Tellus (1986), *38A*, 97–110

Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects

By FRANÇOIS-XAVIER LE DIMET, *Université de Clermont II, Aubière, France* and OLIVIER TALAGRAND, *Laboratoire de Météorologie Dynamique, École Normale Supérieure, 24, rue Lhomond, F-75231 Paris Cedex 05, France*

(Manuscript received January 24; in final form July 8, 1985)

ABSTRACT

Two general algorithms for solving constrained minimization problems are presented and discussed in the context of analysis and assimilation of meteorological observations. In both algorithms, the original constrained problem is transformed by appropriate modifications into one unconstrained problem, or into a sequence of unconstrained problems. The main advantage of proceeding in this way is that the new unconstrained problems can be solved by classical descent algorithms, thus avoiding the need of directly solving the Euler–Lagrange equations of the original constrained problem.

The first algorithm presented in the *augmented lagrangian* algorithm. It generalizes the more classical penalty and duality algorithms. The second algorithm, inspired from optimal control techniques, is based on an appropriate use of an *adjoint* dynamical equation, and seems to be particularly well adapted to the assimilation of observations distributed in time. Simple numerical examples show the ability of these algorithms to solve non-linear minimization problems of the type encountered in meteorology. Their possible use in more complex situations is discussed, in particular in terms of their computational cost.

1. Introduction

The homogeneous definition of meteorological fields over some spatial or temporal domain is often a prerequisite for various works of practical or theoretical interest. An obvious example is numerical weather prediction, for which complete initial conditions must be defined on a regular spatial grid before the prediction itself can be performed. Other examples are diagnostic studies of the atmospheric circulation at various spatial and temporal scales. For instance, the fields produced on regular grids from the observations collected during the First GARP Global Experiment have already been used for a fairly large number of climatological studies.

The procedure which takes observed data and creates homogeneous fields is usually called *analysis*, or *assimilation* when the data are distributed in time and the procedure uses an explicit dynamical model for the time evolution of the atmospheric flow. The fields produced by an analysis or an assimilation must satisfy two basic requirements. On the one hand, they must be close to the observations, to an accuracy which ideally is the accuracy of the observations themselves, at the required spatial and temporal scales. On the other hand, they must verify dynamical and/or statistical relationships which are known to be satisfied by the real atmospheric fields. For instance, it is reasonable to impose that the mass and velocity fields produced by an analysis or an assimilation are in approximate geostrophic balance in mid-latitudes, or that the kinetic energy spectrum of the wind field is similar to the real spectrum. In the specific case of an assimilation, which produces a temporal sequence of meteorological fields, a sensible requirement is that the corresponding temporal evolution is consistent with the evolution equations of the flow. In any case, some information about the dynamical and/or statistical properties of the fields should be introduced into the

analysis of assimilation algorithm. Indeed, most of the research done on analysis and assimilation of meteorological observations has in effect been concerned with the question of how to introduce the appropriate information into the numerical procedures.

Three main classes of algorithms have been used in analysis and assimilation. The first class comprises simple algorithms of spatial interpolation (Cressman, 1959; Barnes, 1973). In these methods, information on the meteorological fields is introduced under the very indirect and simplified form of an "influence function" which characterizes the spatial scale of variation of the fields. These interpolation algorithms, which have the great advantage of being both very simple and very economical to implement, have been widely used for a number of years, and are still often used for various purposes.

The algorithms of the second class are known in meteorology under the collective name of "optimal interpolation", and are in effect linear regression algorithms. The fields they produce are the linear combinations of the observations which minimize the mean quadratic error over a large number of realizations of the analysis. The information to be introduced in the algorithm is therefore statistical, and consists of the first and second order moments of the observed and unknown variables. Optimal interpolation techniques, which have been introduced in meteorology mainly by Gandin (1965), are now widely used, especially as parts of assimilation procedures for large-scale prediction models, see e.g. Lorenc (1981). A major advantage of optimal interpolation is that it provides a practical and internally consistent approach for treating a large set of heterogeneous observations, and it is at present the technique which produces the best results for operational forecasting. However, its extension to the time dimension, and its use in conjunction with an explicit dynamical model for temporal evolution of the atmospheric flow, raise several problems. When the model is linear, optimal interpolation naturally extends to the assimilation of observations distributed in time under the form of *Kalman-Bucy filtering* see e.g. Ghil *et al.*, (1981). For non-linear models, the extension of optimal interpolation to the time dimension is a difficult problem which is being studied in several places (Ghil, Lacarra, personal communications). The procedures now in operational use are already in effect a heuristic generalization of optimal interpolation to the time dimension. But the successive analyses are still performed almost independently of the evolution equations, and the question remains open of how much gain could be achieved by making the assimilation process more consistent with these equations. Another possible defect of optimal interpolation is that, being a statistical method, it tends in some cases to excessively smooth the analysed fields: for example, the amplitudes of rapidly deepening depressions are frequently reduced by optimal interpolation. Excessive smoothing may be particularly troublesome for mesoscale forecasting models, since it may inhibit developments which are unlikely from a statistical point of view, but are very important to predict correctly, precisely because of the rarity of their occurrences.

The algorithms of the third class are variational. They produce fields which minimize a given measure of the "distance" to the observations, while at the same time satisfying an explicit dynamical constraint. The latter will normally be expressed by one or more differential equations. These methods have the great theoretical advantage of providing exact consistency between the analysis and the dynamics, as expressed by the constraint. In that respect, they appear clearly superior to the other methods. On the other hand, their higher mathematical technicality and their probable high computational cost have strongly limited their use. Although the potential usefulness of variational methods for meteorological problems was pointed out very early by Sasaki (1958, 1970), and in spite of a fairly large number of various studies, these methods have not been fully utilized. This is especially true of assimilation studies which address problems containing explicit time derivatives, see e.g. Lewis and Bloom (1978). Variational techniques have thus been applied primarily to analysis, usually of mesoscale fields (see Soliz and Fein, 1980; or Sasaki and Goerss, 1982). Other applications have been concerned with the problem of initialization, i.e. the problem of suppressing unrealistic high frequency gravity oscillations from the initial conditions of a numerical forecast (Daley, 1978; Tribbia, 1982; Temperton, 1982). This also is basically a static problem, from which time derivatives are excluded. But all these applications are still rather limited, and a precise assessment of the possibilities offered by variational methods and

of the quality of the results they can produce is still to be made.

It should be noted that, although these three classes of algorithms were conceived and developed independently, they do have mutual relationships. The interpolation method of Cressman can be described as a simple case of optimal interpolation. Optimal interpolation itself has been shown by Kimeldorf and Wahba (1970) to produce fields which are the solution of a variational problem, in which the function to be minimized is the sum of two terms representing, one the distance to the observations, and the other some measure of the smoothness of the fields (see also Wahba, 1982 and Testud and Chong, 1983). There are therefore more links between these various techniques than could be *a priori* thought, and the study of these links is certainly of great potential interest.

