Tellus (1998), 50A, 657–664
Printed in UK – all rights reserved

Copyright © Munksgaard, 1998

TELLUS
ISSN 0280–6495

Reply to "Comments on Tangent linear and adjoint of 'on–off' processes and their feasibility for use in 4-dimensional variational data assimilation"

By X. ZOU, *Department of Meteorology, The Florida State University, 404 Love Building, Tallahassee, FL32306–4520, USA*

(Manuscript received 5 November 1997; in final form 15 June 1998)

1. Introduction

In 4-dimensional variational data assimilation (4DVAR), a term J_o, which measures the distance between the model solution and observations over a time window $[t_0, T]$, is included in the cost function (J) along with the background term (J_b) and other constraints (J_c), i.e., $J = J_b + J_o + J_c$ (Fischer and Courtier, 1995). Once the 4DVAR problem (J) is defined, a standard unconstrained minimization algorithm is employed to find the minimum (Zou et al., 1993a). In doing so, users of the minimization software are required to supply the values of both the cost function J and the gradient of the cost function with respect to the model initial condition (IC), ∇J. The adjoint of the numerical assimilation model is used to derive the value of the gradient of J_o with respect to the model IC, ∇J_o. Both the paper by Xu (1996, hereafter Xu96) and the paper by Zou (1997, hereafter Zou97) attempt to address problems related to the calculation of ∇J_o using an assimilation model and its adjoint having on–off switches. The former dealt with analytical solution in time and the later is primarily concerned with time-discretized models. Two issues are of major concern: (i) the correctness of the gradient derived from the "classic" adjoint model which keeps on–off switches the same as in the nonlinear model (NLM), and (ii) The continuity of the gradient in the presence of on–off switches.

It was Xu's opinion that "there are significant discrepancies in conclusions and understandings" between Xu96 and Zou97, and he made some specific comments to which my reply follows.

2. Reply to the first comment: Should on–off switches cause discontinuities in the gradient? What kind of generalization is really needed?

(a) The main reason Xu96 derived a different conclusion from that of Zou97 is that his conclusion is not directly relevant to time-discretized models, which Zou97 was addressing. For a 4-dimensional numerical model, the action of switching on and off can occur either at the initial time $t_0 = 0$ or later in time $t = \tau$ when the initial state or model prediction at time τ passes a certain threshold value. If the switch occurs at the initial time t_0 at the initial point x_c, it is agreed by Xu that the gradient is discontinuous at x_c. But Xu disagrees that the gradient can still be discontinuous at $x_0 \neq x_c$ when the switch occurs at a later time within the assimilation window.

Xu used the following example

$$\frac{dz}{dt} = F + GH + (x - x_c), \tag{1a}$$

$$x|_{t=0} = x_0, \tag{1b}$$

to illustrate that "when the initial state varies over the range of $x_a \leqslant x_0 < x_c$, the solution contains an on switch within the data assimilation window ($0 < t \leqslant T$), but the gradient of the cost function is continuous over the range of $x_a \leqslant x_0 < x_c$." In (1), $H_+(x-x_c)$ is defined as

$$H_+(x-x_c) = \begin{cases} 1, & \text{if } x \geqslant x_c, \\ 0, & \text{if } x < x_c. \end{cases}$$

We will use the same example (1) to show that for a discretized model of (1), the gradient of J can become discontinuous at some points of ICs x_0, where $x_0 < x_c$ and the length of the assimilation window $T > (x_c - x_0)/F$, i.e., the switch is turned on later than $t_0 = 0$ during the assimilation window.

