

6-3-2007

Numerically Stable Algorithms for Inversion of Block Tridiagonal and Banded Matrices

Jitesh Jain

Purdue University, jjain@purdue.edu

Hong Li

Purdue University, li73@purdue.edu

Stephen Cauley

Purdue University, stcauley@purdue.edu

Cheng-Kok Koh

Purdue University, chengkok@purdue.edu

Venkataramanan Balakrishnan

Purdue University, ragu@purdue.edu

Follow this and additional works at: <http://docs.lib.purdue.edu/ecetr>

Jain, Jitesh; Li, Hong; Cauley, Stephen; Koh, Cheng-Kok; and Balakrishnan, Venkataramanan, "Numerically Stable Algorithms for Inversion of Block Tridiagonal and Banded Matrices" (2007). *ECE Technical Reports*. Paper 357.
<http://docs.lib.purdue.edu/ecetr/357>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Numerically Stable Algorithms for Inversion of Block Tridiagonal and Banded Matrices

Jitesh Jain*, Stephen Cauley, Hong Li, Cheng-Kok Koh
and Venkataramanan Balakrishnan

School of Electrical and Computer Engineering
Purdue University, West Lafayette, IN 47907-1285
{jjain,stcauley,li73,chengkok,ragu}@ecn.purdue.edu

November 10, 2006

Summary

We provide a new representation for the inverse of block tridiagonal and banded matrices. The new representation is shown to be numerically stable over a variety of block tridiagonal matrices, in addition of being more computationally efficient than the previously proposed techniques. We provide two algorithms for commonly encountered problems that illustrate the usefulness of the results.

KEY WORDS: Block Tridiagonal matrices, semiseparable matrices, representation, stability

1 Introduction

Sparse matrices are encountered in a variety of applications from areas such as applied mathematics and physics to engineering. The underlying numerical problems typically involve solving sparse systems of linear equations, sparse matrix inversion, or sparse eigenvalue computation. There exist several techniques for exploiting general sparsity structures in solving these numerical problems. However, the specific sparsity structure that arises in the solution of integral equations or boundary value problems (block-tridiagonal, block-banded, or variations thereof [17, 18]) is important enough to warrant more specialized approaches. In this paper, we present a compact representation, computable in a numerically stable way, for the inverse of block-tridiagonal and banded matrices. We also demonstrate the advantages in computation with this new representation.

There are a number of elegant theoretical results describing the structure of the inverses of block tridiagonal and block-banded matrices. Representations for the inverses of tridiagonal, banded, and

*Correspondence to: Jitesh Jain, School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907-1285

block tridiagonal matrices can be found in [5, 6, 16, 20, 22, 31]. It has been shown that the inverse of a tridiagonal matrix can be compactly represented by two sequences $\{u_i\}$ and $\{v_i\}$ [2, 3, 4, 21]. This result was extended to the cases of block tridiagonal and banded matrices in [23, 25, 26], where the $\{u_i\}$ and $\{v_i\}$ sequences generalized to matrices $\{U_i\}$ and $\{V_i\}$. Matrices which can be represented in this fashion are more generally known as semiseparable matrices [24, 26].

While the underlying mathematics is theoretically elegant, the computation of parameters $\{u_i\}$ and $\{v_i\}$ is beset by numerical problems for even modest-sized problems [12]. The root cause is that $\{u_i\}$ and $\{v_i\}$ grow exponentially [14, 21] with i . In fact, for matrices with sizes as small as 1000, the computation of $\{u_i\}$ and $\{v_i\}$ becomes unstable (due to overflow and underflow) with computers using standard double-precision arithmetic. One approach that has been successful in eliminating this problem, for the tridiagonal case, is the “ratio” approach [13]. Here, the *ratios* of sequential elements of the $\{u_i\}$ and $\{v_i\}$ sequences are used to describe the inverse of a tridiagonal matrix. Such an approach is numerically stable for matrices of very large sizes, of order millions. The extension of this ratio approach to the general block-tridiagonal case was discussed by the same authors in [20]. The authors used the block factorization of the original block-tridiagonal matrix to construct a block Cholesky decomposition of its inverse. While this approach leads to a stable computation of the matrix ratios, the construction of entries of the inverse from the matrix ratios is unstable. Recently, an alternative definition of semiseparable matrices was introduced in [32]. A new representation based on the alternative definition used $n - 1$ Givens transformations and a vector of length n to represent a semiseparable matrix. Such a representation was shown to preserve all the properties of a general semiseparable matrix; the associated computation was shown to be numerically stable. However, the procedure of calculating the Givens transformations is computationally expensive.