The present article is devoted to the application of *optimization theory* to analysis and assimilation. Optimization theory, which has been greatly developed in the last two decades, has led to the definition of a number of variational techniques which can be applied to a fairly large variety of problems. After a general presentation of variational formalism in Section 2, some of these techniques are presented in Sections 3 and 4, together with numerical results obtained in simple cases of analysis and assimilation with non-linear constraints. Conclusions and comments follow in Section 5.

A number of mathematical notions, such as gradients on functional spaces, convexity, etc., are used in the course of this article. Precise definitions of these notions can be found in most textbooks on optimization theory, for instance in Bertsekas (1982).

2. Variational formalism

Let us first consider the linear advection equation

$$\frac{\partial \phi}{\partial t} + c(x)\frac{\partial \phi}{\partial x} = 0, \tag{2.1}$$

where ϕ is an unknown scalar function, t is time, x a spatial coordinate, and $c(x)$ a known function of x. Given "observations" $\hat{\phi}(x,t)$ over some domain Σ in the (x,t) plane, a simple variational problem is to find a solution of (2.1) which minimizes the quantity

$$J(\phi) = \int_{\Sigma} (\phi - \hat{\phi})^2 \, dx \, dt. \tag{2.2}$$

This is the problem on which Sasaki (1970) illustrated and tested several variational algorithms. The exact solution to that problem is not very difficult to find, and the minimizing ϕ can be shown to be equal, along each characteristic of (2.1), to the average of $\hat{\phi}$ with respect to x. Eqs. (2.1) and (2.2) nevertheless constitute a conceptually and algebraically simple problem, in which some of the algorithms to be presented will be illustrated.

A more general situation appropriate to the use of variational techniques can be described. Assume a computer program is available for solving numerically a given set of differential equations written symbolically as

$$F(U) = 0, \tag{2.3}$$

where U denotes the various meteorological fields (geopotential, wind, . . .) under consideration. These fields must verify (2.3) over some spatial and/or temporal domain Σ, with boundary Γ. Eq. (2.3) may represent a set of dynamic equations for the temporal evolution of the atmospheric flow, or some time independent relationship between the meteorological fields, such as for instance the classical balance equation. Suppose some estimate $\hat{U}$ of the fields U over Σ is also available. $\hat{U}$ may consist of values defined homogeneously over Σ, already obtained through some preliminary analysis or assimilation process, or of local observations at an irregular set of points over Σ. In either case, one is interested in obtaining a solution of (2.3) which is "close" in some sense to $\hat{U}$. This leads to the following typical problem (P). Find the solution of (2.3) which minimizes the scalar function

$$J(U) = \int_{\Sigma} \|U - \hat{U}\|^2 \, d\Sigma \tag{2.4}$$

where $\|\ \|$ denotes a suitable norm, and where the integral must be replaced by an appropriate finite sum if $\hat{U}$ consists of discrete observations. One may also wish that the solution U be smooth in some sense, in which case a term of the form

$$\int_{\Sigma} \|DU\|^2 \, d\Sigma$$

can be added to (2.4), where D denotes an appropriate differential operator.

Eq. (2.3) is the *constraint* of problem (P), while the function J of (2.4) is its *cost-function*. A systematic approach for solving problems of this type is based on a generalization of the classical Lagrange multiplier technique used in finite-dimensional spaces. Let $\{\ ,\ \}$ be an inner product defined on the functional space to which $F(U)$ belongs, and Λ and element of that space. The *Lagrangian* of problem (P) is the scalar function of U and Λ defined by

$$\mathscr{L}(U,\Lambda) = J(U) + \{\Lambda, F(U)\}. \tag{2.5}$$

Under simple regularity conditions (differentiability to a sufficient order) it can be shown (see, e.g. Bertsekas (1982)) that the problem of determining the stationary points of $J(U)$ under the constraint $F(U) = 0$ is equivalent to the problem of determining the stationary points of (2.5) with respect to the two variables U and Λ. This result is at the basis of the *strong constraint formalism* of Sasaki (1970).

The constrained problem (P) is thus replaced by an unconstrained problem with respect to the two variables U and Λ. Λ is called the *Lagrange multiplier* of problem (P). The equations which express that the Lagrangian is stationary at a point (U,Λ) are called the *Euler–Lagrange equations* of problem (P).

For problem (2.1)–(2.2), the Lagrange multiplier Λ must be a regular scalar function defined on the domain Σ. With the inner product $\{\ ,\ \}$ defined as

$$\{\psi,\chi\} = \int_\Sigma \psi\chi \, \mathrm{d}x \, \mathrm{d}t, \tag{2.6}$$

the Lagrangian is

$$\mathscr{L}(\phi,\Lambda) = \int_\Sigma \left[(\phi - \hat{\phi})^2 + \Lambda\left(\frac{\partial\phi}{\partial t} + c\frac{\partial\phi}{\partial x}\right)\right] \mathrm{d}x \, \mathrm{d}t.$$

The first-order variation $\delta\mathscr{L}$ resulting from perturbations $\delta\phi$ and $\delta\Lambda$ in ϕ and Λ respectively is equal to

$$\delta\mathscr{L} = \int_\Sigma \left[2(\phi - \hat{\phi})\delta\phi + \Lambda\left(\frac{\partial\delta\phi}{\partial t} + c\frac{\partial\delta\phi}{\partial x}\right)\right] \mathrm{d}x \, \mathrm{d}t$$

$$+ \int_\Sigma \left(\frac{\partial\phi}{\partial t} + c\frac{\partial\phi}{\partial x}\right) \delta\Lambda \, \mathrm{d}x \, \mathrm{d}t.$$

Integration by parts of the terms containing derivatives of $\delta\phi$ leads to

$$\delta\mathscr{L} = \int_\Sigma \left[2(\phi - \hat{\phi}) - \frac{\partial\Lambda}{\partial t} - \frac{\partial(\Lambda c)}{\partial x}\right] \delta\phi \, \mathrm{d}x \, \mathrm{d}t$$

$$+ \int_\Sigma \left(\frac{\partial\phi}{\partial t} + c\frac{\partial\phi}{\partial x}\right) \delta\Lambda \, \mathrm{d}x \, \mathrm{d}t$$

$$+ \int_\Gamma \Lambda \, \boldsymbol{v}\cdot\boldsymbol{n} \, \delta\phi \, \mathrm{d}\Gamma, \tag{2.7}$$

where $\boldsymbol{n}$ is the unit vector normal to the boundary Γ of Σ, directed outwards, and $\boldsymbol{v}$ is the vector with components c and 1 in the x and t directions respectively.

