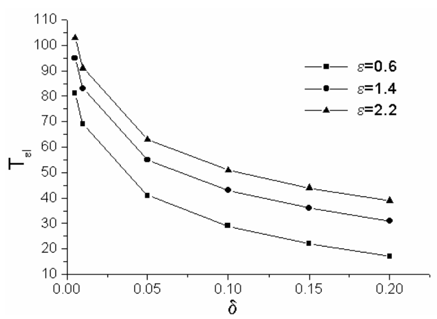

Fig 3. $T_{\varepsilon l}$ *for* O *with norm* $\|\cdot\|_2$

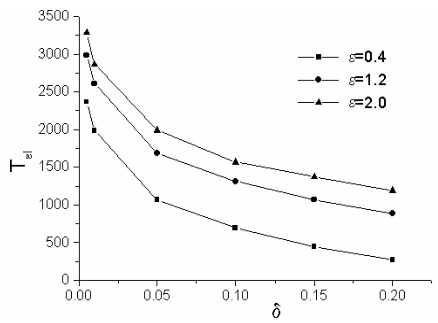

Fig 4. $T_{\varepsilon l}$ *for* C_1 *with norm* $\|\cdot\|_2$

It is clear from Figures 1-4 that the lower bound of the maximum predictable time of the initial observation C_1 is much longer than the corresponding one of the initial observation O. This indicates that there exists stronger predictability around initial observation C_1 than initial observation O.

2.2.2 Problem 2

With the above notations, the upper bound of prediction error of the initial observation O^* at time T is

$$E_u = \max_{\|\vec{X}\| \le \delta} \| M_T(O^* + \vec{X}) - M_T(O^*) \| .$$

The numerical results related to the maximum prediction error are presented in the following figures.

Initial observation O

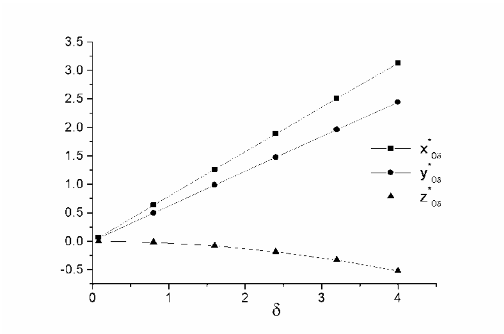

Fig 5. $\vec{X}_\delta^*$ *for T=30 with* $\|\vec{X}\|_2$

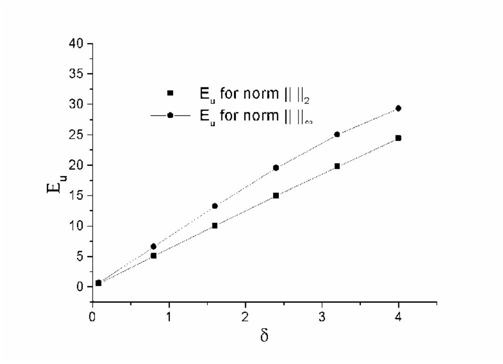

Fig6. E_u *for T=30 with* $\|\vec{X}\|_2, \|\vec{X}\|_\infty$

Initial observation C_1

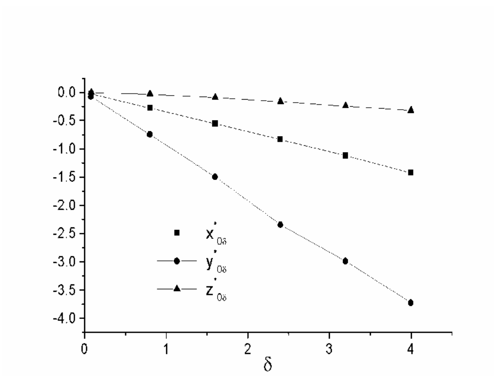

Fig 7. $\vec{X}_\delta^*$ *for T=80 with* $\|\vec{X}\|_2$

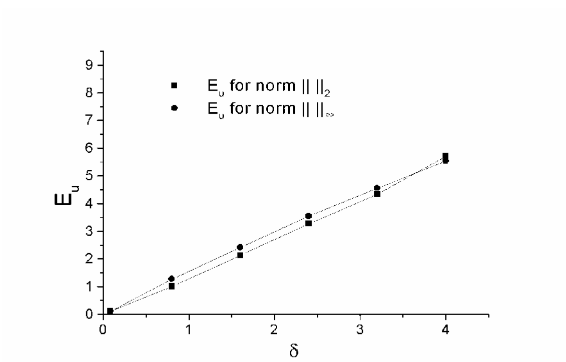

Fig8. E_u *for T=80 with* $\|\vec{X}\|_2, \|\vec{X}\|_\infty$

2.2.3 Problem 3

From section 2.1, we can derive the lower bound of the allowable maximum initial error, i.e.

$$\overline{\delta}_{\max} = \max_{\delta}\{\delta \mid \| M_T(O^* + \vec{X}) - M_T(O^*) \| \le \varepsilon, \|\vec{X}\| \le \delta\}.$$

The numerical results are as follows.

