*Q. J. R. Meteorol. Soc.* (2001), **127**, pp. 2181–2204

# A review on the use of the adjoint method in four-dimensional atmospheric-chemistry data assimilation

By K.-Y. WANG[1,2*], D. J. LARY[2], D. E. SHALLCROSS[3], S. M. HALL[2] and J. A. PYLE[2]

[1]*National Central University, Chung-Li, Taiwan*
[2]*University of Cambridge, UK*
[3]*University of Bristol, UK*

(Received 19 June 2000; revised 30 January 2001)

SUMMARY

In this paper we review a theoretical formulation of the adjoint method to be used in four-dimensional (4D) chemistry data assimilation. The goal of the chemistry data assimilation is to combine an atmospheric-chemistry model and actual observations to produce the best estimate of the chemistry of the atmosphere. The observational dataset collected during the past decades is an unprecedented expansion of our knowledge of the atmosphere. The exploitation of these data is the best way to advance our understanding of atmospheric chemistry, and to develop chemistry models for chemistry-climate prediction. The assimilation focuses on estimating the state of the chemistry in a chemically and dynamically consistent manner (if the model allows online interactions between chemistry and dynamics). In so doing, we can: produce simultaneous and chemically consistent estimates of all species (including model parameters), observed and unobserved; fill in data voids; test the photochemical theories used in the chemistry models. In this paper, the Hilbert space is first formulated from the geometric structure of the Banach space, followed by the development of the adjoint operator in Hilbert space. The principle of the adjoint method is described, followed by two examples which show the relationship of the gradient of the cost function with respect to the output vector and the gradient of the cost function with respect to the input vector. Applications to chemistry data assimilation are presented for both continuous and discrete cases. The 4D data variational adjoint method is then tested in the assimilation of stratospheric chemistry using a simple catalytic ozone-destruction mechanism, and the test results indicate that the performance of the assimilation method is good.

KEYWORDS: Adjoint method Data assimilation 4D-Var

## 1. INTRODUCTION

Over the past decades both (A) a huge number of observations have been accumulated concerning the atmosphere; and (B) a significant improvement in the photochemical theories and models for the studies of important issues such as stratospheric ozone depletion, anthropogenic perturbation of atmospheric chemistry, and the interaction between atmospheric chemistry and terrestrial vegetation have been made.

An increasing amount of effort is being devoted to the studies of the following questions:

(i) How can we bring (A) and (B) together in order to maximize our understanding of atmospheric chemistry?

(ii) How can we use (A) to test and improve (B)? We shouldn't simply use (A) as a mean of verifying (B), because (B) is often too complicated.

Discrepancies between (B) and (A) are in general very difficult to interpret. For example, could the discrepancies be due to the improper use of initial/boundary conditions, or the incompleteness of the photochemistry (and heterogeneous chemistry) theories embedded in the model? However, if we try to 'best fit' (B) to (A), then (i) the best analysis, given the observations and model, can be obtained (i.e. bypassing the uncertainties incurred by the initial conditions and focusing on the model itself); hence, (ii) data voids can be filled (e.g. by applying the 'best fit' of the chemistry model

* Corresponding author: Department of Atmospheric Sciences, National Central University, Chung-Li, Taiwan. e-mail: kuoying@mail.atm.ncu.edu.tw

to sparse and irregular observations in space and time, data voids can be analysed (filled) in a photochemically consistent manner at spatial and temporal points where no observation is available); hence, (iii) photochemical theories can be tested (e.g. via the comparison of best model analyses with measurements). Furthermore, (iv) our understanding of atmospheric chemistry can be improved (e.g. discrepancies between the best possible model assimilation (analyses) and measurements can be attributed to the incompleteness of the model compared with the real world).

In other words, the grand picture of the real world hidden behind those seemingly scattered and sparse observations can be reproduced (one of the ultimate aims of the modelling), thanks to the use of the adjoint method and the framework of the photochemistry theories (models). Only when this man-made picture (based on photochemical theories) is successfully reproduced can the comparison of model results (analyses, simulations, predictions, etc.) with observations become meaningful and an insight into atmospheric photochemical processes be derived. For example, Fisher and Lary (1995) proposed the first Four-Dimensional Variational (4D-Var) chemistry data assimilation to be used in stratospheric chemistry; Khattatov *et al.* (1999) applied the 4D data assimilation method to the analysis of the Upper Atmosphere Research Satellite data; and Elbern *et al.* (1997) used data assimilation for tropospheric chemistry modelling.

Variational data assimilation has long been used in meteorology and oceanography (see e.g. Talagrand and Courtier 1987; Navon *et al.* 1992). However, the introduction of the adjoint method into the study of atmospheric chemistry only began in the mid-1990s, and the increased recognition of the importance of this assimilation method has grown over the past few years. Although some papers had already briefly sketched the adjoint method for chemistry data assimilation (e.g. Fisher and Lary 1995; Elbern *et al.* 1997), a complete overview of the adjoint method has not been developed and presented. In this paper we present a review on the implementation of the adjoint method in 4D-Var chemistry data assimilation, and its robustness when applied to an idealized case of atmospheric-chemistry data analysis. We note that the Kalman filter is also being developed and used as another alternative method in atmospheric data assimilation (e.g. Daley 1991; Khattatov *et al.* 1999).

## 2. ADJOINT THEORY

### (*a*) *The Hilbert space*

The power of adjoint theory and the elegance of its approach are constructed based on the Hilbert space theory. In this section we describe the basis of Hilbert space. Most of our discussions are developed from those of Simmons (1963) and Feintuch (1998).

(i) *Linear space.* A set is generally a collection of elements, without coherence or form. When an algebraic or geometric structure is imposed on a set, so that its elements are organized into a systematic whole, then it becomes a space. Let $L$ be a non-empty set, and assume that each pair of elements $x$ and $y$ in $L$ can be combined by a process called addition to yield an element $z$ in $L$ denoted by $z = x + y$. We now assume that each element $x$ in $L$ can be combined by a process called scalar multiplication to yield an element $y$ in $L$ denoted by $y = ax$. The algebraic system $L$ defined by these operations and axioms is called a linear space. A linear space is often called a vector space, and its elements are spoken of as vectors. Notice that the system of real numbers or the system of complex numbers are the scalars.

(ii) *Normed linear space: Banach space.* Let $X$ be a non-empty set. A metric on $X$ is a real function $\rho$ of ordered pairs of elements of $X$ which satisfies the following conditions:

$$\rho(x, y) \geqslant 0 \quad \text{and} \quad \rho(x, y) = 0 \Longleftrightarrow x = y, \tag{1a}$$

$$\rho(x, y) = \rho(y, x) \quad \text{(symmetry)}, \tag{1b}$$

$$\rho(x, y) \leqslant \rho(x, z) + \rho(z, y) \quad \text{(the triangle inequality)}. \tag{1c}$$

The function $\rho$ assigns to each pair $(x, y)$ of elements of $X$ a non-negative real number $\rho(x, y)$. Here $\rho(x, y)$ is called the distance between $x$ and $y$. A metric space consists of two objects: a non-empty set $X$ and a metric $\rho$ on $X$.

Let $X$ be a metric space with metric $\rho$, and let

$$(x_n) = (x_1, x_2, \ldots, x_n, \ldots) \tag{2}$$

be a sequence of points in $X$. We say that $x_n$ is convergent if there exists a point $x$ in $X$ such that for every $\epsilon > 0$ there exists a positive integer $n_0$ such that

$$n \geqslant n_0 \longrightarrow \rho(x_n, x) < \epsilon. \tag{3}$$

For every convergent sequence $x_n$ has the following property: for each $\epsilon > 0$, there exists a positive integer $n_0$ such that

$$m, n \geqslant n_0 \longrightarrow \rho(x_m, x_n) < \epsilon. \tag{4}$$

A sequence with this property is called a Cauchy sequence. Every convergent sequence is a Cauchy sequence; but a Cauchy sequence is not necessarily convergent. A complete metric space is a metric space in which every Cauchy sequence is convergent.

A normed linear space is a linear space on which there is defined a norm, i.e. a function which assigns to each element $x$ in space a real number $\|x\|$ in such a manner that

$$\|x\| \geqslant 0, \quad \text{and} \quad \|x\| = 0 \Longleftrightarrow x = 0, \tag{5a}$$

$$\|x + y\| \leqslant \|x\| + \|y\|, \tag{5b}$$

$$\|\alpha x\| = |\alpha| \|x\|. \tag{5c}$$

A normed linear space is a metric space with respect to the induced metric defined by

$$\rho(x, y) = \|x - y\|. \tag{6}$$

A Banach space is a normed linear space which is complete as a metric space.

(iii) *Inner-product defined normed space: Hilbert space.* Let $\mathcal{H}$ be a vector space. A function $\langle\ ,\ \rangle$ is called an inner product. Now define a new function $\|\ \|: \mathcal{H} \to R$ by $\|x\| = \langle x, x\rangle^{1/2}$, where $R$ is a real number. It follows that $\|\ \|$ defines a norm on $\mathcal{H}$. This norm allows us to define a metric on $\mathcal{H}$ by means of the metric $\rho$,

$$\rho(x, y) = \|x - y\|. \tag{7}$$

Hence, for every two points $x$ and $y$, the metric $\rho$ gives the distance between them. A sequence $(x_n)$ is said to converge to $x \in \mathcal{H}$ if

$$\rho(x_n, x) = \|x_n - x\| \to 0 \quad \text{as} \quad n \to \infty. \tag{8}$$

A Cauchy sequence is a sequence $x_1, x_2, \ldots$ such that the metric $\rho(x_m, x_n)$ satisfies (Weisstein 1998):

$$\lim_{\min(m,n)\to\infty} \rho(x_m, x_n) = 0. \tag{9}$$

We say $\mathcal{H}$ is *complete* if every Cauchy sequence in $\mathcal{H}$ converges in $\mathcal{H}$. This means that the limit of a convergent sequence of vectors in $\mathcal{H}$ is also in $\mathcal{H}$. Since the Cauchy sequence does converge in the real line (real numbers), the finite-dimensional vector space where the inner product has been defined is always complete. A complete inner-product space is called a *Hilbert space*.

