Q. J. R. Meteorol. Soc. (1987), **113**, pp. 1311–1328 551.509.313:551.511.32

Variational assimilation of meteorological observations with the adjoint vorticity equation. I: Theory

By OLIVIER TALAGRAND
Laboratoire de Météorologie Dynamique, Paris, France
and
PHILIPPE COURTIER*
Direction de la Météorologie, Paris, France

(Received 22 May 1986; revised 18 February 1987. Communicated by Dr A. J. Simmons)

SUMMARY

The following variational approach is taken to the problem of assimilation of meteorological observations: find the solution of the assimilating model which minimizes a given scalar function measuring the 'distance' between a model solution and the available observations. It is shown how the 'adjoint equations' of the model can be used to compute explicitly the 'gradient' of the distance function with respect to the model's initial conditions. The computation of one gradient requires one forward integration of the full model equations over the time interval on which the observations are available, followed by one backward integration of the adjoint equations. Successive gradients thus computed are introduced into a descent algorithm in order to determine the initial conditions which define the minimizing model solution.

The theory is applied to the vorticity equation. Successful numerical experiments performed on a Haurwitz wave are described.

1. INTRODUCTION

The problems raised by the formulation and the implementation of algorithms for the assimilation of meteorological observations are among the important problems encountered in numerical weather prediction. In spite of important progress, two main difficulties have so far prevented the development of fully satisfactory assimilation algorithms. The first of these difficulties results from the strong nonlinearity of the relevant dynamical equations, and the second from the large number of observations to be treated and the large dimension of forecasting models.

A rather detailed account of the present state of the theory of data assimilation has recently been given by Lorenc (1986). In most algorithms proposed or used so far, assimilation is performed as a sequence of corrections performed at successive times on one integration of the forecasting model. Each correction, or analysis, combines background fields produced by the model with new observations. The analysis technique which is generally considered at present as producing the best results is a form of statistical linear regression, usually called 'optimal analysis', or 'optimal interpolation' (McPherson *et al.* 1979; Lorenc 1981; Hollingsworth *et al.* 1985). One great advantage of optimal analysis is that it provides a conceptually simple and internally consistent procedure for treating large quantities of observations which are inhomogeneous in terms of spatial and temporal distribution, nature and accuracy. However, at a more fundamental level, optimal analysis, as it is used now, takes into account the dynamics of the flow only very indirectly. The temporal evolution of the statistical covariances of the forecast error, whose explicit description is required by the principle of the method, is represented by a very simple law. This law is taken independent, or almost independent, of the current state of the atmosphere, and is only a very crude approximation of the complex dynamical processes which determine the real evolution of the forecast error, and which are explicitly represented in the model itself. It ignores in particular all effects due to advection by the

* Present address: European Centre for Medium Range Weather Forecasts, Reading.

flow and provides only a very approximate estimate, at each successive analysis, of the accuracy of the background fields. Indeed, two facts at least (the frequent increase, in the 'initialization' process which follows the assimilation itself, of the difference between the model fields and the corresponding observations on the one hand, and the occurrence of a 'spin-up' period at the beginning of each forecast on the other) clearly show that present analyses are not perfectly adjusted to the dynamics of the model. A natural question is then: could a more accurate description of the dynamics of the assimilating model improve the quality of the assimilation processes and of the subsequent forecasts?

In the case when the basic dynamical equations are linear, optimal interpolation naturally extends to the time dimension under the form of 'Kalman-Bucy filtering' (Petersen 1968; Ghil *et al.* 1981), which provides the theoretical solution to the problem of exact representation of linear dynamics in a sequential assimilation process. Kalman-Bucy filtering is based on the fact that, for linear dynamical systems, the temporal evolution of statistical moments of a given order is independent of moments of higher order. Nonlinear systems, however, do not possess this simple property. The most logical way for extending Kalman-Bucy filtering to nonlinear systems is to use the local 'tangent linear equations' for computing the evolution of the forecast error covariances (Jazwinski 1970; Ghil *et al.* 1982). The tangent linear equations, given an initial perturbation imposed on a model solution, describe the temporal evolution of this perturbation to second order in perturbation amplitude. Although the use of the tangent linear equations seems appropriate as an extension of Kalman-Bucy filtering in the context of data assimilation (Lacarra and Talagrand 1988), it still poses a number of practical problems. In the present state of development of assimilation techniques, any improvement along the lines of optimal interpolation and Kalman-Bucy filtering must be based on an answer to the following question. Is it possible to give a description of the short-term evolution of the forecast error which is at the same time simple enough to be usable in operational practice and accurate enough to improve the quality of present assimilation algorithms? Various works (Lacarra and Talagrand 1988; Ghil, personal communication) are presently in progress in order to answer this question.

Instead of performing successive analyses in the course of one integration of the assimilating model, each individual observation being used only once without feedback to anterior times, an alternative approach to assimilation is to try and adjust globally one model solution to the complete set of available observations. That alternative approach has the theoretical advantage of providing exact consistency between the dynamics of the model and the final results of the assimilation. It led Morel *et al.* (1971) to define and implement a procedure in which the assimilating model is repeatedly integrated forward and backward in time, the observations being constantly introduced into the model. The heuristic idea behind that procedure, supported by a number of numerical results, was that repeated integrations and corrections of the model would converge to a solution which would be compatible, at least to a sufficient degree of accuracy, with the observations. Talagrand (1981a, b), using techniques of stability theory, established criteria for the convergence of such procedures, which explained a number of the numerical results previously obtained by various authors. The same study, based on the use of the local tangent linear equations, also led to the theoretical definition of a numerical procedure maximizing the rate of convergence of the assimilation. It eventually turned out that each iteration of that procedure is essentially the dynamical part (as distinct from the statistical part) of a Kalman-Bucy filter, with which it has in common the property of being based on a linear description of the temporal evolution of the forecast error. It leads to the same basic problem as Kalman-Bucy filtering, namely that of the definition of a practically usable description of the time evolution of the forecast error.

A more systematic and rigorous approach for globally adjusting one nonlinear model solution to a set of observations distributed in time is the following. First define a scalar function which, for any solution of the model equations, measures the 'distance' between that solution and the observations. Then try and determine the particular solution which minimizes that scalar function. This is typically a 'constrained variational' problem, in which the unknowns (the temporal sequence of atmospheric states in the present case) must minimize a given scalar function while at the same time verifying a given set of constraints (the dynamical equations of the model). The first attempts along these lines, in the context of assimilation of meteorological observations, were made on simple cases by Thompson (1969) and Sasaki (1970) (in the now classical terminology of the latter, the problem considered here is a 'strong constraint' problem). The solution of a constrained minimization problem can in principle be obtained by solving the corresponding 'Euler–Lagrange equations' (e.g. Le Dimet and Talagrand 1986). However, in the case of assimilation of meteorological observations, the size and complexity of these equations are such that they can apparently be solved directly only at the price of drastic approximations, which severely limit the possibility of using them in real situations (e.g. Lewis and Bloom 1978, and Bloom 1983). If one notes that a solution of the model is uniquely defined by the corresponding initial state at a given time, the variational problem considered above can be stated in the following terms: find the initial state such that the corresponding model solution minimizes the scalar function measuring the distance to the observations. When stated in these terms, the problem is unconstrained, since no particular condition is imposed on the initial state. Moreover, the numerical dimension of the problem is greatly reduced, since the minimization process must now be performed on the initial state only, and no longer on the whole temporal history of the model. On the other hand, it is now necessary to express in a practically usable form the relationship between the variations of the initial state and the corresponding variations of the distance function. This requires being able to compute numerically the vector of derivatives, i.e. the 'gradient', of the distance function with respect to the parameters which define the initial state. A possible way to determine this gradient is to perturb in turn all components of the initial state and, for each perturbation, to integrate explicitly the model equations and to compute the corresponding variation of the distance function. This approach has been used by Hoffman (1986) who has applied it to a simplified primitive equation model. However, Hoffman himself has pointed out that the numerical cost of this approach is so high as probably to forbid its use in any practical situation.

