ITERATIVE APPROXIMATIONS TO OPTIMAL ANALYSIS

A C Lorenc
Met O 11, Meteorological Office,
Bracknell, England.

1 INTRODUCTION

Equations and notation in this lecture will follow those in a survey of analysis methods for NWP published by Lorenc (1986). One of the themes of this paper was that most analysis methods can be thought of as approximations to an optimal analysis equation. Differences in terminology and generality of definition of 'optimal' can obscure this, as can the different names and buzz-words used. In section 2 I set out and discuss the 'optimal' equation. Of course to do this without long-winded qualifications, I use my own preferred terminology, and apply my own understanding of 'optimal'. In section 3 I discuss iterative solution methods in general, and their possible advantages. In section 4 I survey many of the iterative methods which have been tried, trying to relate them to the 'optimal' equation of section 2, although many of them were not originally presented in this way.

2 DERIVATION OF BASIC ANALYSIS EQUATION

2.1 Optimal analysis

Our objective in analysis for NWP is to find initial conditions for our forecast model which fit the observations, and which are consistent with our prior knowledge as to the behavior of the atmosphere. We might also have requirements imposed by characteristics of the NWP model. Procedures to satisfy the latter I class as initialisation techniques. As the verisimilitude of the NWP model increases, the need for initialization of a good analysis should decrease. However the balance achieved by NWP initialization techniques may be the best numerical way of expressing our prior knowledge about the atmosphere; in this case these initialization techniques should be part of our optimal analysis procedure.

We need to find a single specification of optimal initial conditions, whereas our knowledge in fact only gives us a range of possibilities. We can use Bayesian arguments to find the most likely. This can be done mathematically by a variational minimization of a penalty function which is the sum of penalties measuring the fit to the observations, and the consistency with our prior knowledge. In general an iterative technique will be needed to solve this variational problem.

If the prior constraints are linear, and the error distributions are Gaussian, then these penalties are quadratic, and we are finding the minimum variance solution. In principle, direct (non-iterated) methods exist for this case. One such method is OI. In four-dimensions a direct method is the Kalman-Bucy filter. In practice both these methods require further approximations to be made before they can be applied to NWP analysis. Iterative solution methods, making different approximations, may still be the best choice.

2.2 Fit to observations

In order to compare an analysis specification, or model state ($\mathbf{x}$) with the observations ($\mathbf{y}_o$), to measure the fit between them, we need to be able to convert from one to the other. Where a direct conversion is possible, it is simpler to convert the observations towards the atmospheric parameters used to define a model state. We assume that this has already been done in $\mathbf{y}$. However there remain some observations, particularly from remote sensing, for which a unique conversion is impossible. Since our NWP models are becoming more realistic and comprehensive in their representation of the atmosphere, the transformation ($\boldsymbol{K}$) from $\mathbf{x}$ to $\mathbf{y}$ is in most cases more nearly unique. If the probability distribution functions of observational errors, and of errors in $\boldsymbol{K}$, are approximately Gaussian, with covariances given by $\mathbf{O}$ and $\mathbf{F}$ respectively, then the fit to the observations is measured by $(\mathbf{y}_o-\boldsymbol{K}(\mathbf{x}))^*(\mathbf{O}+\mathbf{F})^{-1}(\mathbf{y}_o-\boldsymbol{K}(\mathbf{x}))$. $\boldsymbol{K}$ is the forward process, or generalized interpolation from the model to the observations. The analysis problem can be thought of as finding the generalized inverse of $\boldsymbol{K}$, to interpolate from observations to model.

2.3 Fit to background

If we had no observations, we would presumably still have a best estimate of the most likely model state, we call this the background $\mathbf{x}_b$. This will of course be consistent with our prior knowledge about atmospheric structures. This prior knowledge will also tell us that some modes of perturbation about $\mathbf{x}_b$ are more likely than others. If this prior knowledge is sufficiently linear and Gaussian, with covariance $\mathbf{B}$, deviations from our prior knowledge can be measured by $(\mathbf{x}_b-\mathbf{x})^*\mathbf{B}^{-1}(\mathbf{x}_b-\mathbf{x})$.

