Journal of Pure and Applied Mathematics: Advances and Applications
Volume 26, Number 1, 2023, Pages 1-23
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/jpamaa_7100122267

A NOTE ON BANDED LINEAR SYSTEMS

D. BARRIOS ROLANÍA and J. C. GARCÍA-ARDILA

Depto. Matemática Aplicada a la Ingeniería Industrial
ETSI Industriales
Universidad Politécnica de Madrid
Spain
e-mail: dolores.barrios.rolania@upm.es

Abstract

In previous works, we studied and analized the Darboux factorization for semi-infinite Hessenberg banded matrices. In this note, we prove that this kind of factorization can be used also for finite matrices. In addition, a new method for solving banded linear systems is provided. Finally, some numerical experiments are reported to show the effectiveness of the proposed method.

1. Introduction

Banded linear systems constitute a relevant kind of linear systems in scientific computing due to its applications in many areas of science and engineering. These systems arise in the study of p-orthogonal polynomials and other fields of approximation theory, as well as the

2020 Mathematics Subject Classification: 15A23, 65F05.
Keywords and phrases: matrix factorization, numerical solutions of linear systems.
Received December 19, 2022; Revised February 11, 2023

discretization and linearization of differential equations [4, 7, 12, 19]. In particular, tridiagonal linear systems are associated with splines [14], quadrature formulas, and other subjects where the zeros of a sequence of orthogonal polynomials have to be located [10].

An extensive class of direct methods for solving a linear system

$$A_N X = b, \tag{1}$$

is based on the LU triangular decomposition

$$A_N = L_N U_N, \tag{2}$$

of the coefficient matrix [3, 18]. It is known that there is no universally best method to solve linear systems. In fact, the choice of one or the other method depends on the problem under consideration, which justifies the construction of new methods in addition to the already known ones [8, 13]. In this sense, we emphasize that, if A_N is a Hessenberg matrix, not necessarily banded, there are several sophisticated methods to deal with (1) (see, for example, [5, 6, 11, 15, 16, 17]).

It is well-known that when A_N is a finite Hessenberg banded matrix of order N,

$$A_N = \begin{pmatrix} a_{0,0} & a_{0,1} & 0 & \cdots & \cdots & 0 \\ a_{1,0} & a_{1,1} & a_{1,2} & & & \vdots \\ \vdots & \vdots & \ddots & \ddots & & \vdots \\ a_{p,0} & a_{p,1} & \cdots & a_{p,p} & a_{p,p+1} & \vdots \\ 0 & a_{p+1,1} & & \ddots & \ddots & \vdots \\ \vdots & 0 & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & a_{N-2,N-p-1} & \cdots & a_{N-2,N-1} \\ 0 & \dots & 0 & a_{N-1,N-p-1} & \cdots & a_{N-1,N-1} \end{pmatrix},$$

with $a_{i+p,i} \neq 0$, $i = 0, 1, \ldots, N - p - 1$, and the factorization (2) can be obtained, then L_N and U_N are triangular banded matrices of the same order N. We assume $N >> p$. Furthermore, the diagonal entries of L_N can be assumed to be equal to 1, being

$$L_N = \begin{pmatrix} 1 & & & & & & & \\ l_{1,0} & 1 & & & & & & \\ \vdots & \vdots & \ddots & & & & & \\ \vdots & \vdots & \ddots & \ddots & & & & \\ l_{p,0} & l_{p,1} & \cdots & l_{p,p-1} & 1 & & & \\ 0 & l_{p+1,1} & \cdots & \cdots & l_{p+1,p} & 1 & & \\ \vdots & \ddots & \ddots & & \ddots & \ddots & & \\ 0 & \cdots & 0 & l_{N-1,N-p-1} & \cdots & l_{N-1,N-2} & 1 \end{pmatrix}, \tag{3}$$

with $l_{p+i,i} \neq 0$, for $i = 0, 1, \ldots, N - p - 1$. In this case U_N is a bi-diagonal upper triangular matrix,

$$U_N = \begin{pmatrix} u_{0,0} & u_{0,1} & & & & \\ & u_{1,1} & \ddots & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & u_{N-2,N-2} & u_{N-2,N-1} & \\ & & & & u_{N-1,N-1} \end{pmatrix}. \tag{4}$$

Under the above conditions, the matrices L_N and U_N are uniquely determined.

On the other hand, the Darboux factorization for a semi-infinite lower triangular $(p+1)$-banded matrix L was introduced and analyzed in [2]. Assuming

$$L = \begin{pmatrix} 1 & & & & & \\ l_{1,0} & 1 & & & & \\ \vdots & \vdots & \ddots & & & \\ \vdots & \vdots & \ddots & \ddots & & \\ l_{p,0} & l_{p,1} & \cdots & l_{p,p-1} & 1 & \\ 0 & l_{p+1,1} & \cdots & \cdots & l_{p+1,p} & 1 \\ \vdots & \ddots & \ddots & & \ddots & \ddots \end{pmatrix}, \tag{5}$$

and $l_{p+i,i} \neq 0$ for $i = 0, 1, \ldots,$ the existence of p bi-diagonal semi-infinite matrices $L^{(i)}$, $i = 1, 2, \ldots, p$,

