Preface

Many important problems in applied science and engineering, such as the Navier–Stokes equations in fluid dynamics, the primitive equations in global climate modeling, the strain-stress equations in mechanical and material engineering, and the neutron diffusion equation in nuclear engineering contain complicated systems of nonlinear partial differential equations (PDEs). When approximated numerically on a discrete grid or mesh, such problems produce large systems of algebraic nonlinear equations, whose numerical solution may be prohibitively expensive in terms of time and storage. High-performance (parallel) computers and efficient (parallelizable) algorithms are clearly necessary.

Three classical approaches to the solution of such systems are: Newton's method, preconditioned conjugate gradients (and related Krylov-subspace acceleration techniques), and multigrid. The first two approaches require the solution of large sparse linear systems at each iteration, which are themselves often solved by multigrid. Developing robust and efficient multigrid algorithms is thus of great importance.

The original multigrid algorithm was developed for the Poisson equation in a square, discretized by finite differences on a uniform grid. For this model problem, multigrid converges rapidly, and actually solves the problem in the minimal possible time (Poisson rate).

The original multigrid algorithm uses rediscretization of the original PDE on each grid in the hierarchy of coarse grids (geometric multigrid). Unfortunately, this approach doesn't work well for more complicated problems with nonrectangular domains, nonuniform grids, variable coefficients, or nonsymmetric or indefinite coefficient matrices. In these cases, matrix-based multigrid methods are required.

Matrix-based (or matrix-dependent) multigrid is a family of methods that use the information contained in the discrete system of equations (rather than the original PDE) to construct the operators used in the multigrid linear system solver. This way, a computer code can be written such that it accepts the coefficient matrix and right-hand side of the discrete system as input and produces the numerical solution as output. The method is automatic in the sense that the above code is independent of the particular application under consideration.

Because the elements in the coefficient matrix contain all the information about the properties of the boundary-value problem and its discretization, matrix-based multigrid methods are efficient even for PDEs with variable coefficients and complicated domains. In fact, matrix-based multigrid methods are the only multigrid methods that converge well even for diffusion problems with discontinuous coefficients, even when the discontinuity lines do not align with the coarse mesh.

This book offers a new approach towards the introduction and analysis of multigrid methods from an algebraic point of view. This approach is independent of the traditional, geometric approach, which is based on rediscretizing the original PDE. Instead, it uses only the algebraic properties of the original linear system to define, analyze, and apply the multigrid iterative method. This way, multigrid methods are well embedded in the family of iterative methods for the numerical solution of large, sparse linear systems. Indeed, as is shown below, multigrid methods can actually be viewed as special cases of domain-decomposition methods.

The present edition of this book introduces the multigrid methods from a unified domain-decomposition point of view. In particular, advanced multigrid versions (such as black-box multigrid, algebraic multigrid, and semicoarsening) can all be interpreted as domain-decomposition methods. Furthermore, it introduces a new semi algebraic approach for systems of PDEs. Each chapter ends with relevant exercises.

The first three parts are introductory. The first part introduces the concept of multilevel/multiscale in many different branches in mathematics and computer science. The second part gives the required background in discretization methods, including finite differences, finite volumes, and finite elements. The third part describes iterative methods for solving the linear system resulting from this discretization. In particular, it introduces multigrid methods from a domain-decomposition point of view.

The next three parts contain the heart of the book. The discussion starts from the simplified but common case of uniform grids, proceeds to the more complicated case of locally refined grids, and concludes with the most general and difficult case of completely unstructured grids. In each of these three parts, we concentrate on a particular multigrid version that fits our framework and method of analysis, and study it in detail. We believe that this study may shed light not only on this particular version but also on other multigrid versions as well.

These three parts are ordered from simple to complex: from the simple case of rectangular uniform grids, where spectral analysis may be used to predict the convergence factors (Part IV), through the more complex case of locally refined (and, in particular, semistructured) grids, where upper bounds for the condition number are available (Part V), to the most general case of completely unstructured grids, where the notions of stability and local anisotropy motivate and guide the actual design of algebraic multilevel methods (Part VI).

The second edition also contains the mathematical background required to make the book self-contained and suitable not only for experienced researchers but also for beginners in applied science and engineering. Thanks to the introductory parts, no background in numerical analysis or multigrid methods is needed. The only prerequisites are linear algebra and calculus. The book can thus serve as a textbook in courses in numerical analysis, numerical linear algebra, scientific computing, and numerical solution of PDEs at the advanced undergraduate and graduate levels.

Acknowledgments

I wish to thank the referees for their thorough and helpful reports, and the Wingate foundation for their kind support. I wish to thank Prof. Moshe Israeli and Prof. Avram Sidi for their valuable advice in the development of the AutoMUG method,

and Prof. Irad Yavneh for his valuable comments. I also wish to thank Dr. Joel Dendy and Dr. Dan Quinlan for useful discussions about semistructured grids, and Dr. Mac Hyman for his valuable support. I wish to thank Dr. Dhavide Aruliah and Prof. Uri Ascher for supplying the coefficient matrix for the Maxwell equations. Finally, I wish to thank my sons Roy and Amir for their constant patience and support.

Yair Shapira
August 2006

2

Preliminaries

In this chapter, we list some basic definitions and prove some standard lemmas used throughout the book. In particular, the lemmas are used to obtain some useful properties of the one- and two-dimensional Fourier (sine) transform, which are used often in the book.

2.1 Preliminary Notation and Definitions

Here we present some elementary notation and definitions from linear algebra that are useful throughout the book.

For every integer i, define

$$i \bmod 2 = \begin{cases} 0 & \text{if } i \text{ is even} \\ 1 & \text{if } i \text{ is odd.} \end{cases}$$

Also, for every two integers i and j, we say that

$$i \equiv j \bmod 2$$

if $i - j$ is even.

Let $A = (a_{i,j})_{1 \le i,j \le K}$ be a square matrix. Then K is called the *order* of A. The reverse direction is also true: saying that A is of order K implies that A is a square matrix.

The lower- (respectively, strictly lower-) triangular part of A is a matrix $L = (l_{i,j})_{1 \le i,j \le K}$ with the elements

$$l_{i,j} = \begin{cases} a_{i,j} & \text{if } i \ge j \text{ (respectively, } i > j) \\ 0 & \text{if } i < j \text{ (respectively, } i \le j). \end{cases}$$

Similarly, the upper- (respectively, strictly upper-) triangular part of A is a matrix $U = (u_{i,j})_{1 \le i,j \le K}$ with the elements

$$u_{i,j} = \begin{cases} a_{i,j} & \text{if } j \ge i \text{ (respectively, } j > i) \\ 0 & \text{if } j < i \text{ (respectively, } j \le i). \end{cases}$$

We say that A is lower- (respectively, strictly lower-) triangular if it is equal to its lower- (respectively, strictly lower-) triangular part. Similarly, we say that A is upper- (respectively, strictly upper-) triangular if it is equal to its upper- (respectively, strictly upper-) triangular part.