For given ϕ and Λ, $\mathscr{L}$ will be stationary if and only if $\delta\mathscr{L}$ is zero for any perturbations $\delta\phi$ and $\delta\Lambda$. This will be the case if and only if the three following conditions are simultaneously verified.

$$2(\phi - \hat{\phi}) - \frac{\partial\Lambda}{\partial t} - \frac{\partial(\Lambda c)}{\partial x} = 0 \qquad \text{on } \Sigma, \tag{2.8a}$$

$$\frac{\partial\phi}{\partial t} + c\frac{\partial\phi}{\partial x} = 0 \qquad \text{on } \Sigma, \tag{2.8b}$$

$$\Lambda = 0, \qquad \text{along } \Gamma \tag{2.8c}$$

Eqs. (2.8a, b), together with the boundary condition (2.8c), are the Euler–Lagrange equations of problem (2.1)–(2.2). Elimination of ϕ through (2.8a) leads for Λ to a second-order parabolic linear partial differential equation with boundary condition (2.8c).

Any constrained problem of type (P) can in principle be solved through its Euler–Lagrange equations. However, except in particular cases, techniques do not exist for directly solving these equations, and the interest of the Euler–Lagrange equations is often more theoretical than practical.

In the algorithms to be presented below, the original problem (P) will be transformed, by changing either the cost-function J, or the variable U, or both, into a sequence of unconstrained minimization problems or even in some cases into one single unconstrained problem. Let W and K be the new variable and cost-function of an unconstrained problem, respectively thus associated with

problem (P). Writing that K is stationary at some point W^* will lead to partial differential equations which are basically of the same nature as the Euler–Lagrange equations. Their direct solution may or may not be easy, depending on the particular problem at hand. If an inner product has been defined on the space to which the unconstrained variable W belongs, it is then possible to define the *gradient* ∇K of the cost-function. This gradient belongs to the same space as W, and is such that the first-order variation δK resulting from a perturbation δW of W is equal to the inner product

$$\delta K = \langle \nabla K, \delta W \rangle. \tag{2.9}$$

If this gradient can be explicitly determined, it is then possible to implement a *descent* algorithm, in which the minimizing W^* is approximated by successive values obtained by varying W along its gradient (or along a direction obtained from a combination of the gradients from each successive iteration). Two classical descent algorithms are the *steepest descent* algorithm and the *conjugate gradient* algorithm (Gill et al., 1982).

In practice, two situations may occur. Either the problem is such that an analytical expression is readily available for ∇K, and the corresponding computations can be performed without difficulty (Section 3). Or the cost-function is a complicated compound function of its argument, in which case *optimal control* techniques will normally lead to an explicit algorithm for determining ∇K (Section 4). The basic advantage of these approaches is that they are very general and can be applied to a very large variety of problems of type (P), once a program is available for solving explicitly the constraint (2.3), without having to resort to a detailed analysis of the properties of the corresponding Euler–Lagrange equations.

An important remark can be made about eqs. (2.8). The condition of stationarity has led for the unknowns ϕ and Λ not only to eqs. (2.8a, b), which must be verified on the interior of Σ, but also to the corresponding boundary condition (2.8c). This situation is by no means particular to problem (2.1)–(2.2). Stationarity conditions will generally lead not only to partial differential equations to be verified on the interior of the domain under consideration, but also to the associated boundary condition.

3. Penalty, duality and augmented lagrangian algorithms

In this section, we will present three algorithms in which the original constrained problem (P) is transformed into a sequence of unconstrained problems. These algorithms are:

— the penalty algorithm
— the duality algorithm
— the augmented Lagrangian algorithm.

The third algorithm is a combination of the other two, which it contains as special cases, and over which it has the advantage of being computationally much more efficient.

3.1. Penalty algorithm

Let us consider problem (P) and, for a given positive real number ε, the scalar function defined by

$$J_\varepsilon(U) = J(U) + \frac{1}{\varepsilon} |F(U)|^2, \tag{3.1}$$

where $|F(U)|$ is a norm for $F(U)$. Therefore $|F(U)| = 0$ if and only if constraint (2.3) is verified. $J_\varepsilon(U)$ is called a *penalized cost-function*, and the number ε is referred to as the *penalty parameter*. $J_\varepsilon(U)$ is non-negative and reaches a minimum for some value U_ε^* of its argument. If $F(U) = 0$, $J_\varepsilon(U)$ is equal to $J(U)$. Therefore, when ε tends to 0, $F(U_\varepsilon^*)$ must also tend to 0 in order to prevent $J_\varepsilon(U_\varepsilon^*)$ from increasing to infinity. It can be shown (see e.g. Ciarlet (1982)) that under simple regularity hypotheses, U_ε^* will tend to the solution U^* of the original problem (P) when ε tends to 0. Given a sequence ε_n of positive numbers tending to 0, the original problem (P) is thus replaced by a sequence of unconstrained problems.

In the case of problem (2.1)–(2.2), where the unknown is ϕ, the norm in (3.1) can be defined as

$$|F(\phi)|^2 = \int_\Sigma \left(\frac{\partial \phi}{\partial t} + c \frac{\partial \phi}{\partial x} \right)^2 \mathrm{d}x\, \mathrm{d}t,$$

and the penalized cost-function (3.1) accordingly becomes

$$J_\varepsilon(\phi) = \int_\Sigma \left[(\phi - \hat{\phi})^2 + \frac{1}{\varepsilon} \left(\frac{\partial \phi}{\partial t} + c \frac{\partial \phi}{\partial x} \right)^2 \right] \mathrm{d}x\, \mathrm{d}t.$$

A development similar to the one which led to expression (2.7) leads here to the following expression for the first order variation δJ_ε resulting from a

perturbation $\delta\phi$ of ϕ

$$\delta J_\varepsilon = 2\int_\Sigma \left[\phi - \hat{\phi} - \frac{1}{\varepsilon}\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}c\right)\left(\frac{\partial \phi}{\partial t} + c\frac{\partial \phi}{\partial x}\right)\right]\partial\phi \, dx\, dt$$

$$+\frac{2}{\varepsilon}\int_\Gamma \left(\frac{\partial \phi}{\partial t} + c\frac{\partial \phi}{\partial x}\right)\boldsymbol{v}\cdot\boldsymbol{n}\,\delta\phi \, d\Gamma,$$

where $\boldsymbol{v}$ and $\boldsymbol{n}$ have the same meaning as in (2.7). Using inner product (2.6), δJ_ε can be written as

$$\delta J_\varepsilon = \{\nabla J_\varepsilon, \delta\phi\},$$

where the gradient ∇J_ε is the sum of two "components", namely

$$2\left[\phi - \hat{\phi} - \frac{1}{\varepsilon}\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial x}c\right)\times\left(\frac{\partial \phi}{\partial t} + c\frac{\partial \phi}{\partial x}\right)\right] \quad \text{on the interior of } \Sigma, \tag{3.2a}$$

$$\frac{2}{\varepsilon}\left(\frac{\partial \phi}{\partial t} + c\frac{\partial \phi}{\partial x}\right)\boldsymbol{v}\cdot\boldsymbol{n} \quad \text{along } \Gamma. \tag{3.2b}$$

This gradient can be used in a descent process in order to determine the function ϕ_ε^* which minimizes J_ε.

