U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Weather Service
National Centers for Environmental Prediction
5200 Auth Road
Camp Springs, MD 20746-4304

Office Note 422

STRUCTURAL PRECONDITIONING IN OPTIMAL ANALYSIS

R. J. Purser *
Environmental Modeling Center

January 1998

* General Sciences Corporation, Laurel, Maryland; e-mail: wd23jp@sun1.wwb.noaa.gov

1. Introduction

Variational data assimilation in numerical weather prediction (NWP) encompasses a wide variety of techniques, from the early pioneering methods of Sasaki (1958) and Gandin (1963) to the important recent developments of optimal control theory employing the forecasting model itself as a strong constraint (Talagrand, 1981a,b; Lewis and Derber, 1985; Thacker and Long, 1988) or weak constraint (Ghil et al., 1981; Bennett and Miller, 1991; Bennett et al., 1996, 1997). As the techniques grow in sophistication and as the data types become more diverse, the numerical problems associated with the necessary large scale linear or nonlinear inversions become correspondingly more severe.

In typical current operational data assimilation schemes the rank of the linear operator of analysis weights, which is essentially the number of distinct data, is at least $\mathcal{O}(10^4)$. On this scale of linear inversion, direct methods are out of the question. Therefore one is forced to resort to iterative methods. The standard methods for large scale linear inversion of a system whose "system matrix" can be put into symmetric (or "self-adjoint") form are the various conjugate gradient (CG) and quasi-Newton (QN) methods, as discussed in Gill et al. (1981). For reasons that will emerge in our discussions below, we will append to this catalog of viable methods the extremely simple method of Chebyshev accelerated iterations, which does not require the system matrix to be symmetric but, when it is, can be regarded as an efficient refinement of the method of steepest descent. For three-dimensional (3-D) variational analysis one iteration by any of the aforementioned methods involves costly spatial convolutions with a modelled covariance function (or its inverse "precision" operator) but, for four-dimensional (4-D) variational assimilation this cost is greatly increased by the additional requirement to perform one forward model integration plus one backward "adjoint model" integration (not necessarily in this order) covering the duration of the assimilation period. It is therefore crucially important to organize the structure of the analysis algorithm so that the total number of such iterations can be kept to the minimum consistent with achieving an adequate approximation to the optimal solution. For 3-D analysis the practical maximum number of iterations is about 50 (Parrish and Derber, 1992); before 4-D variational analysis can become a practical reality it is probably necessary for this number to be reduced to $\mathcal{O}(10)$ iterations.

The task of transforming the system matrix of the linearized problem so that the iterative methods have an accelerated rate of convergence is the so-called "preconditioning" problem. This is dealt with classically by procuring an operator which, while being relatively cheap to generate and apply, acts as a crude inverse of the actual system matrix for the given problem. Unfortunately, this classical approach to preconditioning is extremely difficult to apply effectively in the problems of variational analysis, owing to the inherent complexity of the structure of the system matrix. In fact, the existence of the system matrix is usually only implied, since what is available in practice is typically only a chain of more basic linear operators that act upon vectors and which are applied recursively to produce the final product of system matrix and input vector. The best practical preconditioner of this classical type might typically leave the effective system matrix of the problem with a condition number as large as 10^3 or greater for even a 3-D global assimilation (D. F. Parrish, personal communication). This note proposes a more radical alternative approach, "structural preconditioning", in an effort to transform the inversion problem into a form that can be successfully tackled by an algorithm requiring only

a modest expenditure of the costly iterations.

The essence of the method proposed here is to substitute for the intractably ill-conditioned problem a recursive succession of very well-conditioned problems which provide a sequence of easily computed approximately optimal states converging in the limit to the optimal analysis. Each approximation to the optimal analysis can be regarded as a direct refinement of its predecessor. Provided each sub-problem is solved to within some sufficient tolerance, then the outer-iteration giving rise to the sequence of approximate analyses is shown below to converge (to within a similar tolerance) to the true optimum. By this method, it appears that the otherwise prohibitive cost of achieving a 4-D operational optimal analysis incorporating the constraints of model dynamics might be brought down to an affordable level. Moreover, the theoretical examination of iterative methods we present below suggests that subtle deficiencies of current variational analysis "solutions", which the usual cost-function or gradient-norm diagnostics tend to obscure, can probably be aleviated by the proposed method to provide a more faithful overall representation of the intended optimal analysis than can be attained using the current iterative strategies.

The theoretical methods we use here to examine the convergence behavior of these iterative schemes are based on orthogonal polynomial theory, especially as it relates to Chebyshev polynomials. The application of related methods is reviewed in the recent text by Fischer (1996).

2. Variational analysis

(*a*) *Notation and Definitions*

Where possible, we adopt the following syntactic conventions:

i) Lower-case variables denote (column-) vectors.

ii) Upper-case variables denote operators, script if nonlinear, bold if linear.

iii) "Hatted" variables, e.g., "$\hat{f}$", denote quantities existing in observation-space.

iv) The adjoint of a vector or operator is signified by an asterisk.

v) Angle brackets denote the expectation.

vi) Primes denote error components.

For 4-D analysis, we consider a "period of assimilation", perhaps of about 24 hours duration, commencing at "initial-time" and ending at "data cut-off time". However, it is legitimate to consider the 4-D analysis and background states to extend continuously and indefinitely beyond data cut-off, whereupon both become forecasts.

It is convenient, using the above conventions, to define the main vectors involved in the theory of variational analysis as follows.

a, the analysis state; a', its error component; $\hat{a}$, the "measured" analysis.

b, the background; b', the background error; $\hat{b}$, the "measured" background.

g, (in 4-D analysis only) the concatenation of (i) initial-time instantaneous analysis state and (ii) the dynamical forcing terms for the model's prognostic equations needed to fit the subsequent evolution of the analysis during and after the assimilation period.

h, (in 4-D analysis only) vector with the same structure as g but initiating and guiding the evolution of the background b instead of analysis a; h', the error components of h.

$\hat{e}$, the measurements; $\hat{e}'$, the measurement errors.

$\hat{f}$, the analysis-forcing vector or "representer coefficients" in the terminology of Bennett et al. (1996, 1997).

v, the vector of adjoint model variables for the 4-D variational analysis.

$\hat{r}$, "residuals" vector of the defining equation for the optimal analysis forcing $\hat{f}$.

$\hat{q}$, measurement increments, $\hat{q} = \hat{e} - \hat{b}$.

We define the array of (nonlinear) measurement functionals, forming a mapping from model space to observation space, by the script symbol "$\mathcal{K}$". Thus, the measured analysis $\hat{a}$ is related to a through this mapping:

$$\hat{a} = \mathcal{K}(a),$$

while infinitesimal increments are connected through the Jacobian $\mathbf{K}$ of $\mathcal{K}$:

$$d\hat{a} = \mathbf{K}\, da.$$

Another mapping, employed in 4-D analysis, is what we shall refer to as the "prognostic" operator $\mathcal{P}$. It is this nonlinear operator, representing the initiation and subsequent forcing of the model by the argument of $\mathcal{P}$, that connects a and g, b and h as follows.

$$a = \mathcal{P}(g),$$
$$b = \mathcal{P}(h),$$

The Jacobian, $\mathbf{P}$, of $\mathcal{P}$ expresses the sensitivity of a to g in the neighborhood of the particular evolution a:

$$da = \mathbf{P}\, dg,$$

Note that, unlike $\mathbf{K}$, the operator $\mathbf{P}$, and consequently the mapping $\mathcal{P}$, are always invertible.

