Available online at www.sciencedirect.com
SCIENCE DIRECT®
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
ELSEVIER Journal of Computational and Applied Mathematics 186 (2006) 416–431
www.elsevier.com/locate/cam

Diagonally-striped matrices and approximate inverse preconditioners☆

Dennis C. Smolarski*

Department of Mathematics and Computer Science, Santa Clara University, Santa Clara, CA 95053-0290, USA

Received 22 November 2004

Abstract

The inverse of a banded matrix is, in general, dense. If the structure of the original banded matrix is "striped", that is, the non-zero diagonals are separated by one or more zero diagonals, the inverse may exhibit a similar striped structure. The motivation for studying inverses of striped matrices is to obtain efficient preconditioners for systems arising from radiation transport equations, whose matrices include dominant values along diagonal stripes. Linear systems whose system matrix has a striped inverse lend themselves to the use of a sparse approximate inverse (SPAI) preconditioner whose structure is derived from that of the actual inverse.

MSC: 15A09; 65F10; 65F50

Keywords: Banded matrices; Inverses; Preconditioners; Sparse approximate inverses

1. Introduction

1.1. Sparsity patterns in matrices

This paper discusses matrices similar in sparsity patterns and dominant values to those resulting from radiation transport equations used, for example, to model collapsing neutron stars (see [21]). These

☆ This work is supported by the Department of Energy SciDAC cooperative agreement DE-FC02-01ER41185, http://www.phy.ornl.gov/tsi/.
* Tel.: +1 408 554 4124; fax: +1 40 8554 4795.
E-mail address: dsmolarski@math.scu.edu.

0377-0427/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2005.02.012

transport equations, representing different spatial or radial dimensions, are coupled together to produce a matrix with dense diagonal blocks and linking diagonals. Fig. 1 shows the sparsity pattern of a sample matrix modeling phenomena in one dimension. The two-dimensional model includes multiple copies of one-dimensional matrices forming the diagonal blocks of a larger matrix, coupled by two additional, outlying diagonals, resulting in four diagonals augmenting the values of the diagonal blocks.

The sparsity pattern does not tell the entire story, however. Although the main diagonal blocks are dense, in some cases a block is dominated by elements on its main diagonal, as can be seen by looking at the three-dimensional display, given in Fig. 2, of values of a portion of a test matrix from [21] (the height or depth indicates the values of the element).

This type of matrix can be understood as an overlay of a matrix consisting of non-zero diagonal stripes (i.e., non-zero diagonals separated by one or more zero diagonals) upon a block diagonal matrix. The differences in magnitudes between the elements on the diagonals and the non-diagonal elements in the center blocks result in a striped structure in the inverse. The actual inverse, seen in Fig. 3, though dense, has dominant values along diagonal stripes at regular intervals whose width matches the distance between the main diagonal and the diagonal coupling the center blocks in the original matrix.

Matrices resulting from transport equations are examples of *banded* (*square*) *matrices*, that is, matrices in which all non-zero elements are no more than k columns from the main diagonal, where k is the least such integer. In this case, the matrix is said to have a *bandwidth* of $2k + 1$. A typical example is a tridiagonal matrix in which the bandwidth is 3 and k, therefore, is 1. Matrices with a small bandwidth are also *sparse*, that is, having many more zero elements than non-zeroes. (In contrast, a matrix in which essentially every element is non-zero is said to be *dense*.)

It is well-known that, in general, the inverse, if it exists, of a tridiagonal matrix is dense. If, however, the three non-zero diagonals are not adjacent, the inverse is not necessarily dense and may, in fact, exhibit a diagonally striped structure reflective of the original matrix. (In general, we allow diagonals other than the

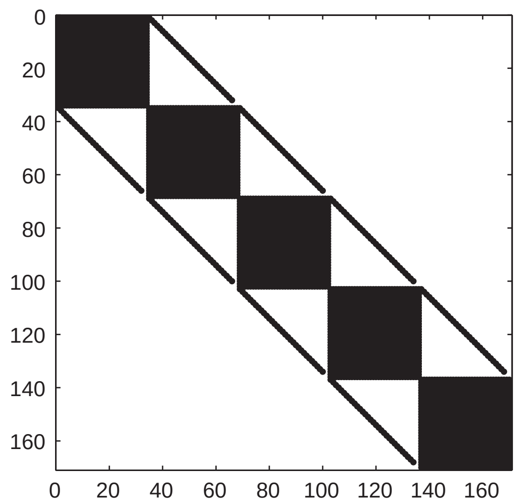

Fig. 1. The sparsity pattern of matrices derived from radiation transport equations. The dense diagonal blocks represent coupling between the various energy groups at the same spatial location. The two outlying diagonals denote coupling between neighboring spatial locations.

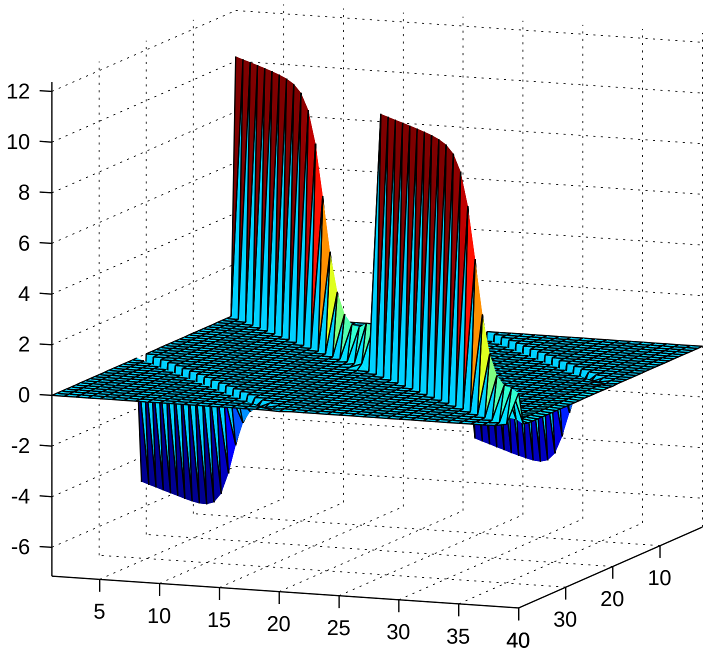

Fig. 2. A three-dimensional display of two diagonal blocks of a radiation transport matrix, where the height indicates the value of an element.

main diagonal to include zero as a value for some, but not all, elements.) In this paper, *structure* informally refers to the pattern of non-zero elements, in particular, the locations of non-adjacent diagonals containing non-zero elements. (Gilbert uses a formal definition of the "structure" of a matrix as a directed graph [9, p. 63]. Also see [7, p. 1806].) For convenience, in this paper, when the inverse of a matrix is referred to, it is assumed that the inverse exists.

