Background from Linear Algebra

- We will consider only matrices with real-valued entries.
- Notations:
  - Matrices will be denoted by capital letters in regular (i.e., not bold) font: e.g., A.
  - Column vectors will be denoted by lower-case letters in bold font: e.g., $\mathbf{u}$.
  - Row vectors will be denoted similarly to column vectors, but with the operation of transposition: e.g., $\mathbf{v}^T$.
  - Scalars will be denoted by lower-case letters in regular font: e.g., a.
- I will denote the identity matrix of appropriate dimension.

Null space, Range (column space), and Rank of a matrix

Null space of an $m \times n$ matrix A, often denoted as $\mathcal{N}(A)$, is the set of all vectors $\mathbf{u} \in \mathbb{R}^n$ such that

$$A\mathbf{u} = \mathbf{0}. \tag{1}$$

Range of an $m \times n$ matrix A, often denoted as $\mathcal{R}(A)$, is the set of all vectors $\mathbf{u} \in \mathbb{R}^m$ such that

$$\mathbf{u} = A\mathbf{v}, \qquad \text{where } \mathbf{v} \text{ can be } \textit{any} \text{ vector in } \mathbb{R}^n. \tag{2}$$

The range of a matrix is sometimes called its *column space*; this is based on the important formula (10a) found below in this document. It is then denoted as $\text{col}(A)$.

Rank of matrix A, often denoted as $\text{rank}(A)$, is the number of its linearly independent[1] columns. Given the aforementioned formula (10a), $\text{rank}(A)$ equals the dimension of its column space (i.e., of $\mathcal{R}(A)$).

Theorem:
The number of linearly independent columns of any matrix equals the number of its linearly independent rows. Equivalently:

$$\text{rank}(A) = \text{rank}(A^T). \tag{3}$$

[1] You may not necessarily remember what "linear independent" means. As with any concept that you don't remember — it's okay. You just have to refresh your memory on this concept on your own. The same applies to all other concepts from your elementary Linear Algebra course that you encounter in this document.

Orthogonal matrices

Any matrix can be viewed as a collection of its columns:

$$Q = [\mathbf{q}_1,\, \mathbf{q}_2, \dots,\, \mathbf{q}_n]. \tag{4}$$

A square matrix is called orthogonal if its columns are orthonormal vectors:

$$\mathbf{q}_i^T\, \mathbf{q}_j = 0 \quad \text{for } i \neq j \;\text{ and } \quad \mathbf{q}_i^T \mathbf{q}_i = 1. \tag{5}$$

As a corollary of (5), any orthogonal matrix satisfies:

$$Q^T\, Q = I. \tag{6a}$$

Thus, the inverse of an orthogonal matrix is simply its transpose:

$$Q^{-1} = Q^T, \tag{6b}$$

and therefore, by a property of the inverse matrix, one also has

$$Q\, Q^T = I. \tag{6c}$$

This implies that Q^T is also an orthogonal matrix, and hence *not only columns, but also rows, of an orthogonal matrix are orthonormal vectors*.

For any $\mathbf{u} \in \mathbb{R}^n$, one has

$$\|Q\mathbf{u}\| = \|\mathbf{u}\|, \tag{7}$$

where $\| \cdots \|$ is the Euclidean norm (a.k.a. length) of a vector.

Alternative view of matrix–matrix multiplication

When multiplying two matrices, you are essentially multiplying the left matrix by each column of the right matrix:

$$A\,B \equiv A\,[\,\mathbf{b}_1,\, \mathbf{b}_2,\, \dots,\mathbf{b}_n\,] = [\,A\mathbf{b}_1,\, A\mathbf{b}_2,\, \dots,A\mathbf{b}_n\,]. \tag{8}$$

Corollary:

Let $A = \mathbf{u}$ with $\mathbf{u} \equiv [u_1,\, u_2,\, \dots, u_m]^T$ and $B = \mathbf{v}^T$ with $\mathbf{v}^T \equiv [v_1,\, v_2,\, \dots, v_n]$. Then the *outer product* of $\mathbf{u}$ and $\mathbf{v}$ is the $m \times n$ matrix

$$\mathbf{u}\,\mathbf{v}^T \overset{(8)}{=} \begin{bmatrix} | & | & & | \\ \mathbf{u}\,v_1 & \mathbf{u}\,v_2 & \cdots & \mathbf{u}\,v_n \\ | & | & & | \end{bmatrix}. \tag{9a}$$