Hence, a Hilbert space is a special Banach space whose norm arises from an inner product, that is, in which there is defined a function $\langle \mathbf{x}, \mathbf{y} \rangle$ of vectors $\mathbf{x}$ and $\mathbf{y}$ with the following properties:

$$\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{z} \rangle + \beta \langle \mathbf{y}, \mathbf{z} \rangle, \tag{10a}$$

$$\overline{\langle \mathbf{x}, \mathbf{y} \rangle} = \langle \mathbf{y}, \mathbf{x} \rangle, \quad \text{if } \mathbf{x}, \mathbf{y} \in C, \tag{10b}$$

$$\langle \mathbf{x}, \mathbf{x} \rangle = \|\mathbf{x}\|^2. \tag{10c}$$

### (*b*) *The adjoint operator*

Following Talagrand (1991), let us consider Hilbert space $\mathcal{E}$ and Hilbert space $\mathcal{F}$, on which inner products have been defined. Given a continuous linear operator L of $\mathcal{E}$ into $\mathcal{F}$, there exists a unique continuous linear operator L* of $\mathcal{F}$ into $\mathcal{E}$, such that for any vector $\mathbf{U} \in \mathcal{E}$ and $\mathbf{V} \in \mathcal{F}$, the following equality between inner products holds:

$$\langle \mathbf{LU}, \mathbf{V} \rangle = \langle \mathbf{U}, \mathbf{L}^*\mathbf{V} \rangle. \tag{11}$$

L* is called the adjoint of L. The property of the adjoint operator L* can be illustrated as follows. Let U and V have finite dimension $n$ and $m$, respectively,

$$\mathbf{U} = \begin{Bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{Bmatrix}, \tag{12}$$

$$\mathbf{V} = \begin{Bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{Bmatrix}, \tag{13}$$

and L be a $m \times n$ matrix. Then the inner product of the left-hand side of (11) is

$$\begin{aligned} \langle \mathbf{LU}, \mathbf{V} \rangle &= \sum_{i=1}^{m} \left( \sum_{j=1}^{n} L_{ij} u_j \right) v_i, \\ &= \sum_{j=1}^{n} u_j \left( \sum_{i=1}^{m} L_{ij} v_i \right). \end{aligned} \tag{14}$$

The exchange in the order of summation indices has been applied to obtain the last line of (14). While the order of the summation in the first line of (14) is first taken on the

input index $j$ then on $i$, the summation in the second line is first performed on the input index $i$ then on $j$. The operations on $\mathbf{L}$ have been turned from row-by-row in the first line into column-by-column in the second line. This shows that the matrix representing $\mathbf{L}^*$ is the transpose of the matrix representing $\mathbf{L}$ (Talagrand 1991). The exchange of the summation operations is allowed given that the element of the matrix $\mathbf{L}$ is a number and not, for instance, another matrix.

### (c) *The principle of the adjoint method*

We follow Talagrand (1991) and Talagrand and Courtier (1987) to derive the general principle of the adjoint method in this section. Let $\mathbf{U} \in \mathcal{E}$ be the input vector, and $\mathbf{V} \in \mathcal{F}$ be the output vector ($\mathcal{E}$ and $\mathcal{F}$ are Hilbert spaces). The components of $\mathbf{U}$ are denoted in (12), and the components of $\mathbf{V}$ are denoted in (13). The output vector $\mathbf{V}$ is a differentiable function of the input vector $\mathbf{U}$ by (Fig. 1):

$$\mathbf{V} = \mathbf{G}(\mathbf{U}). \tag{15}$$

For a perturbation on the output vector $\mathbf{V}$, the corresponding perturbation $\delta\mathbf{V}$ is

$$\delta\mathbf{V} = \mathbf{G}'\delta\mathbf{U}, \tag{16}$$

where the Jacobian matrix $\mathbf{G}'$ is the first derivative of $\mathbf{G}$ with respect to the input vector $\mathbf{G}$:

$$\mathbf{G}' = \frac{\partial \mathbf{G}}{\partial \mathbf{U}} = \begin{pmatrix} \dfrac{\partial G_{11}}{\partial u_1} & \dfrac{\partial G_{12}}{\partial u_2} & \cdots & \dfrac{\partial G_{1n}}{\partial u_n} \\ \dfrac{\partial G_{21}}{\partial u_1} & \dfrac{\partial G_{22}}{\partial u_2} & \cdots & \dfrac{\partial G_{2n}}{\partial u_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial G_{m1}}{\partial u_1} & \dfrac{\partial G_{m2}}{\partial u_2} & \cdots & \dfrac{\partial G_{mn}}{\partial u_n} \end{pmatrix}. \tag{17}$$

Equation (16) is called the tangent linear equation to the nonlinear equation (15) (Talagrand and Courtier 1987), and the equation says that, for a given $\mathbf{U}$, the local sensitivities of the output parameters ($\delta\mathbf{V}$) with respect to the input parameters ($\delta\mathbf{U}$) are given by the Jacobian matrix $\mathbf{G}'(\mathbf{U})$. Now let $J : \mathbf{V} \to J(\mathbf{V})$ be a scalar-valued function of the output vector $\mathbf{V}$. The variation of $J$ with respect to output vector $\mathbf{V}$ is

$$\delta J = \langle \nabla_{\mathbf{V}} J, \delta\mathbf{V} \rangle. \tag{18}$$

Substituting (16) into (18) and using the property of the adjoint operator (11), (18) becomes

$$\delta J = \langle \nabla_{\mathbf{V}} J, \mathbf{G}'\delta\mathbf{U} \rangle = \langle \mathbf{G}'^* \nabla_{\mathbf{V}} J, \delta\mathbf{U} \rangle. \tag{19}$$

By definition of a gradient, this shows that the gradient $\mathbf{G}'^* \nabla_{\mathbf{V}} J$ of $J$ with respect to the input vector $\mathbf{U}$ is equal to

$$\nabla_{\mathbf{U}} J = \mathbf{G}'^* \nabla_{\mathbf{V}} J. \tag{20}$$

Equation (20) is at the heart of the use of adjoint equations, and provides a very efficient way to numerically determine $\nabla_{\mathbf{U}} J$. The principle of the adjoint method is to compute the gradient $\nabla_{\mathbf{U}} J$, not through the direct explicit perturbation of the input parameters ($\mathbf{U}$), but through the use of (20). What is more, the adjoint method gives the instructive

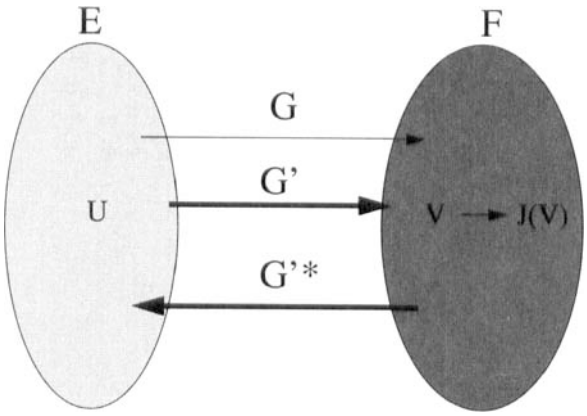


Figure 1. A schematic diagram which shows relations between input vector **U** and output vector **V**: $\mathbf{V} = \mathbf{G}(\mathbf{U})$. The operator $\mathbf{G}'$ is the first derivative of **G** with respect to the input vector **U**. The operator $\mathbf{G}'^*$ is the adjoint of $\mathbf{G}'$. Here $J$ is a differentiable scalar-valued function of the output vector **V**. See text for explanation.

insight obtained from the comparison between (16) and (20). While the direct Jacobian $\mathbf{G}'$ transforms a perturbation on the input vector $\delta\mathbf{U}$ into the first-order corresponding perturbation on the output vector $\delta\mathbf{V}$, the transpose Jacobian $\mathbf{G}'^*$ transforms the gradient of a scalar-valued function $J$ with respect to the output vector $(\nabla_{\mathbf{V}} J)$ into the gradient of the function $J$ with respect to the input vector $(\nabla_{\mathbf{V}} J)$ (Fig. 1).

### (*d*) *An example of the finite-dimensional case*

Let us now consider the finite dimensional vectors **U** and **V** ((12) and (13)). The first-order perturbation $\delta v_i$ $(i = 1, \ldots, m)$ with respect to the input vector $\delta\mathbf{U}$ is given by

$$\delta v_i = \sum_{j=1}^{n} \frac{\partial G_{ij}}{\partial u_j} \delta u_j, \quad i = 1, \ldots, m. \tag{21}$$

The partial derivative of $J$ with respect to the input parameter $u_j$ $(j = 1, \ldots, n)$ is given by the chain rule

$$\left.\begin{aligned} J &= J(\mathbf{V}) \\ &= J(\mathbf{G}(\mathbf{U})), \\ \frac{\partial J}{\partial u_j} &= \sum_{i=1}^{m} \frac{\partial J}{\partial G_{ij}(u_i)} \frac{\partial G_{ij}(u_i)}{\partial u_j} \\ &= \sum_{i=1}^{m} \frac{\partial J}{\partial v_i} \frac{\partial G_{ij}}{\partial u_j}, \quad j = 1, \ldots, n. \end{aligned}\right\} \tag{22}$$

Here, $\partial J/\partial u_j$ is the gradient of $J$ with respect to the input vector **U** and is denoted by $\nabla_{\mathbf{U}} J$; while $\partial J/\partial v_i$ is the gradient of $J$ with respect to the output vector **V** and is denoted by $\nabla_{\mathbf{V}} J$. Notice that the summation in (21) is performed over index $j$ for each $i$ of $\partial G_{ij}/\partial u_j$, i.e. over each row of the same index $i$, and the summation in (22) is performed over the index $i$ for each index $j$. Hence, if we denoted the former summation as the operation over a matrix $\mathbf{G}'$, then the latter summation is exactly the operation over the transpose of the matrix $\mathbf{G}'$ and we denote it as $\mathbf{G}'^*$. Therefore the matrix form of (22) can be written as that shown in (20).

### (e) *An example of catalytic ozone destruction*

Consider the following important catalytic cycles for ozone destruction in polar regions:

$$\begin{array}{lll} \mathrm{ClO} + \mathrm{BrO} \longrightarrow \mathrm{Br} + \mathrm{Cl} + \mathrm{O_2} & (k_1) \\ \mathrm{Cl} + \mathrm{O_3} \longrightarrow \mathrm{ClO} + \mathrm{O_2} & (k_2) \\ \mathrm{Br} + \mathrm{O_3} \longrightarrow \mathrm{BrO} + \mathrm{O_2} & (k_3) \\ \text{Net:} \quad 2\mathrm{O_3} \longrightarrow 3\mathrm{O_2} \end{array}$$

where $k_1$, $k_2$, and $k_3$ are reaction rate coefficients. The gas-phase rate equations for the change of

$$\mathbf{U} = \begin{Bmatrix} \mathrm{O_3} \\ \mathrm{ClO} \\ \mathrm{BrO} \\ \mathrm{Cl} \\ \mathrm{Br} \end{Bmatrix}, \tag{23}$$

where $\mathrm{O_3}$ etc. represent gas concentrations, can be written as

$$\frac{\mathrm{d}\mathbf{U}}{\mathrm{d}t} = F(\mathbf{U}), \tag{24}$$

where

$$F(\mathbf{U}) = \begin{Bmatrix} -k_2[\mathrm{Cl}][\mathrm{O_3}] - k_3[\mathrm{Br}][\mathrm{O_3}] \\ -k_1[\mathrm{ClO}][\mathrm{BrO}] + k_2[\mathrm{Cl}][\mathrm{O_3}] \\ -k_1[\mathrm{ClO}][\mathrm{BrO}] + k_3[\mathrm{Br}][\mathrm{O_3}] \\ k_1[\mathrm{ClO}][\mathrm{BrO}] - k_2[\mathrm{Cl}][\mathrm{O_3}] \\ k_1[\mathrm{ClO}][\mathrm{BrO}] - k_3[\mathrm{Br}][\mathrm{O_3}] \end{Bmatrix} \tag{25}$$