The presence of the boundary component (3.2b) may seem surprising. In numerical applications the constraint (2.3) will be discretized and this additional component will naturally result from the special form necessarily taken by the discretization along the boundary.

In the practice of the penalty algorithm, the iterations on the parameter ε will have to be stopped at some finite value of that parameter, so that the final computed solution will be the solution of the unconstrained problem (3.1) for some finite ε. This is exactly what Sasaki (1970) called *weak constraint formalism*. The advantage of explicitly varying ε, rather than using only one particular value, is that the corresponding variations of U_ε^* contain information as to the accuracy with which the exact solution U^* is approximated by the algorithm.

3.2. *Duality algorithm*

The duality algorithm, introduced by Arrow et al. (1958), is an iterative algorithm which determines stationary points of the lagrangian (2.5). One could think of determining the stationary points of the lagrangian through a descent algorithm. This is not possible, however, because these points are not extrema, but saddle points which cannot be reached by a descent process. The duality algorithm performs alternative descent and ascent steps, which makes it appropriate for determining saddle points. It is defined as follows. U_k and Λ_k being known, U_{k+1} is computed as the value of U which minimizes

$$\mathscr{L}(U, \Lambda_k) = J(U) + \{\Lambda_k, F(U)\},$$

Λ_{k+1} is then defined by

$$\Lambda_{k+1} = \Lambda_k + \rho F(U_{k+1})$$

where ρ is a given positive constant. The convergence of this algorithm to a stationary point (U^*, Λ^*) is proven in Bertsekas (1982), provided the Lagrangian $\mathscr{L}(U, \Lambda)$ is strictly convex with respect to its first argument U in the vicinity of (U^*, Λ^*), and that the constant ρ is properly chosen. It is seen that the solution of the original constrained problem (P) is reduced by the duality algorithm to the solution of a sequence of unconstrained problems with respect to U. Each of these problems can be solved by a descent algorithm. This requires the explicit determination of the gradient of $\mathscr{L}(U, \Lambda)$ with respect to U. In the case of problem (2.1)–(2.2) it is seen from (2.7) that this gradient is made up of the two components

$$2(\phi - \hat{\phi}) - \frac{\partial \Lambda}{\partial t} - \frac{\partial(\Lambda c)}{\partial x} \quad \text{on the interior of } \Sigma$$

$$\Lambda\,\boldsymbol{v}\cdot\boldsymbol{n} \quad \text{along } \Gamma$$

3.3. *The augmented lagrangian algorithm*

The penalty and duality algorithms, although they have been successfully applied to a variety of problems, are not always numerically very efficient, and may be difficult to implement in practice. In the case of the penalty algorithm, the penalized cost-function $J_\varepsilon(U)$, for small values of ε, will vary rapidly along directions perpendicular to the subspace $F(U) = 0$, while it will vary slowly along directions parallel to that subspace. That situation will lead to ill-conditioning and inaccuracy in the determination of the minimizing U_ε^*. For the duality algorithm, the convexity of the Lagrangian may be difficult to check in practice. Moreover, the

efficiency of the convergence may strongly depend on the choice of the initial Λ_0, for which a priori physical considerations will in general be of little help.

For given $\varepsilon > 0$, the *augmented lagrangian* of problem (P) is defined by

$$\mathscr{L}_\varepsilon(U, \Lambda) = J(U) + \{\Lambda, F(U)\} + \frac{1}{\varepsilon} |F(U)|^2 \quad (3.3)$$

where the notations are the same as before. The solution U^* of problem (P) can then be obtained as the limit of U_k in a sequence $(U_k, \Lambda_k, \varepsilon_k)$ constructed as follows. $(U_k, \Lambda_k, \varepsilon_k)$ being known, U_{k+1} is defined as the value of U which minimizes $\mathscr{L}_{\varepsilon_k}(U, \Lambda_k)$. Λ_{k+1} and ε_{k+1} are then given by

$$\Lambda_{k+1} = \Lambda_k + \frac{1}{\varepsilon_k} F(U_{k+1}) \quad (3.4a)$$

$$\varepsilon_{k+1} = c_k \varepsilon_k, \quad (3.4b)$$

where c_k is positive and less than 1. The convergence of this algorithm, which is a generalization of both the penalty and the duality algorithms, is proven in Bertsekas (1982) and Fortin and Glowinski (1982). Bertsekas (1982) gives in addition practical indications for the choice of the sequence c_k.

The main advantage of the augmented lagrangian algorithm is that it is computationally much more efficient than either the penalty or the duality algorithm. Ill-conditioning is still present present in the augmented lagrangian for small ε, but the presence of the Lagrange multiplier introduces additional degrees of freedom. As explained in Bertsekas (1982), proper use of these additional degrees of freedom through (3.4) will lead to a better estimate of the minimizing U for the same decrease of ε. It is even possible to obtain convergence without ε tending to 0. The difficulties associated with ill-conditioning are thus greatly reduced. As for the compared qualities of (3.3) and simple Lagrangian (2.5), the presence in (3.3) of the penalty term avoids the problems which may arise in the duality algorithm if the Lagrangian is not convex. It also leaves more freedom for the choice of the initial Λ_0 and reduces the difficulties which may arise from an inappropriate choice.

The augmented lagrangian algorithm has been used by Navon and De Villiers (1983) in order to maintain integral constraints in the time integration of a meteorological numerical model. We have applied the algorithm to the problem of finding the geopotential field ϕ and the velocity field (u,v) which minimize the cost-function.

$$J(u,v,\phi) = \int_{\mathscr{D}} [(u-\hat{u})^2 + (v-\hat{v})^2 + \alpha(\phi-\hat{\phi})^2] \, dx \, dy \quad (3.5)$$

over a two-dimensional domain $\mathscr{D}$, while satisfying the three scalar constraints

$$E_u \equiv u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + \frac{\partial \phi}{\partial x} - fv = 0, \quad (3.6a)$$

$$E_v \equiv u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + \frac{\partial \phi}{\partial y} + fu = 0, \quad (3.6b)$$

$$E_\phi \equiv \frac{\partial}{\partial x}(\phi u) + \frac{\partial}{\partial y}(\phi v) = 0. \quad (3.6c)$$

In (3.5), the constant α is a dimensional coefficient. The constraints (3.6a–c) are steady-state conditions for the shallow-water equations. The domain $\mathscr{D}$ was a 2500 × 2500 km^2 rectangle centered at 45 N and 5 W. The fields $\hat{u}, \hat{v}, \hat{\phi}$ had been produced at the points of a regular 25 × 25 latitude–longitude grid covering $\mathscr{D}$, through an analysis performed on radiosonde observations of the 500 mb wind and geopotential. The problem considered here can therefore be described as a variational initialization, intended at improving the preliminary analysis $(\hat{u}, \hat{v}, \hat{\phi})$ through the use of constraints (3.6).