Let's define the cost function:

$$J(x_0) = \frac{1}{2}\int_{t_0}^{T} (x - x^{\mathrm{obs}})^2 \, \mathrm{d}t. \tag{2}$$

The gradient ∇J can be expressed as (see eq. (4.7) in Zou97):

$$f(x_0) \equiv \nabla J = \int_{t_0}^{T} (x(t) - x^{\mathrm{obs}}(t))\, \mathrm{d}t + \int_{\tau}^{T} (x(t) - x^{\mathrm{obs}}(t)) \frac{G}{F}\, \mathrm{d}t, \tag{3}$$

where $\tau = (x_c - x_0)/F$, $x(t)$ satisfies the eq. (1), $x_0 \leqslant x_c$ and $t_0 \leqslant \tau \leqslant T$.

For a time-discretized model, the time interval $[0, T]$ is divided into a finite number of intervals depending on the model's spatial resolution and stability condition. Assuming that $\tau = t_{i_\tau}$, eq. (3) can be written as

$$f(x_0) \equiv \nabla J = \sum_{i=0}^{R} (x(t_i) - x^{\mathrm{obs}}(t))\, \Delta t + \sum_{i=i_\tau}^{R} (x(t_i) - x^{\mathrm{obs}}(t_i)) \frac{G}{F}\, \Delta t, \tag{4}$$

where R is the total number of time levels on which observations are available.

For any small perturbation to the IC x_0, Δx_0 (abs$(\Delta x_0) <$ F Δt), if $\Delta x_0 > 0$, we have

$$f(x_0 + \Delta^+ x_0) = \sum_{i=0}^{R} (x^{ptb^+}(t_i) - x^{\mathrm{obs}}(t_i))\, \Delta t + \sum_{i=i_\tau}^{R} (x^{ptb^+}(t_i) - x^{\mathrm{obs}}(t_i)) \frac{G}{F}\, \Delta t, \tag{5}$$

and if $\Delta x_0 < 0$ we have

$$f(x_0 + \Delta^- x_0) = \sum_{i=0}^{R} (x^{ptb^-}(t_i) - x^{\mathrm{obs}}(t_i))\, \Delta t + \sum_{i=i_\tau+1}^{R} (x^{ptb^-}(t_i) - x^{\mathrm{obs}}(t_i)) \frac{G}{F}\, \Delta t, \tag{6}$$

where $x^{pbt^\pm}$ satisfies the same equation (1a) of the nonlinear forward model with a perturbed IC: $x^{pbt^\pm}|_{t=0} = x_0 \pm \Delta x_0$.

From (4)–(6), we obtain

$$\begin{aligned} \lim_{\Delta x_0 \to 0^-} \frac{f(x_0+\Delta x_0) - f(x_0)}{\Delta x_0} &= \sum_{i=0}^{R} \lim_{\Delta x_0 \to 0^-} \frac{x^{ptb^-}(t_i) - x(t_i)}{\Delta x_0} \Delta t + \sum_{i=i_\tau+1}^{R} \lim_{\Delta x_0 \to 0^-} \frac{x^{ptb^-}(t_i) - x(t_i)}{\Delta x_0} \frac{G}{F} \Delta t - \lim_{\Delta x_0 \to 0^-} \frac{x(t_{i_\tau}) - x^{\mathrm{obs}}(t_{i_\tau})}{\Delta x_0} \frac{G}{F} \Delta t \\ &= \sum_{i=0}^{R} \left(1 + \frac{G}{H} H_+(t_i - \tau)\right) \Delta t + \sum_{i=i_\tau+1}^{R} \left(1 + \frac{G}{F} H_+(t_i + \tau)\right) \frac{G}{F} \Delta t - \lim_{\Delta x_0 \to 0^-} \frac{x(t_{i_\tau}) - x^{\mathrm{obs}}(t_{i_\tau})}{\Delta x_0} \frac{G}{F} \Delta t \end{aligned} \tag{7a}$$