Hence, the relationship between the input vector **U** and the output vector **V** during the time interval $\Delta t$ can be written as

$$\mathbf{V} = \mathbf{G}(\mathbf{U}), \tag{26}$$

where

$$\mathbf{G}(\mathbf{U}) = \begin{Bmatrix} \mathrm{O_3}^{(n)} + \Delta t(-k_2[\mathrm{Cl}^{(n)}][\mathrm{O_3}^{(n)}] - k_3[\mathrm{Br}^{(n)}][\mathrm{O_3}^{(n)}]) \\ \mathrm{ClO}^{(n)} + \Delta t(-k_1[\mathrm{ClO}^{(n)}][\mathrm{BrO}^{(n)}] + k_2[\mathrm{Cl}^{(n)}][\mathrm{O_3}^{(n)}]) \\ \mathrm{BrO}^{(n)} + \Delta t(-k_1[\mathrm{ClO}^{(n)}][\mathrm{BrO}^{(n)}] + k_3[\mathrm{Br}^{(n)}][\mathrm{O_3}^{(n)}]) \\ \mathrm{Cl}^{(n)} + \Delta t(k_1[\mathrm{ClO}^{(n)}][\mathrm{BrO}^{(n)}] - k_2[\mathrm{Cl}^{(n)}][\mathrm{O_3}^{(n)}]) \\ \mathrm{Br}^{(n)} + \Delta t(k_1[\mathrm{ClO}^{(n)}][\mathrm{BrO}^{(n)}] - k_3[\mathrm{Br}^{(n)}][\mathrm{O_3}^{(n)}]) \end{Bmatrix} = \begin{Bmatrix} G_1 \\ G_2 \\ G_3 \\ G_4 \\ G_5 \end{Bmatrix}. \tag{27}$$

Equation (27) shows that **G** is nonlinear in **U**. Here **U** is the concentrations (input vector) at time step $n$:

$$\mathbf{U} = \begin{Bmatrix} \mathrm{O_3}^{(n)} \\ \mathrm{ClO}^{(n)} \\ \mathrm{BrO}^{(n)} \\ \mathrm{Cl}^{(n)} \\ \mathrm{Br}^{(n)} \end{Bmatrix}, \tag{28}$$

and **V** is the concentrations (output vector) at time step $n+1$

$$\mathbf{V} = \begin{Bmatrix} O_3^{(n+1)} \\ ClO^{(n+1)} \\ BrO^{(n+1)} \\ Cl^{(n+1)} \\ Br^{(n+1)} \end{Bmatrix}. \tag{29}$$

The tangent linear equation with respect to (24) for a perturbation of **U** is written as (see also Elbern *et al.* 1997)

$$\frac{\delta \mathrm{d}\mathbf{U}}{\mathrm{d}t} = F'\delta\mathbf{U}, \tag{30}$$

or, equivalently from (26)

$$\delta\mathbf{V} = \mathbf{G}'\delta\mathbf{U}, \tag{31}$$

where

$$\mathbf{G}' = \frac{\partial \mathbf{G}}{\partial \mathbf{U}} = \begin{pmatrix} \frac{\partial G_1}{\partial O_3^{(n)}} & \frac{\partial G_1}{\partial ClO^{(n)}} & \frac{\partial G_1}{\partial BrO^{(n)}} & \frac{\partial G_1}{\partial Cl^{(n)}} & \frac{\partial G_1}{\partial Br^{(n)}} \\ \frac{\partial G_2}{\partial O_3^{(n)}} & \frac{\partial G_2}{\partial ClO^{(n)}} & \frac{\partial G_2}{\partial BrO^{(n)}} & \frac{\partial G_2}{\partial Cl^{(n)}} & \frac{\partial G_2}{\partial Br^{(n)}} \\ \frac{\partial G_3}{\partial O_3^{(n)}} & \frac{\partial G_3}{\partial ClO^{(n)}} & \frac{\partial G_3}{\partial BrO^{(n)}} & \frac{\partial G_3}{\partial Cl^{(n)}} & \frac{\partial G_3}{\partial Br^{(n)}} \\ \frac{\partial G_4}{\partial O_3^{(n)}} & \frac{\partial G_4}{\partial ClO^{(n)}} & \frac{\partial G_4}{\partial BrO^{(n)}} & \frac{\partial G_4}{\partial Cl^{(n)}} & \frac{\partial G_4}{\partial Br^{(n)}} \\ \frac{\partial G_5}{\partial O_3^{(n)}} & \frac{\partial G_5}{\partial ClO^{(n)}} & \frac{\partial G_5}{\partial BrO^{(n)}} & \frac{\partial G_5}{\partial Cl^{(n)}} & \frac{\partial G_5}{\partial Br^{(n)}} \end{pmatrix}. \tag{32}$$

Equation (30) (or (31)) is called the tangent linear equation. Equations (32) and (27) show that $\mathbf{G}'$ is now linear in **U**, but $\mathbf{G}'$ is still time dependent. Let $J$ be a scalar-valued function, and $J$ a function of output vector **V**:

$$J = J(\mathbf{V}). \tag{33}$$

For example, $J$ can be the distance function (Talagrand and Courtier 1987) which measures the difference between model and observations at time step $n+1$. If we denote $\widehat{O_3}^{(n+1)}$, $\widehat{ClO}^{(n+1)}$, $\widehat{BrO}^{(n+1)}$, $\widehat{Cl}^{(n+1)}$ and $\widehat{Br}^{(n+1)}$ as observations, then $J$ is

$$\begin{aligned} J = \tfrac{1}{2}\{&(\widehat{O_3}^{(n+1)} - O_3{}^{(n+1)})^2 + (\widehat{ClO}^{(n+1)} - ClO^{(n+1)})^2 \\ &+ (\widehat{BrO}^{(n+1)} - BrO^{(n+1)})^2 + (\widehat{Cl}^{(n+1)} - Cl^{(n+1)})^2 \\ &+ (\widehat{Br}^{(n+1)} - Br^{(n+1)})^2\}. \end{aligned} \tag{34}$$

The gradient of $J$ with respect to $O_3^{(n)}$ at time step $n$ is

$$\begin{aligned}\frac{\partial J}{\partial O_3^{(n)}} &= \frac{\partial \mathbf{G}(\mathbf{U})}{\partial O_3^{(n)}} \frac{\partial J}{\partial \mathbf{G}(\mathbf{U})} \\ &= \frac{\partial \mathbf{G}(\mathbf{U})}{\partial O_3^{(n)}} \frac{\partial J}{\partial \mathbf{V}} \\ &= \left[\frac{\partial G_1}{\partial O_3^{(n)}} \quad \frac{\partial G_2}{\partial O_3^{(n)}} \quad \frac{\partial G_3}{\partial O_3^{(n)}} \quad \frac{\partial G_4}{\partial O_3^{(n)}} \quad \frac{\partial G_5}{\partial O_3^{(n)}}\right] \begin{Bmatrix} \frac{\partial J}{\partial O_3^{(n+1)}} \\ \frac{\partial J}{\partial ClO^{(n+1)}} \\ \frac{\partial J}{\partial BrO^{(n+1)}} \\ \frac{\partial J}{\partial Cl^{(n+1)}} \\ \frac{\partial J}{\partial Br^{(n+1)}} \end{Bmatrix}. \end{aligned} \tag{35}$$

Notice here that $J$ is a scalar-valued function. The gradients of $J$ with respect to the other four species are derived in a similar manner. Hence, at the end we can write the gradient of $J$ with respect to the input vector at time step $n$ as

$$\begin{aligned}\nabla_{\mathbf{U}} J &= \begin{Bmatrix} \frac{\partial J}{\partial O_3^{(n)}} \\ \frac{\partial J}{\partial ClO^{(n)}} \\ \frac{\partial J}{\partial BrO^{(n)}} \\ \frac{\partial J}{\partial Cl^{(n)}} \\ \frac{\partial J}{\partial Br^{(n)}} \end{Bmatrix} \\ &= \begin{pmatrix} \frac{\partial G_1}{\partial O_3^{(n)}} & \frac{\partial G_2}{\partial O_3^{(n)}} & \frac{\partial G_3}{\partial O_3^{(n)}} & \frac{\partial G_4}{\partial O_3^{(n)}} & \frac{\partial G_5}{\partial O_3^{(n)}} \\ \frac{\partial G_1}{\partial ClO^{(n)}} & \frac{\partial G_2}{\partial ClO^{(n)}} & \frac{\partial G_3}{\partial ClO^{(n)}} & \frac{\partial G_4}{\partial ClO^{(n)}} & \frac{\partial G_5}{\partial ClO^{(n)}} \\ \frac{\partial G_1}{\partial BrO^{(n)}} & \frac{\partial G_2}{\partial BrO^{(n)}} & \frac{\partial G_3}{\partial BrO^{(n)}} & \frac{\partial G_4}{\partial BrO^{(n)}} & \frac{\partial G_5}{\partial BrO^{(n)}} \\ \frac{\partial G_1}{\partial Cl^{(n)}} & \frac{\partial G_2}{\partial Cl^{(n)}} & \frac{\partial G_3}{\partial Cl^{(n)}} & \frac{\partial G_4}{\partial Cl^{(n)}} & \frac{\partial G_5}{\partial Cl^{(n)}} \\ \frac{\partial G_1}{\partial Br^{(n)}} & \frac{\partial G_2}{\partial Br^{(n)}} & \frac{\partial G_3}{\partial Br^{(n)}} & \frac{\partial G_4}{\partial Br^{(n)}} & \frac{\partial G_5}{\partial Br^{(n)}} \end{pmatrix} \begin{Bmatrix} \frac{\partial J}{\partial O_3^{(n+1)}} \\ \frac{\partial J}{\partial ClO^{(n+1)}} \\ \frac{\partial J}{\partial BrO^{(n+1)}} \\ \frac{\partial J}{\partial Cl^{(n+1)}} \\ \frac{\partial J}{\partial Br^{(n+1)}} \end{Bmatrix} \\ &= \mathbf{G}'^{*} \nabla_{\mathbf{V}} J. \end{aligned} \tag{36}$$

It is clear from (36) that the adjoint matrix $\mathbf{G}'^*$ is the transpose of the Jacobian matrix $\mathbf{G}'$ of (32). Equation (36) is called the adjoint tangent linear equation (Talagrand and Courtier 1987; Fisher and Lary 1995). If $J$ is the scalar-valued distance function (33), then the problem for atmospheric chemistry data assimilation is to control the input vector $\mathbf{U}$ which will minimize $J$. As will be discussed in more detail in section 3(c), an effective minimization algorithm requires the gradient of $J$ with respect to its input vector $\mathbf{U}$ to be known (Talagrand and Courtier 1985; Giering and Kaminski 1998).

### (*f*) *A remark on the adjoint method*

In the case of atmospheric chemistry modelling, where $\mathbf{V}$ is always an explicit but complicated function of $\mathbf{U}$ (26), it is generally impossible to find a usable numerical expression for $\nabla_{\mathbf{U}} J$ (Talagrand and Courtier 1987). Notice here that the scalar-valued function $J$ is a function of the output vector $\mathbf{V}$, which itself is explicitly determined from $\mathbf{U}$. Now, (36) shows that if a computer code is available to calculate $\mathbf{G}'^*$ for a given $\mathbf{U}$, $\nabla_{\mathbf{U}} J$ can be easily calculated from $\nabla_{\mathbf{V}} J$. Here $\nabla_{\mathbf{V}} J$ can be easily determined whenever $J$ is a simple function of $\mathbf{V}$.

