ADJOINT MODELS

Olivier Talagrand
Laboratoire de Météorologie Dynamique du CNRS
Paris, France

Summary. The general method of adjoint equations, which allow an economical numerical determination of the gradient of one output parameter of a numerical model with respect to all input parameters, is presented. A number of meteorological applications of the method are discussed.

1. INTRODUCTION

A numerical model of the atmospheric circulation, like any numerical algorithm, can be formally described as a process which, starting from the values asssigned to a number of input parameters, computes the corresponding values taken by a number of output parameters. In the case of an atmospheric model, the input parameters are the parameters which define the geometrical and physical conditions of the integration, the spatial and temporal discretization, the initial state from which the model is to be integrated, the lateral boundary conditions whose specification may be required in the course of the temporal integration and, more generally, all parameters whose numerical values must be specified before the integration of the model can be started. What the output parameters exactly are is somewhat conventional, but they basically consist of the history of the flow produced by the integration of the model, together with the various "diagnostic" quantities which one may be led to compute for a given integration. For the most powerful present numerical models of the atmospheric circulation, such as those which are used for operational forecasting in large meteorological centres, the number of input parameters is typically 10^6 - 10^7. As for the number of output parameters, if one chooses to consider as output the sequence of model states produced by the integration of a model over a period of 5 to 10 days, typical of present operational numerical weather forecasting, one obtains values of up to 10^9 - 10^{10}.

In this general context, many problems that one may be led to consider can be described, in one form or another, as problems of sensitivity of output parameters with respect to input parameters. Such problems can arise in many different types of situations, such as for instance the following.

a) Situations in which one actually wants to determine the physical response of the atmospheric system to variations in one or several of the input parameters. A typical example

is encountered in the study of the climatic greenhouse effect, in which one will want to determine the impact, on the various parameters which define the "climate", of the increase of the atmospheric content in various gases.

b) Situations in which one wants to determine the uncertainty on the output parameters resulting from the uncertainty on the input parameters. A typical example is encountered in numerical weather forecasting, where one major problem is at present to determine the uncertainty in the forecast resulting from the uncertainty in the initial conditions.

c) Situations in which one wants to determine what, in the input parameters of the model, was at the origin of some a posteriori observed feature in the output parameters. A typical example is the following : a numerical weather forecast has been erroneous in some respect, for instance by failing to predict the rapid deepening of a depression, and one wants to determine what, in the initial state of the forecast, or in the various physical parameters of the model, was the "cause" of that error.

d) Situations in which one wants to determine the values of the input parameters for which the output parameters satisfy a prescribed condition. This leads to solving so-called optimization problems, intended at determining the values of the input parameters which "optimize" some property of the output parameters. A typical example is variational assimilation of observations, in which one wants to determine the state of the model, at the initial time of the assimilation period, such that the corresponding solution is close, in some appropriate sense, to the observations available over the assimilation period.

In order to solve such sensitivity problems, one might think of using standard analytical methods, based on classical diffferential calculus. However, it is now well established that the atmospheric flow, as well as the numerical models used for simulating it, are "chaotic" systems, for which classical analytical methods are of very limited use. If a numerical model is available, one can, at least in principle, solve sensitivity problems of the different types mentioned above by empirical modification of the input parameters. This is what is done, for instance, when studying the greenhouse problem with numerical models of the atmospheric circulation. A typical procedure is the following : a model is integrated with different values of, say, the atmospheric content of carbon dioxide, and the "climates" produced by the different integrations are compared. Similarly, problems such as the examples mentioned under headings c) and d) above could in principle be solved by empirically modifying the values of the input parameters in order to determine those values which would lead to an accurate forecast of the deepening depression (case c) or to a close fit of the model to the observations (case d). It is clear however that, in these two cases, the numerical cost of this approach would in practice be totally prohibitive.

In order to get some more insight into the problem, let us denote U the vector made up of the input parameters. The components of U, in number n, will be denoted u_i (i = 1, ..., n). Similarly, let us denote V, with components v_j (j = 1..., m), the vector of output parameters (values quoted above are $n \approx 10^6 - 10^7$ and $m \approx 10^9 - 10^{10}$). The output vector V is a uniquely defined function of the input vector, which we will denote

$$V = G(U) \tag{1.1}$$

For given U, the local sensitivities of the output parameters with respect to the input parameters are then given by the jacobian matrix G'(U), made up of the partial derivatives $\partial v_j/\partial u_i$ at point U. One integration of the model, started from the unperturbed input vector U, followed by a second integration, started from an input vector in which one component, say u_1, has been perturbed by an amount Δu_1, gives, for all j's, the sensitivities $\partial v_j/\partial u_1$ under the approximate form of finite-difference ratios $\Delta v_j/\Delta u_1$. One column of the jacobian matrix can therefore be obtained at the cost of two integrations of the model. More generally, p columns of the jacobian can be obtained at the cost of p + 1 integrations of the model (one unperturbed integration, followed by p appropriately perturbed integrations). This "direct perturbation" method for determining sensitivities is practically useful as long as the number of input parameters with respect to which one is interested in determining sensitivities is small (in present meteorological practice, "small" would mean a few units, or at the very most a few tens of units). This approach is for instance perfectly appropriate for studying the climatic impact of the increase of the atmospheric carbon dioxide content.