The present article is devoted to a different numerical approach to the problem of variational data assimilation, which uses techniques of 'optimal control' and, more specifically, the 'adjoint equations' of the assimilating model. The theory of optimal control (see e.g. Lions 1971) deals with the general problem of how the output parameters of a complicated numerical process can be 'controlled' by acting on the input parameters of the process. Among the various tools of optimal control, the adjoint of the local tangent linear equations provides an efficient way for numerically computing the local gradient of a complicated compound function of a set of arguments. The idea of applying adjoint equations to meteorological problems is not new. It was apparently first suggested by Marchuk (1974). Since then, several authors (see e.g. Sadokov and Shteinbok 1977; Hall *et al.* 1982; Cacuci and Hall 1984) have utilized adjoint equations in studying various problems, related in particular to the question of the sensitivity of simple meteorological models with respect to physical parameters. In the context of variational data assimilation, the adjoint of the local tangent linear equations of the assimilating model can be used for computing the gradient of the distance function with respect to the model's initial conditions. This gradient is then used for performing a 'descent step' in the space of

initial conditions, and the process is iterated until some satisfactory approximation of the initial conditions, which minimizes the distance function has been obtained. This specific use of adjoint equations seems to have been first suggested by Penenko and Obraztsov (1976), who performed experiments on a simple linear model. More recently, Derber (1985) and Lewis and Derber (1985) have used adjoint equations in order to adjust independent successive analyses to multi-level quasi-geostrophic models, and have obtained results which clearly show the feasibility and usefulness of this approach. Le Dimet and Talagrand (1986) have given a rather general presentation of the use of adjoint equations in the context of data assimilation, and performed numerical experiments on a one-dimensional nonlinear equation which showed satisfactory numerical convergence of the minimization process.

This paper consists of two parts, devoted respectively to theory and numerical results. In the first part, the theory of adjoint equations is presented in the context of variational assimilation (section 2) under a slightly more general form than in Le Dimet and Talagrand (1986). The theory is then applied in section 3 to the inviscid vorticity equation, and numerical results obtained on a Haurwitz wave are presented in section 4. Section 5 contains a few additional comments. The second part of the paper is devoted to numerical results obtained in the assimilation of real geopotential and wind observations, performed also with the inviscid vorticity equation.

2. The theory of adjoint equations

(a) *General principle*

The most general form so far of the theory of adjoint equations seems to have been presented by Cacuci (1981). We will present it here in the slightly more restrictive context of 'Hilbert spaces', in which it takes a form which is conceptually very simple, and largely sufficient for the applications we will consider. A Hilbert space is essentially a (possibly infinite-dimensional) vector space on which the ordinary concepts of finite-dimensional Euclidean geometry are valid. The theory of partial differential equations makes abundant use of functional Hilbert spaces, i.e. of spaces whose vectors, or points, are functions. Technically (see, e.g., Reed and Simon 1980), a Hilbert space is a vector space on which an inner product has been defined and which is 'complete' with respect to the metric induced by this inner product. This latter condition, which basically means that the limit of a convergent sequence of vectors in the space is necessarily itself in the space, is absolutely essential from a theoretical point of view. Its practical implications will be briefly discussed in the conclusion. Because of the completeness of the real line, a finite-dimensional vector space on which an inner product has been defined is always complete, so that the developments of this section will always be valid in a practical numerical situation. We will make use of the following two basic properties of Hilbert spaces.

(i) Let $\mathcal{F}$ be a Hilbert space, with inner product denoted by $(\ ,\ )$, and $\mathbf{v} \rightarrow \mathcal{J}(\mathbf{v})$ a differentiable scalar function defined on $\mathcal{F}$. At any point in $\mathcal{F}$, the differential $\delta\mathcal{J}$ of $\mathcal{J}$ can be expressed in terms of the differential $\delta\mathbf{v}$ of $\mathbf{v}$ by the inner product

$$\delta\mathcal{J} = (\nabla_{\mathbf{v}}\mathcal{J}, \delta\mathbf{v}) \tag{1}$$

where $\nabla_{\mathbf{v}}\mathcal{J}$ is the 'gradient' of $\mathcal{J}$ with respect to $\mathbf{v}$ and is uniquely defined. When $\mathcal{F}$ has finite dimension and is described by orthonormal coordinates v_i, the components of $\nabla_{\mathbf{v}}\mathcal{J}$ are of course the partial derivatives $\partial\mathcal{J}/\partial v_i$, but the concept of a gradient is much more general. In the following, rather than using explicit coordinates and indices, it will be more convenient to use the general abstract formula (1).

(ii) Let $\mathcal{E}$ be another Hilbert space, with inner product denoted by $\langle\ ,\ \rangle$, and $\mathbf{L}$ a continuous linear operator from $\mathcal{E}$ into $\mathcal{F}$. There exists a unique continuous linear operator $\mathbf{L}^*$ from $\mathcal{F}$ into $\mathcal{E}$ such that the following equality between inner products holds for any $\mathbf{u}$ belonging to $\mathcal{E}$ and any $\mathbf{v}$ belonging to $\mathcal{F}$:

$$(\mathbf{v}, \mathbf{Lu}) = \langle \mathbf{L}^*\mathbf{v}, \mathbf{u} \rangle. \tag{2}$$

$\mathbf{L}^*$ is called the 'adjoint operator' of $\mathbf{L}$. When $\mathcal{E}$ and $\mathcal{F}$ have finite dimensions and are described by orthonormal coordinates, it is well known that the matrix which represents $\mathbf{L}^*$ is the transpose (or transconjugate in case of complex components) of the matrix which represents $\mathbf{L}$. But again, the concept of an adjoint operator is much more general and, rather than using explicit indices, we will use in the following the general abstract formula (2). Appendix A contains a reminder of a number of basic definitions relating to adjoint operators, and of some of their properties.

$\mathcal{F}$, $\mathcal{E}$, and $\mathcal{J}$ still having the same meaning, we now consider a (not necessarily linear) differentiable function $\mathbf{u} \rightarrow \mathbf{v} = \mathbf{G}(\mathbf{u})$ of $\mathcal{E}$ into $\mathcal{F}$. $\mathcal{J}(\mathbf{v}) = \mathcal{J}[\mathbf{G}(\mathbf{u})]$ is a compound function of $\mathbf{u}$. The differential of $\mathcal{J}$ is given by (1), while the differential of $\mathbf{v}$ is equal to

$$\delta\mathbf{v} = \mathbf{G}'\delta\mathbf{u} \tag{3}$$

where $\mathbf{G}'$ is the linear operator from $\mathcal{E}$ into $\mathcal{F}$ obtained by differentiation of $\mathbf{G}$. Substituting (3) into (1) and introducing the adjoint $\mathbf{G}'^*$ of $\mathbf{G}'$, one obtains

$$\delta\mathcal{J} = (\nabla_{\mathbf{v}}\mathcal{J}, \mathbf{G}'\delta\mathbf{u}) = \langle \mathbf{G}'^*\nabla_{\mathbf{v}}\mathcal{J}, \delta\mathbf{u} \rangle.$$

By definition of a gradient, this shows that the gradient $\nabla_{\mathbf{u}}\mathcal{J}$ of $\mathcal{J}$ with respect to $\mathbf{u}$ is equal to

$$\nabla_{\mathbf{u}}\mathcal{J} = \mathbf{G}'^*\nabla_{\mathbf{v}}\mathcal{J}. \tag{4}$$

Formula (4) is at the basis of the use of adjoint equations in control theory, and provides a very efficient way for determining numerically $\nabla_{\mathbf{u}}\mathcal{J}$. In the case when $\mathbf{v}$ is an explicit but complicated function of $\mathbf{u}$ (as in the sequel, where the operation $\mathbf{u} \rightarrow \mathbf{v} = \mathbf{G}(\mathbf{u})$ will denote the integration of a numerical model of the atmospheric flow), it will in general be impossible to find a numerically usable analytic expression for $\nabla_{\mathbf{u}}\mathcal{J}$. A conceivable way to obtain numerical estimates of the components of $\nabla_{\mathbf{u}}\mathcal{J}$ would be to perturb in turn every component of $\mathbf{u}$ and, for each perturbation, to explicitly compute $\mathbf{v} = \mathbf{G}(\mathbf{u})$ and the resulting perturbation on $\mathcal{J}$. But this would require as many explicit computations of $\mathbf{G}(\mathbf{u})$ as there are components in $\mathbf{u}$ and, as said in the introduction, would be totally impracticable for any problem with large dimension. Now, Eq. (4) shows that, if a computer program is available which calculates $\mathbf{G}'^*\mathbf{w}$ for given $\mathbf{w}$, $\nabla_{\mathbf{u}}\mathcal{J}$ can easily be computed from $\nabla_{\mathbf{v}}\mathcal{J}$. $\nabla_{\mathbf{v}}\mathcal{J}$ itself will easily be determined whenever $\mathcal{J}$ is a 'simple' function of $\mathbf{v}$. A program which computes $\mathbf{G}'^*\mathbf{w}$ for given $\mathbf{w}$ will normally be of the same order of complexity as a program which computes $\mathbf{G}(\mathbf{u})$ for given $\mathbf{u}$. Therefore the successive operations needed for determining $\nabla_{\mathbf{u}}\mathcal{J}$ (namely, computing $\mathbf{v} = \mathbf{G}(\mathbf{u})$, then $\nabla_{\mathbf{v}}\mathcal{J}$ and finally $\nabla_{\mathbf{u}}\mathcal{J}$ through (4)) will take at most a few times the cost of a direct computation of $\mathbf{G}(\mathbf{u})$. This achieves an enormous gain over direct explicit perturbations of the components of $\mathbf{u}$.

