Tellus (1986), *38A*, 439–447

Statistical interpolation by means of successive corrections

By ARNE M. BRATSETH, *Institute of Geophysics, University of Oslo, P.O. Box 1022, Blindern 0315 Oslo 3, Norway*

(Manuscript received June 18; in final form November 29, 1985)

ABSTRACT

A theoretical investigation of the successive correction method for objective analysis of meteorological data is presented, and an explicit formula for the interation limit is offered. By means of this result, it is possible to formulate the method so that this limit conforms with statistical optimal interpolation. Simple experiments indicate that desirable properties of statistical interpolations are easily transferred to the successive correction approach. This includes the use of observation error statistics and multivariate analysis.

1. Introduction

In numerical weather prediction, it is necessary to transform atmospheric observations into a form suitable for numerical calculations. This is usually done by interpolation to some kind of regular grid, and is referred to as "objective analysis". During the years a number of methods have been suggested for this problem; see for instance the review by Gustafsson (1981). For a long period this field was dominated by so-called successive correction methods which are of a rather empirical nature (Bergthorsson and Döös, 1955 and Cressman, 1959). In recent years the interest has shifted towards statistical (or optimal) interpolation schemes based on ideas first introduced by Eliassen (1954) and Gandin (1963). Statistical interpolation has the advantage that information from different sources (different observation systems, previous forecasts, climatology etc.) may be combined in a way which pays attention to the assumed reliability of each source. Within this framework it is also possible to make use of cross-correlations between different meteorological variables, for instance pressure and wind. Statistical interpolation, however, is computationally expensive, and for this reason successive correction methods are still of considerable interest.

In this paper, we shall investigate the relationship between successive corrections and statistical interpolation. An explicit expression for the limit of the successive correction iteration procedure will be presented, and it will be shown that the method may be modified so that this limit coincides with statistical interpolations. Thus it seems possible to carry over the desirable properties of statistical interpolations to the simple and computationally inexpensive framework of successive corrections. This is illustrated by means of some examples. A proof of the convergence of the method is outlined in the appendix. Ideas presented in this paper are at present used in an operational numerical weather prediction routine at the Norwegian Weather Service.

2. Statistical interpolation

We shall use the following notation:
F^{O} refers to observations;
F^{P} refers to a first guess (prognoses);
F^{A} refers to an interpolated value (analysis);
F^{T} refers to the true value.
Furthermore deviations from the first guess will be denoted by $f = F - F^{\mathrm{P}}$.

In statistical interpolation, a linear relation is assumed for the deviation from the first guess.

$$f_x^{\mathrm{A}} = \sum_{i=1}^{n} p_i f_i^{\mathrm{O}}. \tag{1}$$

Here the subscript x refers to the analysis point and i refers to n observation points. The weights p_i are

chosen so as to minimize the expected square error:

$$\overline{E^2} = \overline{(f_x^A - f_x^T)^2} = \overline{(f_x^T)^2} - 2\sum_{i=1}^{n} \overline{f_x^T f_i^O} p_i + \sum_{i=1}^{n}\sum_{j=1}^{n} \overline{f_i^O f_j^O} p_i p_j. \quad (2)$$

The bar represents an average over a large number of cases. The minimizing weights satisfy the n equations:

$$\sum_{i=1}^{n} \overline{f_i^O f_j^O} p_i = \overline{f_x^T f_j^O}, \qquad j = 1, \ldots, n. \quad (3)$$

3. The successive correction method

The method of successive corrections is based on an iteration formula:

$$F_x^A(v+1) = F_x^A(v) + \sum_{i=1}^{n} a_{xi}(F_i^O - F_i^A(v)). \quad (4)$$

Here v counts the number of iterations made, and a_{xi} represents weights which are usually based on experience. The iterations are initiated by choosing $F^A(0) = F^P$.

Now we shall study the result of this procedure. It is convenient to introduce vector notation:

$$\boldsymbol{F} = \begin{bmatrix} F_1 \\ \vdots \\ F_n \end{bmatrix} \quad \boldsymbol{a}_k = \begin{bmatrix} a_{k1} \\ \vdots \\ a_{kn} \end{bmatrix}.$$

Eq. (4) can then be written as:

$$F_x^A(v+1) = F_x^A(v) + \boldsymbol{a}_x^t \cdot (\boldsymbol{F}^O - \boldsymbol{F}^A(v)), \quad (5)$$

where $\boldsymbol{a}^t$ is the transpose of $\boldsymbol{a}$. In order to update grid point values by means of (5) we also need to compute estimates for the observed values. This is usually done by interpolation from current grid point values. In this work, however, we shall use (5) to update observation estimates. For the total observation vector we then get:

$$\boldsymbol{F}^A(v+1) = \boldsymbol{F}^A(v) + \boldsymbol{A} \cdot (\boldsymbol{F}^O - \boldsymbol{F}^A(v)), \quad (6)$$

where

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1^t \\ \vdots \\ \boldsymbol{a}_n^t \end{bmatrix}$$