### (*g*) *The essence of the adjoint method*

The essence of the adjoint method is simply to systematically perform computations of (36) (Talagrand 1991). Since at each time step $n$, the adjoint computations proceed from $\nabla_{\mathbf{V}} J$ at later time step $n+1$ to $\nabla_{\mathbf{U}} J$ at earlier time step $n$, the overall integration of the adjoint model is performed in the sense of backward integration into the past. Let $\mathbf{G}^{(i)}$ be the model computation at time step $i$ ($i = 1, \ldots, N$), and $\mathbf{U}^{(i-1)}$ the input vector; if we denote $\circ$ as (Kaminski *et al.* 1999)

$$\mathbf{G}^{(2)}\{\mathbf{G}^{(1)}(\mathbf{U}^{(0)})\} = \mathbf{G}^{(2)} \circ \mathbf{G}^{(1)}(\mathbf{U}^{(0)}), \tag{37}$$

and

$$\mathbf{G} = \mathbf{G}^{(N)} \circ \mathbf{G}^{(N-1)} \circ \cdots \circ \mathbf{G}^{(2)} \circ \mathbf{G}^{(1)}, \tag{38}$$

then

$$\left.\begin{aligned} \mathbf{V}^{(1)} &= \mathbf{G}^{(1)}(\mathbf{U}^{(0)}), \\ \mathbf{V}^{(2)} &= \mathbf{G}^{(2)}(\mathbf{U}^{(1)}) = \mathbf{G}^{(2)} \circ \mathbf{G}^{(1)}(\mathbf{U}^{(0)}), \\ &\vdots \quad \vdots \\ \mathbf{V}^{(N)} &= \mathbf{G}^{(N)}(\mathbf{U}^{(N-1)}) = \mathbf{G}^{(N)} \circ \mathbf{G}^{(N-1)} \circ \cdots \circ \mathbf{G}^{(2)} \circ \mathbf{G}^{(1)}(\mathbf{U}^{(0)}) \\ &= \mathbf{G}(\mathbf{U}^{(0)}). \end{aligned}\right\} \tag{39}$$

If $\mathbf{G}^{(i)}$ is linear in $\mathbf{U}^{(i-1)}$, then $\mathbf{G}^{(i)}(\mathbf{U}^{(i-1)}) = \mathbf{G}^{(i)} \circ \mathbf{U}^{(i-1)}$. Notice that, at each time step $i$, new input vector $\mathbf{U}^{(i)}$ has been updated according to the newly updated output vector $\mathbf{V}^{(i)}$ (39), i.e.

$$\begin{array}{ccc} \mathbf{U}^{(i-1)} & \xrightarrow{\mathbf{G}^{(i)}} & \mathbf{V}^{(i)} \\ & \Downarrow & \\ & \mathbf{U}^{(i)} \xrightarrow{\mathbf{G}^{(i+1)}} & \mathbf{V}^{(i+1)}. \end{array} \tag{40}$$

Hence, at any given $i$, an updated $\mathbf{V}^{(i)}$ is set to a new $\mathbf{U}^{(i)}$, and

$$\delta\mathbf{V}^{(i)} = \delta\mathbf{U}^{(i)}. \tag{41}$$

The tangent linear equation of (39) is written as

$$\delta\mathbf{V}^{(N)} = \mathbf{G}'\delta\mathbf{U}^{(0)}, \tag{42}$$

or

$$\delta\mathbf{U}^{(N)} = \mathbf{G}'\delta\mathbf{U}^{(0)}. \tag{43}$$

By applying the chain rule to the partial derivative of $\mathbf{G}$ with respect to $\mathbf{U}^{(0)}$, the Jacobian of $\mathbf{G}'$ can then be written as

$$\begin{aligned}
\mathbf{G}' &= \frac{\partial \mathbf{G}}{\partial \mathbf{U}^{(0)}} \\
&= \frac{\partial(\mathbf{G}^{(N)}(\mathbf{U}^{(N-1)}))}{\partial \mathbf{U}^{(0)}} \\
&= \frac{\partial(\mathbf{G}^{(N)}(\mathbf{U}^{(N-1)}))}{\partial(\mathbf{G}^{(N-1)} \circ \cdots \circ \mathbf{G}^{(1)}(\mathbf{U}^{(0)}))} \frac{\partial(\mathbf{G}^{(N-1)} \circ \cdots \circ \mathbf{G}^{(1)}(\mathbf{U}^{(0)}))}{\partial(\mathbf{G}^{(N-2)} \circ \cdots \circ \mathbf{G}^{(1)}(\mathbf{U}^{(0)}))} \\
&\quad \times \frac{\partial(\mathbf{G}^{(N-2)} \circ \cdots \circ \mathbf{G}^{(1)}(\mathbf{U}^{(0)}))}{\partial(\mathbf{G}^{(N-3)} \circ \cdots \circ \mathbf{G}^{(1)}(\mathbf{U}^{(0)}))} \cdots \frac{\partial(\mathbf{G}^{(2)} \circ \mathbf{G}^{(1)}(\mathbf{U}^{(0)}))}{\partial(\mathbf{G}^{(1)}(\mathbf{U}^{(0)}))} \frac{\partial(\mathbf{G}^{(1)}(\mathbf{U}^{(0)}))}{\partial \mathbf{U}^{(0)}} \\
&= \frac{\partial(\mathbf{G}^{(N)}(\mathbf{U}^{(N-1)}))}{\partial(\mathbf{G}^{(N-1)}(\mathbf{U}^{(N-2)}))} \frac{\partial(\mathbf{G}^{(N-1)}(\mathbf{U}^{(N-2)}))}{\partial(\mathbf{G}^{(N-2)}(\mathbf{U}^{(N-3)}))} \frac{\partial(\mathbf{G}^{(N-2)}(\mathbf{U}^{(N-3)}))}{\partial(\mathbf{G}^{(N-3)}(\mathbf{U}^{(N-4)}))} \\
&\quad \cdots \frac{\partial \mathbf{G}^{(2)}(\mathbf{U}^{(1)})}{\partial \mathbf{G}^{(1)}(\mathbf{U}^{(0)})} \frac{\partial \mathbf{G}^{(1)}(\mathbf{U}^{(0)})}{\partial \mathbf{U}^{(0)}} \\
&= \frac{\partial(\mathbf{G}^{(N)}(\mathbf{U}^{(N-1)}))}{\partial \mathbf{V}^{(N-1)}} \frac{\partial(\mathbf{G}^{(N-1)}(\mathbf{U}^{(N-2)}))}{\partial \mathbf{V}^{(N-2)}} \frac{\partial(\mathbf{G}^{(N-2)}(\mathbf{U}^{(N-3)}))}{\partial \mathbf{V}^{(N-3)}} \\
&\quad \cdots \frac{\partial \mathbf{G}^{(2)}(\mathbf{U}^{(1)})}{\partial \mathbf{V}^{(1)}} \frac{\partial \mathbf{G}^{(1)}(\mathbf{U}^{(0)})}{\partial \mathbf{U}^{(0)}} \\
&= \mathbf{G}^{(N)'}\mathbf{G}^{(N-1)'}\mathbf{G}^{(N-2)'} \cdots \mathbf{G}^{(2)'}\mathbf{G}^{(1)'}.
\end{aligned} \tag{44}$$

Hence (42) can be rewritten as

$$\delta\mathbf{V}^{(N)} = \mathbf{G}'\delta\mathbf{U}^{(0)} = \mathbf{G}^{(N)'}\mathbf{G}^{(N-1)'}\mathbf{G}^{(N-2)'} \cdots \mathbf{G}^{(2)'}\mathbf{G}^{(1)'}\delta\mathbf{U}^{(0)}. \tag{45}$$

An elementary result of matrix algebra is that the transpose of a product of matrices is the product of the transposes of the factor matrices, taken in reversed order

$$\mathbf{G}'^{*} = \mathbf{G}^{(1)'*}\mathbf{G}^{(2)'*} \cdots \mathbf{G}^{(N-2)'*}\mathbf{G}^{(N-1)'*}\mathbf{G}^{(N)'*}, \tag{46}$$

which again shows the reversed character of the adjoint computations (Talagrand 1991). Hence

$$\begin{aligned}
\nabla_{\mathbf{U}} J &= \mathbf{G}'^{*}\nabla_{\mathbf{V}} J \\
&= \mathbf{G}^{(1)'*}\mathbf{G}^{(2)'*} \cdots \mathbf{G}^{(N-2)'*}\mathbf{G}^{(N-1)'*}\mathbf{G}^{(N)'*}\nabla_{\mathbf{V}} J.
\end{aligned} \tag{47}$$

Equation (47) suggests a useful and systematic way for developing the adjoint of a given code. First, develop the adjoints $\mathbf{G}^{(i)'*}$ of the basic code from the forward Jacobians $\mathbf{G}^{(i)'}$; then combine the various $\mathbf{G}^{(i)'*}$ in reversed order (e.g. see Giering and Kaminski 1998).

## 3. APPLICATION TO DATA ASSIMILATION

### (*a*) *Variational approach I*

There are generally two approaches to applying adjoint methods to variational data assimilation (Daley 1991; Fisher and Lary 1995). The first approach, in terms of Lagrange multipliers, is developed in this section following Daley (1991). The second approach, using the theory of adjoint operators, is developed in the next section following Talagrand and Courtier (1987) and Elbern *et al.* (1997). We note here that the major difference between these two approaches is in the treatment and identification of the gradient of $J$.

Let $H^{(i)}$ be a scalar-valued function which measures the distance between model prediction $\mathbf{V}^{(i)}$ and observations $\widehat{\mathbf{V}}^{(i)}$ available at time step $i$. Hence

$$H^{(i)} = (\widehat{\mathbf{V}}^{(i)} - \mathbf{V}^{(i)})^2, \tag{48}$$

$$J = \frac{1}{2}\sum_{i=0}^{N} w^{(i)} H^{(i)} = \frac{1}{2} w^{(0)} H^{(0)} + \frac{1}{2}\sum_{i=1}^{N} w^{(i)} H^{(i)}, \tag{49}$$

where $w^{(i)}$ is a specific weighting function, which represents the analysis errors of the observations. Following (26), the model forecasts at time step $i$ is given by

$$\mathbf{V}^{(i)} = \mathbf{G}^{(i)}(\mathbf{U}^{(i-1)}). \tag{50}$$

So the problem for atmospheric chemistry data assimilation can be succinctly described as follows: consider the minimization of (49) subject to the strong constraint (50). Equation (50) represents chemistry models which were constructed based on current photochemistry theories. Equation (49) represents the misfits (cost) between model forecasts (50) and observations. By imposing the minimization of (49) under the strong constraint of the model (50), we obtain the best model analysis of atmospheric chemistry.