(*b*) *Application to data assimilation*

From now on, $\mathbf{u}$ will denote the initial conditions from which a numerical model of the atmospheric flow is integrated, the operation $\mathbf{u} \rightarrow \mathbf{v} = \mathbf{G}(\mathbf{u})$ will be the integration of the model, $\mathbf{v}$ denoting the time sequence of successive model states produced by the

integration. The scalar function $\mathcal{J}(\mathbf{v})$ will be the scalar function which measures the distance between $\mathbf{v}$ and the available observations, assumed to be distributed over a time interval $[t_0, t_1]$.

The model evolution equation is written as

$$d\mathbf{x}/dt = \mathbf{F}(\mathbf{x}) \tag{5}$$

where the state vector $\mathbf{x}$ belongs to a Hilbert space $\mathcal{E}$ whose inner product is denoted by $\langle\ ,\ \rangle$, and $\mathbf{F}$ is a regular function of $\mathcal{E}$ into itself. For any $\mathbf{u}$ belonging to $\mathcal{E}$, the initial condition $\mathbf{x}(t_0) = \mathbf{u}$ defines a unique solution $\mathbf{x}(t)$ to (5).

The distance function $\mathcal{J}$ will be taken as

$$\mathcal{J} = \int_{t_0}^{t_1} H[\mathbf{x}(t), t]\, dt \tag{6}$$

where, for given $\mathbf{x}$ belonging to $\mathcal{E}$ and t belonging to $[t_0, t_1]$, $H[\mathbf{x}, t]$ is a scalar measuring the distance between $\mathbf{x}$ and the observations available at time t (for instance, the squared norm of the difference between the observations at time t and the corresponding components of $\mathbf{x}$). For a given initial condition $\mathbf{u}$ and for the corresponding solution $\mathbf{x}(t)$ of (5), the first-order variation $\delta\mathcal{J}$ resulting from a variation $\delta\mathbf{u}$ of $\mathbf{u}$ is equal to

$$\delta\mathcal{J} = \int_{t_0}^{t_1} \langle \nabla_{\mathbf{x}} H(t), \delta\mathbf{x}(t) \rangle\, dt \tag{7}$$

where $\nabla_{\mathbf{x}} H(t)$ is the gradient of $H[\mathbf{x}, t]$ with respect to its first argument $\mathbf{x}$, taken at point $(\mathbf{x}(t), t)$, and $\delta\mathbf{x}(t)$ is the first-order variation of $\mathbf{x}(t)$ resulting from the perturbation $\delta\mathbf{u}$. The variation $\delta\mathbf{x}(t)$ is obtained from $\delta\mathbf{u}$ by integrating the 'tangent linear equation' (also called the 'linearized perturbation equation') relative to the solution $\mathbf{x}(t)$

$$d\delta\mathbf{x}/dt = \mathbf{F}'(t)\, \delta\mathbf{x} \tag{8}$$

where $\mathbf{F}'(t)$ is the operator obtained by differentiating $\mathbf{F}$ with respect to $\mathbf{x}$, and taken at point $\mathbf{x}(t)$. Equation (8) being linear, its solution at a given time t depends linearly on the initial condition at time t_0, which can be written as

$$\delta\mathbf{x}(t) = \mathbf{R}(t, t_0)\, \delta\mathbf{u}$$

where $\mathbf{R}(t, t_0)$ is a well-defined linear operator, called the 'resolvent' of (8) between times t_0 and t. The resolvent $\mathbf{R}(t, t')$ is defined more generally for any two times t and t' lying between t_0 and t_1 and possesses the following two properties:

$$\mathbf{R}(t, t) = \mathbf{I} \qquad \text{for any } t \tag{9a}$$

where $\mathbf{I}$ is the unit operator on $\mathcal{E}$; and

$$\frac{\partial}{\partial t}\mathbf{R}(t, t') = \mathbf{F}'(t)\mathbf{R}(t, t') \tag{9b}$$

for any t and t'.

Equation (7) can now be rewritten as

$$\delta\mathcal{J} = \int_{t_0}^{t_1} \langle \nabla_{\mathbf{x}} H(t), \mathbf{R}(t, t_0)\delta\mathbf{u} \rangle\, dt$$

$$= \int_{t_0}^{t_1} \langle \mathbf{R}^*(t, t_0)\nabla_{\mathbf{x}} H(t), \delta\mathbf{u} \rangle\, dt$$

$$= \left\langle \int_{t_0}^{t_1} \mathbf{R}^*(t, t_0)\nabla_{\mathbf{x}} H(t)\, dt, \delta\mathbf{u} \right\rangle \tag{10}$$

where we have introduced, for all t, the adjoint $\mathbf{R}^*(t, t_0)$ of $\mathbf{R}(t, t_0)$. It is seen from the last line that the gradient of $\mathcal{J}$ with respect to $\mathbf{u}$ is equal to

$$\nabla_{\mathbf{u}}\mathcal{J} = \int_{t_0}^{t_1} \mathbf{R}^*(t, t_0)\nabla_{\mathbf{x}} H(t)\, dt. \tag{11}$$

We now introduce the 'adjoint equation' of (8)

$$-d\delta'\mathbf{x}/dt = \mathbf{F}'^*(t)\delta'\mathbf{x} \tag{12}$$

whose variable $\delta'\mathbf{x}$ also belongs to $\mathcal{E}$, and where $\mathbf{F}'^*(t)$ is the adjoint of $\mathbf{F}'(t)$. Equation (12) is linear, and we will denote by $\mathbf{S}(t', t)$ its resolvent between times t and t'. For any two solutions $\delta\mathbf{x}(t)$ and $\delta'\mathbf{x}(t)$ of the direct and adjoint equations (8) and (12) respectively, the inner product $\langle \delta\mathbf{x}(t), \delta'\mathbf{x}(t)\rangle$ is constant with time since

$$\begin{aligned}\frac{d}{dt}\langle \delta\mathbf{x}(t), \delta'\mathbf{x}(t)\rangle &= \left\langle \frac{d\delta\mathbf{x}}{dt}(t), \delta'\mathbf{x}(t) \right\rangle + \left\langle \delta\mathbf{x}(t), \frac{d\delta'\mathbf{x}}{dt}(t) \right\rangle \\ &= \langle \mathbf{F}'(t)\delta\mathbf{x}(t), \delta'\mathbf{x}(t)\rangle - \langle \delta\mathbf{x}(t), \mathbf{F}'^*(t)\delta'\mathbf{x}(t)\rangle \\ &= 0.\end{aligned}$$

Let $\mathbf{y}$ and $\mathbf{y}'$ be any two elements of $\mathcal{E}$. The solution of the direct equation (8) defined by the initial condition $\mathbf{y}$ at time t assumes at time t' the value $\mathbf{R}(t', t)\mathbf{y}$, while the solution of the adjoint equation (12) defined by the condition $\mathbf{y}'$ at time t' assumes at time t the value $\mathbf{S}(t, t')\mathbf{y}'$. The corresponding equality between inner products therefore reads

$$\langle \mathbf{R}(t', t)\mathbf{y}, \mathbf{y}'\rangle = \langle \mathbf{y}, \mathbf{S}(t, t')\mathbf{y}'\rangle$$

which, being valid for any $\mathbf{y}$ and $\mathbf{y}'$, shows that $\mathbf{S}(t, t')$ is the adjoint of $\mathbf{R}(t', t)$, i.e. the resolvent of the adjoint equation between t' and t is the adjoint of the resolvent of the direct equation between t and t'. Expression (11) then becomes

$$\nabla_{\mathbf{u}}\mathcal{J} = \int_{t_0}^{t_1} \mathbf{S}(t_0, t)\nabla_{\mathbf{x}} H(t)\, dt. \tag{13}$$

We consider next the 'inhomogeneous adjoint equation'

$$-d\delta'\mathbf{x}/dt = \mathbf{F}'^*(t)\delta'\mathbf{x} + \nabla_{\mathbf{x}} H(t). \tag{14}$$

The solution of that equation defined by the condition $\delta'\mathbf{x}(t_1) = 0$ is equal to

$$\delta'\mathbf{x}(t) = \int_{t}^{t_1} \mathbf{S}(t, \tau)\nabla_{\mathbf{x}} H(\tau)\, d\tau \tag{15}$$

as is easily verified by using the resolvent properties (9). Equation (13) now shows that $\nabla_{\mathbf{u}}\mathcal{J}$ is equal to $\delta'\mathbf{x}(t_0)$.