Eq. (6) has the general solution

$$\boldsymbol{F}^A(v) - \boldsymbol{F}^O = (\boldsymbol{I} - \boldsymbol{A})^v \cdot (\boldsymbol{F}^A(0) - \boldsymbol{F}^O). \quad (7)$$

Having solved the problem at the observation points, we can now find the solution at a general interpolation point by using (5) repeatedly and inserting from (7):

$$\begin{aligned} F_x^A(v+1) &= F_x^A(0) + \boldsymbol{a}_x^t \cdot \sum_{m=0}^{v} (\boldsymbol{F}^O - \boldsymbol{F}^A(m)) \\ &= F_x^A(0) + \boldsymbol{a}_x^t \cdot \sum_{m=0}^{v} (\boldsymbol{I} - \boldsymbol{A})^m \cdot (\boldsymbol{F}^O - \boldsymbol{F}^A(0)) \quad (8) \\ &= F_x^A(0) + \boldsymbol{a}_x^t \cdot (\boldsymbol{I} - (\boldsymbol{I} - \boldsymbol{A})^{v+1}) \cdot \boldsymbol{A}^{-1} \cdot \\ &\quad (\boldsymbol{F}^O - \boldsymbol{F}^A(0)). \end{aligned}$$

Now we shall assume that:

$$(\boldsymbol{I} - \boldsymbol{A})^{v+1} \to 0 \qquad \text{as } v \to \infty.$$

If this is not the case, the iteration procedure will not converge. Remembering that $F^A(0) = F^P$, we then get:

$$F_x^A = F_x^P + \boldsymbol{a}_x^t \cdot \boldsymbol{A}^{-1} \cdot (\boldsymbol{F}^O - \boldsymbol{F}^P). \quad (9)$$

This formula corresponds to (1) with the weights:

$$\boldsymbol{p}^t = \boldsymbol{a}_x^t \cdot \boldsymbol{A}^{-1}. \quad (10)$$

Just like the error minimizing weights of the previous section these weights also satisfy a system of linear equations:

$$\boldsymbol{A}^t \cdot \boldsymbol{p} = \boldsymbol{a}_x \qquad \text{or} \qquad \sum_{i=1}^{n} a_{ij} p_i = a_{xj}. \quad (11)$$

This means that if the system (11) can be made equivalent to the system (3), then the successive correction analysis (if it converges) will be equivalent to the statistical interpolation.

4. Improved successive corrections

Eq. (11) is readily transformed into:

$$\sum_{i=1}^{n} \overline{f_i^O f_j^O} [a_{ij} \overline{f_x^T f_j^O} / (a_{xj} \overline{f_i^O f_j^O})] p_i = \overline{f_x^T f_j^O}. \quad (12)$$

When the expression in square brackets is equal to unity, this equation becomes equal to (3). It is interesting to consider the case when:

$$a_{ij} = \overline{f_i^T f_j^T} / \sum_{k=1}^{n} \overline{f_i^T f_k^T}. \quad (13)$$

which corresponds roughly to the simplest applications of the successive correction method. Assuming perfect observations ($f^O = f^T$), the expression in square brackets becomes:

$$D_{ix} = \sum_{k=1}^{n} \overline{f_x^T f_k^T} / \sum_{k=1}^{n} \overline{f_i^T f_k^T}. \tag{14}$$

This expression may be looked upon as a measure of the observation density in the vicinity of the interpolation point x, relative to that of the observation point i. If the interpolation point is sufficiently far removed from the closest observation point, D_{ix} will become much smaller than one. Then the weights p_i determined by (12) will become bigger than the corresponding optimal weights, and the successive correction method pays too much attention to the observations relative to the first guess field. In a similar way we realize that observations occurring in areas of high observation density will be given too much weight relative to observations in areas of low density.

These defects of the successive correction method have been known for a long time, and have often been compensated for by means of ad hoc modifications to the weights a_{ij} based on experiments. Eq. (12), however, provides a theoretical basis for such modifications. In particular we observe that the choice:

$$a_{ij} = \overline{f_i^O f_j^O} / M_j \tag{15a}$$

$$a_{xj} = \overline{f_x^T f_j^O} / M_j, \tag{15b}$$

will make (12) identical to (3). Assuming that M_j can be chosen in such a way as to guarantee fast convergence, these formulae open the possibility of making optimal interpolations by means of successive corrections. This also includes the prospect of a rational treatment of different observation types with different characteristics and of a multivariate analysis where f_i^O represents a mixture of intercorrelated variables, for instance pressure and velocity in the case of numerical weather prediction.