Finally, we introduce here the definitions of the principal covariances. We shall assume background and measurements to be unbiased and both vectors' errors to be mutually uncorrelated:

$$\langle b' \rangle = 0, \quad \langle \hat{e} \rangle = 0, \quad \langle \hat{e} b'^{*} \rangle = 0,$$

For 4-D analysis, the error h' of the model forcing associated with the background has a covariance denoted by

$$\mathbf{H} \equiv \langle h' h'^{*} \rangle,$$

which implies that the corresponding covariance $\mathbf{B}$ of the error of the background state itself is given, exactly for linear dynamics, approximately for nonlinear dynamics, by:

$$\mathbf{B} \equiv \langle b' b'^{*} \rangle \cong \mathbf{P}\mathbf{H}\mathbf{P}^{*},$$

in which $\mathbf{P}^{*}$ signifies the (linear) adjoint-model integration operator. The covariance $\hat{\mathbf{B}}$ of error of the measured background $\hat{b}$, which is referred to by Bennett et al. (1996) as the "representer

matrix", has a corresponding approximate form (exact relative to $\mathbf{B}$ when $\mathcal{K}$ is linear) in terms of $\mathbf{B}$ and $\mathbf{K}$:

$$\hat{\mathbf{B}} \equiv \langle \hat{b}' \hat{b}'^{*} \rangle \cong \mathbf{K}\mathbf{B}\mathbf{K}^{*}.$$

Measurement errors are assumed to be either uncorrelated or correlated only within small groups of related measurements made with the same instrument. Therefore their covariance operator, which we denote by $\hat{\mathbf{E}}$, is either diagonal or block-diagonal with blocks of small size (and therefore trivial to invert):

$$\hat{\mathbf{E}} \equiv \langle \hat{e}' \hat{e}'^{*} \rangle,$$

It is also convenient to supply a symbol for the sum of $\hat{\mathbf{E}}$ and $\hat{\mathbf{B}}$. Since this is just the covariance of the measurement increments $\hat{q}$, we naturally call this covariance,

$$\hat{\mathbf{Q}} = \langle \hat{q} \hat{q}^{*} \rangle = \hat{\mathbf{E}} + \hat{\mathbf{B}}.$$

(b) *Theory of optimal analysis*

In its most general (Bayesian) form, optimal analysis seeks the analysis a which minimizes the "expected loss", computed by integrating some prescribed loss-function with respect to the (*a posteriori*) probability measure associated with the distribution of analysis error a' implied by the choice of a**.** For general probability distributions of the (*a priori*) background errors and measurement errors, the resulting variational principle can be very complicated and implies a nonlinear system of equations for the extrema for the expected loss without guarantees of uniqueness of the possible (local) solutions. In practice, the problem is greatly simplified by quality-controlling the data to remove "outliers" of the measurement error components, choosing a simple idealized loss-model and by then assuming that the various probability distributions are "normal", or at least sufficiently close to being "normal" that they imply a vector equation for the optimal state which, if not exactly linear is, at worst, only weakly nonlinear. The "maximum posterior-probability" (or "penalized maximum-likelihood") solution is perhaps the simplest idealization of this approach and can be expressed as the problem of determining the minimizing state a for the quadratic cost function $\mathcal{L}(a)$ defined (for 3-D analysis) by,

$$\mathcal{L}(a) = \Big\{ (a-b)^{*} \mathbf{B}^{-1} (a-b) + (\hat{e}-\hat{a})^{*} \hat{\mathbf{E}}^{-1} (\hat{e}-\hat{a}) \Big\}, \tag{2.1}$$

whose solution must satisfy

$$a = b + \mathbf{B}\mathbf{K}^{*} \hat{f}, \tag{2.2a}$$

where

$$\hat{\mathbf{E}} \hat{f} = \hat{e} - \hat{a}, \tag{2.2b}$$

and hence, for linear or only weakly nonlinear measurement functionals $\mathcal{K}$,

$$\hat{\mathbf{Q}} \hat{f} \simeq \hat{q} \equiv \hat{e} - \hat{b}. \tag{2.2c}$$

It is worth noting that, with a liberal and adaptive interpretation of the operators $\mathbf{B}$ and $\hat{\mathbf{E}}$ appearing in (2.2a) and (2.2b), these equations can be shown to identify the optimal solutions

for slightly more general maximum-posterior-probability models than can be described by the simple quadratic form (2.1)). Thus (2.2a) and (2.2b) express a fairly general prescription for balancing the estimated error of the background against the estimated errors of the measurements.

For the case of the 4-D variational analysis of Bennett et al. (1996), the nonlinearities inherent in the model dynamics are treated more naturally by defining the cost function to be quadratic in model forcing increments $(g-h)$ instead of model state increments $(a-b)$. Then even a simple covariance model for $\mathbf{H}$ gives rise to an intricately and convincingly structured implied covariance $\mathbf{B}$ through the cumulative effects of the model integration. It is not possible in practice to prescribe such a structured $\mathbf{B}$ by direct means without the explicit involvement of the forecast model. For 4-D variational analysis, we therefore start from the cost function:

$$\mathcal{L}(g) = \Big\{(g-h)^*\mathbf{H}^{-1}(g-h) + (\hat{e}-\hat{a})^*\hat{\mathbf{E}}^{-1}(\hat{e}-\hat{a})\Big\}. \tag{2.3}$$

In this case the formalism leads to a g obtained by incrementing the background model forcing h by a distribution formed by convolving an adjoint model variable with covariance $\mathbf{H}$:

$$g = h + \mathbf{H}v, \tag{2.4a}$$

where v obeys the inhomogeneous adjoint-model equations (which are integrated backward in time):

$$v = \mathbf{P}^* f, \tag{2.4b}$$

with

$$f \equiv \mathbf{K}^* \hat{f}, \tag{2.4c}$$

and, as before,

$$\hat{\mathbf{E}}\hat{f} = \hat{e} - \hat{a}, \tag{2.4d}$$

$$\hat{a} = \mathcal{K}(a), \tag{2.4e}$$

but where a is now prescribed indirectly as the integration forward of the nonlinear model forced by the vector g:

$$a = \mathcal{P}(g). \tag{2.4f}$$

Notice that, for linearized dynamics and measurement functional, the measured-background error covariance implied by this formalism is

$$\hat{\mathbf{B}} = \mathbf{KPHP}^*\mathbf{K}^* \tag{2.5}$$

(*c*) *Strategy for solution*

An iterative algorithm suggested by (2.2a)–(2.2c) in the 3-D case is to start with an initial guess $\hat{f}^{(0)}$ for $\hat{f}$ and set the current iteration index n to zero. Then proceed through the following steps until a sufficiently small residual $\hat{r}^{(n)}$ is obtained.

(i) Form the analysis at iteration n corresponding to $\hat{f}^{(n)}$

$$a^{(n)} = b + \mathbf{B}\mathbf{K}^*\hat{f}^{(n)}; \tag{2.6a}$$

(ii) "Measure" $a^{(n)}$ to evaluate $\hat{a}^{(n)}$ and form the residual $\hat{r}^{(n)}$:

$$\hat{r}^{(n)} = \hat{\mathbf{E}}\hat{f}^{(n)} - (\hat{e} - \hat{a}^{(n)}); \tag{2.6b}$$

(iii) Refine the estimate for the analysis forcing $\hat{f}$ by applying a "Newton iteration" aimed at nullifying the residual:

$$\hat{f}^{(n+1)} = \hat{f}^{(n)} - (\hat{\mathbf{B}} + \hat{\mathbf{E}})^{-1}\hat{r}^{(n)}, \tag{2.6c}$$

where we are using the fact that

$$\frac{\partial \hat{r}}{\partial \hat{f}} = (\hat{\mathbf{B}} + \hat{\mathbf{E}}) \equiv \hat{\mathbf{Q}}. \tag{2.7}$$

As discussed in Bennett et al. (1996), essentially the same iterative strategy should work in principle for the 4-D example, except that evaluating the analysis implied by forcing vector $\hat{f}$ at each stage involves the additional complexity of the two model integrations as described in (2.4a)–(2.4f). The obstacle to direct implementation of this strategy in practice stems from the difficulty of managing the large scale linear inversion implied by each Newton iteration [stage (iii) of the algorithm]. A direct method is impractical here (even the elements of $\hat{\mathbf{Q}}$ remain unknown in practice since this operator is only easily applied implicitly through a sequence of simpler operations). In the absence of a better method, one would conventionally adopt the alternative, iterative approach and employ one of the popular variants of CG or QN method, many of which are described in Gill et al. (1981) (see also Navon and Legler, 1987).