Let R denote the real field, and let C denote the complex field. For every $z \in C$, $\Re(z)$ denotes the real part of z, $\Im(z)$ denotes the imaginary part of z, and $\bar{z}$ denotes the complex conjugate of z. We also denote the k-dimensional vector space over C by

$$C^k \equiv \{(z_1, z_2, \ldots, z_k) \mid z_i \in C,\ 1 \le i \le k\}.$$

Similarly, we denote the 2-D Euclidean plane by

$$R^2 \equiv \{(x, y) \mid x \text{ and } y \text{ are real numbers}\},$$

and the 2-D infinite grid of pairs of integer numbers by

$$Z^2 \equiv \{(k, l) \mid k \text{ and } l \text{ are integer numbers}\}.$$

If a nonzero vector $v \in C^K$ and a complex or real number λ satisfy

$$Av = \lambda v,$$

then we say that v is an eigenvector of A and λ is an eigenvalue of A that corresponds to v. The set of all eigenvalues of A is called the *spectrum* of A. The maximal magnitude of a point in the spectrum of A is called the *spectral radius* of A and is denoted by $\rho(A)$.

The following theorem is the celebrated Gersgorin's Theorem.

Theorem 2.1 *The spectrum of A is contained in the following set.*

$$\bigcup_{i=1}^{K} \left\{ z \text{ is a complex number} \;\middle|\; |z - a_{i,i}| \le \sum_{1 \le j \le K,\ j \ne i} |a_{i,j}| \right\}.$$

Proof. Let v be an eigenvector of A with the corresponding eigenvalue λ. In other words,

$$v \ne \mathbf{0} \quad \text{and} \quad Av = \lambda v. \tag{2.1}$$

Let i be the index of the component in v of maximum modulus; that is,

$$|v_i| \ge |v_j|, \qquad 1 \le j \le K,\ j \ne i. \tag{2.2}$$

Consider the ith equation in (2.1):

$$\sum_{j=1}^{K} a_{i,j} v_j = \lambda v_i,$$

or

$$(a_{i,i} - \lambda) v_i = - \sum_{1 \le j \le K,\ j \ne i} a_{i,j} v_j.$$

By taking absolute values in both sides of the above equation and using also (2.2), one obtains

$$\begin{aligned} |a_{i,i} - \lambda| \ |v_i| &= \left| \sum_{1 \le j \le K, \ j \ne i} a_{i,j} v_j \right| \\ &\le \sum_{1 \le j \le K, \ j \ne i} |a_{i,j}| \ |v_j| \\ &\le \sum_{1 \le j \le K, \ j \ne i} |a_{i,j}| \ |v_i|. \end{aligned}$$

From (2.1) and (2.2), we also have that $|v_i| > 0$. Therefore, one can divide both sides of the above inequality by $|v_i|$, from which the theorem follows.

Furthermore, we say that A is *positive definite* if all its eigenvalues are positive. We say that A is *positive semidefinite* if all its eigenvalues are positive or zero. We say that A is *indefinite* if it has both eigenvalues with positive real parts and eigenvalues with negative real parts.

The square root of a positive-semidefinite matrix is a matrix with the same eigenvectors as the original matrix but with corresponding eigenvalues that are the square roots of the corresponding eigenvalues of the original matrix.

We say that A is diagonal if $a_{i,j} \ne 0$ implies $i = j$, that is, only main-diagonal elements can be nonzero, and all off-diagonal elements vanish.

Let I denote the identity matrix of order K, that is, the diagonal matrix whose main-diagonal elements are all equal to 1. Define the standard unit vector $e^{(i)}$ by

$$e_j^{(i)} = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \ne i. \end{cases}$$

In other words, $e^{(i)}$ is the ith column in the identity matrix I.

We say that A is *tridiagonal* if $a_{i,j} \ne 0$ implies $|i - j| \le 1$. That is, A is a matrix with only three nonzero diagonals: the main diagonal and the two diagonals that are adjacent to it. All other diagonals vanish. In this case, we write

$$A = tridiag(b_i, c_i, d_i);$$

that is, the only nonzero elements in the ith row are b_i, c_i, and d_i, in this order. Of course, the first row contains only two nonzero elements c_1 and d_1, and the last row contains only two nonzero elements b_K and c_K.

The diagonal part of A, $diag(A)$, is the diagonal matrix with the same main-diagonal elements as A:

$$diag(A) = diag(a_{i,i})_{1 \le i \le K} = diag(a_{1,1}, a_{2,2}, \ldots, a_{K,K}).$$

Similarly, $bidiag(A)$, $tridiag(A)$, $pentadiag(A)$, and $blockdiag(A)$ contain, respectively, two diagonals, three diagonals, five diagonals, and a diagonal of blocks with the same elements as in A, and zeroes elsewhere.

We say that A is nonnegative if all its elements are positive or zero. We say that A is an L-matrix if all its off-diagonal elements are either negative or zero. We say that A is an M-matrix if it is an L-matrix with positive main-diagonal elements and a nonnegative inverse A^{-1} [118].

The M-matrix property is particularly important in discrete approximations to differential operators: because the Green function that represents the inverse of the differential operator is usually nonnegative, so should also be the inverse of the matrix that contains the discrete approximation to that differential operator.

Finally, A is diagonally dominant if, for every $1 \leq i \leq K$,

$$a_{i,i} \geq \sum_{1 \leq j \leq K,\ j \neq i} |a_{i,j}|.$$

The following standard lemma is useful in the analysis in Parts III and IV below.

Lemma 2.1 *If A is diagonally dominant, then its spectrum is contained in the set*

$$\bigcup_{i=1}^{K} \{z \text{ is a complex number } \mid |z - a_{i,i}| \leq a_{i,i}\}.$$

As a result,

$$\rho(A) \leq 2\rho(diag(A)).$$

If, in addition, A has a real spectrum, then it is also positive semidefinite.

Proof. The lemma follows from Theorem 2.1.

The transpose of A, A^t, is defined by

$$\left(A^t\right)_{i,j} = a_{j,i}.$$

We say that A is symmetric if A is real and $A = A^t$. If A is symmetric and positive definite, then we say that A is symmetric positive definite or SPD.

The adjoint to A is defined by

$$(A^*)_{i,j} = \bar{a}_{j,i}.$$

In other words,

$$A^* = \bar{A}^t.$$

We say that A is hermitian if $A = A^*$.

In this book, however, we deal with real matrices only, so one can assume $A^* = A^t$. The only place where complex matrices are used is in the numerical examples with indefinite matrices. Therefore, we use "*" only when we would like to emphasize that, for complex matrices, the adjoint should be used rather than the transpose. In fact, it would be possible to replace "^{t}" by "*" and "symmetric" by "hermitian" throughout the book, with only few additional adjustments.