It is interesting to note that the denominator M_j depends on the observation point (index j) only, and can be computed once and for all for each observation, while the corresponding denominator in (13) depends on the interpolation point (index i). It is also important to appreciate the remarkable fact that M_j does not influence the iteration limit (if it exists). Finally it should be noted that (15) suggests different weights depending on whether we are estimating observed values or true values. Thus (15a) should be used to update observation estimates, while (15b) should be used to update grid point values. This distinction is important when the observations are assumed to contain errors. Eq. (7) shows that the observation estimates will converge towards the observed values. Estimates for the corresponding true values, however, may not fit the observations when observation errors are accounted for. An example of this will be discussed in Section 6.

During the final stages of this work, we became aware of a work by Franke and Gordon (1983) in which a variant of the successive correction method is shown to converge towards the optimum solution. Their method corresponds to the special case when M_j is a universal constant independent of j.

5. A suggestion for M_j

We now turn to the problem of choosing the denominator M_j in (15) in such a way as to provide fast convergence of the iteration procedure. As a guideline it seems reasonable that we should try to make as good interpolations as possible in one single iteration. If we have a single observation f_1^O, then $M_1 = \overline{f_1^O f_1^O}$ leads to the optimal weight as given by (3). In the case of two observations this result is still valid provided that the observations are not correlated. On the other hand, if the observations are correlated to the extent that they contain the same information, then the correct result is $M_1 = 2\overline{f_1^O f_1^O}$. Inspired by this example we shall suggest the following formula for the general case:

$$M_j = \overline{f_j^O f_j^O} \sum_k W(f_j^O, f_k^O). \tag{16}$$

Here $W(f_j^O, f_k^O)$ is intended to be a measure of the degree to which observation j may be inferred from observation k. We shall require $W(f_j^O, f_k^O) = 1$ if the two observations contain the same information, and $W(f_j^O, f_k^O) = 0$ if they do not share any information at all. A possible way of defining $W(f_j^O, f_k^O)$ is

$$W(f_j^O, f_k^O) = (1 - \overline{E_{jk}^2} / \overline{f_j^O f_j^O})^{1/2}, \tag{17}$$

where E_{jk}^2 is the mean square error which is to be expected when f_j^O is estimated by means of f_k^O in

the optimal way. $\overline{E_{jk}^2}$ is readily computed from (2). In the case when both observations are scalars, (17) gives:

$$W(f_j^o, f_k^o) = |\overline{f_j^o f_k^o}/(\overline{f_j^o f_j^o}\ \overline{f_k^o f_k^o})^{1/2}| \tag{18}$$

which may be interpreted as the absolute value of a correlation coefficient. The definition (17) may also be applied when one or both of the observations are vectors. The error E_{jk} is defined to be the norm of the error vector whenever f_j^o is a vector.

As an example we shall study the multivariate case when f represents both a scalar field ψ and a vector field with two components $\boldsymbol{v} = (u, v)$. For simplicity we shall assume that $\overline{u_i^2} = \overline{v_i^2}$ and that $\overline{u_i v_i} = 0$. Application of (17) together with (2) and (3) then gives:

$$W(\psi_j, \psi_k) = |\overline{\psi_j \psi_k}/(\overline{\psi_j^2}\ \overline{\psi_k^2})^{1/2}|$$

$$W(\psi_j, \boldsymbol{v}_k) = \{[(\overline{\psi_j u_k})^2 + (\overline{\psi_j v_k})^2]/(\overline{\psi_j^2}\ \overline{u_k^2})\}^{1/2}$$

$$W(\boldsymbol{v}_j, \psi_k) = \{[(\overline{u_j \psi_k})^2 + (v_j \psi_k)^2]/(2\,\overline{u_j^2}\ \overline{\psi_k^2})\}^{1/2}$$

$$W(\boldsymbol{v}_j, \boldsymbol{v}_k) = \{[(\overline{u_j u_k})^2 + (\overline{u_j v_k})^2 + (\overline{v_j u_k})^2 + (\overline{v_j v_k})^2]/(2\,\overline{u_j^2}\ \overline{u_k^2})\}^{1/2}. \tag{19}$$

Note that these quantities are invariants with respect to the choice of the coordinate system. It turns out that this definition of M_j guarantees the convergence of the procedure. A proof is presented in the appendix.

6. Experiments

At this stage, some simple experiments are appropriate. We shall focus on the example of the last section which will be specialized to the case when ψ is a streamfunction of the velocity vector $\boldsymbol{v} = (-\partial\psi/\partial y, \partial\psi/\partial x)$. Furthermore we shall assume that:

$$\overline{\psi_i \psi_j} = \exp\{-[(x_i - x_j)^2 + (y_i - y_j)^2]/2\}. \tag{20}$$

All other statistical properties which we need, can then be computed by means of relations like:

$$\overline{\psi_i \left(\frac{\partial}{\partial x}\psi\right)_j} = \frac{\partial}{\partial x_j}\overline{\psi_i \psi_j} \tag{21}$$

etc.