Others have examined the inverses of banded matrices, proposing, in some cases, formulas for the elements of the inverse (see, for example, [1,13] and, for Toeplitz matrices, [4,18]). This paper will examine a specific type of banded matrix and, instead of proposing an algorithm to compute its inverse, will focus on the structure of the inverse and its relationship to the structure of the original matrix.

1.2. Inverses and preconditioners

Our motivation for determining the structure of the inverse of a banded matrix, A, is to obtain a preconditioner for use in solving a system of linear equations whose system matrix is A, in particular for a matrix arising from the important application of radiation transport. (A preconditioner, M^{-1}, is a matrix used to transform a linear system $Ax = b$ into the preconditioned system $M^{-1}Ax = M^{-1}b$. The ideal preconditioner is one that well approximates A^{-1} yet is relatively inexpensive to compute. See [11].)

Moreover, the particular type of preconditioner we desire is a *sparse approximate inverse* (SPAI) preconditioner (see, for example, [6,7,10,12]), whose structure is based on that of the exact inverse. The observed relationship between the diagonal stripes in the actual inverse and in the original matrix as seen

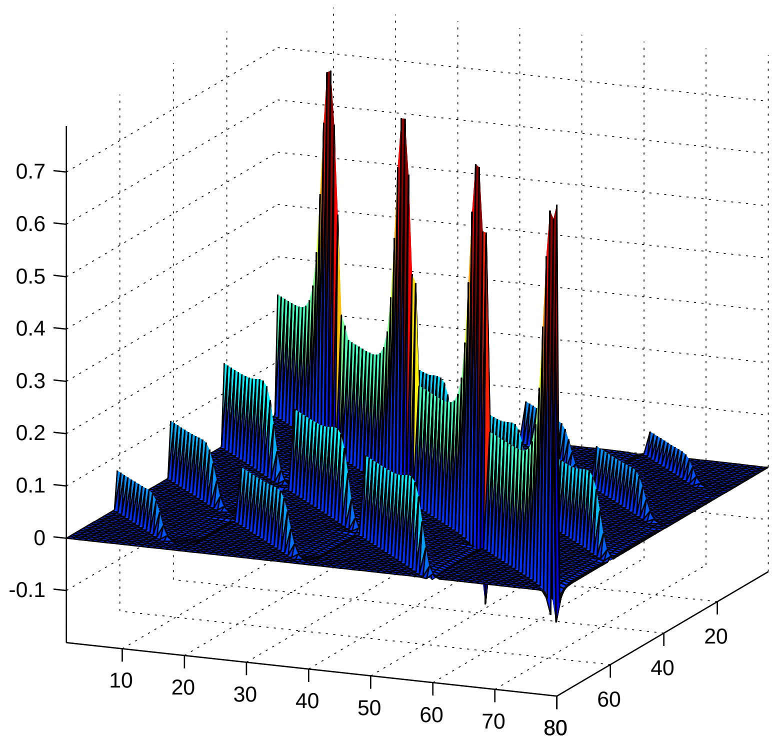

Fig. 3. A three-dimensional display of the inverse of four diagonal blocks of a radiation transport matrix, where the height indicates the value of an element.

in Figs. 2 and 3 gave additional motivation to study the striping phenomenon in an attempt to construct appropriate SPAI matrices with only a specific number of diagonals.

Other approaches exist for analyzing matrix structure and finding the inverse, for example, those by Gilbert [9] or Chow [7]. Gilbert used a graph theoretic approach to analyze the matrix structure of A and then showed that the structure of A^{-1} may be characterized by the transitive closure of the graph describing the structure of A (see [9, pp. 65, 72]). Chow used a similar approach (see [7, pp. 1806, 1808]) adding a "sparsifying" step for A (see [7, p. 1807]) to obtain the SPAI in which elements small in magnitude are dropped before examining the structure. Also see [14,15]. This paper takes a visual approach to showing the structure of the inverses of matrices that exhibit a striped structure, without requiring graph theoretic concepts. Knowing the structure of the exact inverse, an SPAI preconditioner can be constructed by a judicious choice of the locations of the diagonals based on the locations of the non-zero elements in the exact inverse.

SPAI preconditioners have a number of advantages in the iterative solution of linear systems, including their implementation as matrix-vector products and their sparsity. Moreover, computing the parameters of an SPAI preconditioner independently and, therefore, in parallel, takes greater advantage of the architecture of parallel platforms. See, for example, [6,10,12]. One result is that the computation of a preconditioner is a secondary effort and is a low-level contribution to the total work. This further motivates investigating the structure of the inverse of the matrices described in this paper and the relation

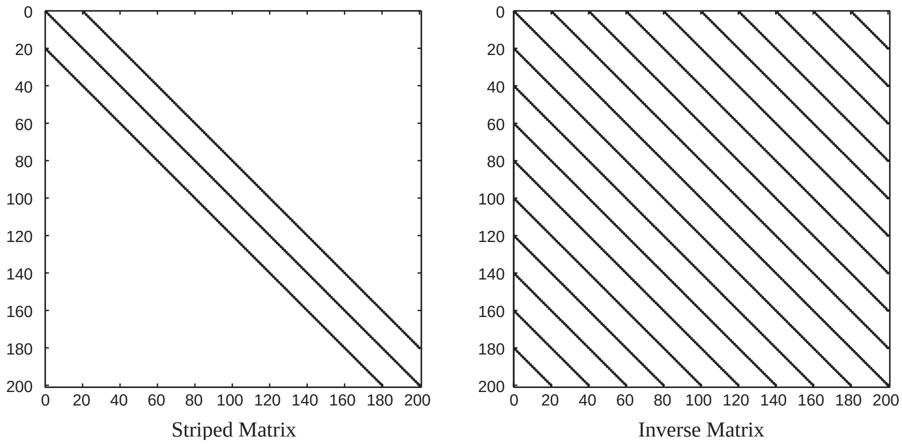

Fig. 4. Non-zero structure of a 200 × 200 striped matrix and its inverse.

of these inverses to SPAI preconditioners. The theorems included below link the structure of a matrix inverse with that of the matrix itself. We see this paper as contributing to the discussion of methods of constructing effective SPAI preconditioners and provides the background for the design of the several successful preconditioners used to solve linear systems derived from radiation transport equations as noted above (see [21]). In addition, one strong advantage of the approach described here is its simplicity.

1.3. Diagonally striped matrices

We define a *diagonally striped matrix* as one consisting of diagonals of non-zero (and possibly some zero) elements such that the adjacent diagonals of at least one non-zero diagonal are zero. (See Melhem who does not require the stripes to be parallel to the main diagonal, [19,20].) One example of a diagonally striped matrix would be a matrix with exactly three non-zero diagonals, one of which is the main diagonal and the other two are equally spaced on either side, though not adjacent to the main diagonal. We will call a matrix of this type a (*structurally symmetric*) 3-*diagonal* (*diagonally*) *striped matrix*.

Fig. 4 displays an example of a striped matrix of size 200 × 200 and of its inverse. The two outer, non-zero diagonals are separated from the main diagonal in the original matrix by 20 columns.

