A Stabilized Sparse-Matrix U-D Square-Root Implementation of a Large-State Extended Kalman Filter

Dale Boggs†, Michael Ghil††, and Christian Keppenne†
†Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, CA 91109-8001, USA
††Department of Atmospheric Sciences and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California, 90024-1565, USA

1. introduction and motivation

The full nonlinear Kalman filter (KF) sequential algorithm is, in theory, well-suited to the four-dimensional data assimilation problem in large-scale atmospheric and oceanic problems (Ghil et al. 1981, Ghil and Malanotte-Rizzoli 1991). Soon after Kalman’s (1960) seminal paper, estimation practitioner’s and numerical analysts became aware of numerical difficulties inherent in the discrete KF recursion equations. A number of early researchers (Bellantoni and Dodge 1967, Schmidt et al. 1968, Leondes 1970) showed that the KF algorithm can be very sensitive to computer roundoff and that results may cease to be meaningful as time advances—even in double precision. A frequent symptom is that an error covariance matrix loses its positive definiteness; this is frequently aggravated by true numerical ill-conditioning which cannot be improved by a simple rescaling of units. Filter “divergence” (Schlee et al. 1967), i.e., numerical results that wildly contradict predicted analytic behavior, encompass at least 3 types of divergence, due to: (1) the effects of computer roundoff, (2) the presence of nonlinearities, and (3) the use of incorrect a priori statistics and an erroneous dynamic model.

These computational shortcomings of the original KF algorithm have motivated alternative formulations of the optimal sequential estimator. Still, many recent KF papers seem oblivious to the conventional discrete KF’s numerical stability problems, or, at the most, consider these to be a computational nuisance. When implementing the extended KF (EKF) for a nonlinear, primitive-equation model with realistic forcing and topography, this issue cannot be ignored: it is a challenging part of the overall problem. The causes of filter-divergence types (1) and (2) play a prominent role in the design of the Kalman-type estimation algorithms discussed below.

A square-root formulation of the KF has inherently better stability and numerical accuracy than the conventional Kalman formulation (Jazwinski 1970, Bierman 1977, Kerr 1990). Square-root filters are algebraically equivalent to the KF, but involve fundamentally different computational methods. A complete suite of square-root type KF recursions was developed primarily for interplanetary spacecraft navigation (Dyer and McReynolds 1969, Thornton and Bierman 1976, Bierman 1977) and yields digitally robust KF implementations. This is, in part, due to the fact that using square-root matrix factors implicitly preserves symmetry and assures nonnegative eigenvalues for the computed covariance; it also reduces the condition number of the matrices that must be inverted.

Square-root algorithms are, nonetheless, not very widely used, because of the erroneous perception that factorization techniques are too complicated compared to the conventional KF; use too much computer storage, and involve too much computation. This perception is due in large part: (1) to an incomplete understanding of the square-root algorithms that are heavily dependent on advanced numerical analysis (Golub and Van Loan 1989, Lawson and Hanson 1974), and (2) on the use of inefficient computer implementations.

The optimality of Kalman-type sequential estimators, whether conventional or square-root, comes at the price of an enormous increase in the number of operations for a full KF implementation. For models with $n$ discrete variables, the cost of advancing the error covariances one time step with the classical Kalman formalism is about $n^3$ times that of integrating the model itself. Since codes simulating or predicting large-scale flows currently have 105–106 variables, KF implementations have so far been mostly experimental, and in low-resolution models with up to a few thousand variables (Jiang and Ghil 1993, and references therein). To mitigate this computational barrier, a number of banded approximations to the KF (Parrish and Cohn 1985; Churnside 1986; Todling and Ghil 1990) have been explored. These approximations to the conventional KF are based on retaining only those elements of the covariance matrix which differ significantly from zero. Since covariances tend to zero with increasing distance (Hallock et al. 1983), it is feasible to calculate—and store—only those diagonals of the forecast-error covariance matrix that contain significant correlations, rather than the entire matrix. The information contained in this greatly reduced set of diagonals is then a good approximation to that contained in the full covariance matrix. These authors also exploited the block-sparseness of the state transition matrix that arises from the finite-difference scheme; the result was a computationally feasible KF for a linear two-dimensional (2-D) shallow-water system. Unfortunately—within the conventional KF formulation—this banded approximation induces a distinct loss of positive-definiteness in the propagated covariance matrix after a few assimilation steps, and the method therefore fails.

