Block Matrices With L-Block-banded Inverse: Inversion Algorithms

Amir Asif, *Senior Member, IEEE,* and José M. F. Moura, *Fellow, IEEE*

***Abstract*—Block-banded matrices generalize banded matrices. We study the properties of positive definite full matrices $\mathcal{P}$ whose inverses $\mathcal{A}$ are L-block-banded. We show that, for such matrices, the blocks in the L-block band of $\mathcal{P}$ completely determine $\mathcal{P}$; namely, all blocks of $\mathcal{P}$ outside its L-block band are computed from the blocks in the L-block band of $\mathcal{P}$. We derive fast inversion algorithms for $\mathcal{P}$ and its inverse $\mathcal{A}$ that, when compared to direct inversion, are faster by two orders of magnitude of the linear dimension of the constituent blocks. We apply these inversion algorithms to successfully develop fast approximations to Kalman–Bucy filters in applications with high dimensional states where the direct inversion of the covariance matrix is computationally unfeasible.**

***Index Terms*—Block-banded matrix, Cholesky decomposition, covariance matrix, Gauss–Markov random process, Kalman–Bucy filter, matrix inversion, sparse matrix.**

I. Introduction

BLOCK-banded matrices and their inverses arise frequently in signal processing applications, including autoregressive or moving average image modeling, covariances of Gauss–Markov random processes (GMrp) [1], [2], or with finite difference numerical approximations to partial differential equations. For example, the point spread function (psf) in image restoration problems has a block-banded structure [3]. Block-banded matrices are also used to model the correlation of cyclostationary processes in periodic time series [4]. We are motivated by signal processing problems with large dimensional states where a straightforward implementation of a recursive estimation algorithm, such as the Kalman–Bucy filter (KBf), is prohibitively expensive. In particular, the inversion of the error covariance matrices in the KBf is computationally intensive precluding the direct implementation of the KBf to such problems.

Several approaches[1] [3]–[10] have been proposed in the literature for the inversion of (scalar, not *block*) banded matrices. In banded matrices, the entries in the band diagonals are scalars. In contrast with banded matrices, there are much fewer results published for *block*-banded matrices. Any block-banded matrix is also banded; therefore, the methods in [3]–[10] could in principle be applicable to block-banded matrices. Extension of inversion algorithms designed for *banded* matrices to *block-banded* matrices generally fails to exploit the sparsity patterns within each block and between blocks; therefore, they are not optimal [11]. Further, algorithms that use the block-banded structure of the matrix to be inverted are computationally more efficient than those that just manipulate scalars because with block-banded matrices more computations are associated with a given data movement than with scalar banded matrices. An example of an inversion algorithm that uses the inherent structure of the constituent blocks in a block-banded matrix to its advantage is in [3]; this reference proposes a fast algorithm for inverting block Toeplitz matrices with Toeplitz blocks. References [12]–[17] describe alternative algorithms for matrices with a similar Toeplitz structure.

This paper develops results for positive definite and symmetric L-block-banded matrices $\mathcal{A}$ and their inverses $\mathcal{P} = \mathcal{A}^{-1}$. An example of $\mathcal{P}$ is the covariance matrix of a Gauss–Markov random field (GMrp); see [2] with its inverse $\mathcal{A}$ referred to as the information matrix. Unlike existing approaches, no additional structure or constraint is assumed on $\mathcal{A}$; in particular, the algorithms in this paper do not require $\mathcal{A}$ to be Toeplitz. Our inversion algorithms generalize our earlier work presented in [18] for tridiagonal block matrices to block matrices with an arbitrary bandwidth L. We show that the matrix $\mathcal{P}$, whose inverse $\mathcal{A}$ is an L-block-banded matrix, is completely defined by the blocks within its L-block band. In other words, when the block matrix $\mathcal{P}$ has an L-block-banded inverse, $\mathcal{P}$ is highly structured. Any block entry outside the L-block diagonals of $\mathcal{P}$ can be obtained from the block entries within the L-block diagonals of $\mathcal{P}$. The paper proves this fact, which is at first sight surprising, and derives the following algorithms for block matrices $\mathcal{P}$ whose inverses $\mathcal{A}$ are L-block-banded:

1) **Inversion of** $\mathcal{P}$: An inversion algorithm for $\mathcal{P}$ that uses only the block entries in the L-block band of $\mathcal{P}$. This is a very efficient inversion algorithm for such $\mathcal{P}$; it is faster than direct inversion by two orders of magnitude of the linear dimension of the blocks used.
2) **Inversion of** $\mathcal{A}$: A fast inversion algorithm for the L-block-banded matrix $\mathcal{A}$ that is faster than its direct inversion by up to one order of magnitude of the linear dimension of its constituent blocks.

Compared with the scalar banded representations, the block-banded implementations of Algorithms 1 and 2 provide computational savings of up to three orders of magnitude of the

Manuscript received January 7, 2003; revised February 24, 2004. This work was supported in part by the Natural Science and Engineering Research Council (NSERC) of Canada under Grants 229862-01 and 228415-03 and by the United States Office of Naval Research under Grant N00014-97-1-0800. The associate editor coordinating the review of this paper and approving it for publication was Prof Trac D. Tran.

A. Asif is with the Department of Computer Science and Engineering, York University, Toronto, ON, M3J 1P3 Canada (Email: asif@cs.yorku.ca)

J. M. F. Moura is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, 15213 USA (e-mail: moura@ece.cmu.edu).

Digital Object Identifier 10.1109/TSP.2004.840709

[1]This review is intended only to provide context to the work and should not be treated as complete or exhaustive.

dimension of the constituent blocks used to represent $\mathcal{A}$ and its inverse $\mathcal{P}$. The inversion algorithms are then used to develop alternative, computationally efficient approximations of the KBf. These near-optimal implementations are obtained by imposing an L-block-banded structure on the inverse of the error covariance matrix (information matrix) and correspond to modeling the error field as a reduced-order Gauss–Markov random process (GMrp). Controlled simulations show that our KBf implementations lead to results that are virtually indistinguishable from the results for the conventional KBf.

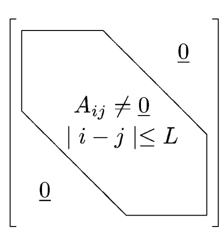
The paper is organized as follows. In Section II, we define the notation used to represent block-banded matrices and derive three important properties for L-block-banded matrices. These properties express the block entries of an L-block-banded matrix in terms of the block entries of its inverse, and vice versa. Section III applies the results derived in Section II to derive inversion algorithms for an L-block-banded matrix $\mathcal{A}$ and its inverse $\mathcal{P}$. We also consider special block-banded matrices that have additional zero block diagonals within the first L-block diagonals. To illustrate the application of the inversion algorithms, we apply them in Section IV to inverting large covariance matrices in the context of an approximation algorithm to a large state space KBf problem. These simulations show an almost perfect agreement between the approximate filter and the exact KBf estimate. Finally, Section V summarizes the paper.

II. Banded Matrix Results

A. Notation

Consider a positive-definite symmetric matrix $\mathcal{P}$ represented by its $(I \times I)$ constituent blocks $\mathcal{P} = \{P_{ij}\}$, $1 \le i, j \le J$. The matrix $\mathcal{P}$ is assumed to have an L-block-banded inverse $\mathcal{A} = \{A_{ij}\}$, $1 \le i,j \le J$, with the following structure:

$\mathcal{A} =$



(1)

where the square blocks A_{ij} and the zero square blocks $\underline{0}$ are of order I. A diagonal block matrix is a 0-block-banded matrix ($L = 0$). A tridiagonal block matrix has exactly one nonzero block diagonal above and below the main diagonal and is therefore a 1-block-banded matrix ($L = 1$) and similarly for higher values of L. Unless otherwise specified, we use calligraphic fonts to denote matrices (e.g., $\mathcal{A}$ or $\mathcal{P}$) with dimensions $(IJ \times IJ)$. Their constituent blocks are denoted by capital letters (e.g., A_{ij} or P_{ij}) with the subscript (i, j) representing their location inside the full matrix in terms of the number of block rows i and block columns j. The blocks A_{ij} (or P_{ij}) are of dimensions $(I \times I)$, implying that there are J block rows and J block columns in matrix $\mathcal{A}$ (or $\mathcal{P}$).

To be concise, we borrow the MATLAB[2] notation to refer to an ordered combination of blocks P_{ij}. A principal submatrix of $\mathcal{P}$ spanning block rows and columns i through j $(1 \le i \le j \le J)$ is given by

$$\mathcal{P}(i:j, i:j) \triangleq \begin{bmatrix} P_{ii} & P_{ii+1} & \cdot & P_{ij} \\ P_{i+1i} & P_{i+1i+1} & \cdot & P_{i+1j} \\ \cdot & \cdot & \ddots & \cdot \\ P_{ji} & P_{ji+1} & \cdot & P_{jj} \end{bmatrix}. \tag{2}$$

The Cholesky factorization of $\mathcal{A} = \mathcal{U}^T\mathcal{U}$ results in the Cholesky factor $\mathcal{U}$ that is an upper triangular matrix. To indicate that the matrix $\mathcal{U}$ is the upper triangular Cholesky factor of $\mathcal{A}$, we work often with the notation $\mathcal{U} = \text{chol}(\mathcal{A})$.

Lemma 1.1 shows that the Cholesky factor of the L-block-banded matrix $\mathcal{A}$ has exactly L nonzero block diagonals above the main diagonal.

Lemma 1.1: A positive definite and symmetric matrix $\mathcal{A} = \mathcal{U}^T\mathcal{U}$ is L-block-banded if and only if (iff) the constituent blocks U_{ij} in its upper triangular Cholesky factor $\mathcal{U}$ are

$$U_{ij} = \underline{0}, \quad \text{for } (j - i) > L, 1 \le i \le (J - L). \tag{3}$$

The proof of Lemma 1.1 is included in the Appendix.

The inverse $\mathcal{U}^{-1}$ of the Cholesky factor $\mathcal{U}$ is an upper triangular matrix

$$\mathcal{U}^{-1} = \begin{bmatrix} U_{11}^{-1} & \underline{*} & \underline{*} & \underline{*} & \cdot & \underline{*} \\ \underline{0} & U_{22}^{-1} & \underline{*} & \underline{*} & \cdot & \underline{*} \\ \underline{0} & \underline{0} & U_{33}^{-1} & \underline{*} & \cdot & \underline{*} \\ & & \ddots & \ddots & & \\ \cdot & \cdot & & & \cdot & \cdot \\ \underline{0} & \cdot & \cdot & \underline{0} & U_{I-1I-1}^{-1} & \underline{*} \\ \underline{0} & \cdot & \cdot & \underline{0} & \underline{0} & U_{II}^{-1} \end{bmatrix} \tag{4}$$

where the lower diagonal entries in $\mathcal{U}^{-1}$ are zero blocks $\underline{0}$. More importantly, the main diagonal entries in $\mathcal{U}^{-1}$ are block inverses of the corresponding blocks in $\mathcal{U}$. These features are used next to derive three important theorems for L-block-banded matrices where we show how to obtain the following:

a) block entry $\{U_{ij}\}$ of the Cholesky factor $\mathcal{U}$ from a selected number of blocks $\{P_{ij}\}$ of $\mathcal{P}$ without inverting the full matrix $\mathcal{P}$;
b) block entries $\{P_{ij}\}$ of $\mathcal{P}$ recursively from the blocks $\{U_{ij}\}$ of $\mathcal{U}$ without inverting the complete Cholesky factor $\mathcal{U}$; and
c) block entries $\{P_{ij}\}$, $|i - j| > L$ outside the first L diagonals of $\mathcal{P}$ from the blocks within the first L-diagonals of $\mathcal{P}$.