Within this framework we have made a number of experiments with different distributions of simulated observations in two dimensions. The general experience obtained from these tests is best illustrated by means of some simple one-dimensional examples.

Fig. 1 shows the univariate case with two observations $\psi_1 = 1$ and $\psi_2 = 0$, represented by dots inside circles. The observations are assumed to be perfect, they are separated by a distance $x_2 - x_1 = \sqrt{2}$, and the first guess is zero. Note that the statistical assumptions employed are revealed by the first iteration which gives $\overline{\psi_x \psi_1}/M_1$. The figure indicates a fast convergence towards the optimal result (full line). Note also the very inadequate limiting result which is obtained by using the standard successive correction weights (13) (dotted line).

The convergence becomes slower when the observations are closer together. Fig. 2 shows the case when the distance is reduced by one half ($x_2 - x_1 = \frac{1}{2}\sqrt{2}$). In this simple case a theoretical explanation based on (8) is worth-while. It is convenient to expand the vector $\boldsymbol{F}^o - \boldsymbol{F}^a(0)$ in eigenfunctions of $\boldsymbol{I} - \boldsymbol{A}$. Eq. (8) then takes the form

$$F_x^a(v+1) = F_x^a(0) + \boldsymbol{a}_x^t \cdot \boldsymbol{A}^{-1} \cdot (\boldsymbol{F}^o - \boldsymbol{F}^a(0)) - \boldsymbol{a}_x^t \cdot \boldsymbol{A}^{-1} \cdot \sum_{i=1}^{n} \lambda_i^{v+1} \boldsymbol{Z}_i, \tag{22}$$

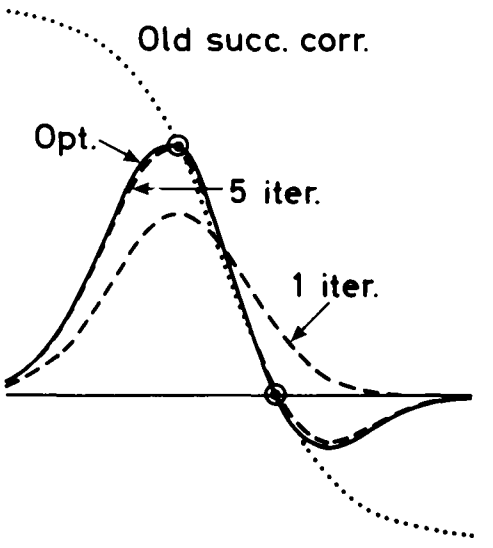

Fig. 1. Analysis of two perfect observations ($\psi_1 = 1$, $\psi_2 = 0$) separated by $\sqrt{2}$. Full line: optimal analysis. Dashed line: proposed scheme. Dotted line: standard successive corrections.

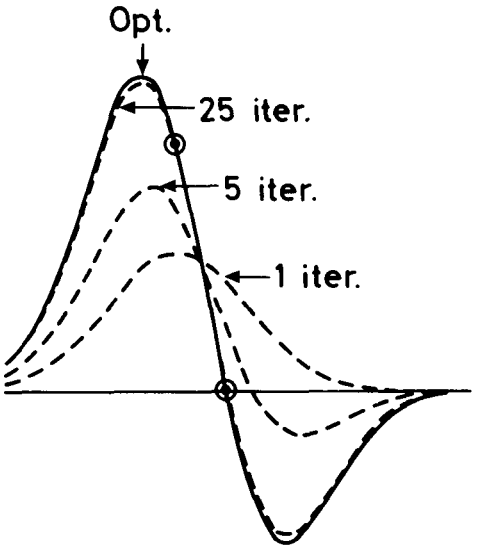

Fig. 2. Analysis of two perfect observations ($\psi_1 = 1$, $\psi_2 = 0$) separated by $\frac{1}{2}\sqrt{2}$. Full line: optimal analysis. Dashed line: proposed scheme.

where $\boldsymbol{Z}_l$ and λ_l represent eigenvectors and corresponding eigenvalues. In our case, the eigenvectors are:

$$\boldsymbol{Z}_M = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}, \quad \boldsymbol{Z}_D = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}, \tag{23}$$

while the eigenvalues are:

$$\lambda_M = 0, \qquad \lambda_D = \frac{2\overline{\psi_1\psi_2}}{1 + \overline{\psi_1\psi_2}} \tag{24}$$

The mode $\boldsymbol{Z}_M$ takes care of the mean value of the observations, while the mode $\boldsymbol{Z}_D$ takes care of the deviations from the mean. Due to the vanishing eigenvalue the mean mode converges after one iteration. The convergence of the deviation mode depends on $\overline{\psi_1\psi_2}$ which may be interpreted as the correlation between the two observations. The eigenvalue λ_D increases monotonically from zero to unity as the correlation increases between the same limits. The procedure therefore never diverges, but the convergence is very slow when the correlation is high. This explains the slow convergence in Fig. 2 where $\overline{\psi_1\psi_2} = 0.8$.

