MATRICES, BLOCK STRUCTURES AND GAUSSIAN ELIMINATION

Numerical linear algebra lies at the heart of modern scientific computing and computational science. Today it is not uncommon to perform numerical computations with matrices having millions of components. The key to understanding how to implement such algorithms is to exploit underlying structure within the matrices. In these notes we touch on a few ideas and tools for dissecting matrix structure. Specifically we are concerned with the *block structure* matrices.

1. Rows and Columns

Let $A \in \mathbb{R}^{m\times n}$ so that A has m rows and n columns. Denote the element of A in the ith row and jth column as A_{ij}. Denote the m rows of A by $A_{1\cdot},\ A_{2\cdot},\ A_{3\cdot},\ \dots, A_{m\cdot}$ and the n columns of A by $A_{\cdot 1},\ A_{\cdot 2},\ A_{\cdot 3},\ \dots, A_{\cdot n}$. For example, if

$$A = \begin{bmatrix} 3 & 2 & -1 & 5 & 7 & 3 \\ -2 & 27 & 32 & -100 & 0 & 0 \\ -89 & 0 & 47 & 22 & -21 & 33 \end{bmatrix},$$

then $A_{2,4} = -100$,

$$A_{1\cdot} = \begin{bmatrix} 3 & 2 & -1 & 5 & 7 & 3 \end{bmatrix},\ A_{2\cdot} = \begin{bmatrix} -2 & 27 & 32 & -100 & 0 & 0 \end{bmatrix},\ A_{3\cdot} = \begin{bmatrix} -89 & 0 & 47 & 22 & -21 & 33 \end{bmatrix}$$

and

$$A_{\cdot 1} = \begin{bmatrix} 3 \\ -2 \\ -89 \end{bmatrix},\ A_{\cdot 2} = \begin{bmatrix} 2 \\ 27 \\ 0 \end{bmatrix},\ A_{\cdot 3} = \begin{bmatrix} -1 \\ 32 \\ 47 \end{bmatrix},\ A_{\cdot 4} = \begin{bmatrix} 5 \\ -100 \\ 22 \end{bmatrix},\ A_{\cdot 5} = \begin{bmatrix} 7 \\ 0 \\ -21 \end{bmatrix},\ A_{\cdot 6} = \begin{bmatrix} 3 \\ 0 \\ 33 \end{bmatrix}.$$

Exercise 1.1. *If*

$$C = \begin{bmatrix} 3 & -4 & 1 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 & 1 & 4 \\ 0 & 0 & 0 & 1 & 0 & 3 \end{bmatrix},$$

what are $C_{4,4}$, $C_{\cdot 4}$ and $C_{4\cdot}$? For example, $C_{2\cdot} = \begin{bmatrix} 2 & 2 & 0 & 0 & 1 & 0 \end{bmatrix}$ and $C_{\cdot 2} = \begin{bmatrix} -4 \\ 2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$.

The block structuring of a matrix into its rows and columns is of fundamental importance and is extremely useful in understanding the properties of a matrix. In particular, for $A \in \mathbb{R}^{m\times n}$ it allows us to write

$$A = \begin{bmatrix} A_{1\cdot} \\ A_{2\cdot} \\ A_{3\cdot} \\ \vdots \\ A_{m\cdot} \end{bmatrix} \text{ and } A = \begin{bmatrix} A_{\cdot 1} & A_{\cdot 2} & A_{\cdot 3} & \dots & A_{\cdot n} \end{bmatrix}.$$

These are called the row and column block representations of A, respectively

1.1. *Matrix vector Multiplication.* Let $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$. In terms of its coordinates (or components), we can also write $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ with each $x_j \in \mathbb{R}$. The term x_j is called the jth component of x. For example if

$$x = \begin{bmatrix} 5 \\ -100 \\ 22 \end{bmatrix},$$

then $n = 3,\ x_1 = 5,\ x_2 = -100,\ x_3 = 22$. We define the matrix-vector product Ax by

$$Ax = \begin{bmatrix} A_{1\cdot} \bullet x \\ A_{2\cdot} \bullet x \\ A_{3\cdot} \bullet x \\ \vdots \\ A_{m\cdot} \bullet x \end{bmatrix},$$

where for each $i = 1, 2, \ldots, m$, $A_{i\cdot} \bullet x$ is the dot product of the ith row of A with x and is given by

$$A_{i\cdot} \bullet x = \sum_{j=1}^{n} A_{ij} x_j \ .$$

For example, if

$$A = \begin{bmatrix} 3 & 2 & -1 & 5 & 7 & 3 \\ -2 & 27 & 32 & -100 & 0 & 0 \\ -89 & 0 & 47 & 22 & -21 & 33 \end{bmatrix} \text{ and } x = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 2 \\ 3 \end{bmatrix},$$

then

$$Ax = \begin{bmatrix} 24 \\ -29 \\ -32 \end{bmatrix}.$$

Exercise 1.2. *If*

$$C = \begin{bmatrix} 3 & -4 & 1 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 2 & 1 & 4 \\ 0 & 0 & 0 & 1 & 0 & 3 \end{bmatrix} \ \textit{and}\ x = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 2 \\ 3 \end{bmatrix},$$

what is Cx?

Note that if $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$, then Ax is always well defined with $Ax \in \mathbb{R}^m$. In terms of components, the ith component of Ax is given by the dot product of the ith row of A (i.e. $A_{i\cdot}$) and x (i.e. $A_{i\cdot} \bullet x$).

The view of the matrix-vector product described above is the *row-space* perspective, where the term *row-space* will be given a more rigorous definition at a later time. But there is a very different way of viewing the matrix-vector product based on a *column-space* perspective. This view uses the notion of the linear combination of a collection of vectors.