Since we operate at a block level, the three theorems offer considerable computational savings over direct computations of $\mathcal{U}$ from $\mathcal{P}$ and vice versa. The proofs are included in the Appendix.

B. Theorems

Theorem 1: The Cholesky blocks $\{U_{ii}, \ldots, U_{i,i+L}\}$'s[3] on block row i of the Cholesky factor $\mathcal{U}$ of an L-block-banded ma-

[2]MATLAB is a registered trademark of Mathworks.

[3]A comma in the subscript helps in differentiating between $P_{ii+2,\tau}$ and $P_{i,i+2\tau}$ that in our earlier notation is written as $P_{ii+2\tau}$. We will use comma in the subscript only for cases where confusion may arise.

trix $\mathcal{A} = \mathcal{U}^T\mathcal{U}$ are determined from the principal submatrix $\mathcal{P}(i:i+L, i:i+L)$ of $\mathcal{P} = \mathcal{A}^{-1}$ by

$$\forall i, 1 \le i \le (J-L):$$
$$\underbrace{\begin{bmatrix} P_{ii} & P_{ii+1} & \cdot & P_{ii+L} \\ P_{i+1i} & P_{i+1i+1} & \cdot & P_{i+1i+L} \\ \cdot & \cdot & \ddots & \cdot \\ P_{i+Li} & P_{i+Li+1} & \cdot & P_{i+Li+L} \end{bmatrix}}_{\mathcal{P}(i:i+L,i:i+L)} \begin{bmatrix} U_{ii}^T \\ U_{i,i+1}^T \\ \vdots \\ U_{i,i+L}^T \end{bmatrix} = \begin{bmatrix} U_{ii}^{-1} \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix} \tag{5}$$

$$\forall i: (J-L+1) \le i \le J:$$
$$\underbrace{\begin{bmatrix} P_{ii} & P_{i,i+1} & \cdot & P_{iJ} \\ P_{i+1,i} & P_{i+1,i+1} & \cdot & P_{i+1,J} \\ \cdot & \cdot & \ddots & \cdot \\ P_{Ji} & P_{J,i+1} & \cdot & P_{JJ} \end{bmatrix}}_{\mathcal{P}(i:J,i:J)} \begin{bmatrix} U_{ii}^T \\ U_{i,i+1}^T \\ \vdots \\ U_{i,J}^T \end{bmatrix} = \begin{bmatrix} U_{ii}^{-1} \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix} \tag{6}$$

$$\vdots$$
$$U_{JJ}^T U_{JJ} = P_{JJ}^{-1} \tag{7}$$

■

Theorem 1 shows how the blocks $\{U_{ij}\}$ of the Cholesky factor $\mathcal{U}$ are determined from the blocks $\{P_{ij}\}$ of the L-banded $\mathcal{P}$. Equations (5) and (6) show that the Cholesky blocks $\{U_{ii} \ldots U_{i,i+L}\}$ on block row i of $\mathcal{U}$, $1 \le i \le (J-L)$ only involve the $(L+1)^2$ blocks in the principal submatrix $\mathcal{P}(i:i+L, i:i+L)$ that are in the neighborhood of these Cholesky blocks $\{U_{ii}, \cdots, U_{i,i+L}\}$. For block rows $i > (J-L)$, the dimensions of the required principal submatrix of $\mathcal{P}$ is further reduced to $\mathcal{P}(i:J, i:J)$, as shown by (6). In other words, all block rows of the Cholesky factor $\mathcal{U}$ can be determined independently of each other by selecting the appropriate principal submatrix of $\mathcal{P}$ and then applying (5). For block row i, the required principal submatrix of $\mathcal{P}$ spans block rows (and block columns) i through $i+L$.

An alternative to (5) can be obtained by right multiplying both sides in (5) by U_{ii} and rearranging terms to get

$$\begin{bmatrix} U_{ii}^T U_{ii} \\ U_{ii+1}^T U_{ii} \\ \vdots \\ U_{ii+L}^T U_{ii} \end{bmatrix} = \underbrace{\begin{bmatrix} P_{ii} & P_{ii+1} & \cdot & P_{ii+L} \\ P_{i+1i} & P_{i+1i+1} & \cdot & P_{i+1i+L} \\ \cdot & \cdot & \ddots & \cdot \\ P_{i+Li} & P_{i+Li+1} & \cdot & P_{i+Li+L} \end{bmatrix}^{-1}}_{\mathcal{P}(i:i+L,i:i+L)} \underbrace{\begin{bmatrix} I_I \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix}}_{I_O} \tag{8}$$

where I_I is the identity block of order I. Equation (8) is solved for $\{U_{ii}^T U_{ii}, U_{ii+1}^T U_{ii}, \ldots, U_{ii+L}^T U_{ii}\}$. The blocks U_{ii} are obtained by Cholesky factorization of the first term $U_{ii}^T U_{ii}$. This factorization to be well defined requires that the resulting first term $U_{ii}^T U_{ii}$ in (8) be positive definite. This is easily verified. Since block $\mathcal{P}(i:i+L, i:i+L)$ is a principal submatrix of $\mathcal{P}$, its inverse is positive definite. The top left entry corresponding to $U_{ii}^T U_{ii}$ on the right-hand side of (8) is obtained by selecting the first $(I \times I)$ principal submatrix of the inverse of $\mathcal{P}(i:i+L, i:i+L)$, which is then positive definite as desired.

We now proceed with Theorem 2, which expresses the blocks P_{ij} of $\mathcal{P}$ in terms of the Cholesky blocks U_{ij}.

Theorem 2: The upper triangular blocks P_{ij} in $\mathcal{P} = \mathcal{A}^{-1}$, with $\mathcal{A}$ being L-block-banded, are obtained recursively from the Cholesky blocks $\{U_{ii}, \ldots, U_{ii+L}\}$ of the Cholesky factor $\mathcal{U} = \text{chol}(\mathcal{A})$ by

$$\forall i, j, 1 \le i \le (J-1), i \le j \le (i+L) \le J:$$
$$P_{ij} = -\sum_{\ell=i+1}^{min(J,i+L)} \left(U_{ii}^{-1} U_{i\ell}\right) P_{\ell j} \tag{9}$$
$$P_{ii} = \left(U_{ii}^T U_{ii}\right)^{-1} - \sum_{\ell=i+1}^{min(J,i+L)} P_{i\ell} \left(U_{ii}^{-1} U_{i\ell}\right)^T \tag{10}$$
$$P_{JJ} = \left(U_{JJ}^T U_{JJ}\right)^{-1} \tag{11}$$

for the last row. ■

Theorem 2 states that the blocks $P_{ii} \ldots P_{ii+L}$ on block row i and within the first L-block diagonals in $\mathcal{P}$ can be evaluated from the corresponding Cholesky blocks $U_{ii} \ldots U_{ii+L}$ in $\mathcal{U}$ and the L-banded blocks in the lower block rows of $\mathcal{P}$, i.e., $P_{mm} \ldots P_{mm+L}$, with $m > i$.

To illustrate the recursive nature of the computations, consider computing the diagonal block P_{J-3J-3} of a matrix $\mathcal{P}$ that, for example, has a 2-block-banded $(L=2)$ inverse. This requires computing the following blocks:

$$\begin{matrix} P_{J-3J-3} & P_{J-3J-2} & P_{J-3J-1} & \\ & P_{J-2J-2} & P_{J-2J-1} & P_{J-2J} \\ & & P_{J-1J-1} & P_{J-1J} \\ & & & P_{JJ} \end{matrix}$$

in the reverse zig-zag order specified as follows:

$$\begin{matrix} 9 & 8 & 7 & \\ & 6 & 5 & 4 \\ & & 3 & 2 \\ & & & 1 \end{matrix}$$

where the number indicates the order in which the blocks are computed. The block P_{JJ} is calculated first, followed by P_{J-1J}, and so on with the remaining entries until P_{J-3J-3} is reached.

Next, we present Theorem 3, which expresses the block entries outside the first L diagonals in $\mathcal{P}$ in terms of its blocks within the first L-diagonals.

Theorem 3: Let $\mathcal{A}$ be L-block-banded and $\mathcal{P} = \mathcal{A}^{-1}$. Then

$$\forall i, j, 1 \le i < (J-L), (i+L) < j \le J:$$
$$P_{ij} = [P_{ii+1} \ldots P_{ii+L}] \times \begin{bmatrix} P_{i+1i+1} & \cdot & P_{i+1i+L} \\ \cdot & \ddots & \cdot \\ P_{i+Li+1} & \cdot & P_{i+Li+L} \end{bmatrix}^{-1} \begin{bmatrix} P_{i+1j} \\ \vdots \\ P_{i+Lj} \end{bmatrix}. \tag{12}$$

■

This theorem shows that the blocks P_{ij}, $|i-j| > L$ outside the L-band of $\mathcal{P}$ are determined from the blocks P_{ij}, $|i-j| \le L$ within its L-band. In other words, the matrix $\mathcal{P}$ is completely specified by its first L-block diagonals. Any blocks outside the L-block diagonals can be evaluated recursively from blocks within the L-block diagonals. In the paper, we refer to the blocks in the L-block band of $\mathcal{P}$ as the *significant* blocks. The blocks outside the L-block band are referred to as the *nonsignificant* blocks. By Theorem 3, the nonsignificant blocks are determined from the significant blocks of $\mathcal{P}$.

To illustrate the recursive order by which the nonsignificant blocks are evaluated from the significant blocks, consider an example where we compute block P_{16} in $\mathcal{P}$, which we assume has a 3-block-banded $(L = 3)$ inverse $\mathcal{A} = \mathcal{P}^{-1}$. First, write P_{16} as given by Theorem 3 as

$$P_{16} = [P_{12} \quad P_{13} \quad P_{14}] \begin{bmatrix} P_{22} & P_{23} & P_{24} \\ P_{32} & P_{33} & P_{34} \\ P_{42} & P_{43} & P_{44} \end{bmatrix}^{-1} \begin{bmatrix} P_{26} \\ P_{36} \\ P_{46} \end{bmatrix}.$$

Then, note that all blocks on the right-hand side of the equation are significant blocks, i.e., these lie in the 3-block band of $\mathcal{P}$, except P_{26}, which is a nonsignificant block. Therefore, we need to compute P_{26} first. By application of Theorem 3 again, we can see that block P_{26} can be computed directly from the significant blocks, i.e., from blocks that are all within the 3-block band of $\mathcal{P}$, so that no additional nonsignificant blocks of $\mathcal{P}$ are needed.

As a general rule, to compute the block entries outside the L-block band of $\mathcal{P}$, we should first compute the blocks on the $(L+1)$th diagonal from the significant blocks, followed by the $(L+2)$ block diagonal entries, and so on, until all blocks outside the band L have been computed.

We now restate, as Corollaries 1.1–3.1, Theorems 1–3 for matrices with tridiagonal matrix inverses, i.e., for $L = 1$. These corollaries are the results in [18].

