

# The Algebra of Elimination

Gilbert Strang

Massachusetts Institute of Technology

[gilstrang@gmail.com](mailto:gilstrang@gmail.com)

## Abstract

Elimination with only the necessary row exchanges will produce the triangular factorization  $A = LPU$ , with the (unique) permutation  $P$  in the middle. The entries in  $L$  are reordered in comparison with the more familiar  $A = PLU$  (where  $P$  is not unique). Elimination with three other starting points  $1, n$  and  $n, n$  and  $n, 1$  produces three more factorizations of  $A$ , including the Wiener-Hopf form  $UPL$  and Bruhat's  $U_1 \pi U_2$  with two upper triangular factors.

All these starting points are useless for doubly infinite matrices. The matrix has no first or last entry. When  $A$  is banded and invertible, we look for a new way to establish  $A = LPU$ . First we locate the pivot rows (and the main diagonal of  $A$ ).  $LPU$  connects to the classical factorization of matrix polynomials developed for the periodic (block Toeplitz) case when  $A(i, j) = A(i + b, j + b)$ .

**Key words.** factorization, elimination, banded matrix, infinite matrix, Bruhat

**AMS subject classification.** 15.01, 15A 23

## 1. Introduction.

The “pedagogical” part of this paper presents the  $LPU$  factorization of an invertible  $n$  by  $n$  matrix  $A$ :

$$A = LPU = (\text{lower triangular})(\text{permutation})(\text{upper triangular}).$$

The reader may feel that everything has been said about the algebra of elimination, which produces  $L$ ,  $P$ , and  $U$ . This is potentially true. But who said it, and where, is not easy to discover. I hope you will feel that some of this is worth saying again. The  $LPU$  form that algebraists like best (with  $P$  in the middle instead of the more practical  $A = PLU$ ) is the least familiar within SIAM.

Once started in this direction, factorizations continue to appear. If elimination begins at the last entry  $A_{nn}$  and works upward, the result is  $UPL$ . Those are new factors of  $A$ , and there must be relations to the original  $L$ ,  $P$ , and  $U$  that we don't know. More inequivalent forms  $A = U_1 \pi U_2$  and  $A = L_1 \pi L_2$  come from starting elimination at  $A_{n1}$  and at  $A_{1n}$ . You may be surprised that the all-time favorite of algebraists is Bruhat's  $U_1 \pi U_2$ : hard to comprehend (but see Section 4).

$$\begin{array}{ccc}
\text{(down and right) } A = LPU \leftarrow & \begin{array}{|c|c|} \hline A_{11} & A_{1n} \\ \hline \end{array} & \rightarrow A = L_1 \pi L_2 \text{ (down and left)} \\
\text{(up and right) } A = U_1 \pi U_2 \leftarrow & \begin{array}{|c|c|} \hline A_{n1} & A_{nn} \\ \hline \end{array} & \rightarrow A = UPL \text{ (up and left)}
\end{array}$$

The more original part of this paper extends  $A = LPU$  to *banded doubly infinite matrices*. What makes this challenging is that elimination has no place to begin.  $A_{11}$  is deep in the middle of  $A$ , and algebra needs help from analysis. The choice of pivot appears to depend on infinitely many previous choices. The same difficulty arose for Wiener and Hopf, because they wanted  $A = UL$  and singly infinite matrices have no last entry  $A_{nn}$ . This was overcome in the periodic (block Toeplitz) case, and in Section 6 we go further.

## 2. The Uniqueness of $P$ in $A = LPU$ .

**Theorem 1.** The permutation  $P$  in  $A = LPU$  is uniquely determined by  $A$ .

*Proof.* Consider the  $s$  by  $t$  upper left submatrices of  $A$  and  $P$ . That part of the multiplication  $A = LPU$  leads to  $a = \ell pu$  for the submatrices, because  $L$  and  $U$  are triangular:

$$\begin{bmatrix} a & * \\ * & * \end{bmatrix} = \begin{bmatrix} \ell & 0 \\ * & * \end{bmatrix} \begin{bmatrix} p & * \\ * & * \end{bmatrix} \begin{bmatrix} u & * \\ 0 & * \end{bmatrix} \text{ gives } a = \ell pu. \quad (1)$$

The submatrix  $\ell$  is  $s$  by  $s$  and  $u$  is  $t$  by  $t$ . Both have nonzero diagonals (therefore invertible) since they come from the invertible  $L$  and  $U$ . Then  $p$  has the same rank as  $a = \ell pu$ . The ranks of all upper left submatrices  $p$  are determined by  $A$ , so the whole permutation  $P$  is uniquely determined [6, 7, 12].

The number of 1's in  $p$  is its rank. Since those 1's produce independent columns (they come from different rows of  $P$ ). The rule is that  $P_{ik} = 1$  exactly where the rank of the upper left submatrices  $a_{ik}$  of  $A$  increases:

$$\text{rank } a_{ik} = 1 + \text{rank } a_{i-1,k-1} = 1 + \text{rank } a_{i-1,k} = 1 + \text{rank } a_{i,k-1}. \quad (2)$$

In words, row  $i$  is dependent on previous rows until column  $k$  is included, and column  $k$  is dependent on previous columns until row  $i$  is included. When  $A = LPU$  is constructed by elimination, a pivot will appear in this  $i, k$  position. The pivot row  $i(k)$  for elimination in column  $k$  will be the first row (the smallest  $i$ ) such that (2) becomes true. Since by convention  $\text{rank } p_{i0} = \text{rank } p_{0k} = \text{rank } a_{i0} = \text{rank } a_{0k} = 0$ , the first nonzero in column 1 and in row 1 of  $A$  will determine  $P_{i1} = 1$  and  $P_{1k} = 1$ .

In case the leading square submatrices  $a_{ii}$  are all nonsingular, which leads to  $\text{rank}(a_{ik}) = \min(i, k)$ , rule (2) puts all pivots on the diagonal:  $P_{ii} = 1$ . This is the case  $P = I$  with no row exchanges and  $A = LU$ .

Elimination by columns produces the same pivot positions (in a different sequence) as elimination by rows. For elimination with different starting points, and also for infinite matrices, rule (2) is to be adjusted. This rule that comes so simply from (1) is all-important.

The map  $P(A)$  from invertible matrices  $A$  to permutations in  $A = LPU$  (a map from  $GL_n$  to  $S_n$ ) is not continuous. We describe below how  $P$  can jump when  $A$  changes smoothly.

### 3. The Algebra of Elimination : $A = LPU = PLU$ .

Suppose elimination starts with  $A_{11} \neq 0$ , and all leading submatrices  $a_{ii}$  are invertible. Then we reach  $A = LU$  by familiar steps. For each  $j > 1$ , subtract a multiple  $\ell_{j1}$  of row 1 from row  $j$  to produce zero in the  $j, 1$  position. The next pivot position 2, 2 now contains the nonzero entry  $\det(a_{22})/\det(a_{11})$ : this is the second pivot.

Subtracting multiples  $\ell_{j2}$  of that second row produces zeros below the pivot in column 2. For  $k = 1, \dots, n$ , the  $k$ th pivot row becomes row  $k$  of  $U$ . The  $k, k$  pivot position contains the nonzero entry  $\det(a_{kk})/\det(a_{k-1,k-1})$ . For lower rows  $j > k$ , subtracting a multiple  $\ell_{jk}$  of row  $k$  from row  $j$  produces zero in the  $j, k$  position. Then the magic of elimination is that the matrix  $L$  of multipliers  $\ell_{jk}$  times the matrix  $U$  of pivot rows equals the original matrix  $A$ . Suppose  $n = 3$ :

$$A = LU \quad \begin{bmatrix} \text{row 1 of } A \\ \text{row 2 of } A \\ \text{row 3 of } A \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \ell_{21} & 1 & 0 \\ \ell_{31} & \ell_{32} & 1 \end{bmatrix} \begin{bmatrix} \text{row 1 of } U \\ \text{row 2 of } U \\ \text{row 3 of } U \end{bmatrix}. \quad (3)$$

The third row of that  $LU$  multiplication correctly states that

$$\text{row 3 of } U = (\text{row 3 of } A) - \ell_{31}(\text{row 1 of } U) - \ell_{32}(\text{row 2 of } U). \quad (4)$$

Now we face up to the possibility of zeros in one or more pivot positions. If  $a_{kk}$  is the first square upper left submatrix to be singular, the steps must change when elimination reaches column  $k$ . A lower row  $i(k)$  must become the  $k$ th pivot row. We have an algebraic choice and an algorithmic choice:

**Algebraic** Choose the first row  $i(k)$  that is not already a pivot row and has a nonzero entry in column  $k$  (to become the  $k$ th pivot). Subtract multiples of this pivot row  $i(k)$  to produce zeros in column  $k$  of all lower nonpivot rows. This completes step  $k$ .