In this paper, we offer a compact representation for the inverses of block tridiagonal and banded matrices that can be constructed in computationally efficient and numerical stable manner. We provide a new representation for the inverse of block-tridiagonal matrices through the use of two sequences $\{D_i\}$ and $\{S_i\}$. Here, $\{D_i\}$ represents the diagonal blocks of the inverse and $\{S_i\}$ the ratios of sequential elements of the sequence $\{V_i\}$. Furthermore, we propose an algorithm for finding individual block entries of the inverse in a numerically stable way. The algorithm applies to general matrices, and is not restricted to diagonally dominant or positive definite matrices. In addition, our representation can be applied to matrices with either singular or rectangular off-diagonal blocks. Finally, diagonal matrices, which are not considered to be in the class of semiseparable matrices [32], can be included as a special case under this representation. As will be seen in §6, this new formulation produces accurate results for large problems that the previously proposed approaches fail to handle due to numerical instability. For simplicity of illustration we will restrict ourselves to the case of real symmetric block-tridiagonal matrices. We note that all results in this paper can be directly extended to the non-symmetric case with only slight modification; as was shown in [27], results for banded matrices can essentially be obtained from a block-tridiagonal formulation.

The remainder of the paper is organized as follows. In §2, we give a brief description of semiseparable matrices and numerical stability issues encountered while forming such matrices. Previous attempts for dealing with such numerical instability are detailed in §3. In §4, we provide a

description of our approach. Stable algorithms for two commonly encountered numerical problems involving semiseparable matrices are presented in §5. Finally, in §6, we establish the effectiveness of the new representation via numerical experiments.

2 Semiseparable Matrices

Let A be a symmetric block-tridiagonal matrix of the form

$$A = \begin{pmatrix} A_1 & -B_1 & & & \\ -B_1^T & A_2 & -B_2 & & \\ & \ddots & \ddots & \ddots & \\ & & -B_{N_y-2}^T & A_{N_y-1} & -B_{N_y-1} \\ & & & -B_{N_y-1}^T & A_{N_y} \end{pmatrix}, \quad (1)$$

where each $A_i, B_i \in \mathbb{C}^{N_x \times N_x}$. Thus, $A \in \mathbb{C}^{N_y N_x \times N_y N_x}$ with N_y diagonal blocks of size N_x each. When A is proper, i.e., when B_i are nonsingular [4], there exists two (non-unique) sequences of matrices $\{U_i\}$ and $\{V_i\}$ such that for $j \geq i$

$$(A^{-1})_{ij} = U_i V_j^T.$$

Hence, A^{-1} can be written as

$$A^{-1} = \begin{pmatrix} U_1 V_1^T & U_1 V_2^T & U_1 V_3^T & \cdots & U_1 V_{N_y}^T \\ V_2 U_1^T & U_1 V_2^T & U_2 V_3^T & \cdots & U_2 V_{N_y}^T \\ V_3 U_1^T & V_3 U_2^T & U_3 V_3^T & \cdots & U_3 V_{N_y}^T \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ V_{N_y} U_1^T & V_{N_y} U_2^T & V_{N_y} U_3^T & \cdots & U_{N_y} V_{N_y}^T \end{pmatrix},$$

where $U_i, V_i \in \mathbb{C}^{N_x \times N_x}$. The $\{U_i\}$ and $\{V_i\}$ sequences can be computed in $O(N_y N_x^3)$ operations in the following manner:

$$\begin{aligned} U_1 &= I_{N_x \times N_x}, & U_2 &= B_1^{-1} A_1, \\ U_{i+1} &= B_i^{-1} (A_i U_i - B_{i-1}^T U_{i-1}), & i &= 2, \dots, N_y - 1, \\ V_{N_y}^T &= (A_{N_y} U_{N_y} - B_{N_y-1}^T U_{N_y-1})^{-1}, & V_{N_y-1}^T &= V_{N_y}^T A_{N_y} B_{N_y-1}^{-1}, \\ V_i^T &= (V_{i+1}^T A_{i+1} - V_{i+2}^T B_{i+1}^T) B_i^{-1}, & i &= N_y - 2, \dots, 1. \end{aligned}$$

Semiseparable matrices arise in a variety of applications that involve integral equations as well as studies of vibrational analysis statistics and rational interpolation. Computationally efficient algorithms have been developed for solving systems of linear equations where the coefficient matrix is a diagonal plus a semiseparable matrix [15, 19]. Algorithms to calculate the eigendecomposition of semiseparable matrices plus block diagonal matrices have been developed in [8, 10].