$$\begin{aligned} \lim_{\Delta x_0 \to 0^+} \frac{f(x_0+\Delta x_0) - f(x_0)}{\Delta x_0} &= \sum_{i=0}^{R} \lim_{\Delta x_0 \to 0^+} \frac{x_{ptb^+}(t_i) - x(t_i)}{\Delta x_0} \Delta t + \sum_{i=i_\tau}^{R} \lim_{\Delta x_0 \to 0^+} \frac{x^{ptb^+}(t_i) - x(t_i)}{\Delta x_0} \frac{G}{F} \Delta t \\ &= \sum_{i=0}^{R} \left(1 + \frac{G}{F} H_+(t_i - \tau)\right) \Delta t + \sum_{i=i_\tau}^{R} \left(1 + \frac{G}{F} H_+(t_i - \tau)\right) \frac{G}{F} \Delta t, \end{aligned} \tag{7b}$$

i.e.

$$\lim_{\Delta x_0 \to 0^-} \frac{f(x_0+\Delta x_0) - f(x_0)}{\Delta x_0} \neq \lim_{\Delta x_0 \to 0^+} \frac{f(x_0+\Delta x_0) - f(x_0)}{\Delta x_0}. \tag{8}$$

Therefore, for a time-discretized model, the on–off switches may cause discontinuity in the gradient at the initial point x_0 where the switch can be turned on at the initial time or later in time during the assimilation window at a time step t_{i_τ}.

In order to better illustrate the above idea, we plot in Fig. 1a the trajectories of model solutions with IC equal to $x_0=2$, $x_0 \pm 0.015$ and $x_0 \pm 0.010$, (b) the gradient of the cost function as a function of the IC x_0 for a range of $1.80 \leqslant x_0 \leqslant 2.20$, and (c) the enlarged picture of (b) centered near $x_0=2.0$. These are numerical results we obtained with parameters $F=2$, $G=5$, $x_c=3$, $t_0=0$, $T=1$, and $\Delta t=1/100$. Observations $x^{\text{obs}}(t)$ are obtained using the same forward model with the IC $x_0^{\text{obs}}=1.5$. Obviously, the gradient $\nabla_{x_0} J$ is discontinuous at $x_0=2$ and any other point which can be expressed as $2.0 \pm kF\,\Delta t$ (k is an integer). The gradient is continuous elsewhere.

In summary, for a time-discretized model for which Zou97 is primarily concerned, on–off switches in the solution can cause discontinuity in the gradient of the cost function. More precisely, for the example shown in eqs. (1) and (2), the gradient of the cost function is discontinuous at the IC x_0 if the switch in the model solution is turned on or off at the time $t_i=t_0+i\,\Delta t$, $i=0, 1, \ldots, N$, where Δt is the time step in a numerical model. If the switch is turned on or off between the two time steps, the gradient can still be continuous. The gradient of the cost function is continuous only if $\Delta t \to 0$.

(b) While Xu in point 2a emphasized the importance of realizing the fact that the cost function and the gradient are functions in the space of the initial state x_0, his comment in point 2b has a direct conflict with this realization. It is true that for a gridded numerical model, the dimension of the control variable of the cost function is in finite-dimensional space, say $J=J(\boldsymbol{x})$, where $\boldsymbol{x}$ is a vector of dimension $N(\boldsymbol{x}=(x_1, x_2, \ldots, x_N)^{\mathrm{T}})$. Here J is viewed as a function of multi-variables $x_1, x_2, \ldots,$ and x_N and each of them (x_i) varies independently in the infinite real space $\mathscr{R}$. Therefore, the generalization of the classic gradient and directional derivative is necessary if the assimilation model contains on–off switches and the gradient of the cost function can become discontinuous as shown before.