**Note.** For  $A = LPU$ , the pivot row  $i(k)$  is *not moved immediately into row  $k$*  of the current matrix. It will indeed be row  $k$  of  $U$ , but it waits for the permutation  $P$  (with  $P_{i(k),k} = 1$ ) to put it there.

**Algorithmic** Choose any row  $I(k)$  that is not already a pivot row and has a nonzero entry in column  $k$ . Our choice of  $I(k)$  may maximize that pivot entry, or not. **Exchange** this new pivot row  $I(k)$  with the current row  $k$ . Subtract multiples of the pivot row to produce zeros in column  $k$  of all later rows.

**Note.** This process normally starts immediately at column 1, by choosing the row  $I(1)$  that maximizes the first pivot. Each pivot row  $I(k)$  moves immediately into row  $k$  of the current matrix and also row  $k$  of  $U$ .

The algebraic choice will lead to  $A = LPU$  and the algorithmic choice to  $A = PLU$ . If the choices coincide, so  $I(k) = i(k)$ , the multipliers will be the same numbers—but they appear in different positions in  $L$  and  $\mathbf{L}$  because row  $I(k)$  has been moved into row  $k$ . Then  $\mathbf{P} = P$  and  $\mathbf{U} = U$  and  $\mathbf{L} = P^{-1}LP$  from the reordering of the rows.

It is more than time for an example.

**Example:** The first pivot of  $A$  is in row  $i(1) = 2$ . The only elimination step is to subtract  $\ell$  times that first pivot row from row 3. This reveals the second pivot in row  $i(2) = 3$ . The order of pivot rows is 2, 3, 1 (and during  $LPU$  elimination they stay in that order!):

$$A = \begin{bmatrix} 0 & 0 & 3 \\ 1 & a & b \\ \ell & \ell a + 2 & \ell b + c \end{bmatrix} \xrightarrow{L^{-1}} \begin{bmatrix} 0 & 0 & 3 \\ 1 & a & b \\ 0 & 2 & c \end{bmatrix} = PU \quad (5)$$

The permutation  $P$  has 1's in the pivot positions. So its columns come from the identity matrix in the order 2, 3, 1 given by  $i(k)$ . Then  $U$  is upper triangular:

$$\begin{bmatrix} 0 & 0 & 3 \\ 1 & a & b \\ 0 & 2 & c \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & a & b \\ 0 & 2 & c \\ 0 & 0 & 3 \end{bmatrix} = PU \quad (6)$$

The lower triangular  $L$  adds  $\ell$  times row 2 of  $PU$  back to row 3 of  $PU$ . That entry  $L_{32} = \ell$  recovers the original  $A$  from  $PU$ :

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \ell & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 3 \\ 1 & a & b \\ 0 & 2 & c \end{bmatrix} = L(PU) = LPU \quad (7)$$

Notice the two zeros below the diagonal of  $L$ . We have  $L_{21} = 0$  because elimination did not subtract a multiple of row  $k = 1$  from row  $j = 2$ . (The first zero in row 1 of  $A$  is the reason that the first pivot was in row  $i(1) = 2$ .) In general  $L_{jk} = 0$  when row  $k$  is a pivot row after row  $j$  is a pivot row. Thus  $L_{jk} = 0$  when  $j > k$  but  $i^{-1}(j) < i^{-1}(k)$ .

The second zero below the diagonal of  $L$  is  $L_{31} = 0$ . Row  $k = 1$  is a pivot row after row  $j = 3$  is a pivot row. Rows 1, 2, 3 were selected as pivot rows in the order 3, 1, 2 given by the inverse of the permutation  $i(k)$ . Consequently  $i^{-1}(1) = 3$  is greater than  $i^{-1}(3) = 2$ .

*For computations.* This rule for zeros in  $L$  becomes important when we compare  $A = LPU$  with the form  $A = PLU$  that elimination codes prefer. When the permutation  $\mathbf{P}$  comes first, it is not unique. The only requirement is that  $\mathbf{P}^{-1}A$  admits an  $LU$  decomposition (the leading principal submatrices must be invertible, because they equal  $\mathbf{L}_k \mathbf{U}_k$ ). We may choose  $\mathbf{P}$  so that all entries of  $\mathbf{L}$  have  $|\mathbf{L}_{jk}| \leq 1$ . If  $|\ell| > 1$  in the matrix  $A$  above, row 3 would become the first pivot row instead of row 2. The multiplier that appears in  $\mathbf{L}$  would change to  $1/\ell$ . This “partial pivoting” aims to prevent small pivots and large multipliers and loss of accuracy.

The MATLAB command  $[B,U] = \ell u(A)$  constructs the upper triangular  $U = U$  and a permuted lower triangular  $B$ . If every step uses the first available pivot row (the algebraic choice), then  $B = LP = PL$ . The full command  $[L,U,\mathbf{inv}P] = lu(A)$  produces an (inverse) permutation for which  $(\mathbf{inv}P)A = LU$ . We can see this permutation as reordering the rows of  $A$  to prepare for a stable factorization.

*Back to algebra.* Consider  $A = PLU$  with no extra row exchanges:  $I(k) = i(k)$ . Then  $P$  and  $U$  are the same as  $P$  and  $U$  in the original  $A = LPU$ . But the lower triangular  $L$  is different from  $L$ . In fact  $PL = LP$  tells us directly that  $L = P^{-1}LP$ . *This reordered matrix  $L$  is still lower triangular.* It is this crucial property that uniquely identifies the specific  $L$  that is constructed by elimination. Other factors  $L$  can enter into  $A = LPU$ , but only the factor produced by elimination is “reduced from the left” with  $P^{-1}LP$  also lower triangular.

The uniqueness of this particular  $L$  is illustrated by an example with many possible  $L$ 's in  $A = LPU$ :

$$A = \begin{bmatrix} 0 & 1 \\ 1 & a \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \ell & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & u \\ 0 & 1 \end{bmatrix} \text{ provided } a = \ell + u. \quad (8)$$

Row 2 must be the first pivot row. There are no rows below that pivot row; the unique “reduced from the left” matrix is  $L = I$  with  $\ell = 0$ . (And  $P^{-1}IP = I$  is lower triangular as required.) To emphasize: All nonzero choices of  $\ell$  are permitted in  $A = LPU$  by choosing  $u = a - \ell$ . But that nonzero entry  $\ell$  will appear *above* the diagonal in  $P^{-1}LP$ . Elimination produced  $\ell = 0$  in the unique reduced factor  $L$ .

The difference between  $L$  and  $L$  in  $A = LPU$  and  $A = PLU$  can be seen in the 3 by 3 example. Both  $L$  and  $L = P^{-1}LP$  come from elimination, they contain the same entries, but these entries are moved around when  $P$  comes first in  $A = PLU$ .