In summary, the gradient $\nabla_{\mathbf{u}}\mathcal{J}$ can be obtained, for given $\mathbf{u}$, by performing the following operations:

(i) Starting from $\mathbf{u}$ at time t_0, integrate the basic evolution equation (5) from t_0 to t_1. Store the values thus computed for $\mathbf{x}(t)$, $t_0 \leqslant t \leqslant t_1$.

(ii) Starting from $\delta'\mathbf{x}(t_1) = 0$, integrate the inhomogeneous adjoint equation (14) backwards in time from t_1 to t_0, the operator $\mathbf{F}'^*(t)$ and the gradient $\nabla_{\mathbf{x}} H(t)$ being

determined, at each time t, from the values $\mathbf{x}(t)$ computed in the direct integration of (5). The final value $\delta'\mathbf{x}(t_0)$ is the required gradient $\nabla_{\mathbf{u}}\mathcal{J}$.

Remark. The approach which has just been presented does not exactly follow the general principle described in subsection (*a*) in that the adjoint equations (12) and (14) have been introduced without *a priori* justification, and have only been verified to lead to the gradient $\nabla_{\mathbf{u}}\mathcal{J}$. The interested reader will find in Talagrand and Courtier (1985) an approach which rigorously leads to (12) and (14).

(*c*) *Descent algorithms*

Once a numerical algorithm is available for computing $\nabla_{\mathbf{u}}\mathcal{J}$, a 'descent algorithm' can be used in order to determine the value $\mathbf{u}_{\min}$ which minimizes $\mathcal{J}$. Successive estimates $\mathbf{u}_n$ of $\mathbf{u}_{\min}$ are obtained through 'descent steps' of the form

$$\mathbf{u}_{n+1} = \mathbf{u}_n - \rho_n \mathbf{D}_n$$

where, for each n, $\mathbf{D}_n$ is a 'descent direction' determined from the successive gradients $\nabla_{\mathbf{u}}\mathcal{J}(\mathbf{u}_n)$, $\nabla_{\mathbf{u}}\mathcal{J}(\mathbf{u}_{n-1}), \ldots$, and ρ_n is an appropriate scalar. If $\mathbf{D}_n$ and ρ_n are properly chosen, the sequence $\mathbf{u}_n$ will tend to $\mathbf{u}_{\min}$. Three classical descent algorithms, which will be used in the numerical experiments described in both parts of this article, are the 'steepest descent' algorithm (in which $\mathbf{D}_n = \nabla_{\mathbf{u}}\mathcal{J}(\mathbf{u}_n)$ for any n), the 'conjugate gradient' algorithm, and the 'quasi-Newton', or 'variable metric', algorithm. These algorithms are described, and their properties compared, in Gill *et al.* (1982).

3. The adjoint of the vorticity equation

(*a*) *The continuous case*

The vorticity equation

$$\partial\zeta/\partial t = J(\zeta + f, \Delta^{-1}\zeta) \tag{16}$$

describes the time evolution of a two-dimensional non-divergent and inviscid flow at the surface of a rotating sphere Σ. In (16), ζ and f are the vorticities of the relative motion and basic rotation respectively, t is time, and Δ^{-1} the inverse two-dimensional Laplacian operator on Σ. J is the Jacobian operator

$$J(a, b) = \nabla a \times \nabla b = \nabla \times (a \nabla b)$$

∇ now denoting first-order differentiation tangent to Σ.

For a given solution ζ of (16), the tangent linear equation, analogous to (8), reads

$$\partial\delta\zeta/\partial t = J(\delta\zeta, \Delta^{-1}\zeta) + J(\zeta + f, \Delta^{-1}\delta\zeta). \tag{17}$$

The determination of the adjoint of (17) requires the prior definition of an inner product on the space of all possible vorticity fields, i.e. the space of all regular functions on the sphere with mean equal to 0. The total kinetic energy K corresponding to a given vorticity field ζ is given by

$$2K = \int_{\Sigma} (\nabla\Delta^{-1}\zeta \cdot \nabla\Delta^{-1}\zeta)\, d\Sigma \tag{18}$$

where the dot denotes scalar product of ordinary vectors in physical space. Expression (18) leads to the following natural definition for an inner product of two vorticity fields

ζ and ζ'

$$\langle \zeta, \zeta' \rangle = \int_\Sigma (\nabla \Delta^{-1} \zeta \cdot \nabla \Delta^{-1} \zeta')\, d\Sigma \tag{19a}$$

which, after integration by parts, can be written as

$$\langle \zeta, \zeta' \rangle = -\int_\Sigma \zeta(\Delta^{-1} \zeta')\, d\Sigma = -\int_\Sigma \zeta'(\Delta^{-1} \zeta)\, d\Sigma. \tag{19b}$$

It immediately results that the Laplacian Δ and its inverse Δ^{-1} are self-adjoint with respect to this inner product, i.e. for any ζ and ζ'

$$\langle \Delta\zeta, \zeta' \rangle = \langle \zeta, \Delta\zeta' \rangle \tag{20a}$$

$$\langle \Delta^{-1}\zeta, \zeta' \rangle = \langle \zeta, \Delta^{-1}\zeta' \rangle. \tag{20b}$$

The Jacobian J has the property that for any three scalar fields a, b, c

$$\int_\Sigma J(a, b)c\, d\Sigma = \int_\Sigma aJ(b, c)\, d\Sigma \tag{21}$$

as seen from the following equalities

$$\int_\Sigma J(a, b)c\, d\Sigma = \int_\Sigma [\nabla \times (a \nabla b)]c\, d\Sigma = \int_\Sigma a(\nabla b \times \nabla c)\, d\Sigma = \int_\Sigma aJ(b, c)\, d\Sigma \tag{22}$$

where integration by parts was used to go from the second to the third term.

Three scalar fields α, $\delta\zeta$, $\delta\zeta'$ being given, use of (19b) and (21) yields

$$\begin{aligned}
\langle J(\delta\zeta, \alpha), \delta\zeta' \rangle &= -\int_\Sigma J(\delta\zeta, \alpha)\Delta^{-1}\delta\zeta'\, d\Sigma \\
&= -\int_\Sigma \delta\zeta\, J(\alpha, \Delta^{-1}\delta\zeta')\, d\Sigma \\
&= -\int_\Sigma \delta\zeta\Delta^{-1}[\Delta J(\alpha, \Delta^{-1}\delta\zeta')]\, d\Sigma \\
&= \langle \delta\zeta, \Delta J(\alpha, \Delta^{-1}\delta\zeta') \rangle
\end{aligned}$$

which shows that, for given α, the adjoint of the linear operator $\delta\zeta \to J(\delta\zeta, \alpha)$ is the operator $\delta\zeta' \to \Delta J(\alpha, \Delta^{-1}\delta\zeta')$. Using the fact that Δ^{-1} is self-adjoint (Eq. (20b)) and the result that the adjoint of the product of two operators is the product of their adjoints taken in reverse order (see appendix A), we obtain the adjoints of the two terms on the right-hand side of (17) and the following expression for the adjoint equation, analogous to (12)

$$\partial\delta'\zeta/\partial t = \Delta J(\Delta^{-1}\delta'\zeta, \Delta^{-1}\zeta) + J(\zeta + f, \Delta^{-1}\delta'\zeta). \tag{23}$$

(*b*) *The discretized case*

In any practical situation, one will use a discretized model of the basic evolution equation (5). If one wants to adjust a solution of the model to a set of observations, it will of course be necessary to use the *adjoint of the model itself*, and the foregoing developments, performed under the implicit hypothesis that the fields under consideration were continuous with respect to both space and time, will have to be derived again for the particular discretization of the model.

