Research

CrossMark
click for updates

Author for correspondence:
Lloyd N. Trefethen
e-mail: trefethen@maths.ox.ac.uk

Continuous analogues of matrix factorizations

Alex Townsend[1] and Lloyd N. Trefethen[2]

[1]Department of Mathematics, MIT, Cambridge, MA 02139, USA
[2]Mathematical Institute, University of Oxford, Woodstock Road, Oxford OX2 6GG, UK

Analogues of singular value decomposition (SVD), QR, LU and Cholesky factorizations are presented for problems in which the usual discrete matrix is replaced by a 'quasimatrix', continuous in one dimension, or a 'cmatrix', continuous in both dimensions. Two challenges arise: the generalization of the notions of triangular structure and row and column pivoting to continuous variables (required in all cases except the SVD, and far from obvious), and the convergence of the infinite series that define the cmatrix factorizations. Our generalizations of triangularity and pivoting are based on a new notion of a 'triangular quasimatrix'. Concerning convergence of the series, we prove theorems asserting convergence provided the functions involved are sufficiently smooth.

1. Introduction

A fundamental idea of linear algebra is matrix factorization, the representation of matrices as products of simpler matrices that may be, for example, triangular, tridiagonal or orthogonal. Such factorizations provide a basic tool for describing and analysing numerical algorithms. For example, Gaussian elimination for solving a system of linear equations constructs a factorization of a matrix into a product of lower- and upper triangular matrices, which represent simpler systems that can be solved successively by forward elimination and back substitution.

In this article, we describe continuous analogues of matrix factorizations for contexts where vectors become univariate functions and matrices become bivariate functions.[1] Mathematically, some of the factorizations

---

[1]Our analogues involve finite or infinite sums of rank 1 pieces and stem from algorithms of numerical linear algebra. Different continuous analogues of matrix factorizations, with a different notion of triangularity related to Volterra integral operators, have been described for example in publications by Gohberg and his co-authors.
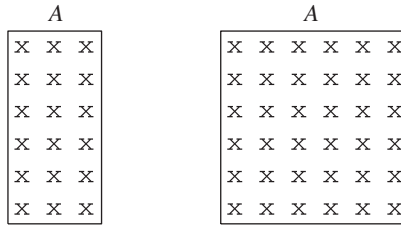
$$A \qquad\qquad A$$

Figure 1. A rectangular and a square matrix. The value of A in row i, column j is $A(i, j)$.

we shall present have roots going back a century in the work of Fredholm [1], Hilbert [2], Schmidt [3] and Mercer [4], which is marvelously surveyed in [5]. Algorithmically, they are related to recent methods of low-rank approximation of matrices and functions put forward by Bebendorf, Geddes, Hackbusch, Tyrtyshnikov and many others; see §8 for more names and references. In particular, we have been motivated by the problem of numerical approximation of bivariate functions for the Chebfun software project [6–8]. The part of Chebfun devoted to this task is called Chebfun2 and was developed by the first author [7, ch. 11–15]. Connections of our results to the low-rank approximations of Chebfun2 are mentioned here and there in this article, and, in particular, see the discussion of Chebfun2 computation in the second half of §8.

Despite these practical motivations, this is a theoretical paper. Although we shall make remarks about algorithms, we do not systematically consider matters of floating-point arithmetic, conditioning or stability.

Some of the power of the matrix way of thinking stems from the easy way in which it connects to our highly developed visual skills. Accordingly, we shall rely on schematic representations, and we shall avoid spelling out precise definitions of what it means, say, to multiply a quasimatrix by a vector when the associated schema makes it obvious to the experienced eye. To begin the discussion, figure 1 suggests the two kinds of discrete matrices we shall be concerned with, rectangular and square. An $m \times n$ matrix is an ordered collection of mn data values, which can be used as a representation of a linear mapping from $\mathbb{C}^n$ to $\mathbb{C}^m$. Our convention will be to show a rectangular matrix by a 6×3 array and a square one by a 6×6 array.

We shall be concerned with two kinds of continuous analogues of matrices. In the first case, one index of a rectangular matrix becomes continuous while the other remains discrete. Such structures seem to have been first discussed explicitly by de Boor [9], Stewart [10, pp. 33–34] and Trefethen & Bau [11, pp. 52–54]. Following Stewart, we call such an object a *quasimatrix*. The notion of a quasimatrix presupposes that a space of functions has been prescribed, and for simplicity we take this to be $C([a, b])$, the space of continuous real or complex functions defined on an interval $[a, b]$ with $-\infty < a < b < \infty$. An '$[a, b] \times n$ quasimatrix' is an ordered set of n functions in $C([a, b])$, which we think of as functions of a 'vertical' variable y. We depict it as shown in figure 2, which suggests how it can be used as a representation of a linear map from $\mathbb{C}^n$ to $C([a, b])$. Its (conjugate) 'transpose', an '$n \times [a, b]$ quasimatrix', is also a set of n functions in $C([a, b])$, which we think of as functions of a 'horizontal' variable x. We use each function as defining a linear functional on $C([a, b])$, so that the quasimatrix represents a linear map from $C([a, b])$ to $\mathbb{C}^n$.

Secondly, we shall consider the fully continuous analogue of a matrix, a *cmatrix*, which can be rectangular or square.[2] A cmatrix is a function of two continuous variables, and again, for simplicity, we take it to be a continuous function defined on a rectangle $[a, b] \times [c, d]$. Thus, a cmatrix is an element of $C([a, b] \times [c, d])$, and it can be used as a representation of a linear map from $C([c, d])$ to $C([a, b])$ (the kernel of a compact integral operator). To emphasize the matrix analogy, we denote a cmatrix generically by A rather than f and we refer to it as a 'cmatrix

[2]We are well aware that it is a little odd to introduce a new term for what is, after all, nothing more than a bivariate function. We have decided to go ahead with 'cmatrix', nonetheless, knowing from experience how useful the term 'quasimatrix' has been.
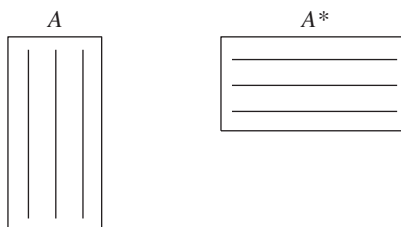
Figure 2. An $[a, b] \times n$ quasimatrix and its $n \times [a, b]$ conjugate transpose. Each column in the first case and row in the second is a function defined on $[a, b]$. For the case of A, on the left, the row index i has become a continuous variable y, and the value of A at vertical position y in column j is $A(y, j)$. Similarly on the right, the value of A^* in row i at horizontal position x is $A^*(i, x)$.
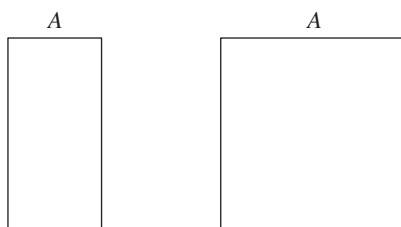

Figure 3. Rectangular and square cmatrices of dimensions $[a, b] \times [c, d]$ and $[a, b] \times [a, b]$, respectively. A cmatrix is just a bivariate function, but the special name is convenient for discussion of factorizations. We think of the vertical variable as y and the horizontal variable as x. For consistency with matrix conventions, a point in the rectangle is written (y, x), and the corresponding value of A is $A(y, x)$.

of dimensions $[a, b] \times [c, d]$'. The 'vertical' variable is y, the 'horizontal' variable is x, and for consistency with matrix notation, the pair of variables is written in the order (y, x), with $A(y, x)$ being the corresponding value of A.

Schematically, we represent a cmatrix by an empty box (figure 3).

A *square cmatrix* is a cmatrix with $c = a$ and $d = b$, in which case, for example, it makes sense to consider eigenvalue problems for the associated operator, although eigenvalue problems are not discussed here. A *Hermitian cmatrix* is a square cmatrix that satisfies $A^* = A$, that is, $\overline{A(x, y)} = A(y, x)$ for each $(y, x) \in [a, b] \times [a, b]$.

Note that this article does not consider infinite discrete matrices, a more familiar generalization of ordinary matrices for which there is also a literature of matrix factorizations. For cmatrix factorizations, we will, however, make use of the generalizations of quasimatrices to structures with infinitely many columns or rows, which will accordingly be said to be quasimatrices of dimensions $[a, b] \times \infty$ or $\infty \times [a, b]$.

Throughout this article, we work with the spaces of continuous functions $C([a, b])$, $C([c, d])$ and $C([a, b] \times [c, d])$, for our aim is to set forth fundamental ideas without getting lost in technicalities of regularity. We trust that if these generalizations of matrix factorizations prove useful, some of the definitions and results may be extended by future authors to less smooth function spaces.

2. Four matrix factorizations

We shall consider analogues of four matrix factorizations described in references such as [11–13]: LU, Cholesky, QR and singular value decomposition (SVD). The Cholesky factorization applies to square matrices (which must in addition be Hermitian and non-negative definite), whereas the other three apply more generally to rectangular matrices. For rectangular matrices, we shall

$$
\begin{array}{ccc}
A & L & U
\end{array}
$$

Figure 4. LU factorization of a matrix (without pivoting). Blank spaces indicate zero entries. The unit lower triangular matrix *L* has 1 on the diagonal and arbitrary entries below the diagonal. The factorization is unique if the columns of *A* are linearly independent, but it does not always exist.

assume $m \geq n$. There are many other factorizations we shall not discuss, such as similarity transformations to diagonal, tridiagonal, Hessenberg or triangular form.