Corollary 1.1: The Cholesky blocks $\{U_{ii}, U_{ii+1}\}$ of a tridiagonal $(L = 1)$ block-banded matrix $\mathcal{A} = \mathcal{U}^T\mathcal{U}$ can be computed directly from the main diagonal blocks $\{P_{ii}\}$ and the first upper diagonal blocks $\{P_{ii+1}\}$ of $\mathcal{P} = \mathcal{A}^{-1}$ using the following expressions:

$$U_{JJ} = \operatorname{chol}\left(P_{JJ}^{-1}\right) \tag{13}$$

$$\left.\begin{aligned} U_{ii} &= \operatorname{chol}\left(\left(P_{ii} - P_{ii+1}P_{i+1i+1}^{-1}P_{ii+1}^T\right)^{-1}\right) \\ U_{ii+1} &= -U_{ii}P_{ii+1}P_{i+1i+1}^{-1} \end{aligned}\right\} \quad \text{for } (J-1) \geq i \geq 1. \tag{14}$$

Corollary 2.1: The main and the first upper block diagonal entries $\{P_{ii}, P_{ii+1}\}$ of $\mathcal{P}$ with a tridiagonal $(L = 1)$ block-banded inverse $\mathcal{A}$ can be evaluated from the Cholesky factors $\{U_{ii}, U_{ii+1}\}$ of $\mathcal{A}$ from the following expressions:

$$P_{JJ} = \left(U_{JJ}^T U_{JJ}^T\right)^{-1} \tag{15}$$

$$\left.\begin{aligned} P_{ii+1} &= \left(-U_{ii}^{-1}U_{ii+1}\right)P_{i+1i+1} \\ P_{ii} &= \left(U_{ii}^T U_{ii}\right)^{-1} + \left(U_{ii}^{-1}U_{ii+1}\right)P_{i+1i+1}\left(U_{ii}^{-1}U_{ii+1}\right)^T \end{aligned}\right\} \quad \text{for } (J-1) \geq i \geq 1. \tag{16}$$

Corollary 3.1: Given the main and the first upper block diagonal entries $\{P_{ii}, P_{ii+1}\}$ of $\mathcal{P}$ with a tridiagonal $(L = 1)$ block-banded inverse $\mathcal{A}$, any nonsignificant upper triangular block entry of $\mathcal{P}$ can be computed from its significant blocks from the following expression:

$$\forall i, j, 1 \leq i \leq J, (i+2) \leq j \leq J:$$
$$P_{ij} = \left(\prod_{\ell=i}^{j-2}\left(P_{\ell+1\ell+1}^{-1}P_{\ell+1\ell}\right)^T\right)P_{j-1j}. \tag{17}$$

In Corollary 3.1, the following notation is used:

$$\prod_{\ell=1}^{J}(A_\ell) = A_1A_2\ldots A_J. \tag{18}$$

We note that in (17), the block P_{ij} is expressed in terms of blocks on the main diagonal and on the first upper diagonal $\{P_{ii}, P_{ii+1}\}$. Thus, any nonsignificant block in $\mathcal{P}$ is computed directly from the significant blocks $\{P_{ii}, P_{ii+1}\}$ without the need for a recursion.

III. Inversion Algorithms

In this section, we apply Theorems 1 and 2 to derive computationally efficient algorithms to invert the full symmetric positive definite matrix $\mathcal{P}$ with an L-block-banded inverse $\mathcal{A}$ and to solve the converse problem of inverting the symmetric positive definite L-block-banded matrix $\mathcal{A}$ to obtain its full inverse $\mathcal{P}$. We also include results from simulations that illustrate the computational savings provided by Algorithms 1 and 2 over direct inversion of the matrices. In this section, the matrix $\mathcal{P}$ is $(N \times N)$, i.e., $N = IJ$ with blocks P_{ij} of order I. We only count the multiplication operations assuming that inversion or multiplication of generic $(I \times I)$ matrices requires I^3 floating-point operations (flops).

A. Inversion of Matrices With Block-Banded Inverses

Algorithm 1—$\mathcal{A} = \mathcal{P}^{-1}$: This algorithm computes the L-block-banded inverse $\mathcal{A}$ from blocks P_{ij} of $\mathcal{P}$ using Steps 1 and 2. Since $\mathcal{A}$ is symmetric $(A_{ij}^T = A_{ji})$ (similarly for P_{ij}), we only compute the upper triangular blocks of $\mathcal{A}$ (or $\mathcal{P}$).

Step 1: Starting with $i = J$, the Cholesky's blocks $\{U_{ii}, \ldots, U_{ii+L}\}$ are calculated recursively using Theorem 1. The blocks on row i, for example, are calculated using (8), which computes the terms $\{U_{ii}^T U_{ii}, U_{ii+1}^T U_{ii}, \ldots, U_{ii+L}^T U_{ii}\}$ for $(J - L) \geq i \geq 1$. The main diagonal Cholesky blocks $\{U_{ii}\}$ are obtained by solving for the Cholesky factors of $\{U_{ii}^T U_{ii}\}$. The off-diagonal Cholesky blocks $U_{i,i+l}$, $1 \leq l \leq L$ are evaluated by multiplying the corresponding entity $U_{i,i+l}^T U_{ii}$ calculated in (8) by the inverse of $\{U_{ii}\}$.

Step 2: The upper triangular block entries $(A_{ij}, j > i)$ in the information matrix $\mathcal{A}$ are determined from the following expression:

$$A_{ij} = \begin{cases} \sum\limits_{\ell=\max(1,j-L)}^{i} U_{\ell i}^T U_{\ell j}, & \text{for } (j-i) \leq L \\ \underline{0}, & \text{for } (j-i) > L \end{cases} \tag{19}$$

obtained by expanding $\mathcal{A} = \mathcal{U}^T\mathcal{U}$ in terms of the constituent blocks of $\mathcal{U}$.

Alternative Implementation: A second implementation of Algorithm 1 that avoids Cholesky factorization is obtained by expressing (19) as

$$A_{ij} = \begin{cases} \sum_{\ell=\max(1,j-L)}^{i} \left(U_{\ell i}^T U_{ii}\right)\left(U_{ii}^T U_{ii}\right)^{-1}\left(U_{\ell j}^T U_{ii}\right)^T, & (j-i) \le L \\ \underline{0}, & (j-i) > L \end{cases} \tag{20}$$

and solving (8) for the Cholesky products $\{U_{ii+k}^T U_{ii}\}$, for $1 \le k \le L$ and $1 \le i \le J$, instead of the individual Cholesky blocks. Throughout the manuscript, we use this implementation whenever we refer to Algorithm 1.

Computations: In (8), the principal matrix $\mathcal{P}(i:i+L, i:i+L)$ is of order LI. Multiplying its inverse with I_O as in (8) is equivalent to selecting its first column. Thus, only L out of L^2 block entries of the inverse of $\mathcal{P}(i:i+L, i:i+L)$ are needed, reducing by a factor of $1/L$ the computations to inverting the principal submatrix. The number of flops to calculate the Cholesky product terms $\{U_{ii+k}^T U_{ii}\}$ on row i of $\mathcal{U}$ is therefore $(LI)^3/L$ or $L^2 I^3$. The total number of flops to compute the Cholesky product terms for N/I rows in $\mathcal{U}$ is then

$$\text{Number of flops in Step 1} = \frac{N}{I} \times (L^2 I^3) = (NL^2 I^2). \tag{21}$$

In Step 2 of Algorithm 1, the number of summation terms in (20) to compute A_{ij} is L (except for the first few initial rows, $i < L$). Each term involves two block multiplications,[4] i.e., $2LI^3$ flops are needed to compute A_{ij}. There are roughly $(L+1)N/I$ nonzero blocks A_{ij} in the upper half of the L-block-banded inverse $\mathcal{A}$ resulting in the following flop count:

$$\text{Number of flops in Step 2} = (L+1)\frac{N}{I} \times (2LI^3) \approx (2NL^2 I^2). \tag{22}$$

The total number of flops to compute $\mathcal{A}$ using Algorithm 1 is therefore given by

$$\text{Number of flops in Algorithm 1} = (3NL^2 I^2) \tag{23}$$

or $(3L^2 I^3 J)$, which is an improvement of $\mathrm{O}((J/L)^2)$ over the direct inversion of $\mathcal{P}$.

As an aside, it may be noted that Step 1 of Algorithm 1 computes the Cholesky factors of an L-block-banded matrix and can be used for Cholesky factorization of $\mathcal{A}$.

B. Inversion of L-Block-Banded Matrices

Algorithm 2—$\mathcal{P} = \mathcal{A}^{-1}$: This algorithm calculates $\mathcal{P}$ from its L-block-banded inverse $\mathcal{A}$ from the following two steps.

Step 1: Calculate the Cholesky blocks $\{U_{ij}\}$ from $\mathcal{A}$. These can be evaluated recursively using the following expressions:

$$\forall i, k, 2 \le i \le J, 1 \le k \le L:$$

$$U_{ii} = \text{chol}\left(A_{ii} - \sum_{\ell=\max(1,i-L)}^{i-1} \left(U_{\ell i}^T U_{\ell i}\right)\right) \tag{24}$$

$$U_{ii+k} = U_{ii}^{-T}\left(A_{ii+k} - \sum_{\ell=\max(1,i-L)}^{i-1} U_{\ell i}^T U_{\ell i+k}\right). \tag{25}$$

The boundary condition (b.c.) for the first row $i = 1$ is

$$\forall k, 1 \le k \le L: U_{11} = \text{chol}(A_{11}) \quad \text{and} \quad U_{11+k} = U_{11}^{-T} A_{11+k}. \tag{26}$$

Equations (24)–(26) are derived by rearranging terms in (19).

Step 2: Starting with P_{JJ}, the block entries $\{P_{ij}\}, 1 \le i \le J$, $j \le i$, and $j \le J$ in $\mathcal{P}$ are determined recursively from the Cholesky blocks $\{U_{ij}\}$ using Theorems 2 and 3.

Alternative Implementation: To compute U_{ii} from (24) demands that the matrix

$$A_{ii} - \sum_{\ell=\max(1,i-L)}^{i-1} \left(U_{\ell i}^T U_{\ell i}\right) \tag{27}$$

be positive definite. Numerical errors with badly conditioned matrices may cause the factorization of this matrix to fail. The Cholesky factorization can be avoided by noting that Theorem 2 requires only terms $(U_{ii}^T U_{ii})$ and $(U_{ii}^{-1} U_{ii+k})$, which in turn use $(U_{ii+m}^T U_{ii+k})$. We can avoid the Cholesky factorization of matrix (27) by replacing Step 1 as follows:

Step 1: Calculate the product terms

$$\left(U_{ii}^T U_{ii}\right) = A_{ii} - \sum_{\ell=\max(1,i-L)}^{i-1} \left(U_{\ell i}^T U_{\ell i}\right) \tag{28}$$

$$\left(U_{ii}^{-1} U_{ii+k}\right) = \left(U_{ii}^T U_{ii}\right)^{-1} \times \left(A_{ii+k} - \sum_{\ell=\max(1,i-L)}^{i-1} U_{\ell i}^T U_{\ell i+k}\right) \tag{29}$$

$$\left(U_{ii+k}^T U_{ii+m}\right) = \left(U_{ii}^{-1} U_{ii+k}\right)^T \left(U_{ii}^T U_{ii}\right)\left(U_{ii}^{-1} U_{ii+m}\right) \tag{30}$$

for $2 \le i \le J$, $1 \le k \le L$, and $k \le m \le L$ with boundary condition $U_{11}^T U_{11} = A_{11}$, $U_{11}^{-1} U_{11+k} = A_{11}^{-1} A_{11+k}$ and $(U_{11+k}^T U_{11+m}) = A_{11+k}^T A_{11}^{-1} A_{11+m}$. We will use implementation (28)–(30) in conjunction with Step 2 for the inversion of L-block-banded matrices.

Computations: Since the term $(U_{\ell i}^T U_{\ell i})$ is obtained directly by iteration of the previous rows, (28) in Step 1 of Algorithm 2 only involves additions and does not require multiplications. Equation (29) requires one $(I \times I)$ matrix multiplication.[5] The number of terms on each block row i of $\mathcal{U}$ is L; therefore, the

[4]Equation (20) also inverts once for each block row i the matrix $(U_{ii}^T U_{ii})$. Such an inversion $1 \le i \le N/I$ times requires NI^2 flops, which is a factor of L^2 less than our result in (21) not affecting the order of the number of computations.