Given k vectors $v^1, v^2, \ldots, v^k \in \mathbb{R}^n$ and k scalars $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}$, we can form the vector

$$\alpha_1 v^1 + \alpha_2 v^2 + \cdots + \alpha_k v^k \in \mathbb{R}^n \ .$$

Any vector of this kind is said to be a *linear combination* of the vectors $v^1, v^2, \ldots, v^k$ where the $\alpha_1, \alpha_2, \ldots, \alpha_k$ are called the coefficients in the linear combination. The set of all such vectors formed as linear combinations of $v^1, v^2, \ldots, v^k$ is said to be the *linear span* of $v^1, v^2, \ldots, v^k$ and is denoted

$$\operatorname{span}\left(v^1, v^2, \ldots, v^k\right) := \left\{ \alpha_1 v^1 + \alpha_2 v^2 + \cdots + \alpha_k v^k \;\middle|\; \alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R} \right\}.$$

Returning to the matrix-vector product, one has that

$$Ax = \begin{bmatrix} A_{11}x_1 + A_{12}x_2 + A_{13}x_3 + \cdots + A_{1n}x_n \\ A_{21}x_1 + A_{22}x_2 + A_{23}x_3 + \cdots + A_{2n}x_n \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ A_{m1}x_1 + A_{m2}x_2 + A_{m3}x_3 + \cdots + A_{mn}x_n \end{bmatrix} = x_1 A_{\cdot 1} + x_2 A_{\cdot 2} + x_3 A_{\cdot 3} + \cdots + x_n A_{\cdot n},$$

which is a linear combination of the columns of A. That is, we can view the matrix-vector product Ax as taking a linear combination of the columns of A where the coefficients in the linear combination are the coordinates of the vector x.

We now have two fundamentally different ways of viewing the matrix-vector product Ax.

Row-Space view of Ax:

$$Ax = \begin{bmatrix} A_{1\cdot} \bullet x \\ A_{2\cdot} \bullet x \\ A_{3\cdot} \bullet x \\ \vdots \\ A_{m\cdot} \bullet x \end{bmatrix}$$

Column-Space view of Ax:

$$Ax = x_1 A_{\cdot 1} + x_2 A_{\cdot 2} + x_3 A_{\cdot 3} + \cdots + x_n A_{\cdot n} \ .$$

2. Matrix Multiplication

We now build on our notion of a matrix-vector product to define a notion of a matrix-matrix product which we call *matrix multiplication*. Given two matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times k}$ note that each of the columns of B resides in $\mathbb{R}^n$, i.e. $B_{\cdot j} \in \mathbb{R}^n$ $i = 1, 2, \ldots, k$. Therefore, each of the matrix-vector products $AB_{\cdot j}$ is well defined for $j = 1, 2, \ldots, k$. This allows us to define a matrix-matrix product that exploits the block column structure of B by setting

$$AB := \begin{bmatrix} AB_{\cdot 1} & AB_{\cdot 2} & AB_{\cdot 3} & \cdots & AB_{\cdot k} \end{bmatrix} . \tag{1}$$

Note that the jth column of AB is $(AB)_{\cdot j} = AB_{\cdot j} \in \mathbb{R}^m$ and that $AB \in \mathbb{R}^{m\times k}$, i.e.

if $H \in \mathbb{R}^{m\times n}$ and $L \in \mathbb{R}^{n\times k}$, then $HL \in \mathbb{R}^{m\times k}$.

Also note that

if $T \in \mathbb{R}^{s\times t}$ and $M \in \mathbb{R}^{r\times \ell}$, then the matrix product TM is only defined when $t = r$.

For example, if

$$A = \begin{bmatrix} 3 & 2 & -1 & 5 & 7 & 3 \\ -2 & 27 & 32 & -100 & 0 & 0 \\ -89 & 0 & 47 & 22 & -21 & 33 \end{bmatrix} \text{ and } B = \begin{bmatrix} 2 & 0 \\ -2 & 2 \\ 0 & 3 \\ 0 & 0 \\ 1 & 1 \\ 2 & -1 \end{bmatrix},$$

then

$$AB = \begin{bmatrix} A \begin{bmatrix} 2 \\ -2 \\ 0 \\ 0 \\ 1 \\ 2 \end{bmatrix} & A \begin{bmatrix} 0 \\ -2 \\ 3 \\ 0 \\ 1 \\ -1 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 15 & 5 \\ -58 & 150 \\ -133 & 87 \end{bmatrix}.$$

Exercise 2.1. *if*

$$C = \begin{bmatrix} 3 & -4 & 1 & 1 \\ 2 & 2 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \end{bmatrix} \textit{ and } D = \begin{bmatrix} -1 & 0 & 2 & 4 & 3 \\ 0 & -2 & -1 & 4 & 5 \\ 5 & 2 & -4 & 1 & 1 \\ 3 & 0 & 1 & 0 & 0 \end{bmatrix},$$

is CD well defined and if so what is it?

The formula (1) can be used to give further insight into the individual components of the matrix product AB. By the definition of the matrix-vector product we have for each $j = 1, 2, \ldots, k$

$$AB_{\cdot j} = \begin{bmatrix} A_{1\cdot} \bullet B_{\cdot j} \\ A_{2\cdot} \bullet B_{\cdot j} \\ A_{m\cdot} \bullet B_{\cdot j} \end{bmatrix}.$$

Consequently,

$$(AB)_{ij} = A_{i\cdot} \bullet B_{\cdot j} \quad \forall\, i = 1, 2, \ldots m,\ j = 1, 2, \ldots, k.$$

That is, the element of AB in the ith row and jth column, $(AB)_{ij}$, is the dot product of the ith row of A with the jth column of B.

2.1. **Elementary Matrices.** We define the *elementary unit coordinate matrices* in $\mathbb{R}^{m\times n}$ in much the same way as we define the elementary unit coordinate vectors. Given $i \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$, the elementary unit coordinate matrix $E_{ij} \in \mathbb{R}^{m\times n}$ is the matrix whose ij entry is 1 with all other entries taking the value zero. This is a slight abuse of notation since the notation E_{ij} is supposed to represent the ijth entry in the matrix E. To avoid confusion, we reserve the use of the letter E when speaking of matrices to the elementary matrices.