The experiments whose results are presented below and in part II of this paper were performed with a pseudo-spectral model of the vorticity equation (16), built on the spherical harmonics Y_n^m, with triangular truncation at degree N ($0 \leq n \leq N$, $-n \leq m \leq n$). A scalar field ψ is then defined by its components ψ_n^m along the spherical harmonics. The spherical harmonic expression for the Laplacian operator

$$(\Delta\psi)_n^m = -\{n(n+1)/a^2\}\psi_n^m$$

where a is the radius of the sphere, leads to the following discretized expression for the inner product (19)

$$\langle \zeta, \zeta' \rangle_N = \sum^N \{a^2/n(n+1)\}\zeta_n^m(\zeta_n'^m)^* \tag{24}$$

where the star denotes imaginary conjugation (this notation is perfectly consistent with the notation used for adjoints, since the adjoint of a complex number z, considered as a linear operator from the field of complex numbers into itself, is the imaginary conjugate of z), and where Σ^N means that the summation extends over all harmonics (n, m) with order $1 \leq n \leq N$.

The partial derivatives appearing in the Jacobian are computed in the model spectral space, and are therefore exact. Following the standard procedure for pseudo-spectral models (Eliasen *et al.* 1970), the quadratic terms are computed in physical space on a collocation grid with enough resolution to ensure that the harmonics resolved in spectral space are free from aliasing errors. The model's discretized Jacobian is therefore also free from aliasing errors. It is shown in appendix B that the adjoint of the unaliased discretized analogue of an operator **L** is the unaliased discretized analogue of the adjoint of **L**. Under these conditions, the model adjoint equation is given by (23), the Jacobian J being simply replaced by the same unaliased approximation as in the integration of the basic equation (16).

The time discretization of the model uses the leapfrog scheme, initialized with a forward Euler step. The corresponding modifications to apply to the developments of section 2, and the resulting time discretization to use for the adjoint equation (23), are described in appendix C.

4. Numerical results: the case of a Haurwitz wave

The numerical model is truncated at degree $N = 21$. A vorticity field is then completely described by 483 independent real parameters (taking into account the fact that the $m = n = 0$ component is necessarily zero), among which 252 parameters are components along harmonics which are symmetric with respect to the equator ($n - m$ even), and 231 are components along harmonics which are antisymmetric ($n - m$ odd). However, all numerical experiments were performed with antisymmetric vorticity fields, so that the effective number of parameters with respect to which the distance function $\mathcal{J}$ was minimized was equal to 231. The time discretization step of the model was one hour.

The integration of the adjoint equation (23) required about twice as much computing time as the integration of the direct equation (16). This doubling is basically due to the fact that differentiation has produced two Jacobians in Eq. (23).

The initial condition obtained by taking the real part of

$$\zeta = (2/\sqrt{3})\alpha\, Y_1^0 + \beta Y_n^m \tag{25}$$

where α is real and β complex, defines a Haurwitz wave with wavenumber m. The corresponding time evolution, under the continuous equation (16), is a solid rotation of

the vorticity field about the polar axis, with angular velocity

$$\alpha - 2(\Omega + \alpha)/(n(n+1))$$

where Ω is the basic rotation. Numerical experiments have been performed using 'observations' taken out of a model integration of a Haurwitz wave with the following values for the relevant parameters: $\Omega = 1$ rotation/day, $n = 5, m = 4$, $\alpha = 7{\cdot}27 \times 10^{-6}\text{s}^{-1}$. The angular velocity of the wave is then $9{\cdot}55°\text{day}^{-1}$. The corresponding vorticity field is shown in Fig. 1.

In a first experiment, the 'observations' consisted of the complete space–time history $\zeta_{\text{OBS}}(t)$ of the vorticity field over a 12-hour time interval $[t_0, t_1]$. The corresponding rotation about the polar axis was 4·77°. The distance-function to be minimized was defined as

$$\mathcal{J} = \sum \langle \zeta(t) - \zeta_{\text{OBS}}(t), \zeta(t) - \zeta_{\text{OBS}}(t) \rangle$$

where the summation extends over all time-steps. The corresponding term $\nabla_x H$ to be added as a 'forcing' in the integration of the inhomogeneous adjoint equation (14) is readily seen to be equal to $2[\zeta(t) - \zeta_{\text{OBS}}(t)]$, i.e. for each harmonic Y_n^m the term $2[\zeta(t) - \zeta_{\text{OBS}}(t)]_n^m$ is added at each time-step to the corresponding component of the field produced by the integration of the adjoint.

The 'observations' $\zeta_{\text{OBS}}(t)$ having been produced by the model itself, the minimizing $\zeta(t)$ is of course $\zeta_{\text{OBS}}(t)$, and the corresponding value of $\mathcal{J}$ is 0. Starting the minimization from an atmosphere at rest, and using the conjugate gradient algorithm for the descent process, the value of $\mathcal{J}$ decreased by about two orders of magnitude at each descent step, and the vorticity field $\zeta_{\text{OBS}}(t_0)$ was reconstructed to within an accuracy of 10^{-9}s^{-1} after five steps. The corresponding difference would of course be invisible in Fig. 1.

In a second experiment, the 'observations' consisted of the complete vorticity field at time t_1, and the distance function was defined as

$$\mathcal{J} = \langle \zeta(t_1) - \zeta_{\text{OBS}}(t_1), \zeta(t_1) - \zeta_{\text{OBS}}(t_1) \rangle.$$

The adjustment was performed on the same 12-hour interval $[t_0, t_1]$ preceding t_1, i.e. it was the vorticity field at time t_0 which was varied in order to minimize $\mathcal{J}$. Starting again from an atmosphere at rest, and using again the conjugate descent algorithm, the initial vorticity field $\zeta_{\text{OBS}}(t_0)$ was reconstructed to within an accuracy of 10^{-8}s^{-1} in eight descent steps. This slower convergence than before is due to the fact that less information was used in order to force the initial conditions towards the minimizing values and, more importantly, that a 12-hour interval separated the 'observations' from the time at which the initial conditions were to be estimated. This of course made accurate estimation of those initial conditions more difficult.

5. Additional comments

We will delay general conclusions until the end of the second part of this paper and limit ourselves for the time being to two theoretical comments.

The first is relative to the condition of completeness, without which the notions of a gradient or of an adjoint operator are not properly defined. It has already been said that a finite-dimensional vector space is complete, so that no difficulty can arise in practical numerical situations as to the existence of gradients or adjoints. Now, it can be very useful, in order to determine the explicit expression of a discretized adjoint, to know the corresponding continuous expression. In this work, the explicit knowledge of the

