International Journal of Control

ISSN: 0020-7179 (Print) 1366-5820 (Online) Journal homepage: https://www.tandfonline.com/loi/tcon20

Implicit Kalman filtering

Mikhail Skliar & W. Fred Ramirez

To cite this article: Mikhail Skliar & W. Fred Ramirez (1997) Implicit Kalman filtering, International Journal of Control, 66:3, 393-412, DOI: 10.1080/002071797224649

To link to this article: https://doi.org/10.1080/002071797224649

INT. J. CONTROL, 1997, VOL. 66, NO. 3, 393–412

Implicit Kalman filtering

MIKHAIL SKLIAR† and W. FRED RAMIREZ‡

For an implicitly defined discrete system, a new algorithm for Kalman filtering is developed and an efficient numerical implementation scheme is proposed. Unlike the traditional explicit approach, the implicit filter can be readily applied to ill-conditioned systems and allows for generalization to descriptor systems. The implementation of the implicit filter depends on the solution of the congruence matrix equation $\mathbf{A}_1\mathbf{P}^x\mathbf{A}_1^{\mathrm{T}} = \mathbf{P}^y$. We develop a general iterative method for the solution of this equation, and prove necessary and sufficient conditions for convergence. It is shown that when the system matrices of an implicit system are sparse, the implicit Kalman filter requires significantly less computer time and storage to implement as compared to the traditional explicit Kalman filter. Simulation results are presented to illustrate and substantiate the theoretical developments.

1. Introduction

The Kalman filter has been extensively studied since its introduction more than thirty years ago (Kalman 1960, Kalman and Bucy 1961). A number of alternative Kalman filtering algorithms have been proposed which are theoretically equivalent or close to the original formulation, and exhibit certain desirable features, such as enhanced numerical stability and computational accuracy, reduced computational requirements, and possibility for parallel computer implementation (Carlson 1990, Chin *et al.* 1995, Jordan 1967, Jover and Kailath 1986, Morf and Kailath 1975, Paige and Saunders 1977, Roy *et al.* 1991, Thornton and Bierman 1980).

The contribution of this paper is the development of a variation of the Kalman filter for a discrete system, specified in the implicit form

$$\mathbf{A}_1(k)\boldsymbol{x}(k+1) = \mathbf{A}_2(k)\boldsymbol{x}(k) + \mathbf{B}(k)\boldsymbol{u}(k) + \mathbf{C}(k)\boldsymbol{w}(k), \tag{1}$$

where $\boldsymbol{x}$ is the state vector of a potentially high dimension n, $\boldsymbol{u}(k)$ is the l-dimensional input vector, $\boldsymbol{w}(k)$ is the p-dimensional stochastic disturbance, $\mathbf{A}_1(k)$ and $\mathbf{A}_2(k)$ are both known $n \times n$ matrices with $\mathbf{A}_1(k)$ possibly ill-conditioned but not singular for all k, $\mathbf{B}(k)$ is the $n \times l$ input transition matrix, $\mathbf{C}(k)$ is the $n \times p$ stochastic disturbance transition matrix, $k = 0, 1, \ldots,$ is the discrete time index. The initial condition $\boldsymbol{x}(0)$ is a Gaussian random vector with mean

$$E\left[\boldsymbol{x}(0)\right] = \hat{\boldsymbol{x}}_0 \tag{2}$$

and covariance

$$E\left[(\boldsymbol{x}(0) - \hat{\boldsymbol{x}}_0)(\boldsymbol{x}(0) - \hat{\boldsymbol{x}}_0)^{\mathrm{T}}\right] = \mathbf{P}_0^x \tag{3}$$

Received 30 January 1995. Revised 9 October 1995. Communicated by Professor M. Morari.

† Department of Chemical and Fuels Engineering, University of Utah, Salt Lake City, UT 84112, U.S.A.; e-mail: Mikhail.Skliar@m.cc.utah.edu.

‡ Department of Chemical Engineering, University of Colorado at Boulder, Boulder, CO 80309-0424, U.S.A.; e-mail: Fred.Ramirez@colorado.edu.

where $E[\cdot]$ is the expectation operator. Measurements are related to states by

$$\boldsymbol{z}(k+1) = \mathbf{H}(k+1)\boldsymbol{x}(k+1) + \boldsymbol{v}(k+1), \tag{4}$$

where $\boldsymbol{z}(k+1)$ is the m-dimensional vector of measurements, $\mathbf{H}(k+1)$ is the known $m \times n$ measurements matrix and $\boldsymbol{v}(k+1)$ is the m-dimensional vector of measurement noise.

Stochastic disturbances $\boldsymbol{w}(k)$ and measurement noise $\boldsymbol{v}(k+1)$ are uncorrelated Gaussian white sequences with zero mean and

$$\left.\begin{aligned} E[\boldsymbol{w}(j)\boldsymbol{w}(k)^{\mathrm{T}}] &= \mathbf{Q}(k)\delta(j-k) \\ E[\boldsymbol{v}(j+1)\boldsymbol{v}(k+1)^{\mathrm{T}}] &= \mathbf{R}(k+1)\delta(j-k) \end{aligned}\right\} \tag{5}$$

where $\mathbf{Q}(k)$ and $\mathbf{R}(k+1)$ are $p \times p$ and $m \times m$ positive definite matrices, and $\delta(j-k) = 1$ if $j = k$ and is zero otherwise.

Discrete implicit systems (both time variant and invariant) and their continuous counterparts naturally arise in many applications (Luenberger 1977, Newcomb and Dziurla 1989, Aplevich 1991). They are also encountered as a result of the discretization of partial differential equations or the discretization of a system of ordinary differential equations using implicit derivative approximation schemes.

If matrix $\mathbf{A}_1$ is allowed to have *rank* $\mathbf{A}_1 < n$, then (1) is often called a descriptor, singular or semistate system. It is also referred to as a generalized system, reflecting the fact that the traditional discrete dynamic system is a particular case of (1) with $\mathbf{A}_1 = \mathbf{I}$. Another special case of system (1) is when $\mathbf{A}_1$ is a diagonal matrix with a large condition number. This is referred to as a singularly perturbed system (Zgurovsky and Skliar 1993a, b).

If $\mathbf{A}_1(k)$ is non-singular for all k, then the implicit system can be cast in an explicit form, and traditional Kalman filtering (Meditch 1969) can be applied to obtain the minimal variance state estimations of (1). However, there is often a strong motivation to avoid matrix inversion required in order to represent an implicit system in the equivalent explicit form. If for some k, matrix $\mathbf{A}_1(k)$ is ill-conditioned, its inverse is calculated with significant error, unless some special measures are built into the filtering algorithm. This would usually involve online calculation of the condition number of $\mathbf{A}_1(k)$ and application of iterative improvement in order to calculate $\mathbf{A}_1^{-1}(k)$ with acceptable accuracy, or the use of a singular value decomposition as a first step in calculating matrix inversion. Clearly, this approach is not well suited for real-time implementation.

Furthermore, as noted by Luenberger (1977), an implicit representation can be simpler than an explicit one in the sense that the interconnection between components of $\mathbf{A}_1(k)\boldsymbol{x}(k+1)$ and $\boldsymbol{x}(k)$ is mild. In terms of matrix structure, this corresponds to the case when matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are sparse. Since matrix inversion destroys any special structure of a matrix, the equivalent explicit representation is a highly interconnected system with system matrix $\mathbf{A}_1^{-1}\mathbf{A}_2$ of a general form. From a computational point of view, operations with sparse matrices are at least an order of n faster than operations with general (full) matrices.

In this paper, we develop an implicit filtering approach for system (1) and (4) which is theoretically equivalent to the optimal Kalman filter but does not require the matrix inversion (see also Skliar and Ramirez 1995). In the next section, known results on Kalman filtering of implicit systems are reviewed, followed by the development of the implicit Kalman filter in § 3. The algorithm of the implicit Kalman

filtering involves prediction and correction steps similar to the traditional Kalman filter. However, the measurement update does not yield the error covariance matrix $\mathbf{P}^x_{k+1|k+1}$ explicitly, but rather it must be determined from the solution of a linear matrix equation $\mathbf{A}_1(k)\mathbf{P}^x_{k+1|k+1}\mathbf{A}_1^{\mathrm{T}}(k) = \mathbf{P}^y_{k+1|k+1}$, where $\mathbf{P}^y_{k+1|k+1}$ is known. In §4, a general method for the iterative solution of this matrix equation is developed, the necessary and sufficient conditions of its convergence are obtained, and the notion of the method's rate of convergence is introduced. In §5, for the case of general matrices $\mathbf{A}_1(k)$ and $\mathbf{A}_2(k)$, we compare the computational efficiency of the developed algorithm with the traditional Kalman filter. An example for sparse banded matrices is considered in §6. Unlike the traditional approach, the implicit filter preserves the structure of $\mathbf{A}_1(k)$ and $\mathbf{A}_2(k)$, allowing for the application of structure specific methods of matrix manipulation. In the case of banded matrices, this results in an implementation of the implicit filter with computational complexity $O(n^2)$, while the traditional equivalent is always of order n^3. In the Conclusions, we give recommendations on the application of this new implicit filter versus the traditional explicit Kalman filter.

2. State estimation of the implicit system

Since it is assumed that the inverse of $\mathbf{A}_1(k)$ exists for all k, the implicit system can be rewritten in the state-space form, allowing for the application of standard state-space methods. The applicability of well known methods inhibited the development of filtering and control methods specifically tailored to non-singular implicit systems. Currently available methods mostly come from the rapidly expanding area of filtering and control of discrete and continuous descriptor systems, which includes non-singular implicit systems as a special case. The initial thrust was to develop methods for the transformation of descriptor systems into the state-space representation of reduced order and to determine the conditions when such a transformation is possible. The resulting state-space system can then be approached using standard state-space methods. A representative list of publications following this approach is Polak (1966), Rosenbrock (1970), Wolovich (1974), Godbout and Jordan (1975).