Now, in many cases, such as for instance in the examples c) and d) above, one will not be interested in the sensitivities of a large number of output parameters with respect to one, or a small number, of the input parameters, but rather in the sensitivities of one, or a small number, of the output parameters with respect to a large number of input parameters. For instance, in example d), one will be interested, in order to determine the initial state which produces the best fit of the model to the observations, in knowing the sensitivities of one output parameter (the fit of the model to the observations) with respect to all parameters defining the initial state of the model. It is in this kind of situations that the adjoint method, to which these notes are devoted, is particularly useful. Just as one integration of the model, followed by one perturbed integration, produces the partial derivatives of all output parameters with respect to one input parameter, one integration of the model, followed by one integration of the so-called adjoint equations of the model (whose cost is of the same order as one integration of the model), produces the partial derivatives of one output parameter with respect to all input parameters. These partial derivatives make up the gradient vector of the output parameter under consideration with respect to the input parameters. The direct perturbation

method produces columns of the jacobian, while the adjoint method produces rows of the jacobian, which clearly shows the "transpose", or "adjoint", aspect of the method. The adjoint method provides a very efficient way for solving many sensitivity or optimization problems. It has stimulated, in the last six years or so, a great interest in numerical modeling of the atmospheric (and also oceanic) circulation, and is the object of many works of various kinds.

The general principle of the adjoint method is presented in detail in Section 2, in both cases of a finite algebraic process, which is in principle sufficient for any numerical application (subsection 2.1), and of differential equations, which leads to instructive insights (subsection 2.3). A number of meteorological applications of the adjoint method, especially to variational assimilation of observations, are presented and briefly discussed in Section 3.

2. THE PRINCIPLE OF THE ADJOINT METHOD

Many different approaches can be followed for presenting the principle of the adjoint method and, depending on one's background and experience, one may prefer one or another of these approaches. We will follow here the approach already followed in *Talagrand* (1989), which seems to require least effort for readers who, like specialists in numerical modeling of the atmosphere, have much experience with large dimension numerical models, but have otherwise little or no experience in fields like optimal control or automatic differentiation, in which the adjoint method has so far been mostly used.

2.1 The finite-dimensional algebraic case

Using the same notations as in the Introduction, let us consider a perturbation δU, with components δu_i ($i = 1, ..., n$), on the input vector U. To first-order with respect to δU, the corresponding perturbation δv_j on the output parameter v_j will be equal to

$$\delta v_j = \sum_{i=1}^{n} \frac{\partial v_j}{\partial u_i} \delta u_i \qquad j = 1, ..., m \tag{2.1}$$

or, in matrix notation

$$\delta V = G' \delta U \tag{2.2}$$

where δV is the output perturbation vector and G' is, as before, the jacobian of (1.1), taken at point U . Eq. (2.2) will be called the linear tangent equation to eq. (1.1), taken at point U.

Let now $V \rightarrow \mathcal{J}(V)$ be a scalar-valued function of the output vector V. The partial derivative of $\mathcal{J}$ with respect to the input parameter u_i is given by the chain rule

$$\frac{\partial \mathcal{J}}{\partial u_i} = \sum_{j=1}^{m} \frac{\partial v_j}{\partial u_i} \frac{\partial \mathcal{J}}{\partial v_j} \qquad i = 1, ..., n \tag{2.3}$$

which shows that the gradient vector of the partial derivatives of $\mathcal{J}$ with respect to the input parameters, which we will denote $\nabla_U \mathcal{J}$, is linked to the gradient vector $\nabla_V \mathcal{J}$ of the partial derivatives of $\mathcal{J}$ with respect to the output parameters by the jacobian matrix G'. But while the summation is performed in eq. (2.1) on the input index i, it is performed in eq. (2.3) on the output index j. The matrix form of eq. (2.3) accordingly reads

$$\nabla_U \mathcal{J} = G'^* \nabla_V \mathcal{J} \tag{2.4}$$

where G'^* denotes the transpose of the jacobian G'.

Comparison between eqs (2.2) and (2.4) is instructive. The direct jacobian G' transforms a perturbation on the input into the first-order corresponding perturbation on the output, while the transpose jacobian G'^* transforms the gradient of a function with respect to the output into the gradient of the same function with respect to the input. It is worth noting that eq. (2.4) is valid for any function $\mathcal{J}$. The principle of the adjoint method (at this stage, we call it "adjoint" in anticipation to what will be presented in subsection 2.2) is to compute the gradient $\nabla_U \mathcal{J}$, not through direct explicit perturbation of the input parameters, but through use of eq. (2.4). If $\mathcal{J}$ is a "simple" enough function of the output vector V (and $\mathcal{J}$ will always be a simple function of some appropriately chosen "output vector", be that $\mathcal{J}$ itself), $\nabla_V \mathcal{J}$ can be determined analytically and introduced as input of the computations represented by (2.4). The important point is that these computations can be performed without having to explicitly determine the complete matrix G'^*. This fact, which may not seem a priori obvious, becomes clear if one realizes that the cost of one adjoint computation (2.4) must be the same as the cost of one tangent linear computation (2.2). The latter does not require the explicit determination of the jacobian G', but can be performed by formally differentiating each step of the computations represented by the basic equation (1.1), and then numerically implementing the corresponding linear computations. To illustrate how the adjoint computations can be implemented in a similar stepwise fashion, let us assume that the basic model (1.1) involves at some stage a computation of the form

$$r = \sqrt{(x^2 + y^2)} \tag{2.5}$$

The corresponding linear perturbation computation will read

$$\delta r = \frac{\partial r}{\partial x}\,\delta x + \frac{\partial r}{\partial y}\,\delta y = \frac{x}{r}\,\delta x + \frac{y}{r}\,\delta y$$

or, in standard matrix notation

$$\delta r = \begin{bmatrix} \frac{x}{r} & \frac{y}{r} \end{bmatrix} \begin{bmatrix} \delta x \\ \\ \delta y \end{bmatrix} \tag{2.6}$$