There has also been substantial research into the specific attributes of semiseparable matrices. It was shown in [14, 21] that the entries of the inverse of a symmetric positive definite matrix decay rapidly away from the diagonal, and are in fact bounded by an exponentially decaying function

along any row or column. This gives rise to a problem with overflow and underflow during the calculation of the $\{U_i\}$ and $\{V_i\}$ sequences. It was shown in [12] that the semiseparable representation suffers from instabilities, making it of limited practical use, especially for large sized problems. This conclusion was further supported in a recent paper [32].

3 Previous Approaches

To address the issue of numerical instability in computation with semiseparable matrices, the authors in [20, 13] proposed the following block factorization of A . Denote by L the block lower part of A . Then

$$A = (\Delta + L) \Delta^{-1} (\Delta + L^T) = (\Sigma + L^T) \Sigma^{-1} (\Sigma + L).$$

Here, Δ and Σ are block diagonal matrices whose diagonal blocks are described by following recurrences

$$\begin{aligned} \Delta_1 &= A_1, & \Delta_i &= A_i - B_{i-1}^T \Delta_{i-1}^{-1} B_{i-1}, \\ \Sigma_{N_y} &= A_{N_y}, & \Sigma_i &= A_i - B_i \Sigma_{i+1}^{-1} B_i^T. \end{aligned} \quad (2)$$

The j^{th} block column of A^{-1} can then be determined as follows,

$$\begin{aligned} A_j^{-1} &= B_{j-1}^{-1} \Delta_{j-1} \dots B_1^{-1} \Delta_1 \Sigma_1^{-1} B_1 \dots \Sigma_{j-1}^{-1} B_{j-1} \Sigma_j^{-1}, \\ A_{j-l}^{-1} &= \left(B_{j-l-1}^{-1} \Delta_{j-l-1} \dots B_1^{-1} \Delta_1 \right) \left(\Sigma_1^{-1} B_1 \dots B_{j-1} \Sigma_j^{-1} \right), \quad l = 1, \dots, j-1, \\ A_{j+l}^{-1} &= \left(B_{j+l}^{-T} \Sigma_{j+l+1} \dots B_{N_y-1}^{-T} \Sigma_{N_y} \right) \left(\Delta_{N_y}^{-1} B_{N_y-1} \dots \Delta_{j+1}^{-1} B_j^T \Delta_j^{-1} \right), \quad l = 1, \dots, N_y - j. \end{aligned} \quad (3)$$

The authors used a block Cholesky decomposition for determining the inverse of diagonally dominant or positive definite block-tridiagonal matrices. Although the above formulations lead to stable factorizations, they still suffer from numerical instabilities during construction of the actual entries of the inverse. Specifically, the block factorizations given in (2) are stable, but combining them to form the entries of A^{-1} in accordance with (3) leads to numerical instability (see §6).

Recently, a new class of semiseparable matrices, called sequentially semiseparable matrices, along with a new set of algorithms has been introduced [9, 11]. Specifically, let C denote a matrix of sequentially semiseparable structure. Then we can block partition C as

$$C_{ij} = \begin{cases} D_i & \text{if } i = j \\ U_i W_{i+1} \dots W_{j-1} V_j^T & \text{if } j > i \\ V_j W_{j-1}^T \dots W_{i+1}^T U_i^T & \text{if } i > j \end{cases}$$

In case when all W_i are identity matrices, C reduces to a semiseparable plus a block diagonal matrix. This formulation inherits the same numerical instability problem of semiseparable matrices.

The authors in [32] propose a new definition for semiseparable matrices in order to address the problem of numerical instability. For a semiseparable matrix of dimension n , this representation

consists of $n - 1$ Givens transformations and a vector of length n . The Given transformations are given by

$$G = \begin{pmatrix} c_1 & c_2 & \cdots & c_{n-1} \\ s_1 & s_2 & \cdots & s_{n-1} \end{pmatrix}, \quad \text{and} \quad D = (d_1 \ d_2 \ \cdots \ d_n). \quad (4)$$