Particularly relevant to the subject of this work, is a number of recent studies which perform estimation of the descriptor system without reducing it to the state-space form. Bender and Laub (1987a, b) considered a dual linear quadratic control problem for time invariant descriptor systems. A distinguishing feature of their approach is the use of a special coordinate system, associated with the singular value decomposition of matrix $\mathbf{A}_1$, which, as before, allows them to apply standard state-space results to a problem of a reduced order $r = rank\ \mathbf{A}_1 \leq n$.

Chisci and Zappa (1992), used the maximum likelihood (ML) approach for descriptor system filtering. Using the ML perspective, they treat system dynamics (1) and the initial condition $\boldsymbol{x}(0)$ as additional measurements. They suggested a 'canonical' form of the augmented measurement equation such that an explicit ML state estimation can be immediately found, and proposed an algorithm for obtaining the canonical representation.

Nikoukhah *et al.* (1990, 1992), used the maximum likelihood approach to obtain a Kalman filter for descriptor systems, applicable in the case of perfect measurements (i.e. when $\mathbf{R}$ is singular), and when $\mathbf{A}_1$ and $\mathbf{A}_2$ are allowed to be non-square (the case of the under and over constrained system). They also accounted for non-causal

phenomena of descriptor systems by augmenting their filter with the information about future dynamics. When applied to an invertible implicit system, the main difference between our results and theirs is that their state estimation is given in explicit form by a 3-block Kalman filter equation and the corresponding 3-block algebraic Riccati equation. Our development leads to an implicit form for the Kalman filter equation and a predictor–corrector recursion algorithm for the estimation error covariance matrix. It involves the solution of a linear matrix equation of the special form of (18).

The filtering methods, most closely related to our results are based on an implicit filter equation. Shin and Kabamba (1988), considered the filtering problem for a continuous time-invariant constrained dynamic system, written in the descriptor form. Using the Wiener-Hopf theory, they obtained an implicit Kalman filter and the descriptor Riccati equations which are continuous and time-invariant counterparts to our results. However, they limited their treatment to the development of the limiting steady-state estimator algorithm, and did not pursue optimal filtering any further.

Wang and Bernhard (1991), obtained an implicit Kalman filter for the time invariant descriptor system in the form

$$\mathbf{A}_1\hat{\mathbf{x}}_{k+1} = (\mathbf{0} \quad \mathbf{A}_2)\mathbf{M}_k^{-1}\begin{pmatrix} \mathbf{A}_1\hat{\mathbf{x}}_k \\ -\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{z}_k \end{pmatrix} \tag{6}$$

for $\mathbf{B} \equiv \mathbf{0}$, $\mathbf{C} \equiv \mathbf{I}$ with

$$\mathbf{M}_k = \begin{pmatrix} \mathbf{P}_{k|k-1} & \mathbf{A}_1 \\ \mathbf{A}_1^{\mathrm{T}} & -\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H} \end{pmatrix}$$

and where the predicted error covariance $\mathbf{P}_{k|k+1}$ is found from the Riccati equation

$$\mathbf{P}_{k|k-1} = (\mathbf{0} \quad \mathbf{A}_2)\mathbf{M}_{k-1}^{-1}\begin{pmatrix} \mathbf{P}_{k-1|k-2} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H} \end{pmatrix}\mathbf{M}_{k-1}^{-1}\begin{pmatrix} \mathbf{0} \\ \mathbf{A}_2^{\mathrm{T}} \end{pmatrix} + \mathbf{Q} \tag{7}$$

If $\mathbf{P}_{k|k-1}$ is invertible the filter equation (6) can be reduced to the following form:

$$\mathbf{A}_1\hat{\mathbf{x}}_{k+1} = \mathbf{A}_2\left[\hat{\mathbf{x}}_k - \mathbf{L}_{WB}\boldsymbol{r}_k\right] \tag{8}$$

where $\mathbf{L}_{WB}$ is the filter gain matrix and the residual $\boldsymbol{r}_k = (\boldsymbol{z}_k - \mathbf{H}\hat{\mathbf{x}}_k)$. This shows that the development by Wang and Bernhard leads to an implicit filter in which estimate correction is based on the time propagated residual at the previous time step.

3. Implicit Kalman filter

Define an auxiliary variable

$$\boldsymbol{y}(k+1) = \mathbf{A}_1(k)\boldsymbol{x}(k+1) \tag{9}$$

Using this definition system (1) and (4) can be rewritten as

$$\boldsymbol{y}(k+1) = \mathbf{A}_2(k)\boldsymbol{x}(k) + \mathbf{B}(k)\boldsymbol{u}(k) + \mathbf{C}(k)\boldsymbol{w}(k) \tag{10}$$

$$\boldsymbol{z}(k+1) = \mathbf{H}_1(k+1)\boldsymbol{y}(k+1) + \boldsymbol{v}(k+1) \tag{11}$$

where $\mathbf{H}_1$ is specified by the matrix equation

$$\mathbf{H}_1\mathbf{A}_1(k) = \mathbf{H} \tag{12}$$

The system (10) and (11) is in the form which allows for the direct application of Kalman filtering. The estimation of $\boldsymbol{y}(k+1)$ is given by the following filter equation

$$\hat{\boldsymbol{y}}_{k+1|k+1} = \mathbf{A}_2(k)\hat{\boldsymbol{x}}_{k|k} + \mathbf{B}(k)\boldsymbol{u}(k) + \mathbf{L}_y(k+1)\left[\boldsymbol{z} - \mathbf{H}_1(k+1)\hat{\boldsymbol{y}}_{k+1|k}\right] \tag{13}$$

The gain matrix in the filter equation is equal

$$\begin{aligned}\mathbf{L}_y(k+1) &= \mathbf{P}^y_{k+1|k+1}\mathbf{H}_1^{\mathrm{T}}(k+1)\mathbf{R}^{-1}(k+1) \\ &= \mathbf{P}^y_{k+1|k}\mathbf{H}_1^{\mathrm{T}}(k+1)\left[\mathbf{H}_1(k+1)\mathbf{P}^y_{k+1|k}\mathbf{H}_1^{\mathrm{T}}(k+1) + \mathbf{R}(k+1)\right]^{1}\end{aligned} \tag{14}$$

where $\mathbf{P}^y_{k+1|k+1}$ is the estimation error covariance matrix of $\boldsymbol{y}(k+1)$. The predicted error covariance matrix $\mathbf{P}^y_{k+1|k}$ is found as a result of the time propagation of the estimation error covariance $\mathbf{P}^x_{k|k}$ of $\boldsymbol{x}$ according to

$$\mathbf{P}^y_{k+1|k} = \mathbf{A}_2(k)\mathbf{P}^x_{k|k}\mathbf{A}_2^{\mathrm{T}}(k) + \mathbf{C}(k)\mathbf{Q}(k)\mathbf{C}^{\mathrm{T}}(k) \tag{15}$$

The covariance measurement update has its usual form

$$\begin{aligned}\mathbf{P}^y_{k+1|k+1} &= \mathbf{P}^y_{k+1|k} - \mathbf{P}^y_{k+1|k}\mathbf{H}_1^{\mathrm{T}}(k+1) \\ &\quad \times \left[\mathbf{H}_1(k+1)\mathbf{P}^y_{k+1|k}\mathbf{H}_1^{\mathrm{T}}(k+1) + \mathbf{R}^{-1}(k+1)\right]^{1}\mathbf{H}_1(k+1)\mathbf{P}^y_{k+1|k}\end{aligned} \tag{16}$$

or equivalently

$$\mathbf{P}^y_{k+1|k+1} = \left[\mathbf{I} - \mathbf{L}_y(k+1)\mathbf{H}_1(k+1)\right]\mathbf{P}^y_{k+1|k} \tag{17}$$

It is straightforward to relate the covariance matrices of the original state $\boldsymbol{x}$ and the auxiliary variable $\boldsymbol{y}$. By definition

$$\begin{aligned}\mathbf{P}^y_{k+1|k+1} &= E\left[\left(\boldsymbol{y}(k+1) - \hat{\boldsymbol{y}}_{k+1|k+1}\right)\left(\boldsymbol{y}(k+1) - \hat{\boldsymbol{y}}_{k+1|k+1}\right)^{\mathrm{T}}\right] \\ &= \mathbf{A}_1(k)E\left[\left(\boldsymbol{x}(k+1) - \hat{\boldsymbol{x}}_{k+1|k+1}\right)\left(\boldsymbol{x}(k+1) - \hat{\boldsymbol{x}}_{k+1|k+1}\right)^{\mathrm{T}}\right]\mathbf{A}_1^{\mathrm{T}}(k)\end{aligned}$$

or

$$\mathbf{P}^y_{k+1|k+1} = \mathbf{A}_1(k)\mathbf{P}^x_{k+1|k+1}\mathbf{A}_1^{\mathrm{T}}(k) \tag{18}$$

Similarly

$$\mathbf{P}^y_{k+1|k} = \mathbf{A}_1(k)\mathbf{P}^x_{k+1|k}\mathbf{A}_1^{\mathrm{T}}(k) \tag{19}$$

We now can establish the relationship between the gain of the implicit filter (13) and the traditional Kalman filter. Substituting (18) into equation (14) and taking into account the definition of $\mathbf{H}_1$ yields

$$\mathbf{L}_y(k+1) = \mathbf{A}_1(k)\mathbf{L}_x(k+1) \tag{20}$$

where $\mathbf{L}_x(k+1) = \mathbf{P}^x_{k+1|k+1}\mathbf{H}^{\mathrm{T}}(k+1)\mathbf{R}^{-1}(k+1)$ is the gain of the traditional explicit Kalman filter.

