FAST INVERSION OF L-BLOCK BANDED MATRICES AND THEIR INVERSES

Amir Asif

Information Technology
Technical University of British Colombia
Surrey, BC V3R 7P8, Canada
amir.asif@techbc.ca

José M. F. Moura

Electrical and Computer Engineering
Carnegie Mellon University
Pittsburgh, PA 15213, U.S.A.
moura@ece.cmu.edu

ABSTRACT

Block banded matrices generalize banded matrices. In this paper, we exploit properties of full matrices whose inverses are L-block banded to derive fast inverses for such matrices, and inverses for matrices that are themselves block banded. We apply these results to design fast implementations for the Kalman-Bucy filter in applications arising often in the physical sciences where the underlying models derive from discretizations of partial differential equations and the observations are sparse.

1. INTRODUCTION

Block banded matrices are used in autoregressive or moving average image models, describe the covariance of Gauss Markov random processes (GMrp), and occur with finite difference representations of partial differential equations. We derive from the properties of these matrices two fast inversion algorithms:

1. to invert a matrix whose inverse is L-block banded;
2. and to invert an L-block banded matrix.

Our algorithms show that the inverse of an L-block banded matrix is completely specified by its first L diagonal blocks; the block entries outside the first L diagonal blocks are determined by the blocks within the first L block-diagonals. These algorithms lead to two orders of magnitude savings over the direct inversion of matrices. Our results differ from [1] where the constituent blocks in the matrices are scalars, and they generalize [2], which considers only tridiagonal ($L = 1$) block matrices. We apply these results to designing fast implementations that approximate well the exact Kalman-Bucy filter (KBf). These near-optimal implementations, referred to as the local KBfs, impose a block-banded structure on the inverse of the error-covariance matrix (information matrix). Banded approximations to information matrices correspond to modeling the error field as a reduced order GMrp.

The paper is organized as follows. Section 2 presents three important properties for L-block banded matrices. Section 3 derives the inversion algorithms, while section 4 describes the local KBf and a frame reconstruction simulation to compare the local KBf with the optimal KBf. In section 5, we conclude the paper.

2. BANDED MATRIX RESULTS

Consider a positive-definite matrix $\mathcal{P}$ that has an L-block banded inverse $\mathcal{A}$. The matrix $\mathcal{P}$ is referred to as the covariance matrix, and its inverse $\mathcal{A}$ is the information matrix, or the potential matrix.

Both $\mathcal{P}$ and $\mathcal{A}$ are represented respectively in terms of their $(I \times I)$ constituent blocks $\{P_{ij}\}$ and $\{A_{ij}\}$, for $1 \leq i, j \leq J$. Since $\mathcal{A}$ is L-block banded, any block entries outside the L-upper and L-lower diagonals in $\mathcal{A}$ are zero blocks, i.e., $A_{ij} = \underline{0}$ for $|i-j| > L$. The overall dimension of both $\mathcal{A}$ and $\mathcal{P}$ is $(IJ \times IJ)$.

The Cholesky factorization of $\mathcal{A} = \mathcal{U}^T\mathcal{U}$, where $\mathcal{U} = \{U_{ij}\}$, is a block upper triangular matrix with exactly L non-zero block diagonals, i.e., $U_{ij} = \underline{0}$ for $(j - i) > L$. Its inverse $\mathcal{U}^{-1}$ is also block upper triangular

$$\mathcal{U}^{-1} = \begin{bmatrix} U_{11}^{-1} & \underline{*} & \underline{*} & \underline{*} & \cdot & \underline{*} \\ \underline{0} & U_{22}^{-1} & \underline{*} & \underline{*} & \cdot & \underline{*} \\ \underline{0} & \underline{0} & U_{33}^{-1} & \underline{*} & \cdot & \underline{*} \\ \cdot & \cdot & \ddots & \ddots & \cdot & \cdot \\ \underline{0} & \cdot & \cdot & \underline{0} & U_{I-1I-1}^{-1} & \underline{*} \\ \underline{0} & \cdot & \cdot & \underline{0} & \underline{0} & U_{II}^{-1} \end{bmatrix}. \quad (1)$$

The lower diagonal entries in $\mathcal{U}^{-1}$ are all zero blocks. More importantly, the diagonal entries in $\mathcal{U}^{-1}$ are the block inverses of the corresponding blocks in $\mathcal{U}$. These features are used next to derive three important results for L-block banded matrices.