**Example** (continued)  $L$  comes from elimination when the pivot rows of  $A$  are moved into 1, 2, 3 order in  $A = (\mathbf{inv}P)A$ :

$$A = \begin{bmatrix} 1 & a & b \\ \ell & \ell a + 2 & \ell b + c \\ 0 & 0 & 3 \end{bmatrix} \xrightarrow{L^{-1}} \begin{bmatrix} 1 & a & b \\ 0 & 2 & c \\ 0 & 0 & 3 \end{bmatrix} = U.$$

We subtracted  $\ell$  times row 1 from row 2, and  $L$  adds it back:

$$L = \begin{bmatrix} 1 & 0 & 0 \\ \ell & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This agrees with (7) after the reordering  $P^{-1}LP$ . The nonzero entry is still below the diagonal, confirming that the  $L$  chosen earlier is “reduced from the left.” No elimination steps were required to achieve zeros in the (3, 1) and (3, 2) positions, so  $L_{31} = L_{32} = 0$ . In terms of the original  $A$  rather than the reordered  $A$ ,  $L_{jk} = 0$  when  $i(j) < i(k)$ .

*To summarize:*  $A = LPU$  has a unique  $P$ , and a unique  $L$  reduced from the left. The permutation in  $A = PLU$  is not unique. But if we exchange rows only when necessary

to avoid zeros in the pivot positions,  $\mathbf{P}$  will agree with  $P$  and  $\mathbf{U} = U$ . The lower triangular  $\mathbf{L}$  in this better known form is  $P^{-1}LP$ .

*Elimination by column operations* To anticipate factorizations that are coming next, it is valuable (and satisfying) to recognize that “column elimination” is equally valid. In this brief digression, multiples of columns are subtracted from later columns. The result will be a *lower* triangular matrix  $L_c$ . Those column operations use *upper* triangular matrices multiplying from the right. The operations are inverted by an upper triangular matrix  $U_c$ .

When the pivot columns come in the natural order 1, 2, 3, elimination by columns produces  $A = L_c U_c$ . This is identical to  $A = LU$  from row operations, except that the pivots now appear in  $L_c$ . When we factor out the diagonal matrix  $D$  of pivots, the uniqueness of  $L$  and  $U$  (from rows) establishes the simple link to  $L_c$  and  $U_c$  from columns:

$$A = L_c U_c = (L_c D^{-1})(D U_c) = LU \quad (9)$$

In our 3 by 3 example, the first pivot (nonzero entry in row 1) is in column  $k(1) = 3$ . Then the second pivot (nonzero in the current row 2) is in column  $k(2) = 1$ . Column operations clear out row 2 in the remaining (later) pivot column  $k(3) = 2$ :

$$A = \begin{bmatrix} 0 & 0 & 3 \\ 1 & a & b \\ \ell & \ell a + 2 & \ell b + c \end{bmatrix} \xrightarrow{U_c^{-1}} \begin{bmatrix} 0 & 0 & 3 \\ 1 & 0 & b \\ \ell & 2 & \ell b + c \end{bmatrix} = L_c P_c \quad (10)$$

The permutation  $P_c$  has the *rows* of the identity matrix in the order 3, 1, 2 given by  $k(i)$ . Then  $L_c$  is lower triangular:

$$\begin{bmatrix} 0 & 0 & 3 \\ 1 & 0 & b \\ \ell & 2 & \ell b + c \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ b & 1 & 0 \\ \ell b + c & \ell & 2 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = L_c P_c \quad (11)$$

The constantly alert reader will recognize that  $k(i)$  is the inverse of  $i(k)$ . The permutation  $P_c$  must agree with  $P$  by uniqueness. The factorization  $A = L_c P_c U_c$  is completed when  $U_c$  undoes the column elimination by adding  $a$  times column 1 back to column 2:

$$A = \begin{bmatrix} 0 & 0 & 3 \\ 1 & 0 & b \\ \ell & 2 & \ell b + c \end{bmatrix} \begin{bmatrix} 1 & a & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = (L_c P_c) U_c = L_c P_c U_c \quad (12)$$

We could move  $P_c$  to the right in  $A = L_c U_c P_c$ . A permutation in this position could do extra column exchanges for the sake of numerical stability. (If  $|a| > 1$  in our example, columns 1 and 2 would be exchanged to keep all entries in  $U_c$  below 1.)

With  $P_c$  equal to  $P$ ,  $U_c P = P U_c$  means that  $U_c$  in the middle is  $P U_c P^{-1}$ . (The nonzero entry  $a$  moves to the 2, 3 position in  $U_c$ .) *This matrix is still upper triangular.* So  $U_c$  is “reduced from the right.” Under this condition the factors in  $A = L_c P_c U_c$  are uniquely determined by  $A$ .

In the 2 by 2 example those factors would move the nonzero entry from  $U$  (earlier) into  $L_c$  (now):

$$\begin{bmatrix} 0 & 1 \\ 1 & a \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ a & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = L_c P_c U_c. \quad (13)$$

*To summarize:* Column elimination produces different triangular factors from row elimination, but  $L$  still comes before  $U$ . In production codes, the practical difference would come from *accessing rows versus columns of  $A$* .

#### 4. Bruhat Decomposition and Bruhat Order

Choosing the 1, 1 entry as the starting point of elimination seems natural. Probably the Chinese who first described the algorithm [13, 20] felt the same. A wonderful history [10] by Grcar describes the sources from antiquity and then Newton’s “extermination” algorithm. (In lecture notes that he didn’t want published, Newton anticipated Rolle and Gauss.) But an algebraist can prefer to start at  $(n, 1)$ , and a hint at the reason needs only a few words.

$A = LPU$  is built on two subgroups (lower triangular and upper triangular) of the group  $GL_n$  of invertible  $n$  by  $n$  real matrices. There is an underlying equivalence relation:  $A \sim B$  if  $A = LBU$  for some triangular  $L$  and  $U$ . Thus  $GL_n$  is partitioned into equivalence classes. Because  $P$  was unique in Theorem 1, each equivalence class contains exactly one permutation (from the symmetric group  $S_n$  of all permutations). Very satisfactory but not perfect.

Suppose the two subgroups are the same (say the invertible upper triangular matrices). Now  $A \sim B$  means  $A = U_1 B U_2$  for some  $U_1$  and  $U_2$ . Again  $GL_n$  is partitioned into (new) equivalence classes, called “double cosets.” Again there is a single permutation matrix  $\pi$  in each double coset from  $A = U_1 \pi U_2$ . But now that the original subgroups are the same (here is the obscure hint, not to be developed further) we can multiply the double cosets and introduce an underlying algebra. The key point is that this “Bruhat decomposition” into double cosets  $U \pi U$  succeeds for a large and important class of algebraic groups (not just  $GL_n$ ).

Actually Bruhat did not prove this. His 1954 note [3] suggested the first ideas, which Harish-Chandra proved. Then Chevalley [5] uncovered the richness of the whole structure. George Lusztig gave more details of this (ongoing!) history in his lecture [14] at the Bruhat memorial conference in Paris.

One nice point, perhaps unsuspected by Bruhat, was the intrinsic partial order of the permutations  $\pi$ . Each  $\pi$  is shared by all the matrices  $U_1 \pi U_2$  in its double coset. We might expect the identity matrix  $\pi = I$  to come first in the “Bruhat order” but instead it comes last. For a generic  $n$  by  $n$  matrix, the permutation in  $A = U_1 \pi U_2$  will be the reverse identity matrix  $\pi = J$  corresponding to  $(n, \dots, 1)$ . Let me connect all these ideas to *upward elimination* starting with the  $n, 1$  entry of  $A$ .

The first steps subtract multiples of row  $n$  from the rows above, to produce zeros in the first column (above the pivot  $a_{n1}$ ). Assume that no zeros appear in the pivot positions along the reverse diagonal from  $n, 1$  to  $1, n$ . Then upward elimination ends

with zeros above the reverse diagonal :

$$\begin{bmatrix} & & \circledast \\ & \circledast & * \\ \circledast & * & * \end{bmatrix} = \begin{bmatrix} & & 1 \\ & 1 & \\ 1 & & \end{bmatrix} \begin{bmatrix} \circledast & * & * \\ & \circledast & * \\ & & \circledast \end{bmatrix} = JU_2. \quad (14)$$

The upward elimination steps are taken by upper triangular matrices. Those are inverted by an upper triangular  $U_1$  (containing all the multipliers). *This generic case has produced  $A = U_1 J U_2$ .* (Stewart suggested to denote the reversal matrix  $J$  by  $I$ .)