The inner product of two vectors $u, v \in C^K$ is defined by

$$(u, v)_2 \equiv u \cdot \bar{v} \equiv \sum_{i=1}^{K} u_i \bar{v}_i.$$

The l_2 norm of a vector $v \in C^K$ is defined by

$$\|v\|_2 \equiv \sqrt{(v, v)_2}.$$

The l_1 norm of a vector $v \in C^K$ is

$$|v|_1 = \sum_{i=1}^{K} |v_i|.$$

The l_∞ norm of a vector $v \in C^K$ is

$$|v|_\infty = \max_{1 \le i \le K} |v_i|.$$

The l_2 norm of the matrix A is defined by

$$\|A\|_2 = \max_{v \in C^K,\ v \neq 0} \frac{\|Av\|_2}{\|v\|_2} = \max_{v \in C^K,\ \|v\|_2 = 1} \|Av\|_2.$$

The l_1 norm of the matrix A is defined by

$$\|A\|_1 = \max_{v \in C^K,\ v \neq 0} \frac{\|Av\|_1}{\|v\|_1} = \max_{v \in C^K,\ \|v\|_1 = 1} \|Av\|_1.$$

The l_∞ norm of the matrix A is defined by

$$\|A\|_\infty = \max_{v \in C^K,\ v \neq 0} \frac{\|Av\|_\infty}{\|v\|_\infty} = \max_{v \in C^K,\ \|v\|_\infty = 1} \|Av\|_\infty.$$

The following lemma states useful matrix-norm inequalities, including the triangle inequality.

Lemma 2.2 *Let $\|\cdot\|$ denote the l_1, l_2, or l_∞ norm. Then we have*

$$\rho(A) \le \|A\|.$$

Moreover,

$$\|Av\| \le \|A\| \cdot \|v\|$$

for every K-dimensional vector v. Furthermore, if B is another matrix of order K, then

$$\|AB\| \le \|A\| \cdot \|B\|$$

and

$$\|A + B\| \le \|A\| + \|B\|.$$

Proof. Let λ be the maximal eigenvalue of A (in terms of magnitude), and let u be the corresponding eigenvector. Then we have

$$\|A\| = \max_{v \in C^K,\ v \neq \mathbf{0}} \frac{\|Av\|}{\|v\|} \ge \frac{\|Au\|}{\|u\|} = |\lambda| = \rho(A).$$

Next, for every nonzero vector $v \in C^K$,

$$\|Av\| = \|A(v/\|v\|)\| \cdot \|v\| \le \max_{w \in C^K,\ \|w\| = 1} \|Aw\| \cdot \|v\| = \|A\| \cdot \|v\|.$$

As a result, we have

$$\|ABv\| \le \|A\| \cdot \|Bv\| \le \|A\| \cdot \|B\| \cdot \|v\|,$$

which implies that

$$\|AB\| \max_{v \in C^K,\ v \neq \mathbf{0}} \frac{\|ABv\|}{\|v\|} \leq \|A\| \cdot \|B\|.$$

Finally, we have

$$\begin{aligned}
\|A+B\| &= \max_{v \in C^K,\ v \neq \mathbf{0}} \frac{\|(A+B)v\|}{\|v\|} \\
&\leq \max_{v \in C^K,\ v \neq \mathbf{0}} \frac{\|Av\| + \|Bv\|}{\|v\|} \\
&\leq \max_{v \in C^K,\ v \neq \mathbf{0}} \frac{\|Av\|}{\|v\|} + \max_{v \in C^K,\ v \neq \mathbf{0}} \frac{\|Bv\|}{\|v\|} \\
&= \|A\| + \|B\|.
\end{aligned}$$

This completes the proof of the lemma.

The following lemma expresses $\|A\|_1$ in terms of the elements in A.

Lemma 2.3

$$\|A\|_1 = \max_{1 \leq j \leq K} \sum_{i=1}^{K} |a_{i,j}|.$$

Proof. Clearly, the l_1 vector norm is convex. Indeed, for any two K-dimensional vectors u and v and a parameter $0 < \alpha < 1$,

$$\|\alpha u + (1-\alpha)v\|_1 \leq \|\alpha u\|_1 + \|(1-\alpha)v\|_1 = \alpha\|u\|_1 + (1-\alpha)\|v\|_1.$$

Thus, the vector v that satisfies $\|v\|_1 = 1$ and maximizes $\|Av\|_1$ is the standard unit vector $e^{(j)}$ for which the jth column in A has a maximum l_1 norm, which is also $\|A\|_1$. This completes the proof of the lemma.

The following lemma expresses $\|A\|_\infty$ in terms of the elements in A.

Lemma 2.4

$$\|A\|_\infty = \max_{1 \leq i \leq K} \sum_{j=1}^{K} |a_{i,j}|.$$

Proof. Clearly, the vector v that satisfies $\|v\|_\infty = 1$ and maximizes $\|Av\|_\infty$ is the vector defined by

$$v_i = \bar{a}_{i,j}/|a_{i,j}|$$

for that i for which $\sum_j |a_{i,j}|$ is maximal (and therefore also equal to $\|A\|_\infty$). This completes the proof of the lemma.

Corollary 2.1

$$\|A^t\|_\infty = \|A\|_1.$$

Proof. The corollary follows from Lemmas 2.3 and 2.4.

Let D be an SPD matrix of order K. The inner product induced by D is defined by

$$(u, v)_D = (u, Dv)_2.$$

The vector norm induced by D is defined by

$$\|v\|_D = \sqrt{(v,v)_D}.$$

The norm of A with respect to $(\cdot,\cdot)_D$ is defined by

$$\|A\|_D = \max_{v\in C^K,\ v\neq 0} \frac{\|Av\|_D}{\|v\|_D} = \max_{v\in C^K,\ \|v\|_D=1} \|Av\|_D.$$

The following lemma is analogous to Lemma 2.2.

Lemma 2.5 *Let D be an SPD matrix. Then*

$$\rho(A) \leq \|A\|_D.$$

Moreover, for any vector $v \in C^K$,

$$\|Av\|_D \leq \|A\|_D \|v\|_D.$$

Furthermore, for any matrix B of order K,

$$\|AB\|_D \leq \|A\|_D \|B\|_D$$

and

$$\|A+B\|_D \leq \|A\|_D + \|B\|_D.$$

Proof. The proof is similar to that of Lemma 2.2.

We say that the matrix A_D^t is the adjoint of A with respect to $(\cdot,\cdot)_D$ if

$$(u, A_D^{tv})_D = (Au, v)_D$$

for every two vectors $u, v \in C^K$. The following lemma states that the adjoint of the adjoint is the original matrix.

Lemma 2.6 *Let D be an SPD matrix. Then*

$$\left(A_D^t\right)_D^t = A.$$

Proof. The lemma follows from the fact that

$$(A_D^t v, u)_D = (v, Au)_D$$

for every two vectors $u, v \in C^K$.

We say that A is symmetric with respect to $(\cdot,\cdot)_D$ if

$$A_D^t = A.$$

Lemma 2.7 *Let D be an SPD matrix. Then $A_D^t A$ is symmetric with respect to $(\cdot,\cdot)_D$.*

Proof. The lemma follows from the fact that, for every two vectors $u, v \in C^K$,

$$(u, A_D^t Av)_D = (Au, Av)_D = (A_D^t Au, v)_D.$$

We say that A is orthogonal if

$$A^* A = I,$$

or, in other words, all the columns of A are orthonormal with respect to the usual inner product $(\cdot,\cdot)_2$.

Let A be a square or rectangular matrix, $A = (a_{i,j})_{1\le i\le K_1, 1\le j\le K_2}$. The transpose of A, A^t, is defined by

$$\left(A^t\right)_{j,i} = a_{i,j}, \quad 1 \le i \le K_1,\ 1 \le j \le K_2.$$

The following standard lemma gives an alternative definition to the notion of the transpose of a matrix.