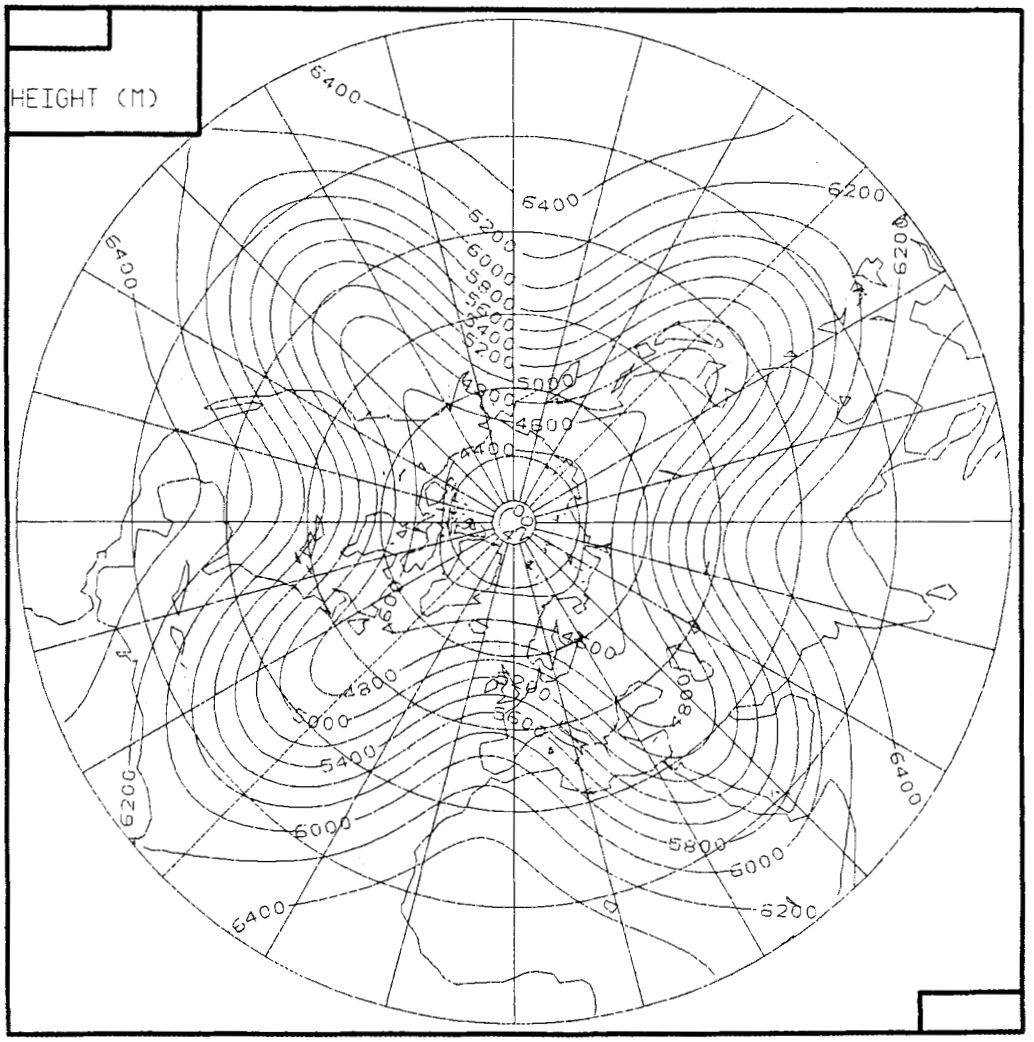

Figure 1. Vorticity field used for the numerical experiments reported in the text. Units: $10^{-6}s^{-1}$; contour interval $200 \times 10^{-6}s^{-1}$.

continuous adjoint (23) has been fundamental for determining its discretized analogue. Infinite-dimensional functional spaces are not necessarily complete, but a basic theoretical result is that they can always be made complete by appropriately enlarging their definitions. However, the 'completion' thus performed will usually require the introduction of 'generalized' functions, such as Dirac distributions, which may then appear in the expression of adjoints. This situation is likely to occur especially when taking the adjoints of differential operators on a bounded domain. These adjoints will usually contain Dirac terms along the boundary.

The question of what the adjoint equations exactly are may seem intriguing. It is important to stress that an adjoint integration is not a backward integration of the basic dynamical equations. The quantities produced at a given time t by an adjoint integration are *not* physical fields at time t, but *partial derivatives* of the function $\mathcal{J}$ with respect to

physical fields at that time. The difference between an adjoint integration and a backward integration of the basic dynamical equations becomes particularly significant when the basic equations contain diffusive or dissipative terms, whose backward integration is usually, from a mathematical point of view, an ill-posed problem. The linear diffusion equation, for instance, contains a Laplacian, which is self-adjoint. The presence of the minus sign in the adjoint equation (14) therefore ensures that the integration of the adjoint diffusion equation is well-posed for integration into the past. More generally, whenever the integration of the basic equation is well-posed only for integration into the future, the integration of the corresponding adjoint equation will be well-posed only for integration into the past.

Along the same lines, it is interesting to compare adjoint assimilation with forward–backward assimilation, as suggested by Morel *et al.* (1971). In both approaches, the assimilating model is integrated repeatedly over the period where the observations are available in order to obtain a progressive adjustment of the model to these observations. The major difference is that the backward integrations were performed by Morel *et al.* with the unmodified model, while they are performed here with the adjoint model. A stability argument developed by Talagrand (1981a) shows that a sufficient condition for convergence of an assimilation of the Morel *et al.* type is that the linearized perturbation equation be antisymmetric. In the present paper's notation, this condition means $\mathbf{F}'^*(t) = -\mathbf{F}'(t)$. With hindsight, this result takes on a new significance. An antisymmetric equation is identical with its adjoint, so that integrating the full equation backwards in time is the same thing (as long as perturbations with respect to some basic solution are considered) as integrating the adjoint. This explains the success of the method of Morel *et al.* in the case of an antisymmetric equation. More generally, it clearly appears that using the adjoint equation is the mathematically proper and rigorous way to achieve the original goal which was heuristically assigned to forward–backward assimilation, namely adjusting a model to observations distributed in time.

APPENDIX A

Basic facts about adjoint operators

Let $\mathscr{E}$, $\mathscr{F}$, $\mathscr{G}$ be three Hilbert spaces, with inner products denoted by $\langle\ ,\ \rangle$, $(\ ,\ )$ and $[\ ,\ ]$ respectively. Let $\mathbf{u} \rightarrow \mathbf{Lu}$ be a continuous linear operator from $\mathscr{E}$ into $\mathscr{F}$, and $\mathbf{v} \rightarrow \mathbf{Mv}$ a continuous linear operator from $\mathscr{F}$ into $\mathscr{G}$. The product $\mathbf{u} \rightarrow \mathbf{MLu}$ is a continuous linear operator from $\mathscr{E}$ into $\mathscr{G}$. Two successive uses of the adjointness equation (2) lead to

$$[\mathbf{w}, \mathbf{MLu}] = (\mathbf{M}^*\mathbf{w}, \mathbf{Lu}) = \langle \mathbf{L}^*\mathbf{M}^*\mathbf{w}, \mathbf{u} \rangle$$

which shows that the adjoint $(\mathbf{ML})^*$ of $\mathbf{ML}$ is equal to

$$(\mathbf{ML})^* = \mathbf{L}^*\mathbf{M}^*.$$

The adjoint of a product of two operators is therefore the product of their adjoints, taken in reverse order. This result immediately generalizes to the product of any number of operators.

Let $\mathbf{L}$ be a continuous linear operator from some Hilbert space $\mathscr{E}$ into itself. Its adjoint $\mathbf{L}^*$ is also an operator from $\mathscr{E}$ into itself. An operator such that $\mathbf{L}^* = \mathbf{L}$ is said to be 'symmetric', or 'self-adjoint'. An operator such that $\mathbf{L}^* = -\mathbf{L}$ is said to be 'antisymmetric', or 'skew symmetric'. It is well known that the eigenvalues of a symmetric

(antisymmetric) operator are real (purely imaginary) and that eigenvectors corresponding to distinct eigenvalues are orthogonal.

Let $\mathbf{F}$ be a continuous linear operator from $\mathscr{E}$ into $\mathscr{F}$ which conserves the inner product in the sense that for any $\mathbf{u}$ and $\mathbf{u}'$ belonging to $\mathscr{E}$

$$(\mathbf{Fu}, \mathbf{Fu}') = \langle \mathbf{u}, \mathbf{u}' \rangle. \tag{A.1}$$

$\mathbf{F}$ is said to be 'unitary'. Introducing the adjoint $\mathbf{F}^*$ of $\mathbf{F}$, (A.1) becomes

$$\langle \mathbf{u}, \mathbf{F}^*\mathbf{Fu}' \rangle = \langle \mathbf{u}, \mathbf{u}' \rangle$$

which, being valid for any $\mathbf{u}$ and $\mathbf{u}'$, shows that $\mathbf{F}^*\mathbf{F}$ is the unit operator for $\mathscr{E}$.

$$\mathbf{F}^*\mathbf{F} = \mathbf{I}$$

$\mathbf{F}^*$ is a left-inverse of $\mathbf{F}$. $\mathbf{F}$ can have a left-inverse if and only if it is injective, i.e. $\mathbf{Fu} \neq \mathbf{Fu}'$ whenever $\mathbf{u} \neq \mathbf{u}'$. Two cases may occur

(i) $\mathbf{F}$ is exactly invertible, i.e. for any $\mathbf{v}$ belonging to $\mathscr{F}$ there is a $\mathbf{u}$ in $\mathscr{E}$ such that $\mathbf{Fu} = \mathbf{v}$. In this case $\mathbf{F}^*$ is the exact inverse of $\mathbf{F}$. Examples of invertible unitary operators are rotations and symmetries in Euclidean space. Another example is Fourier transformation, for which Parseval relationship expresses an equality of type (A.1) between the L^2 norms of a function and its transform. The adjoint of Fourier transformation is its inverse.

(ii) $\mathbf{F}$ is not invertible, i.e. there exist $\mathbf{v}$s in $\mathscr{F}$ which are not images by $\mathbf{F}$ of elements of $\mathscr{E}$. $\mathbf{F}$ has then infinitely many left-inverses, which coincide on the image $\mathbf{F}(\mathscr{E})$ of $\mathscr{E}$. It is easy to see that, among these left-inverses, the adjoint $\mathbf{F}^*$ is the composition of orthogonal projection onto $\mathbf{F}(\mathscr{E})$ and inversion from $\mathbf{F}(\mathscr{E})$ into $\mathscr{E}$.