As for the corresponding adjoint computation, let us assume that we know the partial derivative $\partial J/\partial r$ of J with respect to r (i.e. with respect to the output of the particular simple process represented by eq. 2.5). The derivatives of J with respect to the inputs x and y are given by the chain rule

$$\frac{\partial J}{\partial x} = \frac{\partial r}{\partial x}\frac{\partial J}{\partial r} = \frac{x}{r}\frac{\partial J}{\partial r}$$

$$\frac{\partial J}{\partial y} = \frac{\partial r}{\partial y}\frac{\partial J}{\partial r} = \frac{y}{r}\frac{\partial J}{\partial r}$$

or, in matrix for

$$\begin{bmatrix} \frac{\partial J}{\partial x} \\ \\ \frac{\partial J}{\partial y} \end{bmatrix} = \begin{bmatrix} \frac{x}{r} \\ \\ \frac{y}{r} \end{bmatrix} \frac{\partial J}{\partial r} \tag{2.7}$$

It is seen that the matrix which appears on the right-hand side of this expression is the transpose of the matrix appearing on the right-hand side of (2.6).

The essence of the adjoint method is simply to systematically perform computations of the form (2.7) for all the steps of the basic computations (1.1). Because of their very nature, the adjoint computations, which proceed from the gradient of $\mathcal{J}$ with respect to the output vector V to the gradient of $\mathcal{J}$ with respect to the input vector U, must be performed in reversed order of the basic computations (1.1). In particular, if, as it is the case dynamical models of the atmospheric circulation, the direct model involves some form of temporal integration into the future, the corresponding adjoint model will involve some form of backward integration into the past. More precisely, let us decompose the model computations G into a number of successive steps, *viz.*,

$$G = G_M \circ ..... \circ G_2 \circ G_1 \tag{2.8}$$

Each of these steps can for in instance be an elementary model timestep, which can itself be decomposed into more elementary steps, down to the individual executable coding statements. The chain rule shows that the jacobian G' will be the matrix product of the jacobians of the successive steps, *viz.*,

$$G' = G_M' ..... G_2' G_1'$$

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

$$G'^* = G_1'^* G_2'^* ..... G_M'^* \tag{2.9}$$

which again shows the reversed character of the adjoint computations. In particular, eq. (2.9) suggests a systematic approach for developing the adjoint of a given code : successively develop the adjoints of the various components of the basic code (for instance FORTRAN subroutines), and then connect the elementary adjoints in reversed order.

Two additional properties of the adjoint method clearly appear on eqs (2.2,4,6, and 7).

a) The numerical cost of one adjoint computation (2.4) is the same as the cost of one linear tangent computation (2.2). The equations governing fluid motions, on which numerical

models of the atmospheric circulation are built, contain both linear and quadratic terms, and can be symbolically written as

$$\frac{dX}{dt} = LX + N(X,X) \qquad (2.10)$$

where LX represents the linear terms, and N(X,X) the quadratic terms. Differentiation, analogous to (2.2), will lead to

$$\frac{d\,\delta X}{dt} = L\,\delta X + N(\delta X, X) + N(X, \delta X)$$

While linear terms are unaltered by differentiation, each quadratic term gives rise to <u>two</u> terms. In consequence, the cost of one adjoint computation will be comprised between once and twice the cost of one basic computation (1.1). More generally, it can be shown (see, e.g., *Morgenstern*, 1984) that, for <u>any</u> numerical algorithm, the total operation count of the adjoint computation can be reduced to at most 4 times the total operation count of the direct computation, this ratio being reduced to 2 if one considers only multiplications and divisions.

b) The adjoint computation (2.7) requires the preliminary knowledge of the partial derivatives $\partial r/\partial x = x/r$ and $\partial r/\partial y = y/r$. The same will be true whenever the basic computations will be nonlinear. The jacobian G' and its transpose G'* will in such a case depend on the particular value of the input vector for which one wants to determine the gradient $\nabla_U J$. This means that, before the adjoint computation (2.4) can be implemented, a direct computation (1.1) must have been performed and that all computed intermediary values (or at least all values used in nonlinear computations) either must have been stored in memory in order to be available for the adjoint computations, or will have to be recomputed in the course of the adjoint computation. The corresponding storage requirements may of course be enormous, and that is the price to be paid for the gain in computing time provided by the adjoint method.

2.2 <u>The general notion of an adjoint operator</u>

Rather than using the notion of a transpose matrix, it is much more convenient to use the more general notion of an <u>adjoint operator</u>. Let us consider two (possibly infinite-dimensional) linear spaces $\mathbb{E}$ and $\mathbb{F}$, on which inner products, denoted < , >, have been defined. Given a continuous linear operator L of $\mathbb{E}$ into $\mathbb{F}$, there exists a unique continuous linear operator L^* of $\mathbb{F}$ into $\mathbb{E}$ such that, for any two vectors U and V belonging to $\mathbb{E}$ and $\mathbb{F}$ respectively, the following equality between inner products holds

$$< LU , V > = < U , L^{*}V> \qquad (2.11)$$

L^* is called the adjoint of L. If $\mathbb{E}$ and $\mathbb{F}$ have finite dimensions n and m respectively, and are described by coordinates along orthogonal unitary vectors ($U = (u_i)$, $i = 1, \ldots, n$; $V = (v_j)$, $j = 1, \ldots, m$; $L = (l_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, m$), so that inner products take the familiar form $<U,U'> = \sum_i u_i u'_i$, eq. (2.11) reduces to a change in the order of summation indices

$$\sum_j [\sum_i l_{ij} u_i] v_j = \sum_i u_i [\sum_j l_{ij} v_j]$$

which shows that the matrix representing L^* is the transpose of the matrix representing L.