From (49), (44), (46), and (41), and introducing the following notations as defined from (38)

$$\mathbf{G}^{N} = \mathbf{G}^{(N)} \circ \mathbf{G}^{(N-1)} \circ \cdots \circ \mathbf{G}^{(2)} \circ \mathbf{G}^{(1)}, \tag{51}$$

$$\mathbf{G}^{N'} = \mathbf{G}^{(N)'} \circ \mathbf{G}^{(N-1)'} \circ \cdots \circ \mathbf{G}^{(2)'} \circ \mathbf{G}^{(1)'}, \tag{52}$$

$$\mathbf{G}^{N'*} = \mathbf{G}^{(1)'*} \circ \mathbf{G}^{(2)'*} \circ \cdots \circ \mathbf{G}^{(N-1)'*} \circ \mathbf{G}^{(N)'*}, \tag{53}$$

the variation on $J$ can be written as

$$\begin{aligned}
\delta J &= -\sum_{i=0}^{N} w^{(i)}(\widehat{\mathbf{V}}^{(i)} - \mathbf{V}^{(i)})\delta\mathbf{V}^{(i)} \\
&= -w^{(0)}(\widehat{\mathbf{V}}^{(0)} - \mathbf{V}^{(0)t})\delta\mathbf{V}^{(0)} - \sum_{i=1}^{N} w^{(i)}(\widehat{\mathbf{V}}^{(i)} - \mathbf{V}^{(i)})\mathbf{G}^{i'}\delta\mathbf{U}^{(0)} \\
&= -w^{(0)}(\widehat{\mathbf{V}}^{(0)} - \mathbf{V}^{(0)})\delta\mathbf{V}^{(0)} - \sum_{i=1}^{N} \mathbf{G}^{i'*} w^{(i)}(\widehat{\mathbf{V}}^{(i)} - \mathbf{V}^{(i)})\delta\mathbf{U}^{(0)}.
\end{aligned} \tag{54}$$

Hence the gradient of $J$ with respect to $\mathbf{U}^{(0)}$ is

$$\begin{aligned}\nabla_{\mathbf{U}^{(0)}} J &= -w^{(0)}(\widehat{\mathbf{V}}^{(0)} - \mathbf{V}^{(0)}) - \sum_{i=1}^{N} \mathbf{G}^{i'*} w^{(i)}(\widehat{\mathbf{V}}^{(i)} - \mathbf{V}^{(i)}) \\ &= -w^{(0)}(\widehat{\mathbf{V}}^{(0)} - \mathbf{V}^{(0)}) - \mathbf{G}^{1'*} w^{(1)}(\widehat{\mathbf{V}}^{(1)} - \mathbf{V}^{(1)}) \\ &\quad \cdots - \mathbf{G}^{N'*} w^{(N)}(\widehat{\mathbf{V}}^{(N)} - \mathbf{V}^{(N)}). \end{aligned} \tag{55}$$

Consider the new function

$$\begin{aligned} I &= J + \sum_{i=1}^{N} \lambda^{(i)}(\mathbf{V}^{(i)} - \mathbf{G}^{(i)}(\mathbf{U}^{(i-1)})) \\ &= \frac{1}{2} w^{(0)} H^{(0)} + \sum_{i=1}^{N} \left\{ \frac{1}{2} w^{(i)}(\widehat{\mathbf{V}}^{(i)} - \mathbf{V}^{(i)})^2 + \lambda^{(i)}(\mathbf{V}^{(i)} - \mathbf{G}^{i}(\mathbf{U}^{(i-1)})) \right\}. \end{aligned} \tag{56}$$

The variation of the function $I$ can be written as

$$\begin{aligned} \delta I &= -w^{(0)}(\widehat{\mathbf{V}}^{(0)} - \mathbf{V}^{(0)})\delta\mathbf{V}^{(0)} + \sum_{i=1}^{N} \{ -w^{(i)}(\widehat{\mathbf{V}}^{(i)} - \mathbf{V}^{(i)})\delta\mathbf{V}^{(i)} \\ &\quad + \delta\lambda^{(i)}(\mathbf{V}^{(i)} - \mathbf{G}^{(i)}(\mathbf{U}^{(i-1)})) + \lambda^{(i)}(\delta\mathbf{V}^{(i)} - \mathbf{G}^{(i)'}(\mathbf{U}^{(i-1)})\delta\mathbf{V}^{(i-1)}) \}. \end{aligned} \tag{57}$$

Notice that (41) has been used to obtain the last line of (57). The minimum value of $J$ and $I$ occurs at $\delta I = 0$, which implies that the variations on both $\lambda$ and $\mathbf{V}$ are zero. While the former gives

$$\mathbf{V}^{(i)} - \mathbf{G}^{(i)}(\mathbf{U}^{(i-1)}) = 0, \tag{58}$$

which indicates that variation on $\lambda$ ($\delta\lambda^{(i)}$) yields $N$ constraints of (50); the later variation on $\mathbf{V}$ ($\delta\mathbf{V}^{(i)}$) will give us the following important property. If we expand the following terms with respect to the variations on $\mathbf{V}^{(i)}$,

$$\begin{aligned} &- w^{(0)}(\widehat{\mathbf{V}}^{(0)} - \mathbf{V}^{(0)})\delta\mathbf{V}^{(0)} + \sum_{i=1}^{N} \{ -w^{(i)}(\widehat{\mathbf{V}}^{(i)} - \mathbf{V}^{(i)})\delta\mathbf{V}^{(i)} \\ &\quad + \lambda^{(i)}(\delta\mathbf{V}^{(i)} - \mathbf{G}^{(i)'}(\mathbf{U}^{(i-1)})\delta\mathbf{V}^{(i-1)}) \} \\ &= -w^{(0)}(\widehat{\mathbf{V}}^{(0)} - \mathbf{V}^{(0)})\delta\mathbf{V}^{(0)} - w^{(1)}(\widehat{\mathbf{V}}^{(1)} - \mathbf{V}^{(1)})\delta\mathbf{V}^{(1)} + \lambda^{(1)}(\delta\mathbf{V}^{(1)} - \mathbf{G}^{(1)'}\delta\mathbf{V}^{(0)}) \\ &\quad \cdots - w^{(N)}(\widehat{\mathbf{V}}^{(N)} - \mathbf{V}^{(N)})\delta\mathbf{V}^{(N)} + \lambda^{(N)}(\delta\mathbf{V}^{(N)} - \mathbf{G}^{(N)'}\delta\mathbf{V}^{(N)}) \\ &= \delta\mathbf{V}^{(0)}\{ -w^{(0)}(\widehat{\mathbf{V}}^{(0)} - \mathbf{V}^{(0)}) - \lambda^{(1)}\mathbf{G}^{(1)'} + \underline{\lambda^{(0)}} \} \\ &\quad + \delta\mathbf{V}^{(1)}\{ -w^{(1)}(\widehat{\mathbf{V}}^{(1)} - \mathbf{V}^{(1)}) - \lambda^{(2)}\mathbf{G}^{(2)'} + \lambda^{(1)} \} \\ &\quad \cdots + \delta\mathbf{V}^{(N)}\{ -w^{(N)}(\widehat{\mathbf{V}}^{(N)} - \mathbf{V}^{(N)}) - \underline{\lambda^{(N+1)}\mathbf{G}^{(N+1)'}} + \lambda^{(N)} \} \\ &= 0. \end{aligned} \tag{59}$$

Hence, for each $n$ ($n = 0, \ldots, N$),

$$\lambda^{(n)} = \lambda^{(n+1)}\mathbf{G}^{(n+1)'} + w^{(n)}(\widehat{\mathbf{V}}^{(n)} - \mathbf{V}^{(n)}), \tag{60}$$

with two introduced natural boundary conditions (the terms underlined): $\lambda^{(0)} = \lambda^{(N+1)} = 0$. Hence,

$$\left.\begin{aligned}
\lambda^{(N)} &= \lambda^{(N+1)}\mathbf{G}^{(N+1)'} + w^{(N)}(\hat{\mathbf{V}}^{(N)} - \mathbf{V}^{(N)}) \\
&= w^{(N)}(\hat{\mathbf{V}}^{(N)} - \mathbf{V}^{(N)}), \\
\lambda^{(N-1)} &= \lambda^{(N)}\mathbf{G}^{(N)'} + w^{(N-1)}(\hat{\mathbf{V}}^{(N-1)} - \mathbf{V}^{(N-1)}) \\
&= w^{(N)}(\hat{\mathbf{V}}^{(N)} - \mathbf{V}^{(N)})\mathbf{G}^{(N)'} + w^{(N-1)}(\hat{\mathbf{V}}^{(N-1)} - \mathbf{V}^{(N-1)}), \\
&\vdots \quad \vdots \\
\lambda^{(0)} &= \lambda^{(1)}\mathbf{G}^{(1)'} + w^{(0)}(\hat{\mathbf{V}}^{(0)} - \mathbf{V}^{(0)}) \\
&= w^{(N)}(\hat{\mathbf{V}}^{(N)} - \mathbf{V}^{(N)})\mathbf{G}^{(N)'}\mathbf{G}^{(N-1)'}\mathbf{G}^{(N-2)'} \cdots \mathbf{G}^{(1)'} \\
&\quad + w^{(N-1)}(\hat{\mathbf{V}}^{(N-1)} - \mathbf{V}^{(N-1)})\mathbf{G}^{(N-1)'}\mathbf{G}^{(N-2)'} \cdots \mathbf{G}^{(2)'} \\
&\quad \cdots + w^{(1)}(\hat{\mathbf{V}}^{(1)} - \mathbf{V}^{(1)})\mathbf{G}^{(1)'} + w^{(0)}(\hat{\mathbf{V}}^{(0)} - \mathbf{V}^{(0)}).
\end{aligned}\right\} \tag{61}$$

Hence

$$\begin{aligned}
\lambda^{(0)} = {} & \mathbf{G}^{N'*}w^{(N)}(\hat{\mathbf{V}}^{(N)} - \mathbf{V}^{(N)}) + \mathbf{G}^{N-1'*}w^{(N-1)}(\hat{\mathbf{V}}^{(N-1)} - \mathbf{V}^{(N-1)}) \\
& \cdots + \mathbf{G}^{2'*}w^{(2)}(\hat{\mathbf{V}}^{(2)} - \mathbf{V}^{(2)}) + \mathbf{G}^{1'*}w^{(1)}(\hat{\mathbf{V}}^{(1)} - \mathbf{V}^{(1)}) + w^{(0)}(\hat{\mathbf{V}}^{(0)} - \mathbf{V}^{(0)}).
\end{aligned} \tag{62}$$

Notice that the property of $\mathbf{G}^{N'}$ has been used to derived (62). Therefore, with (55),

$$\nabla_{\mathbf{U}^{(0)}} J = -\lambda^{(0)}. \tag{63}$$

We note that if $\lambda^{(0)}$ is zero, then $\nabla_{\mathbf{U}^{(0)}} J$ is zero. The backward integration of the inhomogeneous adjoint tangent linear equation (60) from $\lambda^{(N+1)} = 0$ to $\lambda^{(0)}$ gives the desired gradient of $J$ with respect to the initial condition $\mathbf{U}^{(0)}$. The overall time coordinates of the integrations can be summarized as

$$\begin{array}{ccccccccccccc}
0 & & \Delta t & & 2\Delta t & & \cdots & & (N-2)\Delta t & & (N-1)\Delta t & & N\Delta t = T \\
0 & \xrightarrow{(1)} & 1 & \xrightarrow{(2)} & 2 & \xrightarrow{(3)} & \cdots & \xrightarrow{(N-2)} & N-2 & \xrightarrow{(N-1)} & N-1 & \xrightarrow{(N)} & N \\
\Uparrow & & & & & & & & & & & & \Downarrow \\
1 & \xleftarrow{(1)} & 2 & \xleftarrow{(2)} & 3 & \xleftarrow{(3)} & \cdots & \xleftarrow{(N-2)} & N-1 & \xleftarrow{(N-1)} & N & \xleftarrow{(N)} & N+1
\end{array} \tag{64}$$

There are $N$ forward integrations of (50), followed by $N$ backward integrations of (60) with imposed natural boundary conditions at both $\lambda^{(N+1)} = 0$ and $\lambda^{(0)} = 0$.