$$L^{(i)} = \begin{pmatrix} 1 & & & \\ \gamma_{i+1} & 1 & & \\ & \gamma_{p+i+2} & 1 & \\ & & \gamma_{2p+i+3} & \ddots \\ & & & \ddots \end{pmatrix}, \quad \gamma_{jp+i+j+1} \neq 0,\ j = 0, 1, \ldots,$$

verifying

$$L = L^{(1)} L^{(2)} \cdots L^{(p)} \tag{6}$$

was proved. When $p \geq 2$ this decomposition is not unique, since it depends on the choice of the set of entries

$$\begin{matrix} \gamma_2 & \cdots & \gamma_{p-1} & \gamma_p \\ \gamma_{p+3} & \cdots & \gamma_{2p} & \\ \vdots & \iddots & & \\ \gamma_{p(p-1)} & & & \end{matrix} \tag{7}$$

(see Table 1). In this note we show that this factorization can be used in the case of finite matrices. As a consequence, a new method is provided to solve a linear system (1).

If $p = 1$ we have that $L = L^{(1)}$ is a bi-diagonal matrix. In this case our method is reduced to the resolution of a tridiagonal system (1) using the classical LU decomposition (see [9]). Therefore, we assume $p \geq 2$ in the sequel.

In Section 2, we analyze the effect of the Darboux factorization on finite banded Hessenberg matrices. This factorization is applied in Section 3 to obtain the solution of finite banded systems, leading to the new method. Some examples are giving in Section 4 to illustrate the proposed method and, finally, some conclusions about our work are comment in Section 5.

2. Darboux Factorization for Finite Matrices

For an infinite lower banded matrix A, we assume $A = LU$ where L is given as in (5) and U is the upper triangular bi-diagonal matrix whose main section of order N is given in (4). Following [1], we assume that L is decomposed as in (6). Each row in Table 1 represents the corresponding entries in that row on the diagonal of U and also on each one of the factors of (6).

We consider the secondary diagonal in Table 1, this is,

$$\gamma_{p+1},\ \gamma_{2p+1},\ \cdots,\ \gamma_{p^2+1},\ \gamma_{(p+1)p+1}.$$

At the top of this secondary diagonal we see some framed entries, which are the starting data (7). Furthermore the entries of U in the first column of Table 1 are well known from the LU factorization. Our main aim in this section is to show that each one of the rest of the entries can be determined from the previous rows.

Table 1. Factors of L

U	$L^{(1)}$	$L^{(2)}$	$\cdots$	$L^{(p-s)}$	$\cdots$	$L^{(p-2)}$	$L^{(p-1)}$	$L^{(p)}$
γ_1	$\boxed{\gamma_2}$	$\boxed{\gamma_3}$	$\cdots$	$\boxed{\gamma_{p-s+1}}$	$\cdots$	$\boxed{\gamma_{p-1}}$	$\boxed{\gamma_p}$	γ_{p+1}
γ_{p+2}	$\boxed{\gamma_{p+3}}$	$\boxed{\gamma_{p+4}}$	$\cdots$	$\boxed{\gamma_{2p-s+2}}$	$\cdots$	$\boxed{\gamma_{2p}}$	γ_{2p+1}	γ_{2p+2}
γ_{2p+3}	$\boxed{\gamma_{2p+4}}$	$\boxed{\gamma_{2p+5}}$	$\cdots$	$\boxed{\gamma_{3p-s+3}}$	$\cdots$	γ_{3p+1}	γ_{3p+2}	γ_{3p+3}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	$\vdots$
$\gamma_{(s-1)p+s}$	$\boxed{\gamma_{(s-1)p+s+1}}$	$\boxed{\gamma_{(s-1)p+s+2}}$	$\cdots$	$\boxed{\gamma_{sp}}$	$\cdots$	γ_{sp+s-2}	γ_{sp+s-1}	γ_{sp+s}
γ_{sp+s+1}	$\boxed{\gamma_{sp+s+2}}$	$\boxed{\gamma_{sp+s+3}}$	$\cdots$	$\gamma_{(s+1)p+1}$	$\cdots$	$\gamma_{(s+1)p+s-1}$	$\gamma_{(s+1)p+s}$	$\gamma_{(s+1)p+s+1}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	$\vdots$
$\gamma_{(p-2)p-2}$	$\boxed{\gamma_{(p-2)p-1}}$	$\boxed{\gamma_{(p-2)p}}$	$\cdots$	$\gamma_{(p-1)p-s-2}$	$\cdots$	$\gamma_{(p-2)p+p-4}$	$\gamma_{(p-2)p+p-3}$	$\gamma_{(p-2)p+p-2}$
$\gamma_{(p-1)p-1}$	$\boxed{\gamma_{(p-1)p}}$	$\gamma_{(p-1)p+1}$	$\cdots$	γ_{p^2-s-1}	$\cdots$	$\gamma_{(p-1)p+p-3}$	$\gamma_{(p-1)p+p-2}$	$\gamma_{(p-1)p+p-1}$
γ_{p^2}	γ_{p^2+1}	γ_{p^2+2}	$\cdots$	$\gamma_{(p+1)p-s}$	$\cdots$	γ_{p^2+p-2}	γ_{p^2+p-1}	γ_{p^2+p}
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	$\vdots$