We summarize the development of the implicit Kalman filter in the following theorem.

Theorem 1:

(*a*) *The optimal state estimate* $\hat{\boldsymbol{x}}_{k+1|k+1}$ *of the implicit system* (1) *and* (4) *is given by the following implicit Kalman filter equation*

$$\mathbf{A}_1(k)\hat{\boldsymbol{x}}_{k+1|k+1} = \hat{\boldsymbol{y}}_{k+1|k} + \mathbf{L}_y(k+1)\left[\boldsymbol{z}(k+1) - \mathbf{H}_1(k+1)\hat{\boldsymbol{y}}_{k+1|k}\right] \tag{21}$$

for $k = 0, 1, \ldots,$ *where*

$$\hat{\boldsymbol{y}}_{k+1|k} = \mathbf{A}_2(k)\hat{\boldsymbol{x}}_{k|k} + \mathbf{B}(k)\boldsymbol{u}(k) \tag{22}$$

with $\hat{\boldsymbol{x}}_{0|0} = \hat{\boldsymbol{x}}_0$.

(*b*) *The implicit Kalman filter gain* $\mathbf{L}_y(k+1)$ *is given by equation* (14). *The relationship between the implicit and explicit filter gains is given by equation* (20).

(*c*) *The estimation error covariance matrix* $\mathbf{P}^y_{k+1|k+1}$ *satisfies matrix equations* (16) *and* (17). *The predicted error covariance matrix* $\mathbf{P}^y_{k+1|k}$ *is governed by equation* (15).

(*d*) *The error covariance matrix of the state estimate* $\hat{\boldsymbol{x}}_{k+1|k+1}$ *is related to the covariance matrix* $\mathbf{P}^y_{k+1|k+1}$ *by the linear matrix equation* (18).

The implicit Kalman filter is theoretically equivalent to the traditional Kalman filtering, provided the inverse of matrix $\mathbf{A}_1$ exists for all k. However, it provides a basis for a new implementation algorithm for the implicit system (1) which does not require matrix inversion. This makes it a superior approach when matrix $\mathbf{A}_1$ is ill-conditioned or sparse.

We now formulate an algorithm to determine the optimal estimate of the state $\boldsymbol{x}(k+1)$ of the system (1) and (4) using the implicit Kalman filter. Given $\boldsymbol{z}(k+1)$, $\hat{\boldsymbol{x}}_{k|k}$ and $\mathbf{L}_y(k+1)$:

Step 1. compute $\hat{\boldsymbol{y}}_{k+1|k}$ by propagating $\hat{\boldsymbol{x}}_{k|k}$ according to equation (22);

Step 2. solve the linear matrix equation (12) for the modified measurement matrix $\mathbf{H}_1(k+1)$;

Step 3. solve the linear equation (21) for the optimal estimate $\hat{\boldsymbol{x}}_{k+1|k+1}$.

The Kalman gain $\mathbf{L}_y(k+1)$ can be calculated in the following way. Given $\mathbf{P}^x_{k|k}$,

Step 1. compute $\mathbf{P}^y_{k+1|k}$ by propagating $\mathbf{P}^x_{k|k}$ according to equation (15);

Step 2. compute the implicit Kalman filter gain using equation (14);

Step 3. calculate $\mathbf{P}^y_{k+1|k+1}$ according to equation (17).

Note that in order to initiate the gain calculation algorithm on the next time step, we need to find $\mathbf{P}^x_{k+1|k+1}$ from the linear equation (18).

Remark 1: Although not explicitly addressed in this paper, the implicit Kalman filter can be generalized for the case of the rank deficient and rectangular matrix $\mathbf{A}_1(k)$. If the matrix pencil of the implicit filter is regular, the solution of such a descriptor system is well known and understood (Lewis 1986). The most straightforward generalization on non-regular (including rectangular) descriptor systems is done if the solution of the involved linear equations is sought in the minimum-norm sense, such as the one generated through the use of the Moore–Penrose matrix inverse (Campbell and Meyer 1979). However, the detailed treatment of implicit filtering for descriptor systems is a topic of future research. □

Remark 2: The implicit formulation of the Kalman filter (as the underlying system itself) does not have a natural orientation, i.e. the state estimation can be

easily propagated in the direction of increasing or decreasing index k. This makes our approach attractive for the fixed-interval two-point boundary value smoothing (Adams et al. 1984), using the standard Luenberger's double sweep method of solution (Luenberger 1977). Furthermore, its extension to non-causal systems can be used as a smoother for two-dimensional discrete systems, such as images (Lewis and Mertzios 1992).The work by Nikoukhah *et al.* (1992) gives a direction on how to account for future dynamics in the smoothing problem for non-causal systems. □

Remark 3: The square root implementation of the implicit Kalman filter is studied in Skliar and Ramirez (1996a). □

4. Iterative solution for the covariance matrix

The direct solution of the equation

$$\mathbf{A}_1(k)\mathbf{P}^x_{k+1|k+1}\mathbf{A}_1^{\mathrm{T}}(k) = \mathbf{P}^y_{k+1|k+1} \tag{18}$$

is equivalent to the solution of the system of $(n^2 + n)/2$ linear equations, and, therefore, is quite involved. In order to make the implicit approach to Kalman filtering practically attractive, we need a method for the efficient solution of equation (18).

The solution of more general linear matrix equations is considered by Chu (1990), and in the references cited therein. The solution is based on the matrix transformations, such as (generalized) singular value and QR decomposition. The least square type solutions are also studied in some detail.

It is known that due to the recurrent nature of the Kalman filter

$$\mathbf{P}^x_{k+1|k+1} \approx \mathbf{P}^x_{k|k} \tag{23}$$

In fact, as $k \to \infty$, equation (23) becomes an equality as the system tends to its steady state. The availability of a good initial approximation for $\mathbf{P}^x_{k+1|k+1}$ suggests the application of iterative methods for the solution of (18).

Let

$$\mathbf{P}^x_{k+1|k+1}\mathbf{A}_1^{\mathrm{T}}(k) = \mathbf{X} \tag{24}$$

and consider the same partition of the matrix $\mathbf{A}_1(k)$ as the one used in the point Jacobi iterative method (Varga 1962)

$$\mathbf{A}_1(k) = \mathbf{D} - \mathbf{L} - \mathbf{U} \tag{25}$$

where $\mathbf{D} = \mathrm{diag}\left[\mathbf{A}_1(k)\right]$ and $\mathbf{L}$ and $\mathbf{U}$ are strictly lower and upper triangular matrices whose non-zero entries are equal to the negative of the corresponding entries of $\mathbf{A}_1(k)$ below and above the main diagonal. With matrix $\mathbf{A}_1(k)$ partitioned according to (25) we can write equation (18) as

$$\mathbf{DX} = (\mathbf{L} + \mathbf{U})\mathbf{X} + \mathbf{P}^y_{k+1|k+1} \tag{26}$$

or, if $\mathbf{D}$ has no zero diagonal elements

$$\mathbf{X} = \mathbf{D}^{-1}(\mathbf{L} + \mathbf{U})\mathbf{X} + \mathbf{D}^{-1}\mathbf{P}^y_{k+1|k+1} \tag{27}$$

Introducing the notation

$$\mathbf{B} = \mathbf{D}^{-1}(\mathbf{L} + \mathbf{U})\mathbf{X} + \mathbf{D}^{-1}\mathbf{P}^y_{k+1|k+1} \tag{28}$$

we obtain that (27) becomes

$$\mathbf{P}^{x}_{k+1|k+1}\mathbf{A}_1^{\mathrm{T}}(k) = \mathbf{B} \tag{29}$$

The application of the matrix partition (25) a second time yields

$$\mathbf{P}^{x}_{k+1|k+1} = \mathbf{P}^{x}_{k+1|k+1}(\mathbf{L}+\mathbf{U})^{\mathrm{T}}\mathbf{D}^{-1} + \mathbf{B}\mathbf{D}^{-1} \tag{30}$$

Following the spirit of the point Jacobi method, we seek the solution of equation (30) using the following successive iteration procedure

$$\begin{aligned}\left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m)} &= \left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m-1)}(\mathbf{L}+\mathbf{U})^{\mathrm{T}}\mathbf{D}^{-1} + \mathbf{B}\mathbf{D}^{-1} \\ &= \left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m-1)}(\mathbf{L}+\mathbf{U})^{\mathrm{T}}\mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{P}^{y}_{k+1|k+1}\mathbf{D}^{-1} \\ &\quad + \mathbf{D}^{-1}(\mathbf{L}+\mathbf{U})\left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m-1)}\mathbf{A}_1^{\mathrm{T}}(k)\mathbf{D}^{-1}\end{aligned} \tag{31}$$

where m is the number of the current iteration.