At stage  $k$  of Bruhat elimination, the pivot row is the *lowest* row that begins with exactly  $k - 1$  zeros. Then that stage produces zeros in column  $k$  for all other rows that began with  $k - 1$  zeros. These upward elimination steps end with a matrix  $\pi U_2$ , where the permutation  $\pi$  is decided by the order of the pivot rows. The steps are inverted by  $U_1$ , so the product  $U_1 \pi U_2$  recovers the original  $A$  and gives its Bruhat decomposition.

In the Bruhat partial order, the reverse identity  $J$  comes first and  $I$  comes last. The permutations  $P$  in  $A = LPU$ , from elimination that starts with  $A_{11}$ , fall naturally in the opposite order. These orders can be defined in many equivalent ways, and this is not the place for a full discussion. But one combinatorial definition fits perfectly with our “rank description” of the pivot positions in equation (2) :

In the Bruhat order for  $LPU$  decomposition (elimination starting at  $A_{11}$ ), two permutations have  $P \leq P'$  when all their upper left  $s$  by  $t$  submatrices have  $\text{rank}(p_{st}) \geq \text{rank}(p'_{st})$ .

**Example:**  $A_n = \begin{bmatrix} 1/n & 1 \\ 1 & 0 \end{bmatrix}$  has  $P_n = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$  but in the limit  $A_\infty = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = P_\infty$ . Here  $P_n < P_\infty$ .

The rank of the upper left 1 by 1 submatrix of  $A_n$  drops to zero in the limit  $A_\infty$ . Our (small) point is that this semicontinuity is always true : ranks can drop but not rise. The rank of a limit matrix never exceeds the limit (or lim inf) of the ranks. The connection between rank and Bruhat order leads quickly to a known conclusion about the map  $P(A)$  from  $A$  in  $\text{GL}_n$  to  $P$  in  $S_n$  :

**Theorem 2.** Suppose  $A_n = L_n P_n U_n$  approaches  $A_\infty = L_\infty P_\infty U_\infty$  and the permutations  $P_n$  approach a limit  $P$ . Then  $P \leq P_\infty$  in the Bruhat order for  $LPU$  (reverse of the usual Bruhat order for the  $\pi$ 's in  $U_1 \pi U_2$ ).

Roughly speaking,  $A_\infty$  may need extra row exchanges because ranks can drop.

## 5. Singly Infinite Banded Matrices

Our first step toward new ideas is to allow infinite matrices. We add the requirement that *the bandwidth  $w$  is finite*:  $A_{ij} = 0$  if  $|i - j| > w$ . Thus  $A$  is a “local” operator. Each row has at most  $2w + 1$  nonzeros. Each component in the product  $Ax$  needs at most  $2w + 1$  multiplications.

To start, assume that no finite combination of rows or of columns produces the zero vector (except the trivial combination). Elimination can begin at the 1, 1 position and



proceed forever. The output is a factorization into  $A = LPU$ . Those three factors are banded, but  $L$  and  $U$  are not necessarily bounded.

An example will show how far we are from establishing that  $L$  and  $U$  are bounded.  $A$  is block diagonal and each block  $B_k$  of  $A$  factors into  $L_k U_k$  with  $P_k = I$ :

$$B_k = \begin{bmatrix} \varepsilon_k & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \varepsilon_k^{-1} & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_k & -1 \\ 0 & \varepsilon_k^{-1} \end{bmatrix} = L_k U_k. \quad (15)$$

If  $\varepsilon_k$  approaches zero in a sequence of blocks of  $A$ , the pivots  $\varepsilon_k$  and  $\varepsilon_k^{-1}$  approach zero and infinity. The block diagonal matrices  $L$  and  $U$  (with blocks  $L_k$  and  $U_k$ ) are unbounded. At the same time  $A$  is bounded with bounded inverse:

$$\text{The blocks in } A^{-1} \text{ are } B_k^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & -\varepsilon_k \end{bmatrix}.$$

To regain control, assume in the rest of this section that  $A$  is **Toeplitz** or **block Toeplitz**. This time invariance or shift invariance is expressed by  $A_{ij} = A_{j-i}$ . The scalars or square blocks  $A_k$  are repeated down the  $k$ th diagonal. It would be hard to overstate the importance of Toeplitz matrices. They can be finite or infinite—in many ways doubly infinite is the simplest of all.

Examples will bring out the intimate link between the matrix  $A$  and its symbol  $a(z)$ , the polynomial in  $z$  and  $z^{-1}$  with coefficients  $A_k$ . Suppose  $A$  is tridiagonal ( $w = 1$ ):

$$A = \begin{bmatrix} 5 & -2 & & \\ -2 & 5 & -2 & \\ & -2 & 5 & \bullet \\ & & \bullet & \bullet \end{bmatrix} \quad \text{corresponds to } a(z) = -2z^{-1} + 5 - 2z.$$

With  $z = e^{i\theta}$ , the “symbol”  $a(e^{i\theta})$  becomes  $5 - 4\cos\theta$ . This is positive for all  $\theta$ , so  $A$  is positive definite. The symbol factors into  $a(z) = (2 - z)(2 - z^{-1}) = u(z)\ell(z)$ . The matrix factors in the same way into  $A = UL$  (and notice the order):

$$A = \begin{bmatrix} 2 & -1 & & \\ & 2 & -1 & \\ & & 2 & -1 \\ & & & \bullet \end{bmatrix} \begin{bmatrix} 2 & & & \\ -1 & 2 & & \\ & -1 & 2 & \\ & & -1 & \bullet \end{bmatrix} = UL. \quad (16)$$

This was a **spectral factorization** of  $a(z)$ , and a **Wiener-Hopf factorization**  $A = UL$ .

When elimination produces  $A = LU$  by starting in the 1, 1 position, the result is much less satisfying:  $L$  and  $U$  are not Toeplitz. (They are asymptotically Toeplitz and their rows eventually approach the good factors  $UL$ .)

One key point is that  $A = UL$  does not come from straightforward elimination—because an infinite matrix has no corner entry  $A_{nn}$  to start upward elimination. We factored  $a(z)$  instead.

Another key point concerns the location of the zeros of  $u(z) = 2 - z$  and  $\ell(z) = 2 - z^{-1}$ . Those zeros  $z = 2$  and  $z = 1/2$  satisfy  $|z| > 1$  and  $|z| < 1$  respectively. Then  $L$  and  $U$  have bounded inverses, and those Toeplitz inverses correspond to  $1/\ell(z)$  and  $1/u(z) = 1/(2 - z) = \frac{1}{2} + \frac{1}{4}z + \frac{1}{8}z^2 + \dots$ .

If we had chosen the factors badly,  $u(z) = 1 - 2z$  and  $\ell(z) = 1 - 2z^{-1}$  still produce  $a = u\ell$  and  $A = UL$ :

$$A = \begin{bmatrix} 1 & -2 & & \\ & 1 & -2 & \\ & & 1 & -2 \\ & & & \bullet \end{bmatrix} \begin{bmatrix} 1 & & & \\ -2 & 1 & & \\ & -2 & 1 & \\ & & -2 & \bullet \end{bmatrix} = UL. \quad (17)$$

The formal inverses of  $U$  and  $L$  have  $1, 2, 4, 8, \dots$  on their diagonals, because the zeros of  $u(z)$  and  $\ell(z)$  are inside and outside the unit circle—the wrong places.

Nevertheless  $U$  is a useful example. It has  $x = (1, \frac{1}{2}, \frac{1}{4}, \dots)$  in its nullspace:  $Ux = 0$  because  $u(\frac{1}{2}) = 0$ . This is a **Fredholm matrix** because the nullspaces of  $U$  and  $U^T$  are finite-dimensional. Notice that  $U^T = L$  has nullspace  $= \{0\}$ . The **Fredholm index** is the difference in the two dimensions:

$$\begin{aligned} \text{index}(U) &= \dim(\text{nullspace of } U) - \dim(\text{nullspace of } U^T) \\ &= 1 - 0. \end{aligned}$$

The index of  $L$  is  $-1$ ; the two nullspaces are reversed. The index of the product  $A = UL$  is  $1 - 1 = 0$ . In fact  $A$  is invertible, as the good factorization  $A = UL$  shows:

$$Ax = b \text{ is solved by } x = A^{-1}b = L^{-1}(U^{-1}b). \quad (18)$$

The key to invertibility is  $a(z) = u(z)\ell(z)$ , with the correct location of zeros to make  $U$  and  $L$  and thus  $A = UL$  invertible. The neat way to count zeros is to use the winding number of  $a(z)$ .