We call s-th *secondary diagonal*, $s = 1, 2, \ldots,$ the set given by the entries

$$\gamma_{sp+s},\ \gamma_{(s+1)p+s},\ \cdots,\ \gamma_{(s+p-1)p+s},\ \gamma_{(s+p)p+s},$$

in Table 1. In particular, for $s = 1$ we have the previously called secondary diagonal. In the following, for each fixed $i \in \mathbb{N}$, we show that the i-th secondary diagonal is determined in terms of the previous s-th secondary diagonals, $s = 1, 2, \ldots, i-1,$ and the starting data (7). Furthermore, we will see that each entry of this i-th secondary diagonal in the k-th row is obtained exclusively in terms of such entries that are in the rows $1, 2, \ldots, k$.

From [1, (35)], we have

$$\delta_k^{(i)}\gamma_{(k+i+1)p+i} = a_{k+i+1,\,i-1} - \sum_{\widetilde{E}_{k+2}^{(0)}} \gamma_{(i-2)p+i+i_1-1}\gamma_{(i-1)p+i+i_2-1}\cdots\gamma_{(k+i)p+i+i_{k+3}-1},$$

$$k = -1,\ 0,\ \ldots,\ p-2,$$

where

$$\delta_k^{(i)} = \gamma_{(i-1)p+i}\gamma_{ip+i}\cdots\gamma_{(k+i)p+i}, \tag{8}$$

and

$$\widetilde{E}_{k+2}^{(0)} = \{(i_1, \ldots, i_{k+3}) : k+3 \le i_{k+3} \le \cdots \le i_1 \le p+1,\ i_{k+3} < p+1\}. \tag{9}$$

For each $k = -1,\ 0,\ \ldots,\ p-2,$ the entry $\gamma_{(k+i+1)p+i}$ is in the $(i+k+1)$-th row and i-th secondary diagonal. Since (8), we can express this entry in terms of $\delta_k^{(i)}$ and

$$\gamma_{(i-2)p+i+i_1-1},\ \gamma_{(i-1)p+i+i_2-1},\ \cdots,\ \gamma_{(k+i)p+i+i_{k+3}-1}, \tag{10}$$

when $(i_1, \ldots, i_{k+3}) \in \widetilde{E}_{k+2}^{(0)}$.

Firstly, from (8), we see that $\delta_k^{(i)}$ is computed from the entries of the same i-th secondary diagonal that is in the rows $i, i+1, \ldots, i+k$.

Second, we analyze the entries (10), this is,

$$\gamma_{(r+i)p+i+i_{r+3}-1}, \quad r = -2, -1, \ldots, k. \tag{11}$$

If $r \leq k-1$ then, taking into account (9),

$$(r+i)p + (i+1) \leq (r+i)p + i + i_{r+3} - 1 \leq (r+i+1)p + i.$$

Hence $\gamma_{(r+i)p+i+i_{r+3}-1}$ is in some row of Table 1 before the $(r+i-1)$-th row. Moreover, when $\gamma_{(r+i-1)p+i+i_{r+2}-1}$ is in the j-th column, then $\gamma_{(r+i)p+i+i_{r+3}-1}$ is in the $(j-1)$-th column or some previous column of the following row. Therefore, if $\gamma_{(r+i-1)p+i+i_{r+2}-1}$ is at the top of the i-th secondary diagonal, the same is true for $\gamma_{(r+i)p+i+i_{r+3}-1}$. Finally, for $r = k$ in (11) the situation is similar but now $\gamma_{(r+i)p+i+i_{r+3}-1} = \gamma_{(k+i)p+i+i_{k+3}-1}$ is in the $(k+i-1)$-th row and not in the i-th secondary diagonal, because $(r+i)p + i + i_{k+3} - 1 < (r+i+1)p + i$. (Just, the entry of this row in the i-th secondary diagonal is that we want to compute.)

In summary, each entry in the i-th secondary diagonal of Table 1 is obtained with the entries of the previous rows that are at the top of the i-th secondary diagonal. Translating this reasoning into matrices $L^{(1)}, \ldots, L^{(p)}, U$, we deduce that the entry in the row i of $L^{(s)}$, $s = 1, \ldots, p$, is obtained using only the rows $1, 2, \ldots, i$ of $L^{(1)}, \ldots, L^{(s)}, U$. As a consequence, $(L^{(1)} \cdots L^{(p)}U)_n = L_n^{(1)} \cdots L_n^{(p)}U_n$, $n \in \mathbb{N}$. In particular,

$$(L^{(1)} \cdots L^{(p)}U)_N = L_N^{(1)} \cdots L_N^{(p)}U_N.$$

From this and from the well-known fact that $(LU)_N = L_N U_N$, we obtain

$$(L^{(1)} \cdots L^{(p)})_N = L_N^{(1)} \cdots L_N^{(p)}. \tag{12}$$

3. Darboux Factorization and Banded Systems

As a consequence of (12), it is possible to use the Darboux factorization for finite matrices. In other words, if there exists the LU factorization for the coefficients matrix A_N in system (1), then we have

$$A_N = L_N^{(1)} \cdots L_N^{(p)} U_N, \tag{13}$$

and we can define

$$X^{(i)} = \begin{cases} L_N^{(i+1)} \cdots L_N^{(p)} U_N X, & i = 1, \ldots, p-1, \\ U_N X, & i = p. \end{cases}$$