In general, this problem is considerably reduced by the fact that the amplitude of the deviation mode is expected to be relatively small in such cases. High correlation means that the observations should be projected mainly on the mean mode. It can therefore be argued that the situation in Fig. 2 has a low probability of occurring. In real life, however, we will often come across such inconsistencies between meteorological observations and their assumed statistical properties. Instrument errors or transmission errors may for instance produce data which are in strong conflict with neighbouring observations. In such cases the successive correction method may not reach the convergence limit within a reasonable number of iterations. The result will then be a smoothed version of the limiting analysis. For many purposes this is quite acceptable. If it is not, we shall have to reduce the conflict either by modifying the observations or by revising our picture of their statistical properties.

A simple way to modify observations is to reject observations which are too incompatible with their neighbours and therefore likely to be wrong. Such observations are conveniently identified by their inability to be assimilated after a small number of iterations of the analysis scheme.

The "statistics" may be modified by assuming larger observational errors or by reducing the influence radius of the covariance function. In the traditional successive correction analysis it is common practice systematically to reduce the influence radius of the weighting functions from one iteration to the next in order to force smaller and smaller scales to converge. Much can be said in favour of this practice even if it does not fit into the statistical interpolation framework. Ideally a statistical interpolation should be based on statistics computed for a large ensemble of states which fit the observations of the actual situation. In real life this cannot be done, and we have to use a much more general ensemble as our guideline. However, many of the weather systems in which we are interested, are far from average systems with regard to size and structure. If the observations indicate a feature of smaller than average size, it therefore seems quite reasonable to try to adapt the "statistics" to the actual situation by reducing the influence radius in that particular area. The practice of reducing the influence radius from one iteration to the next can take care of this in a nice way. This is so even if the influence radius is uniform in space because an area in which the analysis has already converged, will not be influenced by later iterations

with smaller radii. Using this approach it also seems possible to deal with the breakdown of geostrophic balance for smaller scales by relaxing the cross-correlation between pressure and wind fields as the influence radius is reduced.

In the examples shown so far, we have assumed the observations to be perfect. However, the method outlined above allows us to include information about observation error characteristics. We then have to distinguish between observed and true values, and as shown in (15) the weights applied when the true grid point values are estimated, will be different from the weights used to estimate observations. This will be the case even when the observation point coincides with a grid point. If the observation errors and the first guess errors are uncorrelated, we have:

$$\overline{f_i^o f_j^o} = \overline{f_i^t f_j^t} + \overline{(f_i^o - f_i^t)(f_j^o - f_j^t)},$$
$$\overline{f_i^t f_j^o} = \overline{f_i^t f_j^t}. \qquad (25)$$

In our simple example we shall now assume that

$$\overline{(\psi_i^o - \psi_i^t)(\psi_j^o - \psi_j^t)} = \begin{cases} \frac{1}{2} & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}, \qquad (26)$$

We then have all the information we need in order to use (15). Fig. 3 shows the same case as Fig 2, but with observation error statistics included. As we see the convergence towards the optimum result is now much faster. Note that although the grid point analysis does not fit the observations, the observation estimates (marked by dots) converge towards the observed values and the procedure becomes stationary.

Simple examples of multivariate analysis are shown in Figs 4 and 5. The two observation points, separated by $x_2 - x_1 = \sqrt{2}$, now contain both streamfunction and velocity information. The velocity component normal to the coordinate axis (x-axis) is represented by the slope of the graph of ψ. The slopes corresponding to the velocity observations are indicated by arrows in the figures. In Fig. 4 we have focused on the influence of a streamfunction observation by assuming $\psi_1 = 1$, $\psi_2 = 0$, $v_1 = 0$ and $v_2 = 0$. Fig 5. on the other hand demonstrates the influence of a velocity observation ($\psi_1 = 0$, $\psi_2 = 0$, $v_1 = 1$ and $v_2 = 0$. In both cases the procedure converges towards the desired result. These examples contain more conflicting information than the corresponding univariate case

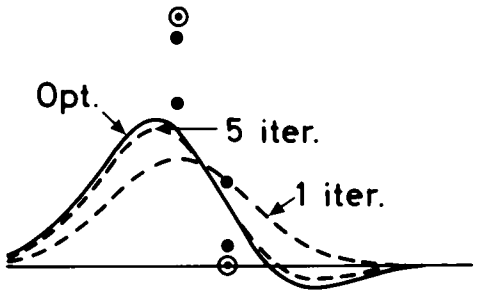

Fig. 3. Analysis of two imperfect observations ($\psi_1 = 1$ $\psi_2 = 0$) separated by $\frac{1}{2}\sqrt{2}$. Full line: optimal analysis. Dashed line: proposed scheme.