Using the fact that

$$(\mathbf{L}+\mathbf{U})^{\mathrm{T}}\mathbf{D}^{-1} = \left(\mathbf{I} - \mathbf{A}_1^{\mathrm{T}}(k)\mathbf{D}^{-1}\right)$$

and

$$\mathbf{D}^{-1}(\mathbf{L}+\mathbf{U}) - \mathbf{I} = -\mathbf{D}^{-1}\mathbf{A}_1(k)$$

equation (31) is therefore reduced to

$$\begin{aligned}\left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m)} &= \left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m-1)} - \mathbf{D}^{-1}\mathbf{A}_1(k)\left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m-1)}\mathbf{A}_1^{\mathrm{T}}(k)\mathbf{D}^{-1} \\ &\quad + \mathbf{D}^{-1}\mathbf{P}^{y}_{k+1|k+1}\mathbf{D}^{-1}\end{aligned} \tag{32}$$

The matrix $\mathbf{A}_1(k)$ can be legitimately partitioned in many different ways leading to iterative schemes different from (32). For instance, the partition

$$\omega\mathbf{A}_1(k) = \mathbf{D} - \omega\mathbf{L} - \left[(1-\omega)\mathbf{D} + \omega\mathbf{U}\right]$$

leads to a family of relaxation methods

$$\begin{aligned}\left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m)} &= \left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m-1)} \\ &\quad - \omega(\mathbf{I} - \omega\mathbf{D}^{-1}\mathbf{L})^{-1}\mathbf{D}^{-1}\mathbf{A}_1(k)\left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m-1)}\left[\omega(\mathbf{I} - \omega\mathbf{D}^{-1}\mathbf{L})^{-1}\mathbf{D}^{-1}\mathbf{A}_1(k)\right]^{\mathrm{T}} \\ &\quad + \omega(\mathbf{I} - \omega\mathbf{D}^{-1}\mathbf{L})^{-1}\mathbf{D}^{-1}\mathbf{P}^{y}_{k+1|k+1}\left[\omega(\mathbf{I} - \omega\mathbf{D}^{-1}\mathbf{L})^{-1}\mathbf{D}^{-1}\right]^{\mathrm{T}}\end{aligned} \tag{33}$$

where $\mathbf{D}$, $\mathbf{L}$, $\mathbf{U}$ have the same meaning as in (25), and $\omega > 0$ is the relaxation factor. An important particular case of the relaxation methods, known as the point Gauss–Seidel method, is obtained by setting $\omega = 1$

$$\begin{aligned}\left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m)} &= \left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m-1)} \\ &\quad - (\mathbf{D} - \mathbf{L})^{-1}\mathbf{A}_1(k)\left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m-1)}\mathbf{A}_1^{\mathrm{T}}(k)\left[(\mathbf{D} - \mathbf{L})^{-1}\right]^{\mathrm{T}} \\ &\quad + (\mathbf{D} - \mathbf{L})^{-1}\mathbf{P}^{y}_{k+1|k+1}\left[(\mathbf{D} - \mathbf{L})^{-1}\right]^{\mathrm{T}}\end{aligned} \tag{34}$$

Let us consider a general iterative method

$$\left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m)} = \left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m-1)} - \mathbf{M}\left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m-1)}\mathbf{M}^{\mathrm{T}} + \mathbf{M}\mathbf{P}^{x}_{k+1|k+1}\mathbf{M}^{\mathrm{T}} \tag{35}$$

where $\mathbf{M}$ is the $n \times n$ iterative matrix, such that

$$\mathbf{M}\mathbf{P}^{x}_{k+1|k+1}\mathbf{M}^{\mathrm{T}} = \mathbf{M}\mathbf{A}_1^{-1}(k)\mathbf{P}^{y}_{k+1|k+1}\mathbf{A}_1^{-1\mathrm{T}}(k)\mathbf{M}^{\mathrm{T}} \tag{36}$$

Equation (36) merely requires that the iterative matrix has a structure $\mathbf{M} = \bar{\mathbf{A}}\mathbf{A}_1(k)$, where $\bar{\mathbf{A}}$ is some chosen matrix. Also, note that the exact solution of the matrix equation (18) satisfies the iterative equation (35).

The specific instances of the matrix $\mathbf{M}$ are:

Jacobi method: $\mathbf{M} = \mathbf{D}^{-1}\mathbf{A}_1(k)$

Gauss- Seidel method: $\mathbf{M} = (\mathbf{D} - \mathbf{L})^{-1}\mathbf{A}_1(k)$

Relaxation methods: $\mathbf{M} = \omega(\mathbf{I} - \omega\mathbf{D}^{-1}\mathbf{L})^{-1}\mathbf{D}^{-1}\mathbf{A}_1(k)$

To analyse the convergence of the iterative process, we define the error matrix $\mathbf{E}^{(m)}_{k+1}$ as

$$\mathbf{E}^{(m)}_{k+1} = \left[\mathbf{P}^{x}_{k+1|k+1}\right]^{m)} - \mathbf{P}^{x}_{k+1|k+1} \tag{37}$$

where $\mathbf{P}^{x}_{k+1|k+1}$ is the solution of (18). It is straightforward to show that

$$\mathbf{E}^{(m+1)}_{k+1} = \mathbf{E}^{(m)}_{k+1} - \mathbf{M}\mathbf{E}^{(m)}_{k+1}\mathbf{M}^{\mathrm{T}} \tag{38}$$

and

$$\mathbf{E}^{(m)}_{k+1} = \mathbf{E}^{(0)}_{k+1} - \mathbf{M}\left[\sum_{i=0}^{m-1}\mathbf{E}^{(i)}_{k+1}\right]\mathbf{M}^{\mathrm{T}} \tag{39}$$

where $\mathbf{E}^{(0)}_{k+1}$ is the initial guess error. When $m \to \infty$ the error series (39) converges to some matrix if there is a norm $\|\cdot\|$ such that numerical series $\|\mathbf{E}^{(0)}_{k+1}\| - \sum_{i=0}^{\infty} \|\mathbf{M}\mathbf{E}^{(i)}_{k+1}\mathbf{M}^{\mathrm{T}}\|$ is convergent. According to d'Alambert's ratio test, the numerical series is convergent if

$$\lim_{m\to\infty} \frac{\|\mathbf{E}^{(m+1)}_{k+1}\|}{\|\mathbf{E}^{(m)}_{k+1}\|} < 1$$

and for (39) this last condition is satisfied only if

$$\|\mathbf{E}^{(m)}_{k+1}\| > \|\mathbf{E}^{(m+1)}_{k+1}\| \tag{40}$$

for all m.

The following theorems provide the necessary and sufficient conditions for the convergence of the general iterative process (35).

Theorem 2: *If the iterative matrix* $\mathbf{M}$ *is non-singular, then the limit of the converging error series* (39) *as* $m \to \infty$ *is a zero matrix.*

Proof: Suppose that the error series (39) is convergent, and that

$$\mathbf{E}^{(\infty)}_{k+1} = \lim_{m\to\infty} \mathbf{E}^{(m)}_{k+1}.$$

Then $\mathbf{E}^{(\infty)}_{k+1}$ satisfies equation (38)

$$\mathbf{E}^{(\infty)}_{k+1} = \mathbf{E}^{(\infty)}_{k+1} - \mathbf{M}\mathbf{E}^{(\infty)}_{k+1}\mathbf{M}^{\mathrm{T}}$$

Since $\mathbf{M}$ is non-singular conclude that $\lim_{m\to\infty} \mathbf{E}^{(m)}_{k+1} = \mathbf{0}$.

Thus if (39) is a convergent series, it converges to a zero matrix, and $\left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m)}$ converges to an exact solution of equation (18). □

Theorem 3—Necessary condition: *If the error series* (39) *converges for an arbitrary symmetric initial guess error* $\mathbf{E}^{(0)}_{k+1}$, *then the spectral radius* $\rho(\mathbf{M})$ *of the iterative matrix* $\mathbf{M}$ *is less than* $\sqrt{2}$.

Proof: From the definition of the error matrix (37) if follows that $\mathbf{E}^{(m)}_{k+1}$ is a symmetric matrix. Then $\mathbf{M}\mathbf{E}^{(m)}_{k+1}\mathbf{M}^{\mathrm{T}}$ is also symmetric. Since the spectral norm of a symmetric matrix is equal to its spectral radius, we obtain

$$\left\|\mathbf{E}^{(m+1)}_{k+1}\right\| = \left\|\mathbf{E}^{(m)}_{k+1} - \mathbf{M}\mathbf{E}^{(m)}_{k+1}\mathbf{M}^{\mathrm{T}}\right\| \geq \left|\left\|\mathbf{E}^{(m)}_{k+1}\right\| - \left\|\mathbf{M}\mathbf{E}^{(m)}_{k+1}\mathbf{M}^{\mathrm{T}}\right\|\right|$$

$$= \left|\rho\left(\mathbf{E}^{(m)}_{k+1}\right) - \rho\left(\mathbf{M}\mathbf{E}^{(m)}_{k+1}\mathbf{M}^{\mathrm{T}}\right)\right|$$

Since

$$\frac{\left\|\mathbf{E}^{(m+1)}_{k+1}\right\|}{\left\|\mathbf{E}^{(m)}_{k+1}\right\|} \geq \frac{\left|\rho\left(\mathbf{E}^{(m)}_{k+1}\right) - \rho\left(\mathbf{M}\mathbf{E}^{(m)}_{k+1}\mathbf{M}^{\mathrm{T}}\right)\right|}{\rho\left(\mathbf{E}^{(m)}_{k+1}\right)}$$

for (40) to hold it is necessary that

$$\frac{\left|\rho\left(\mathbf{E}^{(m)}_{k+1}\right) - \rho\left(\mathbf{M}\mathbf{E}^{(m)}_{k+1}\mathbf{M}^{\mathrm{T}}\right)\right|}{\rho\left(\mathbf{E}^{(m)}_{k+1}\right)} < 1$$

or

$$\left|\rho\left(\mathbf{E}^{(m)}_{k+1}\right) - \rho\left(\mathbf{M}\mathbf{E}^{(m)}_{k+1}\mathbf{M}^{\mathrm{T}}\right)\right| < \rho\left(\mathbf{E}^{(m)}_{k+1}\right)$$

which is satisfied if $\rho\left(\mathbf{M}\mathbf{E}^{(m)}_{k+1}\mathbf{M}^{\mathrm{T}}\right) < 2\rho\left(\mathbf{E}^{(m)}_{k+1}\right)$, or if $\rho(\mathbf{M}) < \sqrt{2}$. □

Theorem 4: *If the iterative matrix* $\mathbf{M}$ *is nonsingular, then matrix* $\left[\mathbf{P}^{x}_{k+1|k+1}\right]^{(m)}$, *calculated by the general iterative method* (35) *tends to an exact solution of the matrix equation* (18) *as* $m \to \infty$ *for all time steps* $k+1$ *and all symmetric initial guesses* $\left[\mathbf{P}^{x}_{k+1|k+1}\right]^{0)}$ *if and only if the eigenvalues* $\lambda^{\mathrm{M}}_1, \lambda^{\mathrm{M}}_2, \ldots, \lambda^{\mathrm{M}}_n$ *of* $\mathbf{M}$ *lie within a complex plane circle with radius* $\sqrt{2} - 1$, *centered at the point* $(1, 0)$.