2.4 Analysis equation

Our basic optimal analysis problem is thus simply to minimize a penalty functional (J) given by the sum of these two terms:

$$J = (\mathbf{y}_o-\boldsymbol{K}(\mathbf{x}))^*(\mathbf{O}+\mathbf{F})^{-1}(\mathbf{y}_o-\boldsymbol{K}(\mathbf{x})) + (\mathbf{x}_b-\mathbf{x})^*\mathbf{B}^{-1}(\mathbf{x}_b-\mathbf{x}) \qquad (1)$$

Note that I have been deliberately vague about the analysis specification, or model state $\mathbf{x}$. In its simplest traditional form it is a gridpoint field. For multivariate three-dimensional analysis it is many fields. It does not have to be represented by gridpoints (eg spectral representation is possible). It does not have to be for a single time. In each of these cases the analysis equation remains the same, although of course the meanings of $\mathbf{x}$, $\boldsymbol{K}$, and $\mathbf{B}$ will change.

The time aspect is worth discussing more fully. Since observations are distributed in time, one way of helping reduce the underdeterminacy is to use a four-dimensional distribution of them. This alone does not help, since we have to make $\mathbf{x}$ four-dimensional as well. However we have prior knowledge about the way the atmosphere behaves; this enables us to say that most four-dimensional states are very unlikely. If we use the NWP model itself in some way to express our knowledge of atmospheric evolution, then the analysis technique is called four-dimensional data assimilation.

3 ITERATIVE SOLUTIONS

There are several iterative solutions to the problem of finding the $\mathbf{x}$ which minimizes (1).

3.1 Advantages and disadvantages

Advantages of iterative methods are:

(a) They can be computationally cheaper.

(b) The need for data selection algorithms can be avoided. i.e. all observations can affect all grid-points.

(c) The need for data search algorithms can be avoided. i.e. it need never be necessary to list the observations near a given point, instead working in terms of the grid-points near each observation.

(d) Nonlinearities in $\boldsymbol{K}$, and non-Gaussian error distributions, can be treated. (The latter can be done by allowing $\mathbf{B}$ and $\mathbf{O}$ to be functions of the latest best estimate of $\mathbf{x}$, rather than of $\mathbf{x}_b$. Alternatively, a non-quadratic form for (1) can be derived and solved.)

(e) Special structure in $\mathbf{K}$ and $\mathbf{B}$ can be used to implement multiplication by them, rather than the semi-empirical covariance function modelling used in OI. In this way it is possible to introduce complex prior constraints into the analysis.

(f) If the algorithm is carefully designed, we can make a virtue out of the slow convergence of some cases; often these cases are those where the observations were in error (Grønås and Midtbø 1986).

Disadvantages of iterative methods are:

(g) Because of the iterations, we can be forced to make approximations in terms which we could afford to calculate more accurately in a direct method.

(h) Because of (g), and also if we have exploited (b), (d) and (e), it is more difficult to analyse the behavior of the scheme. For instance we cannot easily get an analytical expression for the expected analysis error.

(i) Because of (h), it is difficult to perform a preliminary, fully objective, quality control of the observations, using the analysis equations. (f) can be exploited however to integrate a quality control within the analysis, but this needs empirical tuning.

3.2 Descent algorithms

Iterative methods of finding the minimum of a functional are called descent algorithms. Gill *et al.* (1981) is a good textbook for these. Taking partial derivatives of (1) with respect to the elements of **x** gives

$$\mathbf{J'} = -2\{\ \mathbf{K^*}(\mathbf{O}+\mathbf{F})^{-1}(\mathbf{y_o}-\mathit{K}(\mathbf{x})) + \mathbf{B}^{-1}(\mathbf{x_b}-\mathbf{x})\ \} \qquad (2)$$

where $\mathbf{J'}$ is a vector of partial derivatives of J, and **K** is a matrix of partial derivatives of K. The simplest algorithm is to go from the current best estimate **x**[u] to a new estimate **x**[u+1] in the direction indicated by $\mathbf{J'}$, the direction of steepest descent. A more efficient method is to make the direction of search differ from that of previous iterations, as in the conjugate gradient algorithm. In principle the fastest algorithm, in terms of number of iterations, is that of Newton. This requires knowledge of $\mathbf{J''}$, the matrix of second derivatives of J.