We now review the four factorizations to fix notations and normalizations. Let *A* be a real or complex $m \times n$ matrix. An LU factorization is a factorization $A = LU$, where *U* is of dimensions $n \times n$ and upper triangular and *L* is of dimensions $m \times n$ and *unit lower triangular,* which means that it is lower triangular with diagonal entries equal to 1 (figure 4). Such a factorization can be computed by the algorithm of Gaussian elimination without pivoting.

A common interpretation of LU factorization is that it is a change of basis from the columns of *A* to the columns of *L*. Column a_1 of *A* is equal to u_{11} times column ℓ_1 of *L*, column a_2 of *A* is equal to $u_{12}\ell_1 + u_{22}\ell_2$, and so on. Another interpretation is that it is a representation of *A* as a sum of *n* matrices of rank 0 or 1.[3] If u_k^* denotes the *k*th row of *U*, then we have

$$
A = \sum_{k=1}^{n} \ell_k u_k^*. \tag{2.1}
$$

If the columns of *A* are linearly dependent, so that they fail to form a basis, then this will show up as one or more zero entries on the diagonal of *U*, making *U* singular.

Not every matrix has an LU factorization; the factorization exists if and only if Gaussian elimination completes a full *n* steps without an attempted division of a non-zero number by zero. To deal with arbitrary matrices, it is necessary to introduce some kind of *pivoting* in the elimination process. We shall return to this subject in §5.

For our second factorization, let *A* be a square, positive semi-definite Hermitian matrix. In this case, there is a symmetric variant of LU factorization known as *Cholesky factorization*, where $A = R^*R$ and *R* is upper triangular with non-negative real numbers on the diagonal (figure 5). The corresponding representation of *A* as a sum of rank 0 or 1 matrices takes the form

$$
A = \sum_{k=1}^{n} r_k r_k^*, \tag{2.2}
$$

where r_k^* is the *k*th row of *R*. (It is in cases where *A* is only semi-definite that zero vectors r_k may arise [12, §4.2.8].) Such a factorization can be computed by a symmetrized variant of Gaussian elimination which we shall call the Cholesky algorithm (though it is often called just Cholesky factorization). No pivoting is needed, and indeed, it is known that the algorithm when applied to a Hermitian matrix *A* completes successfully (meaning that at no step is the square root of a negative number required) if and only if *A* is non-negative definite. This property makes the Cholesky algorithm a standard method for testing numerically if a matrix is definite.

Our third and fourth factorizations apply to arbitrary rectangular matrices. A *QR factorization* of *A* is a factorization $A = QR$, where *Q* is an $m \times n$ matrix with orthonormal columns and *R* is an $n \times n$ upper triangular matrix with non-negative real numbers on the diagonal (figure 6). Every

---

[3]This interpretation of matrix factorizations can be generalized by the *Wedderburn rank-one reduction formula* [14].

Figure 5. Cholesky factorization of a Hermitian non-negative definite matrix. The diagonal entries of R are real and non-negative, as suggested by the symbol 'r'.

Figure 6. QR factorization of a matrix. The columns of Q are orthonormal, as suggested by the 'q' symbols.

matrix has a *QR* factorization, and if the columns of A are linearly independent, it is unique. The corresponding representation of A is

$$A = \sum_{k=1}^{n} q_k r_k^*. \tag{2.3}$$

A QR factorization can be computed by Gram–Schmidt orthogonalization or by Householder triangularization, as well as by other methods related to Householder triangularization such as Givens rotations. Throughout this article, when we refer to Gram–Schmidt orthogonalization, it is assumed that in cases of rank-deficiency, where a column of a matrix becomes zero after orthogonalization against previous columns, an arbitrary new orthonormal vector is introduced to keep the process going.

Finally, we consider the *SVD*. An SVD of a matrix A is a factorization $A = USV^*$, where U and V have orthonormal columns, known as the left and right *singular vectors* of A, respectively, and S is diagonal with real non-negative diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$, known as the *singular values* (figure 7). An SVD always exists, and the singular values are uniquely determined. The singular vectors corresponding to simple singular values are also unique up to complex signs, i.e. real or complex scaling factors of modulus 1. If some of the singular values are equal, there is further non-uniqueness associated with arbitrary breaking of ties.

The SVD corresponds to the representation

$$A = \sum_{j=1}^{n} \sigma_j u_j v_j^*. \tag{2.4}$$

For any k with $1 \leq k \leq n$, we may consider the partial sum

$$A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^*. \tag{2.5}$$

Figure 7. SVD of a matrix. The columns of U and V (i.e. the rows of V^*) are orthonormal, and the symbols 'σ' denote non-negative real numbers in non-increasing order.

Let $\|\cdot\|$ denote the Frobenius or Hilbert–Schmidt norm

$$\|A\| = \left(\sum_{i,j} |a_{ij}|^2 \right)^{1/2} \tag{2.6}$$

(note that this is not the operator 2-norm). The following property follows from the orthonormality of $\{u_j\}$ and $\{v_j\}$ and the ordering $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$: for each k with $1 \leq k \leq n-1$, A_k is a best rank k approximation to A with respect to $\|\cdot\|$, with corresponding error $E_k = A - A_k$ of magnitude

$$\|E_k\| = \|A - A_k\| = \tau_{k+1} = \left(\sum_{j=k+1}^{n} \sigma_j^2 \right)^{1/2}. \tag{2.7}$$

This fundamental property of the SVD originates with Schmidt [3,15,16] and was generalized to rectangular matrices by Eckart & Young [17]. The SVD also enjoys an optimality property in the 2-norm and other unitarily invariant norms, but our discussion of continuous analogues will be confined to the Frobenius norm.

We defined the SVD algebraically, then stated (2.7) as a consequence. It is also possible to put the reasoning in the other direction. Begin by defining $E_0 = A$. Find a non-negative real number σ_1 and unit vectors u_1 and v_1 such that $\sigma_1 u_1 v_1^*$ is a best approximation to E_0 of rank 1, and define $E_1 = E_0 - \sigma_1 u_1 v_1^*$; then find a non-negative real number σ_2 and unit vectors u_2 and v_2 such that $\sigma_2 u_2 v_2^*$ is a best approximation to E_1, and define $E_2 = E_1 - \sigma_2 u_2 v_2^*$; and so on. Then $\{u_j\}$ and $\{v_j\}$ are orthonormal sets, the scalars satisfy $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$, and we have constructed an SVD of A.

The remainder of this article is devoted to seven tasks. For three of the four matrix factorizations, we shall consider the two generalizations to the cases where A is a quasimatrix, continuous in one direction, or a cmatrix, continuous in both directions. For the Cholesky factorization, which applies only to square matrices, only the cmatrix generalization is relevant.

One of our factorizations can be generalized immediately, the SVD. The other three involve triangular quasimatrices and require pivoting to be introduced in the discussion. A central theme of this article is the generalization of the notions of pivoting and triangular matrices to a continuous setting (beginning in §5). For matrices, one can speak of LU, Cholesky and QR factorizations without pivoting, taking the next row and/or column at each step of the factorization process, but in continuous directions, there is no 'next' row or column. For quasimatrices, we shall see that row pivoting is necessary for LU factorization, which involves a triangular quasimatrix L. For cmatrices, we shall see that column pivoting is necessary for QR factorization, which involves a triangular quasimatrix R, and both row and column pivoting are necessary for LU and Cholesky factorizations, which involve two triangular quasimatrices. No pivoting is needed for the SVD.

For each of our factorizations, we shall consider five aspects: (i) *definition*, (ii) *history*, (iii) *elementary properties*, (iv) *advanced properties* (for the cmatrix factorizations), and (v) *algorithms*.

The elementary properties will be just selections, stated as theorems without proof. (A good foundation for the quasimatrix factorizations is [9].) The advanced properties focus on convergence of infinite series. In particular, our three main new theorems are theorems 6.1, 8.1 and 9.1. As for algorithms, in each case we present idealized methods that produce the required factorizations under the assumptions of exact arithmetic and exact pivoting operations, which will involve selection of globally maximal values in certain columns and/or rows or globally maximal-norm columns or rows. In practice, a computational system like Chebfun2 relies on approximations of these ideals.

3. QR factorization of a quasimatrix

The QR factorization of an $[a, b] \times n$ quasimatrix A is a straightforward extension of the rectangular matrix case. As in that case, column pivoting may be advantageous in some circumstances, but is not necessary for the factorization to make sense mathematically, and we do not include it. Following the schema of figure 8, we define the QR factorization as follows. Orthonormality of functions in $C([a, b])$ is defined with respect to the standard L^2 inner product.

Definition. Let A be an $[a, b] \times n$ quasimatrix. A *QR factorization* of A is a factorization $A = QR$, where Q is an $[a, b] \times n$ quasimatrix with orthonormal columns and R is an $n \times n$ upper triangular matrix with non-negative real numbers on the diagonal.

Note that throughout this article, each column of a quasimatrix is taken to be a function in $C([a, b])$ or $C([c, d])$. Thus, it is implicit in this definition that the columns of Q are continuous functions, though as discussed at the end of §1, this restriction could be relaxed in some cases.