Proof: The necessary condition directly follows from Theorems 2 and 3.To prove sufficiency, let

$$\mathbf{M} = (\mathbf{I} + \mathbf{F})$$

where $\mathbf{I}$ is the unit matrix. Then

$$\mathbf{M}\mathbf{E}^{(m)}_{k+1}\mathbf{M}^{\mathrm{T}} = \mathbf{E}^{(m)}_{k+1} + \mathbf{E}^{(m)}_{k+1}\mathbf{F}^{\mathrm{T}} + \mathbf{F}\mathbf{E}^{(m)}_{k+1} + \mathbf{F}\mathbf{E}^{(m)}_{k+1}\mathbf{F}^{\mathrm{T}}$$

and

$$\left\|\mathbf{E}^{(m+1)}_{k+1}\right\| \leq 2\left\|\mathbf{E}^{(m)}_{k+1}\right\|\|\mathbf{F}\| + \left\|\mathbf{E}^{(m)}_{k+1}\right\|\|\mathbf{F}\|^2$$

This yields the following condition for the error series convergence

$$\frac{\left\|\mathbf{E}^{(m+1)}_{k+1}\right\|}{\left\|\mathbf{E}^{(m)}_{k+1}\right\|} \leq 2\|\mathbf{F}\| + \|\mathbf{F}\|^2$$

or

$$\|\mathbf{F}\| < \sqrt{2} - 1 \tag{41}$$

The spectral radius of a matrix is the largest lower bound for the values of all matrix norms (Horn and Johnson 1985), and there always exists at least one matrix norm $\|\cdot\|$, such that for an arbitrary small positive ϵ

$$\rho(\mathbf{F}) \leq \|\mathbf{F}\| \leq \rho(\mathbf{F}) + \epsilon$$

If $\rho(\mathbf{F}) < \sqrt{2} - 1$, then we can always find $\epsilon < 1$, such that $\rho(\mathbf{F}) + \epsilon$ is still less than $\sqrt{2} - 1$, but larger than at least one norm of $\mathbf{F}$. Thus, the sufficient condition (41) can be replaced by

$$\rho(\mathbf{F}) < \sqrt{2} - 1$$

The spectrum of matrix $\mathbf{M}$ is defined by the following equation

$$\mathbf{M}\mathbf{x}_i = \lambda_i^{\mathrm{M}}\mathbf{x}_i$$

where λ_i^M is $\mathbf{M}$'s eigenvalue, and $\mathbf{x}_i$ is the corresponding eigenvector. The spectrum of $(\mathbf{M} - \mathbf{I})$ can now be defined as

$$\mathbf{M}\mathbf{x}_i - \mathbf{I}\mathbf{x}_i = (\lambda_i^{\mathrm{M}} - 1)\mathbf{x}_i$$

The spectral radius of $(\mathbf{M} - \mathbf{I})$ is therefore equal to $\rho(\mathbf{M} - \mathbf{I}) = \max_i |\lambda_i^{\mathrm{M}} - 1|$. The sufficient condition for convergence can now be written as

$$\max_i |\lambda_i^M - 1| < \sqrt{2} - 1 \tag{42}$$

which limits the eigenvalues of the admissible iterative matrix $\mathbf{M}$ to an open disc with radius $(\sqrt{2} - 1)$, centred around the point $(1, 0)$. □

Figure 1 depicts a typical spectrum of the admissible iterative matrix $\mathbf{M}$, which was used by Skliar and Ramirez (1996b) to filter an implicit system of order 560.

From a practical point of view it is important to be able to predict before the actual computations which admissible iterative matrix $\mathbf{M}$ produces a faster converging iterative process (35). We first note that the fastest possible convergence is achieved when $\mathbf{M} = \mathbf{I}$. The exact solution in this case is reached in one iteration. Thus, the natural way to relate the rate of convergence of the iterative process (35) to the particular form of matrix $\mathbf{M}$ is by measuring the closeness of $\mathbf{M}$ to the unit matrix. In terms of the eigenvalues λ_i^M of matrix $\mathbf{M}$, its closeness to the unit matrix can be characterized by the constant

$$d = \max_i \{|1 - \lambda_i^M|\} \tag{43}$$

These considerations give rise to the following *definition* of the rate of convergence

$$R(\mathbf{M}) = \ln \frac{\sqrt{2} - 1}{d} \tag{44}$$

where constant $\ln(\sqrt{2} - 1)$ is introduced to ensure that $R(\mathbf{M}) > 0$ for all converging iterative schemes (35).

If $\mathbf{M}_1$ and $\mathbf{M}_2$ are two admissible matrices, then we say that $\mathbf{M}_1$ is iteratively faster than $\mathbf{M}_2$ if $R(\mathbf{M}_1) > R(\mathbf{M}_2)$. As in the case of traditional iterative methods, extra caution should be exercised when judging an iterative scheme (35) by its *a priori* rate of convergence.

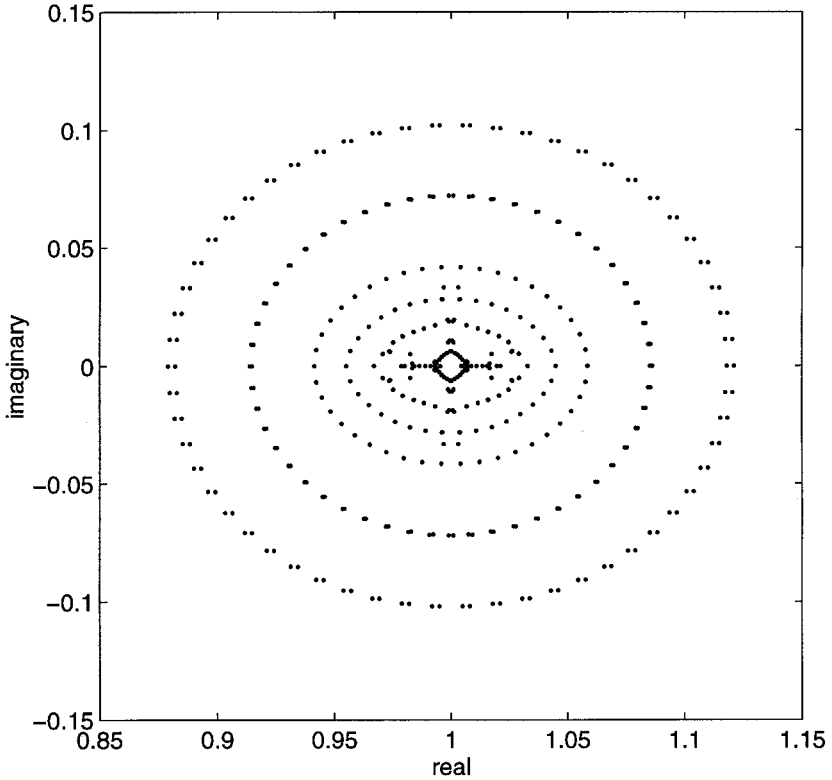

Figure 1. Eigenvalues of the admissible iterative matrix $\mathbf{M}$, $n = 560$.

Let us consider equation (35) for the case of a single iteration and study the behaviour of the introduced error as $k \to \infty$. By doing so we obtain

$$\left[\mathbf{P}^x_{k+1|k+1}\right]^{1)} = \mathbf{P}^x_{k|k} - \mathbf{M}\mathbf{P}^x_{k|k}\mathbf{M}^{\mathrm{T}} + \mathbf{M}\mathbf{P}^x_{k+1|k+1}\mathbf{M}^{\mathrm{T}} \tag{45}$$

where $\mathbf{P}^x_{k|k}$ is used as an initial guess. The error introduced by one iteration is equal to

$$\mathbf{E}^{(1)}_{k+1} = \left[\mathbf{P}^x_{k|k} - \mathbf{P}^x_{k+1|k+1}\right] - \mathbf{M}\left[\mathbf{P}^x_{k|k} - \mathbf{P}^x_{k+1|k+1}\right]\mathbf{M}^{\mathrm{T}} \tag{46}$$

If system (1) and (4) is both (completely) controllable and observable, and $\mathbf{A}_1(k) \to \mathbf{A}_1$ and $\mathbf{A}_2(k) \to \mathbf{A}_2$ as $k \to \infty$, then for any non-negative definite and symmetric initial state $\mathbf{P}^x_0$, $\mathbf{P}^x_{k|k} \to \mathbf{P}^x_{\mathrm{ss}}$ as $k \to \infty$, where $\mathbf{P}^x_{ss}$ is a symmetric positive definite matrix independent of $\mathbf{P}^x_0$. It immediately follows that, in this case, even a single iteration approximation tends to an exact solution as $k \to \infty$, or

$$\lim_{k\to\infty} \mathbf{E}^{(1)}_{k+1} = \mathbf{0} \tag{47}$$

Thus, we have obtained, that the significance of the error in the iterative solution of the matrix equation (18) diminishes with time as the covariance matrix $\mathbf{P}^x_{k|k}$ tends to its steady state, provided the system (1) and (4) is both controllable and observable.