$$\mathbf{J''} = 2\{\ \mathbf{K^*}(\mathbf{O}+\mathbf{F})^{-1}\mathbf{K} + \mathbf{B}^{-1}\ \} \qquad (3)$$

$$\mathbf{x}[u+1] = \mathbf{x}[u] - (\mathbf{J''})^{-1}\mathbf{J'} \qquad (4)$$

(We have neglected differentials of **K**, **O**, and **B** in (3), so strictly this is a Gauss-Newton iteration.) If (4) is evaluated exactly, then algebraic reorganizations are not very important. However the large matrix terms have to be approximated in practice. The grossest approximation is to replace the large matrix inverse by a scalar, taking us back to the steepest descent algorithm. For such approximations it is advantageous to remove a factor of $\mathbf{B}^{-1}$ from $\mathbf{J'}$. This can be thought of as transforming the representation used for **x** into one giving greater weight to those modes of **B** about which we have most prior uncertainty.

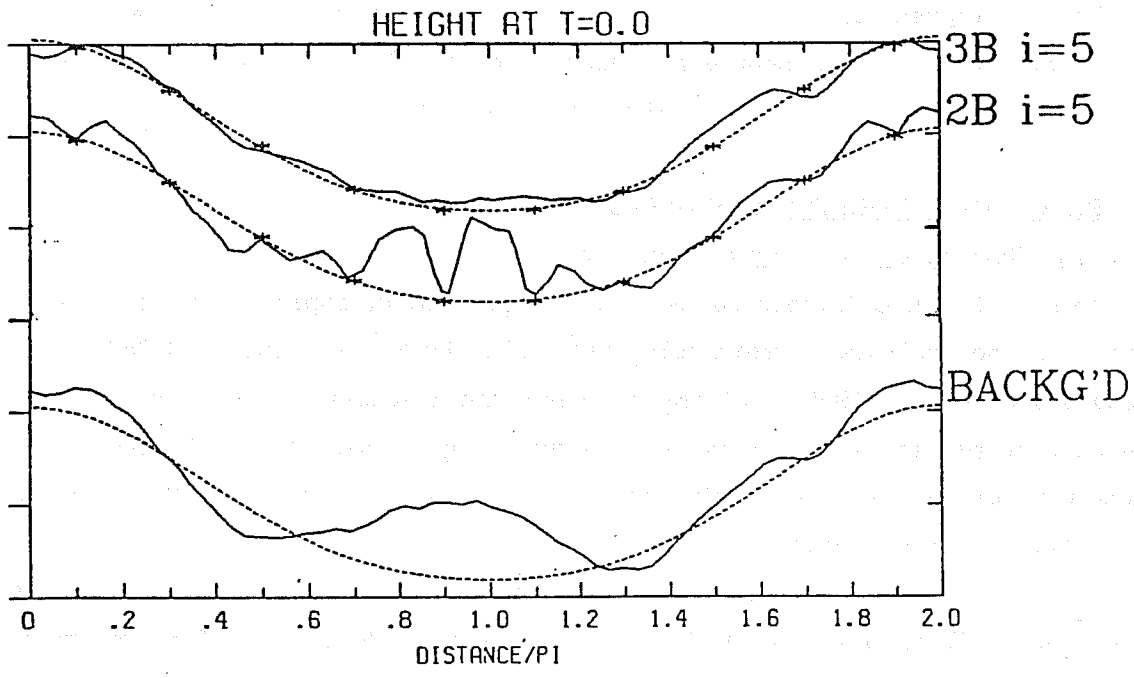

Fig.1 Three graphs are shown, shifted in vertical scale to avoid overplotting.
Dashed lines- 'truth' used in idealized one-dimensional analysis experiment.
Solid lines - *Bottom:* background used,
middle: 5th iteration of an analysis fitting the background, and the observations shown by * , using a conjugate gradient method and (1) and (2),
top: ditto , with control variable transformed by a factor **B**.

An example of the effect this can have is shown in Fig.1, from Lorenc (1988a). In this simple one-dimensional analysis it is assumed that we have prior knowledge that smooth changes to the background $\mathbf{x}_b$ are more likely than rough ones. A straightforward conjugate gradient algorithm first reduces the observational penalty so that the analysis fits the observations, then in later iterations it sets about reducing the roughnesses introduced. As can be seen from curve 2B in Fig.1, after 5 iterations many of these remain. In contrast, if we transform representation for $\mathbf{x}$, to remove the factor $\mathbf{B}^{-1}$, then these roughnesses are not introduced so readily. This is illustrated in curve 3B. Note that the fit to the observations in curve 3B is worse however. Both methods converge to the same limit, but in practice this is less important than ease of computation and speed of convergence for the various modes during the early iterations.