The idea of QR factorization of quasimatrices was perhaps first mentioned explicitly in [11, pp. 52–54], though the underlying mathematics would have been familiar to Schmidt [3,5] and others going back many years.[4] The topic became one of the original capabilities in Chebfun [6,19], invoked by the command qr (which is not limited to continuous functions). Originally, the algorithm employed by Chebfun was Gram–Schmidt orthogonalization, which has the drawback that it is unstable if A is ill-conditioned. This was later replaced by an algorithm stable and applicable in all cases based on a continuous analogue of Householder triangularization [20]. Another less numerically stable approach for the full rank case, mentioned in [19], would be to form the Cholesky factorization R^*R of the $n \times n$ matrix A^*A and then set $Q = AR^{-1}$. As mentioned earlier, this article does not address issues of numerical stability.

Here is a theorem summarizing some basic properties.

Theorem 3.1. *Every $[a, b] \times n$ quasimatrix has a QR factorization, which can be calculated by Gram–Schmidt orthogonalization. If the columns of A are linearly independent, the QR factorization is unique. For each k with $1 \le k \le n$, the columns $q_1, \ldots, q_k$ of Q form an orthonormal basis of a space that contains the space spanned by columns $a_1, \ldots, a_k$ of A. The formula (2.3) gives A as a sum of rank 0 or 1 quasimatrices formed from the columns of Q (functions in $C([a, b])$) and rows of R (vectors in $\mathbb{C}^n$).*

Note that as always, the quasimatrix Q of theorem 3.1 is assumed to have columns that are continuous functions defined on $[a, b]$. In the Gram–Schmidt process, the property of continuity is inherited automatically at each step, so long as zero columns are not encountered as a consequence of rank-deficiency. In that case, an arbitrary new function q_k is introduced that is orthogonal to $q_1, \ldots, q_{k-1}$, and we require q_k to be continuous.

Although it is not our emphasis here, one can also define a QR factorization of an $n \times [c, d]$ quasimatrix, that is, a quasimatrix continuous along rows rather than columns. The factorization process requires column pivoting and yields the product $A = QR$, where Q is an $n \times n$ unitary matrix and R is an $n \times [c, d]$ quasimatrix that is upper triangular and diagonally real in a sense to be defined in §5.

---

[4]A remarkable early pair of figures with a quasimatrix flavour can be found at eqn (33) of part 3 and two pages later in [18].
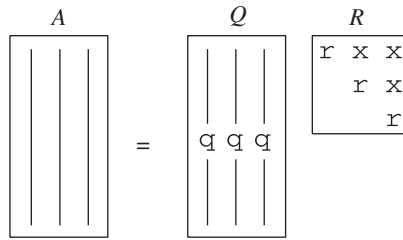
Figure 8. QR factorization of a quasimatrix.

4. Singular value decomposition of a quasimatrix

We define the SVD of a quasimatrix as follows (figure 9).

Definition. Let A be an $[a, b] \times n$ quasimatrix. A *SVD* of A is a factorization $A = USV^*$, where U is an $[a, b] \times n$ quasimatrix with orthonormal columns, S is an $n \times n$ diagonal matrix with diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ and V is an $n \times n$ unitary matrix.

As with QR factorization, it is implicit in this definition that each column of U is a continuous function.

The SVD of a quasimatrix was considered by Battles & Trefethen [6,19]. The following theorem summarizes some of its basic properties, all of which mirror properties of the discrete case. As in §2, $\| \cdot \|$ denotes the Frobenius norm, now defined as in (2.6) but with the sum over i replaced by an integral over y.

Theorem 4.1. *Every $[a, b] \times n$ quasimatrix has an SVD, which can be calculated by computing a QR decomposition $A = QR$ followed by a matrix SVD of the triangular factor, $R = U_1 SV^*$; an SVD of A is then obtained as $(QU_1)SV^*$. The singular values are unique, and the singular vectors corresponding to simple singular values are also unique up to complex signs. The formula (2.4) gives A as a sum of rank 0 or 1 quasimatrices formed from the singular values and vectors. The rank of A is r, the number of non-zero singular values. The columns $u_1, \ldots, u_r$ of $U = QU_1$ form an orthonormal basis for the range of A when regarded as a map from $\mathbb{C}^n$ to $C([a, b])$, and the columns $v_{r+1}, \ldots, v_n$ of V form an orthonormal basis for the nullspace. Moreover, the partial sums A_k defined by (2.5) are best rank k approximations to A, with Frobenius norm errors $\|E_k\| = \|A - A_k\|$ equal to the quantities τ_{k+1} of (2.7).*

Chebfun has included a capability for computing the SVD of a quasimatrix from the beginning in 2003, through the `svd` command. The algorithm used is based on the QR factorization of A, as described in the theorem. (An algorithm based on a continuous analogue of Golub–Kahan bidiagonalization is also possible, though not currently implemented in Chebfun; we hope to discuss this in a later publication.) From these ideas, one can readily define further related notions including the pseudoinverse $VS^{-1}U^*$ (in the full rank case) and the condition number $\kappa(A) = \kappa(S)$ of a quasimatrix, computed by Chebfun commands `pinv` and `cond`.

Following [11, Lecture 4], it is interesting to note the geometrical interpretation of the SVD of a quasimatrix. If A is a quasimatrix of dimensions $[a, b] \times n$, then it can be interpreted as a linear mapping from $\mathbb{C}^n$ to $C([a, b])$. The range of A is a subspace of $C([a, b])$ of dimension at most n, and A maps the unit ball in $\mathbb{C}^n$ (defined with respect to $\| \cdot \|_2$, not $\| \cdot \|$) to a hyperellipsoid in $C([a, b])$, which we may think of as having dimension n if some of the axis lengths are allowed to be zero. The right singular vectors are an orthonormal basis for the unit ball in $\mathbb{C}^n$, the left singular vectors are the semi-axis directions of the hyperellipsoid, and the singular values are the semi-axis lengths.
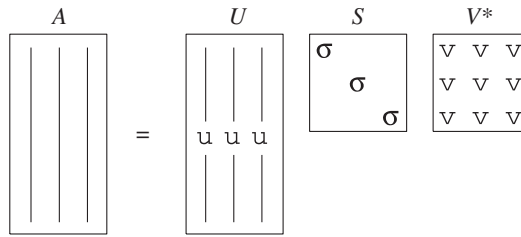
Figure 9. SVD of a quasimatrix. U is a quasimatrix and S and V are ordinary matrices.

5. LU factorization of a quasimatrix

We come now to the first entirely new topic of this article: the generalization of the ideas of pivoting and triangular structure to quasimatrices. Figure 10 shows that we are heading for a factorization $A = LU$, where L is a quasimatrix and U is an upper triangular matrix, but to complete the description we must explain the structure of L.

In §2, LU factorization of matrices was presented without pivoting. Algorithmically, this corresponds to an elimination process in which the first row is used to introduce zeros in the first column, the second row is used to introduce zeros in the second column, and so on. When the row index becomes continuous, however, this approach no longer makes sense. One could take the top of the quasimatrix as a 'first row', but what would be the second row? And so it is that a continuous analogue of row pivoting will be an essential part of our definition of LU factorization of a quasimatrix. (Another term for row pivoting is *partial pivoting*.) When we speak of LU factorization of an $[a, b] \times n$ quasimatrix, row pivoting is always assumed. One could also include column pivoting, but we shall not discuss this variant.

For matrices, the most familiar way to talk about pivoting is in terms of *interchange* of certain rows at each step, leading to a factorization

$$PA = LU, \tag{5.1}$$

where P is a permutation matrix. However, we shall work with a different and mathematically equivalent formulation in terms of *selection* of certain rows at each step, without interchange. In this formulation, we do not move any rows, and there is no permutation matrix P. We get a factorization

$$A = LU, \tag{5.2}$$

but instead of L being lower triangular, it is what Matlab calls *psychologically lower triangular*, meaning that it is a row permutation of a lower triangular matrix.[5]

A choice in our definitions arises at this point. Traditionally in numerical linear algebra, a pivot is chosen corresponding to the *maximal* element in absolute value in a row and/or column, but maximality is not necessary for the factorization to proceed, nor is it always the best choice algorithmically. For example, submaximal pivots may take less work to find than maximal ones and may have advantages for preserving sparsity [21]. In proposing generalized factorizations for quasimatrices and cmatrices, should we use a term like LU factorization for any factorization with a pivot sequence that works (in which case L may take arbitrarily large values off the diagonal), or shall we restrict it to the case where maximal pivots are used (in which case all values off the diagonal are bounded by 1 in absolute value)? In this article, we follow the latter course and insist that pivoting involves maximal values. This makes our factorizations as close as possible to unique and helps us focus on cases where we have the best chance of achieving convergence

---

[5]The term 'psychologically triangular' is not particularly felicitous, but the second author seems to have coined it! He suggested this expression to Matlab inventor Cleve Moler during a conversation years ago, probably during a coffee break at a SIAM meeting.
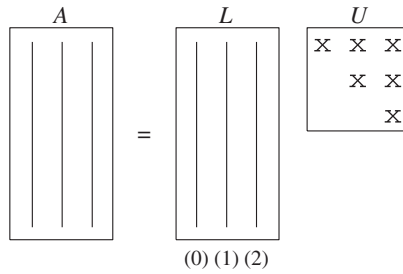
Figure 10. LU factorization of a quasimatrix as a product of a unit lower triangular quasimatrix and an upper triangular matrix. Row pivoting (also known as partial pivoting) is obligatory and is reflected in the digits displayed under *L*, which show the numbers of zeros fixed at nested locations in each column.

theorems for cmatrices. We emphasize that this is only a matter of definitions, however, and one could equally well make the other choice.