### (*b*) *Variational approach II*

Let $H$ be the scalar-valued distance function, and

$$J = \int_0^T H(\mathbf{U})\, \mathrm{d}t. \tag{65}$$

The variation of $J$ with respect to $\mathbf{U}$ is written as

$$\delta J = \int_0^T \langle \nabla_{\mathbf{U}} H, \delta\mathbf{U} \rangle\, \mathrm{d}t. \tag{66}$$

Now define the new function $I$ as

$$I = \int_0^T \left( H + \left\langle \lambda, \frac{\mathrm{d}\mathbf{U}}{\mathrm{d}t} - F(\mathbf{U}) \right\rangle \right) \mathrm{d}t, \tag{67}$$

where $\lambda$ is the Lagrange multiplier, and the constraint model is

$$\frac{\mathrm{d}\mathbf{U}}{\mathrm{d}t} = F(\mathbf{U}). \tag{68}$$

The variation of $I$ is

$$\delta I = \int_0^T \left( \langle \nabla_{\mathbf{U}} H, \delta\mathbf{U} \rangle + \left\langle \delta\lambda, \frac{\mathrm{d}\mathbf{U}}{\mathrm{d}t} - F(\mathbf{U}) \right\rangle + \left\langle \lambda, \frac{\mathrm{d}\delta\mathbf{U}}{\mathrm{d}t} - F'\delta\mathbf{U} \right\rangle \right) \mathrm{d}t. \tag{69}$$

While the variation on $\lambda$ yields the constraint model (68), the variation on $\mathbf{U}$, which follows from

$$\begin{aligned}
\int_0^T \left\langle \lambda, \frac{\mathrm{d}\delta\mathbf{U}}{\mathrm{d}t} - F'\delta\mathbf{U} \right\rangle \mathrm{d}t &= \int_0^T \left( \left\langle \lambda, \frac{\mathrm{d}\delta\mathbf{U}}{\mathrm{d}t} \right\rangle - \int_0^T \langle \lambda, F'\delta\mathbf{U} \rangle \right) \mathrm{d}t \\
&= \int_0^T \lambda \, \mathrm{d}\delta\mathbf{U} - \int_0^T \langle F'^*\lambda, \delta\mathbf{U} \rangle \, \mathrm{d}t \\
&= \lambda\delta\mathbf{U}|_0^T - \int_0^T \left\langle \frac{\mathrm{d}\lambda}{\mathrm{d}t}, \delta\mathbf{U} \right\rangle \mathrm{d}t - \int_0^T \langle F'^*\lambda, \delta\mathbf{U} \rangle \, \mathrm{d}t \\
&= \lambda(T)\delta\mathbf{U}(T) - \lambda(0)\delta\mathbf{U}(0) - \int_0^T \left\langle \frac{\mathrm{d}\lambda}{\mathrm{d}t}, \delta\mathbf{U} \right\rangle \mathrm{d}t \\
&\quad - \int_0^T \langle F'^*\lambda, \delta\mathbf{U} \rangle \, \mathrm{d}t,
\end{aligned} \tag{70}$$

and the first term of the right-hand side of (69), yields the following inhomogeneous adjoint tangent linear equation:

$$-\frac{\mathrm{d}\lambda}{\mathrm{d}t} - F'^*\lambda + \nabla_{\mathbf{U}} H = 0, \tag{71}$$

and the natural boundary conditions $\lambda(0) = \lambda(T) = 0$. Equation (71) can also be written as

$$\frac{\mathrm{d}\lambda}{\mathrm{d}(-t)} = F'^*\lambda - \nabla_{\mathbf{U}} H, \tag{72}$$

which is instructive when compared with the tangent linear equation

$$\frac{\mathrm{d}\delta\mathbf{U}}{\mathrm{d}t} = F'\delta\mathbf{U}. \tag{73}$$

Both (72) and (73) are linear models. But while the later is integrated with respect to the positive increment of time (forwards in time), the former is integrated with respect to the negative increment of time (backwards in time).

Following (45) and (41), the solution to (73) is written as

$$\delta\mathbf{U}^{(n)} = \mathbf{R}(t_n, t_0)\delta\mathbf{U}^{(0)}, \tag{74}$$

where

$$\mathbf{R}(t_n, t_0) = \mathbf{G}^{(n)'}\mathbf{G}^{(n-1)'}\mathbf{G}^{(n-2)'} \cdots \mathbf{G}^{(2)'}\mathbf{G}^{(1)'}$$
$$= \tilde{\mathbf{R}}(t_n, t_{n-1})\tilde{\mathbf{R}}(t_{n-1}, t_{n-2}) \cdots \tilde{\mathbf{R}}(t_2, t_1)\tilde{\mathbf{R}}(t_1, t_0). \tag{75}$$

For every $n$, $\mathbf{G}^{(n)'} = \tilde{\mathbf{R}}(t_n, t_{n-1})$. Here $\mathbf{R}(t_n, t_0)$ is a well-defined linear operator, and is called by Talagrand and Courtier (1987) the resolvent of (73) between times $t_0$ and $t_n$. If set $n = 0$, then (74) gives

$$\delta\mathbf{U}^{(0)} = \mathbf{R}(t_0, t_0)\delta\mathbf{U}^{(0)}, \tag{76}$$

or, in general, for any $t$,

$$\mathbf{R}(t, t) = I. \tag{77}$$

From (74), we can write

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta\mathbf{U} = \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{R}(t, t_0)\delta\mathbf{U}^{(0)}. \tag{78}$$

From (73) and (74), we can also write

$$\frac{\mathrm{d}\delta\mathbf{U}}{\mathrm{d}t} = F'\delta\mathbf{U} = F'\mathbf{R}(t, t_0)\delta\mathbf{U}^{(0)}. \tag{79}$$

Hence the comparison between (78) and (79) shows

$$\frac{\partial\mathbf{R}(t, t_0)}{\partial t} = F'\mathbf{R}(t, t_0). \tag{80}$$

Both (77) and (80) are two important properties for $\mathbf{R}(t, t_0)$.

Now let $\mathbf{S}(t_{n-1}, t_n)$ be the resolvent of the linear adjoint equation

$$-\frac{\mathrm{d}\lambda}{\mathrm{d}t} = F'^{*}\lambda \tag{81}$$

between times $t_n$ and $t_{n-1}$. This means that the solution to (81) can be written as

$$\lambda(t_0) = \mathbf{S}(t_0, t_N)\lambda(t_N)$$
$$= \tilde{\mathbf{S}}(t_0, t_1)\tilde{\mathbf{S}}(t_1, t_2) \cdots \tilde{\mathbf{S}}(t_{N-2}, t_{N-1})\tilde{\mathbf{S}}(t_{N-1}, t_N)\lambda(t_N). \tag{82}$$

For a solution $\delta\mathbf{U}$ of (73) and a solution $\lambda$ of (81) at time $t$, their inner product

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle\delta\mathbf{U}, \lambda\rangle = \left\langle\frac{\mathrm{d}}{\mathrm{d}t}\delta\mathbf{U}, \lambda\right\rangle + \left\langle\delta\mathbf{U}, \frac{\mathrm{d}}{\mathrm{d}t}\lambda\right\rangle$$
$$= \langle F'\delta\mathbf{U}, \lambda\rangle + \langle\delta\mathbf{U}, -F'^{*}\lambda\rangle$$
$$= 0. \tag{83}$$

Hence, the inner product is constant with time. Let $y$ be the solution of the forward equation at time $t$, and $y'$ be the solution of the backward equation at time $t'$ ($t' > t$), observing the following relationship

| | Time $t$ | Time $t'$ |
|---|---|---|
| Solution to $\mathrm{d}\delta\mathbf{U}/\mathrm{d}t = F'\delta\mathbf{U}$ | $y$ | $\mathbf{R}(t', t)y$ |
| Solution to $-\mathrm{d}\lambda/\mathrm{d}t = F'^{*}\lambda$ | $\mathbf{S}(t, t')y'$ | $y'$ |
| Inner product | $\langle y, \mathbf{S}(t, t')y'\rangle$ | $\langle\mathbf{R}(t', t)y, y\rangle$ |

(84)

Equation (83) ensures that the inner product of $\langle y, \mathbf{S}(t, t')y' \rangle$ at time $t$ is equal to the inner product of $\langle \mathbf{R}(t', t)y, y' \rangle$ at time $t'$. Hence

$$\begin{aligned} \langle \mathbf{R}(t', t)y, y' \rangle &= \langle y, \mathbf{S}(t, t')y' \rangle \\ &= \langle y, \mathbf{R}^*(t', t)y' \rangle. \end{aligned} \tag{85}$$

Hence, $\mathbf{S} = \mathbf{R}^*$, and (66) can be rewritten as

$$\begin{aligned} \delta J &= \int_0^T \langle \nabla_{\mathbf{U}} H, \delta \mathbf{U} \rangle \, \mathrm{d}t \\ &= \int_0^T \langle \nabla_{\mathbf{U}} H, \mathbf{R}(t, t_0)\delta \mathbf{U}^{(0)} \rangle \, \mathrm{d}t \\ &= \left\langle \int_0^T \mathbf{R}^*(t, t_0)\nabla_{\mathbf{U}} H \, \mathrm{d}t, \delta \mathbf{U}^{(0)} \right\rangle \\ &= \left\langle \int_0^T \mathbf{S}(t_0, t)\nabla_{\mathbf{U}} H \, \mathrm{d}t, \delta \mathbf{U}^{(0)} \right\rangle. \end{aligned} \tag{86}$$

The exchange of the operation of integration and inner product is allowed given that $\delta \mathbf{U}^{(0)}$ is not dependent on $t$. Therefore, the gradient of $J$ with respect to $\mathbf{U}^{(0)}$ is

$$\nabla_{\mathbf{U}^{(0)}} J = \int_0^T \mathbf{R}^*(t, t_0)\nabla_{\mathbf{U}} H \, \mathrm{d}t = \int_0^T \mathbf{S}(t_0, t)\nabla_{\mathbf{U}} H \, \mathrm{d}t = \int_{t_0}^{t_N} \mathbf{S}(t_0, t)\nabla_{\mathbf{U}} H \, \mathrm{d}t, \tag{87}$$

where at the last term of (87) we have set time $t = 0 = t_0$ and $t = T = t_N$. The properties of $\mathbf{S}(t, t_N)$ are given by

$$\mathbf{S}(t, t) = I, \tag{88}$$

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\lambda(t) &= \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{S}(t, t_N)\lambda(t_N) \\ &= -F'^*\lambda(t) \\ &= -F'^*\mathbf{S}(t, t_N)\lambda(t_N), \end{aligned} \tag{89}$$

$$\frac{\partial}{\partial t}\mathbf{S}(t, t_N), = -F'^*\mathbf{S}(t, t_N).$$

Now consider the solution to the inhomogeneous adjoint tangent linear equation (71), which is written as

$$\lambda(t) = \int_{t_N}^{t} \mathbf{S}(t, \tau)\nabla_{\mathbf{U}} H(\tau) \, \mathrm{d}\tau. \tag{90}$$