In inner product notation, the first-order variation δJ resulting from a perturbation δV of V can be written

$$\delta J = < \nabla_V J , \delta V > \qquad (2.12)$$

We can note that this relationship, which generalizes the finite-dimension relationship $\delta J = \sum_j (\partial J/\partial v_j)\, \delta v_j$, characterizes the gradient, in the sense that, if the first-order variation δJ of a scalar function J of some vector V can be written as the inner product of δV with some vector, then that vector is necessarily equal to $\nabla_V J$.

With this more general system of notations, the relationship (2.4) between gradients can be established as follows. Through uses of eq. (2.2) and of the "adjointness" relationship (2.11), (2.12) becomes

$$\delta J = <\nabla_V J , G'\delta U> = < G'^{*} \nabla_V J , \delta U > \qquad (2.13)$$

The last equality, which expresses δJ as the inner product of δU by the vector $G'^{*} \nabla_V J$, shows that the gradient $\nabla_U J$ is equal to the latter vector, which is what is stated by eq. (2.4).

Use of equations such that (2.11-12) has two distinct advantages : it avoids cumbersome manipulation of indices, and covers the case of infinite-dimensional function spaces, as will be shown in the following subsection. In addition, the general definition (2.11) of an adjoint operator is sometimes very useful, even at the coding level, for developing the adjoint of a numerical model.

2.3 The case of differential equations

In subsection 2.1, we have presented the adjoint method in the case of a process consisting of a finite number of algebraic operations. This of course is in principle sufficient for any numerical application. Now, numerical models of the atmospheric circulation are based on physical principles which can be expressed as partial differential equations verified by fields defined on a space-time continuum, and reference is often made to these partial differential equations, even in the context of numerical simulation of the atmospheric flow. The principle of the adjoint method extends to differential equations, either ordinary or partial, as we will now illustrate on a simple example. The nonlinear advective diffusive equation

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{u^2}{2}\right) - \nu \frac{\partial^2 u}{\partial x^2} = 0 \qquad (2.14)$$

where t is time and x is a one-dimensional spatial coordinate, describes the variations of a one-dimensional velocity field u(x,t) under the action of a viscosity force defined by the positive coefficient ν. Because of the irreversible character of the viscosity term, integration of (2.14) leads to a well-posed problem only for integration into the future. The flow will be assumed to be defined over the rectangular space-time domain $\mathcal{D} = \{x, t\,;\, x_1 \le x \le x_2\,,\ t_1 \le t \le t_2\}$ with boundary Γ. Appropriate initial and boundary conditions for the integration of (2.14) are then values of u at the initial time t_1 for all $x_1 \le x \le x_2$, and at later times $t_1 < t \le t_2$ along both lateral boundaries $x = x_1$ and $x = x_2$. We will call input boundary, and denote Γ_i, that part of Γ consisting of the three segments ($t = t_1$, $x_1 \le x \le x_2$), ($t_1 \le t \le t_2$, $x = x_1$), ($t_1 \le t \le t_2$, $x = x_2$). Under these conditions, integration of eq. (2.14) is a process whose input U is the field u along Γ_i, while its output V is the field u over the entire domain $\mathcal{D}$. V is a uniquely defined, smoothly varying, function of U, of type (1.1).

In order to fix ideas, we now assume that "observations" of u(x,t), denoted $\hat{u}(x,t)$, are available over the whole domain $\mathcal{D}$ and that we want to solve an assimilation type problem, i.e. to determine a solution of (2.14) which is close to the observations. To this end, we define a scalar function $\mathcal{J}$, *viz.*,

$$\mathcal{J} = \frac{1}{2}\int_{\mathcal{D}} (u - \hat{u})^2 \, dx \, dt \qquad (2.15)$$

which, for any field u defined over $\mathcal{D}$, and in particular for any solution of (2.14), measures the "distance" or "misfit", between that solution and the observations. We then want to determine the input U to (2.14) such that the corresponding output minimize $\mathcal{J}$. To this end,

we must be able to relate the variations of U to the corresponding variations of $\mathcal{J}$. Let us assume that we have been able to find a field W = w(x,t), defined along the input boundary Γ_i, such that for any variation $\delta U = \delta u(x,t)$, $(x,t) \in \Gamma_i$, of the input, the corresponding first-order variation of $\mathcal{J}$ reads

$$\delta\mathcal{J} = \int_{\Gamma_i} w\, \delta u\, ds \tag{2.16}$$

where the integral extends over the input boundary. The integral is the inner product, in the L^2 sense, of the field W by the input perturbation δU. W is the gradient , in the L^2 sense, of the scalar function $\mathcal{J}$ with respect to the input U. This gradient explicitly links the variations of U to the corresponding variations of $\mathcal{J}$. For example, a perturbation δU which is everywhere of the same sign as w will result in an increase of $\mathcal{J}$, while a perturbation which is everywhere of the opposite sign will result in a decrease in $\mathcal{J}$. And, for a given amplitude (in the L^2 sense) of δU, the variation of $\mathcal{J}$ will be largest if δu is proportional to w, i. e. if it "directed" along the gradient W.