5. Computational efficiency in the case of general matrices

The comparison of the computational efficiency of the implicit filter with the traditional (explicit) Kalman filter in the case of general matrices $\mathbf{A}_1(k)$ and $\mathbf{A}_2(k)$ is based on the number of floating point arithmetic operations and the computer storage required for the implementation of the filter. The number of floating point operations (flops, as defined by Golub and Van Loan 1989) will somewhat differ, depending on the choice of the method for the matrix inversion and solution of the system of linear equations. We choose to perform first the LU decomposition of the matrix $\mathbf{A}_1(k)$, and use it to calculate $\mathbf{A}_1^{-1}(k)$ and to solve the linear system with multiple right sides. By doing so, we obtain that for each time step the implementation of the explicit filter requires $7\frac{2}{3}n^3 + O(n^2)$ flops, while the implicit filter takes $3\frac{2}{3}n^3 + mI + O(n^2)$ operations, where m is the number of iterations (35) and I is the number of flops in one iteration. The estimation of I for different methods gives:

$$\textit{Jacobi method}: \qquad I = 3n^3 + O(n^2)$$

$$\textit{Gauss- Seidel and relaxation methods}: \qquad I = 6n^3 + O(n^2)$$

As we can see, only the implicit filter with a single Jacobi iteration on each time step is computationally faster than the traditional filter. However, computation of $\mathbf{P}^x_{k+1|k+1}$ requires additional storage for $n^2/2$ elements.

Thus, in the case of a general and well conditioned implicit system (1) the implicit Kalman filter performs slightly faster at the expense of additional computer memory and computational accuracy. However, for an ill-conditioned matrix $\mathbf{A}_1(k)$, the implicit filter can be significantly more accurate than the traditional Kalman filter.

The true power of the implicit filtering is revealed when matrix $\mathbf{A}_1(k)$ (and possibly $\mathbf{A}_2(k)$) has a special structure which allows for more effective matrix manipulation and storage. Unlike the traditional approach, the implicit filter preserves the structure of both $\mathbf{A}_1(k)$ and $\mathbf{A}_2(k)$ making the implicit filtering a superior approach in the practically important cases when $\mathbf{A}_1(k)$ is sparse, and in particular banded.

6. Example: implicit Kalman filter for a banded system

Consider an implicit system (1), resulting from the finite difference approximation of a two-dimensional convection–diffusion partial differential equation as described in the Appendix. Both $\mathbf{A}_1(k)$ and $\mathbf{A}_2(k)$ are banded matrices with the structure shown in Fig. 2 for $n = 24$. When indexed storage is used, only $5n$ elements of each matrix have to be stored. However, matrix $\mathbf{A}_1^{-1}(k)\mathbf{A}_2(k)$ is no longer sparse and therefore requires the storage for all n^2 elements. As a result, the required storage for implementation of the traditional filter is of order $n^2/2$ larger than the storage required for the implicit Kalman filter.

The computer time required for the implementation of the implicit filter can be dramatically reduced by the application of structure specific methods of matrix manipulation. In fact, it is often possible, as in the case of the matrix structure shown in Fig. 2, and banded matrices in general, to develop a computer implementation of the implicit filter, which for large n requires only $O(n^2)$ flops on each time step for any of the discussed methods of the iterative solution of equation (18).

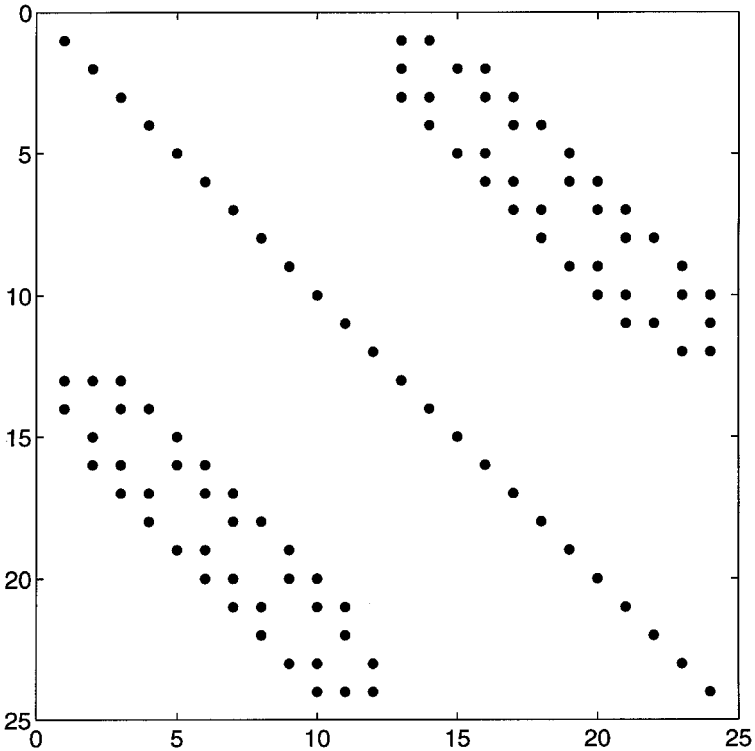

Figure 2. Structure of the banded matrix used in the example; dots indicate non-zero elements.

At the same time, the complexity of the equivalent explicit filter always remains of order n^3.

Consider simulation results for the described implicit system. We assume that estimates generated by the traditional filter are, in essence, optimal (a justified assumption, since matrix $\mathbf{A}_1(k)$ is chosen to be well-conditioned), and use them as a comparison basis for the implicit filter estimates. Figure 3 shows the relative estimation error of filtered measurements calculated as

$$\Delta \hat{z} = \frac{\mathbf{H}(\hat{\mathbf{x}}_{ex} - \hat{\mathbf{x}}_{im})}{\mathbf{H}\hat{\mathbf{x}}_{ex}}$$

where $\hat{\mathbf{x}}_{ex}$ and $\hat{\mathbf{x}}_{im}$ are estimates produced by the explicit and the implicit filters, respectively. The different lines correspond to the application of three different iterative schemes, discussed in § 4. For each of them, a *single* iteration (35) was used to find the solution of matrix equation (18). Note that the rates of convergence according to equation (44) and presented in Table 1, give the correct prediction as to which iterative scheme is converging faster.

Figure 4 illustrates the effect of an increased number of iterations m on the accuracy of the iterative solution of the matrix equation (18). It shows a typical behaviour of the relative error in the calculation of the components of the covariance matrix $\mathbf{P}^x_{k+1|k+1}$ when two different iterative schemes are used with a different number

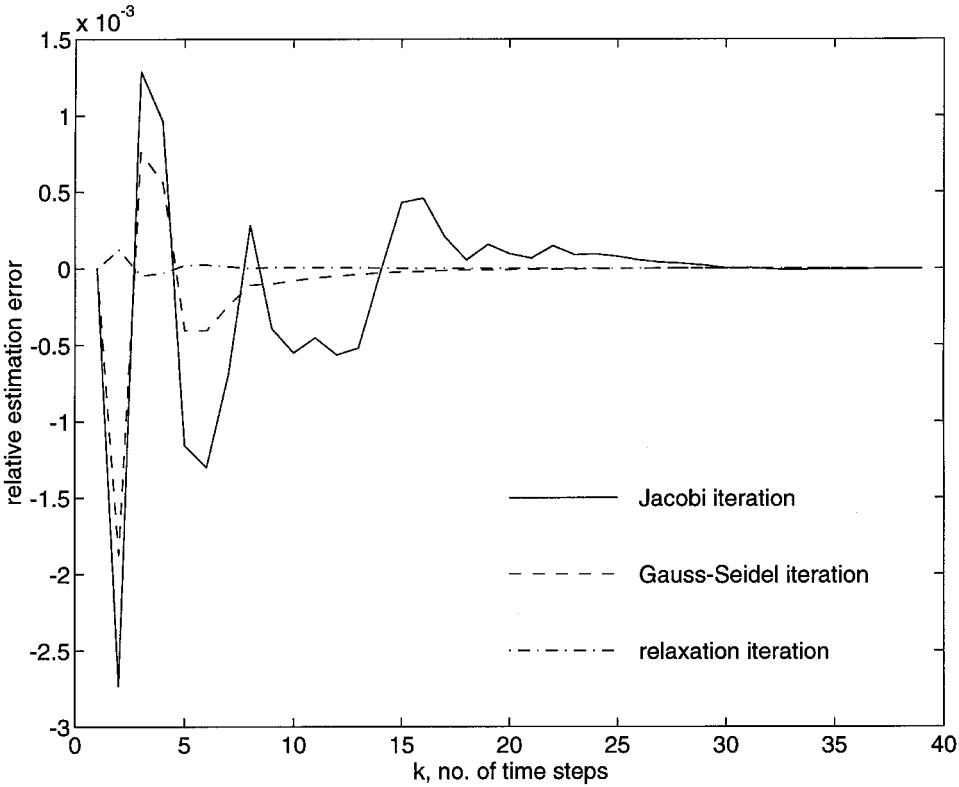

Figure 3. Relative estimation error of the implicit filter.

Method	Rate of convergence, R
Jacobi	0·76
Gauss-Seidel	2·40
Relaxation, $\omega = 1{\cdot}02$	3·03

Table 1. Rates of convergence.

of iterations on each time step. It also supports the conclusion drawn in § 4, that there is no accumulation of errors nor bias due to the approximate nature of the iterative solution of (18).