In addition to $\hat{\mathbf{Q}}$ being a difficult operator for direct linear inversion, it is also very difficult to effectively precondition it for use in an iterative inversion - no close approximation to its inverse can be obtained at reasonable cost. We are left with the woefully inadequate matrix $\hat{\mathbf{E}}$ to serve as preconditioner. Even then, the range of eigenvalues of the preconditioned system matrix is between one and (typically) around 1000 or more. To understand what features of the large scale inversion problem influences the performance of iterative solution strategies, we delve into some (slightly idealized) theoretical aspects of iterative solution methods as seen in the context of problems whose degrees of freedom vastly outnumber the iterations one could ever afford in practice.

3. Convergence behavior of methods for inverting large-scale linear systems

(*a*) *Formal structure of conjugate-gradient and accelerated descent algorithms*

From (2.6c), the problem we need to solve is equivalent to finding, for some given vector, $\hat{q}$, the vector $\hat{f}$ such that:

$$\hat{\mathbf{Q}}\hat{f} - \hat{q} = 0, \tag{3.1}$$

We take the preconditioning matrix to be $\hat{\mathbf{E}}$. The prototype for iterative solution algorithms is the linear preconditioned conjugate gradient (CG) algorithm (Gill et al., 1981). The many minor variants of the CG and QN methods lead to the same sequence of iterates in these linear cases. For large-scale problems, it is not feasible to compute the Hessian directly, so we do not consider such "Newton" methods here. The CG method requires that, at each iteration k, a mixing parameter β_k and step-size parameter α_k be computed, together with a vector $\hat{r}_k$ of residual gradient components and a "conjugate-direction" vector $\hat{p}_k$. Initially, $\hat{r}_0 = \hat{\mathbf{Q}}\hat{f}_0 - \hat{q}$,

$\beta_{-1}=0,\ \hat{p}_{-1}=0$. The algorithm proceeds for $k=0,1,\ldots,$ and is terminated when some suitable convergence criterion for the residual $\hat{r}_k$ is met, or when the pre-set maximum number of iterations has been expended.

$$\hat{p}_k = -\hat{\mathbf{E}}^{-1}\hat{r}_k + \hat{p}_{k-1}\beta_{k-1}, \tag{3.2a}$$

$$\alpha_k = \frac{\hat{r}_k^*\hat{\mathbf{E}}^{-1}\hat{r}_k}{\hat{p}_k^*\hat{\mathbf{Q}}\hat{p}_k}, \tag{3.2b}$$

$$\hat{f}_{k+1} = \hat{f}_k + \hat{p}_k\alpha_k, \tag{3.2c}$$

$$\hat{r}_{k+1} = \hat{r}_k + \hat{\mathbf{Q}}\hat{p}_k\alpha_k, \tag{3.2d}$$

$$\beta_k = \frac{\hat{r}_{k+1}^*\hat{\mathbf{E}}^{-1}\hat{r}_{k+1}}{\hat{r}_k^*\hat{\mathbf{E}}^{-1}\hat{r}_k}. \tag{3.2e}$$

Note that the convergence behavior of (3.2a)–(3.2e) can be analyzed without explicit reference to $\hat{f}_k$ [which appears only in (3.2c)], provided the initial residual $\hat{r}_0$ is given.

Contrast the sophisticated CG algorithm above with the simple preconditioned "accelerated descent" (AD) method:

$$\hat{r}_k = \hat{\mathbf{Q}}\hat{f}_k - \hat{q}, \tag{3.3a}$$

$$\hat{f}_{k+1} = \hat{f}_k - \frac{\hat{\mathbf{E}}^{-1}\hat{r}_k}{x_{k+1}}. \tag{3.3b}$$

in which x_k acts as a step-size parameter at iteration k. Again, we may eliminate reference to $\hat{f}_k$:

$$\hat{r}_{k+1} = \left(\hat{\mathbf{I}} - \frac{\hat{\mathbf{Q}}\hat{\mathbf{E}}^{-1}}{x_{k+1}}\right)\hat{r}_k, \tag{3.3c}$$

and hence, solve explicitly in this case:

$$\hat{r}_k = \prod_{j=1}^{k}\left(\hat{\mathbf{I}} - \frac{\hat{\mathbf{Q}}\hat{\mathbf{E}}^{-1}}{x_j}\right)\hat{r}_0. \tag{3.3d}$$

As evident from Gill et al. (1981), the CG algorithm generates a sequence of residuals orthogonal in the sense,

$$\hat{r}_k^*\hat{\mathbf{E}}^{-1}\hat{r}_j = 0, \quad j\neq k, \tag{3.4}$$

However, the less sophisticated AD method leads to no such properties among *its* successive residuals.

To clarify the consequences of these algorithms, we let $\mathbf{V}$ be the matrix of right-eigenvectors of $(\hat{\mathbf{Q}}\hat{\mathbf{E}}^{-1})$, $\mathbf{U}$ the right-eigenvectors of $(\hat{\mathbf{E}}^{-1}\hat{\mathbf{Q}})$, with $\mathbf{X}$ the associated diagonal matrix of real positive eigenvalues x:

$$(\hat{\mathbf{Q}}\hat{\mathbf{E}}^{-1})\mathbf{V} = \mathbf{V}\mathbf{X}, \tag{3.5a}$$

$$(\hat{\mathbf{E}}^{-1}\hat{\mathbf{Q}})\mathbf{U} = \mathbf{U}\mathbf{X}. \tag{3.5b}$$

The eigenvectors are normalized relative to the "metric" $\hat{\mathbf{E}}^{-1}$ in the sense:

$$\mathbf{V}^*\hat{\mathbf{E}}^{-1}\mathbf{V} \equiv \mathbf{V}^*\mathbf{U} \equiv \mathbf{U}^*\hat{\mathbf{E}}\mathbf{U} = \mathbf{I}. \tag{3.5c}$$