Exercise 2.2. *(Multiplication of square elementary matrices)Let $i, k \in \{1, 2, \ldots, m\}$ and $j, \ell \in \{1, 2, \ldots, m\}$. Show the following for elementary matrices in $\mathbb{R}^{m\times m}$ first for $m = 3$ and then in general.*

(1) $E_{ij}E_{k\ell} = \begin{cases} E_{i\ell} & \text{, if } j = k, \\ 0 & \text{, otherwise.} \end{cases}$

(2) For any $\alpha \in \mathbb{R}$, if $i \neq j$, then $(I_{m\times m} - \alpha E_{ij})(I_{m\times m} + \alpha E_{ij}) = I_{m\times m}$ so that

$$(I_{m\times m} + \alpha E_{ij})^{-1} = (I_{m\times m} - \alpha E_{ij}).$$

(3) For any $\alpha \in \mathbb{R}$ with $\alpha \neq 0$, $(I + (\alpha^{-1} - 1)E_{ii})(I + (\alpha - 1)E_{ii}) = I$ so that

$$(I + (\alpha - 1)E_{ii})^{-1} = (I + (\alpha^{-1} - 1)E_{ii}).$$

Exercise 2.3. *(Elementary permutation matrices)Let $i, \ell \in \{1, 2, \ldots, m\}$ and consider the matrix $P_{ij} \in \mathbb{R}^{m\times m}$ obtained from the identity matrix by interchanging its i and ℓth rows. We call such a matrix an elementary permutation matrix. Again we are abusing notation, but again we reserve the letter P for permutation matrices (and, later, for projection matrices). Show the following are true first for $m = 3$ and then in general.*

(1) $P_{i\ell}P_{i\ell} = I_{m\times m}$ so that $P_{i\ell}^{-1} = P_{i\ell}$.
(2) $P_{i\ell}^T = P_{i\ell}$.
(3) $P_{i\ell} = I - E_{ii} - E_{\ell\ell} + E_{i\ell} + E_{\ell i}$.

Exercise 2.4. *(Three elementary row operations as matrix multiplication)In this exercise we show that the three elementary row operations can be performed by left multiplication by an invertible matrix. Let $A \in \mathbb{R}^{m\times n}$, $\alpha \in \mathbb{R}$ and let $i, \ell \in \{1, 2, \ldots, m\}$ and $j \in \{1, 2, \ldots, n\}$. Show that the following results hold first for $m = n = 3$ and then in general.*

(1) (row interchanges) Given $A \in \mathbb{R}^{m\times n}$, the matrix $P_{ij}A$ is the same as the matrix A except with the i and jth rows interchanged.
(2) (row multiplication) Given $\alpha \in \mathbb{R}$ with $\alpha \neq 0$, show that the matrix $(I + (\alpha - 1)E_{ii})A$ is the same as the matrix A except with the ith row replaced by α times the ith row of A.
(3) Show that matrix $E_{ij}A$ is the matrix that contains the jth row of A in its ith row with all other entries equal to zero.
(4) (replace a row by itself plus a multiple of another row) Given $\alpha \in \mathbb{R}$ and $i \neq j$, show that the matrix $(I + \alpha E_{ij})A$ is the same as the matrix A except with the ith row replaced by itself plus α times the jth row of A.

2.2. **Associativity of matrix multiplication.** Note that the definition of matrix multiplication tells us that this operation is associative. That is, if $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times k}$, and $C \in \mathbb{R}^{k\times s}$, then $AB \in \mathbb{R}^{m\times k}$ so that $(AB)C$ is well defined and $BC \in \mathbb{R}^{n\times s}$ so that $A(BC)$ is well defined, and, moreover,

$$(AB)C = \begin{bmatrix}(AB)C_{\cdot 1} & (AB)C_{\cdot 2} & \cdots & (AB)C_{\cdot s}\end{bmatrix} \tag{2}$$

where for each $\ell = 1, 2, \ldots, s$

$$\begin{aligned}(AB)C_{\cdot\ell} &= \begin{bmatrix}AB_{\cdot 1} & AB_{\cdot 2} & AB_{\cdot 3} & \cdots & AB_{\cdot k}\end{bmatrix} C_{\cdot\ell} \\ &= C_{1\ell}AB_{\cdot 1} + C_{2\ell}AB_{\cdot 2} + \cdots + C_{k\ell}AB_{\cdot k} \\ &= A\left[C_{1\ell}B_{\cdot 1} + C_{2\ell}B_{\cdot 2} + \cdots + C_{k\ell}B_{\cdot k}\right] \\ &= A(BC_{\cdot\ell})\ .\end{aligned}$$

Therefore, we may write (2) as

$$\begin{aligned}(AB)C &= \begin{bmatrix}(AB)C_{\cdot 1} & (AB)C_{\cdot 2} & \cdots & (AB)C_{\cdot s}\end{bmatrix} \\ &= \begin{bmatrix}A(BC_{\cdot 1}) & A(BC_{\cdot 2}) & \ldots & A(BC_{\cdot s})\end{bmatrix} \\ &= A\begin{bmatrix}BC_{\cdot 1} & BC_{\cdot 2} & \ldots & BC_{\cdot s}\end{bmatrix} \\ &= A(BC)\ .\end{aligned}$$

Due to this associativity property, we may dispense with the parentheses and simply write ABC for this triple matrix product. Obviously longer products are possible.

Exercise 2.5. *Consider the following matrices:*

$$A = \begin{bmatrix}2 & 3 & 1\\ 1 & 0 & -3\end{bmatrix} \quad B = \begin{bmatrix}4 & -1\\ 0 & -7\end{bmatrix} \quad C = \begin{bmatrix}-2 & 3 & 2\\ 1 & 1 & -3\\ 2 & 1 & 0\end{bmatrix}$$

$$D = \begin{bmatrix}2 & 3\\ 1 & 0\\ 8 & -5\end{bmatrix} \quad F = \begin{bmatrix}2 & 1 & 1 & 2\\ 1 & 0 & -4 & 0\\ 3 & 0 & -2 & 0\\ 5 & 1 & 1 & 1\end{bmatrix} \quad G = \begin{bmatrix}2 & 3 & 1 & -2\\ 1 & 0 & -3 & 0\end{bmatrix}\ .$$

Using these matrices, which pairs can be multiplied together and in what order? Which triples can be multiplied together and in what order (e.g. the triple product BAC is well defined)? Which quadruples can be multiplied together and in what order? Perform all of these multiplications.