The dependence of the estimation error on the number of iterations, used to solve (18), is best seen in Fig. 5, where the cumulative computational error

$$CE = \sum_{k=0}^{\infty} \left| \left[\mathbf{P}^{x}_{k+1|k+1} \right]_{8,18} - \left[\mathbf{P}^{x}_{k+1|k+1} \right]^{m)}_{8,18} \right|$$

for a particular diagonal element of the error covariance matrix is plotted as a function of m. Similar dependence is observed for all components of $\mathbf{P}^{x}_{k+1|k+1}$, which suggests an exponential rate of convergence of the iterative method (35).

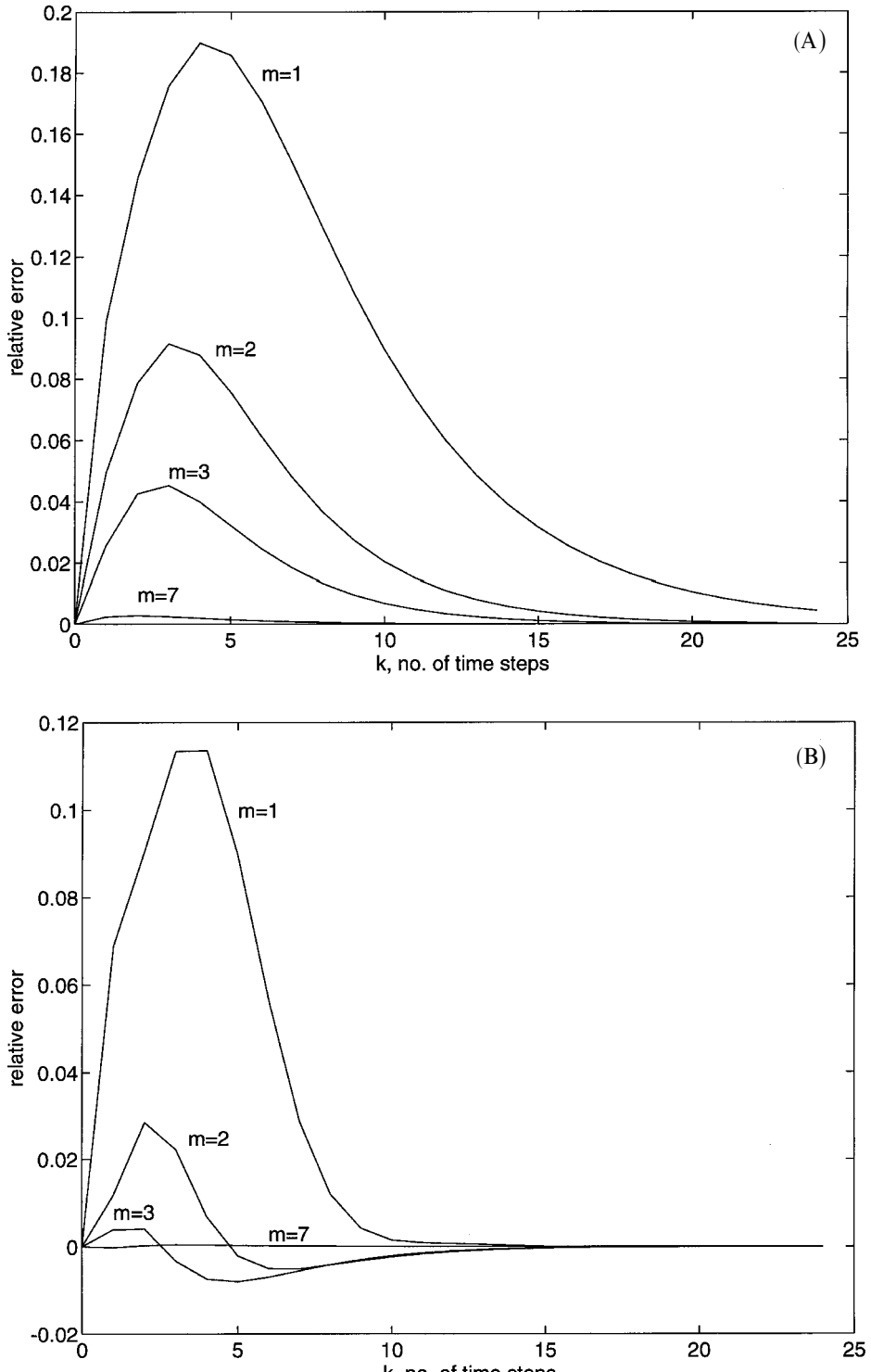

Figure 4. Relative error in the calculation of the component $\left[\mathbf{P}^{x}_{k+1|k+1}\right]_{18,18}$, using (A) Jacobi and (B) Gauss–Seidel iterative schemes; m is the number of iterations.

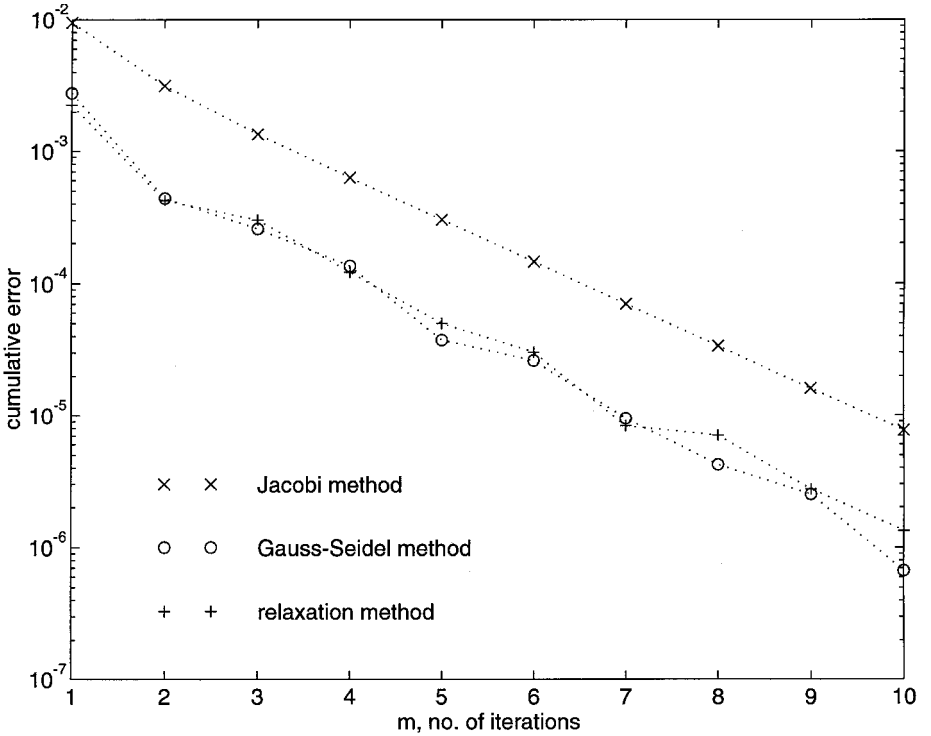

Figure 5. Cumulative error in the calculation of the 18th diagonal element of matrix $\mathbf{P}^{x}_{k+1|k+1}$ as a function of a number of iterations m in the iterative process (35).

7. Conclusions

In this paper, we develop a Kalman filter for the implicit discrete system (1) which does not require computationally expensive and potentially troublesome operation of the inversion of the matrix $\mathbf{A}_1(k)$. The implicit Kalman filter is particularly applicable to an ill-conditioned system (1), where the traditional approach fails.

The implicit filter preserves the structure of matrices $\mathbf{A}_1(k)$ and $\mathbf{A}_2(k)$ allowing for an application of structure specific methods of matrix manipulation. This allows for a potentially significant improvement of the computational efficiency of the filtering process when the matrices of system (1) are sparse.

An algorithm which realizes the advantages of the implicit Kalman filter is also proposed. It is shown that in the case of the banded matrices $\mathbf{A}_1(k)$ and $\mathbf{A}_2(k)$ the implicit filter requires $O(n^2)$ flops on each time step, while the traditional approach remains of order $O(n^3)$ and requires substantially more computer storage. The high efficiency of the proposed algorithm allows us to recommend the application of the implicit filter even when the implicit banded system is time invariant. It can also be used to calculate the steady state error covariance $\mathbf{P}^{x}_{ss}$ and the steady state Kalman gains.

The implicit filter is also shown to be competitive in the case of a general time varying implicit system (1). However, if the matrix $\mathbf{A}_1$ is stationary and not sparse, the application of the implicit filter can be recommended over the explicit approach only if a steady state is quickly reached.

ACKNOWLEDGMENT

The authors acknowledge the support of the NASA Specialized Center for Research and Training on Environmental Health under grant No. 89988–G004, and NASA support under grant No. NRA93–OLMSA–07.

Appendix

Suppose the following two-dimensional partial differential equation is used as a model of the process:

$$\frac{\partial x(t,\xi)}{\partial t} = \mathcal{L}_\xi x(t,\xi) + f(t,\xi) + C(t,\xi)w(t,\xi) \tag{A 1}$$

$$\mathcal{L}_\xi \equiv \mathcal{D}(t,\xi)\left[\frac{\partial^2}{\partial\xi_1^2} + \frac{\partial^2}{\partial\xi_2^2}\right] - u(t,\xi)\frac{\partial}{\partial\xi_1} - v(t,\xi)\frac{\partial}{\partial\xi_2}$$

$$\xi = (\xi_1, \xi_2) \in \Omega, \quad t \in [0, T]$$

$$x(0,\xi) = x_0(\xi)$$

$$\mathcal{L}_{\xi b} x(t,\xi) = C_b(t,\xi)w_b(t,\xi), \quad \xi \in \partial\Omega$$

where the state variable x depends on time t and spatial location ξ within a domain Ω with a boundary $\partial\Omega$. The coefficients $\mathcal{D}$, u and v are (in general) time dependent and presumed to be known. System (A 1) is driven by deterministic (and known) source function f and by a stochastic disturbance w. An operator $\mathcal{L}_{\xi b}$ is used to specify boundary condition; x_0 is an initial condition.