Examples of non-invertible unitary operators will be encountered in part II of this paper. A particular case is truncated Fourier transform when it is performed, as in pseudo-spectral models, with higher resolution in physical space than in spectral space. The result stated above shows that the adjoint of Fourier transform going from an N-dimensional spectral space to an M-point spatial grid in physical space (with $M > N$) is the operation which, starting from a field defined at the M spatial grid points, leads to its first N Fourier components. It is therefore legitimate to say that the adjoint of a Fourier transform is always the inverse transform.

Appendix B

The adjoint of an unaliased discretized operator

$\mathscr{E}$ will denote the space spanned by all the spherical harmonics Y_n^m ($n = 1, 2, \ldots, -n \leq m \leq n$) and $\mathscr{E}_N$ the subspace of $\mathscr{E}$ spanned by harmonics with degree $n \leq N$. Inner products are defined on $\mathscr{E}$ and $\mathscr{E}_N$ by Eqs. (19) and (24) respectively, which can be rewritten as

$$\langle \zeta, \zeta' \rangle = \sum^{\infty} \gamma_n \zeta_n^m (\zeta_n'^m)^*$$

for any ζ and ζ' belonging to $\mathscr{E}$ and

$$\langle \xi, \xi' \rangle_N = \sum^{N} \gamma_n \xi_n^m (\xi_n'^m)^*$$

for any ξ and ξ' belonging to $\mathscr{E}_N$. In these expressions, $\gamma_n = a^2/(n(n+1))$ for any degree

n, and Σ^{∞} denotes summation over all spherical harmonics, while Σ^{N} denotes as in Eq. (24) summation over harmonics with degree $1 \leqslant n \leqslant N$.

Let $\mathbf{L}$ be a continuous linear operator from $\mathscr{E}$ into itself. Saying that a linear operator $\mathbf{L}_N$ from $\mathscr{E}_N$ into itself is an unaliased discretized analogue of $\mathbf{L}$ means that, for any ξ belonging to $\mathscr{E}_N$, $\mathbf{L}_N\xi$ is equal to the result obtained by successively performing the following operations:

(a) extend the expansion of ξ by defining $\xi_n^m = 0$ for $n > N$;

(b) apply the operator $\mathbf{L}$ to the result of (a);

(c) truncate the result of (b) at degree N.

We will denote by $\mathbf{Q}_N$ the operator defined by (a), which is an operator from $\mathscr{E}_N$ into $\mathscr{E}$, and by $\mathbf{P}_N$ the operator defined by (c), which is an operator from $\mathscr{E}$ into $\mathscr{E}_N$. The condition that $\mathbf{L}_N$ be unaliased then reads

$$\mathbf{L}_N = \mathbf{P}_N\mathbf{L}\mathbf{Q}_N \tag{B.1}$$

or, by taking adjoints

$$\mathbf{L}_N^* = \mathbf{Q}_N^*\mathbf{L}^*\mathbf{P}_N^*. \tag{B.2}$$

Now, for any ξ belonging to $\mathscr{E}_N$ and any ζ belonging to $\mathscr{E}$

$$\langle \mathbf{Q}_N\xi, \zeta\rangle = \sum^{\infty} \gamma_n(\mathbf{Q}_N\xi)_n^m(\zeta_n^m)^* = \sum^{N} \gamma_n\xi_n^m(\zeta_n^m)^* = \langle \xi, \mathbf{P}_N\zeta\rangle_N$$

which shows that the operators $\mathbf{Q}_N$ and $\mathbf{P}_N$ are adjoint of each other. Eq. (B.2) accordingly becomes

$$\mathbf{L}_N^* = \mathbf{P}_N\mathbf{L}^*\mathbf{Q}_N$$

which is of the same form as (B.1), and expresses that $\mathbf{L}_N^*$ is obtained from $\mathbf{L}^*$ by unaliased discretization. The adjoint of the unaliased discretized analogue of $\mathbf{L}$ is therefore the unaliased discretized analogue of $\mathbf{L}^*$.

APPENDIX C

The adjoint of the leapfrog time differencing scheme

Starting from the initial condition $\mathbf{u} = \mathbf{x}_0$ at time t_0, numerical integration of Eq. (5) with the leapfrog time differencing scheme, initialized by a forward Euler step, produces the following sequence of estimates $\mathbf{x}_p$ for $\mathbf{x}(t)$ at times $t_p = t_0 + p\Delta t \quad (p = 1, \ldots, M)$.

$$\mathbf{x}_1 = \mathbf{x}_0 + \Delta t\mathbf{F}(\mathbf{x}_0) \qquad \text{(forward Euler step)} \tag{C.1a}$$

$$\mathbf{x}_{p+1} = \mathbf{x}_{p-1} + 2\Delta t\mathbf{F}(\mathbf{x}_p) \qquad p \geqslant 1. \tag{C.1b}$$

The discretized analogue of the distance-function (6) being defined as

$$\mathcal{J} = \sum_{p=0}^{M} H_p(\mathbf{x}_p)$$

where H_p is, for each p, a regular scalar function defined on $\mathscr{E}$, then the first-order variation of $\mathcal{J}$ resulting from a perturbation $\delta\mathbf{x}_0$ of $\mathbf{x}_0$ is equal to

$$\delta\mathcal{J} = \sum_{p=0}^{M} \langle \nabla_{\mathbf{x}}H_p, \delta\mathbf{x}_p\rangle. \tag{C.2}$$

In this expression, $\delta \mathbf{x}_p$ is obtained from $\delta \mathbf{x}_0$ by differentiation of (C.1)

$$\delta \mathbf{x}_1 = (\mathbf{I} + \Delta t \mathbf{F}_0') \delta \mathbf{x}_0 \tag{C.3a}$$

$$\delta \mathbf{x}_{p+1} = \delta \mathbf{x}_{p-1} + 2\Delta t \mathbf{F}_p' \delta \mathbf{x}_p \qquad p \geqslant 1 \tag{C.3b}$$

where $\mathbf{I}$ is the unit operator of $\mathscr{E}$ and $\mathbf{F}_p'$ is the derivative of $\mathbf{F}$ with respect to $\mathbf{x}$, taken at point $\mathbf{x}_p$. After introduction, for $p \geqslant 1$, of the vector in $\mathscr{E}^2$

$$\delta \mathbf{y}_p = \begin{pmatrix} \delta \mathbf{x}_{p-1} \\ \delta \mathbf{x}_p \end{pmatrix}$$

Eqs. (C.3) become

$$\delta \mathbf{y}_1 = \mathbf{T}_0 \delta \mathbf{x}_0$$

$$\delta \mathbf{y}_{p+1} = \mathbf{T}_p \delta \mathbf{y}_p \qquad p \geqslant 1$$

where $\mathbf{T}_0$ is the operator from $\mathscr{E}$ into $\mathscr{E}^2$ represented by the block matrix

$$\mathbf{T}_0 = \begin{pmatrix} \mathbf{I} \\ \mathbf{I} + \Delta t \mathbf{F}_0' \end{pmatrix} \tag{C.4a}$$

and $\mathbf{T}_p$ is, for $p \geqslant 1$, the operator from $\mathscr{E}^2$ into itself represented by the matrix

$$\mathbf{T}_p = \begin{pmatrix} 0 & \mathbf{I} \\ \mathbf{I} & 2\Delta t \mathbf{F}_p' \end{pmatrix}. \tag{C.4b}$$

We denote by $\mathbf{P}$ the projection operator from $\mathscr{E}^2$ onto $\mathscr{E}$ which associates to any vector

$$\begin{pmatrix} \delta \mathbf{x} \\ \delta \mathbf{x}' \end{pmatrix}$$

of $\mathscr{E}^2$ its 'second component' $\delta \mathbf{x}'$

$$\mathbf{P} \begin{pmatrix} \delta \mathbf{x} \\ \delta \mathbf{x}' \end{pmatrix} = \delta \mathbf{x}'$$

or, in matrix form

$$\mathbf{P} = (0 \quad \mathbf{I}). \tag{C.4c}$$

With this notation $\delta \mathbf{x}_p$ ($p \geqslant 1$) is equal to

$$\delta \mathbf{x}_p = \mathbf{P} \mathbf{T}_{p-1} \ldots \mathbf{T}_0 \delta \mathbf{x}_0.$$