[5]As we explained in footnote 4, (29) inverts matrix $U_{ii}^{-1} U_{ii+k}$ for each block row i. For $(1 \le i \le N/I)$, this requires NI^2 flops that do not affect the order of the number of computations.

number of flops for computing all such terms on block row i is LI^3. Equation (30) computes $(U_{ii+k}^T U_{ii+m})$ and involves two matrix multiplications. There are $L^2/2$ such terms in row i, requiring a total of L^2I^3 flops. The number of flops required in Step 1 of Algorithm 2 is therefore given by

$$\text{Number of flops in Step 1} = \frac{N}{I} \times (LI^3 + L^2I^3) \approx (NL^2I^2). \tag{31}$$

Step 2 of Algorithm 2 uses Theorem 2 to compute blocks P_{ij}. Each block typically requires L multiplications of $(I \times I)$ blocks. There are $(N/I)^2$ such blocks in $\mathcal{P}$, giving the following expression for the number of flops:

$$\text{Number of flops in Step 2} = \left(\frac{N}{I}\right)^2 \times (LI^3) = (N^2LI). \tag{32}$$

Adding the results from (31) and (32) gives

$$\text{Number of flops in Algorithm 2} = L(LI + N)NI \tag{33}$$

or $L(L + J)I^3J$ flops, which is an improvement of approximately a factor of $\mathrm{O}(J/L)$ over direct inversion of matrix $\mathcal{A}$.

C. Simulations

In Figs. 1 and 2, we plot the results of Monte Carlo simulations that quantify the savings in floating-point operations (flops) resulting from Algorithms 1 and 2 over the direct inversion of the matrices. The plots are normalized by the total number of flops required in the direct inversion; therefore, the region below the ordinate $y = 1$ in these figures corresponds to the number of computations smaller than the number of computations required by the direct inversion of the matrices. This region represents computational savings of our algorithms over direct inversion. 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 $I = 5$, whereas 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 dimensions of matrices $\mathcal{A}$ and $\mathcal{P}$ in the simulation is (250×250). Except for the few initial cases where the overhead involved in indexing and identifying constituent blocks exceeds the savings provided by Algorithm 2, both algorithms exhibit considerable savings over direct inversion. For Algorithm 1, the computations can be reduced by a factor of 10–100, whereas for Algorithm 2, the savings can be by a factor of 10. Higher savings will result with larger matrices.

D. Choice of Block Dimensions

In this subsection, we compare different implementations of Algorithms 1 and 2 obtained by varying the size of the blocks P_{ij} in the matrix $\mathcal{P}$. For example, consider the following representation for $\mathcal{P}$ with scalar dimensions of $(N \times N)$:

$$\mathcal{P}^{(kI)} = \begin{bmatrix} R_{11} & R_{12} & . & R_{1\frac{N}{kI}} \\ R_{21} & R_{22} & . & R_{2\frac{N}{kI}} \\ & & & \\ . & . & \ddots & . \\ R_{\frac{N}{kI}1} & R_{\frac{N}{kI}2} & . & R_{\frac{N}{kI}\frac{N}{kI}} \end{bmatrix} \tag{34}$$

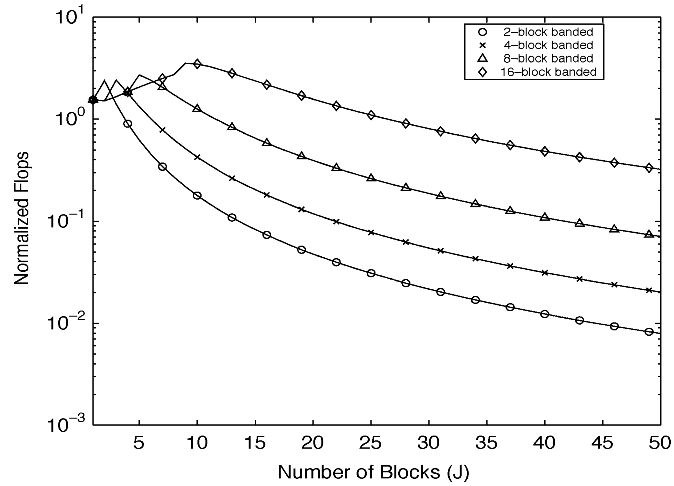


Fig. 1. Number of flops required to invert a full matrix $\mathcal{P}$ with L-block-banded inverse using Algorithm 1 for $L = 2$, 4, 8 and 16. The plots are normalized by the number of flops required to invert $\mathcal{P}$ directly.

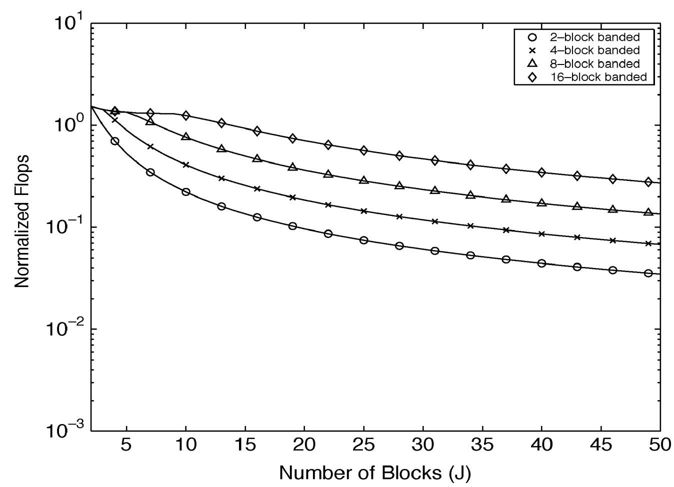


Fig. 2. Number of flops required to invert an L-block-banded $(IJ \times IJ)$ matrix $\mathcal{A}$ using Algorithm 2 for $L = 2$, 4, 8 and 16. The plots are normalized by the number of flops required to invert $\mathcal{A}$ directly.

where the superscript x in $\mathcal{P}^{(x)}$ denotes the size of the blocks used in the representation. The matrix $\mathcal{P}^{(kI)}$ is expressed in terms of $(kI \times kI)$ blocks R_{mn} for $(1 \leq m, n \leq (N/kI))$. For $k = 1$, $\mathcal{P}^{(kI)}$ will reduce to $\mathcal{P}^{(I)}$, which is the same as our representation $\mathcal{P}$ used earlier. Similar representations are used for $\mathcal{A}$: the block-banded inverse of $\mathcal{P}$. Further, assume that the block bandwidth for $\mathcal{A}^{(kI)}$ expressed in terms of $(kI \times kI)$ block sizes is W. Representation $\mathcal{A}^{(I)}$ for the same matrix uses blocks of dimensions $(I \times I)$ and has a block bandwidth of $L = k(W+1)-1$. By following the procedure of Sections III-A and B, it can be shown that the flops required to compute i) the W-block-banded matrix $A^{(kI)}$ from $P^{(kI)}$ using Algorithm 1 and ii) $P^{(kI)}$ from $A^{(kI)}$ using Algorithm 2 are given by

$$\text{Alg. 1: Total number of flops} = 3N(kWI)^2 \tag{35}$$

$$\begin{aligned} \text{Alg. 2: Total number of flops} &= (N + kWI)kWIN \\ &\approx kWIN^2. \end{aligned} \tag{36}$$

To illustrate the effect of the dimensions of the constituent blocks used in $\mathcal{P}$ (or $\mathcal{A}$) on the computational complexity

TABLE I
NUMBER OF FLOPS FOR ALGORITHMS 1 AND 2 EXPRESSED AS A FUNCTION OF THE BLOCK SIZE $(kI \times kI)$. FLOPS FOR ALGORITHM 1 ARE EXPRESSED AS A MULTIPLE OF NI^2 AND FOR ALGORITHM 2 AS A MULTIPLE OF N^2I

Factor k	48	24	16	12	8	6	4	3	2	1
Bandwidth W	1	3	5	7	11	15	23	31	47	95
$\frac{(\#Flops)_{Alg.\ 1}}{NI^2}$	6912	15552	19200	21168	23232	24300	25392	25947	26508	27075
$\frac{(\#Flops)_{Alg.\ 2}}{N^2I}$	48	72	80	84	88	90	92	93	94	95

of our algorithms, we consider an example where $\mathcal{A}^{(kI)}$ is assumed to be 1-block-banded $(W=1)$ with $k=48$. Different implementations of Algorithms 1 and 2 can be derived by varying factor k in dimensions kI of the constituent blocks in $\mathcal{A}^{(kI)}$. As shown in the top row of Table I, several values of k are considered. As dimensions kI of the block size are reduced, the block bandwidth W of the matrix $\mathcal{A}^{(kI)}$ changes as well. The values of the block bandwidth W corresponding to dimensions kI are shown in row 2 of Table I. In row 3, we list, as a multiple of NI^2, the flops required to invert $\mathcal{P}^{(kI)}$ using different implementations of Algorithm 1 for the respective values of k and W shown in rows 1 and 2. Similarly, in row 4, we tabulate the number of flops required to invert $\mathcal{A}^{(kI)}$ using different implementations of Algorithm 2. The number of flops for Algorithm 2 is expressed as a multiple of N^2I.

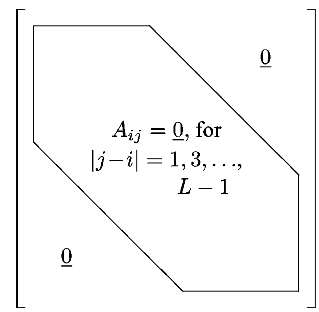
Table I illustrates the computational gain obtained in Algorithms 1 and 2 with blocks of larger dimensions. The increase in the number of computations in the two algorithms using smaller blocks can be mainly attributed to additional zero blocks that are included in the outermost block diagonals of $\mathcal{A}^{(I)}$ when the size kI of the constituent blocks in $\mathcal{A}^{(kI)}$ is reduced. Because the inversion algorithms do not recognize these blocks as zero blocks to simplify the computations, their inclusion increases the overall number of flops used by the two algorithms. If the dimensions of the constituent blocks P_{ij} used in inverting $\mathcal{P}$ with a W-block-banded inverse are reduced from kI to $kI/2$, it follows from (35) that the number of flops in Algorithm 1 is increased by $3N(W+1/4)kI^2$. Similarly, if the dimensions kI of the constituent blocks A_{ij} used in inverting a W-block-banded matrix $\mathcal{A}$ are reduced to $kI/2$, it follows from (36) that the number of flops in Algorithm 2 increases by $kIN^2/2$. Assuming kI to be a power of 2, the above discussion can be extended until $\mathcal{P}$ is expressed in terms of scalars, i.e., P_{ij} is a scalar entry. Using such scalar implementations, it can be shown that the number of flops in Algorithms 1 and 2 are increased by

$$\text{Algorithm 1: Increase in number of flops} = 3NkI(kI-1)\left(2W+1-\frac{1}{kI}\right)$$
$$\text{and Algorithm 2: Increase in number of flops} = (kI-1)N^2,$$

which is roughly an increase by a factor of I^2J over the block-banded implementation based on $(kI \times kI)$ blocks.