We now proceed to demonstrate how, for a given solution u of (2.14), the adjoint method can be used to determine the gradient of $\mathcal{J}$ with respect to U. Let us consider a perturbation δU of the input. The corresponding first-order perturbation $\delta V = \delta u(x,t)$, $(x,t) \in \mathcal{D}$, on the output will verify the tangent linear equation to (2.14), *viz.*,.

$$\frac{\partial\, \delta u}{\partial t} + \frac{\partial}{\partial x}(u\, \delta u) - \nu \frac{\partial^2 \delta u}{\partial x^2} = 0 \tag{2.17}$$

while the corresponding first-order variation of $\mathcal{J}$ will be equal to

$$\delta\mathcal{J} = \int_{\mathcal{D}} (u - \hat{u})\delta u\, dx\, dt \tag{2.18}$$

We now consider a field $\delta' u(x,t)$ defined over the entire domain $\mathcal{D}$, but not otherwise specified at this stage. We multiply eq. (2.17) by $\delta' u$, integrate over $\mathcal{D}$, and subtract the result from the right-hand side of eq. (2.18). This leads to the following expression for $\delta\mathcal{J}$

$$\delta\mathcal{J} = \int_{\mathcal{D}} [(u - \hat{u})\delta u - [\frac{\partial\, \delta u}{\partial t} + \frac{\partial}{\partial x}(u\, \delta u) - \nu \frac{\partial^2 \delta u}{\partial x^2}]\ \delta' u]\, dx\, dt \tag{2.19}$$

Our purpose is to transform this expression into an expression of the form (2.16). This may a priori seem to be doubly difficult because a) the integral in (2.19) is taken over the whole domain $\mathfrak{D}$, while the integral in (2.16) is taken only over the input boundary, and because b) even if the perturbation δu effectively appears as a factor of a product in the first term of the integrand in (2.19), it appears only through its derivatives in the second term. Taking care first of this second difficulty, we observe that the derivative symbols in front of δu can be eliminated and transferred to $\delta' u$ by performing appropriate integrations by parts. One integration by parts will have to be performed for eliminating first-order derivative symbols, and two integrations for eliminating the second-order derivative symbol of the viscosity term. Performing these integrations by parts and rearranging terms lead to

$$\delta \mathcal{J} = \int_{\mathfrak{D}} [\frac{\partial\, \delta' u}{\partial t} + u \frac{\partial\, \delta' u}{\partial x} + \nu \frac{\partial^2\, \delta' u}{\partial x^2} + u - \hat{u}\,]\, \delta u \, dx \, dt$$

$$+ \int_{x_1}^{x_2} [\, \delta' u(x,t_1)\, \delta u(x,t_1) - \underbrace{\delta' u(x,t_2)\, \delta u(x,t_2)}_{1}]\, dx$$

$$+ \int_{t_1}^{t_2} [u(x_1,t)\, \delta' u(x_1,t)\, \delta u(x_1,t) - \underbrace{u(x_2,t)\, \delta' u(x_2,t)\, \delta u(x_2,t)}_{3}]\, dt$$

$$+ \nu \int_{t_1}^{t_2} [\frac{\partial \delta' u}{\partial x}(x_1,t)\, \delta u(x_1,t) - \frac{\partial \delta' u}{\partial x}(x_2,t)\, \delta u(x_2,t)]\, dt$$

$$- \nu \int_{t_1}^{t_2} [\frac{\partial \delta u}{\partial x}(x_1,t)\, \delta' u(x_1,t) - \underbrace{\frac{\partial \delta u}{\partial x}(x_2,t)\, \delta' u(x_2,t)}_{2}]\, dt \qquad (2.20)$$

where the last four integrals are boundary terms along Γ which have appeared because of the integrations by parts. Eq. (2.20) may seem to be still much more complicated than eq. (2.19), thereby raising doubts in the reader's mind as to the wisdom of the whole undertaking, but most terms in (2.20) will presently be eliminated by imposing appropriate conditions on the

field $\delta'u$, which is still at this stage totally unspecified. We can first observe that the first term on the rhs of (2.20), which is an integral over $\mathfrak{D}$, becomes zero if $\delta'u$ is chosen such as to verify the following inhomogeneous linear partial differential equation

$$\frac{\partial\,\delta'u}{\partial t} + u\frac{\partial\,\delta'u}{\partial x} + \nu\frac{\partial^2\,\delta'u}{\partial x^2} + u - \hat{u} = 0 \tag{2.21}$$

This almost entirely takes care of the first of the two difficulties mentioned above, since the expression for $\delta\mathfrak{J}$ now reduces to boundary integrals. If we want to obtain for $\delta\mathfrak{J}$ an expression of the form (2.16), we see that we must eliminate from the rhs of (2.20) term number 1, which is an integral over that part of the boundary Γ which does not belong to Γ_i, and term number 2, which contains a derivative in δu. These two terms can be eliminated by imposing to $\delta'u$ to be zero at the final time t_2 and along the two lateral boundaries $x = x_1$ and $x = x_2$. We will call output boundary, and denote Γ_o, that part of Γ consisting of the three segments $(t = t_2,\ x_1 \le x \le x_2)$, $(t_1 \le t \le t_2,\ x = x_1)$, $(t_1 \le t \le t_2,\ x = x_2)$. The condition imposed on $\delta'u$ therefore reads

$$\delta'u = 0 \quad \text{along } \Gamma_o \tag{2.22}$$

This condition happens to also eliminate term number 3 in eq. (2.20), and there remains