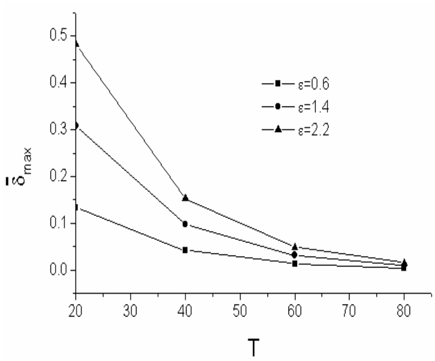

Fig 9. $\bar{\delta}_{\max}$ *for* O *with* $\|\cdot\|_\infty$

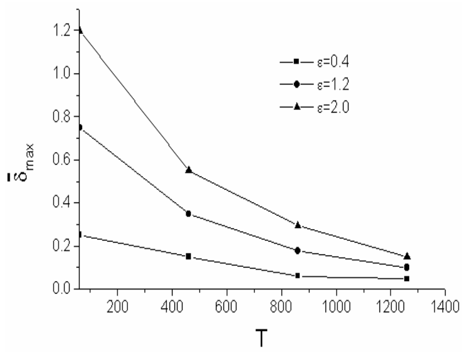

Fig10. $\bar{\delta}_{\max}$ *for* C_1 *with* $\|\cdot\|_\infty$

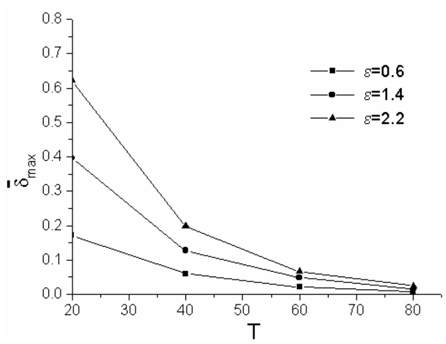

Fig 11. $\bar{\delta}_{\max}$ *for* O *with* $\|\cdot\|_2$

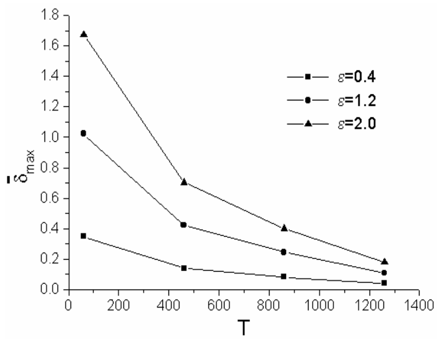

Fig12. $\bar{\delta}_{\max}$ *for* C_1 *with* $\|\cdot\|_2$

3. The variational data assimilation problems with on-off switch condition

3.1 The problem

The variational data assimilation (VDA) is an unconstraint optimization problem mathematically. In optimization process we need to calculate the gradient of the cost function with respect to the initial condition. The adjoint method is an efficient approach to obtain it.

But for the model with discontinuity resulted from the presence of the "on-off" switches**,** the computational efficiency and even the feasibility of the adjoint method used in variational assimilation should be checked. This section is devoted to this problem by using two kinds of humidity equation, which contain threshold processes (represented by "on-off" switches in models).

Model 1 (Xu 1996; Zou 1997):

$$\begin{cases} \dfrac{dq}{dt} = F + \beta H_+(q - q_c), & 0 \le t \le T; \\ q\,|_{t=0} = q_0 \end{cases}$$

Model 2 (Xu 1996; Zou 1997):

$$\begin{cases} \dfrac{dq}{dt} = F + H_+(q - q_c)\alpha(q - q_c), & 0 \le t \le T, \\ q\,|_{t=0} = q_0 \end{cases}$$

where q is the state variable, q_c the critical value of q, the switch is turned on or off, and $H_+(x) = \begin{cases} 1 & x \ge 0 \\ 0 & x < 0 \end{cases}$ is the Heaviside unit step function.

According to the characters of the switch term (represented by $H_+(\cdot)$ in the equation), the "on-off" problems can be divided into two types: discontinuous (Type I) and continuous but not differentiable (Type II).

For these two types of problems, Mu and Wang (2002) recently got the following results.