**Theorem 3.** If  $a(z) = \sum A_k z^k$  starts with  $A_{-m} z^{-m}$  and ends with  $A_M z^M$ , we need  $M$  zeros with  $|z| > 1$  and  $m$  zeros with  $|z| < 1$  (and no zeros with  $|z| = 1$ ). Then  $a(z) = u(z)\ell(z)$  and  $A = UL$  and those factors are invertible.

*The matrix case is harder.*  $A$  is now *block Toeplitz*. The  $A_k$  that go down diagonal  $k$  are square matrices, say  $b$  by  $b$ . It is still true (and all-important) that complete information about the operator  $A$  is contained in the matrix polynomial  $a(z) = \sum A_k z^k$ . The factorization of  $a(z)$  remains the crucial problem, leading as before to  $A = UL$ . Again this achieves “upward elimination without a starting point  $A_{nn}$ .”

The appropriate form for a matrix factorization is a product  $u p \ell$ :

$$a(z) = u(z)p(z)\ell(z) \text{ with } p(z) = \text{diag}(z^{k(1)}, \dots, z^{k(b)}).$$

The polynomial factor  $u(z)$  gives the banded upper triangular block Toeplitz matrix  $U$ . The third factor  $\ell(z)$  is a polynomial in  $z^{-1}$  and it produces  $L$ . The diagonal  $p(z)$  yields a block Toeplitz matrix  $P$ . (It will be a permutation matrix in the doubly infinite case, and we reach  $A = UPL$ .) The diagonal entry  $z^{k(j)}$  produces a 1 in the  $j$ th diagonal entry of the block  $P_k$  of  $P$ .

**Example.** Suppose the  $up\ell$  factorization of  $a(z)$  has  $\ell(z) = I$  :

$$a(z) = \begin{bmatrix} z^{-1} & 0 \\ 1 & z \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ z & 1 \end{bmatrix} \begin{bmatrix} z^{-1} & 0 \\ 0 & z \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \quad (19)$$

For *doubly* infinite block Toeplitz matrices, this gives  $A = \mathbf{UPL}$  with  $L = I$ . Then  $A$  is invertible. But for *singly* infinite matrices, the first row of  $\mathbf{UPL}$  is zero. You see success in rows 3-4, 5-6, ... which are not affected by the truncation to this singly infinite  $\mathbf{UPL}$  with  $L = I$  :

$$\begin{bmatrix} 1 & 0 & 0 & 0 & & & \\ 0 & 1 & 1 & 0 & & & \\ & & 1 & 0 & 0 & 0 & \\ & & 0 & 1 & 1 & 0 & \\ & & & & 1 & 0 & 0 & 0 \\ & & & & 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & & & \\ 0 & 0 & 0 & 1 & & & \\ 1 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 1 & \\ & & 1 & 0 & 0 & 0 & 0 & 0 \\ & & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \\ = \begin{bmatrix} 0 & 0 & 0 & 0 & & & \\ 1 & 0 & 0 & 1 & & & \\ 1 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 1 & 0 & 0 & 1 & \\ & & 1 & 0 & 0 & 0 & 0 & 0 \\ & & 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \begin{array}{l} \text{rows 3-4 of } A \\ \text{rows 5-6 of } A \end{array}$$

The missing nonzero in row 1 comes from the entry  $z^{-1}$  in  $p(z)$ . Invertibility of  $A$  in the singly infinite case requires all the exponents in  $p(z)$  to be  $k(j) = 0$ . Those “partial indices” give the dimensions of the nullspaces of  $A$  and  $A^T$  (here 1 and 1). Invertibility in the doubly infinite case only requires  $\sum k(j) = 0$ . In both cases this sum is the Fredholm index of  $A$  (here 0), equal to the winding number of  $\det a(z)$ .

The matrix factorization  $a(z) = u(z)p(z)\ell(z)$  has a long and very distinguished history. The first success was by Plemelj [17] in 1908. Hilbert and G.D. Birkhoff contributed proofs. Wiener and Hopf found wide applications to convolution equations on a half-line, by factoring  $A$  into  $UL$  when  $P = I$ . The algebraic side was developed by Grothendieck, and the analytic side by the greatest matrix theorist of the 20th century: Israel Gohberg. My favorite reference, for its clarity and its elementary constructive proof, is by Gohberg, Kaashoek, and Spitkovsky [9].

In the banded *doubly infinite* case, a bounded (and block Toeplitz) inverse only requires that  $a(z)$  is invertible on the unit circle:  $\det a(z) \neq 0$  for  $|z| = 1$ . Then  $a = up\ell$  and the reverse factorization  $a = \ell pu$  give  $A = \mathbf{UPL}$  and  $A = \mathbf{LPU}$  with invertible block Toeplitz matrices.  $P$  and  $\mathbf{P}$  are permutations of the integers.

All these are examples of triangular factorizations *when elimination has no starting point*. We presented them as the most important examples of their kind—when the periodicity of  $A$  reduced the problem to factorization of the matrix polynomial  $a(z)$ .

## 6. Elimination on Banded Doubly Infinite Matrices

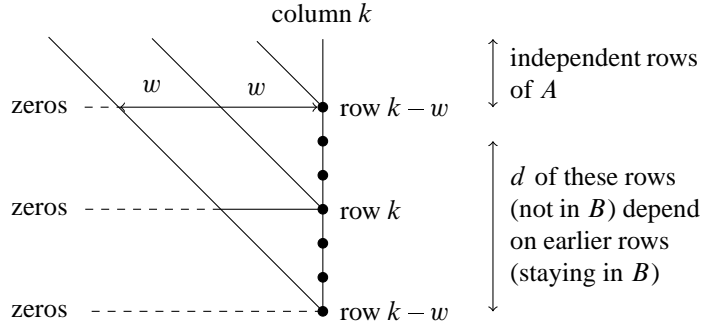
We have reached the question that you knew was coming. *How can elimination get started on a doubly infinite matrix?* To produce zeros in column  $k$ ,  $-\infty < k < \infty$ , we must identify the number  $i(k)$  of the pivot row. When that row is ready for use, its entries before column  $k$  are all zero. Multiples  $\ell_{ji}$  of this row are subtracted from lower rows  $j > i$ , to produce zeros below the pivot in column  $k$  of  $PU$ . The pivot row has become row  $i(k)$  of  $PU$ , and it will be row  $k$  of  $U$ .

Clearly  $i(k) \leq k + w$ , since all lower rows of a matrix with bandwidth  $w$  are zero up to and including column  $k$ . So the submatrix  $C(k)$  of  $A$ , containing all entries  $A_{ij}$  with  $i \leq k + w$  and  $j \leq k$ , controls elimination through step  $k$ . Rows below  $k + w$  and columns beyond  $k$  will not enter this key step: *the choice of pivot row  $i(k)$* .