3.3 Background and first-guess

In iterative optimal analysis it is important to differentiate between these two terms, which are often used synonymously in OI literature.

The background ($\mathbf{x}_b$) contains useful prior information, which should be retained, so that after convergence the analysis should depend on it.

The first-guess is the most convenient starting point for our iteration; ideally the final analysis should not depend on it. Of course in practice the most convenient first-guess is usually the background.

4 PRACTICAL ITERATIVE SCHEMES

4.1 The Successive Correction method

This method was introduced as a rather pragmatic, empirically justified, scheme for the analysis of meteorological fields by Bergthorsson and Döös (1955) and Cressman (1959). Attempts to put the treatment of different scales on a more formal footing were made by Barnes (e.g. Barnes 1973). We shall derive the methods as an approximation to (4). Taking a factor of $\mathbf{B}^{-1}$ from $\mathbf{J}'$, and putting it in $\mathbf{J}''$, gives:

$$\mathbf{x}[u+1] = \mathbf{x}[u] + \mathbf{Q}\{ \mathbf{W}(\mathbf{y}_o - \boldsymbol{K}(\mathbf{x}[u])) + \mathbf{x}_b - \mathbf{x}[u] \} \qquad (5)$$

where $\mathbf{W}$ is an $N_x \times N_y$ matrix of weights, and $\mathbf{Q}$ is an $N_x \times N_x$ matrix of normalization factors, given by:

$$\mathbf{W} = \mathbf{B}\mathbf{K}^* (\mathbf{O}+\mathbf{F})^{-1} \qquad (6)$$

$$\mathbf{Q} = (\mathbf{W}\mathbf{K}+\mathbf{I})^{-1} \qquad (7)$$

Three approximations convert this iteration, which as it stands should converge to the optimal solution, into the successive correction method:

(a) As in OI, we model $\mathbf{B}\mathbf{K}^*$ by a continuous covariance function, so that (6) gives us weights which depend on the distance between observations and grid points and on the observational errors.

(b) We approximate $\mathbf{Q}$ by the reciprocal of the sum of weights at each grid-point, rather than the $N_x \times N_x$ matrix inverse of (7). It is helpful to think of $\mathbf{Q}$ as a function of the data-density; this approximation makes it a local function at each grid-point. Note that $\mathbf{Q}$ only affects the rate of convergence of the iteration, not the final limit.

(c) We start the iteration from $\mathbf{x}_b$, and only continue for a finite number of iterations. Because of our rescaling by $\mathbf{B}$, $\mathbf{x}$ will not differ too much from $\mathbf{x}_b$ in modes which are strongly penalized in our background penalty, and the $\mathbf{x}_b - \mathbf{x}[u]$ term in (5), which arises from the background penalty, can be neglected.

Thus the successive correction equation is

$$\mathbf{x}[u+1] = \mathbf{x}[u] + \mathbf{Q}\,\mathbf{W}\,(\mathbf{y}_o - \boldsymbol{K}(\mathbf{x}[u])) \qquad (8)$$

where the weights (w) which form matrix $\mathbf{W}$ are given by an empirical formula rather similar to (6), and the normalization factors ($\mathbf{Q}$) are given by a diagonal matrix of terms $(\Sigma w+1)^{-1}$, or something similar (sometimes the 1 is omitted). Sometimes smoothing is applied between iterations; this might be regarded as a rather more complicated $\mathbf{Q}$.

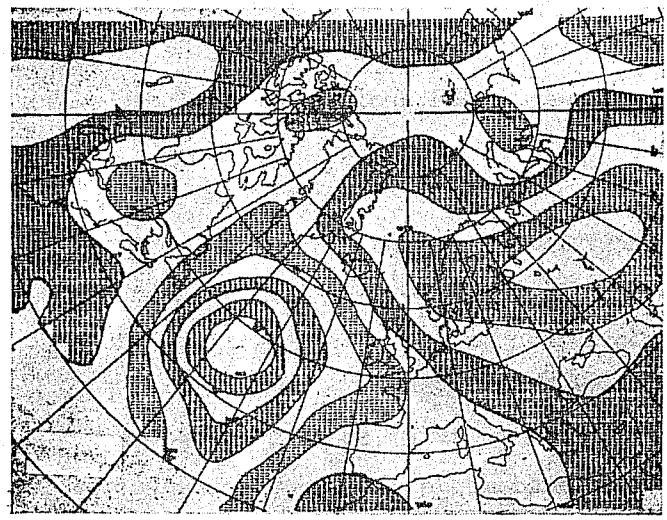

Fig. 4a. First scan in an analysis of the surface pressure. The radius of influence (N) of the stations in this scan is 6 grid units=1800 km.