In row i of a 3-diagonal striped matrix, only elements $a_{i,i-m}$, a_{ii} and $a_{i,i+m}$ are allowed to be non-zero where m is the spacing distance between the diagonals. As is also seen is Fig. 4, the inverse $B = A^{-1}$ is not dense, but instead exhibits a remarkable striped property consisting of multiple non-zero diagonals, each separated by several zero diagonals. In other words, in row i of the inverse, the non-zero elements are b_{ii} and $b_{i,i\pm mk}$ where k is an integer so that $i \pm mk$ is a legitimate column subscript value. It is this striped property that guides us to an SPAI preconditioner.

1.4. Organization

This paper is organized as follows. Section 2 examines structurally symmetric 3-diagonal striped matrices, that is, a banded matrix where there are three non-adjacent, non-zero diagonals and Section 3 examines 5-diagonal striped matrices. Then, Section 4 deals with striped matrices with dense non-zero central bands which is the type of matrix that is associated with radiation transport equations and that has motivated this study. Section 5 deals with structurally non-symmetric striped matrices and Section 6 comments on the standard approach taken to find an inverse when systems can, in fact, be decoupled. Finally, Section 7 reiterates the link between striped matrices and SPAI preconditioners and offers some concluding remarks.

2. Structurally symmetric 3-diagonal striped matrices

As Fig. 4 above demonstrates, the inverse of a (structurally symmetric) 3-diagonal striped matrix, rather than being dense, is also a striped matrix and this striped structure will, in general, be found in inverses of other striped matrices. The inverse will contain the maximum number of diagonals possible for the number of columns of the matrix. (The "maximum number of diagonals possible" is approximately the number of columns of the matrix divided by the number of columns between the main diagonal and the nearest non-zero diagonal, and we omit a longer, more precise definition trusting that the reader will understand this intuitive concept.) In Fig. 4, each diagonal in the inverse is spaced 20 columns away from adjacent diagonals until no more diagonals are possible in the upper right and lower left corners.

As another, smaller, numerical example, let A be the following 10×10 3-diagonal striped matrix

$$\begin{pmatrix}
2 & & 1 & & & & & & & \\
 & 2 & & 1 & & & & & & \\
1 & & 2 & & 1 & & & & & \\
 & 1 & & 2 & & 1 & & & & \\
 & & 1 & & 2 & & 1 & & & \\
 & & & 1 & & 2 & & 1 & & \\
 & & & & 1 & & 2 & & 1 & \\
 & & & & & 1 & & 2 & & 1 \\
 & & & & & & 1 & & 2 & \\
 & & & & & & & 1 & & 2
\end{pmatrix}.$$

Its inverse, A^{-1}, is the diagonally striped matrix

$$\frac{1}{6}\begin{pmatrix}
5 & & -4 & & 3 & & -2 & & 1 & \\
 & 5 & & -4 & & 3 & & -2 & & 1 \\
-4 & & 8 & & -6 & & 4 & & -2 & \\
 & -4 & & 8 & & -6 & & 4 & & -2 \\
3 & & -6 & & 9 & & -6 & & 3 & \\
 & 3 & & -6 & & 9 & & -6 & & 3 \\
-2 & & 4 & & -6 & & 8 & & -4 & \\
 & -2 & & 4 & & -6 & & 8 & & -4 \\
1 & & -2 & & 3 & & -4 & & 5 & \\
 & 1 & & -2 & & 3 & & -4 & & 5
\end{pmatrix}.$$

Note that the dominance of values along the main diagonal is evident in both the original matrix and its inverse, a phenomenon seen above in Figs. 2 and 3. This decay in values motivates the use of only a limited number of central diagonals when we approximate the exact inverse of a striped matrix with an SPAI matrix to use as a preconditioner (as was reported in [21]).

More generally, we have the following theorem.

Theorem 1. *If A is a 3-diagonal striped matrix with the outer two diagonals spaced m columns away from the main diagonal (with $m>1$), the inverse A^{-1} will be a diagonally striped matrix with a maximum number of diagonals per row also spaced at m columns away from each other.*

Proof. Let $A=(a_{ij})$ and let a_{ij} equal zero except for a_{ii} and $a_{i,i\pm m}$ (some of which may also be zero) for $m\neq 1$. Let $B=A^{-1}$ and $B=(b_{ij})$.

Since $A\cdot A^{-1}=A\cdot B=I$, the following relation holds:

$$a_{i,i-m}b_{i-m,j}+a_{ii}b_{ij}+a_{i,i+m}b_{i+m,j}=\begin{cases}1 & \text{if } i=j,\\ 0 & \text{if } i\neq j.\end{cases} \tag{1}$$

(By assumption $a_{ij}=0$ if $j\neq i$ or $j\neq i\pm m$.)

Substituting $i+1$ for j in (1) results in $a_{i,i-m}b_{i-m,i+1}+a_{ii}b_{i,i+1}+a_{i,i+m}b_{i+m,i+1}=0$, which, in turn, implies that $b_{i-m,i+1}$, $b_{i,i+1}$, and $b_{i+m,i+1}$ are all zero for every i. A similar argument can be made for every value of j not equal to $i\pm km$ for an appropriate integer value of k. Thus we can assume that all values of elements of B along diagonals not spaced at multiples of m away from the main diagonal are zero.

To show that the diagonals of B spaced at multiples of m away from the main diagonal are, in general, non-zero, we argue as follows.

If $i=j$, then $a_{i,i-m}b_{i-m,j}+a_{ii}b_{ij}+a_{i,i+m}b_{i+m,j}=a_{i,i-m}b_{i-m,i}+a_{ii}b_{ii}+a_{i,i+m}b_{i+m,i}=1$ and, thus, at least one of $b_{i-m,i}$, b_{ii}, and $b_{i+m,i}$ must be non-zero. Substituting $i-m$ for i and i for j in (1), we have

$$a_{i-m,i-2m}b_{i-2m,i}+a_{i-m,i-m}b_{i-m,i}+a_{i-m,i}b_{ii}=0. \tag{2}$$

If $a_{i-m,i-2m}$ is non-zero, then it follows that, if $b_{i-2m,i}$ is non-zero, it equals

$$-\frac{a_{i-m,i-m}b_{i-m,i}+a_{i-m,i}b_{ii}}{a_{i-m,i-2m}}.$$

If $a_{i-m,i-2m}$ equals zero, then $b_{i-2m,i}$ could, in fact, be any value and (2) would be satisfied. The uniqueness of the inverse means that $b_{i-2m,i}$ is, in fact, zero.

This argument can be repeated to derive a value for $b_{i+2m,i}$ (assuming it is also non-zero). The same argument can be repeated to obtain values for $b_{i\pm km,i}$ (assuming they are non-zero) for any integer k such that $i\pm km$ is a legitimate row value.