Thereby, (1) is reduced to solve iteratively the following $p+1$ simple triangular systems

$$\begin{cases} L_N^{(1)} X^{(1)} = b, \\ L_N^{(k)} X^{(k)} = X^{(k-1)}, \quad k = 2, \ldots, p \\ U_N X = X^{(p)}, \end{cases} \tag{14}$$

In fact, the first p of these systems (corresponding to $L_N^{(k)} X^{(k)} = X^{(k-1)}$, $k = 1, \ldots, p$, with $X^{(0)} = b$) can be solved by forward substitution and the last one (corresponding to $U_N X = X^{(p)}$) can be solved by backward substitution.

We assume (2), where L_N and U_N are given by (3) and (4), respectively. With the purpose of derive an algorithm to obtain the decomposition (13), we write the entries of L_N verifying this decomposition, this is,

$$l_{m,\, m-k} = \sum_{1 \le \sigma_1 < \cdots < \sigma_k \le p} \left(\prod_{j=1}^{k} \gamma_{(m-j)p+\sigma_j+m-j+1} \right), \; k = 1, \ldots, p.$$

We recall that the matrix U_N is known from the LU factorization of A_N. Therefore, we only need to determine the entries $\gamma_{(m-1)p+m+i}$, which are, for $m = 1, 2, \ldots, N-1$, in each row of matrices $L_N^{(i)}$, $i = 1, 2, \ldots, p$. Therefore,

$$l_{m,m-k} = \sum_{\substack{1\le\sigma_1<\cdots<\sigma_k\le p\\ \sigma_1\neq p-k+1}} \left(\prod_{j=1}^{k} \gamma_{(m-j)p+\sigma_j+m-j+1}\right) + \gamma_{m(p+1)-k+1} \prod_{j=2}^{k} \gamma_{(m-j+1)p-k+m+1},\ k = 1, \ldots, p. \tag{15}$$

Besides, because we are assuming $a_{m,m-p} \neq 0$, then all the entries of $L^{(s)}$, $s = 1, \ldots, p$, that are not starting data are necessarily nonzero. This is, $\gamma_{(j-1)p+s+j} \neq 0$ for $s = 1, \ldots, p$, $j = 1, 2, \ldots$. Thus, defining

$$\Gamma_s := \prod_{j=2}^{p-s+1} \gamma_{(m-j)p+m+s}, \quad s = 1, \ldots, p, \tag{16}$$

(with $\Gamma_p = 1$) we have $\Gamma_s \neq 0$ and, from this and (15), taking $k = p - s + 1$ for $s = 1, \ldots, p$, we can write

$$\gamma_{(m-1)p+m+s} = \left(l_{m,\,m-p+s-1} - \sum_{\substack{1\le\sigma_1<\cdots<\sigma_{p-s+1}\le p\\ \sigma_1\neq s}} \prod_{j=1}^{p-s+1} \gamma_{(m-j)p+\sigma_j+m-j+1} \right) / \Gamma_s. \tag{17}$$

Thereby, when all the entries $\gamma_{(k-1)p+k+s}$, $s = 1, 2 \ldots, p$, $k = 1, 2, \ldots, m-1$, have been calculated, $\gamma_{(m-1)p+m+s}$ can be computed using (16)-(17).

In this way, from the starting data (7) we get row by row those of Table 1. If $a \to b$ means that b is obtained from a, in schematic form, we write

$$\to \gamma_{p+1}$$

$$\to \gamma_{2p+1} \to \gamma_{2p+2}$$

$$\to \gamma_{3p+1} \to \gamma_{3p+2} \to \gamma_{3p+3}$$

$$\vdots$$

$$\to \gamma_{p^2+1} \to \gamma_{p^2+2} \to \cdots \to \gamma_{p^2+p}$$

$$\vdots$$

$$\to \gamma_{(N-2)p+N} \to \gamma_{(N-2)p+N+1} \to \cdots \to \gamma_{(N-1)p+N-1}.$$

Remark 1. The free parameters in Table 1, corresponding to the starting data, are characterized by

$$\gamma_{ip+j}, \quad j < i.$$

These parameters are not involved in the computation of value Γ_s in (16). Hence it is possible to take $\gamma_{ip+j} = 0$ for $j < i$. This fact will be used in the sequel section, simplifying the computation of $L^{(i)}$, $i = 1, \ldots, p$, and, consequently, making it easier to solve (1). We recall that we are assuming a unique solution for (1), although it is possible to obtain it through the non-unique systems (14) because of the existence of non-unique factors $L^{(i)}$, $i = 1, \ldots, p$.