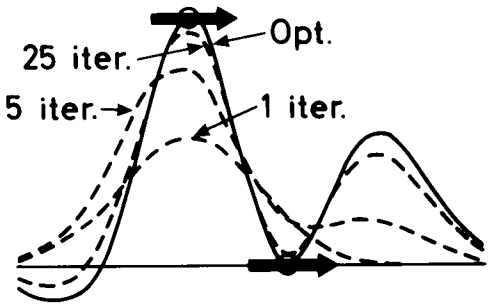

Fig. 4. Analysis of ψ from the perfect observations $\psi_1 = 1$, $\psi_2 = 0$, $v_1 = 0$, $v_2 = 0$. Separation = $\sqrt{2}$. Full line: optimal analysis. Dashed line: proposed scheme.

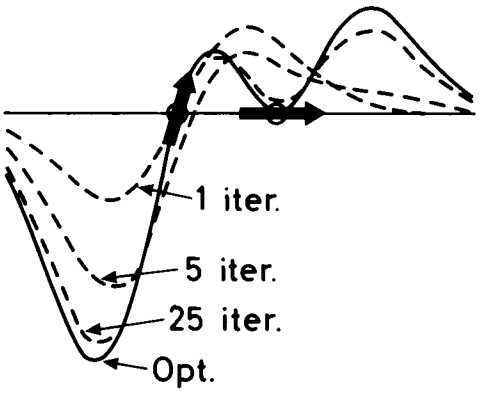

Fig. 5. Analysis of ψ from the perfect observations $\psi_1 = 0$, $\psi_2 = 0$, $v_1 = 1$, $v_2 = 0$. Separation = $\sqrt{2}$. Full line: optimal analysis. Dashed line: proposed scheme.

(Fig. 1), and the convergence is therefore slower. As in the univariate case the convergence can be improved by reducing the radius of influence or by assuming observation errors (not shown).

Within this framework, we have also made a number of experiments with different distributions of a larger number of observations (for instance one hundred observations in two dimensions). These experiments have not revealed any problems beyond what has been discussed in the simple examples above. Clearly, it is difficult to cover all possibilities in experiments like these. A much better test would be to use a larger number of real meteorological data sets. In fact different versions of the method have been used in a numerical weather prediction routine at the Norwegian Weather Service since June 1984. The scheme is multivariate and makes use of observation error statistics. No unexpected problems have been identified, and the general impression is that the method is quite promising (S. Grønaas and H. Midtbø, personal communication). In the present implementation five regular iterations are used, followed by two iterations with a reduced influence radius.

7. Some computational considerations

The computational cost of solving the system of eq. (3) is proportional to the cube of the number of observations allowed to influence a grid point. Most operational statistical interpolation schemes therefore employ severe limitations on this number, with the result that the analysis is not strictly optimal. The cost of the successive correction method is less sensitive to the number of observations involved. In order to keep the number of iterations reasonably low, however, it may be necessary to reduce the radius of influence, or to make other compromises. Thus both approaches have difficulties in producing the optimal solution in data-rich areas. In either case it may be argued that in such areas the analysis is not very sensitive to the method used. Successive correction methods are usually considered to be computationally cheaper than statistical interpolation based on (3), and we believe this to apply to the present method as well. A thorough evaluation of the relative performance of the methods depends on the consequences of the compromises discussed above, and will not be attempted here.

A few comments on the practical implementation of the successive correction method are appropriate. Selection of the observations which are allowed to influence an analysis point requires special algorithms which are often relatively time-consuming. Such algorithms may to a large extent be avoided with successive corrections. Each observation may be allowed to influence all grid points within a certain distance. For each entry of the observation list these grid points are readily identified, and no search or selection is needed at this stage. The method also requires estimates to be made for each of the observations. Usually this is done by interpolation from surrounding grid points. If observation errors are to be accounted for, however, observation estimates should be computed with a separate weight function in accordance with (15a). Fortunately, if errors of different observations are uncorrelated, the only modification of the weights occurs when $i = j$ in (15a). Observation estimates, $F_i^I(v+1)$, interpolated from neighbouring grid points, are then easily corrected for this effect by means of:

$$F_i^{\wedge}(v+1) = F_i^I(v+1) + (\overline{f_i^o f_i^o} - \overline{f_i^I f_i^o}) M_i^{-1} (F_i^o - F_i^{\wedge}(v)). \quad (27)$$

Here we have assumed that the grid is dense enough so that the interpolation can be made without significant errors. In the general case, when observation errors are correlated, it will be necessary to make a separate successive correction interpolation for observations, and some kind of data selection may then be beneficial. This also applies to the computation of the "normalization factor" M_j.