To achieve a non-diverging EKF that is computationally feasible for large geophysical-flow models and is as close to optimal as possible, we have extended PC’S banded algorithm, retaining the computation-saving device of diagonal-wise matrix operations on sparse banded matrices, but within a square-root framework. A less restrictive form of the banded approximation was used, to permit the U-D factorization.
2. The U-D square-root filter

The type of square-root KF we implemented, the so-called U-D filter (Bierman 1977), involves a triangular factorization of the covariance matrix $P$ which requires no actual square-root calculations, thus enhancing efficiency. The U-D filter makes use of the matrix decomposition

$$1 = UDU^T,$$  \hspace{1cm} (1)

where $D$ and $U$ are, respectively, diagonal and unitary upper triangular $(n \times n)$ matrices of the form

$$D = \text{diag}[d_1, d_2, \ldots, d_n], \quad U = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$  \hspace{1cm} (2a,b)

This algorithm is of square-root type since $UDU^T$ is a covariance matrix square root. The numerical stability and accuracy of this algorithm—which relates the U-D observation update cycle to the numerically stable Givens orthogonal transformation method—has been established by Gentleman (1973). Moreover, the U-D algorithm approaches the conventional KF in both timing and storage requirements if coded efficiently, e.g., exploiting the zero lower-triangular portion of the $U$ factor in Eq. (2b) by storing and performing computations on those matrix entries only.

U-D factorization is applied to both the covariance propagation in time (forecast) and data update (assimilation) in Kalman’s sequential algorithm, so that all elements of the discrete KF equations—including the gain matrix—are expressed in terms of these factors. We outline here only the U-D form of the covariance propagation step, since it is the dominant computational burden in the KF. The KF covariance propagation equation, in the notation of Ghil and Malanotte-Rizzoli (1991), is

$$P_{k-1}^f = \Psi_{k-1}^f P_{k-1}^a \Psi_{k-1}^T Q_{k-1},$$  \hspace{1cm} (3)

where $P_{k-1}^f$ and $P_{k-1}^a$ are the forecast covariance at time step $k$ and the analysis covariance at step $k-1,$ and $Y$ and $Q$ are the state-transition and process-noise (model error) matrices. Using Eq. (1), Eq. (3) can be written as

$$U_{k-1}^f D_{k-1}^f U_{k-1}^T = \Psi_{k-1}^f U_{k-1}^a D_{k-1}^a \Psi_{k-1}^T Q_{k-1} + \Lambda_{k-1} Q_{k-1}',$$  \hspace{1cm} (4)

with $\Lambda$ defined so that $Q'$ is diagonal. The right-hand side can be factored as $Y Z Y^T$, where

$$Y = \begin{bmatrix} \Psi_{k-1}^f U_{k-1}^a : \Lambda_{k-1} \end{bmatrix}, \quad Z = \begin{bmatrix} D_{k-1}^a \vdots 0 \vdots 0 \vdots Q_{k-1}' \end{bmatrix}.$$  \hspace{1cm} (5a,b)

This factorization has the same general form as Eq. (1), but not the same dimensions, and $Y$ is no longer upper triangular. The cycle is completed by transforming $Y$ and $Z$ to the desired $(n \times n)$ unitary upper triangular and diagonal form, using a weighted Gram-Schmidt orthogonalization procedure (Golub and Van Loan 1989, Sec. 5.2.8) on $y_i$, the rows of $Y$, that generates a set of $n$ $Z$-orthogonal vectors, $b_i$. The elements of the desired $D_{k}^a$ and $U_{k}^f$ are then given by

$$d_{ij} = b_j^T Z b_i, \quad i = 1, 2, \ldots, n; \quad u_{ij} = \begin{cases} \frac{(y_i^T Z b_j)}{d_{jj}}, & i = 1, 2, \ldots, j-1, \\ 0, & \end{cases}$$  \hspace{1cm} (6a,b)

which completes the propagation from $D_{k}^a$ and $U_{k}^f$ to $D_{k}^a$ and $U_{k}^f$.