We want to establish these facts in Lemma 1 and Lemma 2:

1.  $C(k)$  is a Fredholm matrix: *The nullspaces  $N(C)$  and  $N(C^T)$  are finite-dimensional*: Infinite matrices with this Fredholm property behave in important ways like finite matrices.
2. The index  $-d$  of  $C(k)$ , which is  $\dim N(C) - \dim N(C^T)$ , is independent of  $k$ .
3. In the step from  $C(k-1)$  to  $C(k)$ , the new  $k$ th column is independent of previous columns by the invertibility of  $A$ . (All nonzeros in column  $k$  of  $A$  are included in rows  $k-w$  to  $k+w$  of  $C(k)$ .) Since  $\text{index}(C(k)) = \text{index}(C(k-1))$ , the submatrix  $C(k)$  must contain exactly one row  $i(k)$  that is newly independent of the rows above it. Every integer  $i$  is eventually chosen as  $i(k)$  for some  $k$ .
4. Let  $B(k)$  be the submatrix of  $C(k)$  formed from all rows  $i(j)$ ,  $j \leq k$ . Each elimination step can be described non-recursively, in terms of the original matrix. We have removed the lowest possible  $d$  rows of  $C(k)$  to form this *invertible submatrix  $B(k)$* . Those  $d$  nonpivot rows are combinations of the rows of  $B(k)$ . Elimination subtracts those same combinations of the rows of  $A$  to complete step  $k$ . (The example below shows how these combinations lead to  $L^{-1}$ , where recursive elimination using only the pivot row (and not all of  $B$ ) leads directly to  $L$ .)

The figure shows the submatrix  $C(k)$ . Remaining  $d$  dependent rows leaves the invertible submatrix  $B(k)$ .



The 3 light diagonal lines are at  $45^\circ$  angles, with equal space between them.

5. When elimination is described recursively, the current row  $i(k)$  has all zeros before column  $k$ . It is row  $i(k)$  of  $PU$ . The multipliers  $\ell_{ji}$  will go into column  $i$  of a lower triangular matrix  $L$ , with  $L_{ii} = 1$ . Then  $A = LPU$  with  $P_{k,i(k)} = 1$  in the doubly infinite permutation matrix  $P$ . The pivot row becomes row  $k$  of the upper triangular  $U$ .

We may regard **5** as the execution of elimination, and **1, 2, 3, 4** as the key steps in selecting the pivot rows. Our whole argument will rest on the stability of the index, not changing with  $k$ .

**Lemma 1.**  $C(k)$  is a Fredholm matrix and its index is independent of  $k$ .

**Proof.** The invertible operator  $A$  is Fredholm with index  $\dim N(A) - \dim N(A') = 0 - 0$ . We are assuming that  $A$  is invertible on the infinite sequence space  $\ell_2(\mathbf{Z})$ . Key point : Perturbation by a finite rank matrix like  $D$ , or by any compact operator, leaves index = 0. By construction of  $C(k)$ ,

$$A = \begin{bmatrix} C(k) & D(k) \\ 0 & E(k) \end{bmatrix} \quad \text{and} \quad A' = \begin{bmatrix} C(k) & 0 \\ 0 & E(k) \end{bmatrix}$$

are Fredholm with equal index 0. For banded  $A$ , the submatrix  $D(k)$  contains only finitely many nonzeros (thus  $A - A'$  has finite rank). Clearly we can separate  $C$  from  $E$  :

$$A' = \begin{bmatrix} C(k) & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & E(k) \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & E(k) \end{bmatrix} \begin{bmatrix} C(k) & 0 \\ 0 & I \end{bmatrix}.$$

These two commuting factors are Fredholm since  $A'$  is Fredholm [8]. The indices of the two factors are equal to the indices of  $C(k)$  and  $E(k)$ . Those indices add to index  $(A') = \text{index}(A) = 0$ .

Since  $C(k-1)$  comes from  $C(k)$  by deleting one row and column, the index is the same! Now change  $k$ ! Strictly speaking, the last row and column of  $C(k)$  are replaced

by  $(\dots, 0, 0, 1)$ . This is a finite rank perturbation of  $C(k)$ : no change in the index. And the index of this matrix  $\text{diag}(C(k-1), 1)$  equals the index of  $C(k-1)$ .

Marko Lindner showed me this neat proof of Lemma 1, which he uses to define the “plus-index” and “minus-index” of the outgoing and incoming singly infinite submatrices  $A_+$  and  $A_-$  of  $A$ . These indices are independent of the cutoff position (row and column  $k$ ) between  $A_-$  and  $A_+$ . The rapidly growing theory of infinite matrices is described in [4, 13, 19].

**Lemma 2.** There is a unique row number  $i(k)$ , with  $|i - k| \leq w$ , such that

row  $i(k)$  of  $C(k-1)$  is a combination of previous rows of  $C(k-1)$

row  $i(k)$  of  $C(k)$  is *not* a combination of previous rows of  $C(k)$ .

**Proof.** By Lemma 1, the submatrices  $C(k)$  all share the same index  $-d$ . Each submatrix has nullspace  $= \{\mathbf{0}\}$ , since  $C(k)$  contains all nonzeros of all columns  $\leq k$  of the invertible matrix  $A$ . With index  $-d$ , the nullspace of every  $C(k)^T$  has dimension  $d$ . This means that  $d$  rows of  $C(k)$  are linear combinations of previous rows. Those  $d$  rows of  $C(k)$  must be among rows  $k-w+1, \dots, k+w$  (since the earlier rows of  $C(k)$  contain all nonzeros of the corresponding rows of the invertible matrix  $A$ ).

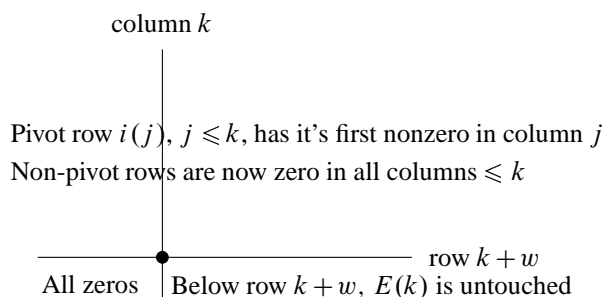
$C(k)$  has one new row and column compared to  $C(k-1)$ . Since  $d$  is the same for both, there must be one row  $i(k)$  that changes from dependent to independent when column  $k$  is included. In  $C(k-1)$ , that row was a combination of earlier pivot rows. In  $A$ , we can subtract that same combination of earlier rows from row  $i(k)$ . This leaves a row whose first nonzero is in column  $k$ . *This is the  $k$ th pivot row.*

Notice this pivot row was not constructed recursively (the used way). This row never changes again, it will be row  $i(k)$  of the matrix  $PU$  when elimination ends, and it will be row  $k$  of  $U$ . Once we have it, we can use its multiples  $\ell_{jk}$  for elimination below—and those numbers  $\ell_{jk}$  will appear in  $L$ . (The example below shows how the  $d$  dependencies lead to  $L^{-1}$ .)

Let  $A(k-1)$  denote the doubly infinite matrix after elimination is complete on columns  $< k$  of  $A$ . Row  $i(k)$  of  $A(k-1)$  is that  $k$ th pivot row. By subtracting multiples  $\ell_{ji}$  of this row from later non-pivot rows, we complete step  $k$  and reach  $A(k)$ . This matrix has zero in columns  $\leq k$  of all  $d$  rows that are combinations of earlier pivot rows. The multipliers are  $\ell_{ji} = 0$  for all rows  $j > k+w$ , since those rows (not in  $C(k)$ ) are and remain zero in all columns  $\leq k$ .

Each row is eventually chosen as a pivot row, because row  $k-w$  of  $C(k)$  has all the nonzeros of row  $k-w$  of  $A$ . That row cannot be a combination of previous rows when we reach step  $k$ ; it already was or now is a pivot row. The bandwidth  $w$  of the permutation  $P$  (associated with the ordering  $i(k)$  of the integers) is confirmed.

This completes the proof of **1, 2, 3, 4** and  $A = LPU$ .



$A(k)$ : current matrix at the end of step  $k$

**Toeplitz example with diagonals  $-2, 5, -2$  (now doubly infinite).** The correct choice of pivot rows is  $i(k) = k$  for all  $k$ . The invertible upper left submatrix  $B(k-1)$  has 5 along its diagonal. The matrix  $C(k-1)$  includes the dependent row  $k$  below ( $w = 1$  and  $d = 1$ ). To see the dependency, multiply rows  $k-1, k-2, k-3, \dots$  by  $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \dots$  and add to row  $k$ :

$$\left[ \begin{array}{ccc|cc} 5 & -2 & & & \\ -2 & 5 & -2 & & \\ & -2 & 5 & -2 & \\ \hline 0 & 0 & -2 & 5 & -2 \end{array} \right] \xrightarrow[k]{k-1} \left[ \begin{array}{ccc|cc} 5 & -2 & & & \\ -2 & 5 & -2 & & \\ & -2 & 5 & -2 & \\ \hline 0 & 0 & 0 & 4 & -2 \end{array} \right] \quad (20)$$

Row  $k$  of  $A$  has become row  $k$  of  $PU$  (also row  $k$  of  $U$ , since  $P = I$ ). The matrix  $L^{-1}$  that multiplies  $A$  to produce  $PU$  has those coefficients  $1, \frac{1}{2}, \frac{1}{4}, \dots$  leftward along each row. Then its inverse, which is  $L$ , has  $1, -\frac{1}{2}, 0, 0, \dots$  down each column.