The Gaussian elimination process for matrices that leads to (5.2)—assuming that pivots are based on maxima—could be described in the following way. Begin with $E_0 = A$. At step $k = 1$, look in the first column of E_0 to find an index i_1 for which $|E_0(i, 1)|$ is maximal and define $\ell_1 = E_0(\cdot, 1)/E_0(i_1, 1)$, $u_1^* = E_0(i_1, \cdot)$ and $E_1 = E_0 - \ell_1 u_1^*$. (If $E_0(i_1, 1) = 0$, ℓ_1 can be any vector with $|\ell_1(i)| \leq 1$ for all i and $\ell_1(i_1) = 1$.) The new matrix E_1 is zero in row i_1 and column 1. At step $k = 2$, look in the second column of E_1 to find an index i_2 for which $|E_1(i, 2)|$ is maximal and define $\ell_2 = E_1(\cdot, 2)/E_1(i_2, 2)$, $u_2^* = E_1(i_2, \cdot)$ and $E_2 = E_1 - \ell_2 u_2^*$. (If $E_1(i_2, 2) = 0$, ℓ_2 can be any vector with $|\ell_2(i)| \leq 1$ for all i, $\ell_2(i_1) = 0$ and $\ell_2(i_2) = 1$.) The matrix E_2 is now zero in rows i_1 and i_2 and columns 1 and 2. Continuing in this fashion, after n steps, E_n is zero in all n columns, so it is the zero matrix, and we have constructed A as a sum (5.3) of n matrices of rank 0 or 1, just as in (2.1),

$$A = \sum_{k=1}^{n} \ell_k u_k^*. \tag{5.3}$$

Equation (5.2) holds if L is the psychologically lower triangular matrix with columns ℓ_k and U is the upper triangular matrix with rows u_k^*.

Gaussian elimination (with row pivoting) for a quasimatrix. When A is an $[a, b] \times n$ quasimatrix, LU factorization is carried out by the analogous n steps. Begin with $E_0 = A$. At step $k = 1$, find a value $y_1 \in [a, b]$ for which $|E_0(y, 1)|$ is maximal and define $\ell_1 = E_0(\cdot, 1)/E_0(y_1, 1)$, $u_1^* = E_0(y_1, \cdot)$ and $E_1 = E_0 - \ell_1 u_1^*$. (If $E_0(y_1, 1) = 0$, ℓ_1 can be any function in $C([a, b])$ with $|\ell_1(y)| \leq 1$ for all y and $\ell_1(y_1) = 1$.) The new quasimatrix E_1 is zero in row y_1 and column 1. At step $k = 2$, find a value $y_2 \in [a, b]$ for which $|E_1(y, 2)|$ is maximal and define $\ell_2 = E_1(\cdot, 2)/E_1(y_2, 2)$, $u_2^* = E_1(y_2, \cdot)$ and $E_2 = E_1 - \ell_2 u_2^*$. (If $E_1(y_2, 2) = 0$, ℓ_2 can be any function in $C([a, b])$ with $|\ell_2(y)| \leq 1$ for all y, $\ell_2(y_1) = 0$ and $\ell_2(y_2) = 1$.) This quasimatrix E_2 is zero in rows y_1 and y_2 and columns 1 and 2. Continuing in this fashion, after n steps all the columns of E_n are zero, so it is the zero quasimatrix, and we have constructed A as a sum (5.3) of n quasimatrices of rank 0 or 1. Equation (5.2) holds if L and U are constructed analogously as before. The matrix U is the $n \times n$ matrix whose kth row is u_k^*, and it is upper triangular. The quasimatrix L is the $[a, b] \times n$ quasimatrix whose kth column is ℓ_k. Column 2 of L has a zero at y_1, column 3 has zeros at y_1 and y_2, column 4 has zeros at y_1, y_2, y_3 and so on—a nested set of $n - 1$ zeros. This is what the digits marked at the bottom in figure 10 indicate.

This brings us to a crucial set of definitions. These are the ideas that all the novel factorizations of this paper are based upon.

Definitions related to triangular quasimatrices. An $[a, b] \times n$ quasimatrix L together with a specified set of distinct values $y_1, \ldots, y_n \in [a, b]$ is *lower triangular* (we drop the word 'psychologically', though in principle it should be there) if column k has zeros at $y_1, \ldots, y_{k-1}$. The *diagonal* of L is the set of values $\ell_1(y_1), \ldots, \ell_n(y_n)$. If the diagonal values are 1, L is *unit lower triangular*. If each diagonal entry dominates the values in its column in the sense that for each k, $|\ell_k(y)| \leq |\ell_k(y_k)|$ for all $y \in [a, b]$, then L is *diagonally maximal,* or *strictly diagonally*

maximal if the inequality is strict. If L is diagonally maximal and its diagonal values are real and non-negative, it is *diagonally real maximal.* Analogous definitions hold in the transposed case of $n \times [a,b]$ quasimatrices, notably the notion of an upper triangular $n \times [a,b]$ quasimatrix, whose rows have nested zeros in a set of distinct points $x_1, \ldots, x_n$.

With these definitions in place, we can state the definition of the LU factorization of an $[a,b] \times n$ quasimatrix A.

Definition. Let A be an $[a,b] \times n$ quasimatrix. An *LU factorization* of A is a factorization $A = LU$, where U is an upper triangular $n \times n$ matrix and L is an $[a,b] \times n$ unit lower triangular diagonally maximal quasimatrix.

If we did not insist on maximal pivots, the definition would be the same except without the condition that L is diagonally maximal. There is no column pivoting in this discussion, so U need not be diagonally maximal. If one did introduce column pivoting, U would be psychologically upper triangular.

We are not aware of any previous literature on the LU factorization of a quasimatrix, and in Chebfun, an overloaded `lu` command to compute it was only introduced in 2013. The following theorem summarizes the most basic properties.

Theorem 5.1. *Every $[a,b] \times n$ quasimatrix has an LU factorization, which can be computed by quasimatrix Gaussian elimination with row pivoting as described above. If the factor L so computed takes only values strictly less than 1 in absolute value off the diagonal, then the factorization is unique.*

As with the Gram–Schmidt process for computing the quasimatrix QR factorization, as noted after theorem 3.1, Gaussian elimination for computing the quasimatrix LU factorization of theorem 5.1 produces columns of L with the required property of continuity, which is inherited from the continuity of the columns of A.

As at the end of §3, we may note that one can also define an LU factorization of an $n \times [c,d]$ quasimatrix, continuous along rows rather than columns. The factorization requires column instead of row pivoting and yields the product $A = LU$, where L is an $n \times n$ unit lower triangular matrix and U is an $n \times [c,d]$ upper triangular quasimatrix. Now it is U rather than L that is diagonally maximal.

6. QR factorization of a cmatrix

We now turn to our first cmatrix factorization and consequently to our first infinite series as opposed to finite sum. Suppose A is a cmatrix of dimensions $[a,b] \times [c,d]$. As suggested in figure 11, we are going to define a QR factorization of A as a factorization $A = QR$ in which Q is an $[a,b] \times \infty$ quasimatrix and R is an $\infty \times [c,d]$ quasimatrix. Such a product corresponds to an infinite series

$$A = \sum_{j=1}^{\infty} q_j r_j^*. \tag{6.1}$$

with $q_j \in C([a,b])$ and $r_j^* \in C([c,d])$, and to give it a precise meaning, we must specify what kind of convergence of the series is asserted. In this article, all series are required to converge absolutely and uniformly with respect to the variables $(y,x) \in [a,b] \times [c,d]$. Accordingly, we define $\| \cdot \|_\infty$ as the supremum norm of a function over $[a,b] \times [c,d]$. (Note that just as $\| \cdot \|$ in this paper is not the operator 2-norm, $\| \cdot \|_\infty$ is not the operator ∞-norm.) The absolute convergence ensures that we need not worry about the order in which the sum is taken. The uniform convergence implies pointwise convergence too and is consistent with the definitions that q_j, r_j^* and A are all continuous. One could require less than absolute and uniform convergence, but as usual, maximal generality is not our aim.

It is hardly surprising that we are going to require the columns of Q to be orthonormal. In addition, R will be upper triangular, but before explaining this, let us consider what a factorization as in figure 11 would amount to if R were not required to have triangular structure. An example
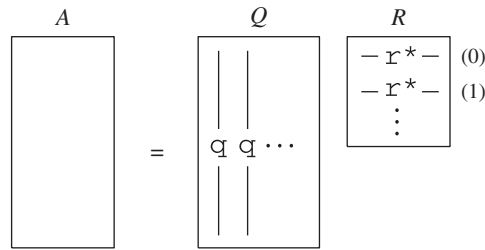
Figure 11. QR factorization of a cmatrix as a product of a column quasimatrix with infinitely many orthonormal columns and an upper triangular row quasimatrix with infinitely many rows. Column pivoting is obligatory and is reflected in the digits displayed on the right of R, which show the numbers of zeros fixed at nested locations in each row. The series implicit in the product QR is assumed to converge absolutely and uniformly.

to bear in mind would be a case in which we began with an $[a, b] \times [c, d]$ cmatrix A and then computed a Fourier series for each x with respect to the 'vertical' variable $y \in [a, b]$. This would give us an $[a, b] \times \infty$ quasimatrix Q with orthonormal columns $Q(\cdot, j)$, $j = 1, 2, \ldots$, corresponding to different Fourier modes on $[a, b]$. The factor $R = Q^*A$ would be the $\infty \times [c, d]$ quasimatrix whose jth row $r^*(j, \cdot)$ would be the function containing the jth Fourier coefficients, depending continuously on x. If A were Lipschitz continuous, say, the series would converge absolutely and uniformly, as required by our definitions. In no sense would R have triangular structure.