A measure of the potential for computational difficulty is the condition number, $\kappa(P)$, defined as $\lambda_{\text{max}}/\lambda_{\text{min}}$, where $\lambda_{\text{max}}$ and $\lambda_{\text{min}}$ are the maximum and minimum eigenvalues of the symmetric, positive-definite matrix $P$. If $\kappa(P)$ is large—the order of $10^\ell$, where $\ell$ is the number of significant digits in a computer word—computational difficulties are certain to arise. Conditioning is improved by using estimation equations that use the square roots of $P$ rather than $P$ itself, since for the square-root matrix $\kappa(U D^{1/2}) = 10^{\ell/2}$.

To introduce sparseness into the discrete estimation equations, we approximate the full $P$ by the symmetric banded approximation $P_B$:

$$P_B = \begin{bmatrix} \beta & \ast & \ast & \ast & \ast \\ \ast & \ddots & \ast & \ast & \ast \\ \ast & \ast & \ddots & \ast & \ast \\ \ast & \ast & \ast & \ddots & \ast \\ \ast & \ast & \ast & \ast & \ast \end{bmatrix}.$$  \hspace{1cm} (7)
where the width of the central band in terms of nonzero diagonals is $2\beta+1; \beta = 3N \text{lon}(2b+1),$ with $N \text{lon}$ the number of longitudinal grid points, and $b$ the bandwidth parameter given by the number of grid points—away from a base point in each coordinate direction—over which nonzero covariances will be calculated. Only thenonzero diagonals of $P_B$ are stored. For $P_B$ to be decomposable into U-D form, it cannot preserve any zero diagonals imbedded in the central band. Thus, our central diagonal band is a priori completely nonzero, while PC’S was interspersed with sub-bands of zero diagonals, due to block-bandedness. This key difference allows us to decompose the sparse approximation $P_B$ into U-D square-root factors: if $P_B$ has the form (7), then $P_B = U_B D U_B^T,$ where $U_B$ is unit upper triangular and $D$ is diagonal, if and only if $U_B$ has the form

$$
U_B = \begin{pmatrix}
1 & xx & \cdots & x \\
1 & xx & \cdots & x \\
1 & xx & \cdots & x \\
& & & \\
& & &
\end{pmatrix}
$$

(8)
The banded covariance approximation (7, 8) is more computationally expensive (i.e., with more nonzero elements needing to be stored and computed) than that of PC, but it allows us to retain the U-D factorized form, which, in turn, yields a near-optimal discrete filter that doesn’t diverge. The new U-D square-root filter still takes full advantage of the induced sparseness by not storing—or doing computations of—the $n \cdot \beta$—lower nonzero diagonals of the upper-right portion of the $U_B$ factor (8).

3. the nonlinear shallow-water model

The model used in this study is a 2-D nonlinear shallow-water model, which describes divergent barotropic motion in a hydrostatic fluid with a free surface, confined to a beta-plane channel centered at 45°N (Todling and Ghil, 1994). The discretization is based on a second-order accurate, quadratic-energy-conserving finite-difference scheme (e.g., Grammeltvedt, 1969), using a modified Euler-backward method in time. The prognostic variables are the zonal and meridional velocities, $u$ and $v$, and the geopotential height $h$ at each gridpoint. The boundary conditions are very simple: $v$ is set to zero at the meridional boundaries of the computational domain, and $u$ and $h$ are independent of longitude along the same boundaries. All fields are periodic in the zonal direction.

For estimation purposes, the dependent variables are organized into an ordered state vector of length $n$, where $n$ is three times the number of grid points. The nonzero diagonals of the sparse $n \times n$ state transition matrix $Y'$ are computed analytically with a tangent-linear approximation. This matrix is used in advancing $P_B$ as in Eq. (3), while the state itself is advanced by the full nonlinear equations of motion.