E. Sparse Block-Banded Matrices

An interesting application of the inversion algorithms is to invert a matrix $\mathcal{A}$ that is not only L-block-banded but is also constrained in having all odd numbered block diagonals within the first L-block diagonals both above and below the main block diagonal consisting of $\underline{0}$ blocks, i.e.,

$\mathcal{A} =$ 

(37)

By appropriate permutation of the block diagonals, the L-block-banded matrix $\mathcal{A}$ can be reduced to a lower order block-banded matrix with bandwidth $L/2$. Alternatively, Lemma 1 and Theorems 1–3 can be applied directly with the following results.

1) The structure of the upper triangle Cholesky block $\mathcal{U}$ is similar to $\mathcal{A}$, with the block entries $U_{i,i+k}$, $1 \le i \le J$, given by

$$U_{i,i+k} = \underline{0} \quad \text{for } k = 1, 3, \ldots, \min\left((L-1), J-i\right) \tag{38}$$
$$\text{and} \quad U_{i,i+k} = \underline{0} \quad \text{for } k > L. \tag{39}$$

In other words, the blocks $U_{i,i+k}$ on all odd numbered diagonals in $\mathcal{U}$ are $\underline{0}$.

2) In evaluating the nonzero Cholesky blocks U_{ij} (or A_{ij} of the inverse $\mathcal{A} = \mathcal{P}^{-1}$), the only blocks required from $\mathcal{P}$ are the blocks corresponding to the nonzero block diagonals in $\mathcal{U}$ (or $\mathcal{A}$). Theorem 1 reduces to

$$\begin{bmatrix} P_{ii} & P_{i,i+2} & \cdot & P_{i,i+L} \\ P_{i+2,i} & P_{i+2,i+2} & \cdot & P_{i+2,i+L} \\ \cdot & \cdot & \ddots & \cdot \\ P_{i+L,i} & P_{i+L,i+2} & \cdot & P_{i+L,i+L} \end{bmatrix} \begin{bmatrix} U_{ii}^T \\ U_{i,i+2}^T \\ \vdots \\ U_{i,i+L}^T \end{bmatrix} = \begin{bmatrix} U_{ii}^{-1} \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix} \tag{40}$$

for $1 \le i \le (J-L)$. Recall that the blocks of $\mathcal{P}$ required to compute $\mathcal{A}$ are referred to as the significant blocks. In our example, the significant blocks are

$$P_{ii+k}, \quad \text{for} \quad k = 0, 2, 4, \ldots, L. \tag{41}$$

3) The blocks on all odd-numbered diagonals in the full matrix $\mathcal{P}$, which is the inverse of a sparse L-block banded matrix $\mathcal{A}$ with zero blocks on the odd-numbered diagonals,

are themselves zero blocks. This is verified from Theorem 2, which reduces to

$$P_{i,i} = \left(U_{i,i}^T U_{i,i}\right)^{-1} - \sum_{\tau=1}^{\frac{\min(J,i+L)}{2}} P_{i,i+2\tau} \left(U_{i,i}^{-1} U_{i,i+2\tau}\right)^T \quad (42)$$

for $1 \leq i \leq J$. The off-diagonal significant entries are expressed as

$$P_{i,i+k} = \begin{cases} \underline{0}, & \text{for } k = 1, 3, \ldots, \min(L, (J-i)) \\ -\sum_{\tau=1}^{\frac{\min(J,i+L)}{2}} \left(U_{i,i}^{-1} U_{i,i+2\tau}\right) P_{i+2\tau,i} & \\ & \text{for } k = 2, 4, \ldots, \min(L, (J-i)) \end{cases} \quad (43)$$

for $1 \leq i \leq (J-2)$ and with boundary conditions

$$P_{J-1J-1} = \left(U_{J-1J-1}^T U_{J-1J-1}\right)^{-1} \quad (44)$$

$$\text{and} \quad P_{JJ} = \left(U_{JJ}^T U_{JJ}\right)^{-1}. \quad (45)$$

4) Theorem 3 simplifies to the following. For $k = L+2, L+4, \ldots, (J-i)$ and $1 \leq i \leq (J-L)$

$$P_{ii+k} = [P_{ii+2} \quad P_{ii+4} \quad \ldots \quad P_{ii+L}] \times \begin{bmatrix} P_{i+2i+2} & P_{i+2i+4} & \cdot & P_{i+2i+L} \\ P_{i+4i+2} & P_{i+4i+4} & \cdot & P_{i+4i+L} \\ \cdot & \cdot & \ddots & \cdot \\ P_{i+Li+2} & P_{i+Li+4} & \cdot & P_{i+Li+L} \end{bmatrix}^{-1} \times \begin{bmatrix} P_{i+2i+k} \\ P_{i+4i+k} \\ \vdots \\ P_{i+Li+k} \end{bmatrix}. \quad (46)$$

Result 4 illustrates that only the even-numbered significant blocks in $\mathcal{P}$ are used in calculating its nonsignificant blocks.

IV. Application to Kalman–Bucy Filtering

For typical image-based applications in computer vision and the physical sciences, the visual fields of interest are often specified by spatial 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 $\mathcal{W}$)

$$\psi^{(k+1)} = \mathcal{C}^{(k)} \psi^{(k)} + \mathcal{D}^{(k)} \mathcal{W}^{(k)} \quad (47)$$

are block-banded and sparse, i.e., $C_{ij} = \underline{0}$ for $|(i-j)| > L_1$. A similar structure exists for $D = \{D_{ij}\}$ with block bandwidth L_2. The dimension of the state vector ψ is on the order of the number of pixels in the field, which is typically 10^4 to 10^6 elements. Due to this large dimensionality, it is usually only practical to observe a fraction of the field. The observations $\mathcal{Y}$ in the observation model with noise $\mathcal{E}$

$$\mathcal{Y}^{(k+1)} = \mathcal{H}^{(k+1)} \psi^{(k+1)} + \mathcal{E}^{(k+1)} \quad (48)$$

and are, therefore, fairly sparse. This is typically the case with remote sensing platforms on board orbiting satellites.

Implementation of optimal filters such as the KBf to estimate the field ψ (the estimated field is denoted by $\widehat{\psi}$) in such cases requires storage and manipulation of $10^4 \times 10^4$ to $10^6 \times 10^6$ matrices, which is computationally not practical. To obtain a practical implementation, we approximate the non-Markov error process $e = (\psi - \widehat{\psi})$ at each time iteration in the KBf by a Gauss–Markov random process (GMrp) of order L. This is equivalent to approximating the error covariance matrix $\mathcal{P} = E\{ee^T\}$, where E is the expectation operator, by a matrix $\tilde{\mathcal{P}}$ whose inverse is L-block-banded. We note that it is the inverse of the covariance that is block-banded—the covariance itself is still a full matrix. In the context of image compression, first-order GMrp approximations have been used to model noncausal images. In [2] and [22], for example, an uncorrelated error field is generated recursively by subtracting a GMrp based prediction of the intensity of each pixel in the image from its actual value.

L-block-banded Approximation: The approximated matrix $\tilde{\mathcal{P}}$ is obtained directly from $\mathcal{P}$ in a single step by retaining the significant blocks P_{ij} of $\mathcal{P}$ in $\tilde{\mathcal{P}}$, i.e.,

$$\tilde{\mathcal{P}}_{ij} = P_{ij}, \quad |i-j| \leq L. \quad (49)$$

The nonsignificant blocks in $\tilde{\mathcal{P}}$, if required, are obtained by applying Theorem 3 and using the significant blocks of $\tilde{\mathcal{P}}$ in (49). In [6], it is shown that the GMrp approximations optimize the Kullback-Leibler mean information distance criterion under certain constraints.

The resulting implementations of the KBf obtained by approximating the error field with a GMrp are referred to as the local KBf [18] and [19], where we introduced the local KBf for a first-order GMrp. This corresponds to approximating the inverse of the error covariance matrix (information matrix) with a 1-block-banded matrix. In this section, we derive several implementations of the local KBf using different values of L of the GMrp approximation and explore the tradeoff between the approximation error versus the block bandwidth L in our approximation. We carry this study using the frame-reconstruction problem studied in [20], where a sequence of (100 $\times$ 100) images of the moving tip of a quadratic cone are synthesized. 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.,

$$\psi(s_1, s_2, k+1) = \psi(s_1+0.2, s_2+0.2, k) + w(s_1, s_2, k), \quad (50)$$

where $w(s_1, s_2, k)$ is the forcing term. 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 (50) with the leap-frog method [21]. The dynamical equation in (50) is a simplified case of the thin-plate model with the spatial coherence constraint that is suitable for surface interpolation. We assume that data is available on a few spatial points on adjacent rows of the field ψ, and a different row is observed at each iteration. The initial conditions used in the local KBf are $\widehat{\psi}(0|0) = \underline{0}$ and $\mathcal{P}(0|0) = I_{IJ}$. The forcing term $\mathcal{W}$ and the observation noise $\mathcal{E}$ are both assumed

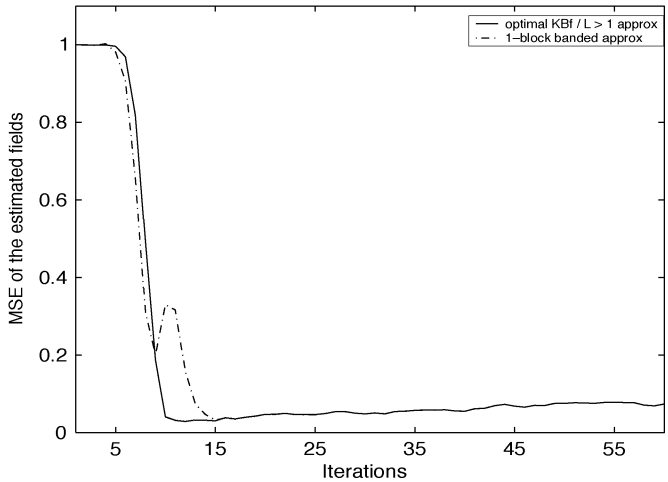


Fig. 3. Comparison of MSEs (normalized with the actual field's energy) of the optimal KBf versus the local KBfs with $L = 1$ to 4. The plots for $(L > 1)$ overlap the plot for the optimal KBf.

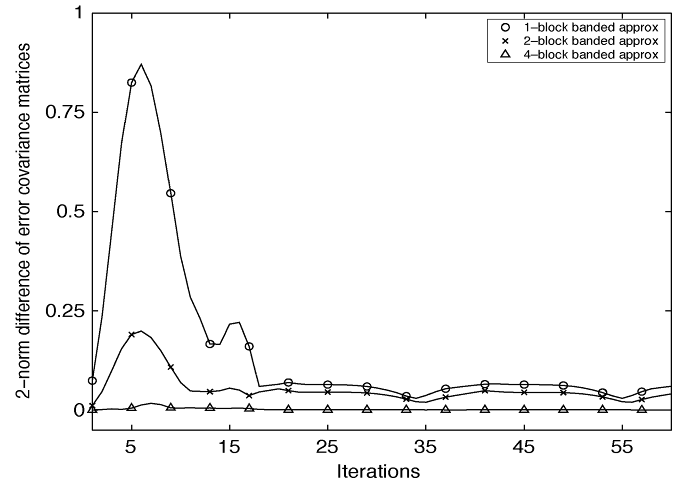


Fig. 4. Comparison of 2-norm differences between error covariance matrices of the optimal KBf versus the local KBfs with $L = 1$ to 4.

independent identically distributed (iid) random variables with Gaussian distribution of zero mean and unit variance or a signal to noise ratio (SNR) of 10 dB. Our discrete state and observation models are different from [20].