4. Numerical Results

In this section, we illustrate with some examples the proposed factorization and its application to solving banded systems. The numerical experiments carried out in Matlab R2020b on a PC equipped with an Intel(R) Core(TM) i7-8550U (CPU @ 1.80GHz-1.99 GHz).

Take a linear system (1), where we are assuming that A_N is a finite Hessenberg $(p+2)$-banded matrix. Consider (2), with L_N and U_N as in (3) and (4), respectively. Our first step is obtaining the factors $L_N^{(i)}$, $i = 1, \ldots, p$ such that

$$L_N = L_N^{(1)} L_N^{(2)} \cdots L_N^{(p)}. \tag{18}$$

In order to implement this decomposition we define the matrix R whose entries γ_i are the values of the free parameters (7),

$$R = \begin{pmatrix} \gamma_2 & \cdots & \gamma_{p-1} & \gamma_p \\ \gamma_{p+3} & \cdots & \gamma_{2p} & 0 \\ \vdots & \iddots & \iddots & \vdots \\ \gamma_{p(p-1)} & 0 & \cdots & 0 \end{pmatrix}. \tag{19}$$

We underline that here the entries γ_i in R are arbitrary, although in the next step we take $R = 0_{p-1}$ the matrix of order $p-1$ whose entries are all equal to 0.

The factorization (18) can be obtained using the following Matlab code:

Listing 1. Factorization of L in terms of $L^{(1)} \cdots L^{(p)}$

```
function[L1]=fac_L(L,R,p)
N=length(L);
L1=eye(N);
l=eye(N);
for k=0:p-1
   S=eye(N);
 if (p==1)
    L1=L;
 M=L1;
       break
end
if(p-1-k>0)
  S(2:N+1:(p-1-k)*N)=R([1:p-1-k],k+1);
end
   if(k<p-1)
       for t=1:p-k-1
           L(t+1,[1:t+k])=L(t+1,[1:t+k])-S(t+1,t)*L(t,[1:t+k]);
    end
  end
   for i=0:N-(p+1-k)
            l(p+1-k+i,p-k+i)=L(p+1-k+i,i+1)/L(p-k+i,i+1);
            L(p+1-k+i,[i+1:i+p+1-k])=L(p+1-k+i,[i+1:i+p+1-
            k])-l(p+1-k+i,p-k+i).*L(p-k+i,[i+1:i+p+1-k]);
    S(p+1-k+i,p-k+i)=l(p+1-k+i,p-k+i);
  end
  L1(:,:,k+1)=S;
end
end
```

The number of operations required on the loops are $2(t+k)$ and $2(p-k)+3$, respectively. Hence the total cost is given by

$$\begin{aligned}\mathrm{Cost}(N,\ p) &= \sum_{k=0}^{p-2}\left(\sum_{t=1}^{p-k-1} 2(t+k)\right) + \sum_{k=0}^{p-2}\left(\sum_{i=0}^{N-(p-k+1)} (2(p-k)+3)\right)\\ &= \frac{p(p-1)(2p-1)}{3} + (p-1)(N-p)(2p+3)\\ &\quad + (p-1)(p-2)\frac{(8p-6N+15)}{6}.\end{aligned}$$

In particular, if we take all the free parameters (7) as $\gamma_{ip+j} = 0$, this is, $R = 0_{p-1}$, the lines 12 to 14 of the above code are simplify and the computational cost is reduced to

$$(p-1)(N-p)(2p+3) + (p-1)(p-2)\frac{(8p-6N+15)}{6}.$$

Taking into account this fact, with the goal to solve the linear system (1) we assume $R = 0_{p-1}$ in the sequel.

We can use the classical LU decomposition to obtain (2) and then to apply (18) using the above code. However, with a similar structure to used there, it is possible to find directly (13) without to have to know U_N previously. This is done in the following code, where, from A_N and b, the matrices $L_N^{(i)}$, $i = 1, \ldots, p$, and the solution x of (1) are obtained. Moreover, the error $\|A_N x - b\|$ is computed (here and in the sequel we use the Euclidean norm $\|\cdot\|$).

Listing 2. Solution of the system $Ax = b$

```
function[L,U,x,r]=Sband(A,b)
N=length(A);
y=b;
U=A;
x=zeros(N,1);
L=eye(N);;
for s=1:N
    if A([s:N],1)==zeros(N-s+1,1)
     t=s-1;
    break
    end
 end
p=t-1;
for k=0:p-1
    L(:,:,k+1)=eye(N);
        for i=0:N-(p+1-k) %computation of matrices L^(i).
           L(p+1-k+i,p-k+i,k+1)=U(p+1-k+i,i+1)/U(p-k+i,i+1);
           U(p+1-k+i,[i+1:i+p+1-k])=U(p+1-k+i,[i+1:i+p+1-
           k])-L(p+1-k+i,p-k+i,k+1)*U(p-k+i,[i+1:i+p+1-k]);
       end
for t=2+p-(k+1):N
    y(t)=y(t)-L(t,t-1,k+1)*y(t-1); %forward substitution.
end
end

x(N)=y(N)/U(N,N);
for s=N-1:-1:1
x(s)=(y(s)-U(s,s+1)*x(s+1))/U(s,s);%backward substitution.
end
r=norm(A*x-b);
end
```