3. Reply to the second comment: Problems in interpretation and comparisons with the "practical" linearization that ignores the variation of switch time

"Keeping the on–off switch the same as in the NLM" means to keep the IF statement in the tangent linear and adjoint models the same as in the NLM by using the basic state for the IF statement (Zou et al., 1993b). It is equivalent to fixing only the Heaviside function, not the switch time τ outside the Heaviside function fixed. The paragraph after eqs. (4.10)–(4.11), along with eqs. (4.12)–(4.16), are used to reveal the possible reason behind the resulting difference between "keeping the on–off switch the same as in the NLM" in the nonlinear model solution and "keeping the on–off switch the same as in the NLM" in the time tendency equation. Finding the difference in the direct linearization of the nonlinear model solution (with and without ignoring the variation of the switch) and the "practical" linearization of the nonlinear model — keeping the on–off switch the same as in the NLM in the tendency equation — is part of the purpose of Section 4.1 and was clearly shown on p. 12 of Zou97. Here, "practical" means what was actually done in the so-called "classic" adjoint method (Xu96).

It was not claimed in Zou97 that eqs. (4.12ab) are the "practical" tangent linear model (TLM) equations. It was said explicitly in the paragraph following eq. (4.16) that "in the practical adjoint development, eqs. (12b) and (13b) are not considered." Eq. (4.12) is used to analyze what could have caused error in the gradient calculation using the "classic" adjoint method and to search for a way to correct the error if it does produce significant errors in the gradient calculation.

In the gradient calculation, one does not solve the TLM equations. However, the TLM solution is used for deriving the gradient of the cost function using the perturbation method (Navon et al., 1992). The "initial condition" at $t=\tau$ in (4.12b) was assigned in order to obtain a correct TLM solution as shown in (4.11). The entire derivation from (4.1) to (4.16) serves to provide a clear picture of (i) How the existence of an on–off switch generates an error in the gradient calculation using the "classic" adjoint model; (ii) In what circumstances can the gradient calculation using the "classic" adjoint method still be correct in the