Lemma 2.8 *Let $Z = (z_{i,j})_{1\le i\le K_2, 1\le j\le K_1}$ be a square or rectangular matrix. Then Z is the transpose of A if and only if*

$$(Au, v)_2 = (u, Zv)_2 \tag{2.3}$$

for every two vectors $u \in C^{K_2}$ and $v \in C^{K_1}$.

Proof. Let us first prove the "only if" part. Assume that $Z = A^t$. Then,

$$(Au, v)_2 = \sum_{i=1}^{K_1}\sum_{j=1}^{K_2} a_{i,j}u_j\bar{v}_i = \sum_{j=1}^{K_2}\sum_{i=1}^{K_1} u_j z_{j,i}\bar{v}_i = (u, Zv)_2.$$

Let us now prove the "if" part. Assume that (2.3) holds. Then, the assertion that $z_{j,i} = a_{i,j}$ follows by picking a vector u whose all components vanish except its jth component, u_j, which is equal to 1, and a vector v whose all components vanish except its ith component, v_i, which is equal to 1. This completes the proof of the lemma.

We say that the matrices A and Z are the transpose of each other with respect to $(\cdot,\cdot)_D$ if

$$(Au, v)_D = (u, Zv)_D$$

for every two vectors $u \in C^{K_2}$ and $v \in C^{K_1}$.

More standard linear algebra lemmas can be found in Section 2.3.

Define the absolute value of A by

$$|A|_{i,j} = |a_{i,j}|, \quad 1 \le i \le K_1,\ 1 \le j \le K_2. \tag{2.4}$$

Define the diagonal matrix of row sums of A by

$$rs(A) = diag\left(\sum_{j=1}^{K_2} a_{i,j}\right)_{1\le i\le K_1}. \tag{2.5}$$

By "grid" we mean a finite set of points in R^2. Let g be a grid; for example, the grid in (3.6) below. Let s be a subset of g (a subgrid). Let $l_2(s)$ be the set of complex-valued functions defined on s. In particular, $l_2(g)$ is the set of complex-valued functions defined on g, also referred to as "grid functions."

Define the injection operator $J_s : l_2(g) \to l_2(s)$ by

$$(J_s v)(k) \equiv v(k), \qquad v \in l_2(g), \quad k \in s. \tag{2.6}$$

In particular, J_c is the injection onto the coarse grid c defined in (6.1) below.

2.2 Application in Pivoting

Here we present an application of the algorithm in Section 1.18 to the pivoting of tridiagonal matrix. Let

$$A = tridiag(b_i, 1, d_i)$$

be a tridiagonal matrix of order N. The pivoting process is defined by $p_1 = 1$ and, for $i = 2, 3, \ldots, N$,

$$p_i = 1 - \frac{b_i d_{i-1}}{p_{i-1}}. \tag{2.7}$$

The pivots p_i are useful in the factorization of A as the product of triangular matrices. Indeed, define the diagonal matrix

$$X = diag(p_1, p_2, \ldots, p_N).$$

Let L be the strictly lower triangular part of A, and U be the strictly upper triangular part of A. Then we have

$$A = (L + X)X^{-1}(X + U).$$

This factorization is useful in solving linear systems with A as a coefficient matrix.

Now, the pivoting process in (2.7) can be reformulated as a continued fraction. Indeed, define

$$q_i = p_i - 1.$$

The pivoting process in (2.7) is thus equivalent to the process defined by $q_1 = 0$ and, for $i = 2, 3, \ldots, N$,

$$q_i = \frac{-b_i d_{i-1}}{1 + q_{i-1}}. \tag{2.8}$$

It is well known (see, e.g., [93]) that the convergents in a continued fraction may be interpreted as a sequence of products of 2×2 matrices. In fact, the current convergent is obtained from the previous one by multiplying it on the right by some 2×2 matrix. Here, however, the q_is are the "mirror image" of a continued fraction, in which q_i corresponds to the 2×2 matrix obtained from multiplying the matrix corresponding to q_{i-1} by some 2×2 matrix on the left rather than on the right. Still, this structure is well suited for the parallel algorithm in Section 1.18. The p_is are then obtained from the q_is, and the pivoting is completed.

The solution of the tridiagonal linear system also requires forward elimination in $L + X$ and back substitution in $X + U$. These processes can by themselves be formulated as products of affine transformations, and thus are also suitable for the parallel algorithm in Section 1.18. (See Section 2.2 in [96] for the details.)

2.3 Standard Lemmas about Symmetric Matrices

This section contains standard lemmas that are useful in the sequel. Basically, these lemmas show that matrices that are symmetric with respect to an inner product

of the form $(\cdot,\cdot)_D$ for some SPD matrix D enjoy the same properties as standard symmetric matrices. The first lemma shows that the inverse of a symmetric matrix is also symmetric.

Lemma 2.9 *Let D be an SPD matrix, and assume that A is symmetric with respect to $(\cdot,\cdot)_D$. Then A^{-1} is also symmetric with respect to $(\cdot,\cdot)_D$.*

Proof. For every two vectors x and y,

$$(x, A^{-1}y)_D = (AA^{-1}x, A^{-1}y)_D = (A^{-1}x, AA^{-1}y)_D = (A^{-1}x, y)_D.$$

This completes the proof of the lemma.

The next lemma shows that matrices that are symmetric with respect to an induced inner product have real spectra.

Lemma 2.10 *Let D be an SPD matrix, and assume that A is symmetric with respect to $(\cdot,\cdot)_D$. Then A has a real spectrum.*

Proof. If

$$Av = \lambda v$$

for a nonzero vector z and a scalar λ, then

$$\lambda(v,v)_D = (Av,v)_D = (v,Av)_D = \bar{\lambda}(v,v)_D.$$

This completes the proof of the lemma.

The next lemma says that a matrix is symmetric with respect to some inner product if and only if there exists a basis of orthonormal eigenvectors (with respect to this inner product) that correspond to real eigenvalues.

Lemma 2.11 *Let D be an SPD matrix. Then A is symmetric with respect to $(\cdot,\cdot)_D$ if and only if there exists a basis of eigenvectors of A that correspond to real eigenvalues and are orthonormal with respect to $(\cdot,\cdot)_D$ (i.e., they are orthogonal to each other with respect to $(\cdot,\cdot)_D$ and their D-induced norm is equal to 1).*