The number of operations required in the inner loops are $2(p-k)+3$ and 2, respectively. Moreover in the external loop we have 3 operations. Hence the operational count is

$$\begin{aligned}\mathrm{Cost}(N,\ p) &= 1+\sum_{s=1}^{N-1} 3+\sum_{k=0}^{p-1}\left(\sum_{i=0}^{N-(p-k+1)}(2(p-k)+3)+\sum_{t=p-k+1}^{N} 2\right)\\ &= 3N-2+(2p+4)(N-p)\,p+(p-1)p\,\frac{(8p-6N+14)}{6}.\end{aligned}$$

In the rest of this section we give some examples where we apply the above programs to show the proposed method.

Example 1. Let L_N be a lower triangular matrix of order N obtained by using Matlab as

$$L_N = triu(tril(rand(N),\ -1),\ -p)+eye(N).$$

In this example, taking $p=3$ and $N=6$, we obtain

$$L_6=\begin{pmatrix} 1.0000e{+}00 & 0 & 0 & 0 & 0 & 0\\ 6.8893e{-}01 & 1.0000e{+}00 & 0 & 0 & 0 & 0\\ 9.4602e{-}01 & 6.1273e{-}01 & 1.0000e{+}00 & 0 & 0 & 0\\ 8.7354e{-}01 & 3.0081e{-}01 & 7.0870e{-}01 & 1.0000e{+}00 & 0 & 0\\ 0 & 7.9814e{-}01 & 9.9293e{-}01 & 8.0991e{-}01 & 1.0000e{+}00 & 0\\ 0 & 0 & 1.6248e{-}01 & 1.8676e{-}01 & 2.7750e{-}02 & 1.0000e{+}00 \end{pmatrix}.$$

Using the code of `fac_L(L,R,p)` with $R=\begin{pmatrix}1 & 2\\ 1 & 0\end{pmatrix}$ we arrive to $L_6 = L_6^{(1)}L_6^{(2)}L_6^{(3)}$, where

$$L_6^{(1)} = \begin{pmatrix} 1.0000e+00 & 0 & 0 & 0 & 0 & 0 \\ 1.0000e+00 & 1.0000e+00 & 0 & 0 & 0 & 0 \\ 0 & 1.0000e+00 & 1.0000e+00 & 0 & 0 & 0 \\ 0 & 0 & 6.9489e-01 & 1.0000e+00 & 0 & 0 \\ 0 & 0 & 0 & 1.4004e+00 & 1.0000e+00 & 0 \\ 0 & 0 & 0 & 0 & 1.6689e-01 & 1.0000e+00 \end{pmatrix},$$

$$L_6^{(2)} = \begin{pmatrix} 1.0000e+00 & 0 & 0 & 0 & 0 & 0 \\ 2.0000e+00 & 1.0000e+00 & 0 & 0 & 0 & 0 \\ 0 & -5.4394e-01 & 1.0000e+00 & 0 & 0 & 0 \\ 0 & 0 & 3.6377e+00 & 1.0000e+00 & 0 & 0 \\ 0 & 0 & 0 & -2.6865e-01 & 1.0000e+00 & 0 \\ 0 & 0 & 0 & 0 & -8.8640e-01 & 1.0000e+00 \end{pmatrix},$$

$$L_6^{(3)} = \begin{pmatrix} 1.0000e+00 & 0 & 0 & 0 & 0 & 0 \\ -2.3111e+00 & 1.0000e+00 & 0 & 0 & 0 & 0 \\ 0 & 1.5667e-01 & 1.0000e+00 & 0 & 0 & 0 \\ 0 & 0 & -3.6239e+00 & 1.0000e+00 & 0 & 0 \\ 0 & 0 & 0 & -3.2187e-01 & 1.0000e+00 & 0 \\ 0 & 0 & 0 & 0 & 7.4727e-01 & 1.0000e+00 \end{pmatrix}.$$

In fact, computing $L_6^{(1)}L_6^{(2)}L_6^{(3)}$ and compare with L_6, we see the error

$$\| L_6 - L_6^{(1)}L_6^{(2)}L_6^{(3)} \| = 2.8448e-16,$$

which is very close to the machine epsilon in Matlab.

Example 2. As in Example 1, L_N is generated as

$$L_N = triu(tril(rand(N), -1), -p) + eye(N).$$

Now we use `fac_L(L,R,p)` with $R = O_{p-1}$ and we study the error $\| L_N - L_N^{(1)} \cdots L_N^{(p)} \|$ for several values of p and N. The results are summarized in the following table:

Table 2. Error $\| L_N - L_N^{(1)} \cdots L_N^{(p)} \|$ for several values of N and p

N	p	Error
100	2	5.9962e-15
100	49	4.6883e-12
100	69	9.8936e-12
100	94	4.3101e-11

N	P	Error
200	2	2.7756e-15
200	99	6.7055e-11
200	149	1.2626e-10
200	194	3.9741e-11

N	p	Error
300	2	2.2204e-15
300	99	6.2859e-11
300	199	7.2198e-10
300	294	1.0849e-10

In the following examples, the Matlab function `Sband(A,b)` defined in the second program is used to solve the linear banded system (1).