The semiseparable matrix in this case can be simply written as

$$S = \begin{pmatrix} c_1 d_1 & c_2 s_1 d_1 & c_3 s_2 s_1 d_1 & \cdots & c_{n-1} s_{n-2} \cdots s_1 d_1 & s_{n-1} s_{n-2} \cdots s_1 d_1 \\ c_2 s_1 d_1 & c_2 d_2 & c_3 s_2 s_1 d_1 & \cdots & \vdots & \vdots \\ c_3 s_2 s_1 d_1 & c_3 s_2 s_1 d_1 & c_3 d_3 & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & c_{n-1} s_{n-2} d_{n-2} & c_n s_{n-1} s_{n-2} d_{n-2} \\ c_{n-1} s_{n-2} \cdots s_1 d_1 & \cdots & \cdots & \cdots & c_{n-1} d_{n-1} & s_{n-1} d_{n-1} \\ s_{n-1} s_{n-2} \cdots s_1 d_1 & \cdots & \cdots & \cdots & s_{n-1} d_{n-1} & d_n \end{pmatrix}. \quad (5)$$

This factorization was shown to be numerically stable. However, the procedure for calculating the Givens transformations and the process of determining the vector elements d_i both involve calculating norms, which can become computationally prohibitive. Also, it is not evident how the procedure for calculating these entries can be modified to include the case when the off-diagonal blocks of the block tridiagonal matrix are rectangular.

4 Our Algorithm

We propose the use of ratios of sequential elements of the sequences $\{U_i\}$ and $\{V_i\}$ for the stable computation of A^{-1} . These sequences, $\{R_i\}$ and $\{S_i\}$, are defined as follows:

$$U_i = R_i U_{i+1}, \quad V_{i+1}^T = V_i^T S_i.$$

The sequences $\{R_i\}$ and $\{S_i\}$ can be computed in $O(N_y N_x^3)$ operations by use of the following numerically stable recursions:

$$\begin{aligned} R_1 &= A_1^{-1} B_1, & R_i &= (A_i - B_{i-1}^T R_{i-1})^{-1} B_i, \quad i = 2, \dots, N_y - 1, \\ S_{N_y-1} &= B_{N_y-1} A_{N_y}^{-1}, & S_i &= B_i (A_{i+1} - S_{i+1} B_{i+1}^T)^{-1}, \quad i = N_y - 2, \dots, 1. \end{aligned} \quad (6)$$

It is important to note that the above recursions can be seen as an extension of the ratio sequences introduced for tridiagonal matrices in [13]. However, there still remains a need for an algorithm to determine the individual block entries of the inverse in a numerically stable fashion.

It is readily verified that the diagonal blocks of A^{-1} , denoted D_i , are given by the recursion

$$\begin{aligned} D_1 &= (A_1 - B_1 S_1^T)^{-1}, & D_{i+1} &= (A_{i+1} - B_{i+1} S_{i+1}^T)^{-1} (I + B_i^T D_i S_i), \quad i = 1, \dots, N_y - 2, \\ D_{N_y} &= A_{N_y}^{-1} \left(I + B_{N_y-1}^T D_{N_y-1} S_{N_y-1} \right). \end{aligned}$$

Then, the remaining block entries can be computed in a numerically stable way as follows.

$$A_{ij}^{-1} = D_i S_{i+1} S_{i+2} \cdots S_j, \quad j > i, \quad A_{ij}^{-1} = R_{j+1} R_{j+2} \cdots R_i D_i, \quad j < i.$$

While the Σ and Δ sequences in (2) and the R and S sequences in (6) may appear to share similarities, we highlight the differences. First, the algorithm in (2) requires the matrix to be proper, whereas there are no such restrictions imposed on the ratio sequences. This enables the ratio sequences to be used for matrices with either rectangular or singular off-diagonal blocks. Also, the approach used in finding actual entries of the inverse, from the respective sequences, is clearly different. As will be shown in §6, the numerical stability of the method described above differs significantly from the algorithm based upon the Σ , Δ representation suggested in [20]. Also, given a semiseparable matrix, the proposed procedure for calculating $\{D_i, R_i, S_i\}$ is straightforward and less computationally expensive than the computation of the Givens transformations for the new representation proposed in [32]. Calculation of the Givens transformation involves calculating norms and is hence time consuming. For our representation, the elements of sequence $\{D_i\}$ are equal to the corresponding block diagonal of the semiseparable matrix. Calculating $\{R_i, S_i\}$ involves multiplying the inverse of D_i with a corresponding off-diagonal block of the semiseparable matrix. Hence, the computation involved for our representation is considerably less than the one required with the approach in [32]; we will support this assertion via numerical examples in §6.

In the case under consideration (real, symmetric), either of the two sets $\{D_i, R_i\}$ or $\{D_i, S_i\}$ can be used as a valid representation of the inverse. In the remainder of the paper we will be using the $\{D_i, S_i\}$ sequences as our representation for a semiseparable matrix.

5 Fast Multiplications

We demonstrate the utility of the new parametrization of the inverse on two standard problems that typically arise in simulation.