Theorem 1 *The Cholesky's blocks $\{U_{ii}, \ldots, U_{ii+L}\}$'s and the covariance blocks P_{ij} for $1 \leq i \leq (J - L)$ are related by*

$$\begin{bmatrix} P_{ii} & \cdot & \cdot & P_{ii+L} \\ P_{i+1i} & \cdot & \cdot & P_{i+1i+L} \\ \cdot & \cdot & \ddots & \cdot \\ P_{i+Li} & \cdot & \cdot & P_{i+Li+L} \end{bmatrix} \begin{bmatrix} U_{ii}^T \\ U_{ii+1}^T \\ \vdots \\ U_{ii+L}^T \end{bmatrix} = \begin{bmatrix} U_{ii}^{-1} \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix}. \quad (2)$$

The boundary conditions (b.c.) are

$$U_{JJ}^T U_{JJ} = P_{JJ}^{-1}$$

$$\begin{bmatrix} P_{J-1J-1} & P_{J-1J} \\ P_{JJ-1} & P_{JJ} \end{bmatrix} \begin{bmatrix} U_{J-1J-1}^T \\ U_{J-1J}^T \end{bmatrix} = \begin{bmatrix} U_{J-1J-1}^{-1} \\ \underline{0} \end{bmatrix}$$

$$\cdots$$

$$\begin{bmatrix} P_{J-LJ-L} & \cdot & \cdot & P_{J-LJ} \\ P_{J-L+1J-L} & \cdot & \cdot & P_{J-L+1J} \\ \cdot & \cdot & \ddots & \cdot \\ P_{JJ-L} & \cdot & \cdot & P_{JJ} \end{bmatrix} \begin{bmatrix} U_{J-LJ-L}^T \\ U_{J-LJ-L+1}^T \\ \vdots \\ U_{J-LJ}^T \end{bmatrix} = \begin{bmatrix} U_{J-LJ-L}^{-1} \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix}$$

Proof: Theorem 1 is proven by induction.
($L = 1$): For $L = 1$, theorem 1 reduces to

$$\begin{bmatrix} P_{ii} & P_{ii+1} \\ P_{i+1i} & P_{i+1i+1} \end{bmatrix} \begin{bmatrix} U_{ii}^T \\ U_{ii+1}^T \end{bmatrix} = \begin{bmatrix} U_{ii}^{-1} \\ \underline{0} \end{bmatrix} \quad (3)$$

for $1 \leq i \leq (J-1)$ with b.c. $U_{JJ}^T U_{JJ} = P_{JJ}^{-1}$. This is the tridiagonal block banded matrix proven in [2].
($L = k$): By the induction step, theorem 1 is valid for a k-banded matrix with $k < L$.
($L = k+1$): From the equality

$$\mathcal{P} = (\mathcal{U}^T \mathcal{U})^{-1}, \text{ we get } \mathcal{P}\mathcal{U}^T = \mathcal{U}^{-1} \tag{4}$$

where we replace $\mathcal{U} = \{U_{ij}\}$, substitute $\mathcal{U}^{-1}$ from (1), and express $\mathcal{P}$ as $\{P_{ij}\}$. Then we multiply out the left hand side of (4) and equate the diagonal and lower diagonal block entries. Equating the (J, J) block elements results in the first b.c. for theorem 1. Equating the $(J-1, J-1)$ and the $(J, J-1)$ block elements in (4) gives