Before we set out to verify (90), let's have a close look at equations (71), (81), and (89). The solution to the inhomogeneous equation (71) is

$$\lambda = \lambda_{\mathrm{h}} + \lambda_{\mathrm{p}}, \tag{91}$$

where $\lambda_{\mathrm{h}}$ is the solution to the homogeneous equation (81), and $\lambda_{\mathrm{p}}$ is a particular solution to the inhomogeneous equation (71). For every $t'(t < t' < t_N)$, if

$$\lambda_{\mathrm{h}}(t) = \mathbf{S}(t, t')\nabla_{\mathbf{U}(t')} H(t'), \tag{92}$$

then

$$\frac{\mathrm{d}}{\mathrm{d}t}\lambda_{\mathrm{h}}(t) = -F'^{*}\mathbf{S}(t, t')H(t') = -F'^{*}\lambda_{\mathrm{h}}(t). \tag{93}$$

Hence, (92) is the general solution to (81). Since **S** is a linear operator, let's consider the following summation

$$\begin{aligned}\sum_{i=t_N}^{t} \mathbf{S}(t, i)\nabla_{\mathbf{U}}H(i) &= \mathbf{S}(t, t)\nabla_{\mathbf{U}}H(t) + \mathbf{S}(t, t+1)\nabla_{\mathbf{U}}H(t+1) \\ &\quad + \mathbf{S}(t, t+2)\nabla_{\mathbf{U}}H(t+2) \\ &\quad \cdots + \mathbf{S}(t, t_N - 1)\nabla_{\mathbf{U}}H(t_N - 1) + \mathbf{S}(t, t_N)\nabla_{\mathbf{U}}H(t_N) \\ &= \nabla_{\mathbf{U}}H(t) + \lambda_h, \end{aligned} \tag{94}$$

were $\lambda_{\mathrm{h}}$ is from (92). Hence, $\lambda_{\mathrm{p}}$ has a form relating to $\nabla_{\mathbf{U}}H(t)$. Now, we write (94) in a continuous form as

$$\int_{t_N}^{t} \mathbf{S}(t, \tau)\nabla_{\mathbf{U}(\tau)}H(\tau)\,\mathrm{d}\tau = \lambda(t). \tag{95}$$

Following Boas (1983), the differentiation of an integral, following *Leibniz's rule*

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{u(t)}^{v(t)} f(t, \tau)\,\mathrm{d}\tau = f(t, v)\frac{\mathrm{d}v}{\mathrm{d}t} - f(t, u)\frac{\mathrm{d}u}{\mathrm{d}t} + \int_{u}^{v}\frac{\partial f}{\partial t}\,\mathrm{d}\tau. \tag{96}$$

Let

$$\begin{aligned} f(t, \tau) &= \mathbf{S}(t, \tau)\nabla_{\mathbf{U}(\tau)}H(\tau), \\ v(t) &= t, \\ u(t) &= t_N. \end{aligned} \tag{97}$$

Then

$$\left.\begin{aligned} f(t, v)\frac{\mathrm{d}v}{\mathrm{d}t} &= \mathbf{S}(t, t)\nabla_{\mathbf{U}(t)}H(t) \\ &= \nabla_{\mathbf{U}(t)}H(t), \\ f(t, u)\frac{\mathrm{d}u}{\mathrm{d}t} &= 0, \\ \int_{u}^{v}\frac{\partial f}{\partial t}\,\mathrm{d}\tau &= -F'^{*}\int_{u}^{v}\mathbf{S}(t, \tau)\nabla_{\mathbf{U}(\tau)}H(\tau)\,\mathrm{d}\tau. \end{aligned}\right\} \tag{98}$$

Here (89) is used to derived the last line of (98). Hence,

$$\frac{\mathrm{d}\lambda}{\mathrm{d}t} = \nabla_{\mathbf{U}(t)}H(t) - F'^{*}(t')\lambda(t). \tag{99}$$

Here we have demonstrated that, with the boundary condition $\lambda(t_N) = 0$, (90) is the solution of the inhomogeneous adjoint tangent linear equation (71). Most importantly, if we set $t = t_0$, then (90) is

$$\lambda(t_0) = \int_{t_N}^{t_0}\mathbf{S}(t, \tau)\nabla_{\mathbf{U}}H(\tau)\,\mathrm{d}\tau. \tag{100}$$

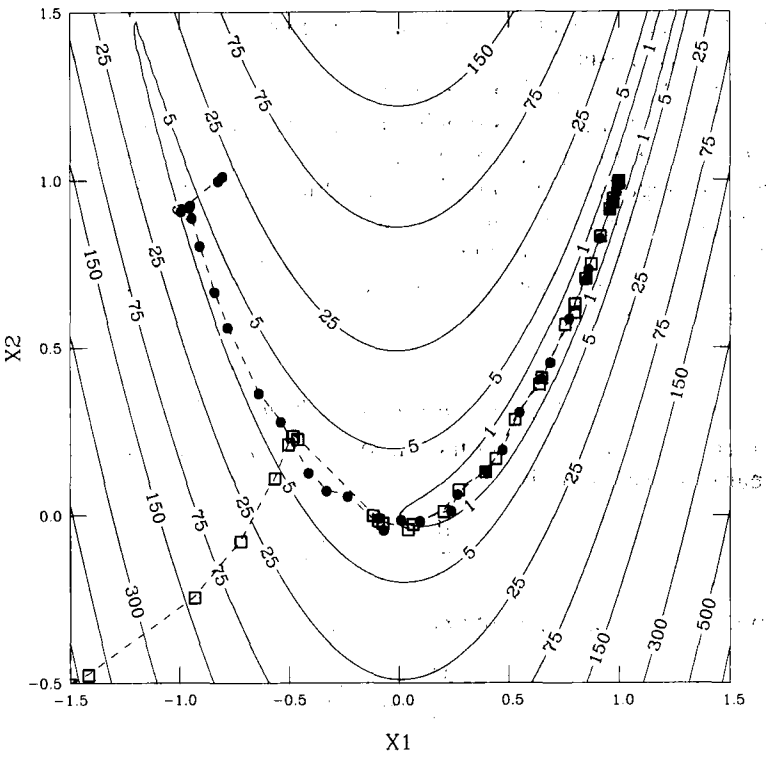


Figure 2. Contours of the two-dimensional Rosenbrock function with respect to two variables $x_1$ and $x_2$. Also shown on the plot are two steepest descent minimization paths. For the path denoted by solid circles, the steepest descent starts from (−0.8, 1.0); while for the path denoted by open squares, the steepest descent starts from (−1.5, −0.5). Both paths converge to (1, 1) of the minimum point.

Hence, when compared with (87), it follows that

$$\nabla_{\mathbf{U}^{(0)}} J = -\lambda(t_0). \tag{101}$$

We note that if $\lambda(t_0)$ is zero, then $\nabla_{\mathbf{U}^{(0)}} J$ is zero. Equation (101) now shows that the desired gradient of $J$ with respect to $\mathbf{U}^{(0)}$ is equal to the negative of the backward integration of the inhomogeneous adjoint tangent linear equation from boundary condition $\lambda(t_N) = 0$ to $\lambda(t_0)$. This result is consistent with the result described in the previous section using Lagrange multipliers.

### (c) *Minimization procedures*

Once the gradient of the cost function $J$ is found, a steepest descent algorithm can be used to find its minimum value (Daley 1992). Figure 2 shows contours of a Rosenbrock function,

$$f(x_1, x_2) = (1 - x_2)^2 + 100(x_2 - x_1^2)^2, \tag{102}$$

and two steepest descent paths. The paths are calculated based on the descent algorithm M1QN3 of Gilbert and Lemarechal (1989), which used special Broyden–Fletcher–Goldfarb–Shanno method matrices (Nocedal 1980) in updating quasi-Newton matrices for the calculation of the Hessian at each iteration. The advantage of this method is that the large-scale problem (with many control variables) can be employed on the platform

with limited storage. It is clear that both paths efficiently converge to the minimum point of (1, 1) where $f_{\min}$ exists.

Since chemical concentrations can vary over many orders of magnitude, a proper scaling of each control variable is crucial to the success of the minimization procedure (Navon *et al.* 1992). From a geometrical point of view, if the constant-cost contours are circular, then the negative gradient of the cost function will point radially towards the minimum, and only one descent iteration will be needed. In practice, the cost function with respect to the original coordinates of control variables is multiplied by a scaling matrix, which is chosen so that the resulting vector points more nearly towards the minimum (Thacker 1989).

### (*d*) *Implementation of a chemical 4D-Var*

Here we conclude the discussion of adjoint theory and outline its implementation. The nonlinear forecast model is

$$\frac{\mathrm{d}\mathbf{U}}{\mathrm{d}t} = F(\mathbf{U}). \tag{103}$$

The corresponding forward tangent linear model is written as

$$\frac{\mathrm{d}\delta\mathbf{U}}{\mathrm{d}t} = F'\delta\mathbf{U}. \tag{104}$$

The backward inhomogeneous adjoint tangent linear model is then formulated as

$$\frac{\mathrm{d}\lambda}{\mathrm{d}(-t)} = F'^{*}\lambda - \nabla_{\mathbf{U}}H, \tag{105}$$

with the boundary conditions

$$\lambda(0) = \lambda(T) = 0. \tag{106}$$

Here $H$ is the scalar-valued distance function (section 3(b)). The discretized form of the adjoint model is written as

$$\lambda^{(n)} = \lambda^{(n+1)}\mathbf{G}^{(n+1)'} + w^{(n)}(\hat{\mathbf{V}}^{(n)} - \mathbf{V}^{(n)}). \tag{107}$$

Finally, a steepest descent minimization algorithm (as described in section 3(c)) is used to find the value which minimizes $J$, the distance between model prediction and observation.

The implementation of a chemical 4-D var is described as follows. The $\nabla J(\mathbf{U}_{t_0})$ can be obtained, for a given $\mathbf{U}_{t_0}$, by performing the following operations:

(i) Starting from $\mathbf{U}_{t_0}$ at time $t_0$, integrate (103) forwards in time from $t = t_0$ to $t = T$. Stores the values thus computed for $\mathbf{U}(t)$, $t_0 \leqslant t \leqslant T$.

(ii) Starting from $\lambda(t = T) = 0$, integrate (107) backwards in time from $t = T$ to $t = t_0$. The adjoint operator $\mathbf{G}'^{*}(t)$ and the gradient $\nabla_{\mathbf{U}}J(t)$ are determined, at each $t$, from the values $\mathbf{U}(t)$ computed in the direct forward integration of equation.

(iii) The final value of $\lambda(t = t_0)$ is the negative of the required gradient $\nabla J_{\mathbf{U}_{t_0}}$. Notice that if $\lambda(t = t_0)$ is zero, then $\nabla J_{\mathbf{U}_{t_0}}$ is zero.

(iv) Use a descent algorithm to determine the value $\mathbf{U}_{\min}$ which minimizes $J$.