Proof. Let us first prove the "only if" part. Assume that A is symmetric with respect to $(\cdot,\cdot)_D$. Assume also that

$$Av = \lambda v \quad \text{and} \quad Au = \mu u$$

for some nonzero vectors v and u and scalars $\lambda \neq \mu$. Using Lemma 2.10, we then have

$$\lambda(v,u)_D = (Av,u)_D = (v,Au)_D = \bar{\mu}(v,u)_D = \mu(v,u)_D,$$

which implies

$$(v,u)_D = 0.$$

Consider now the Jordan block of A corresponding to the eigenvalue λ, and assume that the corresponding Jordan subspace is spanned by the vectors $w_1, w_2, \ldots, w_{n+1}$ in the Jordan basis. By induction, assume also that every invariant subspace of A of dimension at most n can be spanned by n eigenvectors of A that are orthonormal with respect to $(\cdot,\cdot)_D$. In particular, the span of $w_1, w_2, \ldots, w_n$ can

be spanned by such orthonormal eigenvectors $v_1, v_2, \ldots, v_n$ corresponding to the eigenvalue λ. For some scalars $\alpha_1, \alpha_2, \ldots, \alpha_n$, we have

$$Aw_{n+1} = \lambda w_{n+1} + \sum_{i=1}^{n} \alpha_i v_i.$$

Therefore, for every $1 \leq j \leq n$, we have

$$\lambda(w_{n+1}, v_j)_D = (w_{n+1}, Av_j)_D = (Aw_{n+1}, v_j)_D = \lambda(w_{n+1}, v_j)_D + \alpha_j,$$

implying that

$$\alpha_j = 0.$$

Thus, w_{n+1} is also an eigenvector of A corresponding to the eigenvalue λ. By applying a Gramm–Schmidt process to w_{n+1} with respect to $(\cdot,\cdot)_D$, one obtains an eigenvector v_{n+1} that corresponds to the eigenvalue λ and is also orthogonal to $v_1, v_2, \ldots, v_n$ with respect to $(\cdot,\cdot)_D$ and satisfies $\|v_{n+1}\|_D = 1$. This completes the induction step and the proof of the "only if" part.

Let us now prove the "if" part. Assume that there exists a basis of eigenvectors of A that correspond to real eigenvalues and are orthonormal with respect to $(\cdot,\cdot)_D$. Let V be the matrix whose columns are these eigenvectors and Λ the diagonal matrix whose diagonal elements are the corresponding real eigenvalues. In other words,

$$AV = V\Lambda,$$

and, hence,

$$A = V\Lambda V^{-1}, \tag{2.9}$$

with

$$V^* D V = I. \tag{2.10}$$

From (2.10) we have

$$V^{-1} = V^* D. \tag{2.11}$$

By using (2.11) in (2.9), we get

$$A = V\Lambda V^* D. \tag{2.12}$$

As a result, for every two vectors x and y we have

$$\begin{aligned}
(Ax, y)_D &= (V\Lambda V^* Dx, y)_D \\
&= (DV\Lambda V^* Dx, y)_2 \\
&= (x, DV\Lambda V^* Dy)_2 \\
&= (Dx, V\Lambda V^* Dy)_2 \\
&= (x, V\Lambda V^* Dy)_D \\
&= (x, Ay)_D .
\end{aligned}$$

This completes the proof of the "if" part and the whole lemma.

The next lemma gives alternative definitions to the terms "positive definite" and "positive semidefinite" defined in Section 2.1 above. When the condition in the lemma is satisfied, we may say that A is SPD with respect to $(\cdot,\cdot)_D$.

Lemma 2.12 *Let D be an SPD matrix, and assume that A is symmetric with respect to $(\cdot,\cdot)_D$. Then A is positive definite (respectively, positive semidefinite) if and only if for every nonzero vector x $(Ax,x)_D$ is positive (respectively, positive or zero).*

Proof. Let us first prove the "only if" part. Let x be a nonzero vector. From Lemma 2.11, we have

$$x = \sum \alpha_i v_i,$$

where the α_is are some scalars and the v_is are the orthonormal eigenvectors of A with corresponding eigenvalues λ_i. Thus, we have

$$(Ax,x)_D = \sum \lambda_i |\alpha_i|^2.$$

Because x is nonzero, at least one of the α_is is also nonzero. Therefore, if A is positive definite (respectively, positive semidefinite), then all the λ_is are positive (respectively, positive or zero), implying that $(Ax,x)_D$ is positive (respectively, positive or zero). This completes the proof of the "only if" part.

Let us now prove the "if" part. Indeed, if $(Ax,x)_D$ is positive (respectively, positive or zero) for every nonzero vector x, then this is particularly true for an eigenvector of A, implying that the corresponding eigenvalue is positive (respectively, positive or zero). This completes the proof of the lemma.

The next lemma is a version of the "minimax" theorem.

Lemma 2.13 *Let D be an SPD matrix, and assume that A is symmetric with respect to $(\cdot,\cdot)_D$. Then*

$$\rho(A) = \max_{x \neq \mathbf{0}} \left| \frac{(Ax,x)_D}{(x,x)_D} \right|. \tag{2.13}$$

Proof. Let us solve a problem that is equivalent to the right-hand side in (2.13).

$$\text{maximize} \quad |(Ax,x)_D| \quad \text{subject to} \quad (x,x)_D = 1. \tag{2.14}$$

From Lemma 2.11, we can represent every vector x as

$$x = \sum \alpha_i v_i,$$

where the α_is are some scalars and the v_is are the orthonormal eigenvectors of A with corresponding eigenvalues λ_i. Furthermore, we have

$$(Ax,x)_D = \sum \lambda_i |\alpha_i|^2.$$

Thus, the problem in (2.14) can be reformulated as

$$\text{maximize} \quad \left| \sum \lambda_i |\alpha_i|^2 \right| \quad \text{subject to} \quad \sum |\alpha_i|^2 = 1, \tag{2.15}$$

where the λ_is are the eigenvalues of A and the α_is are the unknowns to be found. From Lagrange theory, we have that the extreme vectors $(\alpha_1, \alpha_2, \ldots)^t$ for the function to be maximized in (2.15) are the ones for which the gradient of that function is a scalar multiple of the gradient of the constraint; that is, $(\lambda_1\alpha_1, \lambda_2\alpha_2, \ldots)^t$ is a scalar multiple of $(\alpha_1, \alpha_2, \ldots)^t$. This could happen only when all the α_is vanish

except those that correspond to one of the eigenvalues, say λ_j. At this extreme vector we have

$$\sum \lambda_i |\alpha_i|^2 = \lambda_j \sum |\alpha_i|^2 = \lambda_j.$$

This implies that the solution to (2.14) is the eigenvector of A with the eigenvalue of the largest possible magnitude, namely, $\rho(A)$. This completes the proof of the lemma.

Lemma 2.14 *Let D be an SPD matrix, and assume that A is symmetric with respect to $(\cdot,\cdot)_D$. Then*

$$\|A\|_D = \rho(A).$$

Proof. The proof is similar to the proof of Lemma 2.13, except that here one should solve the problems

$$\text{maximize} \quad (Ax, Ax)_D \quad \text{subject to} \quad (x,x)_D = 1 \tag{2.16}$$

and

$$\text{maximize} \quad \sum \lambda_i^2 |\alpha_i|^2 \quad \text{subject to} \quad \sum |\alpha_i|^2 = 1 \tag{2.17}$$

rather than (2.14) and (2.15), respectively. The solution of (2.16) and (2.17) leads to

$$\|A\|_D^2 = \rho(A)^2,$$

which completes the proof of the lemma.

The next lemma shows that the energy norm induced by a positive definite matrix is bounded in terms of the energy norm induced by a yet "more positive definite" matrix:

Lemma 2.15 *Let D be an SPD matrix, and assume that A, $\acute{A}$ and $A - \acute{A}$ are symmetric with respect to $(\cdot,\cdot)_D$ and positive semidefinite. Then for every vector x we have*

$$(x, \acute{A}x)_D \le (x, Ax)_D \tag{2.18}$$

Furthermore,

$$\|\acute{A}\|_D \le \|A\|_D. \tag{2.19}$$

Proof. Equation (2.18) follows from Lemma 2.12 and

$$(x, Ax)_D = (x, \acute{A}x)_D + (x, (A - \acute{A})x)_D \ge (x, \acute{A}x)_D.$$

Then, (2.19) follows from (2.18) and Lemmas 2.12 through 2.14. This completes the proof of the lemma.