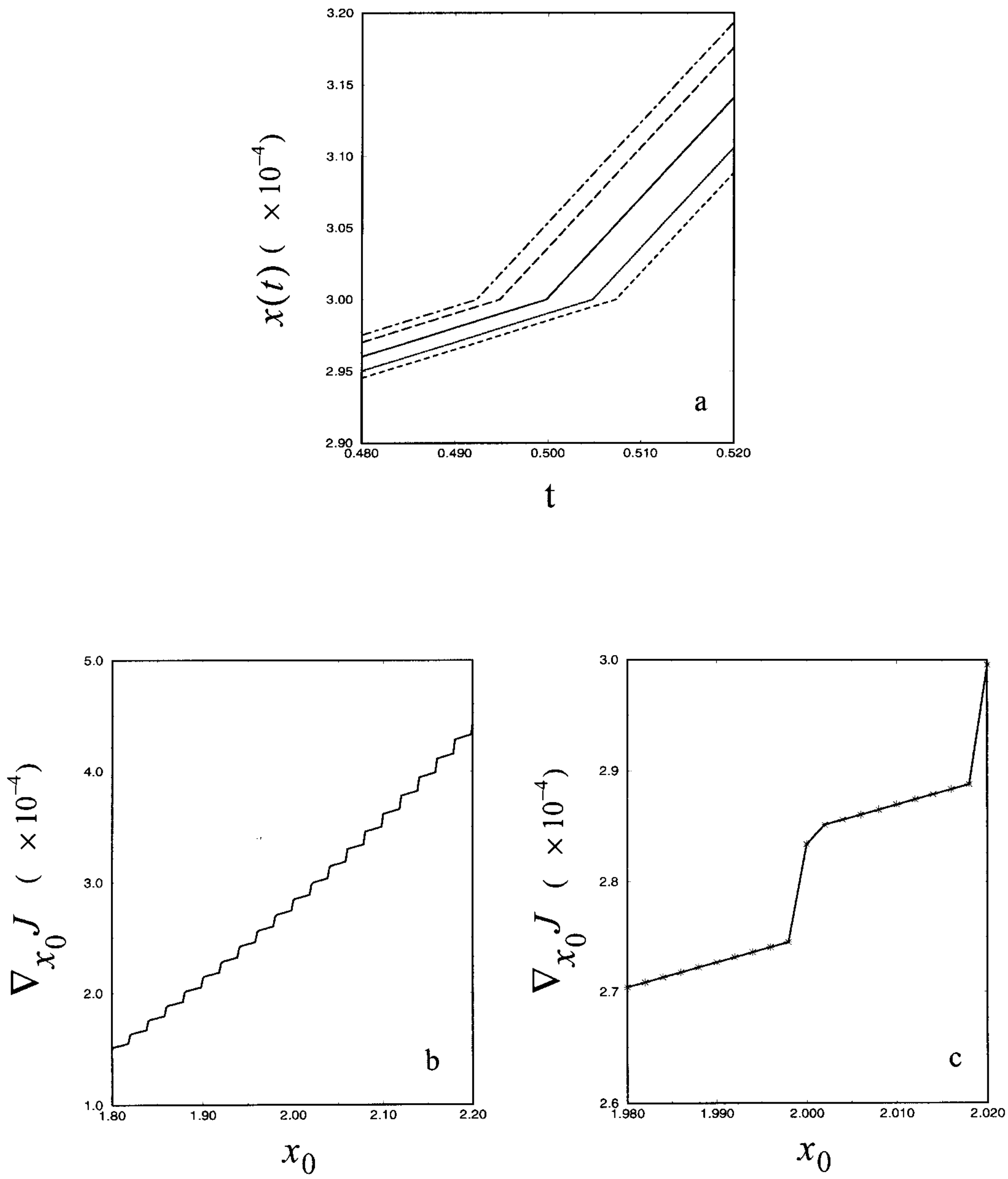

Fig. 1. (a) The trajectories of model (eq. (1)) solutions with IC equal to $x_0 = 2$, $x_0 \pm 0.015$ and $x_0 \pm 0.010$. (b) Variation of the gradient of the cost function (eq. (2)) with respect to the IC x_0 with its values ranging from $1.80 \leqslant_0 \leqslant 2.20$, and (c) the enlarged picture of (b) centered near $x_0 = 2.0$. Also plotted in (c) is the gradient using the formula (4.15) in Zou97 (marked by "*").

presence of on–off switches? and (iii) If the error occurs in the gradient calculation using the "classic" adjoint method due to the existence of on–off switches, can one find a way to obtain a correct gradient still using the values of the adjoint variable through the integration of the "classic" adjoint model? The error in the gradient calculation for eq. (1) using the "classic" adjoint method results from the discontinuity in the perturbation solution at the switching time. From eq. (4.15) we also see that a correct gradient can still be obtained using the values of the "classic" adjoint method and the difference of the two nonlinear model solutions $(x(\tau, x_0+\varepsilon)-x(\tau, x_0))$ at the time τ (switching time), where ε is a small perturbation in the ICs. Here ε is "small" enough to ignore the higher-order terms and "large" enough to avoid machine round-off errors.

4. Reply to the 4th comment

If eq. (1) is modified to the following equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = F + H_+(x-x_\mathrm{c})G(x-x_\mathrm{c}), \quad x|_{t=0} = x_0, \tag{9}$$

the nonlinear model perturbation solution becomes continuous at the switching time τ. As indicated in Zou97, the "classic" adjoint model will provide a correct gradient of the cost function, which we will now show in more detail as follows:

Keeping the on–off switch the same as in the NLM, the TLM equation of (9) becomes

$$\frac{\mathrm{d}x'}{\mathrm{d}t} = H_+(x-x_\mathrm{c})Gx', \quad x'_0|_{t=t_0} = x'_0, \tag{10}$$

which has the following analytic solution

$$x'(t) = x'_0 + H_+(x(t)-x_\mathrm{c})x'_0\,\mathrm{e}^{G(t-\tau)}. \tag{11}$$

This is exactly the 1st-order approximation to the difference between the two forward model solutions.