Carrying this expression into (C.2) and taking adjoints leads to

$$\delta \mathscr{J} = \sum_{p=1}^{M} \langle \mathbf{T}_0^* \ldots \mathbf{T}_{p-1}^* \mathbf{P}^* \nabla_{\mathbf{x}} H_p, \delta \mathbf{x}_0 \rangle + \langle \nabla_{\mathbf{x}} H_0, \delta \mathbf{x}_0 \rangle$$

which is the discretized analogue of (10) and shows that the gradient of $\mathscr{J}$ with respect to $\mathbf{x}_0$ is equal to

$$\nabla_{\mathbf{x}_0} \mathscr{J} = \sum_{p=1}^{M} \mathbf{T}_0^* \ldots \mathbf{T}_{p-1}^* \mathbf{P}^* \nabla_{\mathbf{x}} H_p + \nabla_{\mathbf{x}} H_0. \tag{C.5}$$

The adjoints $\mathbf{T}_p^*$ ($p \geqslant 1$) and $\mathbf{P}^*$ are represented by the transposes of the corresponding

matrices (C.4), i.e.

$$\mathbf{T}_0^* = (\mathbf{I} \quad \mathbf{I} + \Delta t \mathbf{F}_0'^*)$$

$$\mathbf{T}_p^* = \begin{pmatrix} 0 & \mathbf{I} \\ \mathbf{I} & 2\Delta t \mathbf{F}_p'^* \end{pmatrix}$$

$$\mathbf{P}^* = \begin{pmatrix} 0 \\ \mathbf{I} \end{pmatrix}.$$

It is seen that, except for the change of $\mathbf{F}_p'$ into $\mathbf{F}_p'^*$, $\mathbf{T}_p^*$ is identical with $\mathbf{T}_p$. $\mathbf{T}_p^*$ therefore also represents a leapfrog integration step, performed with $\mathbf{F}_p'^*$. More generally it is seen that the successive multiplications of $\nabla_{\mathbf{x}} H_p$ by $\mathbf{P}^*, \mathbf{T}_{p-1}^*, \ldots, \mathbf{T}_0^*$ are equivalent to the following operations:

(1) define $\delta' \mathbf{x}_{p+1} = 0$ and $\delta' \mathbf{x}_p = \nabla_{\mathbf{x}} H_p$;

(2) compute

$$\delta' \mathbf{x}_k = \delta' \mathbf{x}_{k+2} + 2\Delta t \mathbf{F}_k'^* \delta' \mathbf{x}_{k+1} \tag{C.6a}$$

successively for $k = p - 1, \ldots, 1$;

(3) compute

$$\delta' \mathbf{x}_0 = \delta' \mathbf{x}_2 + \delta' \mathbf{x}_1 + \Delta t \mathbf{F}_0'^* \delta' \mathbf{x}_1. \tag{C.6b}$$

Equation (C.6a) is the adjoint of the leapfrog step (C.3b) and is a backward leapfrog step performed on the adjoint equation (12). (C.6b), which is the adjoint of the forward Euler step represented by $\mathbf{T}_0$, is distinctly different from either a forward or a backward Euler step.

Now, the adjoint integration being linear, it is sufficient, in order to compute the sum (C.5) to perform only *one* backward integration, started from $\delta' \mathbf{x}_{M+1} = 0$ and $\delta' \mathbf{x}_M = \nabla_{\mathbf{x}} H_M$, and in the course of which the quantity $\nabla_{\mathbf{x}} H_p$ is added, for all $p < M$, to the currently computed $\delta' \mathbf{x}_p$.

Note that, in (C.6a), the operator $\mathbf{F}_k'^*$, i.e. the adjoint operator at time $k\Delta t$, is applied to $\delta' \mathbf{x}_{k+1}$, i.e. the adjoint variable at time $(k + 1)\Delta t$. In this respect (C.6a) is different from (C.3b). This should not be surprising. In the same way $\delta \mathbf{x}_{p+1}$ is obtained in the direct integration by applying $\mathbf{F}_p'$ to $\delta \mathbf{x}_p$, $\delta' \mathbf{x}_k$ is obtained in the adjoint integration by applying $\mathbf{F}_k'^*$ to $\delta' \mathbf{x}_{k+1}$.

REFERENCES

Bloom, S. C. 1983 The use of dynamical constraints in the analysis of mesoscale rawinsonde data. *Tellus*, **35A**, 363–378

Cacuci, D. G. 1981 Sensitivity theory for nonlinear systems. I: Nonlinear functional analysis approach. *J. Math. Phys.*, **22**, 2794–2802

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

Derber, J. C. 1985 'The variational four-dimensional assimilation of analyses using filtered models as constraints'. Ph.D. dissertation, University of Wisconsin, Madison, Wisconsin, U.S.A.

Eliasen, E., Machenhauer, B. and Rasmussen, E. 1970 'On a numerical method for integration of the hydrodynamical equations with a spectral representation of the horizontal fields'. Report No 2, Institute for Theoretical Meteorology, Copenhagen University, Denmark

Ghil, M., Cohn, S., Tavantzis, J., Bube, K. and Isaacson, E. 1981 'Application of estimation theory to numerical weather prediction'. Pp. 139–224 in *Dynamic meteorology. Data assimilation methods* (Editors L. Bengtsson, M. Ghil and E. Källén). Springer-Verlag, New York

Ghil, M., Cohn, S. E. and Dalcher, A. 1982 Sequential estimation, data assimilation and initialization. Pp. 83–97 in *The interaction between objective analysis and initialization* (Editor D. Williamson). Proceedings of the Fourteenth Stanstead Seminar, McGill University, Montreal, Canada

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

Hall, M. C. G., Cacuci, D. G. and Schlesinger, M. E. 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. Weather Rev.*, **114,** 388–397

Hollingsworth, A., Lorenc, A. C., Tracton, M. S., Arpe, K., Cats, G., Uppala, S. and Kållberg, P. 1985 The response of numerical weather prediction systems to FGGE level IIb data. Part I: Analyses *Q. J. R. Meteorol. Soc.*, **111,** 1–66

Jazwinski, A. H. 1970 *Stochastic processes and filtering theory.* Academic Press, New York, U.S.A.

Lacarra, J. F. and Talagrand, O. 1988 Short range evolution of small perturbations in a barotropic model. *Tellus*, (to appear)

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

Lewis, J. M. and Bloom, S. C. 1978 Incorporation of time continuity into subsynoptic analysis by using dynamical constraints. *ibid.*, **30,** 496–516

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

Lions, J. L. 1971 *Optimal control of systems governed by partial differential equations* (English translation). Springer-Verlag, Berlin

Lorenc, A. C. 1981 A global three-dimensional multivariate statistical interpolation scheme. *Mon. Weather Rev.*, **109,** 701–721

1986 Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112,** 1177–1194

McPherson, R. D., Bergman, K. H., Kistler, R. E., Rasch, G. E. and Gordon, D. S. 1979 The NMC operational global data assimilation scheme. *Mon. Weather Rev.*, **107,** 1445–1461

Marchuk, G. I. 1974 *Numerical solution of the problems of the dynamics of the atmosphere and the ocean* (In Russian). Gidrometeoizdat, Leningrad, U.S.S.R.

Morel, P., Lefevre, G. and Rabreau, G. 1971 On initialization and non-synoptic data assimilation. *Tellus*, **23,** 197–206

Penenko, V. V. and Obraztsov, N. N. 1976 A variational initialization method for the fields of the meteorological elements (English translation). *Soviet Meteorol. Hydrol.*, **1976**(11), 1–11

Petersen, D. P. 1968 On the concept and implementation of sequential analysis for linear random fields. *Tellus*, **20,** 673–686

Reed, M. and Simon, B. 1980 *Methods of modern mathematical physics. I: Functional analysis.* Academic Press, New York, U.S.A.

Sadokov, V. P. and Shteinbok, D. B. 1977 Application of conjugate functions in analysis and forecasting of the temperature anomaly (English translation). *Soviet Meteorol. Hydrol.*, **1977**(10), 16–21

Sasaki, Y. 1970 Some basic formalisms in numerical variational analysis. *Mon. Weather Rev.*, **98,** 875–883

Talagrand, O. 1981a A study of the dynamics of four-dimensional data assimilation. *Tellus*, **33,** 43–60

1981b On the mathematics of data assimilation. *Tellus*, **33,** 321–339

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

Thompson, P. D. 1969 Reduction of analysis error through constraints of dynamical consistency. *J. Appl. Meteorol.*, **8,** 738–742