5.1 Fast computation of $A^{-1}x$

The first problem is the computation of the product of the inverse of a block-tridiagonal matrix A and a vector x , i.e. evaluating $A^{-1}x$. For general matrices A , the best methods solving $Ay = x$ are iterative methods such as GMRES [29]. For block-tridiagonal A , the parametrization of A^{-1} given in §4 provides a *direct* method for computing $A^{-1}x$. We will establish via numerical experiments in §6 that this direct method is competitive with GMRES for solving a *single* block-tridiagonal system of equations $Ax = b$. Thus, there are clear advantages to our direct method when solving $Ay = x$ for multiple right-hand sides x .

We present pseudo-code for computing the product $A^{-1}x$, where A^{-1} is represented by $\{D_i, S_i\}$ sequences. For any vector z , we will use $z_{[i]}$ to denote the i th block-vector of size N_x , i.e., with z_i denoting the i th component of z , $z_{[i]} = [z_{(i-1)N_x+1}, z_{(i-1)N_x+2}, \dots, z_{iN_x}]^T$, $i = 1, 2, \dots, N_y$.

```
function y = Ainvx(D_i, S_i, x)
    P[N_y] = x[N_y];
    for i = (N_y - 1) downto 1 {
        P[i] = x[i] + S_i P[i+1];
```

```

    }
    y[1] = D1p[1];
    q[1] = S1TD1x[1];
    for i = 2 to (Ny - 1) {
        q[i] = SiT(q[i-1] + Dix[i]);
        y[i] = Dip[i] + q[i-1];
    }
    y[Ny] = DNyp[Ny] + q[Ny-1];
return y;

```

The above algorithm takes computational time of $O(N_y N_x^2)$ as compared to $O(N_y^2 N_x^2)$ otherwise.

5.2 Fast computation of the diagonal blocks of $A^{-1}\Sigma A^{-*}$

The second problem is the computation of the block diagonal entries of a matrix $A^{-1}\Sigma A^{-*}$, where Σ is another block diagonal matrix. This problem plays an important role in computational nano-electronics, specifically while solving for the current-voltage relationships of nanotransistors [30]. The following pseudocode provides an efficient algorithm for this computation. The following notation is used: With Σ regarded as a block-matrix with blocks of size $N_x \times N_x$, Σ_{a_i} denotes the (i, i) block of Σ , and Σ_{b_i} the $(i, i+1)$ block.

```

function M = ASA*(Di, Si, Σai, Σbi)
    for i = 1 to Ny {
        Ji = DiΣaiDi*;
    }
    KNy = JNy;
    for i = (Ny - 1) downto 1 {
        Ki = Ji + SiTKi+1Si*T - SiTDi+1ΣbiTDi* - DiΣbiDi+1*Si*T;
    }
    M1 = K1;
    L1 = J1;
    for i = 2 to Ny {
        Li = Si-1TLi-1Si-1*T + Ji - Si-1TDi-1Σbi-1TDi* - DiΣbi-1TDi-1Si-1*T;
        Mi = Ki + Li - Ji;
    }
return M;

```

The above algorithm requires $O(N_y N_x^3)$ computation as compared to $O(N_y^3 N_x^3)$ otherwise.

6 Numerical Experiments

We present a comparison of the ratio-based approach presented in this paper with several existing techniques. We begin by comparing the stability and computational efficiency of the algorithm

presented in §4 for obtaining a ratio-based parametrization of semiseparable matrices, against existing techniques. We then demonstrate the stability of the ratio-based approach for increasing problem sizes, for various classes of matrices. Indeed, we show that for some cases, the direct solution of $Ax = b$ by computing $A^{-1}b$ using our ratio-based approach competes favorably with the best-known iterative methods for solving systems of linear equations.

All numerical experiments were performed in MATLAB running on a Intel[®] Pentium[®] 4 CPU 1.5GHz machine.

6.1 Generating a compact representation of a semiseparable matrix

We provide results on computational requirements for the calculation of our compact representation starting from either a semiseparable matrix or its inverse. For the comparison with [32], we used the author’s implementation, taken from [1]. Given the semiseparable matrix, we calculate the D_i and S_i sequences for our “ratio-based” representation, and the d_i and Givens rotation matrices for the new representation proposed in [32]. The results are shown in Table 1.

N_y	Algorithm in [32]	Ratio-based
200	0.19	0.00
400	3.28	0.01
800	24.12	0.02
1600	201.94	0.04

Table 1: Computation time in seconds ($N_x = 1$; tridiagonal case).