After discretization, problem (3.5–6) is reduced to the minimization of a scalar function with 1875 scalar arguments linked by 1587 independent scalar constraints. The details of the implementation of the augmented lagrangian algorithm, including the choice of the sequence c_k, were taken from Bertsekas (1982). The descent algorithm used for minimizing $\mathscr{L}_{\varepsilon_k}(U, \Lambda_k)$ with respect to U was an algorithm of the so-called *bundle* type, described in Lemaréchal and Strodiot (1981). Each minimization required an average of 20 steps of the descent algorithm.

Table 1 gives the variations, with the number k of iterations of the algorithm, of the spatially averaged norms of the constraints (3.6a–c). The quantities E_u and E_v, starting from values which, although large, were already meteorologically

Table 1. *Variations of the root-mean-squared quantities E_u, E_v, E_ϕ with the iteration number k of the augmented lagrangian algorithm*

k	1	2	5	10
r.m.s. E_u ($m \cdot s^{-2}$)	0.29×10^{-3}	0.19×10^{-3}	0.42×10^{-5}	0.18×10^{-5}
r.m.s. E_v ($m \cdot s^{-2}$)	0.12×10^{-3}	0.11×10^{-3}	0.24×10^{-5}	0.24×10^{-5}
r.m.s. E_ϕ ($m \cdot s^{-1}$)	0.13×10^{-1}	0.10×10^{-1}	0.61×10^{-3}	0.21×10^{-3}

acceptable, decreased by two orders of magnitude in five to ten iterations. This clearly demonstrates the numerical convergence of the process. The changes imposed by the variational procedure on the geopotential height and the wind components were about 2 m and 1 ms^{-1} respectively. These values are less than the corresponding observational errors and show that the final adjusted fields remained within the accuracy of the original observations.

4. Reduction of the control variable, optimal control

4.1. Principle

In the three algorithms presented in the preceding section, each of the intermediary minimization processes is performed on the complete unknown field U. In the terminology of optimization theory, U is said to be the *control variable* of the algorithms. In many problems of type (P), a solution of the constraint (2.3) will be uniquely defined by the specification of some appropriate condition V along the boundary Γ of Σ, or along part of Γ. Depending on the particular problem under consideration, V will be the value of U itself along Γ, and/or its derivatives. The constrained problem (P) can then be stated in the following equivalent terms: find the boundary condition V such that the corresponding solution of (2.3) minimizes the cost-function J. When expressed in these terms, the problem is unconstrained, since no particular condition is imposed on the boundary condition V. V will now be varied in order to minimize J and will be the new control variable of the process. The change which is thus made is referred to as a *reduction of the control variable.*

A great advantage of reducing the control variable is that the effective dimension of a discretized problem, i.e., the number of parameters which will have to be explicitly varied in the minimization process, is substantially decreased. On the other hand, it is necessary to determine explicitly how J varies as a function of V, i.e., to determine the gradient of J with respect to V. J is a function of V through the constraint (2.3), and standard differentiation techniques will not in general lead to a practically usable form for the gradient of J. One conceivable way to get an estimate of the gradient of J in a discretized case would be to perturb in turn all the components of V and, for each perturbation, to solve explicitly the constraint (2.3) and compute the corresponding variation of J. But it is clear that, even for problems of relatively small size, the amount of computation involved would make that approach totally impractical. The techniques of optimal control (Lions, 1971) make it possible to explicitly compute the gradient of J with respect to V at a much smaller cost. In order to compute the gradient of J for a given V, it is only necessary to solve the constraint equation (2.3) once, and then to solve the *adjoint* of the linearized form of the same equation. This approach can be used to solve a problem such as (2.1)–(2.2), but it seems to be particularly well suited to the problem of assimilation of observations distributed in time, as was discussed by Penenko and Obraztsov (1976). The technique will be demonstrated in this context.

Let us consider a system whose state at any time t is defined by a state vector $Y(t)$ belonging to some space $\mathscr{E}$ and whose time evolution is described by the equation

$$\frac{dY}{dt} = H(Y), \tag{4.1}$$

where H is a regular function of $\mathscr{E}$ into itself.

Assuming that observations $\hat{Y}((t_1), \hat{Y}(t_2), \ldots, \hat{Y}(t_n)$ of the state of the system are available at times $t_1 < t_2 < \ldots < t_n$, we want to find a solution $Y(t)$ of (4.1) minimizing the functional

$$J[Y(t)] = \sum_{i=1}^{n} \langle Y(t_i) - \hat{Y}(t_i), Y(t_i) - \hat{Y}(t_i) \rangle, \quad (4.2)$$

where $\langle \ , \ \rangle$ denotes an inner product defined over $\mathscr{E}$. We do not assume that the observations $\hat{Y}(t_i)$ are exactly compatible with a solution of (4.1), so that the minimum of J will not in general be equal to 0.

The problem just stated is of type (P), with (4.1) and (4.2) as constraint and cost-function respectively. Any solution $Y(t)$ of (4.1) is uniquely defined by the specification of the initial condition $Y(t_1)$, which we will use as control variable. The problem is then: find the initial condition $Y(t_1)$ such that the corresponding solution of (4.1) minimizes the cost-function (4.2).

Given a solution $Y(t)$ of (4.1), the first order variation δJ resulting from a perturbation $\delta Y(t_1)$ of the initial condition is equal to

$$\delta J = 2 \sum_{i=1}^{n} \langle Y(t_i) - \hat{Y}(t_i), \delta Y(t_i) \rangle, \quad (4.3)$$

where the first-order variations $\delta Y(t_i)$ are themselves obtained by integrating the *linearized perturbation equation*

$$\frac{d\delta Y}{dt} = A(t)\delta Y. \quad (4.4)$$

starting from the initial condition $\delta Y(t_1)$. System (4.4) is obtained by linearizing the basic evolution eq. (4.1) about the solution $Y(t)$. $A(t)$ is the *Jacobian operator* made up of the derivatives, taken at point $Y(t)$, of the function H with respect to Y. For any solution of (4.4), the value $\delta Y(t)$ at a given time t depends linearly on the initial condition $\delta Y(t_1)$. The corresponding linear operator is called the *resolvent* of (4.4) between times t_1 and t, and will be denoted $R(t, t_1)$. Therefore for any i

$$\delta Y(t_i) = R(t_i, t_1)\delta Y(t_1),$$

which, after substitution into (4.3), leads to

$$\delta J = 2 \sum_{i=1}^{n} \langle Y(t_i) - \hat{Y}(t_i), R(t_i, t_1)\delta Y(t_1) \rangle. \quad (4.5)$$