$$P_{J-1J-1} U_{J-1J-1}^T + P_{J-1J} U_{J-1J}^T = U_{J-1J-1}^{-1} \tag{5}$$

$$P_{JJ-1} U_{J-1J-1}^T + P_{JJ} U_{J-1J}^T = \underline{0} \tag{6}$$

that prove the second b.c. for theorem 1. The remaining equations are proved similarly. For example, equation (2) follows by equating the (i, i), $(i+1, i)$, ..., $(i+k+1, i)$ block elements on both sides of (4). ∎
Theorem 1 illustrates a special relationship between the blocks $\{U_{ij}\}$ of the Cholesky factor $\mathcal{U}$ and the blocks $\{P_{ij}\}$ of the covariance matrix $\mathcal{P}$. Evaluating the Cholesky blocks $\{U_{ii} \ldots U_{ii+L}\}$ on row i of $\mathcal{U}$ does not involve the entire matrix $\mathcal{P}$, only the following L^2 blocks in its neighborhood

$$\begin{bmatrix} P_{ii} & P_{ii+1} & \cdot & P_{ii+L} \\ P_{i+1i} & P_{i+1i+1} & \cdot & P_{i+1i+L} \\ & & \ddots & \\ \cdot & \cdot & & \cdot \\ P_{i+Li} & P_{i+Li+1} & \cdot & P_{i+Li+L} \end{bmatrix}$$

Theorem 2 *The blocks P_{ij} in the covariance matrix $\mathcal{P}$ relate recursively to the Cholesky blocks $\{U_{ii}, \ldots, U_{ii+L}\}$ by*

$$P_{ii} = (U_{ii}^T U_{ii})^{-1} - \sum_{\ell=i+1}^{min(J,i+L)} P_{i\ell} \, (U_{ii}^{-1} U_{i\ell})^T \tag{7}$$

$$P_{ij} = - \sum_{\ell=i+1}^{min(J,i+L)} (U_{ii}^{-1} U_{i\ell}) \, P_{\ell j} \tag{8}$$

for $1 \leq i \leq (J-1)$, $i \leq j \leq (i+L) \leq J$, and with b.c.

$$P_{JJ} = (U_{JJ}^T U_{JJ})^{-1}. \tag{9}$$

Proof: Theorem 2 is proved by induction.
($L = 1$): For $L = 1$, theorem 2 reduces to

$$P_{ii} = (U_{ii}^T U_{ii})^{-1} - P_{ii+1} \, (U_{ii}^{-1} U_{ii+1})^T \tag{10}$$

$$P_{ij} = -(U_{ii}^{-1} U_{ii+1}) \, P_{i+1j} \tag{11}$$

for $1 \leq i \leq (J-1)$ with b.c. $P_{JJ} = (U_{JJ}^T U_{JJ})^{-1}$. This is proven for tridiagonal block-banded matrices in [2].
($L = k$): By the induction step, theorem 1 is valid for a k-block banded matrix with $k < L$.
($L = k+1$): The b.c. $P_{JJ} = (U_{JJ}^T U_{JJ})^{-1}$ is proven directly by rearranging terms in the first b.c. in theorem 1.
Equations (7) and (8) are proven directly by expanding (2) for $L = k+1$, which results in $k+1$ simultaneous expressions. Equation (7) results by rearranging terms in the first expression. The remaining expressions in (2) are special cases of (8) for $j = i+1, \ldots, j = (i+k+1)$. ∎
Theorem 2 states that the blocks $P_{ii} \ldots P_{ii+L}$ on row i within the first L-block diagonals in $\mathcal{P}$ can be evaluated from the corresponding Cholesky blocks $U_{ii} \ldots U_{ii+L}$ on block row i of $\mathcal{U}$ and the L-banded blocks in the lower block rows of $\mathcal{P}$, i.e., $P_{mm} \ldots P_{mm+L}$, with $m > i$.