As the problem size increases, the computational cost for calculating the Givens matrix and d_i sequence becomes increasingly prohibitive. As expected, the computational complexity for calculating the ratio based representation is linear with respect to N_y .

6.2 Stability issues

We now explore the stability of the compact representation of the inverse of block-tridiagonal matrices proposed in this paper. We first present a comparison with the algorithm in [13, 20]. For this analysis we consider the case of a block-tridiagonal matrix, where the diagonal blocks are tridiagonal and the off-diagonal blocks are diagonal. The results are shown in Table 2, where accuracy is measured by the square of the Frobenius norm for the matrix product AA^{-1} . As problem size increases, the numerical instabilities in the computation of desired entries of A^{-1} using [20] become evident.

We next explore the numerical stability and performance of our representation for large problem sizes, for a number of different classes of matrices. For large problem sizes, it is difficult to directly quantify the accuracy with which $AA^{-1} = I$ is satisfied. Therefore, we have taken the route of comparing a random column in the inverse computed using our ratio-based method with the solution obtained via GMRES with an incomplete LU factorization-based preconditioner [28]. Given

Size ($N_x \times N_y$)	Algorithm in [20]	Ratio-based
10 × 50	500.00	500.00
20 × 50	1000.00	1000.00
10 × 100	1000.00	1000.00
20 × 100	2000.00	2000.00
10 × 200	2000.00	2000.00
20 × 200	4000.00	4000.00
10 × 400	7.1067×10^{12}	4000.00
20 × 400	2.1007×10^{14}	8000.00

Table 2: Values of $(AA^{-1})_F^2$.

a matrix A the error in finding the p th column of A^{-1} is defined as $\|(Ax - b)\|$, where b is the p th column of identity matrix and $x = A^{-1}b$. For GMRES, we solve $Ax = b$ with a relative tolerance of 10^{-10} and the reported error values are again $\|(Ax - b)\|$. We also present the computation time with each method. It is worth noting that our method generates the complete inverse in the implicit form (D, S sequences) and calculating any other column of the inverse after this requires very less computational effort.

- *Sparse positive-definite Toeplitz matrices:* We consider finding the inverse of a block-tridiagonal matrix A with $A_i = A_j, B_i = B_j \forall i, j$ (see (1)). We also assume that A_i is tridiagonal and B_i is diagonal. The results are shown in Table 3. It is evident that preconditioned GMRES easily outperforms the ratio-based approach (as a tool to solve linear equations) in this case. This is consistent with the excellent performance of GMRES for solving sparse positive-definite systems of equations. We note that the time taken by the ratio-based algorithm is consistent with the associated computation of $O(N_y N_x^3)$.

Size ($N_x \times N_y$)	Error = $\ (Ax - b)\ $		Time (in sec)	
	GMRES	Ratio-based	GMRES	Ratio-based
20 × 1000	3.56×10^{-13}	1.91×10^{-17}	.68	.80
20 × 2000	4.09×10^{-13}	2.24×10^{-16}	1.69	1.37
40 × 1000	4.04×10^{-13}	2.22×10^{-16}	1.27	1.95
40 × 2000	2.43×10^{-13}	4.60×10^{-17}	2.41	3.60
80 × 1000	5.79×10^{-14}	1.84×10^{-17}	2.61	8.34
80 × 2000	4.75×10^{-14}	1.70×10^{-17}	6.72	16.57
160 × 1000	1.17×10^{-12}	2.25×10^{-16}	9.17	57.63
160 × 2000	8.16×10^{-13}	1.34×10^{-17}	16.1	142.03

Table 3: Sparse positive-definite Toeplitz case: Error values and computational time as compared with results obtained with GMRES.

- *General sparse positive-definite matrices:* Here we consider finding the inverse of a positive definite matrix A , such that each of the diagonal block, A_i is tridiagonal and each of the off-diagonal block B_i is diagonal. Results are shown in Table 4.

Size ($N_x \times N_y$)	Error = $\ (Ax - b)\ $		Time (in sec)	
	GMRES	Ratio-based	GMRES	Ratio-based
20×1000	3.60×10^{-11}	2.94×10^{-10}	.55	.68
20×2000	6.50×10^{-15}	2.89×10^{-15}	.74	1.19
40×1000	9.90×10^{-13}	5.61×10^{-11}	.85	1.69
40×2000	4.50×10^{-15}	8.55×10^{-15}	1.61	3.13
80×1000	1.70×10^{-11}	4.32×10^{-10}	2.17	8.63
80×2000	1.40×10^{-11}	5.06×10^{-10}	4.30	18.61
160×1000	3.60×10^{-11}	1.29×10^{-11}	5.98	65.97
160×2000	1.80×10^{-11}	4.91×10^{-11}	9.93	153.00

Table 4: General sparse positive-definite case: Error values and computational time as compared with results obtained with GMRES.