The next lemma provides some properties for the square root of an SPD matrix.

Lemma 2.16 *Let D be an SPD matrix, and assume that A is symmetric with respect to $(\cdot,\cdot)_D$ and positive semidefinite. Then*

$$(A^{1/2})^2 = A, \tag{2.20}$$

$A^{1/2}$ is also symmetric with respect to $(\cdot,\cdot)_D$, and

$$\|A^{1/2}\|_D^2 = \|A\|_D. \tag{2.21}$$

Proof. From the "only if" part in Lemma 2.11, we have that (2.9) holds (with the notation used there). From the definition of the square root of a matrix in Section 2.1, we have

$$A^{1/2} = V\Lambda^{1/2}V^{-1},$$

which implies (2.20). From the "if" part in Lemma 2.11, we have that $A^{1/2}$ is also symmetric with respect to $(\cdot,\cdot)_D$. Finally, from Lemma 2.14, we have

$$\|A^{1/2}\|_D^2 = \rho(A^{1/2})^2 = \rho(A) = \|A\|_D,$$

which completes the proof of the lemma.

The following lemma gives an alternative definition to the norm of a matrix.

Lemma 2.17 *Let D be an SPD matrix. Then*

$$\|A\|_D = \|D^{1/2}AD^{-1/2}\|_2.$$

Proof. Using Lemma 2.16 (with D there being the identity matrix I and A there being the SPD matrix D from this lemma), we have

$$\begin{aligned}\|A\|_D &= \max_{v\neq 0}\frac{\|Av\|_D}{\|v\|_D}\\ &= \max_{D^{1/2}v\neq 0}\frac{\|D^{1/2}AD^{-1/2}(D^{1/2}v)\|_2}{\|D^{1/2}v\|_2}\\ &= \|D^{1/2}AD^{-1/2}\|_2.\end{aligned}$$

This completes the proof of the lemma.

We conclude with two lemmas regarding more general matrices, which are not necessarily symmetric with respect to any inner product. The following lemma states that the norm of the adjoint is the same as the norm of the original matrix:

Lemma 2.18 *Let D be an SPD matrix, and let A_D^t be the adjoint of A with respect to $(\cdot,\cdot)_D$. Then*

$$\|A_D^t\|_D = \|A\|_D.$$

Proof. On one hand, we have from Lemmas 2.5, 2.7, and 2.13 that

$$\begin{aligned}\|A\|_D^2 &= \max_{v\in C^K,\ v\neq\mathbf{0}}\frac{\|Av\|_D^2}{\|v\|_D^2}\\ &= \max_{v\in C^K,\ v\neq\mathbf{0}}\frac{(Av,Av)_D}{(v,v)_D}\\ &= \max_{v\in C^K,\ v\neq\mathbf{0}}\frac{(A_D^tAv,v)_D}{(v,v)_D}\\ &= \rho(A_D^tA)\\ &\leq \|A_D^tA\|_D\\ &\leq \|A_D^t\|_D\|A\|_D,\end{aligned}$$

which implies

$$\|A\|_D \leq \|A_D^t\|_D.$$

On the other hand, using Lemma 2.6, we also have

$$\begin{aligned}
\|A_D^t\|_D^2 &= \max_{v \in C^K,\ v \neq \mathbf{0}} \frac{\|A_D^{tv}\|_D^2}{\|v\|_D^2} \\
&= \max_{v \in C^K,\ v \neq \mathbf{0}} \frac{(A_D^{tv}, A_D^{tv})_D}{(v,v)_D} \\
&= \max_{v \in C^K,\ v \neq \mathbf{0}} \frac{(AA_D^{tv}, v)_D}{(v,v)_D} \\
&= \rho(AA_D^t) \\
&\leq \|AA_D^t\|_D \\
&\leq \|A\|_D \|A_D^t\|_D,
\end{aligned}$$

which implies

$$\|A_D^t\|_D \leq \|A\|_D.$$

This completes the proof of the lemma.

Finally, the following lemma bounds the norm induced by an SPD matrix D.

Lemma 2.19 *Let D be an SPD matrix, and let A_D^t be the adjoint of A with respect to $(\cdot,\cdot)_D$. Then*

$$\|A\|_D \leq \sqrt{\|A_D^t\|_\infty \|A\|_\infty}.$$

Proof. From Lemmas 2.2, 2.7, and 2.13, we have that

$$\begin{aligned}
\|A\|_D^2 &= \max_{v \in C^K,\ v \neq \mathbf{0}} \frac{\|Av\|_D^2}{\|v\|_D^2} \\
&= \max_{v \in C^K,\ v \neq \mathbf{0}} \frac{(Av, Av)_D}{(v,v)_D} \\
&= \max_{v \in C^K,\ v \neq \mathbf{0}} \frac{(A_D^t Av, v)_D}{(v,v)_D} \\
&= \rho(A_D^t A) \\
&\leq \|A_D^t A\|_\infty \\
&\leq \|A_D^t\|_\infty \|A\|_\infty.
\end{aligned}$$

This completes the proof of the lemma.

Corollary 2.2 *Let D be an SPD matrix, and assume that A is symmetric with respect to $(\cdot,\cdot)_D$. Then*

$$\|A\|_D \leq \|A\|_\infty.$$

Proof. The corollary follows from Lemma 2.19 by setting $A_D^t = A$.

Corollary 2.3 *Assume that $A = (a_{i,j})$ is symmetric and diagonally dominant with positive main-diagonal elements. Define $D = diag(A)$. Then*

$$\|D^{-1/2}AD^{-1/2}\|_2 = \|D^{-1}A\|_D \leq 2.$$

Proof. Because A is symmetric, $D^{-1}A$ is symmetric with respect to $(\cdot,\cdot)_D$. The corollary follows from Lemma 2.17, Corollary 2.2, and Lemma 2.4, when applied to $D^{-1}A$.

2.4 The Fourier Transform

In many problems in pure and applied science, the notion of "scales" or "levels" is helpful not only in the solution process but also in a deep understanding of the problem and its complexity. Most often, each scale has its own contribution to and influence on the mathematical or physical phenomenon. A most useful tool in distinguishing between different scales is the Fourier transform. This transform interprets a given function in terms of frequencies or waves rather than numerical nodal values. Low frequencies or long waves describe the variation of the function in coarse scales, whereas high frequencies or short waves correspond to fine-scale, delicate changes in the function. The Fourier transform has many applications in pure and applied mathematics, from classical, analytic fields such as functional analysis to modern, practical fields such as digital signal processing.