Theorem 3 *Given blocks $\{P_{ij}\}$ on the first L-block diagonals of $\mathcal{P}$, the upper triangular blocks beyond the L'th-block diagonal are given by*

$$P_{ii+k} = [P_{ii+1} \ldots P_{ii+L}] \cdot \begin{bmatrix} P_{i+1i+1} & \cdot & P_{ii+L} \\ & \ddots & \\ P_{i+Li+1} & \cdot & P_{i+Li+L} \end{bmatrix}^{-1} \begin{bmatrix} P_{i+1i+k} \\ \vdots \\ P_{i+Li+k} \end{bmatrix} \tag{12}$$

for $1 \leq i < (J-L)$ and $L < k \leq (J-i)$.

Proof: Theorem 3 is proven by induction from theorems 1 and 2. The case $L = 1$ has been proven in [2]. ∎
Theorem 3 states that the matrix $\mathcal{P}$ is completely specified by its first L-block diagonals. Any blocks outside the L-block diagonals can be evaluated directly from blocks within the L-block diagonals using theorem 3.

3. INVERTING L-BLOCK BANDED MATRICES

In this section, theorems 1 and 2 are used to derive algorithms to invert a full matrix with an L-block banded inverse and to solve the converse problem of inverting an L-block banded matrix.

3.1. Algorithm 1

In terms of the earlier notation, Algorithm 1 computes the potential matrix $\mathcal{A}$ from blocks P_{ij} of of the covariance matrix $\mathcal{P}$ in two steps.
Step 1: Starting with the J'th row, calculate the Cholesky's blocks $\{U_{ii}, \ldots, U_{ii+L}\}$, are calculated recursively using theorem 1. The blocks on the i'th row, for example, are calculated by rearranging (2) as

$$\begin{bmatrix} P_{ii} & \cdot & \cdot & P_{ii+L} \\ P_{i+1i} & \cdot & \cdot & P_{i+1i+L} \\ & & \ddots & \\ \cdot & \cdot & & \cdot \\ P_{i+Li} & \cdot & \cdot & P_{i+Li+L} \end{bmatrix} \begin{bmatrix} U_{ii}^T U_{ii} \\ U_{ii+1}^T U_{ii} \\ \vdots \\ U_{ii+L}^T U_{ii} \end{bmatrix} = \begin{bmatrix} I_I \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix} \tag{13}$$

for $(J-L) \geq i \geq 1$. The block I_I is the identity matrix of order I. The main diagonal Cholesky blocks $\{U_{ii}\}$ are now obtained by solving for the Cholesky factors of $\{U_{ii}^T U_{ii}\}$. The off diagonal Cholesky blocks are evaluated by multiplying the corresponding entity calculated in (13) with the inverse of $\{U_{ii}\}$.
Step 2: The upper triangular block entries in the information matrix $\mathcal{A}$ is determined from the following expression

$$A_{ij} = \begin{cases} \sum_{\ell=\max(1,j-L)}^{i} U_{\ell i}^T U_{\ell j} & \text{for } (j-i) \leq L \\ \underline{0} & \text{for } (j-i) > L \end{cases} \tag{14}$$

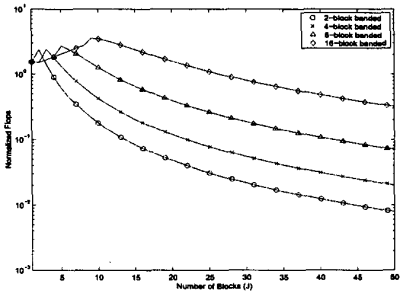

Fig. 1. Normalized flops required to invert a full matrix $\mathcal{P}$ with L-block banded inverse using Algorithm 1 for $L = 2, 4, 8$ and 16.