4. Numerical results

4.1 U-D filter test

We have run several assimilation experiments on a 17 x 24 grid with boundaries at 25°N and 65°N (5000 km meridionally), so that $Ax = 1178$ km and $Ay = 312.5$ km, with an integration step size of $At = 15$ rein, and with $n = 3 \times 17 \times 24 = 1224$ state variables. This modelsize is comparable to advanced linear KF implementations (Jiang and Ghil, 1993, Miller et al., 1994, Todling and Ghil, 1994), but the present method should permit, once tested, much larger applications. The height $h$ of the model’s free surface is given initially, and the initial data for the velocity components $u$ and $v$ are calculated from the geostrophic relation. The initial state is a westerly jet with north-south perturbations of different wavelengths and amplitudes along its zonal axis. Synthetic observations are taken every 3 hours, and are processed using the U-D square-root EKF of Sec. 2 with the bandwidth parameter $b$ ranging from 2 to 4 and full (i.e., no approximation, $\beta = n = 1224$). A diagonal model-error (process-noise) matrix $Q$ with nonzero variances for all variables was used in all the experiments.

For the relatively small model size used in these experiments, the banded approximation yields modest computational savings. For example, for $b = 3$, the number of nonzero diagonals of the covariance matrix that must be stored and calculated is 504 out of a total of $n = 1224$. This yields a savings of about $1/2$ in the floating-point operation count, since the centrally located diagonals retained are longer than the ones omitted. However, for a more realistically sized atmospheric or oceanic model, the savings can be more dramatic: for a global model with a 2° x 2° grid using a bandwidth parameter $b = 4$, only 10% of the diagonals of the full covariance matrix are retained and calculated with the approximation. As an accuracy check of the banded U-D filter, we note that the resulting state estimates and their associated error-covariance estimates—obtained using $b = 2, 3$, and 4—did not differ in any significant way from those obtained using the full U-D filter.
On the other hand, a check of the U-D filter's overall performance relative to that of a conventional EKF (hereafter) yielded dramatic results. Every U-D filter experiment was successful in the filter's performance in a numerically stable, non-divergent manner; the same cases run with the EKF on the same CRAY YMP using the same precision—invariably failed due to filter divergence after a few forecast-update steps. The failures were all similar: each would terminate due to a sudden rapid increase in the norm of the computed $P$ matrix, culminating in computer overflow within one day after initial time. It is interesting that comparison EKF runs replicated almost exactly the full U-D filter results $(h=1)$ during the first few assimilation steps; soon, however, results would begin to drift apart due to the accrual of numerical errors in the unstable EKF computations, followed rapidly by the EKF's blow-up.

4.2 State and covariance estimates

We show the results of an "identical-twin" experiment where, starting with an initial "guess" $w'(t_0)$ of the "true" initial state $w(t_0)$, we use noisy observations of the "true" evolution $w'(t)$, $t \geq t_0$, to obtain a U-D filter estimate $w(t)$ that converges to the true field $w(t)$ as assimilation proceeds. The initial error $w'(t_0) - w(t_0)$ was generated with a geostrophically balanced perturbation that has approximate root-mean-square (RMS) values of $\Delta u = 3.8$ m/s and $\Delta v = 1.29$ m/s for the velocities, and $\Delta h = 370$ m for the geopotential height (Fig. 1), with the RMS taken over all grid points. The observing pattern used was highly idealized and the observations rather plentiful in this preliminary experiment: observations of $u$, $v$, and $h$ were made every 12 At = 3 hours at all grid points.

![RMS Error](image)

Figure 1. The 3-day evolution of the RMS assimilation error taken over all the grid points of the beta-plane channel. The estimated error for the wind components $u$ and $v$ and the geopotential height $h$ are shown individually at each forecast step. There are 96 forecast steps per day, and the fields are updated every 12 At (3 hours).