8. Concluding remarks

We have presented a successive correction approach which converges towards a limit which is optimal in a statistical sense. Experiments performed seem to justify the belief that statistical interpolations can be produced quite conveniently by this approach provided that the observations are not too incompatible with the basic statistical assumptions. If they are, the convergence may become too slow to be of practical interest. The result after a reasonable number of iterations will then be a smoothed version of the limiting result.

Often this will be acceptable and even preferable to the exact limit. We also have the option to reconsider the validity of the observations or the statistical assumptions. In particular slow convergence provides a convenient criterion for identification of dubious observations.

Often supplementary information is available which does not fit into the statistical interpolation framework. Examples are complicated constraints like the non-linear balance equation or the requirement of hydrostatic stability. Such constraints may be imposed through iterative procedures. Williamson and Daley (1983) suggest an iteration cycle of alternating statistical interpolation and non-linear normal mode initialization. It seems likely that procedures like this can be made much more efficient by the use of the approach outlined in this paper. In fact the present work was initially motivated by a similar idea based on a dynamical initialization method described in Bratseth(1982). Thus the successive correction approach not only has the ability to imitate statistical interpolations, but also has an interesting potential beyond the statistical interpolation framework.

9. Acknowledgements

We wish to thank Sigbjørn Grønås and Helge Midtbø at the Norwegian Weather Service for many useful discussions in connection with this work, and we are also grateful to Andrew Lorenc who provided important encouragement for the convergence proof in the appendix.

10. Appendix. Convergence proof.

Eq. (15) implies that the matrix $\boldsymbol{A}$ is of the form

$$\boldsymbol{A} = \boldsymbol{C} \cdot \boldsymbol{D}, \tag{A1}$$

where $\boldsymbol{C}$ is the symmetric matrix $\{\overline{f_i^o f_j^o}\}$ and $\boldsymbol{D}$ is the diagonal matrix $\{M_j^{-1} \partial_{ij}\}$. Now assume that x is an eigenvector of $\boldsymbol{A}$ corresponding to the eigenvalue μ so that:

$$\boldsymbol{A} \cdot \boldsymbol{x} = \boldsymbol{C} \cdot \boldsymbol{D} \cdot \boldsymbol{x} = \mu \boldsymbol{x}. \tag{A2}$$

Multiplying by $\boldsymbol{D}^{1/2}$ from the left we get:

$$(\boldsymbol{D}^{1/2} \cdot \boldsymbol{C} \cdot \boldsymbol{D}^{1/2}) \cdot (\boldsymbol{D}^{1/2} \cdot \boldsymbol{x}) = \mu (\boldsymbol{D}^{1/2} \cdot \boldsymbol{x}). \tag{A3}$$

The matrix $\boldsymbol{D}^{1/2} \cdot \boldsymbol{C} \cdot \boldsymbol{D}^{1/2}$ is symmetric and it then follows that the eigenvalues μ_k are real and that the eigenvectors $\boldsymbol{D}^{1/2} \cdot \boldsymbol{x}_k$ form an orthogonal basis. If these vectors are normalized, the observation vector may be expanded as:

$$\boldsymbol{f} = \boldsymbol{F}^o - \boldsymbol{F}^p = \sum_k (\boldsymbol{x}_k^t \cdot \boldsymbol{D} \cdot \boldsymbol{f}) \boldsymbol{x}_k. \tag{A4}$$

When this is introduced in (7) and (8), we realize that the iteration procedure will converge if the eigenvalues of the matrix $\boldsymbol{I} - \boldsymbol{A}$ are smaller than one in modulus, and since the eigenvalues of $\boldsymbol{A}$ are shown to be real, they must satisfy:

$$0 < \mu < 2. \tag{A5}$$

If (A3) is multiplied from the left by $(\boldsymbol{D}^{1/2} \cdot \boldsymbol{x})^t$ we get:

$$\mu = \boldsymbol{x}^t \cdot \boldsymbol{D} \cdot \boldsymbol{C} \cdot \boldsymbol{D} \cdot \boldsymbol{x} = \boldsymbol{x}^t \cdot \boldsymbol{D} \cdot \overline{\boldsymbol{f}\boldsymbol{f}^t} \cdot \boldsymbol{D} \cdot \boldsymbol{x}$$
$$= \overline{(\boldsymbol{x}^t \cdot \boldsymbol{D} \cdot \boldsymbol{f})^2} \geqslant 0, \tag{A6}$$

showing that μ is never negative. Furthermore this result shows that eigenvectors with vanishing eigenvalues do not contribute to the expansion in (A4). The lower limit in (A5) is therefore always satisfied. From (A6) we also conclude that vanishing eigenvalues imply the existence of vanishing linear combinations of observations. Thus convergence problems due to the lower limit in (A5) are always connected with redundant information in the observation vector.