3. Block Matrix Multiplication

To illustrate the general idea of block structures consider the following matrix.

$$A = \begin{bmatrix}3 & -4 & 1 & 1 & 0 & 0\\ 0 & 2 & 2 & 0 & 1 & 0\\ 1 & 0 & -1 & 0 & 0 & 1\\ 0 & 0 & 0 & 2 & 1 & 4\\ 0 & 0 & 0 & 1 & 0 & 3\end{bmatrix}\ .$$

Visual inspection tells us that this matrix has structure. But what is it, and how can it be represented? We re-write the the matrix given above *blocking* out some key structures:

$$A = \left[\begin{array}{ccc|ccc}3 & -4 & 1 & 1 & 0 & 0\\ 0 & 2 & 2 & 0 & 1 & 0\\ 1 & 0 & -1 & 0 & 0 & 1\\ \hline 0 & 0 & 0 & 2 & 1 & 4\\ 0 & 0 & 0 & 1 & 0 & 3\end{array}\right] = \left[\begin{array}{c|c}B & I_{3\times 3}\\ \hline 0_{2\times 3} & C\end{array}\right]\ ,$$

where

$$B = \begin{bmatrix}3 & -4 & 1\\ 0 & 2 & 2\\ 1 & 0 & -1\end{bmatrix}, \quad C = \begin{bmatrix}2 & 1 & 4\\ 1 & 0 & 3\end{bmatrix}\ ,$$

$I_{3\times 3}$ is the 3×3 identity matrix, and $0_{2\times 3}$ is the 2×3 zero matrix. Having established this structure for the matrix A, it can now be exploited in various ways. As a simple example, we consider how it can be used in matrix multiplication.

Consider the matrix

$$M = \begin{bmatrix} 1 & 2 \\ 0 & 4 \\ -1 & -1 \\ 2 & -1 \\ 4 & 3 \\ -2 & 0 \end{bmatrix}.$$

The matrix product AM is well defined since A is 5×6 and M is 6×2. We show how to compute this matrix product using the structure of A. To do this we must first *block decompose M conformally with the block decomposition of A.* Another way to say this is that we must give M a block structure that allows us to do block matrix multiplication with the blocks of A. The correct block structure for M is

$$M = \begin{bmatrix} X \\ Y \end{bmatrix},$$

where

$$X = \begin{bmatrix} 1 & 2 \\ 0 & 4 \\ -1 & -1 \end{bmatrix}, \quad \text{and} \quad Y = \begin{bmatrix} 2 & -1 \\ 4 & 3 \\ -2 & 0 \end{bmatrix},$$

since then X can multiply $\begin{bmatrix} B \\ 0_{2\times 3} \end{bmatrix}$ and Y can multiply $\begin{bmatrix} I_{3\times 3} \\ C \end{bmatrix}$. This gives

$$\begin{aligned}
AM &= \begin{bmatrix} B & I_{3\times 3} \\ 0_{2\times 3} & C \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} BX + Y \\ CY \end{bmatrix} \\
&= \begin{bmatrix} \begin{bmatrix} 2 & -11 \\ 2 & 12 \\ -1 & -2 \end{bmatrix} + \begin{bmatrix} -2 & 6 \\ 4 & 3 \\ -2 & 0 \end{bmatrix} \\ \\ \begin{bmatrix} 0 & 1 \\ -4 & -1 \end{bmatrix} \end{bmatrix} \\
&= \begin{bmatrix} 4 & -12 \\ 2 & 9 \\ 0 & 3 \\ 0 & 1 \\ -4 & -1 \end{bmatrix}.
\end{aligned}$$

Block structured matrices and their matrix product is a very powerful tool in matrix analysis. Consider the matrices $M \in \mathbb{R}^{n\times m}$ and $T \in \mathbb{R}^{m\times k}$ given by

$$M = \begin{bmatrix} A_{n_1\times m_1} & B_{n_1\times m_2} \\ C_{n_2\times m_1} & D_{n_2\times m_2} \end{bmatrix}$$

and

$$T = \begin{bmatrix} E_{m_1\times k_1} & F_{m_1\times k_2} & G_{m_1\times k_3} \\ H_{m_2\times k_1} & J_{m_2\times k_2} & K_{m_2\times k_3} \end{bmatrix},$$

where $n = n_1 + n_2$, $m = m_1 + m_2$, and $k = k_1 + k_2 + k_3$. The block structures for the matrices M and T are said to be *conformal* with respect to matrix multiplication since

$$MT = \begin{bmatrix} AE + BH & AF + BJ & AG + BK \\ CE + DH & CF + DJ & CG + DK \end{bmatrix}.$$

Similarly, one can conformally block structure matrices with respect to matrix addition (how is this done?).

Exercise 3.1. *Consider the the matrix*

$$H = \begin{bmatrix} -2 & 3 & 2 & 0 & 0 & 0 & 0 \\ 1 & 1 & -3 & 0 & 0 & 0 & 0 \\ 2 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 & -1 & 0 & 0 \\ 0 & 0 & 0 & 2 & -7 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 2 & 3 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 8 & -5 \end{bmatrix}.$$

Does H have a natural block structure that might be useful in performing a matrix-matrix multiply, and if so describe it by giving the blocks? Describe a conformal block decomposition of the matrix

$$M = \begin{bmatrix} 1 & 2 \\ 3 & -4 \\ -5 & 6 \\ 1 & -2 \\ -3 & 4 \\ 1 & 1 \\ 1 & 1 \end{bmatrix}$$

that would be useful in performing the matrix product HM. Compute the matrix product HM using this conformal decomposition.

Exercise 3.2. *Let $T \in \mathbb{R}^{m\times n}$ with $T \neq 0$ and let I be the $m \times m$ identity matrix. Consider the block structured matrix $A = [\, I \quad T\,]$.*

(i) If $A \in \mathbb{R}^{k\times s}$, what are k and s?
(ii) Construct a non-zero $s \times n$ matrix B such that $AB = 0$.

The examples given above illustrate how block matrix multiplication works and why it might be useful. One of the most powerful uses of block structures is in understanding and implementing standard *matrix factorizations* or reductions.