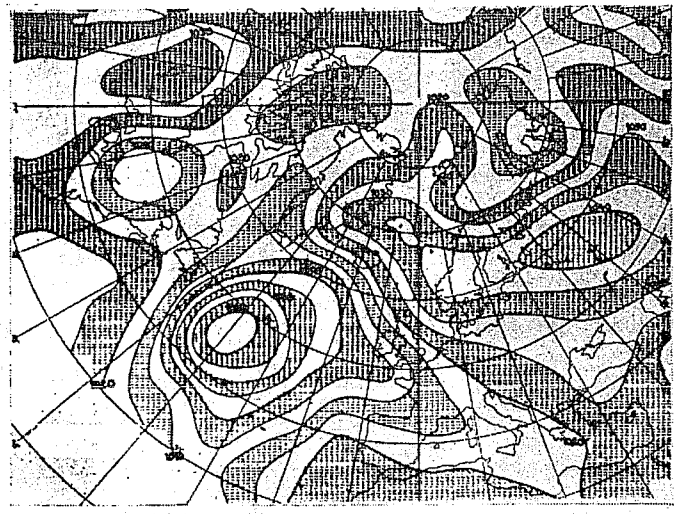

Fig. 4c. Third scan. N=2.5 grid units=750 km.

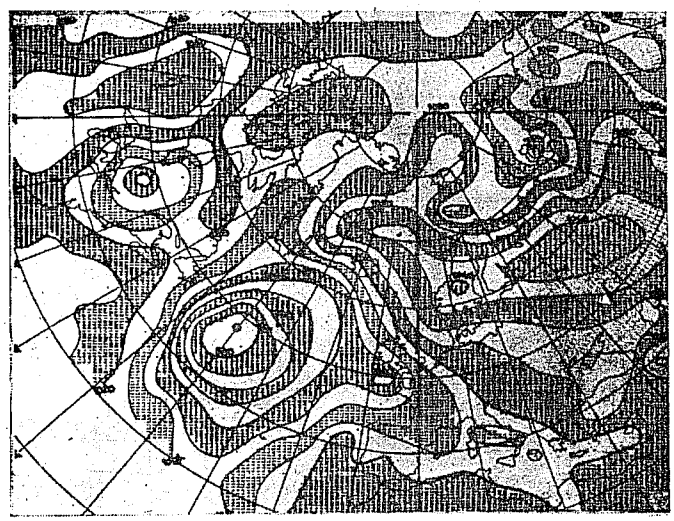

Fig 4e. Fifth scan. N=1 grid unit=300 km.

Fig.2 Surface pressure fields after 1, 3, and 5 scans of a successive correction analysis method (from Döös 1969).

Note that because of the omission of the forcing towards the background, the method will converge to fit the observations exactly, if this is possible. (8) thus only converges to the optimal analysis in the case where the observations are perfect.

Generally it has been found that convergence is faster if large horizontal scales are done in early iterations, with an appropriate form for the weights and a large radius of influence for observations. Later iterations put in the detail by having a rapid drop to zero in the weights formula. This is illustrated in Fig.2, from Döös (1969).

It was also noted empirically at an early stage that the successive correction method as outlined above gives each observation in data dense areas equal weight to isolated observations. Although the net weight after a large number of iterations will correct this, convergence will be speeded by making the weights dependent on the local observation density (e.g. Döös 1969). We can go to the extreme in doing this by replacing the gridpoint normalization $\mathbf{Q}$, which we have approximated by the the reciprocal of the sum of weights (and hence the observation density) at each gridpoint, by a similar observation normalization $\mathbf{Q}'$, evaluated at observation positions:

$$\mathbf{x}[u+1] = \mathbf{x}[u] + \mathbf{W}\,\mathbf{Q}'\,(\mathbf{y}_o - K(\mathbf{x}[u])) \qquad (9)$$