With this convention, the "characteristic modes" of the respective iterative schemes can be studied separately. The modes of residuals $\hat{r}_k$ are expressed in terms of eigenvectors $\mathbf{V}$, those of $\hat{p}_k$ in terms of eigenvectors $\mathbf{U}$:

$$\hat{r}_k = \mathbf{V} r_k, \tag{3.6a}$$

$$\hat{p}_k = \mathbf{U} p_k. \tag{3.6b}$$

With substitutions (3.6a)–(3.6b), cycles similar to (3.2a)–(3.2e) and (3.3a)–(3.3c) describe the corresponding evolution of coefficient vectors r_k and p_k , except for the simplification that $\hat{\mathbf{Q}}$ is replaced by the diagonal operator $\mathbf{X}$, while $\hat{\mathbf{E}}^{-1}$ is replaced by the identity operator. But, in addition to this simplification, diagonalization of the system operator reveals a connection between successive iterates r_k of the CG method and the successive polynomials defined to be orthogonal relative to some particular weight distribution. To see this, first observe that r_k in either the CG or AD iterations relates to the initial residual vector r_0 through some real-valued polynomial $\tilde{r}_k$ of degree k in the system operator, $\mathbf{X}$, and, for the CG method, p_k also relates to r_0 through some other polynomial of degree k:

$$r_k = \tilde{r}_k(\mathbf{X}) r_0, \tag{3.7a}$$

$$p_k = \tilde{p}_k(\mathbf{X}) r_0, \tag{3.7b}$$

where the constant term of polynomial $\tilde{r}_k$ is unity. Orthogonality condition (3.4) may now be interpreted as a statement of the orthogonality of distinct pairs of polynomials $\tilde{r}_k$ in the weighted integral sense:

$$\int_{\mathcal{S}} \tilde{r}_k(x)\, \tilde{r}_j(x)\, w_0(x)\, dx = 0, \quad j \neq k, \tag{3.8}$$

where we define, for each k, "weight" profile $w_k(x)$ to be a generalized function of x with support within the interval $\mathcal{S} \equiv [x^-, x^+]$ marking the range of eigenvalues of X, and with w_k defined by,

$$\int_{x^-}^{x} w_k(x')dx' = \sum_{x' \leq x} r_k^2(x'). \tag{3.9}$$

In this sense, the CG algorithm is just a recurrence relation for the generation of the polynomials $\tilde{r}_k$, of unit constant term, orthogonal relative to w_0:

$$\tilde{p}_k(x) = -\tilde{r}_k(x) + \tilde{p}_{k-1}(x)\beta_{k-1}, \tag{3.10a}$$

$$\alpha_k = \frac{\int \tilde{r}_k^2(x) w_0(x)\, dx}{\int \tilde{p}_k^2(x) x\, w_0(x)\, dx}, \tag{3.10b}$$

$$\tilde{r}_{k+1}(x) = \tilde{r}_k(x) + \tilde{p}_k(x)\, x\, \alpha_k, \tag{3.10c}$$

$$\beta_k = \frac{\int \tilde{r}_{k+1}^2(x)\, w_0(x)\, dx}{\int \tilde{r}_k^2(x)\, w_0(x)\, dx} \equiv \frac{\int w_{k+1}(x)\, dx}{\int w_k(x)\, dx}, \tag{3.10d}$$

Consistent with the integral definitions of the inner-product (3.8), a cost-function for which (3.10a)–(3.10d) defines the CG solution method is the quadratic functional given by:

$$\mathcal{C} = \int_{\mathcal{S}} \frac{\tilde{r}_k^2(x) w_0(x)}{x}\, dx \equiv \int_{\mathcal{S}} \frac{w_k(x)}{x}\, dx. \tag{3.11}$$

The square-root of $\mathcal{C}$ at any stage of an iterative inversion defines one particularly useful measure of the "distance" remaining to the true solution. Although other measures are also admissible, we shall exclusively refer to this measure when we speak of the norm of the residuals. The *log-norm* is then defined at iteration k as the quantity,

$$\gamma_k \equiv \log(\mathcal{C}^{1/2}). \tag{3.12}$$

For the CG algorithm, γ_k decreases monotonically with iteration index k, but this is not generally true for AD methods.

(*b*) *Theorem 1.*

If, at any given stage k of the CG method seeking the minimum of a quadratic functional, a non-trivial residual remains, then: (i) there exists an AD method for the same problem yielding an identical approximate solution at iteration k; (ii) there exists no other AD method for this problem that yields a smaller residual norm at iteration k.

(*c*) *Proof of theorem 1*

(i) From a general property of orthogonal polynomials we infer that the k zeroes of $\tilde{r}_k$ for the CG method are real and distinct and lie on the support of w_0 . By taking the first k step-size parameters x_j of method (3.3d) to be these zeroes, the same polynomial $\tilde{r}_k(x)$ is obtained for the descent method (3.3a)–(3.3d) as for the CG method (3.2a)–(3.2e).

(ii) By considering variations of $\mathcal{C}$ with respect to the locations x_j of the k zeroes of a polynomial $\tilde{r}_k$ of degree k with unit constant term, we find that $\mathcal{C}$ is minimized only if $\tilde{r}_k$ is orthogonal, relative to weight w_0 , to all polynomials of degree less than k. Such a polynomial is uniquely the $\tilde{r}_k$ associated with the CG method at iteration k.

(*d*) *Theorem 2*

If the log-norm for the residual at each iteration k is γ_k and the *condition number* for the problem is $c = x^+/x^-$, then:

(i) for each iteration k, there is an upper bound γ_k^+ given by

$$\gamma_k^+ = \gamma_0 - \log(\cosh k\phi), \tag{3.13a}$$

where

$$\phi = \log\left(\frac{1+c^{-1/2}}{1-c^{-1/2}}\right) \equiv \operatorname{arccosh}\left(\frac{c+1}{c-1}\right). \tag{3.13b}$$

such that,

$$\gamma_k \leq \gamma_k^+ \tag{3.13c}$$

(ii) for given k there exists a problem for which the CG method only manages to attain this bound:

$$\gamma_k = \gamma_k^+. \tag{3.13d}$$

(*e*) *Remarks*

When the eigenvalues $x \in \mathcal{S} \equiv [x^-, x^+]$ are so numerous and are distributed so that they may be considered effectively continuous throughout this interval, we may complement theorem 2 (for the upper bound on γ_k) by the following theorem (for a lower bound on γ_k):

(*f*) *Theorem 3*

When the initial residual weight profile w_0 can be regarded as positive over $\mathcal{S}$, there exists a constant, $\bar{\gamma}_0$, that depends only on w_0 , such that no method of the type (3.3a)–(3.3b) yields a log-norm at iteration k smaller than,

$$\bar{\gamma}_k = \bar{\gamma}_0 - k \log(\phi), \tag{3.14}$$

with ϕ defined by (3.13b).

(*g*) *Remarks*

The proofs of theorems 2 and 3, which rely heavily on the properties of Chebyshev polynomials, are developed in the appendix. For asymptotically large k, graphs of the bounds γ^+ and $\bar{\gamma}_k$, plotted against k itself, become parallel with the same constant slope. The Chebyshev polynomials (of the first kind) in x with constant term unity, and with interval of orthogonality defined by $\mathcal{S}$ have the unique property, at each degree k, of possessing the smallest maximum absolute value in this interval of any polynomial of degree k with unit constant term. This is what is referred to as the "mini-max" property. It can be shown that the reciprocal of this maximum absolute value, which the polynomial attains $k+1$ times in $\mathcal{S}$, is the quantity, $\cosh(k\phi)$, appearing in (3.13a). Therefore, under conditions that effectively pertain to the large-scale linear inversion problems of data assimilation, the rate of decrease in the log-norm of the residual for a CG method eventually becomes no better than the rate obtained in an AD method that uses, for its step-size parameters x , the roots of a high-degree Chebyshev polynomial. Table 1 lists some of these asymptotic convergence rate in decades per iteration (using familiar base-10 logarithms to express the magnitude of the residual norm).

For large k and an initial weight distribution w_0 which is continuous and sufficiently regular within its interval of support $\mathcal{S}$ it is instructive to consider a generalization of the functions $|\tilde{r}_k|$ to the more inclusive class of what we might call "pseudo-polynomials" in which the "quantized" distribution of zeroes in $\mathcal{S}$ of a proper polynomial is replaced by an idealized "concentration" $\nu_k(x)$ with support in $\mathcal{S}$, and such that the pseudo-polynomial of degree k has the definition :

$$\tilde{r}_k(x) \equiv \exp\left\{\int_{\mathcal{S}} \log\left|1 - \frac{x}{x'}\right| \, \nu_k(x')\, dx'\right\}, \tag{3.15}$$

with the degree k given by the integral of the concentration,

$$\int_{\mathcal{S}} \nu_k(x)\, dx = k. \tag{3.16}$$

TABLE 1. ASYMPTOTIC CONVERGENCE OF CHEBYSHEV AND CONJUGATE GRADIENT METHODS.