4. Gauss-Jordan Elimination Matrices and Reduction to Reduced Echelon Form

In this section, we show that Gaussian-Jordan elimination can be represented as a consequence of left multiplication by a specially designed matrix called a *Gaussian-Jordan elimination matrix.*

Consider the vector $v \in \mathbb{R}^m$ block decomposed as

$$v = \begin{bmatrix} a \\ \alpha \\ b \end{bmatrix}$$

where $a \in \mathbb{R}^s$, $\alpha \in \mathbb{R}$, and $b \in \mathbb{R}^t$ with $m = s + 1 + t$. In this vector we refer to the α entry as the *pivot* and assume that $\alpha \neq 0$. We wish to determine a matrix G such that

$$Gv = e_{s+1}$$

where for $j = 1, \ldots, n$, e_j is the unit coordinate vector having a one in the jth position and zeros elsewhere. We claim that the matrix

$$G = \begin{bmatrix} I_{s\times s} & -\alpha^{-1}a & 0 \\ 0 & \alpha^{-1} & 0 \\ 0 & -\alpha^{-1}b & I_{t\times t} \end{bmatrix}$$

does the trick. Indeed,

$$Gv = \begin{bmatrix} I_{s\times s} & -\alpha^{-1}a & 0 \\ 0 & \alpha^{-1} & 0 \\ 0 & -\alpha^{-1}b & I_{t\times t} \end{bmatrix} \begin{pmatrix} a \\ \alpha \\ b \end{pmatrix} = \begin{bmatrix} a - a \\ \alpha^{-1}\alpha \\ -b + b \end{bmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = e_{s+1}. \tag{3}$$

The matrix G is called a *Gaussian-Jordan Elimination Matrix*, or GJEM for short. Note that G is invertible since

$$G^{-1} = \begin{bmatrix} I & a & 0 \\ 0 & \alpha & 0 \\ 0 & b & I \end{bmatrix},$$

Moreover, for any vector of the form $w = \begin{pmatrix} x \\ 0 \\ y \end{pmatrix}$ where $x \in \mathbb{R}^s$ $y \in \mathbb{R}^t$, we have

$$Gw = w.$$

The GJEM matrices perform precisely the operations required in order to execute Gauss-Jordan elimination. That is, each elimination step can be realized as left multiplication of the augmented matrix by the appropriate GJEM.

For example, consider the linear system

$$\begin{array}{rcrcrcr} 2x_1 & + & x_2 & + & 3x_3 & = & 5 \\ 2x_1 & + & 2x_2 & + & 4x_3 & = & 8 \\ 4x_1 & + & 2x_2 & + & 7x_3 & = & 11 \\ 5x_1 & + & 3x_2 & + & 4x_3 & = & 10 \end{array}$$

and its associated augmented matrix

$$A = \begin{bmatrix} 2 & 1 & 3 & 5 \\ 2 & 2 & 4 & 8 \\ 4 & 2 & 7 & 11 \\ 5 & 3 & 4 & 10 \end{bmatrix}.$$

The first step of Gauss-Jordan elimination is to transform the first column of this augmented matrix into the first unit coordinate vector. The procedure described in (3) can be employed for this purpose. In this case the pivot is the $(1,1)$ entry of the augmented matrix and so

$$s = 0,\ a \text{ is void},\ \alpha = 2,\ t = 3,\ \text{and } b = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix},$$

which gives

$$G_1 = \begin{bmatrix} 1/2 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ -5/2 & 0 & 0 & 1 \end{bmatrix}.$$

Multiplying these two matrices gives

$$G_1A = \begin{bmatrix} 1/2 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -2 & 0 & 1 & 0 \\ -5/2 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 3 & 5 \\ 2 & 2 & 4 & 8 \\ 4 & 2 & 7 & 11 \\ 5 & 3 & 4 & 10 \end{bmatrix} = \begin{bmatrix} 1 & 1/2 & 3/2 & 5/2 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 1/2 & -7/2 & -5/2 \end{bmatrix}.$$

We now repeat this process to transform the second column of this matrix into the second unit coordinate vector. In this case the $(2,2)$ position becomes the pivot so that

$$s = 1,\ a = 1/2,\ \alpha = 1,\ t = 2,\ \text{and } b = \begin{bmatrix} 0 \\ 1/2 \end{bmatrix}$$

yielding

$$G_2 = \begin{bmatrix} 1 & -1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1/2 & 0 & 1 \end{bmatrix}.$$

Again, multiplying these two matrices gives

$$G_2G_1A = \begin{bmatrix} 1 & -1/2 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1/2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & 3/2 & 5/2 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 1/2 & -7/2 & -5/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -4 & -4 \end{bmatrix}.$$

Repeating the process on the third column transforms it into the third unit coordinate vector. In this case the pivot is the $(3,3)$ entry so that

$$s=2,\ a=\begin{bmatrix}1\\1\end{bmatrix},\ \alpha=1,\ t=1,\ \text{and } b=-4$$

yielding

$$G_3=\begin{bmatrix}1&0&-1&0\\0&1&-1&0\\0&0&1&0\\0&0&4&1\end{bmatrix}.$$

Multiplying these matrices gives

$$G_3G_2G_1A=\begin{bmatrix}1&0&-1&0\\0&1&-1&0\\0&0&1&0\\0&0&4&1\end{bmatrix}\begin{bmatrix}1&0&1&1\\0&1&1&3\\0&0&1&1\\0&0&-4&-4\end{bmatrix}=\begin{bmatrix}1&0&0&0\\0&1&0&2\\0&0&1&1\\0&0&0&0\end{bmatrix},$$

which is in reduced echelon form. Therefore the system is consistent and the unique solution is

$$x=\begin{bmatrix}0\\2\\1\end{bmatrix}.$$

Observe that

$$G_3G_2G_1=\begin{bmatrix}3&-1/2&-1&0\\1&1&-1&0\\-2&0&1&0\\-10&-1/2&4&1\end{bmatrix}$$

and that

$$\begin{aligned}(G_3G_2G_1)^{-1}&=G_1^{-1}G_2^{-1}G_3^{-1}\\&=\begin{bmatrix}2&0&0&0\\2&1&0&0\\4&0&1&0\\5&0&0&1\end{bmatrix}\begin{bmatrix}1&1/2&0&0\\0&1&0&0\\0&0&1&0\\0&1/2&0&1\end{bmatrix}\begin{bmatrix}1&0&1&0\\0&1&1&0\\0&0&1&0\\0&0&-4&1\end{bmatrix}\\&=\begin{bmatrix}2&1&3&0\\2&2&4&0\\4&2&7&0\\5&3&4&1\end{bmatrix}.\end{aligned}$$

In particular, reduced Gauss-Jordan form can always be achieved by multiplying the augmented matrix on the left by an invertible matrix which can be written as a product of Gauss-Jordan elimination matrices.

Exercise 4.1. *What are the Gauss-Jordan elimination matrices that transform the vector* $\begin{bmatrix}2\\3\\-2\\5\end{bmatrix}$ *in to* e_j *for* $j=1,2,3,4$*, and what are the inverses of these matrices?*

5. Some Special Square Matrices

We say that a matrix A is square if there is a positive integer n such that $A\in\mathbb{R}^{n\times n}$. For example, the Gauss-Jordan elimination matrices are a special kind of square matrix. Below we give a list of some square matrices with special properties that are very useful to our future work.