Observations were generated by adding random, uncorrelated Gaussian noise to the true trajectory, with standard deviations $\sigma_u = 2$ m/s, $\sigma_v = 0.75$ m/s, and $\sigma_h = 75$ m. The observations were processed sequentially at each assimilation time (Ghil and Malanotte-Rizzoli 1991, p. 185, and references there), with each row of the observation matrix $H$ containing a single entry equal to one, and the rest zeros. Model errors were also assumed to be uncorrelated, with standard deviations of 10 m/s, 5 m/s, and 200 m per half-day for $u$, $v$, and $h$, respectively. The initial analysis-error covariance matrix, $P_0^{-1}$, was taken to be diagonal, with standard deviations given by 2S rids, 8 m/s, and 1000 m for $u$, $v$, and $h$. Figure 1 shows the RMS error of a 3-day assimilation run at each forecast and analysis step: the RMS errors in the $u$, $v$, and $h$ fields are computed from the vector $w'(t) - w(t)$ by summing over all grid points. The estimated state is seen to converge to the true state in all components.

Figure 2 depicts the evolution of the estimated error in $u$. Each panel in the figure shows the state-estimate error at a specific forecast time over the rectangular beta-channel region with a gray-scale range depicting errors in the wind component ranging from -4 m/s (black) to +4 m/s (white). The images are shown in time from 6 hours to 48 hours past initial time, and show the estimated state converging over the entire spatial domain to the true trajectory.

A closer inspection of Fig. 1 reveals a manifestation of a problem that will be at once: (1) new to manyKF practitioners; (2) explainable only with advanced numerical analysis concepts; (3) difficult, but not impossible, to solve; (4) probably at the heart of what will be the major difficulty in applying the EKF to data assimilation for highly resolved geophysical flow models. In Fig. 1, at the update time itself, the state estimation error increases, rather than decreases. This problem arises from the fact that errors in the observational residual—the difference between an observed value and the model's prediction of that value—are transformed by the Kalman inversion process, even in the square-root form, into large errors in the state estimate along those directions in the
Figure 2. Time evolution of the estimated state error in \( u \); gray scale. Each image depicts the error over the rectangular beta-channel region which constitutes the domain of the cm-layer shallow-water model used. The 17 parallels run 25° to 65° N, and the 24 meridians run from 0° to 360° E.

The \( r \)-dimensional phase space that correspond to eigenvectors of \( P_t \) associated with large eigenvalues. A heuristic explanation of this fact in the context of large-scale geophysical, mid-latitude flows, is that the divergence of the flow field is much smaller than the curl. Hence, errors in observed divergence will affect the assimilated result much more seriously than errors of similar size in the curl. This problem is ubiquitous in EKF estimates of large nonlinear systems that (1) require linearized approximations, and (2) possess covariance matrices and hence their inverses, the system information; matrices that are ill-conditioned.

Figure 3 shows the evolution in time of the spectrum: each curve is the 

Figure 3. Magnitude-ordered spectrum of the 12,24 x 1224 forecast-error covariance matrix at 6 different time steps of the assimilation. Curve 1 (bold) shows the eigenvalues for the 3-tiered initial a priori uncertainty for each of the 3 variables \( u, v \), and \( h \). Curves 2, 3, and 4 are the eigenvalues for forecast steps 2, 3, 4, immediately following the initial time. Curve 5 results at the last forecast step immediately preceding the first analysis time. Curve 6 shows the eigenvalues preceding the third analysis time (step 37).
5. Summary

We have shown that a square-root implementation of the EKF for a fairly realistic shallow-water model is stable. This can overcome the stability problems of the CKF encountered in space-navigation problems, many other engineering areas (Kerr 1990, and references there), by Evensen (1992) in an ocean model, and in the present atmospheric model.

Using U-D factorization, a modified version of Parrish and Cohn's (1989) banded approximation is stable. Until the CKF begins to diverge, it agrees quite well with the banded U-D square-root filter.

A square-root EKF has the best chance of dealing successfully with the problems posed by ill-conditioning of large error-covariance matrices. The difficulties still encountered with the present U-D implementation will be surmounted in the future by changing to a square-root information filter (SRIF: Bierman 1977).

6. References