This is the form used in the Meteorological Office's analysis correction scheme (Lorenc *et al.* 1988, Macpherson 1987,1988). Note that it has the advantage of preserving any linear properties of the weights, such as a constraint that wind increments should be non-divergent, while (8) does not. Like (8), (9) treats the observations as perfect, and converges to fit them exactly. Bratseth (1986) proposed a way round this, by replacing $\mathbf{y}_o$ in the iteration by an (iterated) estimate of what the perfect observations would be:

$$\mathbf{x}[u+1] = \mathbf{x}[u] + \mathbf{W}\,\mathbf{Q}'\,(\mathbf{y}[u] - K(\mathbf{x}[u])) \qquad (10)$$

$$\mathbf{y}[u+1] = \mathbf{y}[u] + \mathbf{Q}'\,(\mathbf{y}[u] - K(\mathbf{x}[u])) \qquad (11)$$

if we go back to (6) for the weights, and use

$$\mathbf{Q}' = (\mathbf{KW}+\mathbf{I})^{-1} \qquad (12)$$

for the normalization factor, then this iteration also converges to the optimal analysis. As for (5), any approximations made in $\mathbf{Q}'$ only affect the rate of convergence, not the final limit. Fig.3 (from Bratseth 1986) illustrates this scheme for a simple one-dimensional analysis.

This convergence to the optimal result can appear surprising, if we only look at the weight formula for a single iteration, since it is impossible to get negative weights, or to get normalized weights which add to more than one. In the OI formula for weights both can occur, to give the extrapolation effects seen in fig.3, and to give extrema greater than any single observation. As long

as we take proper account of the accumulated effect over several iterations, which occurs through the use of $-K(x[u])$ in the correction formula for $x[u+1]$, then the same is true for the **net** weight in the successive correction method. Similar comments apply to other iterative schemes described below. It can be very misleading only to look at the mathematical properties of the first iteration; extrapolation effects occur only in the accumulated result of several iterations.

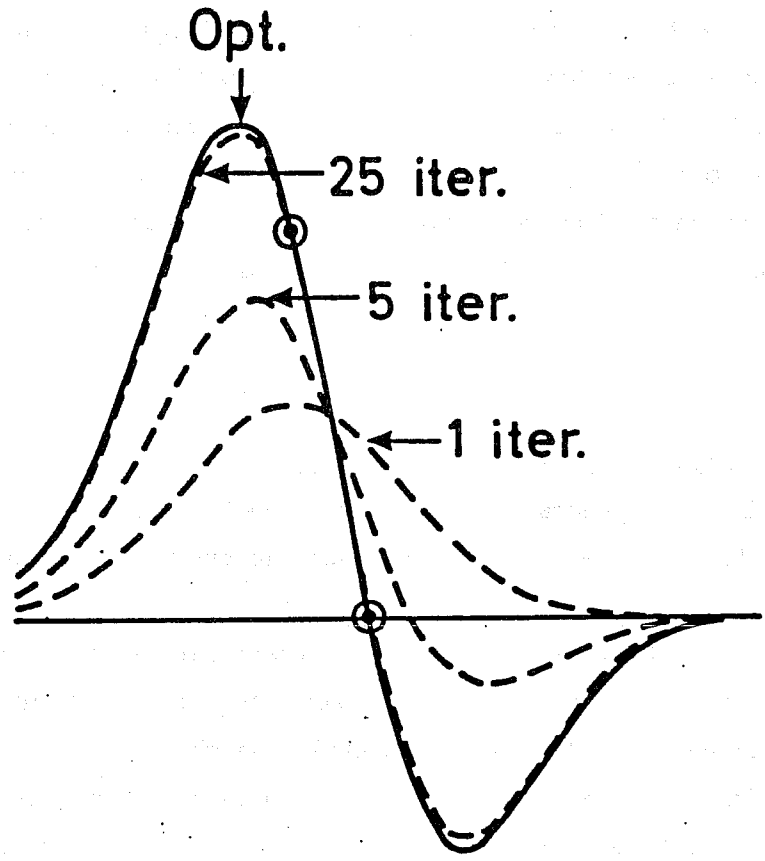

Fig.3 Analysis of two perfect observations ⊙ .
Solid lines: background, and OI analysis.
Dashed lines: modified successive correction analysis after 1, 5, and 25 iterations (from Bratseth 1986).