Diagonal Matrices: The diagonal of a matrix $A=[A_{ij}]$ is the vector $(A_{11},A_{22},\dots,A_{nn})^T\in\mathbb{R}^n$. A matrix in $\mathbb{R}^{n\times n}$ is said to be diagonal if the only non-zero entries of the matrix are the diagonal entries. Given a vector $v\in\mathbb{R}^n$, we write $\text{diag}(v)$ the denote the diagonal matrix whose diagonal is the vector v.

The Identity Matrix: The identity matrix is the diagonal matrix whose diagonal entries are all ones. We denote the identity matrix in $\mathbb{R}^k$ by I_k. If the dimension of the identity is clear, we simply write I. Note that for any matrix $A\in\mathbb{R}^{m\times n}$ we have $I_mA=A=AI_n$.

Inverse Matrices: The inverse of a matrix $X \in \mathbb{R}^{n\times n}$ is any matrix $Y \in \mathbb{R}^{n\times n}$ such that $XY = I$ in which case we write $X^{-1} := Y$. It is easily shown that if Y is an inverse of X, then Y is unique and $YX = I$.

Permutation Matrices: A matrix $P \in \mathbb{R}^{n\times n}$ is said to be a permutation matrix if P is obtained from the identity matrix by either permuting the columns of the identity matrix or permuting its rows. It is easily seen that $P^{-1} = P^T$.

Unitary Matrices: A matrix $U \in \mathbb{R}^{n\times n}$ is said to be a unitary matrix if $U^TU = I$, that is $U^T = U^{-1}$. Note that every permutation matrix is unitary. But the converse is note true since for any vector u with $\|u\|_2 = 1$ the matrix $I - 2uu^T$ is unitary.

Symmetric Matrices: A matrix $M \in \mathbb{R}^{n\times n}$ is said to be symmetric if $M^T = M$.

Skew Symmetric Matrices: A matrix $M \in \mathbb{R}^{n\times n}$ is said to be skew symmetric if $M^T = -M$.

6. The LU Factorization

In this section we revisit the reduction to echelon form, but we incorporate permutation matrices into the pivoting process. Recall that a matrix $P \in \mathbb{R}^{m\times m}$ is a *permutation matrix* if it can be obtained from the identity matrix by permuting either its rows or columns. It is straightforward to show that $P^TP = I$ so that the inverse of a permutation matrix is its transpose. Multiplication of a matrix on the left permutes the rows of the matrix while multiplication on the right permutes the columns. We now apply permutation matrices in the Gaussian elimination process in order to avoid zero pivots.

Let $A \in \mathbb{R}^{m\times n}$ and assume that $A \neq 0$. Set $\widetilde{A}_0 := A$. If the $(1,1)$ entry of $\widetilde{A}_0$ is zero, then apply permutation matrices P_{l0} and P_{r0} to the left and and right of $\widetilde{A}_0$, respectively, to bring *any* non-zero element of $\widetilde{A}_0$ into the $(1,1)$ position (e.g., the one with largest magnitude) and set $A_0 := P_{l0}\widetilde{A}_0P_{r0}$. Write A_0 in block form as

$$A_0 = \begin{bmatrix} \alpha_1 & v_1^T \\ u_1 & \widehat{A}_1 \end{bmatrix} \in \mathbb{R}^{m\times n},$$

with $0 \neq \alpha_1 \in \mathbb{R}$, $u_1 \in \mathbb{R}^{n-1}$, $v_1 \in \mathbb{R}^{m-1}$, and $\widetilde{A}_1 \in \mathbb{R}^{(m-1)\times(n-1)}$. Then using α_1 to zero out u_1 amounts to left multiplication of the matrix A_0 by the Gaussian elimination matrix

$$\begin{bmatrix} 1 & 0 \\ -\frac{u_1}{\alpha_1} & I \end{bmatrix}$$

to get

$$\begin{bmatrix} 1 & 0 \\ -\frac{u_1}{\alpha_1} & I \end{bmatrix} \begin{bmatrix} \alpha_1 & v_1^T \\ u_1 & \widehat{A}_1 \end{bmatrix} = \begin{bmatrix} \alpha_1 & v_1^T \\ 0 & \widetilde{A}_1 \end{bmatrix} \in \mathbb{R}^{m\times n}\ , \tag{4}$$

where

$$\widetilde{A}_1 = \widehat{A}_1 - u_1v_1^T/\alpha_1\ .$$

Define

$$\widetilde{L}_1 = \begin{bmatrix} 1 & 0 \\ \frac{u_1}{\alpha_1} & I \end{bmatrix} \in \mathbb{R}^{m\times m} \quad \text{and} \quad \widetilde{U}_1 = \begin{bmatrix} \alpha_1 & v_1^T \\ 0 & \widetilde{A}_1 \end{bmatrix} \in \mathbb{R}^{m\times n}\ .$$

and observe that

$$\widetilde{L}_1^{-1} = \begin{bmatrix} 1 & 0 \\ -\frac{u_1}{\alpha_1} & I \end{bmatrix}$$

Hence (4) becomes

$$\widetilde{L}_1^{-1}P_{l0}\widetilde{A}_0P_{r0} = \widetilde{U}_1, \text{ or equivalently,} \quad A = P_{l0}\widetilde{L}_1\widetilde{U}_1P_{r0}^T\ . \tag{5}$$

Note that $\widetilde{L}_1$ is *unit* lower triangular (ones on the mail diagonal) and $\widetilde{U}_1$ is block upper-triangular with one nonsingular 1×1 block and one $(m-1)\times(n-1)$ block on the block diagonal.