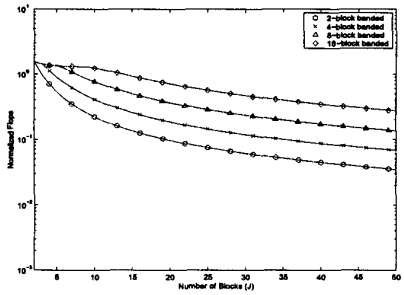

Fig. 2. Normalized flops required to invert an L-block banded $(IJ \times IJ)$ matrix $\mathcal{A}$ using Algorithm 2 for $L = 2, 4, 8$ and 16.

Equation (14) is obtained by expanding the expression $\mathcal{A} = \mathcal{U}^T\mathcal{U}$ in terms of the constituent blocks in $\mathcal{A}$ and $\mathcal{U}$.
By counting the number of operations, we verify that the above algorithm is of order $O(LJI^3)$. For small values of L, this implies an improvement of a factor of J^2/L over direct inversion.

3.2. Algorithm 2

This algorithm calculates the covariance matrix $\mathcal{P}$ from its L-block banded inverse $\mathcal{A}$ in two steps.
Step 1: Calculate the Cholesky blocks $\{U_{ij}\}$ from $\mathcal{A}$. These are evaluated recursively by

$$U_{ii} = \text{chol}\Big(A_{ii} - \sum_{\ell=\max(1,i-L)}^{i-1} \big(U_{\ell i}^T U_{\ell i}\big)\Big) \tag{15}$$

$$U_{ii+k} = U_{ii}^{-T}\Big(A_{ii+k} - \sum_{\ell=\max(1,i-L)}^{i-1} U_{\ell i}^T U_{\ell i+k}\Big) \tag{16}$$

for $2 \le i \le J$ and $1 \le k \le L$. The b.c. for the first row $i = 1$ is

$$U_{11} = \text{chol}\big(A_{11}\big) \quad \text{and} \quad U_{11+k} = U_{11}^{-T} A_{11+k} \tag{17}$$

for $1 \le k \le L$. Equations (15-17) are derived by rearranging terms in (14).
Step 2: Starting with P_{JJ}, the block entries $\{P_{ij}\}$ $J \ge i \ge 1$, $J \ge j \le i$ in the covariance matrix $\mathcal{P}$ are now determined recursively using theorem 2 by substituting the values for the Cholesky blocks $\{U_{ij}\}$ calculated in step 1.
A concern in (15) is the right hand term that must be positive definite for Cholesky factorization. Although not included here, it can be shown that the right hand term is indeed positive definite. Alternatively, the Cholesky factorization step is avoided by noting that theorem 2 requires products $(U_{ii}^T U_{ii})$ and $(U_{ii}^{-1} U_{ii+k})$. We solve (15) for these product terms, thus avoiding performing Cholesky factorization. The complexity of algorithm 2 is on the order of $O(LJ^2I^3)$, an improvement of a factor of J/L over direct inversion.

3.3. Simulations

In Figs. 1 and 2, we plot results of Monte Carlo simulations run to quantify savings in floating point operations (flops) resulting from Algorithms 1 and 2 over direct inversion. The plots are normalized by the total number of flops required in direct inversion, therefore, the region below the ordinate $y = 1$ represents computational savings. In each case, the dimension I of the constituent blocks $\{P_{ij}\}$ in $\mathcal{P}$ (or of $\{A_{ij}\}$ in $\mathcal{A}$) is kept constant at $I = 5$ while the parameter J denoting the number of $(I \times I)$ blocks on the main diagonal in $\mathcal{P}$ (or $\mathcal{A}$) is varied from 1 to 50. The maximum dimension of matrices $\mathcal{A}$ and $\mathcal{P}$ in the simulation is (250×250). Except for the initial cases where the overhead involved in indexing and identifying constituent blocks exceeds the savings, both algorithms exhibit considerable savings over direct inversion. For Algorithm 1, the computations are reduced by a factor of 10 - 100 while in algorithm 2, the savings is by a factor of about 10. We expect higher savings for larger matrices.