Condition number	Decades reduction in residual norm per iteration
2	.766
5	.418
10	.284
20	.198
50	.124
100	.0872
200	.0615
500	.0389
1000	.0275
2000	.0194
5000	.0123
10000	.00869

The proof of theorem 3 uses functions of the form (3.15) constrained by (3.16) to define the lower bound $\bar{\gamma}_k$ of (3.14). But a proper concentration should also be non-negative:

$$\nu_k(x) \geq 0, \; x \in \mathcal{S}. \tag{3.17}$$

The log-norm γ_k^- obtained by minimizing $\mathcal{C}$ over the class of pseudo-polynomials of degree k constained by (3.17) is clearly not larger than the γ_k obtained when the minimization is carried out over the more restrictive class of true polynomials of degree k. Likewise, the removal of constraint (3.17) leads to a log-norm $\bar{\gamma}_k$ not larger than γ_k^-. For large k the difference, $\gamma_k - \gamma_k^-$, where γ_k is associated with the CG method, tends to remain close to $(\log 2)/2$ when the true polynomials $\tilde{r}_k$ behave almost sinusoidally inside $\mathcal{S}$. Apart from this almost constant difference, γ_k^- typically gives a good indication of the evolution of the CG algorithm's γ_k . The next theorem is therefore relevant in practice to the qualitative description of the convergence behavior of the CG algorithm.

(*h*) *Theorem 4*

When γ_k^- and $\bar{\gamma}_k$ are defined as above, the difference, $\gamma_k^- - \bar{\gamma}_k$, is a non-negative, non-increasing, function of k.

(*i*) *Remarks*

The theory given in the appendix that justifies theorem 4 leads also to the surprising implication that, while the precipitous decline in γ_k characteristic of the early transient phase of CG iterations is reducing some of the residual components at a faster rate than can be achieved by the Chebyshev iterations, there remain other analysis array modes which are being virtually neglected during this transient phase. For these neglected modes (which, incidently, may be quite important for the quality of the assimilation) the convergence is slower than the characteristic Chebyshev rate and, for some modes, may actually be divergent initially. Thus, the common strategy of applying the CG iterations only during this initial phase of rapid residual reduction and terminating the iterations when, for large condition numbers, the inevitable asymptotic

stage is reached, is a strategy almost guaranteed to ensure the effective neglect of some of the modes of the analysis.

4. Structural preconditioning of the analysis

From section 2 we see that, in order to effect a good preconditioning of the inversion problem by conventional means, we would need to discover an inexpensive approximation to the operation of multiplying a vector by the inverse of either $(\hat{\mathbf{B}} + \hat{\mathbf{E}})$ or $(\hat{\mathbf{I}} + \hat{\mathbf{E}}^{-1}\hat{\mathbf{B}})$. Even in 3-D analysis, this is almost impossible to accomplish efficiently, owing to the highly structured nature of these matrices.

Instead, we turn our attention to a reconsideration of the fundamental objective of data analysis. It is when we recognize that minimizing the "cost function" is not the true objective that we can begin to make real progress with this problem. An objective analysis can be quite acceptable as the starting point for operational forecasting provided that it is everywhere within a reasonable tolerance of the formally "optimal" analysis. The appropriate criterion for this tolerance is not the component of measurement error for those modes of the analysis measured with extremely high precision. But note that it is these modes that are responsible for the conditioning difficulties. Here, we propose to alleviate the conditioning problems directly, by intervening in the definitions of the effective measurement precisions to whatever extent is required to bring the condition number of the associated inversion problem down to some pre-established value, typically about five. However, to ensure that we do not thereby do unnecessary damage to the analysis, we propose in addition to wrap the usual cycle of iterations (which now converge extremely rapidly with even the simple Chebyshev method) inside an "outer iteration" in which the effective data values are themselves manipulated to compensate for their otherwise diminished strength of influence.

(*a*) *Linearized analysis algorithm*

Consider the modification of the error covariance,

$$\hat{\mathbf{G}} = \hat{\mathbf{E}} + \hat{\mathbf{F}}, \tag{4.1}$$

where $\hat{\mathbf{F}}$ is positive, symmetric and, in most cases, diagonal, with elements sufficiently large to bring the condition number of

$$(\hat{\mathbf{I}} + \hat{\mathbf{G}}^{-1}\hat{\mathbf{B}}), \tag{4.2}$$

down to the target value [$\mathcal{O}(10)$ or less]. Now we are able to solve the linear inversion,

$$(\hat{\mathbf{G}} + \hat{\mathbf{B}})\hat{f}^{(1)} = (\hat{e} - \hat{b}), \tag{4.3}$$

to high accuracy with few iterations of any of the standard solution methods. The modified analysis forcing $\hat{f}^{(1)}$ is, for most components, a surprisingly good approximation to the true optimal forcing $\hat{f}$, even for those components ("measurement array modes" in the terminology of Bennett et al., 1996) associated with the largest eigenvalues x of the original inversion problem. Indeed, the original problem could even possess infinitely strong (Lagrange multiplier) constraints and yet the corresponding components of $\hat{f}^{(1)}$ typically remain close to the intended Lagrange multipliers.

However, even those components of $\hat{f}^{(1)}$ that are in error can be remedied in subsequent outer iterations when we adopt the scheme:

$$\hat{e}^{(n)} = \hat{e} + \hat{\mathbf{F}}\hat{f}^{(n-1)}, \tag{4.4a}$$

$$(\hat{\mathbf{G}} + \hat{\mathbf{B}})\hat{f}^{(n)} = (\hat{e}^{(n)} - \hat{b}). \tag{4.4b}$$

Let

$$\hat{f}'^{(n)} = \hat{f}^{(n)} - \hat{f}. \tag{4.5}$$

Then,

$$\begin{aligned}(\hat{\mathbf{G}} + \hat{\mathbf{B}})\hat{f}'^{(n)} &= (\hat{\mathbf{G}} + \hat{\mathbf{B}})\hat{f}^{(n)} - (\hat{\mathbf{G}} + \hat{\mathbf{B}})\hat{f}, \\ &= \hat{e} + \hat{\mathbf{F}}\hat{f}'^{(n-1)} - \hat{b} - (\hat{\mathbf{G}} + \hat{\mathbf{B}} - \hat{\mathbf{F}})\hat{f}, \\ &= \hat{\mathbf{F}}\hat{f}'^{(n-1)}. \end{aligned} \tag{4.6}$$

Hence, the rate of convergence of $\hat{f}^{(n)}$ in the outer iteration to the true optimal solution $\hat{f}$ is determined by the eigenvalues ζ of the matrix,

$$\hat{\mathbf{Z}} = (\hat{\mathbf{G}} + \hat{\mathbf{B}})^{-1}\hat{\mathbf{F}} = (\hat{\mathbf{Q}} + \hat{\mathbf{F}})^{-1}\hat{\mathbf{F}}. \tag{4.7}$$

Since $\hat{\mathbf{Q}}$ and $\hat{\mathbf{F}}$ are both positive definite (even for "strong" constraints, for which the corresponding diagonal elements of $\hat{\mathbf{E}}$ vanish), all the eigenvalues ζ lie in the range,

$$0 \leq \zeta < 1,$$

and the outer iterations converge to give the optimal forcing.