For any continuous linear operator L defined on $\mathscr{E}$, it is possible to define its adjoint L^*, which is also a linear operator on $\mathscr{E}$, characterized by the property that, for any two vectors X and Z in $\mathscr{E}$ the following equality holds between inner products (see Courant and Hilbert (1962) pp. 234–237):

$$\langle X, LZ \rangle = \langle L^*X, Z \rangle. \quad (4.6)$$

Introducing the adjoint $R^*(t_i, t_1)$ of $R(t_i, t_1)$, (4.5) becomes

$$\delta J = \langle 2 \sum_{i=1}^{n} R^*(t_i, t_1)[Y(t_i) - \hat{Y}(t_i)], \delta Y(t_1) \rangle,$$

which, by the definition (2.9) of the gradient, shows that the gradient of J with respect to $Y(t_1)$ is

$$\nabla J = 2 \sum_{i=1}^{n} R^*(t_i, t_1)[Y(t_i) - \hat{Y}(t_i)]. \quad (4.7)$$

We introduce at this point the *adjoint equation*

$$\frac{d\delta^* Y}{dt} = -A^*(t)\delta^* Y, \quad (4.8)$$

whose variable $\delta^* Y$ belongs to the space $\mathscr{E}$, and where $A^*(t)$ is the adjoint of $A(t)$. Let $S(t', t)$ be the resolvent of (4.8) between times t and t'. For any two solutions $X(t)$ and $Z(t)$ of the direct and adjoint equations (4.4) and (4.8) respectively, the inner product $\langle X(t), Z(t) \rangle$ is constant with time since

$$\frac{d}{dt}\langle X(t), Z(t) \rangle = \langle \frac{dX}{dt}(t), Z(t) \rangle + \langle X(t), \frac{dZ}{dt}(t) \rangle$$

$$= \langle A(t)X(t), Z(t) \rangle + \langle X(t), -A^*(t)Z(t) \rangle$$

$$= \langle A(t)X(t), Z(t) \rangle - \langle A(t)X(t), Z(t) \rangle = 0.$$

The solution of (4.4) defined by the condition $X(t_1)$ at time t_1 takes at time t_i the value $R(t_i, t_1)X(t_1)$, while the solution of (4.8) defined by the condition $Z(t_i)$ at time t_i takes at time t_1 the value $S(t_1, t_i)Z(t_i)$. Therefore

$$\langle X(t_1), S(t_1, t_i)Z(t_i) \rangle = \langle R(t_i, t_1)X(t_1), Z(t_i) \rangle.$$

This equality being verified for any $X(t_1)$ and $Z(t_i)$, it results that

$$S(t_1, t_i) = R^*(t_i, t_1).$$

$R^*(t_i, t_1)$ is thus the resolvent of the adjoint

equation (4.8) between times t_i and t_1. In eq. (4.7), the term $R^*(t_i,t_1)\,[Y(t_i) - \hat{Y}(t_i)]$ can therefore be obtained by integrating (4.8) *backwards* in time from t_i to t_1, starting from $Y(t_i) - \hat{Y}(t_i)$. Indeed, since eq. (4.8) is linear, the sum on the right-hand side of (4.7) can be obtained by *one* backward integration of (4.8) from t_n to t_1, the term $Y(t_i) - \hat{Y}(t_i)$ being added at time t_i to the current value $\delta^*Y(t_i)$.

In summary, the gradient ∇J corresponding to some initial condition $Y(t_1)$ can be explicitly obtained by performing the following operations: (i) starting from $Y(t_1)$, integrate the basic eq. (4.1) from t_1 *to* t_n. Store the values thus computed for $t_1 \leqslant t \leqslant t_n$, which will make up the coefficients of the operator $A^*(t)$. (ii) starting from $\delta^*Y(t_n) = Y(t_n) - \hat{Y}(t_n)$, integrate the adjoint equation (4.8) backwards in time from t_n to t_1, the *forcing* term $Y(t_i) - \hat{Y}(t_i)$ being added to the currently computed solution $\delta^*Y(t_i)$ at each observation time t_i. The final result at time t_1 is half the gradient ∇J.

4.2. *A simple numerical example*

The applicability of optimal control to data assimilation is being studied in several places (Lewis and Derber, 1985, Courtier, private communication). We will present here numerical results obtained with the following equations

$$\frac{\partial \phi}{\partial t} + \frac{\partial}{\partial x}(\phi u) = 0, \tag{4.9a}$$

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}(\phi + \tfrac{1}{2}u^2) = 0, \tag{4.9b}$$

where the "geopotential" ϕ and the scalar "velocity" u are functions of time t and of a one-dimensional periodic spatial coordinate x. Eqs. (4.9), which can be considered as a one-dimensional version of the shallow-water equations, are the analogue of (4.1), the state vector $Y(t)$ being now made up of the two fields $\phi(x,t)$ and $u(x,t)$ at time t

$$Y(t) = \begin{pmatrix} \phi(x,t) \\ u(x,t) \end{pmatrix}.$$

The inner product of two state vectors

$$Y_1 = \begin{pmatrix} \phi_1 \\ u_1 \end{pmatrix} \quad \text{and} \quad Y_2 = \begin{pmatrix} \phi_1 \\ u_2 \end{pmatrix}$$

is defined as

$$\langle Y_1, Y_2 \rangle = \int (\phi_1\phi_2 + \Phi_0 u_1 u_2)\,dx, \tag{4.10}$$

where Φ_0 is an appropriate constant geopotential.

Given a solution $(\phi(x,t),\ u(x,t))$ of (4.9), the corresponding linearized perturbation equations are

$$\frac{\partial \delta\phi}{\partial t} + \frac{\partial}{\partial x}(u\delta\phi + \phi\delta u) = 0,$$

$$\frac{\partial \delta u}{\partial t} + \frac{\partial}{\partial x}(\delta\phi + u\delta u) = 0,$$

while the adjoint equations (4.8) with respect to the inner product (4.10) are

$$\frac{\partial \delta^*\phi}{\partial t} + u\frac{\partial \delta^*\phi}{\partial x} + \Phi_0\frac{\partial \delta^* u}{\partial x} = 0, \tag{4.11a}$$

$$\frac{\partial \delta^* u}{\partial t} + \frac{\phi}{\Phi_0}\frac{\partial \delta^*\phi}{\partial x} + u\frac{\partial \delta^* u}{\partial u} = 0, \tag{4.11b}$$

as can be verified, for instance, by directly checking the adjointness condition (4.6).