Our aim, however, is not Fourier series but QR factorization. The signal property of QR factorization is nesting of column spaces, as described in theorem 3.1 for the quasimatrix case: for each k, the first k columns of Q must form a basis of a space that contains the 'first k columns' of A. When A is a cmatrix, to make sense of the idea of its first k columns, we will have to introduce column pivoting. As with the LU factorization of §5, we shall restrict our attention to pivots based on maximality, giving a correspondingly narrow definition of the factorization. And thus we are led to the following algorithm for computing the QR factorization of a cmatrix, which corresponds to what the matrix computations literature calls 'modified' Gram–Schmidt orthogonalization with column pivoting.

Modified Gram–Schmidt orthogonalization (with column pivoting) for a cmatrix. Let A be an $[a, b] \times [c, d]$ cmatrix and set $E_0 = A$. At step $k = 1$, find a value $x_1 \in [c, d]$ for which $\|E_0(\cdot, x)\|$ is maximal, define $q_1 = E_0(\cdot, x_1)/\|E_0(\cdot, x_1)\|$ and $r_1^*(x) = q_1^*E_0(\cdot, x)$, and set $E_1 = E_0 - q_1r_1^*$. Each column of the new cmatrix E_1 is orthogonal to q_1. As mentioned at the beginning of §3, orthonormality and the norm $\|\cdot\|$ for functions in $C([a, b])$ are defined by the standard L^2 inner product. At step $k = 2$, find a value $x_2 \in [c, d]$ for which $\|E_1(\cdot, x)\|$ is maximal, define $q_2 = E_1(\cdot, x_2)/\|E_1(\cdot, x_2)\|$ and $r_2^*(x) = q_2^*E_1(\cdot, x)$, and set $E_2 = E_1 - q_2r_2^*$. Each column of E_2 is now orthogonal to both q_1 and q_2. Continuing in this fashion, we construct a series corresponding to a factorization $A = QR$; the update equation is

$$E_{k+1} = E_k - q_{k+1}q_{k+1}^*E_k, \quad k = 0, 1, 2, \ldots . \tag{6.2}$$

If A has infinite rank, the process goes forever as described. If A has finite rank r, then E_r will become zero at step r, and in subsequent steps one may choose new points x_j arbitrarily together with arbitrary continuous orthonormal vectors q_k and rows r_k^* identically equal to zero.

This algorithm of cmatrix modified Gram–Schmidt orthogonalization with column pivoting produces a quasimatrix R that is upper triangular according to our definitions. Specifically, the sequence of distinct numbers $x_1, x_2, \ldots$ has the property that row 2 of R has a zero at x_1, row 3 of R has zeros at x_1 and x_2, and so on. Moreover, R is diagonally real maximal.

We can now state the general definition. (Recall that a diagonally real maximal quasimatrix was defined in §5.)

Definition. Let A be an $[a, b] \times [c, d]$ cmatrix. A *QR factorization* of A is a factorization $A = QR$, where Q is an $[a, b] \times \infty$ quasimatrix with orthonormal columns and R is an upper triangular diagonally real maximal $\infty \times [c, d]$ quasimatrix.

We are not aware of any previous literature on QR factorization of a cmatrix. The qr command of Chebfun2 constructs the factorization from the LU factorization (up to a finite precision of 16 digits), to be described in §8.

It is easily seen that the algorithm we have described produces quasimatrices Q and R with the required continuous columns. What is not clear whether the series represented by the product QR converges absolutely and uniformly to A. This brings us to our first substantial point of analysis. What smoothness conditions on A ensure that the quasimatrices Q and R that we have constructed correspond to a QR factorization $A = QR$? One might guess that a relatively mild smoothness condition on A might be enough, but we do not know if this is true or not. Here is what we can prove. Given a number $\rho > 1$, the *Bernstein ρ-ellipse* is the region in the complex plane bounded by the ellipse with foci ± 1 and semi-axis lengths summing to ρ. The *Bernstein ρ-ellipse scaled to $[c, d]$* is the region bounded by the ellipse with foci c and d and semi-axis lengths summing to $\rho(d - c)/2$.

Theorem 6.1. *Let A be an $[a, b] \times [c, d]$ cmatrix. Suppose that $A(\cdot, x)$ is a Lipschitz continuous function on $[a, b]$, uniformly with respect to $x \in [c, d]$. Moreover, suppose there is a constant $M > 0$ such that for each $y \in [a, b]$, the row function $A(y, \cdot)$ can be extended to a function in the complex x-plane that is analytic and bounded in absolute value by M throughout a neighbourhood Ω (independent of y) of the Bernstein $2\sqrt{2}\rho$-ellipse scaled to $[c, d]$ for some $\rho > 1$. Then the series constructed by QR factorization (with column pivoting) converges absolutely and uniformly at the rate $\|E_k\|_\infty = O(\rho^{-2k/3})$, giving a QR factorization $A = QR$. If the factor R so computed is strictly diagonally maximal, then the factorization is unique.*

Note that this theorem requires less smoothness in y than in x. So far as we know this distinction may be genuine, reflecting the fact that a QR factorization takes norms with respect to y while working with individual columns indexed by x.

Together with theorems 8.1 and 9.1, theorem 6.1 is one of the three main theorems of this paper. Proofs of those two theorems are given in §§8 and 9. The three results have different smoothness assumptions, but there is some overlap in the proofs. A proof of theorem 6.1, which has elements in common with both of the others, can be found in Section 4.9 of the first author's PhD thesis [22].

7. Singular value decomposition of a cmatrix

Following figure 12, we define the SVD of a cmatrix as follows.

Definition. Let A be an $[a, b] \times [c, d]$ cmatrix. A *SVD* of A is a factorization $A = USV^*$, where U is an $[a, b] \times \infty$ quasimatrix with orthonormal columns, S is an $\infty \times \infty$ diagonal matrix with diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ and V is an $[c, d] \times \infty$ quasimatrix with orthonormal columns.

This corresponds to a series

$$A = \sum_{j=1}^{\infty} \sigma_j u_j v_j^*, \tag{7.1}$$

which as usual is required to converge absolutely and uniformly. Just as the SVD of a quasimatrix A can be computed from the quasimatrix QR decomposition $A = QR$ followed by the matrix SVD $R = USV^*$, the SVD of a cmatrix A could be computed from the cmatrix QR decomposition $A = QR$ followed by the quasimatrix SVD $R = USV^*$ (the transpose of the SVD described in §4), at least up to a certain accuracy if R is truncated to a finite number of rows.

Although our notation is new, the mathematics of the SVD of a cmatrix goes back a century, beginning with the work of Schmidt [3,5]. In particular, it is known that a small amount of smoothness suffices to make the SVD series converge pointwise, absolutely and uniformly. To explain this effect, consider the classical approximation theory problem of a continuous function f defined on the interval $[-1, 1]$. It is known that the Chebyshev series of f, which expands f in Chebyshev polynomials, converges absolutely and uniformly if f is Lipschitz continuous. If, for some $\nu \geq 1$, f has a νth derivative of bounded variation, the ∞-norm of the error of the degree k partial sum of the Chebyshev series is $O(k^{-\nu})$. Similarly if, for some $\rho > 1$, f is analytic and bounded in the region bounded by the Bernstein ρ-ellipse, the ∞-norm of the error is $O(\rho^{-k})$.

Figure 12. SVD of a cmatrix. This infinite series has a long history going back to Schmidt [3].

(See Theorems 3.1, 7.2 and 8.2 of [23].) It follows that if a cmatrix A has such smoothness with respect to either variable x or y, its rank k approximation errors $\{\tau_k\}$ of (2.7), hence likewise its singular values $\{\sigma_k\}$, must converge at the same rate. (For the analytic case, such an argument was possibly first published by Little & Reade [24].) The following theorem summarizes some of this information. The existence and uniqueness statement is standard and is just included for completeness. As always, $\|\cdot\|$ is the continuous analogue of the Frobenius norm (2.6).

Theorem 7.1. *Let A be an $[a,b] \times [c,d]$ cmatrix that is uniformly Lipschitz continuous with respect to x and y. Then an SVD of A exists, the singular values are unique, with $\sigma_k \to 0$ and $\tau_k \to 0$ as $k \to \infty$, the singular vectors corresponding to simple singular values are unique up to complex signs, and the series (7.1) converges absolutely and uniformly to A. Moreover, the partial sums A_k defined by (2.5) are best rank k approximations to A, with Frobenius norm errors $\|E_k\| = \|A - A_k\|$ equal to the numbers τ_{k+1} of (2.7). If, for some $\nu \geq 1$, the functions $A(\cdot, x)$ have a νth derivative of variation uniformly bounded with respect to x, or if the corresponding assumption holds with the roles of x and y interchanged, then the singular values and approximation errors satisfy $\sigma_k, \tau_k = O(k^{-\nu})$. If, for some $\rho > 1$, the functions $A(\cdot, x)$ can be extended in the complex y-plane to analytic functions in the Bernstein ρ-ellipse scaled to $[a,b]$ uniformly bounded with respect to x, or if the corresponding assumption holds with the roles of x and y interchanged, then the singular values and approximation errors satisfy $\sigma_k, \tau_k = O(\rho^{-k})$.*

Proof. The existence and uniqueness of the SVD series is due to Schmidt in 1907 [3], but his analysis does not fully meet our needs since he assumed only that A is continuous, in which case the singular functions need not converge absolutely or uniformly (or indeed pointwise). The situation where A has some smoothness was addressed by Hammerstein in 1923, who proved uniform convergence of (7.1) under an assumption that is implied by Lipschitz continuity [25], and Smithies in 1938, who proved absolute convergence under a weaker assumption essentially of Hölder continuity with exponent more than $\frac{1}{2}$ [26, theorem 14]. These results establish the existence and uniqueness claims of this theorem. The rank k approximation property is due to Schmidt [3]; see also [16]. The proofs of the $O(k^{-\nu})$ and $O(\rho^{-k})$ results were outlined above. ∎

If A is a non-negative definite Hermitian cmatrix, whose Cholesky factorization we shall consider in §9, then the SVD is known to exist without the extra assumption of Lipschitz continuity (i.e. continuity of A is enough to ensure continuity and absolute and uniform convergence of its finite-rank approximations). This is *Mercer's theorem* [4].