Thus, B consists of (possible) non-zero values along diagonals spaced every m columns away from the main diagonal. □

We note that matrix A could be block 3-diagonally striped and corresponding results would hold.

Corollary 2. *If A is an $n\times n$ 3-diagonal striped matrix with the outer two diagonals spaced $m>n/2$ columns away from the main diagonal, the sparsity patterns for A and A^{-1} coincide.*

Proof. Since Theorem 1 concludes that A^{-1} has the maximum number of diagonals per row spaced at m columns apart, given that $m > n/2$, it is only possible to have two non-zero diagonals in the inverse apart from the main diagonal. These diagonals are spaced m columns away from the main diagonal as in the original matrix. □

3. Structurally symmetric 5-diagonal striped matrices

As noted earlier, a matrix representing a two-dimensional model employing radiation transport equations includes diagonals coupling multiple copies of the pattern displayed in Fig. 1. Thus, the dominant values in this model produce a 5-diagonal striped matrix. The inverse of this matrix is not always striped.

Suppose A is a 5-diagonal striped matrix with non-zero diagonals at a_{ii}, $a_{i,i\pm m_1}$ and $a_{i,i\pm m_2}$ with $m_1 = 3$ and $m_2 = 5$, as in the following 10×10 matrix:

$$\begin{pmatrix} 2 & & & 1 & & 1 & & & & \\ & 2 & & & 1 & & 1 & & & \\ & & 2 & & & 1 & & 1 & & \\ 1 & & & 2 & & & 1 & & 1 & \\ & 1 & & & 2 & & & 1 & & 1 \\ 1 & & 1 & & & 2 & & & 1 & \\ & 1 & & 1 & & & 2 & & & 1 \\ & & 1 & & 1 & & & 2 & & \\ & & & 1 & & 1 & & & 2 & \\ & & & & 1 & & 1 & & & 2 \end{pmatrix}.$$

In this case, the inverse is, in fact, a *dense* matrix, rather than a diagonally striped matrix.

In contrast, if A were a 5-diagonal striped matrix with non-zero diagonals at a_{ii}, $a_{i,i\pm m_1}$ and $a_{i,i\pm m_2}$ with $m_1 = 4$ and $m_2 = 6$, its inverse $B = A^{-1}$ would also be a diagonally striped matrix with non-zero diagonals at b_{ii} and $b_{i,i\pm 2k}$ where, as above, k is are arbitrary integer value so that $i \pm 2k$ is a legitimate column subscript value. Fig. 5 depicts a larger version of this latter matrix, where A is 200×200, and $m_1 = 40$ and $m_2 = 60$. As Fig. 5 shows, the multiple diagonals of the inverse are spaced 20 columns apart.

These two examples motivate the following theorem.

Theorem 3. *Let A be an $n \times n$ 5-diagonal striped matrix with non-zero diagonals at a_{ii}, $a_{i,i\pm m_1}$, and $a_{i,i\pm m_2}$ with $0 < m_1 < m_2$. Assume $m_1 \leqslant n/2$. If $\gcd(m_1, m_2) = 1$, then A^{-1} is dense. If, however, $\gcd(m_1, m_2) = m > 1$, then $B = A^{-1}$ will be a diagonally striped matrix with non-zero diagonals at b_{ii} and $b_{i,i\pm mk}$ where, as above, k is an arbitrary integer value so that $i \pm mk$ is a legitimate column subscript value.*

Comment: If $m_1 > n/2$, then A can be partitioned into blocks such that the upper left and lower right blocks are simple diagonal matrices. In this case, the inverse will consist of four striped diagonal sub-blocks separated by spacing twice the distance between m_1 and $n/2$. In addition, the center part of the matrix will continue a diagonal from the diagonal of the upper left block to the diagonal of the lower right block. See Figs. 6 and 7 as examples.

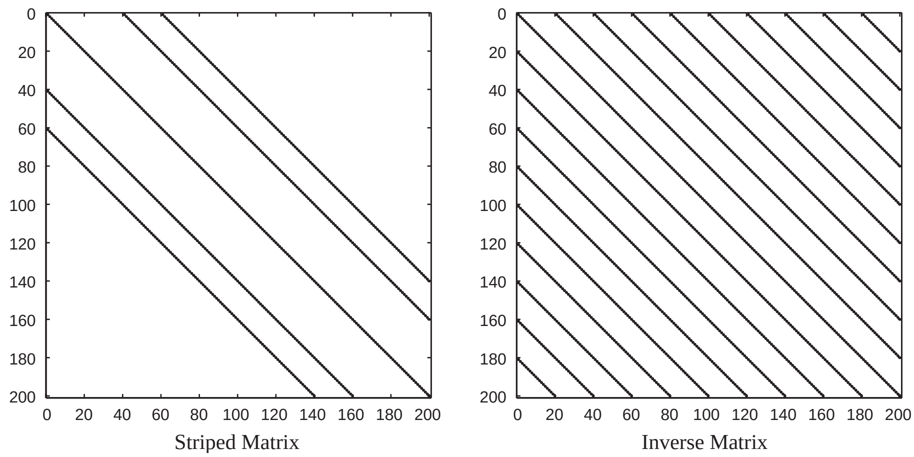

Fig. 5. Non-zero structure of a 200 × 200 striped matrix and its inverse.

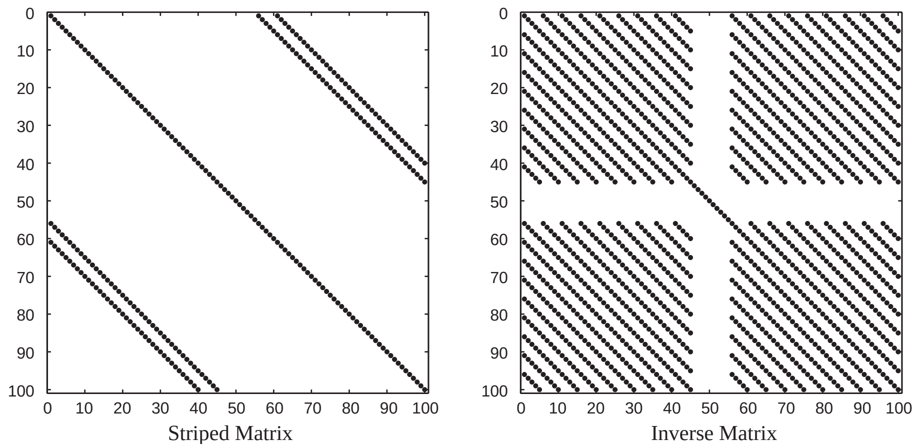

Fig. 6. Non-zero structure of a 100 × 100 striped matrix and its inverse where $m_1 = 55$ and $m_2 = 60$.

Proof. This proof follows the argument given above for Theorem 1, with adaptations for two different values for column spacing.