For Type I switch: 1) In the case of time continuous model, the reason that the conventional approach, which ignores the variation at the switch point, fails to give a correct gradient is that the propagator of the conventional tangent linear model in this approach differs from the tangent linear operator of the propagator of the nonlinear forward model, in spite of the fact that the gradient of the cost function with respect to the initial condition exists except for the threshold. 2) In the case of time discrete model obtained by traditional schemes, the conventional approach can always be used to gain the correct gradient where it exists. But there might exist local minimum for the discrete cost function, which do not exist in the continuous case, and the optimization with the gradient obtained by the conventional approach could yield the local minimum rather than the global minimum as we require. 3) A new approach, which is based on the nonlinear perturbation equation directly, rather than on the conventional tangent linear model, is proposed. This approach can be used to obtain the accurate gradient in the continuous case. 4) In the discrete case, this new approach can be employed to obtain correct descent direction of the cost function, which yields the global minimum, and overcomes the drawback of the conventional approach. Further, the advantage of this new approach is that it is not necessary to modify the forward model, or the tangent linear and adjoint models. For Type II switch, the conventional approach can be used correctly both for the time continuous and discrete cases.

In the case of time discrete model obtained by traditional schemes, for the discontinuous switch, the cost function becomes piecewise and may have multiple minima. So the result of assimilation with convectional approach is sensitive to the first guess. This phenomenon can be elminated by our new approach. For the Type II switch, the conventional approach is feasible.

3.2 Numerical results on Type I switch

The numerical experiments based on the optimization with the conventional approach (Zou et al., 1993; Zou, 1997) (Algorithm 1) and the new one of Mu and Wang (Algorithm 2) are made. For the discrete cost function , it is found that there are multiple minima in the case of Type I switch.

In the following experiment, the continuous cost function has unique minimum $q_0^* = 0.3$, while the minima of the discrete cost function are: $q_0^{(1)} = 0.19$, $q_0^{(2)} = 0.25$. The first is a local minimum and the second is the global one. The global minimum is close to the analytical global minimum $q_0^* = 0.3$ but does not coincide with it due to model error.

Using different initial guess values, we compare the effectiveness of Algorithms 1 and 2.

Experiment 1: The first guess is 0.150, which is at the left of the local and global minimums.

Table 13. *Results obtained by Algorithm 1*

k	$q_0^{(k)}$	$J_d^{'}(q_0^{(k)})$	$J^{'}(q_0^{(k)})$	$J_d(q_0^{(k)})$
0	1.500000E-01	-4.134046E-2	-4.628400E-2	3.181858E-03
1	1.653934E-01	-2.579297E-2	-4.141818E-2	2.665143E-03
2	1.909308E-01	-1.172721E-7	-3.340439E-2	2.335806E-03

Table 14. *Results obtained by Algorithm 2*

k	$q_0^{(k)}$	$J_d^{'}(q_0^{(k)})$	$J^{'}(q_0^{(k)})$	$J_d(q_0^{(k)})$
0	1.500000E-01	-4.247610E-2	-4.628400E-2	3.181858E-03
1	1.649819E-01	-3.585407E-2	-4.154791E-2	2.675852E-03
2	2.460991E-01	-1.230505E-2	-1.634159E-2	5.263018E-04
3	2.497602E-01	-1.071611E-2	-1.522132E-2	5.261876E-04
4	2.497611E-01	-1.071578E-2	-1.522105E-2	5.261853E-04

Experiment 2. The first guess is 0.21, which is between the local and global minimal values. In Alogorithm 1, $q_0^{(k)}$ converges to 0.19, which is the local minimum. The cost function varies from 2.5E-03 to 2.3E-03. In Algorithm 2, the final result is 0.25, which is the global minimum. The cost function varies from 2.5E-03 to 5.4E-04. At the beginning of the iteration, the derivative obtained by the conventioal approach has different sign with analytical one, which explains why the optimization process yields the local minimum.

Experiment 3. The first guess is 0.31, at the right to the local and global minimal values. In Algorithm 1, $q_0^{(k)}$ comes through 0.29 to 0.25, the global minimum. Algorithm 2 ends with 0.25, which is also the global minimum.

4. Sensisivity analysis of the model with respect to initial condition

In this section we discuss how to make sensitivity analysis of the model with respect to initial condition by assuming that the model error can be ignored.

Let

$$E = \min J(U_0) = \min \frac{1}{2} \| M_T(U_0) - U_T^{obs}) \|^2 , \tag{4.1}$$

where M_T is the propagator (the numerical model), U_0 the initial value and U_T^{obs} the observational value at time T. By using the adjoint method and the optimization algorithms, the initial value U_0 and the minimum E satisfied (4.1) can be found. For a given error bound ε, there are two cases for E:

$$\begin{cases} E > \varepsilon & (1) \\ 0 \le E \le \varepsilon & (2). \end{cases}$$

If case (1) appears, it is easily shown that E is greater than the given allowable error, which means that the numerical model cannot predict the observational value U_T^{obs} within the allowable error even if we have the optimal initial value. In other words, no matter how we adjust the initial value, the numerical model can not yield a satisfactory forecast results. In this case, we know that there exists considerable model error, and the model needs to be improved. In case (2), E is small enough . The numerical model solution and the observational value U_T^{obs} have no essential difference. It implies that the forecast results of the numerical model are satisfactory. That is to say, if we adjust the initial value of the numerical model U_0, the forecast result can be improved.