System (A 1) is assumed to be observable and the measurements z are related to the state by

$$z(t,\xi) = H(t,\xi)x(t,\xi) + v(t,\xi)$$

where H is a measurements operator and v is a measurement noise.

Stochastic perturbations w, w_b and v are assumed to be uncorrelated Gaussian processes with zero means and known covariances.

In order to develop a discrete analogue of system (A 1) we first find a finite difference approximation of the operator $\mathcal{L}_\xi$. For the simulation example we have used a five-point stencil approximation of the second-order spatial derivative and upwind differencing of the first order spatial derivative. The application of the Crank–Nicolson scheme for time derivative approximation gives us a discrete model of the process in the form of equation (1) with

$$\mathbf{A}_1(k) = \left(\mathbf{I} - \frac{\Delta t}{2}\mathbf{A}(k)\right)$$

$$\mathbf{A}_2(k) = \left(\mathbf{I} + \frac{\Delta t}{2}\mathbf{A}(k)\right)$$

where $\mathbf{A}$ is a discrete analogue of $\mathcal{L}_\xi$, and Δt is a discretization step in the time domain.

The appearance of the matrices $\mathbf{A}_1(k)$ and $\mathbf{A}_2(k)$ also depends on the ordering of the equations in (1). The alternating diagonals, or D4 ordering (Aziz and Settari, 1979) produces $\mathbf{A}_1(k)$ and $\mathbf{A}_2(k)$ in the form shown in Figure 2.

The simulation of § 6 was conducted with $\mathcal{D} = 10$, $u = 0$, $\nu = -1$ and no-flux boundary conditions. The discretization steps in time and space are $\Delta t = 30$ and $\Delta \xi_1 = \Delta \xi_2 = 50$, and the size of the rectangular domain is 200 by 300. These data are used to obtain a discrete analogue of (48), which has a dimension $n = 24$.

We further assume that there is a single point-wise source applied to the system, such that

$$\mathbf{B} = \overbrace{[0 \ldots 0\,0\,0\,0\,1\,0\,0\,0]}^{24 \text{ elements}}{}^{\mathrm{T}},$$

and $\boldsymbol{u}\Delta \xi_1 \Delta \xi_2 = 600$.

A single point-wise sensor is used to generate measurements (4), where

$$\mathbf{H} = \overbrace{[0 \ldots 0\,1\,0\,0\,0\,0\,0\,0]}^{24 \text{ elements}}$$

Zero mean Gaussian stochastic model disturbance has a variance $E[\boldsymbol{w}(k)\boldsymbol{w}(k)^{\mathrm{T}}] = 40\mathbf{I}$ and $\mathbf{C}(k) = \mathbf{I}$. The measurement noise is assumed to have a variance $E[v(k+1)v(k+1)^{\mathrm{T}}] = 4$ and a zero mean.

The structure specific method used to solve linear systems (12) and (21) is described in Aziz and Settari (1979, pp. 258-259).

References

Adams, M. B., Levy, B. C., and Willsky, A. S., 1984, Linear smoothing for descriptor systems. *Proceedings of the Conference on Decision and Control*, Las Vegas, Nevada, U.S.A., pp. 1–6.

Aplevich, J. D., 1991, *Implicit Linear Systems* (Berlin: Springer-Verlag).

Aziz, K., and Settari, A., 1979, *Petroleum Reservoir Simulation* (London, U.K.: Applied Science Publishers).

Bender, D. J., and Laub, A. J., 1987a, The linear-quadratic optimal regulator for descriptor systems: discrete-time case. *Automatica*, **23**, 71–85; 1987b, The linear-quadratic optimal regulator for descriptor systems. *IEEE Transactions on Automatic Control*, **32**, 672–688.

Campbell, S. L., and Meyer, C. D., 1979, *Generalized Inverses of Linear Transformations* (London, U.K.: Pitman).

Carlson, N. A., 1990, Federated square root filter for decentralized parallel processes. *IEEE Transactions on Aerospace and Electronic Systems*, **26**, 517–525.

Chin, T. M., Karl, W. C., and Willsky, A. S., 1995, A distributed and iterative method for square root filtering in space-time estimation. *Automatica*, **31**, 67–82.

Chisci, L., and Zappa, G., 1992, Square-root Kalman filtering of descriptor systems. *Systems & Control Letters*, **19**, 325–334.

Chu, K.-W. E., 1990, On linear matrix equations. *Computational Techniques and Applications: CTAC-89*, edited by W. L. Hogarth and B. J. Noye (New York: Hemisphere Publishing) 591–598.

Godbout, L. F., and Jordan, D., 1975, On state equation descriptions of linear differential systems. *Journal of Dynamic Systems, Measurement and Control*, **97**, 333–344.

Golub, G. H., and Van Loan, C. F., 1989, *Matrix Computations*, second edition (Baltimore, Maryland, U.S.A.: Johns Hopkins University Press).

Horn, R. A., and Johnson, C. R., 1985, *Matrix Analysis* (Cambridge, U.K.: Cambridge University Press).

Jordan, T. J., 1967, Experiments on error growth associated with some linear least-squares procedures. *Mathematics of Computation*, **22**, 579–588.

Jover, J. M., and Kailath, T., 1986, A parallel architecture for Kalman filter measurement update and parameter estimation. *Automatica*, **22**, 43–57.

Kalman, R. E.,1960, A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, **82**, 35–45.

KALMAN, R. E., and BUCY, R. S., 1961, New results in linear filtering and prediction theory. *Journal of Basic Engineering*, **83**, 95–100.

LEWIS, F. L., 1986, A survey of linear singular systems. *Journal of Circuits Systems Signal Processing*, **5**, 3–36.

LEWIS, F. L., and MERTZIOS, B. G., 1992, On the analysis of two-dimensional discrete singular systems. *Journal of Circuits Systems Signal Processing*, **11**, 399–419.

LUENBERGER, D. G., 1977, Dynamic equations in descriptor form. *IEEE Transactions on Automatic Control*, **22**, 312–321.

MEDITCH, J. S., 1969, *Stochastic Optimal Linear Estimation and Control* (New York: McGraw-Hill).

MORF, M., and KAILATH, T., 1975, Square-root algorithms for least-squares estimation. *IEEE Transactions on Automatic Control*, **20**, 487–497.

NEWCOMB, R. W., and DZIURLA, B., 1989, Some circuits and system applications of semistate theory. *Journal of Circuits Systems Signal Processing*, **8**, 235–260.

NIKOUKHAH, R., WILLSKY, A. S., and LEVY, B. C., 1990, Kalman filtering and Riccati equations for descriptor systems. *Proceedings of the Conference on Decision and Control*, Honolulu, Hawaii, U.S.A., pp. 2886–2891; 1992, Kalman filter and Riccati equations for descriptor systems. *IEEE Transactions on Automatic Control*, **37**, 1325–1342.

PAIGE, C. C., and SAUNDERS, M. A., 1977, Least squares estimation of discrete linear dynamic systems using orthogonal transformations. *SIAM Journal of Numerical Analysis*, **14**, 180–193.

POLAK, E., 1966, An algorithm for reducing a linear, time-invariant system to state form. *IEEE Transactions on Automatic Control*, **11**, 577–579.

ROSENBROCK, H. H., 1970, *State Space and Multivariable Theory* (New York: Wiley).

ROY, S., HASHEMI, R. H., and LAUB, A. J., 1991, Square root parallel Kalman filtering using reduced-order local filters. *IEEE Transactions on Aerospace and Electronic Systems*, **27**, 276–289.

SHIN, K.-C., and KABAMBA, P. T., 1988, Observation and estimation in linear descriptor systems with application to constrained dynamical systems. *Journal of Dynamic Systems, Measurement, and Control*, **110**, 255–265.

SKLIAR, M., and RAMIREZ, W. F., 1995, Kalman filter for discrete implicit systems. *Proceedings of the American Control Conference*, Seattle, Washington, U.S.A., pp. 524–528; 1996 a, Square root implicit Kalman filtering. *Proceedings of the 13th IFAC World Congress*, San Francisco, California, U.S.A., Vol. J., pp. 251–256; 1996 b, Air quality monitoring and early detection of air contamination events in an enclosed environment. *Journal of Spacecraft and Rockets*, submitted for publication.

THORNTON, C. L., and BIERMAN, G. J., 1980, UDU^{T} covariance factorization for Kalman filtering. *Control and Dynamic Systems*, edited by C. T. Leondes, **16**, 178–248.

VARGA, R. S., 1962, *Matrix Iterative Analysis* (Englewood Cliffs, New Jersey, U.S.A.: Prentice Hall).

WANG, X.-M., and BERNHARD, P., 1991, Filtrage et lissage des systèmes implicites discrets. *RAIRO APII, Automatique Productique Informatique Industrielle, Paris*, **25**, 205–228.

WOLOVICH, W. A., 1974, *Linear Multivariable Systems* (New York: Springer-Verlag).

ZGUROVSKY, M. Z., and SKLIAR, M. V., 1993 a, Multiscale expansion of the solution of a singularly perturbed system of evolutional operator equations—Part 1. *Soviet Journal of Automation and Information Sciences*, **24** (5), 61–71; 1993 b, Multiscale expansion of the solution of a singularly perturbed system of evolutional operator equations—Part 2. *Ibid.*, **24** (6), 18–29.