$$\delta\mathfrak{J} = \int_{x_1}^{x_2} \delta'u(x,t_1)\,\delta u(x,t_1)\,dx + \nu\int_{t_1}^{t_2} \left[\frac{\partial\delta'u}{\partial x}(x_1,t)\,\delta u(x_1,t) - \frac{\partial\delta'u}{\partial x}(x_2,t)\,\delta u(x_2,t)\right] dt \tag{2.23}$$

which is of the form (2.16), and defines the gradient of $\mathfrak{J}$ with respect to the input U. This gradient is equal to $\delta'u(x,t)$ at the initial time t_1, and to $\nu\,\partial'u/\partial x$ (resp. $-\nu\,\partial'u/\partial x$) along the boundary $x = x_1$ (resp. $x = x_2$), where $\delta'u$ is the solution of (2.21) verifying the "final" and lateral boundary conditions expressed by (2.22). Eq. (2.21) is the adjoint equation of (2.14) for the particular solution u. Integrating it amounts to performing a computation of form (2.4), the gradient $\nabla_V\mathfrak{J}$ being here the term $u - \hat{u}$, which originates in the differentiation of the integrand in (2.15). Changing $\mathfrak{J}$ to another expression would modify only that term, without other modification to either (2.21) or (2.22). The "adjoint" character of the derivations performed above must be clear : the transformation from (2.19) to (2.23) is essentially equivalent to the transformation from the second to the third members of eq. (2.13). In the case of eqs (2.19-23), this transformation is essentially performed through integrations by

parts, which carry derivatives from one factor of the inner product to the other. This is absolutely general, and adjoints of differential equations are always obtained by performing integrations by parts.

It has been mentioned that, because of the presence of the viscosity term, eq. (2.14) leads to a well-posed problem only for integration into the future. In the adjoint equation (2.21), the viscosity term has changed sign, so that eq. (2.21) leads to a well-posed problem only for integration into the past (mathematically, this is due to the fact that the time-derivative in eq. (2.14) is first-order, while the space derivative in the viscosity term is second-order, leading to one more change of sign for the viscosity term in the integrations by parts). There can therefore be no difficulty in the adjoint integration because of the presence of an irreversible term in the basic equation (2.14). The fact that a dissipative term, like the viscosity term in (2.14), must remain dissipative in the adjoint integration can be easily understood. In a purely dissipative integration, any solution will tend to an asymptotic state of rest, and the final state of the integration will become less and less sensitive to the initial state as the length of the integration period increases, with the consequence that the gradient of a given output with respect to the initial state will tend to zero. Conversely, the result of an adjoint integration, which is the gradient of the output with respect to the initial state, must tend to zero as the length of the adjoint integration period increases, which means that there must be some form of dissipation in the adjoint integration. More generally, whenever a basic equation leads to a well-posed problem only for integration into the future, the corresponding adjoint equation will lead to a well-posed problem only for integration into the past.

In any practical situation, one would have at one's disposal a discretized model of the basic equation (2.14). In order to obtain the discretized adjoint model, one could start from the non-discretized adjoint equation (2.21), and discretize it. However, it may not be obvious which discretization of (2.21) is the exact adjoint of the original discretization of (2.14). One could also perform on the direct discretized model the same manipulations which have been performed above on (2.14) and (2.17). This has been done in *Talagrand* (1989) for a particular discretization of (2.14) (and for periodic boundary conditions in the spatial coordinate x). But experience strongly suggests that the most efficient way for developing the discretized adjoint is to actually start from the computer code, and to write the adjoint instruction per instruction, along the lines suggested by eqs (2.7-9). This is a lengthy and tedious, but very straightforward task. All this is not to mean that determining the exact expression of the non-discretized adjoint equation is without practical interest. It helps in understanding the significance of the adjoint method, and it can be very useful for identifying and establishing properties of the adjoint equation, as has just been shown for instance on the particular question of the effect of irreversible terms.

3. METEOROLOGICAL APPLICATIONS

3.1 Variational assimilation

The use of the adjoint method in the general context of numerical modeling of the atmospheric (and oceanic) circulation was first advocated by the Soviet mathematician Marchuk and his collaborators (see, e.g., *Marchuk*, 1974, and, for a more recent reference, *Marchuk and Skiba*, 1990). Later, Cacuci and collaborators (see, e.g., *Hall* et al.,1982, or *Cacuci and Hall*, 1984) stressed the potential utility of the adjoint method for direct sensitivity studies. It is however to the problem of variational assimilation of observations that the adjoint method has been mostly used so far in meteorology. The principle of variational assimilation, which has been described in detail in *Talagrand* (1989), has been illustrated in subsection 2.3 : one wants to minimize a scalar function which measures the "distance" between a model solution and the available observations. This function will typically be of the form (2.15), i.e. will be an integral of squared model-minus-observations differences. One particular choice, consistent with the theory of least-square linear statistical estimation, is to weight the squared differences with coefficients which are inversely proportional to the statistical variances of the corresponding observational errors. In order to minimize the "distance-function" (often called a "cost-function"), an iterative minimization algorithm is implemented on the model input, usually limited in this particular type of application to the initial (and possibly lateral boundary) conditions whose specification uniquely defines a model solution. The minimization algorithm proceeds by successive approximations, each iteration consisting in the computation of the local gradient of the distance-function with respect to the input, followed by a descent step leading to a better approximation of the minimizing input. Classical minimization algorithms are the conjugate gradient or quasi-Newton algorithms (see, e. g., *Gill* et al., 1982, or *Navon and Legler*, 1987). In variational assimilation, the adjoint model is only used as a (very efficient) numerical tool for numerically computing the gradient of the distance-function with respect to the model input. In principle, this gradient could also be determined by other methods, for instance by direct perturbation, as has indeed been done by *Hoffman* (1986). But the corresponding numerical cost, as already mentioned, would be totally prohibitive in any practical situation.