Let $A = (a_{ij})$ and assume a_{ij} are zero except for a_{ii}, $a_{i,i\pm m_1}$, and $a_{i,i\pm m_2}$ for $m_1, m_2 \neq 1$. Let $B = A^{-1}$ and $B = (b_{ij})$.

Since $A \cdot A^{-1} = A \cdot B = I$, the following relation holds:

$$\begin{aligned} &a_{i,i-m_2}b_{i-m_2,j} + a_{i,i-m_1}b_{i-m_1,j} + a_{ii}b_{ij} \\ &\quad + a_{i,i+m_1}b_{i+m_1,j} + a_{i,i+m_2}b_{i+m_2,j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \end{aligned} \tag{3}$$

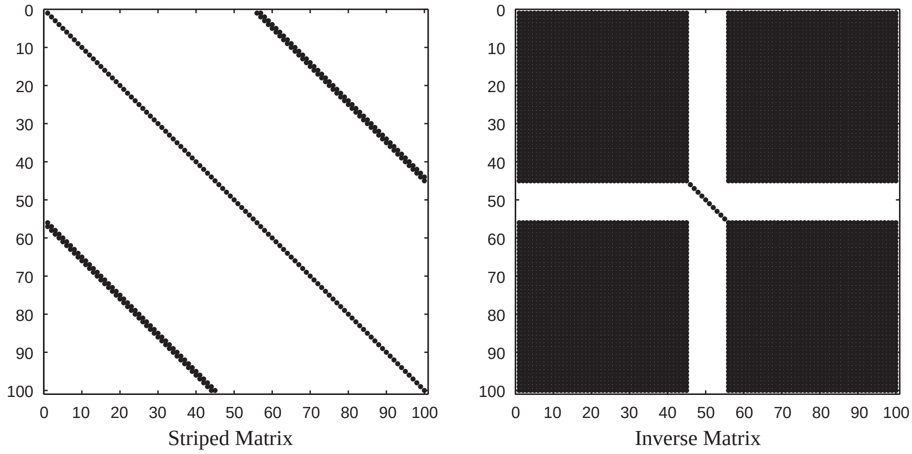

Fig. 7. Non-zero structure of a 100 × 100 striped matrix and its inverse where $m_1 = 55$ and $m_2 = 56$.

(As in Theorem 1, by assumption $a_{ij} = 0$ if $j \neq i$ or $j \neq i \pm m_1$ or $j \neq i \pm m_2$. Again, as in Theorem 1, we can conclude that all values of elements of B along diagonals other than those demanded by this theorem will be zero.)

To show that B is either dense or the diagonals of B spaced at multiples of $m = \gcd(m_1, m_2) > 1$ are, in general, non-zero, we argue as follows:

If $i = j$, then $a_{i,i-m_2}b_{i-m_2,j} + a_{i,i-m_1}b_{i-m_1,j} + a_{ii}b_{ij} + a_{i,i+m_1}b_{i+m_1,j} + a_{i,i+m_2}b_{i+m_2,j} = a_{i,i-m_2}b_{i-m_2,i} + a_{i,i-m_1}b_{i-m_1,i} + a_{ii}b_{ii} + a_{i,i+m_1}b_{i+m_1,i} + a_{i,i+m_2}b_{i+m_2,i} = 1$ and at least one of $b_{i-m_2,i}$, $b_{i-m_1,i}$, b_{ii}, $b_{i+m_1,i}$, $b_{i+m_2,i}$ must be non-zero. Substituting $i - m_1$ for i and i for j in (3), we have

$$a_{i-m_1,i-(m_1+m_2)}b_{i-(m_1+m_2),i} + a_{i-m_1,i-2m_1}b_{i-2m_1,i} + a_{i-m_1,i-m_1}b_{i-m_1,i} + a_{i-m_1,i}b_{ii} + a_{i-m_1,i-(m_1-m_2)}b_{i-(m_1-m_2),i} = 0. \tag{4}$$

If $a_{i-m_1,i-(m_1+m_2)}$ is non-zero, it follows that, if $b_{i-(m_1+m_2),i}$ is also non-zero, it equals

$$-\frac{a_{i-m_1,i-2m_1}b_{i-2m_1,i} + a_{i-m_1,i-m_1}b_{i-m_1,i} + a_{i-m_1,i}b_{ii} + a_{i-m_1,i-(m_1-m_2)}b_{i-(m_1-m_2),i}}{a_{i-m_1,i-(m_1+m_2)}}.$$

If $a_{i-m_1,i-(m_1+m_2)}$ equals zero, then $b_{i-(m_1+m_2),i}$ is also zero, as in Theorem 1.

Also as with Theorem 1, this argument can be repeated to show that $b_{i-(2m_1+m_2),i}$ and $b_{i-(m_1+2m_2),i}$ are also non-zero and then that $b_{i+k_1m_1+k_2m_2,i}$ is non-zero for any integers k_1 and k_2 such that $i + k_1m_1 + k_2m_2$ is a legitimate row value.

Since $\gcd(m_1, m_2) = m$ implies that there exist integers k_1 and k_2 such that $m = k_1m_1 + k_2m_2$, if $m = 1$ then the subdiagonal and superdiagonal of the inverse matrix have non-zero values. By taking appropriate multiples of k_1 and k_2, one can show that every diagonal has non-zero values. If $m > 1$, the inverse matrix will be striped with stripes at distances $\pm m$ from the main diagonal.

Thus, B consists of (possible) non-zero values along diagonals spaced every m columns and rows away from the main diagonal if $\gcd(m_1, m_2) = m > 1$ and dense if $\gcd(m_1, m_2) = 1$. □

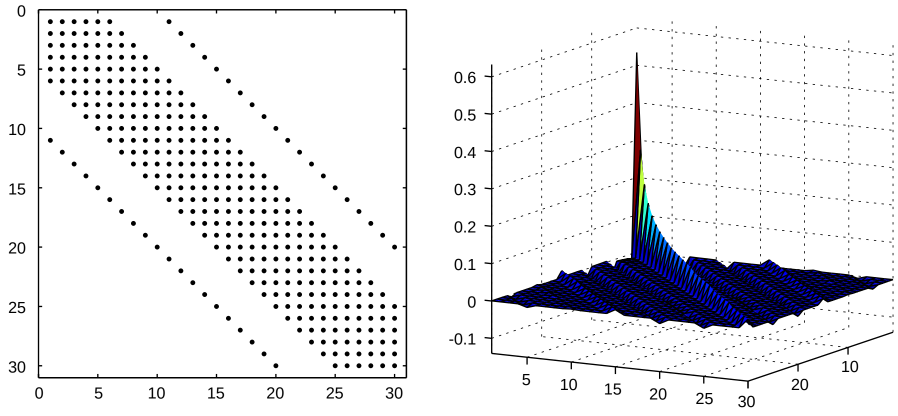

Fig. 8. Non-zero structure of a 30×30 striped matrix with dense center band and its inverse. For this matrix, $m_1 = 5$ and $m_2 = 10$.