This was non-recursive elimination. It produced the pivot row  $\dots, 0, 4, -2, 0, \dots$  by solving one infinite system. We can see normal recursive elimination by using this pivot row to remove the  $-2$  that still lies below the pivot 4. The multiplier is  $-\frac{1}{2}$ . This is the correct entry of  $L$ , found in the previous section by factoring the polynomial symbol  $a(z) = -2z^{-1} + 5 - 2z$ .

Suppose we make the *incorrect pivot choice*  $i(k) = k-1$  for all  $k$ . That gives  $P =$  doubly infinite shift. It leads to an  $LPU$  factorization of  $A$  that we don't want, with  $L = (A)$ (inverse shift) and  $P =$  (shift) and  $U = I$ . This lower triangular  $L$  has  $-2, 5, -2$  down each column. (To maintain the convention  $L_{ii} = 1$ , divide this  $L$  by  $-2$  and compensate with  $U = -2I$ .)

Recursively, this looks innocent. We are using the  $-2$ 's above the diagonal to eliminate each 5 and  $-2$  below them. But when the singly infinite submatrix in (20) loses its last row  $\dots, -2, 5$  (and becomes lower triangular with  $-2$  on its diagonal instead of 5), it is *no longer invertible*. The vector  $(\dots, \frac{1}{4}, \frac{1}{2}, 1)$  is in its nullspace. The correct choice had bidiagonal  $L$  and  $U$  as in (16).

In the language of Section 5, this lower triangular matrix has roots at 2 and  $\frac{1}{2}$ . It cannot have a bounded inverse. The misplaced root produces that vector in the nullspace.

**Theorem 4.** The nonzero entries of  $P, L, U$  lie in bands of width  $2w$  :

$$\begin{aligned} P_{ik} &= 0 & \text{if } |i - k| > w \\ L_{ik} &= 0 & \text{if } i - k > 2w \quad (\text{and if } i < k) \\ U_{ik} &= 0 & \text{if } k - i > 2w \quad (\text{and if } k < i) \end{aligned}$$

**Proof.** For finite matrices, the rank conditions (2) assure that  $P_{ik} = 1$  cannot happen outside the diagonal band  $|i - k| \leq w$  containing all nonzeros of  $A$ . Then

$$A = LPU \quad \text{gives} \quad L = AU^{-1}P^{-1} = AU^{-1}P^T.$$

The factor  $U^{-1}P^T$  cannot have nonzeros below subdiagonal  $w$ , since  $U^{-1}$  is upper triangular. Then  $L$  cannot have nonzeros below subdiagonal  $2w$ .

Similarly the matrices  $P^TL^{-1}$  and  $A$  are zero above superdiagonal  $w$ . So their product  $U = P^TL^{-1}A$  is zero above superdiagonal  $2w$ .

For infinite matrices, the choice of row  $i(k)$  as pivot row in Lemma 2 satisfies  $|i - k| \leq w$ . Thus  $P$  again has bandwidth  $w$ . The entries  $\ell_{ji}$  multiply this pivot row when it is subtracted from lower rows of  $C(k)$ . Since row  $k + w$  is the last row of  $C(k)$ , its distance from the pivot row cannot exceed  $2w$ .

Pivot rows cannot have more than  $2w$  nonzeros beyond the pivot. So when they move into  $U$  with the pivot on the diagonal,  $U$  cannot have nonzeros above superdiagonal  $2w$ .

The extreme cases are matrices with all nonzeros on subdiagonal and superdiagonal  $w$ . These show that the bands allowed by Theorem 4 can be attained.

## 7. Applications of $A = LPU$

In this informal final section, we comment on the doubly infinite  $A = LPU$  and a few of its applications.

**7.1** If  $A$  is a block Toeplitz matrix, so that  $A(i, j) = A(i + b, j + b)$  for all  $i$  and  $j$ , then  $L, P$ , and  $U$  will have the same block Toeplitz property. The multiplication  $A = LPU$  of doubly infinite matrices translates into a multiplication  $a(z) = \ell(z)p(z)u(z)$  of  $b$  by  $b$  matrix polynomials. Our result can be regarded as a new proof of that classical factorization.

This new proof is non-constructive because the steps from original rows (of  $A$ ) to pivot rows (of  $PU$ ) require the solution of singly-infinite systems with matrices  $B(k)$ . The constructive solution of those systems would require the Wiener-Hopf idea that is itself based on  $a(z) = u(z)p(z)\ell u(z)$ : a vicious circle.

**7.2 Infinite Gram-Schmidt.** From the columns  $a_1, \dots, a_n$  of an invertible matrix  $A$  we can produce the orthonormal columns  $q_1, \dots, q_n$  of  $Q$ . Normally each  $q_k$  is a combination of  $a_k$  and the preceding  $a_j$  (or equivalently the preceding  $q_j$ ,  $j < k$ ). Then  $A$  is factored into  $Q$  times an upper triangular. The question is how to start the process when  $A$  is doubly infinite.



Notice that  $Q^T Q = I$  leads to  $A^T A = (QR)^T (QR) = R^T R$ . This is a special  $LU$  factorization (Cholesky factorization) of the symmetric positive definite matrix  $A^T A$ . The factors  $R^T$  and  $R$  will have the same main diagonal, containing the square roots of the pivots of  $A^T A$  (which are all positive).

If  $A$  is doubly infinite and banded, so is  $A^T A$ . Then the factorization in Section 6 produces  $R^T R$ . The submatrices  $B(k)$  in the proof share the main diagonal of  $A^T A$ , and  $\|B(k)^{-1}\| \leq \| (A^T A)^{-1} \|$ . No permutation  $P$  is needed and we reach the banded matrix  $R$ .

Finally  $Q = AR^{-1}$  has orthonormal columns  $q_k$  as required. Each  $q_k$  is a combination of the original  $a_j$ ,  $j \leq k$ .  $Q$  is banded below its main diagonal but not above—apart from the exceptional cases when  $R$  is banded with banded inverse.

**7.3** Theorem 1 came from the observation that the upper left submatrices of  $A$ ,  $L$ ,  $P$ ,  $U$  satisfy  $a = \ell pu$ . With doubly infinite matrices and singly infinite submatrices, this remains true. The ranks of diagonal blocks  $A_+$  and  $A_-$  are now infinite, so we change to nullities. But as the block diagonal example in Section 5 made clear,  $L$  and  $U$  and their inverses may not be bounded operators. At this point the uniqueness of  $P$  comes from its construction (during elimination) and not from Theorem 1.

**7.4** In recent papers we studied the group of banded matrices *with banded inverses* [21-23]. These very special matrices are products  $A = F_1 \dots F_N$  of block diagonal invertible matrices. Our main result was that  $A = F_1 F_2$  if we allow blocks of size  $2w$ , and then  $N \leq Cw^2$  when the blocks have size  $\leq 2$ . The key point is that the number  $N$  of block diagonal factors is controlled by  $w$  and not by the size of  $A$ . The proof uses elimination and  $A$  can be singly infinite.

We have no proof yet when  $A$  is doubly infinite. It is remarked in [23] that  $A = LPU$  reduces the problem to banded triangular matrices  $L$  and  $U$  with banded inverses. We mention Panova's neat factorization [16] of  $P$  (whose inverse is  $P^T$ ). With bandwidth  $w$ , a singly infinite  $P$  is the product of  $N < 2w$  parallel exchanges of neighbors (block diagonal permutations with block size  $\leq 2$ ).