Variational assimilation, with use of the adjoint of the assimilating model, was first advocated by *Penenko and Obraztsov* (1976), who demonstrated the method on a small dimension linear model. Later, a number of authors applied the same approach to both meteorological (see, e.g., *Le Dimet*, 1980, *Lewis and Derber*, 1985, *Le Dimet and Talagrand*, 1986, *Talagrand and Courtier*, 1987, *Courtier and Talagrand*, 1987, 1990, *Lorenc*, 1988, and references in these various papers) and oceanographical models (see, e. g., *Schröter and Wunsch*, 1986, *Wunsch*, 1987, *Thacker*, 1988, *Thacker and Long*, 1988). The most recent meteorological results have been obtained with medium resolution primitive equation models, for which the dimension of the input vector with respect to which the minimization is performed reaches

values of up to 10^5. See, e.g., *Courtier* et al. (1990), *Navon* et al. (1990), *Chao and Chang* (1990), *Thépaut and Courtier* (1991). The models used so far contained no physics, or only very simplified physics.

In a typical application, the distance-function is estimated over a 24-hour period, and results obtained so far show first that the adjoint method works successfully in the sense that it is capable of determining the gradient of the distance function with respect to the model initial conditions. Repeated tests have shown that the gradient is determined to computer accuracy. In addition, the minimization process is capable of effectively determining the model solution which minimizes the distance-function. With standard minimization algorithms, such as quasi-Newton algorithms, ten to fifteen iterations are typically necessary for reaching the minimum. In addition, in all experiments performed so far, the minimum appears to be unique, and no problem has been encountered which could be due to the existence of secondary minima. As for the physical properties of the minimizing solution, it appears that, if the function to be minimized is only a measure of the misfit to the observations, as would for instance be the function defined by eq. (2.15), the minimizing solution is contaminated by irrealistic gravity wave noise. The minimization solution becomes physically acceptable only if an appropriate penalty term, measuring the deviation of the model solution from geostrophic balance, is included in the distance function. This point is discussed in detail in *Courtier and Talagrand* (1990). Finally, the quality of the forecasts produced from the results of variational assimilation seems to be as good as one can expect from the models used so far. As far as the meteorological quality of the results is concerned, variational assimilation seems so far to be fully successful.

It goes beyond the scope of these notes to discuss in detail the respective qualities and defects of variational assimilation and of the alternative approach to assimilation of observations, namely sequential assimilation, which is used in present operational numerical weather forecasting. Indeed, a detailed and accurate comparison between those two approaches, in terms of both numerical cost and intrinsic quality of the results, remains to be made. As concerns a possible operational implementation of variational assimilation, the main difficulty seems to be its numerical cost. Even though the use of the adjoint method allows to implement variational assimilation on models on which it would be totally unthinkable to implement it with other methods, ten to fifteen iterations of the minimization process (remember one iteration consists of one forward model integration followed by one adjoint integration) go beyond what could be performed in present operational practice. Progress in numerical methods and computing power may of course drastically change the situation in the future. Another possible difficulty may be due to the presence of "threshold" processes in models. : such processes may be responsible for the existence of secondary minima in the distance-

function (see, e. g., *Douady and Talagrand*, 1990). This is a problem which still has to be studied in detail.

A problem which in many respects is similar to the problem of assimilation of observations is the problem of inversion of satellite radiances (indeed, it can be considered as being one aspect of assimilation of observations). A variational approach to that problem, with explicit use of the adjoint of the radiative transfer equation, has been studied by *Eyre* (1989a, b) and *Thépaut and Moll* (1990). The results obtained by these authors clearly show the efficiency of the variational approach.

3.2 Direct sensitivity studies

If variational assimilation has so far been the meteorological application of the adjoint method which has stimulated most interest, it nevertheless is costly in that it requires the implementation of a large dimension iterative minimization process. A much simpler class of possible applications is provided by direct sensitivity studies, in which one simply wants to determine, for its own sake, the sensitivity of one output parameter of the model with respect to a number of input parameters (see for instance example c) in the Introduction). A very convincing example was given early by *Courtier* (1987), who discovered the detrimental influence of a diurnal thermal wave on an initialization process from the spatial structure of the gradient of an appropriately defined diagnostic quantity with respect to the geopotential field before the initialization. Other sensitivity studies have been performed recently by *Marchuk and Skiba*, (1990) and *Errico and Vukicevic* (1991). Adjoint equations have also been used for studying various aspects of the general problem of the prediction of forecast error by *Urban* (1985), *Lacarra and Talagrand* (1988) and *Veyre* (1990).

4. REFERENCES

Cacuci, D. G., and M. C. G. Hall, 1984: Efficient estimation of feedback effects with application to climate models. J. Atmos. Sci., 41, 2063-2068.

Chao, W. C., and L. P. Chang, 1990: Status of the development of a variational data assimilation system using the adjoint method at Goddard Laboratory for Atmospheres. In: Proceedings of International Symposium on Assimilation of Observations in Meteorology and Oceanography (Clermont-Ferrand, France, July 1990), World Meteorological Organization, Geneva, Switzerland, 355-358.

Courtier, P., 1987: Application of optimal control to numerical meteorological forecasting. Doctoral Thesis, Université Pierre-et-Marie-Curie, Paris, France, (in French).