Chebfun2 has an `svd` command that computes the SVD of a cmatrix down to the usual Chebfun accuracy of about 16 digits. The algorithm uses a cmatrix LU factorization (§8) and two quasimatrix QR factorizations to reduce the problem to a matrix SVD.

8. LU factorization of a cmatrix

Our final two factorizations involve both row and column pivoting, with triangular quasimatrices on both sides. Both are implemented in Chebfun2. In fact, the basic method by which Chebfun2 represents a function $f(x,y)$ is cmatrix LU decomposition, which was the starting motivation for us to write this article, and we shall say more about this application at the end of this section.

To apply Gaussian elimination to a cmatrix, as it is continuous in both directions, we need to pick both a row and a column with which to eliminate. Various strategies for such choices have been applied in the literature of low-rank matrix approximations, including a method with a quasi-optimality property based on volume maximization; some references to such methods can be found in [22,27]. The method we shall consider is the one that follows most directly from the classical matrix computations literature, where a pivot row and column are chosen based on the maximal entry in the cmatrix: the traditional term is *complete pivoting*. The ingredients have appeared in the earlier factorizations, so we can go directly to the definition, as depicted schematically in figure 13.

Definition. Let A be an $[a, b] \times [c, d]$ cmatrix. An *LU factorization* of A is a factorization $A = LU$, where L is an $[a, b] \times \infty$ unit lower triangular diagonally maximal quasimatrix and U is an upper triangular diagonally maximal $\infty \times [c, d]$ quasimatrix.

We can describe the algorithm as follows.

Gaussian elimination (*with row and column pivoting*) *for a cmatrix.* Let A be an $[a, b] \times [c, d]$ cmatrix, and begin with $E_0 = A$. At step $k = 1$, find a pair $(y_1, x_1) \in [a, b] \times [c, d]$ for which $|E_0(y, x)|$ is maximal and define $\ell_1 = E_0(\cdot, x_1)/E_0(y_1, x_1)$, $u_1^* = E_0(y_1, \cdot)$ and $E_1 = E_0 - \ell_1 u_1^*$. (If $E_0(y_1, x_1) = 0$, A is the zero cmatrix and ℓ_1 can be any function in $C([a, b])$ with $|\ell_1(y)| \leq 1$ for all y and $\ell_1(y_1) = 1$; u_1^* will necessarily be zero.) The new cmatrix E_1 is zero in row y_1 and column x_1. At step $k = 2$, find a pair $(y_2, x_2) \in [a, b] \times [c, d]$ for which $|E_1(y, x)|$ is maximal and define $\ell_2 = E_1(\cdot, x_2)/E_1(y_2, x_2)$, $u_2^* = E_1(y_2, \cdot)$, and $E_2 = E_1 - \ell_2 u_2^*$. (If $E_1(y_2, x_2) = 0$, E_1 is the zero cmatrix and ℓ_2 can be any function in $C([a, b])$ with $|\ell_2(y)| \leq 1$ for all y, $\ell_2(y_1) = 0$ and $\ell_2(y_2) = 1$; u_2^* will necessarily be zero.) This cmatrix E_2 is zero in rows y_1 and y_2 and columns x_1 and x_2. We continue in this fashion, generating the LU decomposition (5.2) step by step; the update equation is

$$E_{k+1} = E_k - \ell_{k+1} u_{k+1}^*, \quad k = 0, 1, 2, \ldots. \tag{8.1}$$

The quasimatrix U is the $\infty \times [c, d]$ quasimatrix whose kth row is u_k^*, and it is upper triangular and diagonally maximal with pivot sequence $x_1, x_2, \ldots$. The quasimatrix L is the $[a, b] \times n$ quasimatrix whose kth column is ℓ_k, and it is unit lower triangular and diagonally maximal with pivot sequence $y_1, y_2, \ldots$.

When does the series constructed by Gaussian elimination converge, so we can truly write $A = LU$? As with theorem 6.1, one might guess that a relatively mild smoothness condition on A should suffice, but we do not know if this is true. Note that in the following theorem, A has to be smooth with respect to either x or y; it need not be smooth in both variables.

Theorem 8.1. *Let A be an $[a, b] \times [c, d]$ cmatrix. Suppose there is a constant $M > 0$ such that for each $x \in [c, d]$, the function $A(\cdot, x)$ can be extended to a function in the complex y-plane that is analytic and bounded in absolute value by M throughout a neighbourhood Ω (independent of x) of the closed region K consisting of all points at distance $\leq 2\rho(b - a)$ from $[a, b]$ for some $\rho > 1$ (or analogously with the roles of y and x reversed and also the roles of (a, b) and (c, d)). Then the series constructed by Gaussian elimination converges absolutely and uniformly at the rate $\|E_k\|_\infty = O(\rho^{-k})$, giving an LU factorization $A = LU$. If the factors L and U are strictly diagonally maximal, then the factorization is unique.*

Proof. Fix $x \in [c, d]$, and for each step k, let e_k denote the error function at step k,

$$e_k = A(\cdot, x) - \sum_{j=1}^{k} \ell_j(\cdot) u_j^*(x), \tag{8.2}$$

a function of $y \in [a, b]$. The elimination process is such that e_k is also analytic in K, with magnitude at worst doubling at each step,

$$|e_k(y)| \leq 2^k M, \quad y \in K. \tag{8.3}$$

(Note that in this formula y is taking complex values.) Because of the elimination, e_k has at least k zeros $y_1, \ldots, y_k$ in $[a, b]$. Let p_k be the polynomial $(y - y_1) \cdots (y - y_k)$. Then, e_k/p_k is analytic in K,
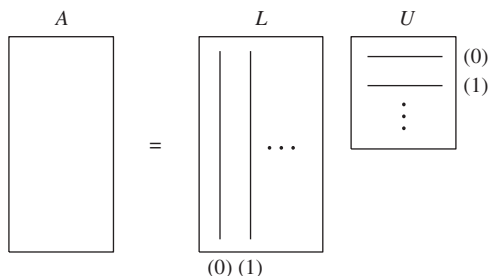
Figure 13. LU factorization of a cmatrix. Now both row and column pivoting are obligatory, and both *L* and *U* have triangular structure. Again *L* has unit diagonal entries. As in figures 10 and 11, the digits displayed at the margins indicate the numbers of zeros fixed at nested locations in the columns of *L* and rows of *U*.

hence satisfies the maximum modulus principle within K. For any $y \in [a, b]$, this implies

$$|e_k(y)| \le |p_k(y)| \sup_{s \in \partial K} \frac{|e_k(s)|}{|p_k(s)|} \le 2^k M \sup_{s \in \partial K} \frac{|p_k(y)|}{|p_k(s)|}. \tag{8.4}$$

In this quotient of polynomials, each of the k factors in the denominator is at least 2ρ times bigger in modulus than the corresponding factor in the numerator. We conclude that $|e_k(y)| \le \rho^{-k} M$ for $y \in [a, b]$. As this error estimate applies independently of x and y, it establishes uniform convergence. It also implies that the next term $\ell_{k+1} u_{k+1}^*$ in the series is bounded in absolute value by $\rho^{-k} M$, which implies absolute convergence since $\sum_{k=0}^{\infty} \rho^{-k} M < \infty$. ∎

We now comment on how Chebfun2 uses cmatrix LU factorization to represent functions on rectangles.

The LU factorization of a cmatrix is an infinite series, and if the cmatrix is somewhat smooth, one may expect the series to converge at a good rate. The principle of Chebfun2 is that functions are represented to approximately 16-digit accuracy by finite-rank representations whose ranks adjust as necessary to achieve this accuracy. For functions defined on a two-dimensional rectangle, the representation chosen is a finite section of the cmatrix LU factorization

$$A \approx \tilde{A} = \sum_{j=1}^{k} \ell_j u_j^*. \tag{8.5}$$

For a typical function A arising in our Chebfun2 explorations, k might be 10 or 100. In principle, Chebfun2 follows precisely the algorithm of cmatrix Gaussian elimination (thus using both row and column pivoting) to find this approximation, though in practice, grid-based approximations are employed to diminish the work that would be involved in computing the true global extremum of $|E(y, x)|$ at each step. For details of the algorithms and numerical examples, see [8,22].

The representation (8.5) is based on one-dimensional functions ℓ_j of y and u_j^* of x. In Chebfun2, these are represented as standard Chebfun objects, i.e. global polynomial interpolants through data in a Chebyshev grid in the interval $[a, b]$ or $[c, d]$ that is adaptively determined for 16-digit accuracy. Thus in Chebfun2 calculations, the philosophy of floating point arithmetic is replicated at three levels:

— *numbers* are represented by binary expansions of fixed length 64 bits;
— *one-dimensional functions* are represented by polynomial interpolants of adaptively determined degree; and
— *two-dimensional functions* are represented by LU approximations (8.5) of adaptively determined rank.