A doubly infinite  $P$  will require a power of the infinite shift matrix  $S$ , in addition to  $F_1 \dots F_N$ . This power  $s(P)$  is the "shifting index" of  $P$  and  $|s| \leq w$ . The main diagonal is not defined for doubly infinite matrices, until the shifting index  $s(A) = s(P)$  tells us where it ought to be. This agrees with the main diagonal located by de Boor [ ] for a particular family of infinite matrices.

**7.5** For singly infinite Fredholm matrices the main diagonal is well defined. It is located by the Fredholm index of  $A$ . When the index is zero, the main diagonal is in the right place. (Still  $A$  may or may not be invertible. For a block Toeplitz matrix invertibility requires all partial indices  $k(j)$  to be zero, not just their sum.)

The proof of Lemma 1 showed why the Fredholm indices of the incoming  $A_-$  and outgoing  $A_+$  are independent of the cutoff position (row and column  $k$ ). When  $A$  is invertible, that "minus-index" and "plus-index" add to zero. The connection to the shifting index was included in [23].

**Theorem 5.** The shifting index of a banded invertible matrix  $A$  (and of its permutation  $P$ ) equals the Fredholm index of  $A_+$  (the plus-index).

Check when  $A$  is the doubly infinite shift matrix  $S$  with nonzero entries  $S_{i,i+1} = 1$ . Then  $P$  coincides with  $S$  and has shifting index 1 (one  $S$  in its factorization into bandwidth 1 matrices). The outgoing submatrix  $A_+$  is a singly infinite shift with  $(1, 0, 0, \dots)$  in its nullspace. Then  $A_+^T x = 0$  only for  $x = 0$ , so the Fredholm index of  $A_+$  is also 1.

A deep result from the theory of infinite matrices [18, 19] concerns the Fredholm indices of the *limit operators* of  $A$ .

**7.6** I would like to end with a frightening example. It shows that the associative law  $A(Bx) = (AB)x$  can easily fail for infinite matrices. I always regarded this as the most fundamental and essential law! It defines  $AB$  (by composition), and it is the key to so many short and important proofs that I push my linear algebra classes to recognize and even anticipate a “proof by moving the parentheses.”

The example has  $Bx = 0$  but  $AB = I$ . And  $0 = A(Bx) = (AB)x = x$  is false.

$$A = \begin{bmatrix} 1 & 1 & 1 & \bullet \\ 0 & 1 & 1 & \bullet \\ 0 & 0 & 1 & \bullet \\ 0 & 0 & 0 & \bullet \end{bmatrix} \quad B = \begin{bmatrix} 1 & -1 & 0 & \bullet \\ 0 & 1 & -1 & \bullet \\ 0 & 0 & 1 & \bullet \\ 0 & 0 & 0 & \bullet \end{bmatrix} \quad x = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \bullet \end{bmatrix} \quad (21)$$

This is like the integral of the derivative of a constant.  $A$  is an unbounded operator, the source of unbounded difficulty. A direct proof of the law  $A(Bx) = (AB)x$  would involve rearranging series. Riemann showed us that without absolute convergence, which is absent here, all sums are possible if  $a_n \rightarrow 0$ .

This example has led me to realize that grievous errors are all too possible with infinite matrices. I hope this paper is free of error. But when elimination has no starting point (and operator theory is not developed in detail), it is wise to be prepared for the worst.

### Acknowledgements

This paper was begun on a visit to the beautiful Hong Kong University of Science and Technology. We thank its faculty and its president Tony Chan for their warm welcome. On my final morning, George Lusztig clarified the history of Bruhat. The books of Marko Lindner (and our email correspondence) introduced me to the limit analysis of non-Toeplitz infinite matrices.

## References

1. C. Albert, C.-K. Li, G. Strang and G. Yu, *Permutations as products of parallel transpositions*, submitted to SIAM J. Discrete Math (2010).
2. E. Asplund, *Inverses of matrices  $\{a_{ij}\}$  which satisfy  $a_{ij} = 0$  for  $j > i + p$* , Math. Scand. 7 (1959) 57-60.
3. F. Bruhat, *Representations induites des groupes de Lie semisimples complexes*, Comptes Rendus Acad. Sci. Paris 238 (1954), 437-439.
4. S.N. Chandler-Wilde and M. Lindner, *Limit Operators, Collective Compactness, and the Spectral Theory of Infinite Matrices*, Amer. Math. Society Memoirs 210 (2011) no. 989.
5. C. Chevalley, *Sur certains groupes simples*, Tohoku Math. J. 7 (1955).
6. C. de Boor,
7. L. Elsner, *On some algebraic problems in connection with general eigenvalue algorithms*, Lin. Alg. Appl. 26 (1979) 123-138.
8. I. Gohberg and S. Goldberg, *Finite dimensional Wiener-Hopf equations and factorizations of matrices*, Lin. Alg. Appl. 48 (1982) 219-236.
9. I. Gohberg, S. Goldberg, and M.A. Kaashoek, *Basic Classes of Linear Operators*, Birkhäuser, 2003.
10. I. Gohberg, M. Kaashoek, and I. Spitkovsky. *An overview of matrix factorization theory and operator applications*, Operator Th. Adv. Appl., Birkhäuser (2003) 1-102.
11. J.F. Grcar, *How ordinary elimination became Gaussian elimination*, Historia Math. 38 (2011).
12. Harish-Chandra, *On a lemma of Bruhat*, J. Math. Pures Appl. 35 (1956), 203-210.
13. R. Hart, *The chinese Roots of Linear Algebra*, Johas Hopkins Univ. Press, 2011.
14. L. Yu. Kolotilina and A. Yu. Yeremin, *Bruhat decomposition and solution of sparse linear algebra systems*, Soviet J. Numer. Anal. Math. Modelling 2 (1987) 421-436.
15. M. Lindner, *Infinite Matrices and Their Finite Sections*, Birkhäuser, 2006.
16. G. Lusztig, Bruhat decomposition and applications, [math.mit.edu/~gyuri](http://math.mit.edu/~gyuri).
17. V. Olshevsky, P. Zhlobich, and G. Strang, *Green's matrices*, Lin. Alg. Appl. 432 (2010) 218-241.
18. G. Panova, *Factorization of banded permutations*, [arXiv.org/abs/1007.1760](http://arXiv.org/abs/1007.1760) (2010); Proceedings Amer. Math. Soc., to appear.

19. J. Plemelj, *Riemannsche Funktionenscharen mit gegebener Monodromiegruppe*, Monat. Math. Phys. 19 (1908) 211-245.
20. V.S. Rabinovich, S. Roch, and J. Roe, *Fredholm indices of band-dominated operators*, Integral Eqns. Oper. Th. 49 (2004) 221-238.
21. V.S. Rabinovich, S. Roch, and B. Silbermann, *The finite section approach to the index formula for band-dominated operators*, Operator Theory 187 (2008) 185-193.
22. K. Shen, J.N. Crossley, A.W.C. Lun, *The Nine Chapters of the Mathematical Art*, Oxford University Press, 1999.
23. G. Strang, *Fast transforms: Banded matrices with banded inverses*, Proc. Natl. Acad. Sci. 107 (2010) 12413-12416.
24. G. Strang, *Groups of banded matrices with banded inverses*, Proceedings Amer. Math. Soc., to appear.
25. G. Strang, *Banded matrices with banded inverses and  $A = LPU$* , International Congress of Chinese Mathematicians, Beijing, 2010. Proceedings to appear.

Let me add a (non-infinite!) example with  $i(1) = 2$ . The matrix  $C(1)$  contains the first column of  $A$ . Its 1 by 1 submatrix  $B(1) = [1]$  is invertible. The multiplier  $\ell_{32} = 4$  goes into  $L$ , in the first and only elimination step:

$$\begin{aligned}
 A &= \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 4 & 8 & 13 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 0 & 0 & 1 \end{bmatrix} = L(PU) \\
 &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix}.
 \end{aligned}$$