Courtier, P., and O. Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. Q. J. R. Meteorol. Soc., 113, 1329-1347.

Courtier, P., and O. Talagrand, 1990: Variational assimilation of meteorological observations with the direct and adjoint shallow-water equations. Tellus, 42A, 531-549.

Courtier, P., J. N. Thépaut and O. Talagrand, 1990: 4 dimensional data assimilation using the adjoint of a primitive equation model. In: Proceedings of International Symposium on Assimilation of Observations in Meteorology and Oceanography (Clermont-Ferrand, France, July 1990), World Meteorological Organization, Geneva, Switzerland, 337-340.

Douady, D., and O. Talagrand, 1990: The impact of threshold processes on variational assimilation. In: Proceedings of International Symposium on Assimilation of Observations in Meteorology and Oceanography (Clermont-Ferrand, France, July 1990), World Meteorological Organization, Geneva, Switzerland, 486-487.

Errico, R. M., and T. Vukisevic, 1991: Sensitivity analysis using an adjoint of the PSU/NCAR mesoscale model, Report, National Center for Atmospheric Research, Boulder, CO, USA.

Eyre, J. R., 1989a: Inversion of cloudy satellite sounding radiances by non-linear optimal estimation I: Theory and simulation for TOVS. Q. J. R. Meteorol. Soc., 115, 1001-1026.

Eyre, J. R., 1989b: Inversion of cloudy satellite sounding radiances by non-linear optimal estimation I: Application to TOVS data. Q. J. R. Meteorol. Soc., 115, 1027-1068.

Gill, P.E., W. Murray and M. H. Wright, 1982: Practical optimization, Academic Press, London.

Hall, M. C. G., D. G. Cacuci and M. E. Schlesinger, 1982: Sensitivity analysis of a radiative-convective model by the adjoint method. J. Atmos. Sci., 39, 2038-2050.

Hoffman, R. N., 1986: A four-dimensional analysis exactly satisfying equations of motion. Mon. Wea. Rev., 114, 388-397.

Lacarra, J. F., and O. Talagrand, 1988: Short-range evolution of small perturbations in a barotropic model. Tellus, 40A, 81-95.

Le Dimet, F.X., 1980: A general method for variational objective analysis of meteorological fields. Scientific Report No 30, Laboratoire Associé de Météorologie Physique, Clermont-Ferrand, France (in French).

Le Dimet, F.X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations : theoretical aspects. Tellus, 38A, 97-110.

Lewis, J. M., and J. C. Derber, 1985: The use of adjoint equations to solve a variational adjustment problem with advective constraints. Tellus, 37A, 309-322.

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

Marchuk, G. I., 1974: Numerical solution of the problems of dynamics of the atmosphere and the ocean. Gidrometeoizdat, Leningrad, USSR (in Russian).

Marchuk, G. I., and Yu. K. Skiba, 1990: Role of adjoint functions in studying the sensitivity of a model of the thermal interaction of the atmosphere and ocean to variations in input data. Atm. Ocean. Phys. 26, 335-342 (Translation from the Russian edition).

Morgenstern, J., 1984: How to compute fast a function and all its derivatives. A variation on the theorem of Baur-Stressen. Report no 49, Laboratoire CNRS 168, Université de Nice, France.

Navon, I. M., and D. M. Legler, 1986: Conjugate-gradient methods for large-scale minimization in meteorology. Mon. Wea. Rev., 115, 1479-1502.

Navon, I. M., Xiaolei Zou, K. Johnson, J. Derber and J. Sela, 1990: Variational real-data assimilation with the N.M.C. spectral model. Part I : adiabatic model tests. In: Proceedings of International Symposium on Assimilation of Observations in Meteorology and Oceanography (Clermont-Ferrand, France, July 1990), World Meteorological Organization, Geneva, Switzerland, 341-348.

Penenko, V. V., and N. N. Obraztsov, 1976: A variational initialization method for the fields of the meteorological elements. Soviet Meteorol. Hydrol.,11, 1-11 (Translation from the Russian edition).

Schröter, J., and C. Wunsch, 1986: Solution of nonlinear finite difference ocean models by optimization methods with sensitivity and observational strategy analysis. J. Phys. Ocean., 16, 1855-1874.

Talagrand, O., 1989: Four-dimensional variational assimilation. In: Proceedings of Seminar on Data assimilation and the use of satellite data, ECMWF, Reading, England, September 1988.

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

Thacker, W. C., 1988: Fitting models to inadequate data by enforcing spatial and temporal smoothness. J. Geophys. Res., 93, 10,655-10,665.

Thacker, W. C., and R. B. Long, 1988: Fitting dynamics to data. J. Geophys. Res., 93, 1227-1240.

Thépaut, J.N., and P. Courtier, 1991: submitted for publication in Q. J. R. Meteorol. Soc.

Thépaut, J.N., and P. Moll, 1990: Variational inversion of simulated TOVS radiances using the adjoint technique. Q.J. R. Meteorol. Soc., 116, 1425-1448.

Urban, B., 1985: Maximum error growth in simple meteorological models. Internal report, Ecole Nationale de la Météorologie, Toulouse, France, (in French).

Veyre, P., 1990: Study of the forecast of the forecast skill with the Kalman filter technique. In: Proceedings of International Symposium on Assimilation of Observations in Meteorology and Oceanography (Clermont-Ferrand, France, July 1990), World Meteorological Organization, Geneva, Switzerland, 460.

Wunsch, C., 1987: Using transient tracers : the regularization problem. Tellus, 39B, 477-492.