Numerical experiments suggest that this theorem also generalizes to diagonally striped matrices with a higher number of diagonals.

4. Striped matrices with dense non-zero bands

The striping phenomenon can also exist in an inverse that is dense as was seen in the introduction in the example given in Figs. 1–3 of a matrix derived from radiation transport equations (and its inverse). This phenomenon will occur, for example, in a 5-diagonal striped matrix in which values between the two inner, off-main diagonals are allowed to be non-zero, but small in magnitude relative to the magnitude of the values on the other diagonals. As another example, the matrix and inverse depicted in Fig. 8 is 30×30, with $m_1 = 5$ and $m_2 = 10$. The values of the elements on the five primary diagonals range from 1 to 30, but the non-zero elements between the two diagonals at distance $m_1 = 5$ from the main diagonal are all 0.100. Thus the primary diagonals are of magnitude 10 to 100 times greater than other elements. This difference in magnitude is enough to retain the striped diagonal phenomenon in the inverse even though the inverse is, in fact, dense.

5. Structurally non-symmetric matrices

The matrices discussed in the previous sections were assumed to be structurally symmetric, i.e., to have diagonals spaced at equal distances on either side of the main diagonal, although no assumption about the values of those diagonals has been made. Matrices motivated by physical problems, for example those generated by radiation transport equations, are usually structurally symmetric.

Nevertheless, there are two types of structurally *non-symmetric* matrices that can also be examined. One type has diagonals on either side of the main diagonal but spaced differently, and the other has diagonals only on one side of the main diagonal (and is, thus, upper or lower triangular). Although a

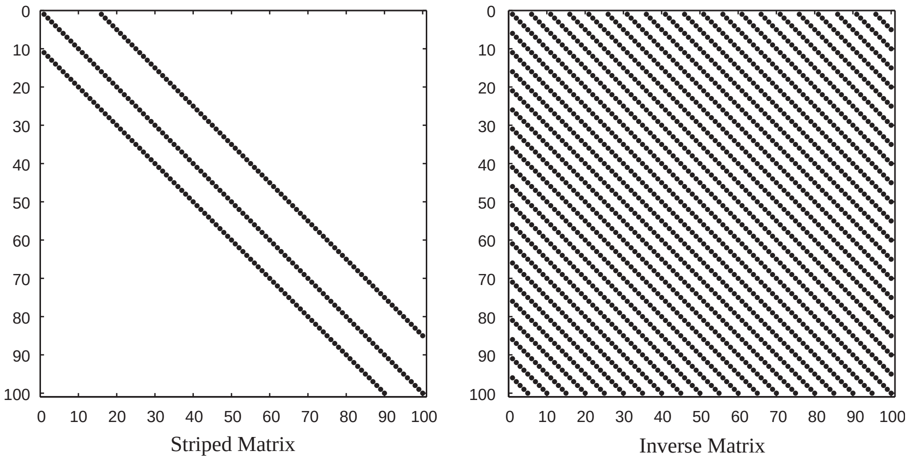

Fig. 9. Non-zero structure of a 100×100 striped matrix and its inverse where $m_1 = 10$ and $m_2 = 15$. Stripes in the inverse are separated by five columns.

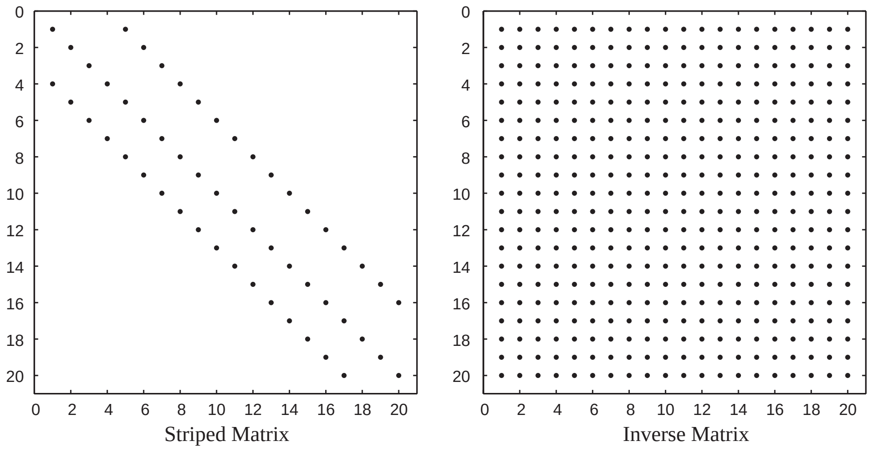

Fig. 10. Non-zero structure of a 20×20 striped matrix and its inverse where $m_1 = 3$ and $m_2 = 4$. Inverse is dense.

lower triangular system can be easily solved by backward reduction, it may not be practical to use this approach and some preconditioner may be needed. For this reason, we comment briefly on the structure of inverses of structurally non-symmetric striped matrices.

When A has non-zero diagonals on either side of the main diagonal but spaced differently, the results of Theorem 3 also hold with a slight modification of the statement as follows: if A is an $n \times n$ 3-diagonal striped matrix (rather than a 5-diagonal matrix) with non-zero diagonals at a_{ii}, $a_{i,i+m_1}$, and $a_{i,i-m_2}$, then the inverse is dense if $\gcd(m_1, m_2) = 1$ and a striped diagonal matrix otherwise. For examples, see Figs. 9 and 10.

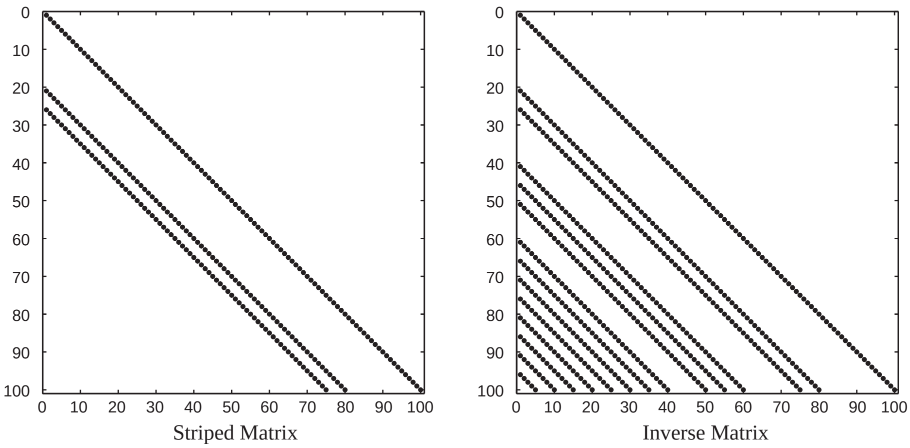

Fig. 11. Non-zero structure of a 100×100 striped matrix A and its inverse B where $m_1 = 20$ and $m_2 = 25$. Stripes in the inverse are separated by five columns. There are only zero diagonals in $B = A^{-1}$ between b_{ii} and $b_{i,i-20}$, between $b_{i,i-25}$ and $b_{i,i-2\times 20} = b_{i,i-40}$, and between $b_{i,i-2\times 25} = b_{i,i-50}$ and $b_{i,i-3\times 20} = b_{i,i-60}$.