Example 3. Let A_N be the $(p+2)$-banded Hessenberg matrix of order $N \times N$ generated with the Matlab command $A_N = triu(tril(rand(N),1),-p)$, and let b be defined as b = rand(1, N). In the case $p = 3$ and $N = 5$ we have obtained $b = b_5 = (0.5788,\ 0.8670,\ 0.4067,\ 0.1126,\ 0.4438)^t$ and

$$A_5 = \begin{pmatrix} 0.8487 & 0.1008 & 0 & 0 & 0 \\ 0.9168 & 0.5078 & 0.5170 & 0 & 0 \\ 0.9870 & 0.5856 & 0.1710 & 0.6559 & 0 \\ 0.5051 & 0.7629 & 0.9386 & 0.4519 & 0.3672 \\ 0 & 0.0830 & 0.5905 & 0.8397 & 0.2393 \end{pmatrix}.$$

Then applying `Sband(A,b)` for $A = A_5$ and $b = b_5$, we have

$$L_5^{(1)} = \begin{pmatrix} 1.0000 & 0 & 0 & 0 & 0 \\ 0 & 1.0000 & 0 & 0 & 0 \\ 0 & 0 & 1.0000 & 0 & 0 \\ 0 & 0 & 0.5118 & 1.0000 & 0 \\ 0 & 0 & 0 & 0.1791 & 1.0000 \end{pmatrix},$$

$$L_5^{(2)} = \begin{pmatrix} 1.0000 & 0 & 0 & 0 & 0 \\ 0 & 1.0000 & 0 & 0 & 0 \\ 0 & 1.0765 & 1.0000 & 0 & 0 \\ 0 & 0 & 11.9054 & 1.0000 & 0 \\ 0 & 0 & 0 & 0.0805 & 1.0000 \end{pmatrix},$$

$$L_5^{(3)} = \begin{pmatrix} 1.0000 & 0 & 0 & 0 & 0 \\ 1.0803 & 1.0000 & 0 & 0 & 0 \\ 0 & 0.0975 & 1.0000 & 0 & 0 \\ 0 & 0 & -12.4810 & 1.0000 & 0 \\ 0 & 0 & 0 & 2.9126 & 1.0000 \end{pmatrix},$$

$$U = \begin{pmatrix} 0.8487 & 0.1008 & 0 & 0 & 0 \\ 0 & 0.3990 & 0.5170 & 0 & 0 \\ 0 & 0 & -0.4359 & 0.6559 & 0 \\ 0 & 0 & 0 & 0.4938 & 0.3672 \\ 0 & 0 & 0 & 0 & -0.9255 \end{pmatrix}.$$

Table 3. Error $\|A_N x - b\|$ solving the system

N	p	$\|A_N x - b\|$	Cond(A_N)
100	2	7.9062e-12	1.1798e+05
100	9	1.9265e-12	8.6089e+07
100	49	1.0565e-06	4.0811e+09
100	69	1.1231e-06	3.2712e+09
100	94	3.7707e-06	1.4124e+11
200	2	2.6245e-06	1.9270e+11
300	2	9.8204e-05	3.0101e+12

We obtained the approximate solution $x = (0.8481, -1.3984, 1.5465, 0.1892, -2.1404)^t$ with an error $\|A_5 x - b_5\| = 3.5544e-16$.

Example 4. As in Example 3, A_N is a $(p+2)$-banded Hessenberg matrix generated with the Matlab command $A_N = triu(tril((N), 1), -p)$, and b is generated as $b = \text{rand}(1, N)$. We use the code of `Sband(A,b)` to obtain the approximate solution x and we compute the error $\|A_N x - b\|$ for several values of N and p. We summarize the results in the Table 3, where the last column indicates the condition number of A_N.

5. Conclusions

In this paper a method to solve banded linear systems is proposed, which is based on the decomposition of the matrix of coefficients in product of bi-diagonal matrices. This new method (14) is an extension of the classical method commonly used to solve tridiagonal systems based in the LU factorization of the coefficients matrix (see [9]).

One remarkable advantage of the new method is its low computational cost. Several numerical experiments have been given, showing this fact and the excellent results obtained.

Although in this paper only the case of Hessenberg matrices is studied, the method is easily extended to general banded matrices. In fact, if A_N is a banded matrix, but not a Hessenberg matrix, then its LU factorization leads to a $(q+1)$-banded upper triangular matrix U_N and, making use of the same idea of this work, the lower triangular matrix U_N^T can be decomposed as a product of bi-diagonal lower triangular matrices. This is,

$$U_N^T = U_N^{(q)^T} \cdots U_N^{(1)^T}.$$

Hence

$$A_N = L_N^{(1)} \cdots L_N^{(p)} U_N^{(1)} \cdots U_N^{(q)},$$

and one algorithm similar to (14) can be applied to solve (1).