The adjoint model of eq. (10) can be written as (Zou et al., 1992)

$$-\frac{\mathrm{d}\hat{x}}{\mathrm{d}t} = H_+(x-x_\mathrm{c})G\hat{x} + (x-x^\mathrm{obs}), \quad \hat{x}|_{t=T} = 0. \tag{12}$$

The adjoint eq. (12) is solved sequentially. First we integrate the following adjoint equation

$$-\frac{\mathrm{d}\hat{x}}{\mathrm{d}t} = G\hat{x} + (x-x^\mathrm{obs}), \quad (\tau \leqslant t \leqslant T), \quad \hat{x}|_{t=T} = 0, \tag{13a}$$

backward in time from T to τ to obtain the value of $\hat{x}$ at time τ, $\hat{x}(\tau)$. Then we continue the backward adjoint integration from τ to 0 of the following equation

$$-\frac{\mathrm{d}\hat{x}}{\mathrm{d}t} = (x-x^\mathrm{obs}), \quad (0 \leqslant t \leqslant \tau), \quad \hat{x}|_{t=\tau} = \hat{x}(\tau). \tag{13b}$$

The adjoint eq. (13a) has an analytic solution as

$$\hat{x}(\tau) = \int_\tau^T (x-x^\mathrm{obs})\,\mathrm{e}^{G(t-\tau)}\,\mathrm{d}t. \tag{14}$$

Using (14) we can derive the analytic solution of (13b)

$$\hat{x}(0) = \int_0^\tau (x-x^\mathrm{obs})\,\mathrm{d}t + \int_\tau^T (x-x^\mathrm{obs})\,\mathrm{e}^{G(t-\tau)}\,\mathrm{d}t. \tag{15}$$

The value of $\hat{x}(0)$ is exactly the same value of the G-differential of J defined in (2) that one can derive using the perturbation method described in Subsection 4.2 of Zou97.

In order to examine exactly what will happen for a time-discretized model, we coded a time-stepping nonlinear model of eq. (9), its TLM and adjoint model. The method we used to develop the adjoint model is similar to the one we used for the NCEP (National Center for Environmental Prediction) global spectral adjoint model (Zou et al., 1993b) and the MM5 (the Penn State/NCAR nonhydrostatic mesoscale model version 5) adjoint model (Zou and Kuo, 1996) with moisture physics. We use the following procedure: (1) Build a discretized forward model based on eq. (9); (2) Linearize the forward model around the basic state at the coding level and obtain the TLM; (3) Write the adjoint model as the transpose of the TLM operator; and (4) Integrate the discretized adjoint model backward in time from T to t_0 while forcing terms: $(x(t_i)-x^\mathrm{obs}(t_i))\,\Delta t$ $(i=N, N-1, \ldots, 0)$ are added to the adjoint variable $\hat{x}(t)$ at every time step. Notice that the cost function (2) is now modified to

$$J(x_0) = \frac{1}{2}\sum_{i=0}^{N} (x(t_i)-x^\mathrm{obs})^2\,\Delta t, \tag{16}$$

observations are generated by the forward model integration. The model parameters take the same values as those in Fig. 1.

Fig. 2 shows the variation of the true perturbation solution (dashed line) and the results of numerical integration of the discretized TLM (solid line marked by "*") for model eqs. (1) and (9). We find that the TLM solution which keeps