Let us further consider case (2), defining a maximum allowable initial error ε_0 , we investigate the difference between the optimal initial value U_0 and the observational value U_0^{obs} . There are three cases:

$$\begin{cases} \| U_0 - U_0^{obs} \|^2 << \varepsilon_0 & \text{(a)} \\ \| U_0 - U_0^{obs} \|^2 \approx \varepsilon_0 & \text{(b)} \\ \| U_0 - U_0^{obs} \|^2 >> \varepsilon_0 & \text{(c)} \end{cases}.$$

In (a), it is indicated that the numerical model has sufficient ability to forecast the corresponding weather or climate process. The satisfactory forecast results can be obtained from the existing observational data. So we don't need to treat particularly with the initial field of the model. However, for (b), it is different from (a). Although the numerical model has the ability to forecast the corresponding weather or climate process, the satisfactory prediction results cannot be obtained with U_0^{obs} . Therefore we must improve the initial field by some methods (for example, assimilation method). From the improved initial field, the satisfactory prediction results can be obtained too. In the case of (c), similar to (b), the numerical model has the ability to forecast the corresponding weather or climate process. But the satifactory forecast results cannot be obtained from the existing observational data. This can be due to the lack of the information on the initial field. The existing observational data cannot reprsent the weather and climate processes. It is needed to intensify the observational network and obtain more detailed observational data than the existing one. Of course in the case of (1), it is also possible that U_0 is too far away from U_0^{obs} to be true, and we should also investigate the model error.

5. Discussion

In this paper, we classify the predictability problems of numerical weather and climate prediction into three problems on the basis of practical demands, investigate the variational data assimilation problems with "on-off" switch condition, and discuss the sensitivity analysis of the model with respect to initial conditions. All these problems can be formulated as nonlinear optimization problems. In operational numerical weather and climate prediction, the models are often very complicated and of high dimensions. To solve the above nonlinear optimization problems with such high dimensions, the capacity of the existing computers (speed, memory, etc) must be taken into account. In addition, the models governing the motions of atmosphere and oceans are generally nonlinear. In some cases, the problems are non-smooth with complex constraint conditions.

With the development of the economy and society, quantifying study becomes more and more important in atmosphere and ocean research. It is expected that the rapid development in computational mathematics and in computers will serve our purpose in the future not too far, and it is time to devote our energies to these NOPs.

Acknowledgments. This work was supported by the National Nature Scientific Foundation of China (No. 40023001 and 40075015), KZCX2-208 of the Chinese Academy of Sciences, and LIAMA.

Reference

[1] P.D., Thompson, "Uncertainty of initial state as a factor in the predictability of large-scale atmospheric flow patterns", *Tellus*, Vol. 9, 1957. pp. 275-295.

[2] E.N. Lorenz, "Deterministic Nonperiodic Flow", *J. Atmos. Sci.,* Vol. 20, 1963. pp. 78-89.

[3] O. Talagrand, "Assimilation of Observation, an Introduction", *J. Meteor, Soc. Japan*, Vol. 1B, 1997. pp.191-209.

[4] F.X. Ledimet and O.Talagrand, "Varaiational algorithm for analysis and assimilation of meteorological Observations: Theoretical aspects", *Tellus*, Vol. 38A, 1986. pp. 97-110."

[5] M. Mu, W.S. Duan and J.C. Wang, "The Predictability Problem in Numerical Weather and Climate Prediction", *Adv. Atmos. Sci.,* Vol. 19, 2002. pp. 191-204.

[6] Q. Xu, "Generalized Adjoint for Physical Processes with Parameterized Discontinuities. Part I: Basic Issues and Heuristic Examples", *J. Atmos. Sci.,* Vol. 53, 1996. pp. 1123-1142.

[7] X. Zou, I.M., Navon, and J.G., Sela, "Variational Data Assimilation with Moist Threshold Processes Using The NMC Spectral Model", *Tellus*, Vol. 45A, 1993. pp. 370-387.

[8] X. Zou, "Tangent Linear and Adjoint of 'on-off' Processes and Their Feasibility for use in 4-Dimensional Variational Data Assimilation", *Tellus,* Vol. 49A, 1997. pp. 3-31.

[9] M. Mu and J.F. Wang, "An Adjoint Approach for Variational Data Assimilation with Physical 'on-off' Processes", preprint, 2002.