Next consider the matrix $\widetilde{A}_1$ in $\widetilde{U}_1$. If the $(1,1)$ entry of $\widetilde{A}_1$ is zero, then apply permutation matrices $\widetilde{P}_{l1} \in \mathbb{R}^{(m-1)\times(m-1)}$ and $\widetilde{P}_{r1} \in \mathbb{R}^{(n-1)\times(n-1)}$ to the left and and right of $\widetilde{A}_1 \in \mathbb{R}^{(m-1)\times(n-1)}$, respectively, to bring *any* non-zero element of $\widetilde{A}_0$ into the $(1,1)$ position (e.g., the one with largest magnitude) and set $A_1 := \widetilde{P}_{l1}\widetilde{A}_1P_{r1}$. If the element of $\widetilde{A}_1$ is zero, then stop. Define

$$P_{l1} := \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{P}_{l1} \end{bmatrix} \quad \text{and} \quad P_{r1} := \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{P}_{r1} \end{bmatrix}$$

so that P_{l1} and P_{r1} are also permutation matrices and

$$P_{l1}\widetilde{U}_1 P_{r1} = \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{P}_{l1} \end{bmatrix} \begin{bmatrix} \alpha_1 & v_1^T \\ 0 & \widetilde{A}_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{P}_{r1} \end{bmatrix} = \begin{bmatrix} \alpha_1 & v_1^T P_{r1} \\ 0 & \widetilde{P}_{l1}\widetilde{A}_1 P_{r1} \end{bmatrix} = \begin{bmatrix} \alpha_1 & \tilde{v}_1^T \\ 0 & A_1 \end{bmatrix}, \tag{6}$$

where $\tilde{v}_1 := P_{r1}^T v_1$. Define

$$U_1 := \begin{bmatrix} \alpha_1 & \tilde{v}_1^T \\ 0 & A_1 \end{bmatrix}, \quad \text{where} \quad A_1 = \begin{bmatrix} \alpha_2 & v_2^T \\ u_2 & \widehat{A}_2 \end{bmatrix} \in \mathbb{R}^{(m-1)\times(n-1)},$$

with $0 \neq \alpha_2 \in \mathbb{R}$, $u_2 \in \mathbb{R}^{n-2}$, $v_1 \in \mathbb{R}^{m-2}$, and $\widetilde{A}_2 \in \mathbb{R}^{(m-2)\times(n-2)}$. In addition, define

$$L_1 := \begin{bmatrix} 1 & 0 \\ \widetilde{P}_{l1}\frac{u_1}{\alpha_1} & I \end{bmatrix},$$

so that

$$\begin{aligned} P_{l1}^T L_1 &= \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{P}_{l1}^T \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \widetilde{P}_{l1}\frac{u_1}{\alpha 1} & I \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ \frac{u_1}{\alpha 1} & \widetilde{P}_{l1}^T \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ \frac{u_1}{\alpha 1} & I \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \widetilde{P}_{l1}^T \end{bmatrix} \\ &= \widetilde{L}_1 P_{l1}^T, \end{aligned}$$

and consequently

$$L_1^{-1} P_{l1} = P_{l1}\widetilde{L}_1^{-1}\ .$$

Plugging this into (5) and using (6), we obtain

$$L_1^{-1} P_{l1} P_{l0}\widetilde{A}_0 P_{r0} P_{r1} = P_{l1}\widetilde{L}_1^{-1} P_{l0}\widetilde{A}_0 P_{r0} P_{r1} = P_{l1}\widetilde{U}_1 P_{r1} = U_1,$$

or equivalently,

$$P_{l1} P_{l0} A P_{r0} P_{r1} = L_1 U_1.$$

We can now repeat this process on the matrix A_1 since the $(1,1)$ entry of this matrix is non-zero. The process can run for no more than the number of rows of A which is m. However, it may terminate after $k < m$ steps if the matrix $\widehat{A}_k$ is the zero matrix. In either event, we obtain the following result.

Theorem 6.1. *[The LU Factorization] Let $A \in \mathbb{R}^{m\times n}$. If $k = \operatorname{rank}(A)$, then there exist permutation matrices $P_l \in \mathbb{R}^{m\times m}$ and $P_r \in \mathbb{R}^{n\times n}$ such that*

$$P_l A P_r = LU,$$

where $L \in \mathbb{R}^{m\times m}$ is a lower triangular matrix having ones on its diagonal and

$$U = \begin{bmatrix} U_1 & U_2 \\ 0 & 0 \end{bmatrix}$$

with $U_1 \in \mathbb{R}^{k\times k}$ a nonsingular upper triangular matrix.

Note that a column permutation is only required if the first column of $\hat{A}_k$ is zero for some k before termination. In particular, this implies that the $\operatorname{rank}(A) < m$. Therefore, if $\operatorname{rank}(A) = m$, column permutations are not required, and $P_r = I$. If one implements the LU factorization so that a column permutation is *only* employed in the case when the first column of $\hat{A}_k$ is zero for some k, then we say the LU factorization is obtained through partial pivoting.

Example 6.1. *We now use the procedure outlined above to compute the LU factorization of the matrix*

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 4 & 2 \\ -1 & 1 & 3 \end{bmatrix}.$$

$$
\begin{aligned}
L_1^{-1}A &= \begin{bmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 2 & 4 & 2 \\ -1 & 1 & 3 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & -3 \\ 0 & 2 & 5 \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
L_2^{-1}L_1^{-1}A &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & -3 \\ 0 & 2 & 5 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 8 \end{bmatrix}
\end{aligned}
$$

We now have

$$U = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 8 \end{bmatrix},$$

and

$$L = L_1 L_2 = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -1 & 1 & 1 \end{bmatrix}.$$

7. Solving Equations with the LU Factorization

Consider the equation $Ax = b$. In this section we show how to solve this equation using the LU factorization. Recall from Theorem 6.1 that the algorithm of the previous section produces a factorization of A of the form $P_l \in \mathbb{R}^{m\times m}$ and $P_r \in \mathbb{R}^{n\times n}$ such that

$$A = P_l^T L U P_r^T,$$

where $P_l \in \mathbb{R}^{m\times m}$ and $P_r \in \mathbb{R}^{n\times n}$ are permutation matrices, $L \in \mathbb{R}^{m\times m}$ is a lower triangular matrix having ones on its diagonal, and

$$U = \begin{bmatrix} U_1 & U_2 \\ 0 & 0 \end{bmatrix}$$

with $U_1 \in \mathbb{R}^{k\times k}$ a nonsingular upper triangular matrix. Hence we may write the equation $Ax = b$ as

$$P_l^T L U P_r^T x = b.$$

Multiplying through by P_l and replacing $UP_r^T x$ by w gives the equation

$$Lw = \hat{b}, \quad \text{where} \quad \hat{b} := P_l b\ .$$

This equation is easily solved by forward substitution since L is a nonsingular lower triangular matrix. Denote the solution by $\overline{w}$. To obtain a solution x we must still solve $UP_r^T x = \overline{w}$. Set $y = P_r x$. The this equation becomes

$$\overline{w} = Uy = \begin{bmatrix} U_1 & U_2 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

where we have decomposed y to conform to the decomposition of U. Doing the same for $\overline{w}$ gives

$$\begin{pmatrix} \overline{w}_1 \\ \overline{w}_2 \end{pmatrix} = \begin{bmatrix} U_1 & U_2 \\ 0 & 0 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

or equivalently,

$$
\begin{aligned}
\overline{w}_1 &= U_1 y_1 + U_2 y_2 \\
\overline{w}_2 &= 0.
\end{aligned}
$$