4.2 Alternative methods of expressing B

B is our way of expressing our prior knowledge of which modes of variation about $\mathbf{x}_b$ are most likely. In OI, and in the successive correction method, it is represented by an empirically determined background error covariance function, which is the major term in the expression for the weights. However we might say that our only prior knowledge is that small scales are less likely. Multiplication by **B** is then equivalent to a filtering of small scales (and a scaling of the values by the background error variance). $\mathbf{K}^*$ is the adjoint or transpose of the interpolation (K) from grid to observations; it

can be represented exactly, or approximated by accumulating observation increments at surrounding gridpoints. Then, again omitting the background forcing term, we get:

$$\mathbf{x}[u+1] = \mathbf{x}[u] + \mathbf{QB}\ \mathbf{K}^{*}(\mathbf{O}+\mathbf{F})^{-1}(\mathbf{y}_{o}-K(\mathbf{x}[u])) \qquad (13)$$

An old Meteorological Office scheme used orthogonal polynomials to do this filtering (Dixon 1976). An old NMC scheme used Hough functions (Flattery 1971). Experience with both these methods, as with the successive correction method, was that rapid convergence was aided by doing large scales in the early iterations. The latter scheme also used prior knowledge about balance, since in early iterations wind and height fields were coupled by the use of balanced funtions. Phillips (1986) has shown that there are similarities between empirically determined background error covariance functions, and those obtained assuming equipartition of energy between geostrophic normal modes. In this case multiplication by **B** is equivalent to projecting on these modes, scaling appropriately, and projecting back again. Parrish (1988) has proposed a scheme along these lines.

4.3 Four-dimensional data assimilation

(1) is completely general; **x** can be expressed in any way, including a four-dimensional field, and **y** can contain observations over a period of time. To get an optimal analysis we must then use our prior knowledge about the evolution of the atmosphere. The best way to do this in practical NWP is to use a forecast model. Since forecast models are not perfect, the constraint imposed on our four-dimensional analysis, that its time evolution should be consistent with the model's, should not be strictly enforced. Practical implementation of a optimal four-dimensional analysis scheme with such a constraint is severely limited by available computer resources. The storage and manipulation of high-resolution four-dimensional fields requires many more resources than running a NWP model, which only manipulates three-dimensional fields. The traditional NWP method of approximating four-dimensional data assimilation is the analysis-forecast cycle, in which observations are inserted using a purely three-dimensional analysis procedure into a background forecast from the results of the previous analysis. In this case our $\mathbf{x}_b$ summarises the information from all previous observations; it often contains more information than the current observations. If the evolution of errors in the forecast can be modelled by a linear K, then matrix **B** of the error covariances at each time can be forecast, and the optimal final state can be found explicitly, using a Kalman filter. This method does allow for deficiencies in our forecast model, since errors in $\mathbf{x}_b$ are represented by **B**. However **B** is so large as to make its accurate manipulation impracticable.

An alternative to the intermittent analysis-forecast cycle is repeated insertion data assimilation. In this too the forecast model is run forward only, with the model state being nudged towards the observations over a period of time (Lorenc 1984). This can be regarded as an analysis-forecast cycle, with the analysis being done by an iterative method, and with iterations of the analysis and time-steps of the model intermingled (Lorenc *et al* 1988). The approximations used to model the time variation of **B**, and the fact that observations from different times are not analysed together, mean that analysis-forecast cycles are not optimal in their use of the time-dimensional information (e.g. tendency information, and advection of tracers) from the observations.

If we ignore imperfections in our forecast model, and impose the model's equations as a strong constraint, then many of the computational difficulties of a four-dimensional analysis can be avoided, since all possible four-dimensional fields are defined by their initial three-dimensional fields and the NWP model. We can thus reduce the analysis problem to that of finding the best initial state (le Dimet and Talagrand 1986). $\boldsymbol{K}$ is extended to include the forecast from this initial state to the observation times, as well the interpolations in space to the observation positions. Since we can run the NWP model which forms the major component of $\boldsymbol{K}$, it is not too difficult also to run its adjoint, and evaluate explicitly $\mathbf{K}^*$. We are including more observations in the analysis, so the background information is less important, and we can approximate (or even ignore) the background penalty. Then an iterative solution to the analysis equation (1) becomes feasible, using (2) and a descent algorithm such as the conjugate gradient method. Talagrand (1988) and Courtier (1988) describe these variational four-dimensional methods. They are computationally expensive; each iteration requires an integration of the NWP model, the storage of all its results, and integration of its adjoint. Thus the full analysis will consume at least an order of magnitude more computer resources than the NWP model used.