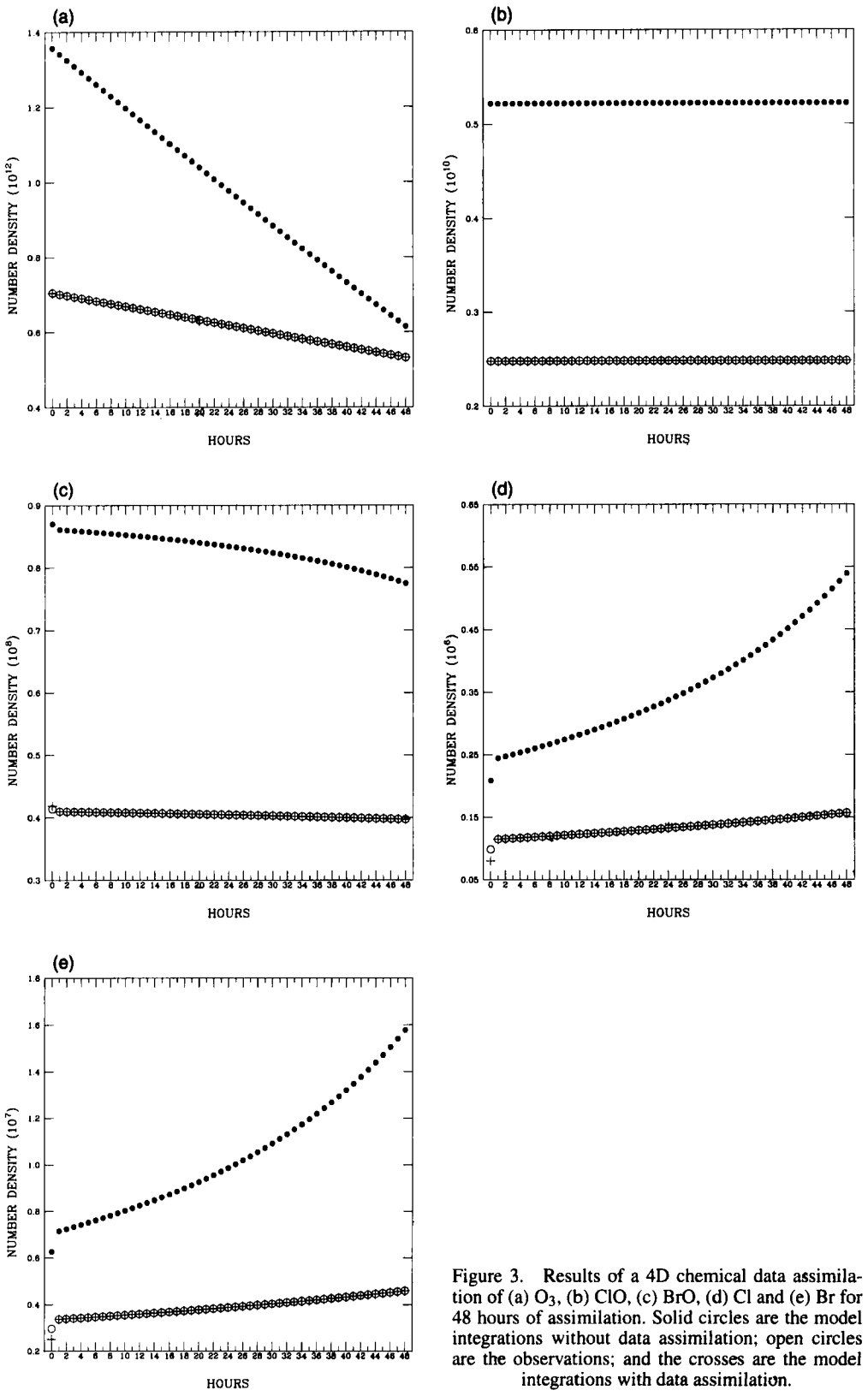


Figure 3. Results of a 4D chemical data assimilation of (a) $O_3$, (b) ClO, (c) BrO, (d) Cl and (e) Br for 48 hours of assimilation. Solid circles are the model integrations without data assimilation; open circles are the observations; and the crosses are the model integrations with data assimilation.

## 4. APPLICATION TO A CHEMISTRY 4D-VAR

Following the formulation and implementation procedures described previously, a chemical data assimilation has been performed using the catalytic ozone destruction mechanism described in section 2(e). A stiff ordinary differential equation integrator (Brown *et al.* 1989) was used to integrate (24), with rate coefficients taken from De More *et al.* (1997). The model was first run with initial conditions close to the typical stratospheric concentrations found at a height of 25 km. A 48-hour model integration was performed and the output stored as the observations. A second integration was then performed where the initial conditions used were double those of the first integration, and the output treated as the modelled results. Finally a third integration was performed using the modelled output from integration 2 and the observations obtained from integration 1. Here the integration used the iterative variational data assimilation method to minimize the difference between initial modelled values and observations.

Figure 3 shows the results from integrations (1–3). The model—integration 2—shows that the high ozone concentrations have been significantly reduced to less than 50% of their initial value after two days of catalytic ozone destruction. During this period of time, the radicals ClO and BrO have remained relatively constant, whereas Cl and Br atoms have risen dramatically as the availability of ozone decreases.

For integration 1, where initial ozone levels were half those of integration 2, the rate of ozone loss is lower, but still a considerable amount (30%) is destroyed during the integration. Comparing integrations 1 and 2 confirms the earlier work of Crutzen *et al.* (1995) that higher ozone levels leads to increased catalytic loss. It should be noted of course that in these integrations gas-phase chemistry only has been considered. If heterogeneous chemistry (i.e. polar stratospheric clouds) had been included these would have led to increased levels of chlorine and bromine in the system and therefore increased loss of ozone.

Integration 3 uses data assimilation (crosses) to minimize the difference between the model (solid circles) and the observations (open circles), and Fig. 3 shows that this has been achieved remarkably well. The longer-lived species ($O_3$ and ClO) are in fact a better fit than the shorter lived species (BrO, Cl and Br). However, since the concentration of the short-lived species is basically controlled by the change of the long-lived species (e.g. see Fisher and Lary 1995; Khattatov *et al.* 1999), a good estimate of the long-lived species will produce a good analysis for the short-lived species.

In the actual implementation of the technique to real data, the observation error should be taken into account when comparing differences between the 4D-Var analyses and the observations. In a separate paper we will discuss the practical application of the chemical 4D-Var in analysing Atmospheric Trace Molecule Spectroscopy data and for studies on $NO_y$ partitioning in the stratosphere. We will also explore the use of the Kalman filter, which considers errors from both model and observations, in 4D variational chemistry data assimilation.

## 5. CONCLUSION

In this review paper we described a detailed formulation of the adjoint method to be used in 4D variational chemistry data assimilation. The implementation procedure is outlined and the method is tested using a simple catalytic ozone destruction mechanism which is relevant to the ozone depletion during the early spring over the polar regions. As shown above, the 4D-Var method is very effective in fitting the model to the observations. Hence the use of this method will enable better analysis of stratospheric and tropospheric chemistry to be made. In a following paper we will describe the

application of the 4D-Var method in fitting a complicated stratospheric chemistry model to ATMOS data, and to the study of $NO_y$ partioning in the stratosphere.

## ACKNOWLEDGEMENTS

We are very grateful to Olivier Talagrand, Hendrik Elbern, Boris Khattatov, and European Centre for Medium-Range Weather Forecasts for sending us their publications, and M. Fisher for his 4D-Var code. We acknowledge the use of data from the Met Office and British Atmospheric Data Centre. David Lary like to thank the Natural Environment Research Council (NERC) and the European Union for research support. We also thank two anonymous reviewers for their insightful comments that greatly improved the clarity of this paper. The Centre for Atmospheric Science is a joint initiative of the Department of Chemistry and the Department of Applied Mathematics and Theoretical Physics. This work forms part of the NERC United Kingdom Universities Global Atmospheric Modelling Programme.

## REFERENCES


Boas, M. L. 1983 *Mathematical methods in the physical sciences*, second edition. John Wiley & Sons, New York

Brown, P. N., Byrne, G. D. and Hindmarsh, A. C. 1989 VODE: A variable coefficient ODE solver. *SIAM J. Sci. Stat. Comput.*, **10**, 1038–1051

Crutzen, P. J., Grooß, J.-U., Brühl, C., Müller, R. and Russell III, J. M. 1995 A re-evaluation of the ozone budget with HALOE UARS data: No evidence for the ozone deficit. *Science*, **268**, 705–708

Daley, R. 1991 *Atmospheric data analysis*. Cambridge University Press, Cambridge

DeMore, W. B., Sander, S. P., Golden, D. M., Hampson, R. F., Kurylo, M. J., Howard, C. J., Ravishankara, A. R., Kolb, C. E. and Molina, M. J. 1997 Chemical kinetics and photochemical data for use in stratospheric modeling—Evaluation 12. *JPL Publication*, 97–94

Elbern, H., Schmidt, H. and Ebel, A. 1997 Variational data assimilation for tropospheric chemistry modeling. *J. Geophys. Res.*, **102**, 15967–15985

Feintuch, A. 1998 *Robust control theory in Hilbert space*. Springer-Verlag, New York

Fisher, M. and Lary, D. J. 1995 Lagrangian four-dimensional variational data assimulation of chemical species. *Q. J. R. Meteorol. Soc.*, **121**, 1681–1704

Giering, R. and Kaminski, T. 1998 Recipes for adjoint code construction. *ACM TOMS*, **24**, 437–474

Gilbert, J. C. and Lemarechal, C. 1989 Some numerical experiments with variable-storage quasi-Newton algorithms. *Math. Program.*, **45**, 407–435

Kaminski, T., Heimann, M. and Giering, R. 1999 A coarse grid three-dimensional global inverse model of the atmospheric transport. I: Adjoint model and Jacobian matrix. *J. Geophys. Res.*, **104**, 18535–18553

Khattatov, B. V., Gille, J. C., Lyjak, L. V., Brasseur, G. B., Dvortsov, V. L., Roche, A. E. and Waters, J. W. 1999 Assimilation of photochemically active species and a case analysis of UARS data. *J. Geophys. Res.*, **104**, 18715–18737

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

Nocedal, J. 1980 Updating quasi-Newton matrices with limited storage. *Math. Comput.*, **35**, 773–782

Simmons, G. F. 1963 *Introduction to topology and modern analysis*. McGraw-Hill, London

Talagrand, O. 1991 Adjoint models. Pp. 73–91 in *Numerical Methods in Atmospheric Models, Vol. II*. European Centre for Medium-Range Weather Forecasts, Reading, UK

Talagrand, O. and Courtier, P. 1985 'Formalisation de la méthode du modele adjoint. Applications météorologiques'. Note No. 117, Etablissement d'Etudes et de Recherches Météorologiques, Paris, France

Talagrand, O. and Courtier, P. 1987 Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory. *Q. J. R. Meteorol. Soc.*, **113**, 1311–1328

Thacker, W. C. 1989 The role of the Hessian matrix in fitting models to measurements. *J. Geophys. Res.*, **94**, 6177–6196

Weisstein, E. W. 1998 *CRC concise encyclopedia of mathematics*. CRC Press, Boca Raton, Fla., USA