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. In each case, the MSE are normalized with the energy present in the field. The solid line in Fig. 3 corresponds to the MSE for the exact KBf, whereas the dotted line is the MSE obtained for the local KBf using a 1-block-banded approximation. The MSE plots for higher order $(L > 1)$ block-banded approximations are so close to the exact KBf that they are indistinguishable in the plot from the MSE of the optimal KBf. It is clear from the plots that the local KBf follows closely the optimal KBf, showing that the reduced-order GMrp approximation is a fairly good approximation to the problem.

To quantify the approximation to the error covariance matrix $\mathcal{P}$, we plot in Fig. 4 the 2-norm difference 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 iterations 6 to 10. The bump reduces and subsequently disappears for higher values of L.

Discussion: The experiments included in this section were performed to make two points. First, we apply the inversion algorithms to derive practical implementations of the KBf. For applications with large dimensional fields, the inversion of the error covariance matrix is computationally intensive precluding the direct implementation of the KBf to such problems. By approximating the error field with a reduced order GMrp, we impose a block-banded structure on the inverse of the covariance matrix (information matrix). Algorithms 1 and 2 invert the approximated error covariance matrix with a much lower computational cost, allowing the local KBf to be successfully implemented. Second, we illustrate that the estimates from the local KBf are in almost perfect agreement with the direct KBf, indicating that the local KBf is a fairly good approximation of the direct KBf. In our simulations, a relatively small value of the block bandwidth L is sufficient for an effective approximation of the covariance matrix. Intuitively, this can be explained by the effect of the strength of the state process and noise on the structure of the error covariance matrix in the KBf. When the process noise $\mathcal{W}$ is low, the covariance matrix approaches the structure imposed by the state matrix $\mathcal{C}$. Since $\mathcal{C}$ is block-banded, it makes sense to update the L-block diagonals of the error covariance matrix. On the other hand, when the process noise is high, the prediction is close to providing no information about the unknown state. Thus, the structural constraints on the inverse of the covariance matrix have little effect. As far as the measurements are concerned, only a few adjacent rows of the field are observed during each time iteration. In the error covariance matrix obtained from the filtering step of the KBf, blocks corresponding to these observed rows are more significantly affected than the others. These blocks lie close to the main block diagonal of the error covariance matrix, which the local KBf updates in any case. As such, little difference is observed between the exact and local KBfs.

V. Summary

The paper derives inversion algorithms for L-block-banded matrices and for matrices whose inverse are L-block-banded. The algorithms illustrate that the inverse of an L-block-banded matrix is completely specified by the first L-block entries adjacent to the main diagonal and any outside entry can be determined from these significant blocks. Any block entry outside the L-block diagonals can therefore be obtained recursively from the block entries within the L-block diagonals. Compared to direct inversion of a matrix, the algorithms provide computational savings of up to two orders of magnitude of the dimension of the constituent blocks used. Finally, we apply our inversion algorithms to approximate covariance matrices in signal processing applications like in certain problems of the Kalman–Bucy filtering (KBf), where the state is large, but a block-banded structure occurs due to the nature of the state equations and the sparsity of the block measurements. In these problems, direct inversion of the covariance matrix is computationally intensive due to the large dimensions of the state fields. The block-banded approximation to the inverse of the error covariance matrix makes the KBf implementation practical, reducing the computational complexity of the KBf by at least two orders of the linear dimensions of the estimated field. Our simulations show that the resulting KBf implementations are practically feasible and lead to results that are virtually indistinguishable from the results of the conventional KBf.

APPENDIX

In the Appendix, we provide proofs for Lemma 1.1 and Theorems 1–3.

Lemma 1.1: Proved by induction on L, the block bandwidth of $\mathcal{A}$.

Case $(L=0)$*:* For $(L=0)$, $\mathcal{A}$ is a block diagonal matrix. Lemma 1.1 implies that $\mathcal{A}$ is block diagonal iff its Cholesky factor $\mathcal{U}$ is also block diagonal. This case is verified by expanding $\mathcal{A}=\mathcal{U}^T\mathcal{U}$ in terms of the constituent blocks $\mathcal{A}=\{A_{ij}\}$ and $\mathcal{U}=\{U_{ij}\}$. We verify the *if* and *only if* statements separately.

If statement: Assume that $\mathcal{U}$ is a block diagonal matrix, i.e., $U_{ij}=0$ for $i\neq j$. By expanding $\mathcal{A}=\mathcal{U}^T\mathcal{U}$, it is straightforward to derive

$$A_{ii}=U_{ii}^TU_{ii} \quad \text{and} \quad A_{ij}=\underline{0}, \quad \text{for} \quad i\neq j.$$

$\mathcal{A}$ is therefore a block diagonal matrix if its Cholesky factor $\mathcal{U}$ is block diagonal.

only if statement: To prove $\mathcal{U}$ to be block diagonal for case $(L=0)$, we use a nested induction on block row i. Expanding the expression $\mathcal{U}^T\mathcal{U}=\mathcal{A}$ gives

$$\sum_{\ell=1}^{\min(i,j)} U_{\ell i}^TU_{\ell j}=A_{ij} \tag{51}$$

for $1\leq i\leq J$ and $i\leq j\leq J$.

For block row $i=1$, (51) reduces to $U_{11}^TU_{1j}=A_{1j}$ for $1\leq j\leq J$. The first block $(j=1)$ on block row $(i=1)$ is given by $U_{11}^TU_{11}=A_{11}$. The block A_{11} is a principal submatrix of a positive definite matrix $\mathcal{A}$; hence, $A_{11}>\underline{0}$. Since U_{11} is a Cholesky factor of A_{11}, hence, $U_{11}>\underline{0}$. The remaining upper triangular blocks $(2\leq j\leq J)$ in the first block row of $\mathcal{U}$ are given by $U_{11}^TU_{1j}=\underline{0}$. Since $U_{11}>\underline{0}$ and is invertible, the off-diagonal entries U_{1j}, $2\leq j\leq J$ on block row $(i=1)$ are, therefore, zero blocks.

By induction, assume that all upper triangular off-diagonal entries on row $i=\tau$ in $\mathcal{U}$ are zero blocks, i.e., $U_{\tau j}=\underline{0}$ for $\tau+1\leq j\leq J$.

For $i=\tau+1$, expression (51)[6]

$$U_{1\tau+1}^TU_{1j}+\ldots+U_{\tau\tau+1}^TU_{\tau j}+U_{\tau+1\tau+1}^TU_{\tau+1j}=A_{\tau+1j} \tag{52}$$

for $(\tau+1)\leq j\leq J$. From the previous induction steps, we know that $U_{1j},U_{2j},\ldots,U_{\tau j}$ in (52) are all zero blocks for $(\tau+1)\leq j\leq J$. Equation (52) reduces to

$$U_{\tau+1\tau+1}^TU_{\tau+1j}=A_{\tau+1j} \quad \text{for} \quad (\tau+1)\leq j\leq J. \tag{53}$$

For $j=\tau+1$, the block $A_{\tau+1\tau+1}>\underline{0}$, implying that its Cholesky factor $U_{\tau+1\tau+1}>\underline{0}$. For $(\tau+2)\leq j\leq J$, (53) reduces to $U_{\tau+1\tau+1}^TU_{\tau+1j}=\underline{0}$, implying that the off-diagonal entries in $\mathcal{U}$ are zero blocks. This completes the proof of Lemma 1.1 for the diagonal case $(L=0)$.

Case $(L=k-1)$*:* By induction, we assume that the matrix $\mathcal{A}$ is $k-1$ block-banded iff its Cholesky factors $\mathcal{U}$ is upper triangular with only its first $k-1$ diagonals consisting of nonzero blocks.

Case $(L=k)$*:* We prove the *if* and *only if* statements separately.

If statement: The Cholesky block $\mathcal{U}$ is upper triangular with only its main and upper k-block diagonals being nonzeros. Expanding $\mathcal{A}=\mathcal{U}^T\mathcal{U}$ gives

$$A_{ij}=\sum_{\ell=1}^{\min(i,j)} U_{\ell i}^TU_{\ell j}. \tag{54}$$

for $1\leq i\leq J$ and $i\leq j\leq J$. We prove the *if* statement for a k block-banded matrix by a nested induction on block row i.

For $(i=1)$, (54) reduces to $A_{1j}=U_{11}^TU_{1j}$. Since the Cholesky blocks U_{1j} are zero blocks for $i+k+1\leq j\leq J$, therefore, A_{1j} must also be $\underline{0}$ for $i+k+1\leq j\leq J$ on block row $(i=1)$.

By induction, we assume $A_{\tau j}=\underline{0}$ for $i=\tau$ and $i+k+1\leq j\leq J$.

For block row $i=\tau+1$, (54) reduces to

$$A_{\tau+1j}=U_{1\tau+1}^TU_{1j}+\ldots+U_{\tau\tau+1}^TU_{\tau j}+U_{\tau+1\tau+1}^TU_{\tau+1j} \tag{55}$$

for $\tau+1\leq j\leq J$.

With block row $i=\tau+1$ and for $(\tau+1)\leq j\leq J$, note that the Cholesky blocks $U_{1j},U_{2j},\ldots,U_{\tau+1j}$ in (55) lie outside the k-block diagonals and are, therefore, $\underline{0}$. Substituting the value of the Cholesky blocks in (55), proves that $A_{\tau+1j}=\underline{0}$, for $(\tau+1+k+1)\leq j\leq J$. Matrix $\mathcal{A}$ is therefore k-block-banded.

Only if statement: Given that matrix $\mathcal{A}$ is k-block-banded, we show that its Cholesky factor $\mathcal{U}$ is upper triangular with only the main and upper k-block diagonals being nonzeros. We prove by induction on the block row i.

For block row $(i=1)$, (51) reduces to

$$U_{11}^TU_{1j}=A_{1j} \quad \text{for} \quad 1\leq j\leq J. \tag{56}$$

For the first entry $(j=1)$, $A_{11}>\underline{0}$, which proves that the Cholesky factor $U_{11}>\underline{0}$ and, hence, is invertible. Substituting $A_{1j}=\underline{0}$ for $(k+2)\leq j\leq J$ in (56), it is trivial to show that the corresponding entries U_{1j} and $(k+2)\leq j\leq J$, in $\mathcal{U}$ are also $\underline{0}$.

By induction, assume $U_{\tau j}=\underline{0}$ for block row $(i=\tau)$ and $(\tau+k+1)\leq j\leq J)$. Substituting block row $i=\tau+1$ in (51) gives (52). It is straightforward to show that the Cholesky blocks $U_{\tau+1j}$ outside the k-bands for $(\tau+k+2)\leq j\leq J$

[6]There is a small variation in the number of terms at the boundary. The proofs for the b.c. follow along similar lines and are not explicitly included here.

are $\underline{0}$ by noting that blocks $U_{1j}, U_{2j}, \ldots, U_{\tau j}$ in (52) are $\underline{0}$ for $(\tau + k + 2) \leq j \leq J$. ■

Theorem 1: Theorem 1 is proved by induction.

Case $(L = 1)$*:* For $L = 1$, Theorem 1 reduces to Corollary 1.1, proved in [18].