The hierarchy of Fourier functions can be introduced as eigenfunctions of a boundary-value problem. Consider the ordinary differential equation

$$-u''(x) = \mathcal{F}(x), \qquad 0 < x < 1, \tag{2.22}$$

where $\mathcal{F}$ is a given function on $(0,1)$ and u is the unknown function on $[0,1]$. Assume also that Dirichlet boundary conditions that specify the values of u at the endpoints are given:

$$u(0) = u(1) = 0. \tag{2.23}$$

The boundary-value problem (2.22) and (2.23) is called a Sturm–Liouville problem. This problem has the set of eigenfunctions

$$\{\sin(\pi kx)\}_{k=1}^{\infty} \tag{2.24}$$

with the corresponding eigenvalues π^2k^2. In other words, the functions in (2.24) satisfy the boundary conditions in (2.23) and are also eigenfunctions of the second-derivative operator in (2.22):

$$-\sin''(\pi kx) = \pi^2k^2\sin(\pi kx).$$

It is well known that the eigenfunctions of a Sturm–Liouville problem are orthogonal to each other with respect to integration. Indeed, for every distinct $k \geq 1$ and $l \geq 1$,

$$\begin{aligned}\int_0^1 \sin(\pi kx)\sin(\pi lx)dx &= \frac{1}{2}\int_0^1 (\cos(\pi(k-l)x) - \cos(\pi(k+l)x))dx \\ &= \frac{1}{2\pi(k-l)}\left[\sin(\pi(k-l)x)\right]_0^1 - \frac{1}{2\pi(k+l)}\left[\sin(\pi(k+l)x)\right]_0^1 \\ &= 0 - 0 = 0.\end{aligned}$$

Let us now describe the discrete Fourier (sine) transform in an N-dimensional vector space. Let N be a positive integer, and define $h = 1/(N+1)$. The kth Fourier vector, or mode, is obtained by sampling the kth eigenfunction in (2.24) at the points

$$h, 2h, 3h, \ldots, Nh$$

In other words, for every $1 \leq k \leq N$, the Fourier mode (discrete wave) $v^{(k)}$ is defined by

$$v^{(k)} \equiv (2h)^{1/2}(\sin(\pi kh), \sin(2\pi kh), \ldots, \sin(N\pi kh))^t. \tag{2.25}$$

Define the symmetric, tridiagonal matrix of order N

$$A = tridiag(-1, 2, -1). \tag{2.26}$$

In fact, the matrix A is obtained from the finite-difference discretization method applied to the boundary-value problem (2.22) and (2.23) [see Section 3.5 below]. The Fourier modes $v^{(k)}$ are eigenvectors of A:

$$Av^{(k)} = 4\sin^2(\pi kh/2)v^{(k)}. \tag{2.27}$$

From the symmetry of A, we have that the $v^{(k)}$s form an orthogonal basis in C^N (Lemma 2.11).

Let $\mathcal{V}$ be the matrix whose columns are the $v^{(k)}$s. That is,

$$\mathcal{V} = \left(v^{(1)} \mid v^{(2)} \mid \cdots \mid v^{(N)}\right). \tag{2.28}$$

The matrix $\mathcal{V}$ represents an operator in C^N defined by $\mathcal{V}\colon C^N \to C^N$ defined by

$$u \to \mathcal{V}u, \quad u \in C^N. \tag{2.29}$$

This operator is referred to as the sine transform in one dimension. The following standard lemma shows that $\mathcal{V}$ is symmetric and orthogonal, thus equal to its inverse.

Lemma 2.20 *$\mathcal{V}$ is an orthogonal symmetric matrix. That is,*

$$\mathcal{V}^{-1} = \mathcal{V}^t = \mathcal{V}.$$

Proof. The symmetry of $\mathcal{V}$ follows from

$$\mathcal{V}_{j,k} = (2h)^{1/2}\sin(\pi jkh) = (2h)^{1/2}\sin(\pi kjh) = \mathcal{V}_{k,j}.$$

To show that $\mathcal{V}$ is also orthogonal, one needs to show that the $v^{(k)}$s are orthonormal, that is,

$$(v^{(k)}, v^{(l)})_2 = \begin{cases} 0 & \text{if } k \neq l \\ 1 & \text{if } k = l. \end{cases}$$

Now, the orthogonality of the $v^{(k)}$s follows from the discussion that follows (2.26) and (2.27) above. It is only left to show that these vectors are also normal, that is,

$$(v^{(k)}, v^{(k)})_2 = 1$$

for every $1 \le k \le N$. Indeed,

$$\begin{aligned}(v^{(k)}, v^{(k)})_2 &= 2h \sum_{j=1}^{N} \sin^2(\pi kjh) \\ &= h \sum_{j=1}^{N} \left(1 - \Re\left(\exp(2\sqrt{-1}\pi kjh)\right)\right) \\ &= hN + h - h \sum_{j=0}^{N} \Re\left(\exp(2\sqrt{-1}\pi kjh)\right) \\ &= 1 - h\Re\left(\sum_{j=0}^{N} \exp(2\sqrt{-1}\pi kjh)\right) \\ &= 1 - h\Re\left(\frac{1 - \exp(2\sqrt{-1}\pi k)}{1 - \exp(2\sqrt{-1}kh)}\right) \\ &= 1 - 0 = 1.\end{aligned}$$

This completes the proof of the lemma.

Lemma 2.20 shows that the sine transform in (2.29) may be thought of as a change of coordinates or basis. Instead of specifying the nodal values in the original vector u in (2.29), one obtains the various frequencies contained in u in the vector $\mathcal{V}u$. Each coordinate in $\mathcal{V}u$ provides the amplitude of a certain wave contained in u.

The Fourier transform defined above can also be slightly modified to handle boundary conditions different from those in (2.23). Consider, for example, the Dirichlet–Neumann boundary conditions

$$u(0) = u'(1) = 0 \tag{2.30}$$

that specify the value of u at one end point and the value of its derivative at the other end point. The eigenfunctions of the corresponding Sturm–Liouville problem (2.22), (2.30) are

$$\{\sin(\pi(k - 1/2)x)\}_{k=1}^{\infty}. \tag{2.31}$$

If h is redefined as $h = 1/(N + 1/2)$, then the Fourier modes $v^{(k)}$ are obtained from sampling the kth function in (2.31) at the points

$$h, 2h, 3h, \ldots, Nh.$$

These modes are the eigenvectors of the matrix whose elements are the same as in A, except of its lower-right element, which is equal to 1, rather than $A_{N,N} = 2$. Because this matrix is still symmetric, its eigenvectors are still orthogonal to each other as before.

The sine transform provides the representation of a vector in C^N in terms of its oscillations rather than its components or nodal values. In many cases, however, one is interested in oscillations in more than one spatial direction. Consider for example, a rectangular, uniform $N \times N$ grid and vectors in C^{N^2} that are defined on this grid, hence also referred to as grid functions. For $1 \le k, l \le N$, define the (k, l)-wave or

mode to be the grid function that is the tensor product of $v^{(k)}$ and $v^{(l)}$:

$$v_{j,m}^{(k,l)} = v_j^{(k)} v_m^{(l)}, \quad 1 \le j, m \le N.$$

Let $\mathcal{V}$ denote now the matrix of order N^2 whose columns are the $v^{(k,l)}$s, $1 \le k, l \le N$. The transform

$$u \to \mathcal{V}u, \qquad u \in C^{N^2} \tag{2.32}$$

is the 2-D (two-dimensional) sine transform. The following standard lemma establishes that the 2-D sine-transform matrix $\mathcal{V}$ is also orthogonal, so the 2-D sine transform in (2.32) may also be thought of as a change of basis in C^{N^2} from the usual nodal basis to the Fourier basis of waves or modes.