(b) *Fully non-linear analysis algorithm*

Note that, if we adopt the Chebyshev AD method for the inner iterations, then it is no longer necessary to ensure that the system operator is symmetric, or even linear. We can therefore tackle directly the full non-linearity of the variational assimilation equations. For example, in solving the equations of 4-D variational analysis given by Bennett et al. (1996), the residual $\hat{r}_k^{(n)}$ corresponding to forcing $\hat{f}_k^{(n)}$ at inner iteration k, outer iteration n, can be calculated without having to construct the linearized operator $\hat{\mathbf{B}}$ explicitly. However, in the AD algorithm, it then is algebraically a little simpler to use "search direction", $\hat{s}_k^{(n)}$, instead of $\hat{r}_k^{(n)}$, where:

$$\hat{s}_k^{(n)} \equiv \hat{\mathbf{G}}^{-1}\hat{r}_k^{(n)}, \tag{4.8}$$

so that the steps involved in the inner iteration of the 4-D variational analysis become:

$$\nu_k^{(n)} = \mathbf{P}^*\mathbf{K}^*\hat{f}_k^{(n)}, \tag{4.9a}$$

$$g_k^{(n)} = h + \mathbf{H}\nu_k^{(n)}, \tag{4.9b}$$

$$a_k^{(n)} = \mathcal{P}(g_k^{(n)}), \tag{4.9c}$$

$$\hat{a}_k^{(n)} = \mathcal{K}(a_k^{(n)}), \tag{4.9d}$$

$$\hat{s}_k^{(n)} = \hat{f}_k^{(n)} - \hat{\mathbf{G}}^{-1}(\hat{e}^{(n)} - \hat{a}_k^{(n)}), \tag{4.9e}$$

$$\hat{f}_{k+1}^{(n)} = \hat{f}_k^{(n)} - \frac{\hat{s}_k^{(n)}}{x_{k+1}}. \tag{4.9f}$$

(*c*) *Estimation and regulation of the condition number*

It is clearly advantageous to use the simpler Chebyshev accelerated iterations with such nonlinear equations. There are variants of the CG or QN methods designed to handle the more general non-quadratic minimization problems, but they often require extra iterative steps in order to determine the step-sizes for each line-search. The AD method is unencumbered by such additional expense. However, successful implementation of an AD method such as the Chebyshev scheme requires that the range of eigenvalues, $[x^-, x^+]$ of the modified system operator, $(\hat{\mathbf{B}} + \hat{\mathbf{G}})\hat{\mathbf{G}}^{-1}$, be known. We can always take $x^- \simeq 1$, and at least a lower bound on x^+ can be estimated dynamically during the iterations by employing the readily verified inequality,

$$x^+ \geq \max_j ((1 - \rho_j)x_j), \tag{4.10a}$$

where ρ_{k+1} is the ratio:

$$\rho_{k+1} = \frac{\hat{s}_k^* \hat{\mathbf{G}} \hat{s}_{k+1}}{\hat{s}_k^* \hat{\mathbf{G}} \hat{s}_k} \equiv \frac{\hat{s}_k^* \hat{r}_{k+1}}{\hat{s}_k^* \hat{r}_k}. \tag{4.10b}$$

By monitoring the bound obtained in (4.10a) during the course of the iterations, it is possible to identify whether the actual condition number of the modified system operator is exceeding its target value. Then, by inspection of the dominant components of the scaled residual $\hat{s}_k$ associated with the ρ_k found from (4.10a) to violate the intended inequality, it is possible to locally "fine-tune" the components $\hat{\mathbf{F}}$, and hence $\hat{\mathbf{G}}$, in order to enforce the maximum limit on the condition number of the modified algorithm that will ensure the rapid convergence of the Chebyshev cycle of AD iterations comprising the "inner iteration".

5. Preconditioning for satellite data

In the handling of satellite radiometric data we have a situation in which a group of related observations comprising the "sounding" have intrinsically small and approximately independent errors, but which measure strongly overlapping thermal characteristics of the atmosphere. In this case, the measurement space covariance $\hat{\mathbf{B}}$ will tend to be strongly and positively correlated within the set of components of the sounding, while the corresponding components $\hat{\mathbf{E}}$ will typically be uncorrelated and of small magnitudes. This isolated part of the inversion process (that is, conventional one-dimensional satellite sounding retrieval) by itself tends to be extremely ill-conditioned, with the very large components of $(\hat{\mathbf{I}} + \hat{\mathbf{E}}^{-1}\hat{\mathbf{B}})$ corresponding to the deepest observed structures and the smallest components (eigenvalues hardly greater than one) corresponding to the high vertical wavenumber structures. This problem is further exacerbated in 3-D or 4-D variational assimilation when many such soundings lie together within distances

comparable to the horizontal scale of the background error covariance $\mathbf{B}$, for then, the precision-weights of these soundings accumulate almost additively to produce condition numbers several times larger than those obtained for each isolated sounding.

In this case, it is not appropriate to augment $\hat{\mathbf{E}}$ by a matrix $\hat{\mathbf{F}}$ that is diagonal, but rather, by one which serves to make the "shape" of the distribution of elements of the submatrix $\hat{\mathbf{G}}$ for each sounding resemble that of the corresponding $\hat{\mathbf{B}}$. A singular-value decomposition of the inversion profile of such soundings taken in isolation reveals the appropriate correlated structure of the matrix increments $\hat{\mathbf{F}}$ required to make the large eigenvalues of $\hat{\mathbf{G}}^{-1}\hat{\mathbf{B}}$ smaller than those of $\hat{\mathbf{E}}^{-1}\hat{\mathbf{B}}$ without reducing any further the eigenvalues that are already small enough (comparable to one). Without this precaution, the structural preconditioning will successfully reduce the inversion problem's condition number, but will seriously slow the convergence of outer iterations for those analysis array modes whose eigenvalues were originally small - those modes being the important high-wavenumber vertical structures near the limit of detectability by the satellite instrument.

6. Discussion

An examination of the theory of 3-D and 4-D optimal analysis, and of the asymptotic nature of the standard iterative algorithm used to solve the equations of optimal analysis, suggests that the major obstacle to achieving the benefits that an optimal analysis offers is the problem of poor conditioning of the system matrix. In operational meteorology it is not feasible to spend more than a few tens of iterations at most on solving for the analysis and, in the case of 4-D analysis, the limit on iterations is even more stringent. In spite of the precipitous drop in the magnitude of the diagnosed residual gradient typically seen in the first few iterations of the standard conjugate gradient or quasi-Newton iterations, the asymptotic analysis presented in section 3 indicates that, in large scale inversions, this initial rate of residual-reduction is often confined to only a sub-range of the analysis modes - those for which the product of initial squared-residuals and eigenvalue concentration just happens to be relatively large. Other more subtle modes of the analysis, possibly of equal importance, are suffering a relative neglect during this early transient "super-Chebyshev" phase of the iterations. Consequently, when the iterations of the solution method are terminated before the asymptotic ("Chebyshev") phase of the procedure is reached, we must seriously question whether the resulting analysis is really sufficiently close to the intended optimum for all analysis modes contributing.