- *General sparse matrices:* Here we consider finding the inverse of a matrix A (not necessarily positive-definite), such that each of the diagonal block, A_i is tridiagonal and each of the off-diagonal block B_i is diagonal. The results are shown in Table 5. It is evident that the ratio-based technique outperforms preconditioned GMRES in this case.

Size ($N_x \times N_y$)	Error = $\ (Ax - b)\ $		Time (in sec)	
	GMRES	Ratio-based	GMRES	Ratio-based
20×1000	7.19×10^{-12}	5.02×10^{-13}	2.21	.78
20×2000	1.49×10^{-11}	1.51×10^{-13}	4.23	1.38
40×1000	1.21×10^{-12}	2.04×10^{-12}	10.70	1.97
40×2000	1.91×10^{-12}	1.16×10^{-9}	21.53	3.94
80×1000	7.16×10^{-12}	3.07×10^{-9}	43.19	8.85
80×2000	1.07×10^{-12}	4.77×10^{-9}	88.78	17.78
160×1000	9.32×10^{-12}	1.98×10^{-10}	175.53	53.21
160×2000	6.64×10^{-13}	4.58×10^{-10}	396.48	140.52

Table 5: General sparse case: Error values and computational time as compared with results obtained with GMRES.

- *General Matrices:* Here we consider finding the inverse of a general block-tridiagonal matrix A . The results are shown in Table 6. We were unable to run cases with higher values of N_x with GMRES, owing to prohibitive memory requirements; the corresponding entries in the table are marked ”-”.

Size ($N_x \times N_y$)	Error = $\ (Ax - b)\ $		Time (in sec)	
	GMRES	Ratio-based	GMRES	Ratio-based
20×1000	1.60×10^{-13}	9.02×10^{-12}	5.08	.75
20×2000	2.90×10^{-14}	2.40×10^{-11}	8.03	1.46
40×1000	4.50×10^{-15}	8.55×10^{-15}	23.79	2.29
40×2000	5.70×10^{-12}	2.87×10^{-11}	42.24	4.45
80×1000	1.40×10^{-11}	5.06×10^{-10}	119.01	12.28
80×2000	—	7.55×10^{-9}	—	24.50
160×1000	—	4.37×10^{-7}	—	85.67
160×2000	—	1.75×10^{-8}	—	205.04

Table 6: General case: Error values and computational time as compared with results obtained with GMRES.

7 Conclusion

We have presented a compact representation, computable in a numerically stable way, for the inverse of block-tridiagonal and banded matrices. We have also demonstrated the advantages in computation with this new representation. The stability of algorithms based on the new representation appears to continue for very large problem sizes, making it practical to directly exploit the structure of block-tridiagonal matrices for large-scale modeling and simulation problems. A significant advantage with the compact representation presented herein is that the computation of the ratio sequences can be parallelized. This offers the possibility of parallel algorithms for computation with semiseparable matrices, which are explored elsewhere [7].

References

- [1] <http://www.cs.kuleuven.ac.be/~raf/homepage/software/phd/BRSS.m>.
- [2] E. Asplund. Inverses of matrices a_{ij} which satisfy $a_{ij} = 0$ for $j > i + p$. *Mathematica Scandinavica*, 7:57–60, 1959.
- [3] S.O. Asplund. Finite boundary value problems solved by green’s matrix. *Mathematica Scandinavica*, 7:49–56, 1959.
- [4] R. Bevilacqua, B. Codenotti, and F. Romani. Parallel solution of block tridiagonal linear systems. *Linear Algebra Appl.*, 104:39–57, 1988.
- [5] K. Bowden. A direct solution to the block tridiagonal matrix inversion problem. *International Journal of General Systems*, 15:185–198, 1989.
- [6] B. Bukhberger and G. A. Emel’yanenko. Methods of inverting tridiagonal matrices. *Computational Mathematics and Mathematical Physics*, 13:10–20, 1973.