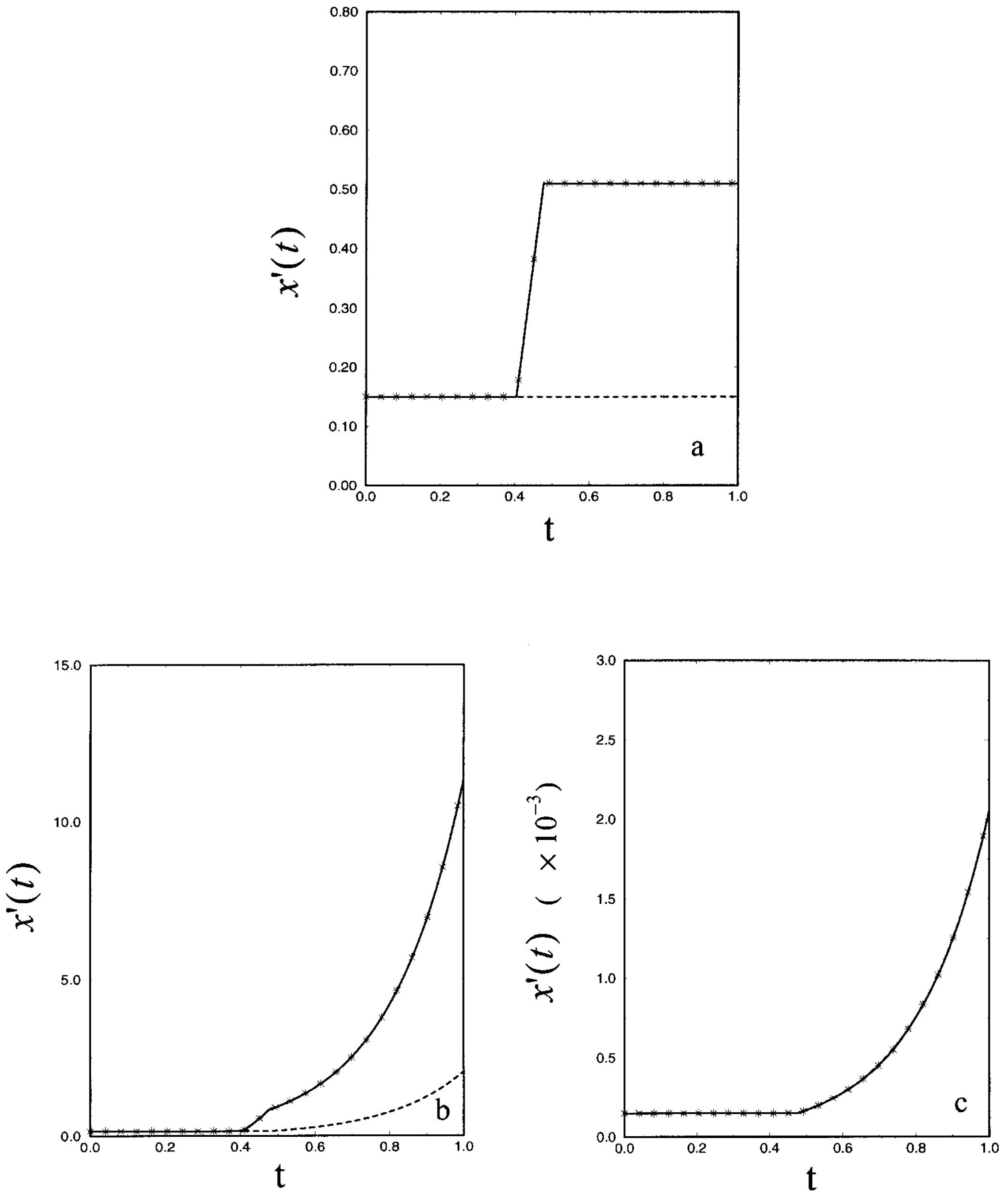

Fig. 2. The variation of the true perturbation solution (solid line marked by "*") and the results of numerical integration of the discretized TLM (dashed line) for (a) model eq. (1), (b) model eq. (9) with $x_0'=0.15$ and (c) model eq. (9) with $x_0'=0.0015$.