A spectral numerical model of eqs. (4.9) has been developed, truncated at wavenumber 64. "Observations" $\hat{\phi}(x,t_1)$ and $\hat{\phi}(x,t_2)$ of the exact complete geopotential field have been extracted from an integration of the model at two times t_1 and t_2, and the gradient computed through the adjoint equations (4.11) has been used in a descent algorithm in order to find the solution of (4.9) minimizing the cost-function

$$J = \int \{[\phi(x,t_1) - \hat{\phi}(x,t_1)]^2 + [\phi(x,t_2) - \hat{\phi}(x,t_2)]^2\}\,dx.$$

The "observations" being, in this particular case, exactly compatible with one model solution it was known beforehand that the absolute minimum of J was equal to 0. However, the descent process was started from a first-guess which was not compatible with the "observed" solution, and the initial value of J was accordingly different from 0. Fig. 1 shows the variations of J with the number of descent steps, i.e. the number of successive explicit computations of the gradient. One sees that the value of J is decreased by a factor of about 2 at each step. In these experiments, the descent

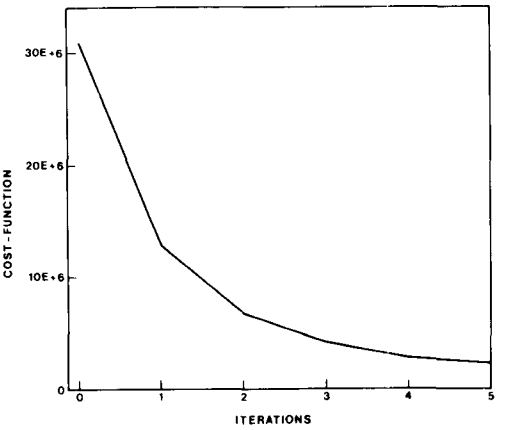

Fig. 1. Variation of the value of the optimal control cost-function J with the number of computations of the gradient. Units on the vertical axis are arbitrary.

algorithm, which we will not describe in detail, was very simple. Courtier (private communication) has obtained a significantly more rapid descent by using more efficient algorithms.

4.3. Additional comments

Some comments can be made about the use of optimal control in data assimilation. For the sake of simplicity, and in order to keep the approach sufficiently general, we have assumed that the basic evolution eq. (4.1) was continuous with respect to time. In any practical situation, one will of course use a discretized numerical model, and the adjoint model to be integrated must be the adjoint of the discretized model. This adjoint model can be obtained by following the derivation of Subsection 4.1, replacing time derivatives by the corresponding finite time differences. The final result will often be different from what would have been obtained from a direct discretization of the continuous adjoint (4.8). Some care must therefore be exercised in deriving the adjoint system.

Again for the sake of simplicity, we have assumed that the available observations were complete at each observation time t_i, i.e. that all the components of $Y(t_i)$ had been observed. This condition is not necessary, and the method can easily be extended to the case when only some components of $Y(t_i)$ have been observed, i.e., to the case of *any* space-time distribution of observations. Indeed, in the numerical example given above, only the geopotential ϕ was assumed to have been observed.

Finally, we have assumed that the cost-function J was quadratic with respect to the unknown solution $Y(t)$ (eq. (4.2)). This is not necessary either, and the approach given above can be generalized to any functional form for the cost-function J. It is not even necessary that the latter be a function of only the unknown solution $Y(t)$ and the case when J would depend also on the time derivatives of $Y(t)$ can be covered by the same general approach. A change in the cost-function J would only alter the forcing term added to the variable in the course of the integration of the adjoint equation (4.8), but not the adjoint operator $A^*(t)$ itself. The possibility of using a cost-function which depends not only on $Y(t)$, but also on its time derivatives may in practice be very useful: it will, for instance, allow the inclusion in J of an additional term limiting the amount of high frequency oscillations in the solution $Y(t)$. This opens the possibility that the "initialization" problem could be solved, at least to some extent, together with the "analysis" problem.

Thus, from a theoretical point of view, optimal control appears to be extremely powerful and to open a broad range of possible applications. The numerical results already obtained show that it is indeed capable of decreasing the functional on simple non-linear systems. Whether optimal control can produce satisfactory results in a practical situation at an acceptable computational cost remains of course to be determined, and additional work must be done to answer that question.

5. Conclusion

Two general algorithms for solving constrained minimization problems have been presented, together with numerical results obtained on simple non-linear meteorological examples. The presentation has been intended primarily at giving understanding of how these algorithms work. Therefore, some theoretical aspects, such as the precise conditions under which convergence is ensured, have not been considered. On the other hand, some emphasis has been put on the advantages of the algorithms presented. They ensure exact consistency between the constraint and the

reconstructed fields. They are founded on perfectly sound theoretical bases, which, in the authors' opinion, should guarantee that no mathematical difficulties will be encountered once these algorithms are properly implemented. They are very general and do not impose more than simple regularity conditions on the constraint and cost-function. These simple conditions will be verified in most, if not all, meterological problems, and once standard routines are developed, it will be possible to use them on a large variety of different problems. Finally, the numerical results presented in this article, although preliminary, do show that the algorithms which have been studied can effectively be used for solving problems of the type encountered in analysis or assimilation of meteorological observations.

Exact consistency between the constraint and the reconstructed fields is useful only if the constraint is itself a sufficiently accurate model of the physical reality. Generally speaking, the better the model, the more beneficial exact consistency between the analysis and the model will be. Operational numerical forecasting models now represent many atmospheric features and phenomena (such as baroclinic instability, three-dimensional structure of mid-latitude depressions, boundary layer physics, cloudiness and related effects) to a degree of accuracy which is not accounted for in the covariance functions used in operational optimal interpolation. Indeed, the constantly observed occurrence of a "spin-up" time of one day or so in numerical forecasts is an obvious proof of discrepancy between analyses and models. It is reasonable to think that better consistency between assimilations and models would improve the quality of numerical forecasts. Variational techniques provide one way of achieving better consistency.

An important remark must be made about the optimal control algorithms presented in Section 4. These algorithms require the preliminary development of the adjoint of the original constraint (2.3). However, once this adjoint is available, it can be used for many diverse applications. Adjoint techniques can generally be utilized for computing the gradient of a compound scalar function of a set of arguments, whether this gradient is needed for implementing a descent algorithm, as in the example discussed above, or for another purpose. In order to explicitly determine the gradient of a scalar function $u \rightarrow G(u)$, it will be necessary to perform the adjoint of the computations which, starting from u, lead to $G(u)$. Whenever these computations include the temporal integration of a dynamical model such as (4.1), the corresponding adjoint computations will include the backward integration of the adjoint equation (4.8). The forcing term added in the backward integration will depend on the particular function whose gradient is to be computed, but the adjoint eq. (4.8) will always be the same for a given dynamical model. This equation will be the appropriate tool to use whenever one needs to compute the gradient of an "output" parameter of the model with respect to its "input" parameters (or part of these parameters), such as initial and boundary conditions, and physical parameters. Kontarev (1980) and Hall and Cacuci (1983), for instance, have shown how the adjoint of a dynamical model can be used for sensitivity studies of the kind encountered in climatology and predictability problems. The development of the adjoint of a dynamical model may require a large amount of work, but once that work has been done, the adjoint will be available for a large variety of applications, of which analysis and assimilation are only examples.