- [7] S. Cauley, J. Jain, C.-K. Koh, and V. Balakrishnan. Parallelization of quantum-scale device simulation, submitted for consideration, Gordon Bell prize. In *SuperComputing*, 2006.
- [8] S. Chandrasekaran and M. Gu. Fast and stable eigen decomposition of symmetric banded plus semi-separable matrices. *Linear Algebra Appl.*, 313:107–114, 2000.
- [9] S. Chandrasekaran and M. Gu. A fast and stable solver for recursively semiseparable system of linear equations. *Contemporary Mathematics*, 281:39–53, 2001.
- [10] S. Chandrasekaran and M. Gu. A divide and conquer algorithm for eigen decomposition of symmetric block diagonal plus semi-separable matrices. *Numerische Mathematik*, 96(4):723–731, 2004.
- [11] S. Chandrasekaran, Dewilde P., M. Gu, and A-J. Pals, T.and Van der Veen. Fast stable solver for sequentially semi-separable linear systems of equations. *Lecture Notes in Computer Science*, 2552:545–554, 2002.
- [12] P. Concus and Meurant G. On computing INV block preconditionings for the conjugate gradient method. *BIT*, 26:493–504, 1986.
- [13] P. Concus, G. H. Golub, and Meurant G. Block preconditionings for the conjugate gradient method. *SIAM Journal on Scientific Computing*, 6:220–252, 1985.
- [14] S. Demko, W. F. Moss, and Smith P. W. Decay rates for inverses of band matrices. *Mathematics of Computation*, 43:491–499, 1984.
- [15] Y. Eidelman and I. Gohberg. Inversion formulas and linear complexity algorithm for diagonal plus semiseparable matrices. *Computer and Mathematics with Applications*, 33(4):69–79, 1996.
- [16] E. M. Godfrin. A method to compute the inverse of an n-block tridiagonal quasi-hermitian matrix. *Journal of Physics:Condensed Matter*, 3:7843–7848, 1991.
- [17] R. A. Gonzales, J. Eisert, I. Koltracht, M. Neumann, and Rawitscher G. Integral equation method for the continuous spectrum radial Schrodinger equation. *Journal of Computational Physics*, 134:134–139, 1997.
- [18] L. Greengard and Rokhlin V. On the numerical solution of two-point boundary value problems. *Communications on Pure and Applied Mathematics*, 44:419–452, 1991.
- [19] N. Mastronardi, S. Chandrasekaran, and S. V. Huffel. Fast and stable two-way algorithm for diagonal plus semi-separable system of linear equations. *Numerical Linear Algebra Appl.*, 8(1):7–12, 2001.
- [20] G. Meurant. A review on the inverse of symmetric block tridiagonal and block tridiagonal matrices. *SIAM Journal on Matrix Analysis and Applications*, 13(3):707–728, 1992.

- [21] R. Nabben. Decay rates of the inverses of nonsymmetric tridiagonal and band matrices. *SIAM Journal on Matrix Analysis and Applications*, 20(3):820–837, 1999.
- [22] T. Oohashi. Some representation for inverses of band matrices. *TRU Mathematics*, 14:39–47, 1978.
- [23] F. Romani. On the additive structure of inverses of banded matrices. *Linear Algebra Appl.*, 80:131–140, 1986.
- [24] P. Rozsa. Band matrices and semi-separable matrices. *Colloquia Mathematica Societatis Janos Bolyai*, 50:229–237, 1986.
- [25] P. Rozsa. On the inverse of band matrices. *Integral Equations and Operator Theory*, 50:82–95, 1987.
- [26] P. Rozsa, R. Bevilacqua, P. Favati, and F. Romani. On the inverse of block tridiagonal matrices with applications to the inverses of band matrices and block band matrices. *Operator Theory: Advances and Applications*, 40:447–469, 1989.
- [27] P. Rozsa, R. Bevilacqua, F. Romani, and P. Favati. On band matrices and their inverses. *Linear Algebra Appl.*, 150:287–295, 1991.
- [28] Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS Publishing, 1996.
- [29] Y. Saad and M. Schultz. GMRES: A generalized minimal residual algorithm for solving non-symmetric linear systems. *SIAM Journal on Scientific Computing*, 7:856–869, 1986.
- [30] A. Svizhenko, M. P. Anantram, T. R. Govindan, B. Biegel, and R. Venugopal. Two-dimensional quantum mechanical modeling of nanotransistors. *Journal of Applied Physics*, 91(4):2343–2354, 2002.
- [31] T. Torii. Inversion of tridiagonal matrices and the stability of tridiagonal systems of linear equations. *Technology Reports of the Osaka University*, 16:403–414, 1966.
- [32] R. Vandebril, M. Van Barel, and N. Mastronardi. A note on the representation and definition of semiseparable matrices. *Numerical Linear Algebra Appl.*, 12(8):839–858, 2005.