The case when additional diagonals are on the same side of the main diagonal, however, is not a simple generalization of the structurally symmetric case. We will examine two subcases independently and state, without proof, conjectures concerning the structure of the inverses. The conjectures assume A is lower triangular; the upper triangular case is similar.

The first subcase occurs when the additional diagonals are at distances not relatively prime. In this case, the inverse will be sparse and striped. See Fig. 11 as an example. This case motivates the following conjecture.

Conjecture 4. *Let A be an $n \times n$ 3-diagonal (lower triangular) striped matrix with non-zero diagonals at a_{ii}, $a_{i,i-m_1}$, and $a_{i,i-m_2}$ and $m_1 < m_2$. If $\gcd(m_1, m_2) = m > 1$, then $B = A^{-1}$ will be a diagonally striped (lower triangular) matrix with non-zero diagonals at b_{ii} and $b_{i,i-mk}$ where k is a positive integer so that $i - mk$ is a legitimate column subscript value, with these exceptions: there are no non-zero diagonals between b_{ii} and $b_{i,i-m_1}$, between $b_{i,i-m_2}$ and $b_{i,i-2m_1}$, between $b_{i,i-2m_2}$ and $b_{i,i-3m_1}, \ldots,$ between $b_{i,i-pm_2}$ and $b_{i,i-(p+1)m_1}$ (for positive integer p), until $(p+1)m_1 - pm_2 = m$.*

The second subcase occurs when the additional diagonals are at distances that are relatively prime. In this case, the inverse will be generally dense, and most dense in the lower left portion (assuming A is a lower triangular matrix). But diagonal stripes may appear between that portion and the main diagonal. See Fig. 12 as an example. This case motivates the following conjecture.

Conjecture 5. *Let A be an $n \times n$ 3-diagonal (lower triangular) striped matrix with non-zero diagonals at a_{ii}, $a_{i,i-m_1}$, and $a_{i,i-m_2}$ and $m_1 < m_2$. If $\gcd(m_1, m_2) = 1$, then $B = A^{-1}$ will be a lower triangular dense matrix, with these exceptions: there are zero diagonals between b_{ii} and $b_{i,i-m_1}$ and between $b_{i,i-m_1}$ and $b_{i,i-m_2}$; in addition, there are zero diagonals spaced at $m_2 - m_1$ intervals (at irregular patterns) between $b_{i,i-m_2}$ and $b_{i,i-m_1 m_2}$.*

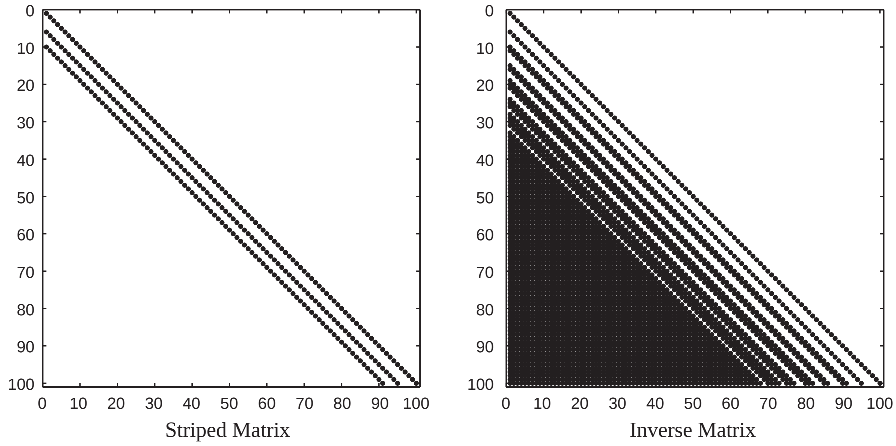

Fig. 12. Non-zero structure of a 100×100 striped matrix A and its inverse where $m_1 = 5$ and $m_2 = 9$. There are only zero diagonals in $B = A^{-1}$ between b_{ii} and $b_{i,i-20}$, between $b_{i,i-25}$ and $b_{i,i-2\times 20} = b_{i,i-40}$, and between $b_{i,i-2\times 25} = b_{i,i-50}$ and $b_{i,i-3\times 20} = b_{i,i-60}$.

6. Decoupled systems

We briefly mention decoupled matrices and how techniques employed for these matrices can also be used to obtain an SPAI preconditioner for a tridiagonal matrix.

Matrices considered in this paper consist of diagonals corresponding to the location of the *dominant* values, and assume the existence of non-zero values elsewhere (as there are in main diagonal blocks in matrices derived from radiation transport equations). Without such additional non-zero values, it may be possible to decouple the equations and rearrange the non-zeros into a matrix of smaller bandwidth.

For example, the following 6×6 3-diagonal striped matrix A (with values corresponding to the subscripts in the natural ordering) can be transformed into a tridiagonal matrix $\tilde{A}$.

$$A = \begin{pmatrix} 11 & & 13 & & & \\ & 22 & & 24 & & \\ 31 & & 33 & & 35 & \\ & 42 & & 44 & & 46 \\ & & 53 & & 55 & \\ & & & 64 & & 66 \end{pmatrix}, \quad \tilde{A} = \begin{pmatrix} 11 & 13 & & & & \\ 31 & 33 & 35 & & & \\ & 53 & 55 & & & \\ & & & 22 & 24 & \\ & & & 42 & 44 & 46 \\ & & & & 64 & 66 \end{pmatrix}.$$

(The rearranged matrix $\tilde{A}$ consists of the columns and rows of the original matrix with the column and row order being 1, 3, 5, 2, 4, and 6.)

$\tilde{A}$ is also block diagonal, resulting in a block diagonal inverse in which each block is dense. It is obvious that rearranging the order of rows and columns in the inverse of the rearranged matrix to correspond to the original ordering leads, in this case, to a 5 diagonal striped matrix. (We also note that whether an inverse is sparse is independent of the ordering that is used and, thus, independent of the striping.)

Where the values in a 3-diagonal striped matrix can be reordered to create a tridiagonal system matrix, various efficient techniques, including parallel preconditioners other than SPAI, exist to solve the new system. Nevertheless, the equivalence of A and $\tilde{A}$ suggests that it is possible to take a tridiagonal matrix, duplicate the values to obtain a larger matrix with the block pattern of $\tilde{A}$, reorder the values to obtain a 3-diagonal striped matrix, and then construct an SPAI striped preconditioner. In some situations, using a 3-diagonal striped preconditioner may lead to faster convergence than methods specifically designed for tridiagonal matrices.