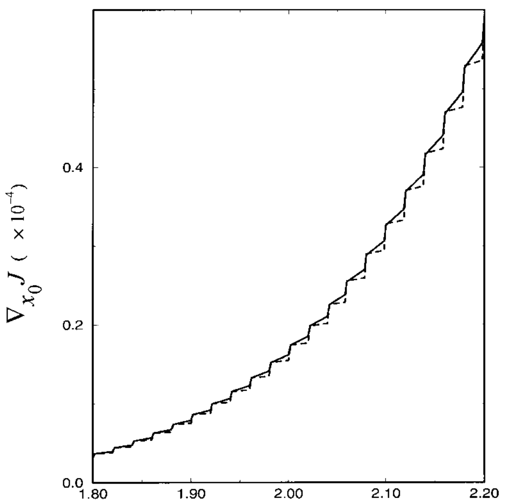

Fig. 3. Variation of the gradient of the cost function with respect to the IC obtained using the "classic" adjoint model which keeps the on–off switches the same as in the NLM (dashed line) and the true solution of the gradient (solid line). Eq. (9) is used as an assimilation model.

the "on–off" switch the same as in the NLM does not approximate the NLM perturbation solution well when the perturbation solution is discontinuous (Fig. 2a). However, when the NLM perturbation solution is continuous at the switching time, the TLM solution which keeps the "on–off" switch the same as in the NLM still provides a correct first-order approximation to the NLM solution (Figs. 2b and c). Now let's examine the correctness of the gradient of the cost function obtained using the adjoint model which is developed as a transpose operation of the forward TLM model with on–off switches kept the same as in the NLM.

Figure 3 shows the variation of the gradient of the cost function with respect to the IC obtained using the "classic" adjoint model (dashed line), which is compared with the true solution of the gradient (solid line) for eq. (9). The dash line is the numerical result obtained by integrating the adjoint model backward in time. We found that the variation of the gradient with respect to the IC using the "classic" adjoint method approximates the true solution very well. The difference between the two lines depends on the accuracy of the finite-difference scheme in time. If we reduce the time step Δt from 0.01 to 0.001, we obtain a greatly improved accuracy for the gradient calculation using the adjoint model. Therefore, the gradient of the cost function obtained by the adjoint model which keeps the "on–off" switches the same as in the NLM is correct in the presence of on–off switches if the nonlinear perturbation solution is continuous at the switching points in time.

5. Summary

I hope that both issues: (i) the continuity of the gradient of the cost function with respect to the IC in the presence of on–off switches, and (ii) the correctness of the gradient of the cost function using the "classic" adjoint model — an adjoint model developed by keeping on-off switches the same as in the TLM, have been addressed more clearly through this reply and the corresponding review of the paper Zou97 that precipitated this comment and reply event.

6. Acknowledgments

The author thanks Dr. Bin Wang for his help in developing all the discretized numerical models and in making the corresponding plots needed for this reply.

REFERENCES

Fischer, M. and P. Courtier, 1995. Estimating the covariance matrices of analysis and forecast error in variational data assimilation. *Technical Memorandum*, no. 220, ECMWF, Reading, UK.

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

Xu, Q., 1996. Generalized adjoint for physical processes with parameterized discontinuities — Part I: basic issues and heuristic examples. *J. Atmos. Sci.* **53**, 1123–1142.

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

Zou, X., I. M. Navon, M. Berger, Paul, K. H. Phua, T. Schlick, and F. X. LeDimet, 1993a. Numerical experience with limited-memory quasi-Newton and truncated-Newton methods. *SIAM Journal on Optimization* **3**, 582–608.

Zou, X., Navon, I. M. and J. G. Sela, 1993b. Variational data assimilation with moist threshold processes using the NMC spectral model. *Tellus* **45A**, 370–387.

Zou, X. and Y.-H. Kuo, 1996. Rainfall assimilation through an optimal control of initial and boundary conditions in a limited-area mesoscale model. *Mon. Wea. Rev.* **124**, 2859–2882.

Zou, X., 1997. Tangent linear and adjoint of "on–off" processes and their feasibility for use in 4-dimensional variational data assimilation. *Tellus* **49A**, 3–31.