The proposal to precondition by altering the structure of the problem is motivated by the need to keep the total number of iterations of the analysis as small as possible consistent with the attainment of an adequate approximation, for all modes, of the formally optimal analysis state. Evidence from preliminary idealized analysis studies reveals that the proposed new method does indeed work as intended and that the more sophisticated CG and QN iterative techniques, which are restricted in their applications to systems whose matrices are exactly symmetric, can be replaced by a simple Chebyshev iteration which converges well regardless of the non-symmetry or non-linearity of this operator. However, the success of the Chebyshev acceleration method does rely on prior knowledge of the minimum and maximum eigenvalues of the effective linearized system operator. Fortunately, structural preconditioning, as described in section 4, leaves the smallest eigenvalues close to one, and the largest eigenvalues can be estimated

dynamically through the iterative process by what amounts to a variant of the power method. The separation of the solution method into outer and inner iterations permits adaptive changes to the operators $\hat{\mathbf{F}}$, and hence $\hat{\mathbf{G}}$, between successive outer iterations, so that the monitoring of the largest eigenvalue at each stage can be used to make whatever adjustments are necessary to bring the condition number of the modified problem down to below some small pre-selected target value on which a standardized Chebyshev iterative scheme can be guaranteed to work efficiently and robustly.

Appendix A

Properties and applications of Chebyshev polynomials

The theory of convergence of linear iterative methods is intimately connected with the properties of the Chebyshev polynomials of the first kind (Ambramowitz and Stegun, 1965, Chap. 22). For argument $z = \cos(\theta)$ or $z = \cosh(\phi)$ and interval of orthogonality, $[-1, 1]$, the standard definitions of these polynomials are:

$$T_k(z) = \cos(k\theta) = \cosh(k\phi). \tag{A.1}$$

For the weight distribution $W_N(z)$ defined by

$$W_N(z) = \frac{1}{2N} \sum_{j=0}^{2N-1} \delta(z - \cos\frac{\pi j}{N}), \tag{A.2}$$

that is, a superposition of delta-functions forming $N+1$ distinct "impulses" within the standard interval, the polynomials $T_k(z)$ for $k \leq N$ form an orthogonal set. In the limit $N \to \infty$ the discrete weight function (A.2) effectively becomes the piecewise-continuous function:

$$W_\infty(z) = \begin{cases} \frac{1}{\pi}(1-z^2)^{-1/2} & \text{if } z \in (-1,1) \\ 0 & \text{otherwise} \end{cases} \tag{A.3}$$

with respect to which, all the polynomials $T(z)$ are mutually orthogonal:

$$\int_{-1}^{+1} T_j(z)\, T_k(z)\, W_\infty(z)\, dz = \begin{cases} 1, & j = k = 0 \\ 1/2, & j = k > 0 \\ 0, & j \neq k \end{cases} \tag{A.4}$$

We shall have occasion to use the following identities:

$$\int_{-1}^{+1} \log|z - z'|\, T_n(z')\, W_\infty(z')\, dz' = \begin{cases} -\log 2 & , n = 0, \quad |z| \leq 1, \\ -\log 2 + \operatorname{arccosh}|z| & , \ n = 0, \quad |z| \geq 1, \\ -\frac{T_n(z)}{n} & , n > 0, \quad |z| \leq 1, \\ -\frac{1}{n}\Big\{z\Big(1 + (1 - 1/z^2)^{1/2}\Big)\Big\}^{-n} & , n > 0, \quad |z| \geq 1. \end{cases} \tag{A.5}$$

Integrals of the type (A.5) arise in two-dimensional potential theory, where they are most easily studied through the use of elliptic coordinates (Morse and Feshbach, 1953).

For our purposes we need to translate these Chebyshev polynomials so that their standard interval becomes $\mathcal{S} \equiv [x^-, x^+]$. The necessary transformation relating $x \in \mathcal{S}$ to $z \in [-1, +1]$ is:

$$z = z_0 - \frac{2x}{x^+ - x^-} \tag{A.6a}$$

where,

$$z_0 = \frac{x^+ + x^-}{x^+ - x^-} \tag{A.6b}$$

We may also rescale the translated Chebyshev polynomials $\tau_k(x)$ to have unit constant term at $x = 0$:

$$\tau_k(x) = \frac{T_k(z)}{T_k(z_0)} \tag{A.7}$$

Of all real kth degree polynomials with unit constant term, it is easy to show that τ_k is uniquely the one with the "mini-max" property discussed in section 3. The maximum absolute value of $\tau_k(x)$ for x in $\mathcal{S} = [x^-, x^+]$ is $1/T_k(z_0) = \mathrm{sech}(k\phi)$, where ϕ is defined in terms of condition number $c = x^+/x^-$ by (3.13b). This maximum is attained by $|\tau_k(x)|$ $k+1$ times in the interval $\mathcal{S}$ at precisely the locations of the impulsive components of the weight function $W_k(z)$.

Proof of Theorem 2

(i) Defining the "k-cycle Chebyshev method" to be the AD method whose step-size parameters x_j comprise the zeroes of τ_k , the mini-max property, $|\tau_k(x)| \leq \mathrm{sech}(k\phi)$, implies (3.13a) for this Chebyshev method. But we saw in section 3 that, of all k-step AD methods, none reduces log-norm γ_k faster than the CG method, which consequently satisfies the assertion (3.13a) also.
(ii) Construct an initial residual weight $w_0(x) = W_k(z)$ with x and z related through (A.6a)–(A.6b). The first k polynomials $\tilde{r}_j$ of the CG algorithm are then precisely the Chebyshev polynomials τ_j , $j = 1, \ldots, k$. Since, at step k, the weight $w_0(x)$ is concentrated at the $k+1$ locations where $|\tilde{r}_k| = |\tau_j|$ achieves its maximum value of $\mathrm{sech}(k\phi)$, while at these same locations $\tilde{r}_0 = 1$, the equality (3.13c) follows for this specially constructed problem.

Proof of Theorem 3

Consider minimization of $\mathcal{C}$ of (3.11) over the set of pseudo-polynomials defined by (3.15), (3.16) and, optionally, by (3.17). Including the constraints through the use of Lagrange multipliers, we seek the extremum of

$$\begin{aligned} \mathcal{C}(\nu_k, \Lambda, \eta) = & \\ & \int_{\mathcal{S}} \exp\left\{ 2 \int_{\mathcal{S}} \log \left| 1 - \frac{x}{x'} \right| \, \nu_k(x') \, dx' \right\} \frac{w_0(x)}{x} \, dx \\ & + \Lambda \left\{ \int_{\mathcal{S}} \nu_k(x') \, dx' - k \right\} + \int_{\mathcal{S}} \eta(x') \, \nu_k(x') \, dx', \end{aligned} \tag{A.8}$$

with respect to independent variations of Λ, and of $\nu_k(x')$ and $\eta(x')$ for $x' \in \mathcal{S}$, where $\eta(x')$ vanishes unless "activated" and is only activated on sets of x' for which $\nu_k(x') = 0$. Stationary first variations of $\mathcal{C}$ with respect to $\nu_k(x')$ lead to the Euler-Lagrange condition:

$$\int_{\mathcal{S}} 2 \log \left| 1 - \frac{x}{x'} \right| \, \frac{w_k(x)}{x} \, dx + \Lambda + \eta(x') = 0, \quad x' \in \mathcal{S}. \tag{A.9}$$

Second order variation of $\mathcal{C}$ in the "direction" $\delta\nu$ is always non-negative,

$$\delta^2\mathcal{C} \equiv \frac{1}{2}\int\int \mathcal{H}(x', x'')\,\delta\nu_k(x')\,\delta\nu_k(x'')\,dx'\,dx'' \geq 0, \tag{A.10a}$$

owing to the form of the kernel $\mathcal{H}$,

$$\mathcal{H}(x', x'') = \int_{\mathcal{S}} 4\log\left|1-\frac{x}{x'}\right|\,\log\left|1-\frac{x}{x''}\right|\,\frac{w_k(x)}{x}\,dx, \quad x'\in\mathcal{S}, \quad x''\in\mathcal{S}. \tag{A.10b}$$

This confirms the convexity of $\mathcal{C}$ and guarantees the uniqueness of each of the two minimum values of $\mathcal{C}$ obtained with constraint (3.17), and without it. In the case when we do not invoke (3.17), and using overbars to distinguish the variables associated with this less constrained minimization, (A.9) with $\eta = 0$ simplifies to

$$\int_{\mathcal{S}} \log|x'-x|\;\bar{w}_k(x)\,dx = K_k, \quad x'\in\mathcal{S}, \tag{A.11}$$

K_k being some constant for each k. Defining the normalized "Chebyshev weight" for interval $\mathcal{S}$ to be:

$$\Delta\bar{\nu}(x) = \frac{1}{\pi[(x-x^-)(x^+-x)]^{1/2}} = \frac{2W_\infty(z)}{(x^+-x^-)}, \tag{A.12}$$

and using the fact that the Chebyshev polynomials T_n constitute a complete set of basis functions for their standard interval, we can infer from (A.5), (A.11) and (A.12), that

$$\bar{w}_k(x) \propto \Delta\bar{\nu}(x) \tag{A.13}$$

for any k.