Acknowledgements

The work of D. Barrios Rolanía was partially supported by Agencia Estatal de Investigación, Ministerio de Ciencia e Innovación, Spain, under grant PID2021-122154NB-I00.

The work of J. C. García-Ardila was partially supported by Comunidad de Madrid multiannual agreement with the Universidad Rey Juan Carlos, Spain, under grant Proyectos I+D para Jóvenes Doctores, Ref. M2731, project NETA-MM.

References

[1] D. Barrios Rolanía, On the Darboux transform and the solutions of some integrable systems, Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales: Serie A. Matemáticas 113(2) (2019), 1359-1378.

DOI: https://doi.org/10.1007/s13398-018-0553-5

[2] D. Barrios Rolanía and D. Manrique, On the existence of Darboux transformations for banded matrices, Applied Mathematics and Computation 253 (2015), 116-125.

DOI: https://doi.org/10.1016/j.amc.2014.12.053

[3] I. S. Duff, A review of frontal methods for solving linear systems, Computer Physics Communications 97(1-2) (1996), 45-52.

DOI: https://doi.org/10.1016/0010-4655(96)00020-3

[4] J. M. McDonough. Lectures on Computational Numerical Analysis of Partial Differential Equations, Mechanical Engineering Textbook Gallery 3, University of Kentucky, Lexington, KY, 2008.

https://uknowledge.uky.edu/me textbooks/3/

[5] L. Gemignani and G. Lotti, Efficient and stable solution of *M*-matrix linear systems of (block) Hessenberg form, SIAM Journal on Matrix Analysis and Applications 24(3) (2003), 852-876.

DOI: https://doi.org/10.1137/S0895479801387085

[6] L. Gemignani and F. Poloni, Comparison theorems for splittings of *M*-matrices in (block) Hessenberg form, BIT Numerical Mathematics 62(3) (2022), 849-867.

DOI: https://doi.org/10.1007/s10543-021-00899-4

[7] S. Ghosh, Skew-orthogonal polynomials, differential systems and random matrix theory, Journal of Physics A: Mathematical and Theoretical 40(4) (2007), 711-740.

DOI: https://doi.org/10.1088/1751-8113/40/4/009

[8] X. M. Gu, H. W. Sun, Y. L. Zhao and X. Zheng, An implicit difference scheme for time-fractional diffusion equations with a time-invariant type variable order, Applied Mathematics Letters 120 (2021); Article 107270.

DOI: https://doi.org/10.1016/j.aml.2021.107270

[9] E. Isaacson and H. B. Keller, Analysis of Numerical Methods, Dover Publications, Inc., New York, 1994.

[10] C. Jagels and L. Reichel, On the computation of Gauss quadrature rules for measures with a monomial denominator, Journal of Computational and Applied Mathematics 286 (2015), 102-113.

DOI: https://doi.org/10.1016/j.cam.2015.02.042

[11] M. A. Jandron, A. A. Ruffa and J. Baglama, An asynchronous direct solver for banded linear systems, Numerical Algorithms 76(1) (2017), 211-235.

DOI: https://doi.org/10.1007/s11075-016-0251-3

[12] H. B. Li, M. Y. Song, E. J. Zhong and X. M. Gu, Numerical gradient schemes for heat equations based on the collocation polynomial and Hermite interpolation, Mathematics 7(1) (2019); Article 93.

DOI: https://doi.org/10.3390/math7010093

[13] W. H. Luo, X. M. Gu and B. Carpentieri, A hybrid triangulation method for banded linear systems, Mathematics and Computers in Simulation 194 (2022), 97-108.

DOI: https://doi.org/10.1016/j.matcom.2021.11.012

[14] W. H. Luo, T. Z. Huang, L. Li, H. B. Li and X. M. Gu, Quadratic spline collocation method and efficient preconditioner for the Helmholtz equation with the Sommerfeld boundary conditions, Japan Journal of Industrial and Applied Mathematics 33(3) (2016), 701-720.

DOI: https://doi.org/10.1007/s13160-016-0225-9

[15] A. A. Ruffa, A solution approach for lower Hessenberg linear systems, ISRN Applied Mathematics (2011); Article ID 236727.

DOI: https://doi.org/10.5402/2011/236727

[16] G. W. Stewart, On the solution of block Hessenberg systems, Numerical Linear Algebra with Applications 2(3) (1995), 287-296.

DOI: https://doi.org/10.1002/nla.1680020309

[17] H. S. Stone, An efficient parallel algorithm for the solution of a tridiagonal linear system of equations, Journal of the Association for Computing Machinery 20(1) (1973), 27-38.

DOI: https://doi.org/10.1145/321738.321741

[18] L. N. Trefethen and D. Bau, Numerical Linear Algebra, In: SIAM, Philadelphia, PA, 1997.

[19] Y. L. Zhao, P. Y. Zhu, X. M. Gu and X. L. Zhao, A second-order accurate implicit difference scheme for time fractional reaction-diffusion equation with variable coefficients and time drift term, East Asian Journal on Applied Mathematics 9(4) (2019), 723-754.

DOI: https://doi.org/10.4208/eajam.200618.250319

■