Hence, if $\overline{w}_2 \neq 0$, the system is inconsistent, i.e., no solution exists. On the other hand, if $\overline{w}_2 = 0$, we can take $y_2 = 0$ and solve the equation

$$\overline{w}_1 = U_1 y_1 \tag{7}$$

for $\overline{y}_1$, then

$$\overline{x} = P_r^T \begin{pmatrix} \overline{y}_1 \\ 0 \end{pmatrix}$$

is a solution to $Ax = b$. The equation (7) is also easy to solve since U_1 is an upper triangular nonsingular matrix so that (7) can be solved by back substitution.

8. The Four Fundamental Subspaces and Echelon Form

Recall that a subset W of $\mathbb{R}^n$ is a subspace if and only if it satisfies the following three conditions:

(1) The origin is an element of W.
(2) The set W is closed with respect to addition, i.e. if $u \in W$ and $v \in W$, then $u + v \in W$.
(3) The set W is closed with respect to scalar multiplication, i.e. if $\alpha \in \mathbb{R}$ and $u \in W$, then $\alpha u \in W$.

Exercise 8.1. *Given $v^1, v^2, \ldots, v^k \in \mathbb{R}^n$, show that the linear span of these vectors,*

$$\operatorname{span}\left(v^1, v^2, \ldots, v^k\right) := \left\{\alpha_1 v^1 + \alpha_2 v^2 + \cdots + \alpha_k v^k \,\middle|\, \alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}\right\}$$

is a subspace.

Exercise 8.2. *Show that for any set S in $\mathbb{R}^n$, the set*

$$S^\perp = \{v \,:\, w^T v = 0 \text{ for all } w \in S\}$$

is a subspace. If S is itself a subspace, then $S^\perp$ is called the subspace orthogonal (or perpendicular) to the subspace S.

Exercise 8.3. *If S is any suset of $\mathbb{R}^n$ (not necessarily a subspace), show that $(S^\perp)^\perp = \operatorname{span}(S)$.*

Exercise 8.4. *If $S \subset \mathbb{R}^n$ is a subspace, show that $S = (S^\perp)^\perp$.*

A set of vectors $v^1, v^2, \ldots, v^k \in \mathbb{R}^n$ are said to be *linearly independent* if $0 = a_1 v^1 + \cdots + a_k v^k$ if and only if $0 = a_1 = a_2 = \cdots = a_k$. A *basis* for a subspace in any maximal linearly independent subspace. An elementary fact from linear algebra is that the subspace equals the linear span of any basis for the subspace and that every basis of a subspace has the same number of vectors in it. We call this number the *dimension* for the subspace. If S is a subspace, we denote the dimension of S by $\dim S$.

Exercise 8.5. *If $s \subset \mathbb{R}^n$ is a subspace, then any basis of S can contain only finitely many vectors.*

Exercise 8.6. *Show that every subspace can be represented as the linear span of a basis for that subspace.*

Exercise 8.7. *Show that every basis for a subspace contains the same number of vectors.*

Exercise 8.8. *If $S \subset \mathbb{R}^n$ is a subspace, show that*

$$\mathbb{R}^n = S + S^\perp \tag{8}$$

and that

$$n = \dim S + \dim S^\perp. \tag{9}$$

Let $A \in \mathbb{R}^{m \times n}$. We associate with A its four fundamental subspaces:

$$\begin{array}{ll} \operatorname{Ran}(A) := \{Ax \,|\, x \in \mathbb{R}^n\} & \operatorname{Null}(A) := \{x \,|\, Ax = 0\} \\ \operatorname{Ran}(A^T) := \left\{A^T y \,\middle|\, y \in \mathbb{R}^m\right\} & \operatorname{Null}(A^T) := \left\{y \,\middle|\, A^T y = 0\right\}. \end{array}$$

where

$$\begin{array}{ll} \operatorname{rank}(A) := \dim \operatorname{Ran}(A) & \operatorname{nullity}(A) := \dim \operatorname{Null}(A) \\ \operatorname{rank}(A^T) := \dim \operatorname{Ran}(A^T) & \operatorname{nullity}(A^T) := \dim \operatorname{Null}(A^T) \end{array} \tag{10}$$

Exercise 8.9. *Show that the four fundamental subspaces associated with a matrix are indeed subspaces.*

Observe that

$$\begin{aligned}
\mathrm{Null}(A) &:= \{x \mid Ax = 0\} \\
&= \{x \mid A_{i\cdot} \bullet x = 0,\ i = 1, 2, \ldots, m\} \\
&= \{A_{1\cdot}, A_{2\cdot}, \ldots, A_{m\cdot}\}^{\perp} \\
&= \mathrm{span}\,(A_{1\cdot}, A_{2\cdot}, \ldots, A_{m\cdot})^{\perp} \\
&= \mathrm{Ran}(A^T)^{\perp}\ .
\end{aligned}$$

Since for any subspace $S \subset \mathbb{R}^n$, we have $(S^{\perp})^{\perp} = S$, we obtain

$$\mathrm{Null}(A)^{\perp} = \mathrm{Ran}(A^T) \text{ and } \mathrm{Null}(A^T) = \mathrm{Ran}(A)^{\perp}. \tag{11}$$

The equivalences in (11) are called the *Fundamental Theorem of the Alternative*.

One of the big consequences of echelon form is that

$$n = \mathrm{rank}(A) + \mathrm{nullity}(A). \tag{12}$$

By combining (12), (9) and (11), we obtain the equivalence

$$\mathrm{rank}(A^T) = \dim \mathrm{Ran}(A^T) = \dim \mathrm{Null}(A)^{\perp} = n - \mathrm{nullity}(A) = \mathrm{rank}(A).$$

That is, the row rank of a matrix equals the column rank of a matrix, i.e., the dimensions of the row and column spaces of a matrix are the same!