The Chebfun2 technology is closely related to the low-rank matrix approximations (often hierarchical, though not hierarchical in Chebfun2) developed over the years by many authors
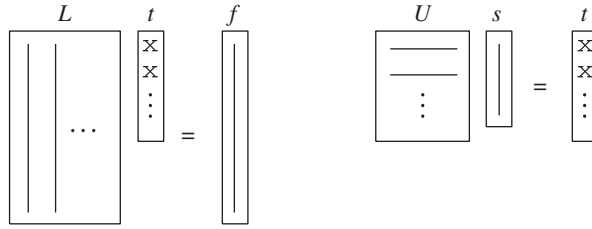
Figure 14. The cmatrix analogue (an integral equation) of the familiar matrix technique of solving $As = f$ via two problems $Lt = f$ and $Us = t$. First, an $\infty \times 1$ vector t is constructed element-by-element by enforcing discrete conditions at the diagonal points $y_1, y_2, \ldots$ of L. Then, a sequence of values is computed of an $[c, d] \times 1$ function $s \in C([c, d])$ at the sample points $x_1, x_2, \ldots$, the diagonal points of U. If these diagonal points are dense in $[c, d]$ and the sample values behave appropriately, a candidate solution $s \in C([c, d])$ is determined.

including Bebendorf, Drineas, Goreinov, Maday, Mahoney, Martinsson, Oseledets, Rokhlin, Savostyanov, Tyrtyshnikov and Zamarashkin. Entries into this literature can be found in [8,27,28]. A 2000 paper of Bebendorf is particularly close to our work [29]. For applications to functions rather than matrices, besides Bebendorf, another set of predecessors are Geddes and his students [30–32]. We are not aware of previous explicit discussions of the LU factorization of a cmatrix.

Thus in Chebfun2, every function starts from an LU factorization of finite rank, so that cmatrices are reduced to finite-dimensional quasimatrices. On this foundation, exploiting the finite rank, algorithms are built for further operations including QR and Cholesky factorization, SVD, integration, differentiation, vector calculus and application of integral operators.

For matrices, the basic application of LU factorization is solution of a system of equations $As = f$, that is, $LUs = f$ (we avoid using the usual letters x and b as they have other meanings in our discussion). If we set $t = Us$, this reduces to the two triangular problems

$$Lt = f \quad \text{and} \quad Us = t, \tag{8.6}$$

which can be solved successively first for t and then for s. Figure 14 sketches what the analogous sequence looks like for the integral equation defined by an $[a, b] \times [c, d]$ cmatrix A and a right-hand side $f \in C([a, b])$. Whether and when this process converges to a solution $s \in C([c, d])$ is a question of analysis that we shall not consider.

9. Cholesky factorization of a cmatrix

For our final factorization, suppose A is a Hermitian cmatrix of dimensions $[a, b] \times [a, b]$. We can always consider its LU factorization as described in §8. However, it is natural to wish to preserve the symmetry, and this is where the idea of a Cholesky factorization applies. Here is the definition, shown schematically in figure 15. Using different language ('Geddes series'), Cholesky factorizations of cmatrices have been studied by Geddes and his students [30–32].

Definition. Let A be an $[a, b] \times [a, b]$ square cmatrix. A *Cholesky factorization* of A is a factorization $A = R^*R$, where R is an $\infty \times [a, b]$ diagonally real maximal upper triangular quasimatrix.

Suppose A is a cmatrix with a Cholesky factorization $A = R^*R$. Then A is Hermitian, and for any $u \in C([a, b])$, we may compute

$$u^*Au = u^*R^*Ru = (Ru)^*(Ru) \geq 0. \tag{9.1}$$

A Hermitian cmatrix A which satisfies this inequality for every $u \in C([a, b])$ is said to be *non-negative definite.* We have just shown that if A has a Cholesky factorization, it is non-negative definite. We would like to be able to say that the converse holds too, so that a Hermitian cmatrix
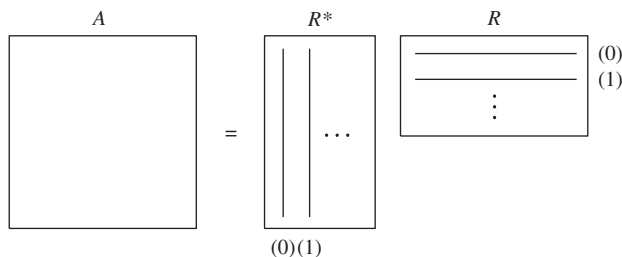
Figure 15. Cholesky factorization of a Hermitian non-negative definite cmatrix.

has a Cholesky factorization if and only if it is non-negative definite. Below we shall prove that this is true if A is sufficiently smooth.

The lower- and upper triangular factors of a Cholesky factorization are conjugate transposes of one another, and in particular, they have the same pivoting sequence. Consequently, to select a pivot, the Cholesky algorithm searches only along the diagonal. We can describe the algorithm as follows.

Cholesky algorithm (with pivoting) for a Hermitian cmatrix. Let A be a Hermitian $[a, b] \times [a, b]$ cmatrix, and begin with $E_0 = A$. At step $k = 1$, find a value x_1 for which $E_0(x, x)$ is maximal. (The diagonal entries are necessarily real since E_0 is Hermitian.) If $E_0(x_1, x_1) < 0$, A is not non-negative definite; the algorithm terminates. Otherwise, let γ_1 be the non-negative real square root of $E_0(x_1, x_1)$ and define $r_1 = E_0(\cdot, x_1)/\gamma_1$ and $E_1 = E_0 - r_1 r_1^*$. (If $E_0(x_1, x_1) = 0$, A is the zero cmatrix and we take r_1 to be the zero function in $C([a, b])$.) The new cmatrix E_1 is zero in row x_1 and column x_1. At step $k = 2$, find a value $(x_2, x_2) \in [a, b] \times [a, b]$ for which $E_1(x, x)$ is maximal. If $E_1(x_2, x_2) < 0$, A is not non-negative definite; the algorithm terminates. Otherwise, let γ_2 be the non-negative real square root of $E_1(x_2, x_2)$ and define $r_2 = E_1(\cdot, x_2)/\gamma_2$ and $E_2 = E_1 - r_2 r_2^*$. This cmatrix E_2 is zero in rows and columns x_1 and x_2. We continue in this fashion, generating the Cholesky factorization step by step, with the update equation

$$E_{k+1} = E_k - r_{k+1} r_{k+1}^*, \quad k = 0, 1, 2, \ldots . \tag{9.2}$$

The quasimatrix R is the $\infty \times [a, b]$ quasimatrix whose kth row is r_k^*, and it is upper triangular and diagonally real maximal with pivot sequence $x_1, x_2, \ldots$.

We now turn to a theorem about Cholesky factorization of a cmatrix. This is a special case of LU factorization, and theorem 8.1 could be applied here again. However, the slightly stronger result below can be proved by a continuous analogue of an argument given by Harbrecht *et al.* [33]. Like theorem 8.1, this theorem requires A to be analytic in a sizeable region in the complex plane with respect to one or the other of its arguments, but the region is smaller than before. A convergence result with this flavour for symmetric functions was announced in a talk by Geddes [32] and attributed to himself and Chapman, with the less explicit hypothesis that the region of analyticity is 'sufficiently large'. It appears that there is no publication giving details of the Chapman–Geddes result.

Theorem 9.1. *Let A be an $[a, b] \times [a, b]$ Hermitian cmatrix. Suppose there is a constant $M > 0$ such that for each $x \in [a, b]$, the function $A(\cdot, x)$ can be extended to a function in the complex y-plane that is analytic and bounded in absolute value by M throughout a neighbourhood Ω (independent of x) of the closed Bernstein 4ρ-ellipse scaled to $[a, b]$ for some $\rho > 1$ (or analogously with the roles of y and x reversed). Then if A is non-negative definite, the series constructed by the Cholesky algorithm converges absolutely and uniformly at the rate $\|E_k\|_\infty = O(\rho^{-k})$, giving a Cholesky factorization $A = R^*R$. If A is not non-negative definite, the Cholesky algorithm breaks down with the square root of a negative number.*

Proof. It is readily seen that if A is non-negative definite, then this property is preserved by steps of the Cholesky algorithm; thus if the algorithm breaks down, A must not be non-negative definite. Conversely, if the algorithm does not break down, then as we are about to show, it yields

a Cholesky factorization $A = R^*R$, and as shown in (9.1), the existence of such a factorization implies that A is non-negative definite. From here on, accordingly, we assume A is non-negative definite.