Now we shall turn to the right-hand inequality in (A5). In order to simplify the discussion we shall assume that f has been normalized so that $\overline{f_i^o f_i^o} = 1$, and we shall also drop the superscript O. If we define $\boldsymbol{y} = \{y_i\} = \boldsymbol{D} \cdot \boldsymbol{x}$ then the i'th component of (A2) may be written as:

$$\sum_j \overline{f_i f_j} y_j = \mu M_i y_i. \tag{A7}$$

Here i may be chosen so that $|y_j| \leqslant |y_i|$ for all j. It is then easy to show that:

$$|\mu M_i - 1| \leqslant \sum_{j \neq i} |\overline{f_i f_j}| |y_j| / |y_i$$
$$\leqslant \sum_{i \neq j} |\overline{f_i f_j}|. \tag{A8}$$

A similar result is known in the mathematical literature as Gerschgorin's theorem. If all the f's are scalars, (16) and (18) give:

$$M_i = \sum_j |\overline{f_i f_j}| \tag{A9}$$

and (A8) is reduced to:

$$|\mu M_i - 1| \leqslant M_i - 1 \tag{A10}$$

which implies that $\mu \leqslant 1$ so that (A5) is satisfied. This conclusion is also valid when the variables are not normalized. To see this it is sufficient to repeat the arguments above with $|\overline{f_j^2}|^{1/2} y_j$ playing the role of y_j.

If (A9) is used for vector components, M_j will depend on the orientation of the coordinate system. This arbitrariness is avoided by using the invariant formulas suggested in (19). In fact the argument given above is easily extended to cover these cases. Using the notation of Section 5, we realize that some of the components of (A7) may be written as:

$$\sum_j \overline{\psi_i \psi_j} y_j + \sum_k \overline{\psi_i v_k} \cdot z_k = \mu M_i y_i \tag{A11}$$

with

$$M_i = \sum_j |\overline{\psi_i \psi_j}| + \sum_k |\overline{\psi_i v_k}|,$$

while the rest of the components may be organized as vector equations of the form:

$$\sum_j \overline{v_i \psi_j} y_j + \sum_k \overline{v_i v_k} \cdot z_k = \mu M_i z_i, \tag{A12}$$

with

$$M_i = \{\sum_j |\overline{v_i \psi_j}| + \sum_k (|\overline{u_i v_k}|^2 + |\overline{v_i v_k}|^2)^{1/2}\} 2^{-1/2}.$$

Here some of the components of y have been organized as two-dimensional vectors z. At least one of the quantities $|y|$ and $|z|$ has the property that it is not smaller than any of the others. If $|y_i|$ satisfies this condition,, (A11) immediately gives:

$$|\mu M_i - 1| \leqslant \sum_{j \neq i} |\overline{\psi_i \psi_j}| + \sum_k |\overline{\psi_i v_k}| = M_i - 1, \tag{A13}$$

which implies that $\mu \leqslant 1$. On the other hand, if $|z_i|$ is the dominating modulus then (A12) gives:

$$|\mu M_i - 1| \leqslant \sum_j |\overline{v_i \psi_j}| + \sum_{k \neq i} (|\overline{u_i v_k}|^2 + |v_i v_k|^2)^{1/2} = 2^{1/2}(M_i - 1) \tag{A14}$$

which implies that $\mu \leqslant 2^{1/2} + (1 - 2^{1/2}) M_i^{-1} < 2^{1/2}$. In either case, (A5) is satisfied with a wide margin.

REFERENCES

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

Bratseth, A. M. 1982. A simple and efficient approach to the initialization of weather prediction models. *Tellus 34*, 352–357.

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

Eliassen, A. 1954. Provisional report on calculation of spatial covariance and autocorrelation of the pressure field. Inst. Weather and Climate Res., Acad. Sci. Oslo, Rep. No. 5. Reprinted in: *Data assimilation methods*. Ed. L. Bengtsson, M. Ghil, E. Källén), Springer-Verlag, 319–330.

Franke, R. and Gordon, W. J. 1983. The structure of optimum interpolation functions. *Tech. Rep. NPS-53-83-0005*, Naval Postgraduate School. Monterey, CA, 23 pp.

Gandin, L.S. 1963. *Objective analysis of meteorological fields*. Leningrad. Hydromet. Press. Translated from Russian by Israel Program for Scientific Translations. Jerusalem 1965.

Gustafsson, N. 1981. A review of methods for objective analysis. *Data assimilation methods*. (Ed. L. Bengtsson, M. Ghil, E. Källén), Springer-Verlag, 17–76.

Williamson, D. L. and Daley, R. 1983. A unified analysis-initialization technique. *Mon. Wea. Rev. 111*, 1517–1536.