Variational techniques therefore offer a number of theoretical advantages, but the corresponding price to be paid is the large amount of computing power required, both in terms of time and of memory size. For example, in the optimal control algorithm presented in Section 4, each iteration of the descent process will require at least one integration of the direct eq. (4.1) and one integration of the adjoint eq. (4.8). The computing costs of these two operations will be comparable. In the case of assimilation of observations performed over a 24-h time period, each iteration will therefore require the equivalent of at least 48 h of model integration. Moreover, to integrate the adjoint equation it is in principle necessary to store the entire history $Y(t)$, $t_1 \leqslant t \leqslant t_n$, produced by the corresponding forward integration of the direct equation. Considerable storage place will therefore be necessary in addition to fast computation. In any case, it is clear that, with present models and computers, optimal control could be used for data assimilation, without simplification, only if satisfactory convergence could be reached in a few iterations at most.

However, many possibilities can be thought of for reducing the cost of variational techniques. One

of these possibilities is for the successive descent steps of an optimal control assimilation to be performed with models of increasing resolution. That would make the first steps of the process, which do not require much accuracy, very economical. Also, experience of optimization techniques has shown that the choice of an appropriate descent algorithm can be very efficient in reducing the total cost of a minimization process. In the authors' opinion, the cost of variational techniques, while high, is not prohibitive, at least not for analyses or assimilations which are not part of an operational weather forecasting system, and therefore not subject to strict limitations in computing time. Moreover, future progress in computing power and theoretical methods will probably make feasible computations which are at present difficult or impossible to perform. For all these reasons, a detailed study of variational techniques, both in terms of their intrinsic qualities and of their computational cost, is of great interest.

6. Acknowledgments

Most of this article was written during a stay of one of the authors at the Cooperative Institute for Meteorological Satellite Studies (NOAA and University of Wisconsin). The authors thank John Lewis and John Derber not only for helpful discussions and useful comments on an earlier version of the manuscript, but also for assistance in the development of the methods described in this article. They also thank Claude Lemaréchal (INRIA, Paris) and Philippe Courtier (Météorologie Nationale, Paris) for useful discussions. The former provided the authors with codes for various descent algorithms (MODULOPT). The successive versions of the manuscript were typed with diligence and efficiency by Mrs. Gail M. Turluck, Mlle Marie-Christine Cally and Mme Irene Fontaine-Gilmour.

This research was supported by contract INAG ATP Recherches Atmosphériques.

REFERENCES

Arrow, K. J., Hurwicz, L. and Uzawa, M. 1958. *Studies in linear and non-linear programming*. Stanford University Press, Stanford, 229 pp.

Barnes, S. L. 1973. *Mesoscale objective analysis using weighted time-series observations*. NOAA Technical Memorandum ERL NSSL-62. Environmental Research Laboratories, National Severe Storms Laboratory, Norman, Okla., 60 pp.

Bertsekas, D. P. 1982. *Constrained optimization and Lagrange multiplier methods*. Academic Press, London, 395 pp.

Ciarlet, P. G. 1982. *Introduction to matrix numerical analysis and optimization* (in French). Masson, Paris, 279 pp.

Cressman, G. P. 1959. An operational objective analysis system. *Mon. Wea. Rev. 87*, 367–374.

Courant, R. and Hilbert, D. 1962. *Methods of mathematical physics*, vol. II. Interscience Publish., New York, 830 pp.

Daley, R. 1978. Variational nonlinear normal mode initialization. *Tellus 30*, 201–218.

Fortin, M. and Glowinski, R. 1982. *Augmented lagrangian methods* (in French). Dunod, Paris, 322 pp.

Gandin, L. S. 1965. *Objective analysis of meteorological fields* (translated from the Russian). Israel Program for Scientific Translations, Jerusalem, 242 pp.

Gill, P. E., Murray, W. and Wright, M. H. 1982. *Practical optimization*. Academic Press, London, 401 pp.

Ghil, M., Cohn, S., Tavantzis, J., Bube, K. and Isaacson, E. 1981. Applications of estimation theory to numerical weather prediction, in *Dynamical meteorology. Data assimilation methods*, eds. Bengtsson, L., Ghil, M. and Källen, E. Springer-Verlag, New York, 139–224.

Hall, M. C. G. and Cacuci, D. G. 1983. Physical interpretation of the adjoint functions for sensitivity analysis of atmospheric models. *J. Atmos. Sci. 40*, 2537–2546.

Kimeldorf, G. and Wahba, G. 1970. A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist. 41*, 495–502.

Kontarev, G. 1980. *The adjoint equation technique applied to meteorological problems*. ECMWF Technical Report No. 21, Reading U.K., 21 pp.

Lemaréchal, C. and Strodiot, J. J. 1981. On a bundle algorithm for non-smooth optimization. In *Nonlinear programming* eds. Magasarian, Meyer and Robinson. Academic Press, London, 245–282.

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

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

Lions, J. L. 1971. *Optimal control of systems governed by partial differential equations*. Springer-Verlag, Berlin, 396 pp.

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

Navon, I. M. and De Villiers, R. 1983. Combined penalty multiplier optimization methods to enforce integral invariants conservation. *Mon. Wea. Rev. 111*, 1228–1243.

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

Sasaki, Y. 1958. An objective analysis based on the variational method, *J. Meteorol. Soc. Japan 36*, 738–742.

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

Sasaki, Y. and Goerss, J. S. 1982. Satellite data assimilation using NASA data systems test 6 observations. *Mon. Wea. Rev. 110*, 1635–1644.

Soliz, P. and Fein, J. 1980. Kinematic analyses for a major winter storm during Amtex 75. *J. Meteorol. Soc. Japan. 58*, 16–32.

Temperton, C. 1982. Variational normal mode initialization for the ECMWF gridpoint model, in *The interaction between objective analysis and initialization*, ed. Williamson, D. Publication in Meteorology 127, McGill University, Montreal, 160–164.

Testud, J. and Chong, M. 1983. Three-dimensional wind field analysis from dual-Doppler radar data, part I. Filtering, interpolating and differentiating the raw data. *J. Clim. Appl. Meteorol. 22*, 1204–1215.

Tribbia, J. J. 1982. On variational normal mode initialization. *Mon. Wea. Rev. 109*, 455–470.

Wahba, G. 1982. Variational methods in simultaneous optimum interpolation and initialization, in *The interaction between objective analysis and initialization*, ed. Williamson, D. Publication in Meteorology 127, McGill University, Montreal, 178–185.