Lemma 2.21 *The 2-D sine-transform matrix $\mathcal{V}$ is an orthogonal symmetric matrix. That is,*

$$\mathcal{V}^{-1} = \mathcal{V}^t = \mathcal{V}.$$

Proof. The symmetry of $\mathcal{V}$ follows from the symmetry of $\mathcal{V}$ established in Lemma 2.20.

$$v_{m,n}^{(k,l)} = v_m^{(k)} v_n^{(l)} = v_k^{(m)} v_l^{(n)} = v_{k,l}^{(m,n)}.$$

The orthonormality of the $v^{(k,l)}$s also stems from Lemma 2.20. Indeed, consider the (k,l)- and (m,n)-waves:

$$\begin{aligned}
\left(v^{(k,l)}, v^{(m,n)}\right)_2 &= \sum_{1 \le i,j \le N} v_{i,j}^{(k,l)} v_{i,j}^{(m,n)} \\
&= \sum_{i=1}^{N} v_i^{(k)} v_i^{(m)} \sum_{j=1}^{N} v_j^{(l)} v_j^{(n)} \\
&= \left(v^{(k)}, v^{(m)}\right)_2 \left(v^{(l)}, v^{(n)}\right)_2 \\
&= \begin{cases} 1 & \text{if } k = m \text{ and } l = n \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

This completes the proof of the lemma.

2.5 Exercises

1. Repeat the exercises at the end of Chapter 1, only this time replace the abstract problem in Section 1.18 by the more concrete pivoting problem in Section 2.2. (If your code from these exercises is written as a template function in C++, then all you have to do is to use it in conjunction with the "matrix2" class in Section 2.20 in [103].)
2. Define the Fourier vectors $w^{(k)}$ $(0 \le k < N)$ by

$$w_j^{(k)} = N^{-1/2} \exp(2\pi\sqrt{-1}jkh), \qquad 1 \le j \le N,$$

where $h = 1/N$. Show that the $w^{(k)}$s are the eigenvectors of the symmetric $N \times N$ matrix

$$A = \begin{pmatrix} 2 & -1 & & & & & -1 \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & \ddots & & & \\ & & \ddots & \ddots & -1 & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ -1 & & & & & -1 & 2 \end{pmatrix}$$

with distinct positive (or zero) eigenvalues. Conclude that A is positive semidefinite, and that the $w^{(k)}$s form an orthonormal basis of C^N.

3. Define the Fourier (cosine) vectors $w^{(k)}$ $(0 \leq k < N)$ by

$$w_j^{(k)} = \sin(\pi/2 + \pi(j - 1/2)kh), \qquad 1 \leq j \leq N,$$

where $h = 1/N$. Show that the $w^{(k)}$s are the eigenvectors of the symmetric tridiagonal $N \times N$ matrix

$$A = \begin{pmatrix} 1 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & \ddots & & & \\ & & \ddots & \ddots & -1 & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{pmatrix}$$

with distinct positive (or zero) eigenvalues. Conclude that A is positive semidefinite, and that the $w^{(k)}$s form an orthogonal basis of R^N. Normalize the $w^{(k)}$s so that they are also orthonormal.

4. Define the 2-D Fourier vectors $w^{(k,l)}$ $(0 \leq k, l < N)$ by

$$w_{j,m}^{(k,l)} = w_j^{(k)} w_m^{(l)}, \qquad 1 \leq j, m \leq N.$$

Define the operators $X : R^{N^2} \to R^{N^2}$ and $Y : R^{N^2} \to R^{N^2}$ by

$$\begin{aligned} (Xv)_{i,j} &= v_{i,j} - v_{i,j+1} && \text{if } j = 0 \\ (Xv)_{i,j} &= 2v_{i,j} - v_{i,j-1} - v_{i,j+1} && \text{if } 0 < j < N \\ (Xv)_{i,j} &= v_{i,j} - v_{i,j-1} && \text{if } j = N \\ (Yv)_{i,j} &= v_{i,j} - v_{i+1,j} && \text{if } i = 0 \\ (Yv)_{i,j} &= 2v_{i,j} - v_{i-1,j} - v_{i+1,j} && \text{if } 0 < i < N \\ (Yv)_{i,j} &= v_{i,j} - v_{i-1,j} && \text{if } i = N, \end{aligned}$$

for every 2-D vector $v \in R^{N^2}$. Show that the $w^{(k,l)}$s are the eigenvectors of the symmetric operator $X+Y$ with distinct positive (or zero) eigenvalues. Conclude that $X + Y$ is positive semidefinite, and that the $w^{(k,l)}$s form an orthonormal basis of R^{N^2} (provided that the $w^{(k)}$s have been normalized in the previous exercise).

5. Define the Fourier (sine) vectors $w^{(k)}$ $(0 \le k < N)$ by

$$w_j^{(k)} = \sin(\pi(k+1/2)jh), \qquad 1 \le j \le N,$$

where $h = 1/(N+1/2)$. Show that the $w^{(k)}$s are the eigenvectors of the symmetric tridiagonal $N \times N$ matrix

$$A = \begin{pmatrix} 2 & -1 & & & & & \\ -1 & 2 & -1 & & & & \\ & -1 & 2 & \ddots & & & \\ & & \ddots & \ddots & -1 & & \\ & & & -1 & 2 & -1 & \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{pmatrix}$$

with distinct positive eigenvalues. Conclude that A is SPD, and that the $w^{(k)}$s form an orthogonal basis of R^N. Normalize the $w^{(k)}$s so that they are also orthonormal.

6. Define the 2-D Fourier vectors $w^{(k,l)}$ $(0 \le k, l < N)$ by

$$w_{j,m}^{(k,l)} = w_j^{(k)} w_m^{(l)}, \qquad 1 \le j, m \le N.$$

Define the operators $X : R^{N^2} \to R^{N^2}$ and $Y : R^{N^2} \to R^{N^2}$ by

$$\begin{aligned} (Xv)_{i,j} &= 2v_{i,j} - v_{i,j+1} && \text{if } j = 0 \\ (Xv)_{i,j} &= 2v_{i,j} - v_{i,j-1} - v_{i,j+1} && \text{if } 0 < j < N \\ (Xv)_{i,j} &= v_{i,j} - v_{i,j-1} && \text{if } j = N \\ (Yv)_{i,j} &= 2v_{i,j} - v_{i+1,j} && \text{if } i = 0 \\ (Yv)_{i,j} &= 2v_{i,j} - v_{i-1,j} - v_{i+1,j} && \text{if } 0 < i < N \\ (Yv)_{i,j} &= v_{i,j} - v_{i-1,j} && \text{if } i = N, \end{aligned}$$

for every 2-D vector $v \in R^{N^2}$. Show that the $w^{(k,l)}$s are the eigenvectors of the symmetric operator $X + Y$ with distinct positive eigenvalues. Conclude that $X + Y$ is SPD, and that the $w^{(k,l)}$s form an orthonormal basis of R^{N^2} (provided that the $w^{(k)}$s have been normalized in the previous exercise).