4. APPLICATION TO KALMAN-BUCY FILTERING

For typical image-based applications in computer vision and the physical sciences, the visual fields can normally be specified by spatially local interactions. In other cases, these fields are modeled by finite difference equations obtained from discretizing partial differential equations. Consequently, the state matrices, $\mathcal{C}$ and $\mathcal{D}$, in the state equation (with forcing term $\mathcal{W}$)

$$\psi^{(k+1)} = \mathcal{C}^{(k)}\psi^{(k)} + \mathcal{D}^{(k)}\mathcal{W}^{(k)} \tag{18}$$

are block banded and sparse, i.e., $C_{ij} = \underline{0}$ for $|(i-j)| > L$. A similar structure exists for $D = \{D_{ij}\}$. The dimension of the state vector ψ obtained by lexicographic ordering the field is on the order of the number of pixels in the field, typically 10^4 to 10^6 elements. Due to the large dimensionality, it is only practical to observe only a portion of the field. The observations $\mathcal{Y}$ in the observation model with noise $\mathcal{E}$

$$\mathcal{Y}^{(k)} = \mathcal{H}^{(k)}\psi^{(k)} + \mathcal{E}^{(k)} \tag{19}$$

are therefore also fairly sparse. Implementation of optimal filters in such cases requires storage and manipulation of $10^4 \times 10^4$ to $10^6 \times 10^6$ matrices. As a result, practical implementations require some sort of approximate sub-optimal scheme. We describe such practical implementations for the Kalman-Bucy filter (KBf).

Due to the banded structure of the state and observation equations, it is possible to express the KBf in terms of the constituent blocks C_{ij} and D_{ij} of the state and observation matrices. As a result, updating a portion (say row i) of the field ψ at instant $k+1$ requires only local information in the neighborhood of the field at instant k and knowledge of a few selected blocks P_{ij} in the covariance matrix $\mathcal{P}$. In our implementation of the KBf (local KBf), we

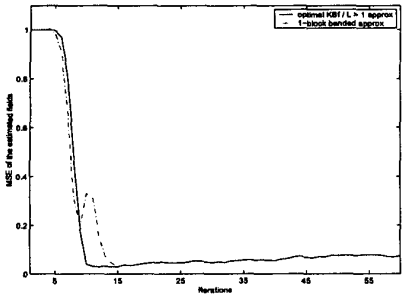

Fig. 3. Comparison of MSEs (normalized with the actual field's energy) of the optimal KBf versus the local KBfs with $L = 1$ to 4. The plots for $L > 1$ overlap the plot for the optimal KBf.

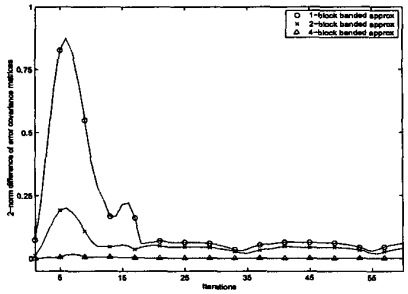

Fig. 4. Comparison of 2-norm differences between error covariance matrices of the optimal KBf and local KBfs with $L = 1$ to 4.

use this feature to our advantage and update only a few diagonals of the error covariance matrix. The exact number of updated diagonals depends on the level of the approximation. For a 1-block banded approximation, only the main diagonal P_{ii} entries and the first upper diagonal entries P_{ii+1} are updated. In general, for an L-block banded approximation, the blocks on the main diagonal and the first L upper diagonals are updated. Any other blocks P_{ij}'s if required are obtained from the updated blocks using theorem 3. Since fewer blocks are being updated, the local KBf requires less computations than the optimal KBf. For a $(J \times J)$ field, the savings is on the order of $O(J^2/L)$. Our local approximation corresponds to modeling the error field in the spatial dimensions as a reduced order acausal GMrp. It is shown in [1] that the GMrp approximations optimize the Kullback-Leibler mean information distance criterion.