Since explicit computation of terms like $\mathbf{K}^*\mathbf{B}\,\mathbf{K}$ is completely impractical for present forecast models and computers, an iterative method which avoids evaluation of $\mathbf{J}''$ is necessary. However it is possible to combine the OI approach in space with an iteration for the time-dimension (Lorenc 1988b).

4.4 Nonlinear problems

Iteration is necessary whenever (1) is not quadratic in $\mathbf{x}$. This occurs when $\boldsymbol{K}$ is nonlinear, as it is for the NWP models used in the four-dimensional assimilations described above. Another example is the inversion of cloudy radiances. This can be regarded as a multivariate one-dimensional optimal analysis using (1). Eyre (1987,1988) solves the equation using an iterative

Newton method.

If the probability distribution function of observational errors is non-Gaussian, the observational penalty in (1) is non-quadratic, and iteration is necessary (Purser 1984).

The properties of these nonlinear problems are discussed more fully in Lorenc (1988a,c).

References

Barnes, S.L. 1973
'Mesoscale objective analysis using weighted time-series observations'. NOAA Tech Memorandum ERL NSSL-62.

Bergthorsson, P and Döös B.R. 1955
'Numerical weather map analysis'. Tellus 7 329-340.

Bratseth, A.M. 1986
'Statistical interpolation by means of successive corrections'. Tellus 38A 439-447.

Cressman, G.P. 1959
'An operational objective analysis scheme'. Mon, Wed, Rev 87 367-374.

Dixon, R. 1976
'An objective analysis system using orthogonal polynomials'. GARP WGNE Report No 11 pp 73-85.

Döös, B.R. 1969
'Numerical analysis of meteorological data'. Lectures on numerical short-range weather prediction, Moscow 1965 pp 679-706. Pub. Hydrometeoizdat Leningrad 1969.

Eyre,J.R. 1987
'Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation; theory and simulation for TOVS.' submitted to Quart. J. Roy. Met. Soc.

Eyre,J.R. 1988
'Satellite retrieval research at the UK Metorological Office' ECMWF Seminars on *Data assimilation and the use of satellite data.* September 1988, (This volume).

Flattery, T.W. 1971
'Spectral models for global analysis and forecasting'. Proc Sixth AWS Technical Exchange Conf. US Naval Academy, 1970, Air Weather Service Tech. Rept 242 pp42-53.

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

Grønås,S. and Midtbø,K.H. 1986
'Operational multivariate analyses by successive corrections' in *Short- and medium-range NWP.* Collection of papers presented at WMO/IUGG NWP Symposium, Tokyo 4-8 Aug 1986. J. Met. Soc. Japan.

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

Lorenc, A.C. 1984
'Data assimilation by repeated insertion into a forecast model - Principles, practice, problems and plans.' ECMWF Seminar on Data Assimilation systems and OSE with particular emphasis on FGGE. September 1984. Met O 11 Technical Note No 193.

Lorenc, A.C. 1986
'Analysis methods for numerical weather prediction.' Quart. J. Roy. Met. Soc. 112, 1177-1194.

Lorenc, A.C., R.S.Bell, and B.Macpherson 1988
'Developments in data assimilation at the UK Meteorological Office' Preprints, 8th AMS Conference on Numerical Weather Prediction. Baltimore, Md. Feb.22-26, 1988. 398-401

Lorenc, A.C. 1988a
'Optimal nonlinear Objective Analysis.' Quart. J. Roy. Met. Soc. 114, 205-240.

Lorenc, A.C. 1988b
'A practical approximation to optimal 4-dimensional objective analysis' Mon. Wea. Rev. 116, 730-745.

Lorenc, A.C. 1988c
'Optimal Nonlinear Analysis' ECMWF Seminars on *Data assimilation and the use of satellite data*. September 1988, (This volume).

Phillips, N.A. 1985
'The spatial statistics of random geostrophic modes and first-guess errors'. Tellus 38A 314-332.

Purser, R.J. 1984
'A new approach to the optimal assimilation of meteorological data by iterative Bayersian analysis'. Preprints, 10th conference on weather forecasting and analysis. Am Met Soc. 102-105.