Case $(L = k)$*:* By the induction step, Theorem 1 is valid for a k-block-banded matrix, i.e.,

$$\underbrace{\begin{bmatrix} P_{ii} & P_{ii+1} & \cdot & P_{ii+k} \\ P_{i+1i} & P_{i+1i+1} & \cdot & P_{i+1i+k} \\ \cdot & \cdot & \ddots & \cdot \\ P_{i+ki} & P_{i+ki+1} & \cdot & P_{i+ki+k} \end{bmatrix}}_{\mathcal{P}(i:i+k,i:i+k)} \underbrace{\begin{bmatrix} U_{ii}^T \\ U_{i,i+1}^T \\ \vdots \\ U_{i,i+k}^T \end{bmatrix}}_{\mathcal{U}^T(i:i,i:i+k)} = \begin{bmatrix} U_{ii}^{-1} \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix} \quad (57)$$

where the Cholesky blocks in $\mathcal{U}^T(i:i, i:i+k)$ on row i of the Cholesky factor $\mathcal{U}$ are obtained by left multiplying the inverse of the $(kJ \times kJ)$ principal submatrix $\mathcal{P}(i:i+k, i:i+k)$ of $\mathcal{P}$ with the $(kJ \times J)$-block column vector with only one $(J \times J)$ nonzero block entry U_{ii}^{-1} at the top. The dimensions of the principal submatrix $\mathcal{P}(i:i+k, i:i+k)$ and the block column vector containing U_{ii}^{-1} as its first block entry depend on the block bandwidth k or the number of nonzero Cholesky blocks U_{ij} on row i of the Cholesky factor $\mathcal{U}$. Below, we prove Theorem 1 for a $(k+1)$ block-banded matrix.

Case $(L = k + 1)$*:* From Lemma 1.1, a $(k+1)$ block-banded matrix $\mathcal{P}$ has the Cholesky factor $\mathcal{U}$ that is upper triangular and is $(k+1)$ block-banded. Row i of the Cholesky factor $\mathcal{U}$ now has $(k+1)$ nonzero entries $[U_{ii}, \ldots, U_{ii+k+1}]$. By induction from the previous case, these Cholesky blocks can be calculated by left multiplying the inverse of the $((k+1)J \times (k+1)J)$ principal submatrix $\mathcal{P}(i:i+k+1, i:i+k+1)$ of $\mathcal{P}$ with the $((k+1)J \times J)$-block column vector with one nonzero block entry U_{ii}^{-1} at the top as in

$$\underbrace{\begin{bmatrix} P_{ii} & P_{ii+1} & \cdot & P_{ii+k+1} \\ P_{i+1i} & P_{i+1i+1} & \cdot & P_{i+1i+k+1} \\ \cdot & \cdot & \ddots & \cdot \\ P_{i+k+1i} & P_{i+k+1i+1} & \cdot & P_{i+k+1i+k+1} \end{bmatrix}}_{\mathcal{P}(i:i+k+1,i:i+k+1)} \underbrace{\begin{bmatrix} U_{ii}^T \\ U_{ii+1}^T \\ \vdots \\ U_{ii+k+1}^T \end{bmatrix}}_{\mathcal{U}^T(i:i,i:i+k+1)} = \begin{bmatrix} U_{ii}^{-1} \\ \underline{0} \\ \vdots \\ \underline{0} \end{bmatrix}. \quad (58)$$

To prove this result, we express the equality

$$\mathcal{P} = \left(\mathcal{U}^T\mathcal{U}\right)^{-1}, \quad \text{in the form} \quad \mathcal{P}\mathcal{U}^T = \mathcal{U}^{-1}, \quad (59)$$

replace $\mathcal{U} = \{U_{ij}\}$ and $\mathcal{P}$ as $\{P_{ij}\}$, and substitute $\mathcal{U}^{-1}$ from (4). We prove by induction on the block row i, $(J \geq i \geq 1)$.

For block row $(i = J)$, we equate the (J, J) block elements in (59). This gives $P_{JJ}U_{JJ}^T = U_{JJ}^{-1}$, which results in the first b.c. for Theorem 1.

By induction, assume that Theorem 1 is valid for the block row $(i = \tau + 1)$.

To prove Theorem 1 for the block row $(i = \tau)$, equate the $(\tau, \tau), (\tau + 1, \tau), \ldots, (\tau + k + 1, \tau)$ block elements on both sides of (59) as

$$\begin{array}{ll} \text{Block} & \text{Expression} \\ (\tau,\tau) & P_{\tau\tau}U_{\tau\tau}^T + \ldots + P_{\tau\tau+k+1}U_{\tau\tau+k+1}^T = U_{\tau\tau}^{-1} \quad (60) \\ (\tau+1,\tau) & P_{\tau+1}U_{\tau\tau}^T + \ldots + P_{\tau+1\tau+k+1}U_{\tau\tau+k+1}^T = \underline{0} \quad (61) \\ & \vdots \\ (\tau+k+1,\tau) & P_{\tau+k+1\tau}U_{\tau\tau}^T + \ldots \\ & \quad + P_{\tau+k+1\tau+k+1}U_{\tau\tau+k+1}^T = \underline{0} \quad (62) \end{array}$$

which expressed in the matrix-vector notation proves Theorem 1 for block row $(i = \tau)$. ■

Theorem 2: Theorem 2 is proved by induction.

Case $(L = 1)$*:* For $L = 1$, Theorem 2 reduces to Corollary 2.1 proved in [18].

Case $(L = k)$*:* By the induction step, Theorem 2 is valid for a k-block-banded matrix. By rearranging terms, Theorem 2 for $(L = k)$ is expressed as

$$P_{ii} = \left(U_{ii}^T U_{ii}\right)^{-1} - U_{ii}^{-1} \underbrace{[U_{ii+1} \ \ldots \ U_{ii+k}]}_{\mathcal{U}(i,i+1:i+k)} \underbrace{\begin{bmatrix} P_{i+1i} \\ \vdots \\ P_{i+ki} \end{bmatrix}}_{\mathcal{P}(i+1:i+k,i)} \quad (63)$$

$$P_{ij} = -U_{ii}^{-1} \underbrace{[U_{ii+1} \ \ldots \ U_{ii+k}]}_{\mathcal{U}(i,i+1:i+k)} \underbrace{\begin{bmatrix} P_{i+1j} \\ \vdots \\ P_{i+kj} \end{bmatrix}}_{\mathcal{P}(i+1:i+k,j)} \quad (64)$$

where the dimensions of the block row vector $\mathcal{U}(i, i+1:i+k))$ are $(I \times kI)$. The dimensions of the block column vector $\mathcal{P}(i+1:i+k, j))$ are $(kI \times I)$. In evaluating P_{ij}, the procedure for selecting the constituent blocks in the block row vector of $\mathcal{U}$ and the block column vector of $\mathcal{P}$ is straightforward. For $\mathcal{U}(i, i+1:i+k))$, we select the blocks on block row i spanning columns $(i+1)$ through $(i+k)$. Similarly, for $\mathcal{P}(i+1:i+k, j))$, the blocks on block column j spanning rows $(i+1)$ to $(i+k)$ are selected. The number of spanned block rows (or block columns) depends on the block bandwidth k.

Case $(L = k + 1)$*:* By induction, from the $(L = k)$ case, the dimensions of the block column vectors or block row vectors in (63) and (64) would increase by one $(I \times I)$ block. The block row vector derived from $\mathcal{U}$ in (63) now spans block columns i through $(i + k + 1)$ along block row i of $\mathcal{P}$ and is given by $\mathcal{U}(i, i+1:i+k+1)$. The block column vector involving $\mathcal{P}$ in (63) is now given by $\mathcal{P}(i+1:i+k+1, i)$ and spans block rows $(i+1)$ through $(i+k+1)$ of block column i. Below, we verify Theorem 2 for the $(L = k+1)$ case by a nested induction on block row i, $(J \geq i \geq 1)$.

For block row $(i = J)$, Theorem 2 becomes $P_{JJ} = \left(U_{JJ}^T U_{JJ}\right)^{-1}$, which is proven directly by rearranging terms of the first b.c. in Theorem 1.

Assume Theorem 2 is valid for block row $i = \tau + 1$.

TABLE II
CASE ($L = k$): SUBMATRICES NEEDED TO CALCULATE THE NONSIGNIFICANT ENTRY $\mathcal{P}_{ii+\tau}$ IN A k BLOCK BANDED MATRIX

Entry	Dimensions
Block row vector $\mathcal{P}(i, i+1 : i+k)$	$(I \times kI)$
Principal submatrix $\mathcal{P}(i+1 : i+k, i+1 : i+k)$	$(kI \times kI)$
Block column vector $\mathcal{P}(i+1 : i+k, i+r)$	$(I \times kI)$

TABLE III
CASE ($L = k+1$): SUBMATRICES NEEDED TO CALCULATE THE NONSIGNIFICANT ENTRY $\mathcal{P}_{ii+\tau}$ IN A $k+1$ BLOCK BANDED MATRIX

Entry	Dimensions
Block row vector $\mathcal{P}(i, i+1 : i+k+1)$	$(I \times (k+1)I)$
Principal submatrix $\mathcal{P}(i+1 : i+k+1, i+1 : i+k+1)$	$(k+1)I \times (k+1)I)$
Block column vector $\mathcal{P}(i+1 : i+k+1, i+r)$	$(I \times (k+1)I)$

For block row $i = \tau$, Theorem 2 can be proven by right multiplying (60)–(62) with U_{ii}^{-T} and solving for $P_{\tau j}$ for $\tau \leq jJ$. To prove (9), right multiplying (60) by U^{-T} gives

$$P_{\tau\tau} + P_{\tau\tau+1}\left(U_{\tau\tau}^{-1}U_{\tau\tau+1}\right)^T + \dots + P_{\tau\tau+k+1}(U_{\tau\tau}U_{\tau\tau+k+1})^T = \left(U_{\tau\tau}^T U_{\tau\tau}\right)^{-1} \quad (65)$$

which proves Theorem 2 for block $P_{\tau\tau}$. Equation (10) is verified individually for block row $(i = \tau)$ with $(\tau+1) \leq j \leq (\tau+k+1)$ by right multiplying (61) and (62) on both sides by $U_{\tau\tau}^{-T}$ and solving for $P_{\tau\tau+1}, \dots, P_{\tau\tau+k+1}$. ■

Theorem 3: Theorem 3 is proved through induction.

Case $(L = 1)$*:* For $L = 1$, Theorem 3 reduces to Corollary 3.1 proved in [18].

Case $(L = k)$*:* By induction, assume Theorem 3 is valid for $L = k$

$$P_{ij} = [P_{ii+1} \dots P_{ii+k}] \begin{bmatrix} P_{i+1i+1} & \cdot & P_{i+1i+k} \\ \cdot & \ddots & \cdot \\ P_{i+ki+1} & \cdot & P_{i+ki+k} \end{bmatrix}^{-1} \begin{bmatrix} P_{i+1j} \\ \vdots \\ P_{i+kj} \end{bmatrix} \quad (66)$$

for $1 \leq i < (J-k)$ and $(i+k) < j \leq J$. To calculate the nonsignificant entries P_{ii+r}, the submatrices of $\mathcal{P}$ are shown in Table II. The number of blocks selected in each case depends on block bandwidth k.

Case $(L = k+1)$*:* By induction from the previous case, the submatrices required to compute the nonsignificant entry P_{ii+r} of $\mathcal{P}$ with an $(k+1)$ block-banded inverse are appended with the additional blocks shown in Table III, i.e., the block row vector $\mathcal{P}(i, i+1 : i+k+1)$ is appended with an additional block P_{ii+k+1}. The principal submatrix of $\mathcal{P}$ is appended with an additional row consisting of blocks from row $(i+k+1)$ spanning block columns $(i+1)$ through $(i+k+1)$ of $\mathcal{P}$ and an additional column from column $(i+k+1)$ spanning block rows $(i+1)$ through $(i+k+1)$ of $\mathcal{P}$. Similarly, the block column vector $\mathcal{P}(i+1 : i+k+1, i+r)$ has an additional block $P_{i+k+1i+r}$.

We prove Theorem 3 for a $(k+1)$ block-banded matrix by a nested induction on the block row i, $(J-(k+1)) \geq i \geq 1$.