4.1. Numerical Results

To illustrate the effect of our approximations, we apply the local KBf to a dynamical frame reconstruction problem, [3], where a sequence of (100×100) images of the moving tip of a quadratic cone are synthesized. The covariance matrix has dimensions of $(10^4 \times 10^4)$. The surface $\psi(t)$ translates across the image frame with a constant velocity whose components along the two frame axes are both 0.2 pixels/frame, i.e.,

$$\psi(s_1, s_2, k+1) \quad = \quad \psi(s_1 + 0.2, s_2 + 0.2, k). \qquad (20)$$

Since the spatial coordinates (s_1, s_2) take only integer values in the discrete dynamical model on which the filters are based, we use a finite difference model obtained by discretizing (20) with the leap-frog method. The dynamical equation in (20) is a simplified case of the thin-plate model with the spatial coherence constraint suitable for surface interpolation, [3]. The observation model is

$$\mathcal{Y}^{(k+1)} \quad = \quad \mathcal{H}^{(k+1)}\psi^{(k+1)} + \mathcal{E}^{(k+1)} \qquad (21)$$

where $\mathcal{H}$ is the time varying observation matrix and $\mathcal{E}$ is the observation noise, assumed Gaussian. We assume that data is available on a single row of the field ψ and a different row is observed at each iteration. Such sparsity arises in remote sensing applications with satellite observations and mimics the Topex/Poseidon altimetry measurements of the ocean surface. Although the experiment is borrowed from [3], our state and observation models are different.

Fig. 3 shows the evolution over time of the mean square error (MSE) for the estimated fields ψ obtained from the optimal KBf and the local KBfs. The MSE plots are normalized by the energy present in the actual field. As it is observed, the local KBf follows closely the optimal KBf showing that the GMrp approximation is a fairly good approximation to the problem. For implementations of the local KBf with the number of block diagonals $L > 1$, the MSE of the local KBf is so small that it is undistinguishable in the plot from the optimal KBf MSE.

To quantify the approximation of the error covariance matrix $\mathcal{P}$, we plot in figure 4 the 2-norm differences between the error covariance matrix of the optimal KBf and the local KBfs with $L = 1, 2$, and 4 block banded approximations. The 2-norm differences are normalized with the 2-norm magnitude of the error covariance matrix of the optimal KBf. The plots show that, after a small transient, the difference between the error covariance matrices of the optimal KBf and the local KBfs is small, with the approximation improving as the value of L is increased. An interesting feature for $L = 1$ is the sharp bump in the plots around the 6 – 10'th iterations. The bump reduces and subsequently disappears for higher values of L.

5. CONCLUSIONS

The paper presents fast algorithms to invert full matrices with L-block banded inverses and vice versa. In both cases, only blocks corresponding to the non-zero blocks of the L-block banded matrix are used. These algorithms provide computational savings of up to two orders of magnitude. The properties of L-block banded matrices are used to derive efficient implementations of the KBf, referred to as the local KBfs. The local KBfs reduce computations by $O(J^2/L)$, where J is the linear dimension of the field and L is the block bandwidth of the matrix approximation, and their performance is virtually indistinguishable from the optimal KBf in our simulations.

6. REFERENCES

[1] A. Kavcic, and J. M. F. Moura, "Matrix with Banded Inverses: Algorithms and Factorization of Gauss-Markov Processes," in *IEEE Trans. on Inf. Theory*, 46(4), pp. 1495-1509, Jul. 2000.

[2] A. Asif, and J. M. F. Moura, "Data Assimilation in Large Time-Varying Multidimensional Fields," in *IEEE Trans. on Image Processing*, 8(11), pp. 1593-1607, Nov. 1999.

[3] T. M. Chin, W. C. Karl, and A. S. Willsky "Sequential Filtering for Multi-frame Visual Reconstruction," in *Signal Processing*, 28, pp. 311-333, 1992.