Let $\bar{\nu}_k(x)$ be the concentration profile associated with $\bar{w}_k(x)$. From the invariance of the shape of $\bar{w}_k(x)$ implied by (A.13) we obtain:

$$\frac{1}{2}\frac{d}{dk}\log|\bar{w}_k(x)| = \int_{\mathcal{S}} \log\left|1-\frac{x}{x'}\right|\,\frac{d\bar{\nu}_k(x')}{dk}\,dx' \equiv \text{constant for } x\in\mathcal{S}$$

This integral, whose form is similar to (A.11), uniquely specifies the shape of the k-derivative of $\bar{\nu}_k$, while normalization condition (3.16) specifies its amplitude:

$$\frac{d\bar{\nu}_k(x)}{dk} = \Delta\bar{\nu}(x). \tag{A.14}$$

The left side of (A.14), being independent of x, must also equal the rate of change of the corresponding log-norm $\bar{\gamma}_k$. Using (A.5) we discover:

$$\frac{d\bar{\gamma}_k}{dk} \equiv \frac{1}{2}\frac{d}{dk}\log|\bar{w}_k(x)| = -\text{arccosh}\left(\frac{c+1}{c-1}\right) = -\log\left(\frac{1+c^{-1/2}}{1-c^{-1/2}}\right), \tag{A.15}$$

which verifies theorem 3.

Proof of Theorem 4

From (A.14), we have

$$\bar{\nu}_k(x) = \bar{\nu}_0 + k\,\Delta\bar{\nu}(x), \tag{A.16}$$

but the form of $\nu_0(x)$ itself remains unknown. Expanding $\nu_0(x)$ in terms of Chebyshev polynomials:

$$\bar{\nu}_0(x) = \sum_{n=1}^{\infty} \Delta\bar{\nu}(x) b_n T_n(z) \tag{A.17}$$

where z is related to x through (A.6a)–(A.6b), then employing (A.13), it follows that,

$$\frac{1}{2}[\log \Delta\bar{\nu}(x) - \log w_0(x)] = \int_{-1}^{1} \log|z - z'|\; W_\infty(z') \sum_{n=1}^{\infty} b_n\, T_n(z')\, dz' + \text{constant} \tag{A.18}$$

hence, by (A.4) and (A.5),

$$b_n = -n \int_{\mathcal{S}} [\log \Delta\bar{\nu}(x) - \log w_0(x)]\, W_\infty(z)\, T_n(z)\, dz \tag{A.19}$$

[Note that, because of (3.16), a contribution $T_0(z)$ to $\bar{\nu}_0$ does not appear.]

Except when the initial residual weight profile, $w_0(x)$, already has the shape of the Chebyshev weight, $\Delta\bar{\nu}(x)$, some coefficients b_n are non-zero and then the corresponding reconstructed concentration, $\bar{\nu}_0(x)$), together with the subsequent evolving concentration, $\bar{\nu}_k(x)$, for sufficiently small k, violate the inequality (3.17) for some subset of $x \in \mathcal{S}$. For these small enough k, the inclusion of the additional constraint (3.17) must lead to a concentration $\nu_k^-(x)$ different from $\bar{\nu}_k(x)$ and a log-norm γ_k^- larger than $\bar{\gamma}_k$.

The difference is non-zero only when the constraint (3.17) is activated, that is, when $\nu_k^-(x) = 0$ for a nontrivial set of $x \in \mathcal{S}$. Addition of extra constraints cannot decrease the norm. For any $k' > k$, a concentration $\nu_{k'}^*$ defined by adding to ν_k^- the appropriate proportion of the Chebyshev concentration $\Delta\bar{\nu}$:

$$\nu_{k'}^*(x) = \nu_k(x) + (k' - k)\Delta\bar{\nu}, \tag{A.20}$$

implies a log-norm $\gamma_{k'}^*$ that satisfies

$$\gamma_{k'}^* - \bar{\gamma}_{k'} = \gamma_k - \bar{\gamma}_k. \tag{A.21}$$

But this $\nu_{k'}^*(x)$, being positive throughout cS, does *not* activate the constraint (3.17). Therefore,

$$\gamma_{k'}^* \le \gamma_{k'}^*, \tag{A.22}$$

and the assertion of theorem 4 is justified.

Acknowledgments

The author is grateful to Prof. Andrew F. Bennett and to Dr. David F. Parrish for stimulating discussions relating to this work and to Drs. John C. Derber and David F. Parrish for their internal reviews.

References

Abramowitz, M., and I. A. Stegun, 1965 *Handbook of mathematical functions*, Dover, New York. 1046pp.

Bennett, A. F., 1992 *Inverse methods in physical oceanography*, Cambridge University Press. 346pp.

Bennett, A. F., and R. N. Miller, 1991 Weighting initial conditions in variational assimilation schemes. *Mon. Wea. Rev.*, **119,** 1098–1102.

Bennett, A. F., B. S. Chua, and L. M. Leslie, 1996 Generalized inversion of a global numerical weather prediction model. *Meteor. Atmos. Phys.*, **60,** 165–178.

1997 Generalized inversion of a global numerical weather prediction model, II: Analysis and implementation. *Meteor. Atmos. Phys.*, **62,** 129–140.

Fischer, B., 1996 *Polynomial based iteration methods for symmetric linear systems.* Wiley/Teubner, 283pp.

Gandin, L. S., 1963 *Objective analysis of meteorological fields.* Leningrad, Gidromet; (Jerusalem, Israel Program for Scientific Translations; 1965, 242pp).

Ghil, M., S. Cohn, J. Tavantzis, K. Bube and E. Isaacson, 1981 Applications of estimation theory to numerical weather prediction. In: Dynamical meteorology: data assimilation methods, L. Bengtsson, M. Ghil, and E. Kallen, eds. Spinger-Verlag, New York), pp 139–224.

Gill, P. E., W. Murray, and M. H. Wright, 1981 *Practical optimization.* Academic Press, London, 401pp.

Lewis, J. M., and J. C. Derber, 1985 The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37,** 309–327.

Morse, P. M., and H. Feshbach 1953 *Methods of Theoretical Physics*, McGraw-Hill. New York. Vol I, 997pp.

Navon, I. M., and D. M. Legler, 1987 Conjugate-gradient methods for large-scale minimization in meteorology. *Mon. Wea. Rev.*, **115,** 1479–1502.

Parrish, D. F., and J. C. Derber, 1992 The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.*, **120,** 1747–1763.

Sasaki, Y. 1958 An objective analysis based on the variational method. *J. Meteor. Soc. Japan*, **36,** 77–88.

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

Talagrand, O., 1981b On the mathematics of data assimilation. *Tellus*, **33,** 321-339.

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