For block row $i = J-(k+1)$, the only nonsignificant entry on the block row $J-(k+1)$ in $\mathcal{P}$ are $P_{J-(k+1),J}$. Theorem 3 for $P_{J-(k+1),J}$ is verified by equating the $(J, J-(k+1))$ block on both sides of the expression $\mathcal{P}\mathcal{U}^T = \mathcal{U}^{-1}$

$$P_{J,J-(k+1)}U_{J-(k+1),J-(k+1)}^T + P_{JJ-k}U_{J-(k+1),J-k}^T + \dots + P_{JJ}U_{J-(k+1)J} = \underline{0} \quad (67)$$

which can be expressed in the form

$$P_{\tilde{J},J} = -\left[\left(U_{\tilde{J},\tilde{J}}^{-1}U_{\tilde{J},\tilde{J}+1}\right) \dots \left(U_{\tilde{J},\tilde{J}}^{-1}U_{\tilde{J},J}\right)\right] \begin{bmatrix} P_{\tilde{J}+1,J} \\ \vdots \\ P_{JJ} \end{bmatrix} \quad (68)$$

where $\tilde{J} = J-(k+1)$. To express the Cholesky product terms in terms of P_{ij}, put $\tau = \tilde{J}$ in (61) and (62) and left multiply with $U_{\tilde{J}\tilde{J}}^{-T}$. By rearranging terms, the resulting expression is

$$\left[\left(U_{\tilde{J}\tilde{J}}^{-1}U_{\tilde{J}\tilde{J}+1}\right) \dots \left(U_{\tilde{J}\tilde{J}}^{-1}U_{\tilde{J}J}\right)\right] = -[P_{\tilde{J},\tilde{J}+1} \dots P_{\tilde{J}J}] \cdot \begin{bmatrix} P_{\tilde{J}+1\tilde{J}+1} & \cdot & P_{\tilde{J}+1J} \\ \cdot & \ddots & \cdot \\ P_{J\tilde{J}+1} & \cdot & P_{JJ} \end{bmatrix}^{-1}. \quad (69)$$

Substituting the value of the Cholesky product terms from (69) in (68) proves Theorem 3 for block row $i = \tilde{J}$.

By induction, assume that Theorem 3 is valid for block row $i = \tau + 1$.

The proof for block row $i = \tau$ is similar to the proof for block row $i = \tilde{J}$. The nonsignificant blocks on row $i = \tau$ are $P_{\tau\tau+k+2}, \dots, P_{\tau J}$. Equating the (j, τ) for $(\tau+k+2 \leq j \leq J)$ block entries on both sides of the equation, gives

$$P_{j,\tau}U_{\tau\tau}^T + P_{j,\tau+1}U_{\tau\tau+1}^T + \dots + P_{j,i+L}U_{\tau\tau+k+1} = \underline{0}. \quad (70)$$

Solving (70) for $P_{j,\tau}$ and then taking the transpose of (70) gives

$$P_{\tau,j} = -\left[\left(U_{\tau\tau}^{-1}U_{\tau\tau+1}\right) \ \dots \ \left(U_{\tau\tau}^{-1}U_{\tau\tau+k+1}\right)\right] \begin{bmatrix} P_{\tau+1,j} \\ \vdots \\ P_{\tau+k+1,j} \end{bmatrix}. \tag{71}$$

To express the Cholesky product terms in terms of P_{ij}, we left multiply (61) and (62) with $U_{\tau\tau}^{-T}$, and rearrange terms. The resulting expression is

$$\begin{aligned}\left[\left(U_{\tau\tau}^{-1}U_{\tau\tau+1}\right) \ \dots \ \left(U_{\tau\tau}^{-1}U_{\tau\tau+k+1}\right)\right] &= -\left[P_{\tau,\tau+1} \ \dots \ P_{\tau,\tau+k+1}\right] \\ &\times \begin{bmatrix} P_{\tau+1\tau+1} & \cdot & P_{\tau+1\tau+k+1} \\ \cdot & \ddots & \cdot \\ P_{\tau+k+1\tau+1} & \cdot & P_{\tau+k+1\tau+k+1} \end{bmatrix}^{-1}. \end{aligned} \tag{72}$$

Substituting the Cholesky product terms from (72) in (71) proves Theorem 3 for block row $i = \tau$. ∎

REFERENCES

[1] J. W. Woods, "Two-dimensional discrete Markovian fields," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 2, pp. 232–240, Mar. 1972.

[2] J. M. F. Moura and N. Balram, "Recursive structure of noncausal Gauss Markov random fields," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 334–354, Mar. 1992.

[3] S. J. Reeves, "Fast algorithm for solving block-banded Toeplitz systems with banded Toeplitz blocks," in *Proc. Int. Conf. Acoust., Speech, Sgnal Process.*, vol. 4, May 2002, pp. 3325–3329.

[4] M. Chakraborty, "An efficient algorithm for solving general periodic Toeplitz system," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 784–787, Mar. 1998.

[5] G. S. Ammar and W. B. Gragg, "Superfast solution of real positive definite Toeplitz systems," *SIAM J. Matrix Anal. Applicat.*, vol. 9, pp. 61–76, Jan. 1988.

[6] A. Kavcic and J. M. F. Moura, "Matrix with banded inverses: algorithms and factorization of Gauss-Markov processes," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1495–1509, Jul. 2000.

[7] S. Chandrasekaran and A. H. Syed, "A fast stable solver for nonsymmetric Toeplitz and quasi-Toeplitz systems of linear equations," *SIAM J. Matrix Anal. Applicat.*, vol. 19, no. 1, pp. 107–139, Jan. 1998.

[8] P. Dewilde and A. J. van der Veen, "Inner-outer factorization and the inversion of locally finite systems of equations," *Linear Algebra Applicat.*, vol. 313, no. 1–3, pp. 53–100, Jul. 2000.

[9] A. M. Erishman and W. Tinney, "On computing certain elements of the inverse of a sparse matrix," *Commun. ACM*, vol. 18, pp. 177–179, Mar. 1975.

[10] I. S. Duff, A. M. Erishman, and W. Tinney, *Direct Methods for Sparse Matrices*. Oxford, U.K.: Oxford Univ. Press, 1992.

[11] G. H. Golub and C. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: ohn Hopkins Univ. Press, 1996, pp. 133–192.

[12] J. Chun and T. Kailath, *Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms*. New York: Springer-Verlag, 1991, pp. 215–236.

[13] N. Kalouptsidis, G. Carayannis, and D. Manolakis, "Fast algorithms for block-Toeplitz matrices with Toeplitz entries," *Signal Process.*, vol. 6, no. 3, pp. 77–81, 1984.

[14] T. Kailath, S.-Y. Kung, and M. Morf, "Displacement ranks of matrices and linear equations," *J. Math. Anal. Applicat.*, vol. 68, pp. 395–407, 1979.

[15] D. A. Bini and B. Meini, "Toeplitz system," *SIAM J. Matrix Anal. Applicat.*, vol. 20, no. 3, pp. 700–719, Mar. 1999.

[16] A. E. Yagle, "A fast algorithm for Toeplitz-block-Toeplitz linear systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, Phoenix, AZ, May 2001, pp. 1929–1932.

[17] C. Corral, "Inversion of matrices with prescribed structured inverses," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, May 2002, pp. 1501–1504.

[18] A. Asif and J. M. F. Moura, "Data assimilation in large time-varying multidimensional fields," *IEEE Trans. Image Process.*, vol. 8, no. 11, pp. 1593–1607, Nov. 1999.

[19] ——, "Fast inversion of L-block-banded matrices and their inverses," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, Orlando, FL, May 2002, pp. 1369–1372.

[20] T. M. Chin, W. C. Karl, and A. S. Willsky, "Sequential filtering for multi-frame visual reconstruction," *Signal Process.*, vol. 28, no. 1, pp. 311–333, 1992.

[21] J. C. Strikwerda, *Finite Element Schemes and Partial Differential Equations*. Pacific Grove, CA: Wadsworth, Brooks/Cole, 1989.

[22] A. Asif and J. M. F. Moura, "Image codec by noncausal prediction, residual mean removal, and cascaded VQ," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 1, pp. 42–55, Feb. 1996.

Amir Asif (M'97–SM'02) received the B.S. degree in electrical engineering with the highest honor and distinction in 1990 from the University of Engineering and Technology, Lahore, Pakistan, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA in 1993 and 1996, respectively.

He has been an Associate Professor of computer science and engineering at York University, North York, ON, Canada, since 2002. Prior to this, he was on the faculty of Carnegie Mellon University, where he was a Research Engineer from 1997 to 1999 and the Technical University of British Columbia, Vancouver, BC, Canada, where he was an Assistant/Associate Professor from 1999 to 2002. His research interests lie in statistical signal processing (one and multidimensional), image and video processing, multimedia communications, and genomic signal processing. He has authored over 30 technical contributions, including invited ones, published in international journals and conference proceedings.

Dr. Asif has been a Technical Associate Editor for the IEEE SIGNAL PROCESSING LETTERS since 2002. He has served on the advisory committees of the Science Council of British Columbia from 2000 to 2002. He has organized two IEEE conferences on signal processing theory and applications and served on the technical committees of several international conferences. He has received several distinguishing awards including the 2004 Excellence in Teaching Award from the Faculty of Science and Engineering and the 2003 Mildred Baptist teaching excellence award from the Department of Computer Science and Engineering. He is a member of the Professional Engineering Society of Ontario.

José M. F. Moura (S'71–M'75–SM'90–F'94) received the engenheiro electrotécnico degree in 1969 from Instituto Superior Técnico (IST), Lisbon, Portugal, and the M.Sc., E.E., and the D.Sc. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology (M.I.T.), Cambridge, in 1973 and 1975, respectively.

He is a Professor of electrical and computer engineering at Carnegie Mellon University, Pittsburgh, PA, since 1986. From 1999 to 2000, he was a Visiting Professor of electrical engineering at M.I.T. Prior to this, he was on the faculty of IST from 1975 to 1984. He was visiting Genrad Associate Professor of electrical engineering and computer science at M.I.T. from 1984 to 1986 and a Visiting Research Scholar at the Department of Aerospace Engineering, University of Southern California, Los Angeles, during the Summers of 1978 to 1981. His research interests include statistical signal, image, and video processing, and communications. He has published over 270 technical contributions, is the co-editor of two books, holds six patents on image and video processing, and digital communications with the US Patent Office, and has given numerous invited seminars at US and European Universities and Laboratories.

Dr. Moura served as Vice-President for Publications for the IEEE Signal Processing Society (SPS), where he was a Board of Governors elected member from 2000 to 2002. He was also Vice-President for Publications for the *IEEE Sensors Council* from 2000 to 2002. He is on the Editorial Board of the IEEE PROCEEDINGS, the IEEE SIGNAL PROCESSING MAGAZINE, and the *ACM Transactions on Sensor Networks*. He chaired the IEEE TAB Transactions Committee from 2002 to 2003 that joins the about 80 Editors-in-Chief of the IEEE Transactions. He was the Editor-in-Chief for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1995 to 1999 and Acting Editor-in-Chief for the IEEE SIGNAL PROCESSING LETTERS from December 2001 to May 2002. He has been a member of several Technical Committees of the SPS. He was on the IEEE Press Board from 1991 to 1995.

Dr. Moura is corresponding member of the Academy of Sciences of Portugal (Section of Sciences). He was awarded the 2003 IEEE Signal Processing Society meritorious service award, and in 2000, the IEEE Millenium medal. He is affiliated with several IEEE societies, Sigma Xi, AMS, AAAS, IMS, and SIAM.