Take k steps of the Cholesky algorithm. This yields a $k \times [a, b]$ upper triangular quasimatrix R_k and a corresponding rank k approximation $A_k = R_k^*R_k$ to A with error $E_k = A - A_k$. If $E_k = 0$, the algorithm has successfully produced a Cholesky factorization in the form of a finite sum, and the assertions of the theorem hold trivially. Assume then $E_k \neq 0$. Let the diagonal entries of R, which are necessarily real and positive and non-increasing, be denoted by $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_k > 0$. Because of the pivoting in the Cholesky algorithm, we have

$$\|E_k\|_\infty \leq \gamma_{k+1}^2 \leq \gamma_k^2. \tag{9.3}$$

Now R_k is a $k \times [a, b]$ quasimatrix, but it contains within it the $k \times k$ matrix $\tilde{R}_k$ of entries from columns $1, \ldots, k$, and this matrix is psychologically upper triangular and diagonally real maximal, with diagonal entries $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_k > 0$. It is readily seen that the entries on the jth diagonal of the inverse of a unit triangular matrix are bounded in absolute value by 2^j. Similarly, for a triangular matrix with minimal diagonal entry γ_k, they are bounded in absolute value by $2^j/\gamma_k$. By regarding $\tilde{R}_k^{-1}$ as the sum of its diagonals, it follows that

$$\|\tilde{R}_k^{-1}\|_2 \leq \frac{2^k}{\gamma_k},$$

where $\| \cdot \|_2$ is the matrix 2-norm. (A more precise estimate is given in theorem 6.1 and the remark that follows it in [34], with discussion of the relevant literature.) This implies

$$\|\tilde{A}_k^{-1}\|_2 \leq \frac{4^k}{\gamma_k^2},$$

where $\tilde{A}_k = \tilde{R}_k^*\tilde{R}_k$ is the $k \times k$ Hermitian positive definite submatrix of A extracted from its rows and columns $1, \ldots, k$. Another way to say this is that the kth singular value of $\tilde{A}_k$ (i.e. kth eigenvalue in absolute value) satisfies

$$\sigma_k(\tilde{A}_k) \geq \frac{\gamma_k^2}{4^k}.$$

Thus by combining with (9.3), we get

$$\|E_k\|_\infty \leq 4^k \sigma_k(\tilde{A}_k) = 4^k \inf \|\tilde{A}_k - C_{k-1}\|_2,$$

where the infimum is over $k \times k$ matrices C_{k-1} of rank $k - 1$, or by switching from the 2-norm of $\tilde{A}_k - C_{k-1}$ to its maximum entry,

$$\|E_k\|_\infty \leq k4^k \inf \max_{i,j} |(\tilde{A}_k - C_{k-1})_{i,j}|. \tag{9.4}$$

Now by the observations in the paragraph above theorem 7.1, our analyticity assumption implies that A can be approximated to accuracy $O((4\rho)^{-k})$ by functions that are polynomials of degree $k - 2$ with respect to one of the variables, and indeed to accuracy $O((4\rho + \varepsilon)^{-k})$ for some sufficiently small $\varepsilon > 0$ since we assumed analyticity in a neighbourhood Ω of the Bernstein ellipse. When such a function is sampled on a $k \times k$ grid, the resulting matrix is of rank at most $k - 1$. Combining this observation with (9.4) completes the proof. ∎

Theorem 9.1, like theorems 6.1 and 8.1, makes a very strong smoothness assumption. Experience with Chebfun2 shows that in practice, QR, LU and Cholesky factorizations all proceed without difficulty for cmatrices that have just a minimal degree of smoothness. We do not know whether it can be proved that this must always be the case.

Chebfun2 does not compute Cholesky factorizations in the manner we have described in this section. Instead, its `chol` command starts from the cmatrix LU factorization already computed when a chebfun2 is first realized (of finite rank, accurate to 16 digits), and the Cholesky factors are

then obtained by appropriately rescaling the columns of *L* and rows of *U*. Just as with matrices, `chol` applied to Hermitian cmatrices proves a highly practical way of testing for non-negative definiteness (up to 16-digit precision).

10. Conclusion

In closing, we would like to emphasize that this article is intended as a contribution to conceptual and mathematical foundations. Our work sprang from the Chebfun2 software project, as described especially in §8, and it has connections with low-rank matrix approximation algorithms developed by many authors in many applications, but the new content of this paper is theoretical. We have proposed concepts of matrix factorizations of quasimatrices and cmatrices that we hope others will find useful and established the first theorems (theorems 6.1, 8.1 and 9.1) asserting that the QR, LU and Cholesky cmatrix factorizations exist.

References

1. Fredholm I. 1903 Sur une classe d'équations fonctionnelles. *Acta Math.* **27**, 365–390. (doi:10.1007/BF02421317)
2. Hilbert D. 1904 Grundzüge einer allgemeinen Theorie der Integralgleichungen. *Nachr. Königl. Ges. Gött.* 49–91.
3. Schmidt E. 1907 Zur Theorie der linearen und nichtlinearen Integralgleichungen. I. Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener. *Math. Ann.* **63**, 433–476. (doi:10.1007/BF01449770)
4. Mercer J. 1909 Functions of positive and negative type and their connection with the theory of integral equations. *Phil. Trans. R. Soc. Lond. A* **209**, 415–446. (doi:10.1098/rsta.1909.0016)
5. Stewart GW. 2011 Fredholm, Hilbert, Schmidt: three fundamental papers on integral equations. See www.cs.umd.edu/~stewart/FHS.pdf.
6. Battles Z, Trefethen LN. 2004 An extension of MATLAB to continuous functions and operators. *SIAM J. Sci. Comput.* **25**, 1743–1770. (doi:10.1137/S1064827503430126)
7. Driscoll TA, Hale N, Trefethen LN (eds). 2014 *Chebfun guide*. Oxford, UK: Pafnuty Publications. See http://www.chebfun.org.
8. Townsend A, Trefethen LN. 2013 An extension of Chebfun to two dimensions. *SIAM J. Sci. Comput.* **35**, C495–C518. (doi:10.1137/130908002)
9. de Boor C. 1991 An alternative approach to (the teaching of) rank, basis, and dimension. *Lin. Alg. Appl.* **146**, 221–229. (doi:10.1016/0024-3795(91)90026-S)
10. Stewart GW. 1998 *Afternotes goes to graduate school: lectures on advanced numerical analysis.* Philadelphia, PA: SIAM.
11. Trefethen LN, Bau III, D. 1997 *Numerical linear algebra.* Philadelphia, PA: SIAM.
12. Golub GH, Van Loan CF. 2013 *Matrix computations*, 4th edn. Baltimore, MD: Johns Hopkins.
13. Stewart GW. 1998 *Matrix algorithms, v. 1: basic decompositions.* Philadelphia, PA: SIAM.
14. Chu MT, Funderlic RE, Golub GH. 1995 A rank-one reduction formula and its applications to matrix factorizations. *SIAM Rev.* **37**, 512–530. (doi:10.1137/1037124)
15. Stewart GW. 1993 On the early history of the singular value decomposition. *SIAM Rev.* **35**, 551–566. (doi:10.1137/1035134)
16. Weyl H. 1912 Das asymptotische Verteilungsgesetz der Eigenwerte linearer partieller Differentialgleichungen (mit einer Anwendung auf die Theorie der Hohlraumstrahlung). *Math. Ann.* **71**, 441–479. (doi:10.1007/BF01456804)

17. Eckart C, Young G. 1936 The approximation of one matrix by another of lower rank. *Psychometrika* **1**, 211–218. (doi:10.1007/BF02288367)
18. Born M, Heisenberg W, Jordan C. 1926 Zur Quantenmechanik. II. *Zeit. Phys.* **35**, 557–615. (doi:10.1007/BF01379806)
19. Battles Z. 2005 Numerical linear algebra for continuous functions. DPhil. thesis, Oxford University Computing Lab., Oxford, UK.
20. Trefethen LN. 2010 Householder triangularization of a quasimatrix. *IMA J. Numer. Anal.* **30**, 887–897. (doi:10.1093/imanum/drp018)
21. Davis TA. 2006 *Direct methods for sparse linear systems.* Philadelphia, PA: SIAM.
22. Townsend A. 2014 Computing with functions in two dimensions. D.Phil. thesis, Mathematical Institute, University of Oxford, Oxford, UK. See http://math.mit.edu/~ajt/.
23. Trefethen LN. 2013 *Approximation theory and approximation practice.* Philadelphia, PA: SIAM.
24. Little G, Reade JB. 1984 Eigenvalues of analytic kernels. *SIAM J. Math. Anal.* **15**, 133–136. (doi:10.1137/0515009)
25. Hammerstein A. 1923 Über die Entwickelung des Kernes linearer Integralgleichungen nach Eigenfunktionen. *Sitzungsberichte Preuss. Akad. Wiss.*, 181–184.
26. Smithies F. 1938 The eigen-values and singular values of integral equations. *Proc. Lond. Math. Soc.* **2**, 255–279. (doi:10.1112/plms/s2-43.4.255)
27. Bebendorf M. 2008 *Hierarchical matrices: a means to efficiently solve elliptic boundary value problems*. New York, NY: Springer.
28. Hackbusch W. 2009 *Hierarchische matrizen: algorithmen und analysis.* Berlin, Germany: Springer.
29. Bebendorf M. 2000 Approximation of boundary element matrices. *Numer. Math.* **86**, 565–589. (doi:10.1007/PL00005410)
30. Carvajal OA. 2004 A hybrid symbolic-numeric method for multiple integration based on tensor-product series approximations. Master's thesis, University of Waterloo, Waterloo, ON, Canada.
31. Chapman FW. 2003 Generalized orthogonal series for natural tensor product interpolation. PhD thesis, University of Waterloo, Waterloo, ON, Canada.
32. Geddes KO. 2008 Convergence theory for Geddes–Newton series expansions of positive definite kernels. Presentation at Symbolic and Numerical Computing 2008, Hagenberg, Austria.
33. Harbrecht H, Peters M, Schneider R. 2012 On the low-rank approximation by the pivoted Cholesky decomposition. *Appl. Numer. Math.* **62**, 428–440. (doi:10.1016/j.apnum.2011.10.001)
34. Higham NJ. 1987 A survey of condition number estimation for triangular matrices. *SIAM Rev.* **29**, 575–596. (doi:10.1137/1029112)