7. Conclusion

This paper has focussed on a particular type of banded matrix, namely a diagonally striped matrix, and on the structure of its inverse. As shown in Theorems 1 and 3, the inverse of a striped matrix generally reflects the striped structure of the original matrix. The motivation was to obtain efficient preconditioners to solve linear systems arising from radiation transport equations, where the dominant values in the system matrix exhibit a striped structure.

When the actual inverse does exhibit a diagonally striped pattern, an SPAI preconditioner could be constructed, based on the structure of the actual inverse, but with a reduced number of diagonals, for example, having a 3-diagonal or 5-diagonal striped structure. This type of preconditioner would be very efficient to compute (since the elements in each row could be computed independently) and could take advantage of parallel platforms. In addition, the preconditioner may approximate the actual inverse (which may, in fact, be dense) well enough to accelerate convergence of an iterative method.

Experiments conducted in [21] validate diagonally striped SPAI preconditioners for matrices derived from radiation transport equations (also see comparisons and discussion in [2]). SPAI preconditioners with 3 and 5 diagonals were successfully tested on matrices derived from a one-dimensional case and preconditioners with 5 and 9 diagonals on matrices from a two-dimensional case.

Striped diagonal SPAI preconditioners have particular advantages when the system matrix A is not explicitly formed (as in some of our experiments) and other preconditioning approaches, such as an LU decomposition, are not feasible.

Finally, we note that Toeplitz matrices may seem to belong to the class of matrices studied in this paper and finding inverses of Toeplitz matrices has been an active area of research (e.g., [3,5,8,16–18,22,23]). We have not found, however, the Toeplitz structure to be of any advantage.

Acknowledgements

Thanks are due to Victor Eijkhout (University of Tennessee, Knoxville) and to a referee who noted that striped diagonal matrices (without center dense blocks) are decoupled systems. Additional thanks go to a referee who pointed out Corollary 2 and to other referees for valuable suggestions and references.

Special thanks go to Edmond F. D'Azevedo (Oak Ridge National Laboratory) for suggesting the following alternative proof to Theorem 1: It is easily seen that if A is a 3-diagonal striped matrix, then A^k is also a striped matrix with diagonals spaced m columns away from each other with the number of non-zero diagonals equalling $2k + 1$. From the Cayley-Hamilton theorem, since A satisfies its characteristic polynomial, we have $a_0 I + a_1 A + a_2 A^2 + \cdots + a_{n-1} A^{n-1} + a_n A^n = 0$. Rearranging terms, dividing by a_0 and multiplying by A^{-1}, we obtain an expression for A^{-1} which is a sum of powers

of A, all of which are diagonally striped matrices (with diagonals spaced *m* columns apart), namely, $A^{-1} = -a_1/a_0 I - a_2/a_0 A - \cdots - a_{n-1}/a_0 A^{n-2} - a_n/a_0 A^{n-1}$. Thus A^{-1} itself must be a diagonally striped matrix with diagonals spaced *m* columns apart, with the maximum number of diagonals possible.

References

[1] E.L. Allgower, Exact inverses of certain band matrices, Numer. Math. 21 (1973) 279–284.
[2] R. Balakrishnan, E.F. D'Azevedo, J.W. Fettig, B. Messer, A. Mezzacappa, F. Saied, P.E. Saylor, D.C. Smolarski, F.D. Swesty, On the performance of SPAI and ADI-like preconditioners for core collapse supernova simulations in one spatial dimension, Technical Report UIUCDCS-R-2003-2335, Department of Computer Science, University of Illlinois, Urbana-Champaign, 2003, J. Sci. Comput., 2004, Submitted for publication.
[3] A. Ben-Artzi, T. Shalom, On inversion of toeplitz and close to toeplitz matrices, Linear Algebra Appl. 75 (1986) 173–192.
[4] L. Berg, Lineare Gleichungssysteme mit Bandstruktur und ihr asymptotisches Verhalten, VEB Deutscher Verlag der Wissenschaften, Berlin, 1986.
[5] A. Calderón, F. Spitzer, H. Widom, Inversion of toeplitz matrices, Illinois J. Math. 3 (1959) 490–498.
[6] E.T.-F. Chow, Robust Preconditioning For Sparse Linear Systems, Ph.D. Thesis, The University of Minnesota, 1997.
[7] E.T.-F. Chow, A priori sparsity patterns for parallel sparse approximate inverse preconditioners, SIAM J. Sci. Comput. 21 (2000) 1804–1822.
[8] B. Friedlander, M. Morf, T. Kailath, L. Ljung, New inversion formulas for matrices classified in terms of their distance from toeplitz matrices, Linear Algebra Appl. 27 (1979) 31–60.
[9] J.R. Gilbert, Predicting structure in sparse matrix computations, SIAM J. Matrix Anal. Appl. 15 (1994) 62–79.
[10] N.I.M. Gould, J.A. Scott, On approximate-inverse preconditioners, Technical Report RAL 95-026, Computing and Information Systems Department, Rutherford Appleton Laboratory, Oxfordshire, England, 1995.
[11] A. Greenbaum, Iterative Methods for Solving Linear Systems, SIAM, Philadelphia, 1997.
[12] M.J. Grote, T. Huckle, Parallel preconditioning with sparse approximate inverses, SIAM J. Sci. Comput. 18 (1997) 838–853.
[13] G.Y. Hu, R.F. O'Connell, Analytical inversion of symmetric tridiagonal matrices, J. Phys. A 29 (1996) 1511–1513.
[14] I.E. Kaporin, A preconditioned conjugate gradient method for solving discrete analogs of differential problems, Differential Equations 26 (1990) 897–906.
[15] L.Yu. Kolotilina, Explicit preconditioning of systems of linear algebraic equations with dense matrices, J. Soviet Math. 43 (1988) 2566–2573.
[16] G. Labahn, T. Shalom, Inversion of toeplitz matrices with only two standard equations, Linear Algebra Appl. 175 (1992) 143–158.
[17] G. Labahn, T. Shalom, Inversion of toeplitz structured matrices using only standard equations, Linear Algebra Appl. 207 (1994) 49–70.
[18] D.A. Lavis, B.W. Southern, The inverse of a symmetric banded toeplitz matrix, Rep. Math. Phys. 39 (1997) 137–146.
[19] R. Melhem, Determination of stripe structures for finite element matrices, SIAM J. Numer. Anal. 24 (1987) 1419–1433.
[20] R. Melham, Parallel solution of linear systems with striped sparse matrices, Parallel Comput. 6 (1988) 165–184.
[21] F.D. Swesty, D.C. Smolarski, P.E. Saylor, A comparison of algorithms for the efficient solution of the linear systems arising from multi-group flux-limited diffusion problems, Astrophys. J. Suppl. Ser. 153 (2004) 369–387.
[22] W.F. Trench, An algorithm for the inversion of finite toeplitz matrices, J. Soc. Indust. Appl. Math. 12 (1964) 515–522.
[23] H. Widom, Inversion of toeplitz matrices II, Illinois J. Math. 4 (1960) 88–99.