Alternatively, if we use the more conventional, row-by-column way of matrix multiplication, we can rewrite the left-hand side of (9a) as:

$$\mathbf{u}\,\mathbf{v}^T = \begin{pmatrix} \text{——} & u_1\,\mathbf{v}^T & \text{——} \\ \text{——} & u_2\,\mathbf{v}^T & \text{——} \\ & \cdots & \\ \text{——} & u_m\,\mathbf{v}^T & \text{——} \end{pmatrix}. \tag{9b}$$

Here and below in this document, the vertical and horizontal lines are added to help you visualize the (vertical or horizontal) arrangement of columns or rows, respectively. Equations (9) show that $\mathbf{u}\,\mathbf{v}^T$ has rank one (another way to say the same is that it is a *rank-1 matrix*). Indeed, all of its columns are proportional to the first column, and all rows are proportional to the first row.

Alternative view of matrix–vector multiplication

If $A = [\,\mathbf{a}_1,\, \mathbf{a}_2,\, \ldots\,,\mathbf{a}_n\,]$ and $\mathbf{u} \equiv [u_1,\, u_2,\, \ldots\,, u_n]^T$, then

$$A\,\mathbf{u} \;=\; u_1\,\mathbf{a}_1 + u_2\,\mathbf{a}_2 + \cdots + u_n\,\mathbf{a}_n. \tag{10a}$$

You should convince yourself of this fact by considering an example of some 2×2 matrix (say, with entries 1, 2, 3, 4) and $\mathbf{u} = [u_1,\, u_2]^T$.

By taking the transpose of formula (10a) (and renaming the matrix and the vector accordingly), one can obtain an *alternative form of* (10a). Namely, if $B^T = [\,\mathbf{b}_1,\, \mathbf{b}_2,\, \ldots\,, \mathbf{b}_m\,]$ (i.e., $\mathbf{b}_1^T,\, \mathbf{b}_2^T,\, \ldots\,, \mathbf{b}_m^T$ are the rows of B) and $\mathbf{v}^T \equiv [v_1,\, v_2,\, \ldots\,, v_m]$, then

$$\mathbf{v}^T\,B \;=\; v_1\,\mathbf{b}_1 + v_2\,\mathbf{b}_2 + \cdots + v_m\,\mathbf{b}_m. \tag{10b}$$

Formulas (10) are extensively used in Linear Algebra.

Rank-k matrices

Above you have seen that $\mathbf{u}\,\mathbf{v}^T$ is a rank-1 matrix. One can construct a rank-2 matrix (i.e. a matrix whose rank equals 2) as follows. Let the set of two vectors $\{\mathbf{u}_1,\, \mathbf{u}_2\}$ be linear independent, and so be the set $\{\mathbf{v}_1,\, \mathbf{v}_2\}$. Then the following outer product produces a rank-2 matrix:

$$[\,\mathbf{u}_1,\, \mathbf{u}_2\,] \begin{pmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \end{pmatrix} \;=\; \mathbf{u}_1\,\mathbf{v}_1^T \;+\; \mathbf{u}_2\,\mathbf{v}_2^T. \tag{11}$$

You should convince yourself of correctness of this formula by considering a simple example, e.g.:

$$\begin{aligned} \begin{pmatrix} 11 & 12 \\ 21 & 22 \end{pmatrix} \begin{pmatrix} 101 & 102 & 103 \\ 201 & 202 & 203 \end{pmatrix} &= \\ \begin{pmatrix} 11\cdot 101 + 12\cdot 201 & 11\cdot 102 + 12\cdot 202 & 11\cdot 103 + 12\cdot 203 \\ 21\cdot 101 + 22\cdot 201 & 21\cdot 102 + 22\cdot 202 & 21\cdot 103 + 22\cdot 203 \end{pmatrix} &= \\ \begin{pmatrix} 11\cdot 101 & 11\cdot 102 & 11\cdot 103 \\ 21\cdot 101 & 21\cdot 102 & 21\cdot 103 \end{pmatrix} + \begin{pmatrix} 12\cdot 201 & 12\cdot 202 & 12\cdot 203 \\ 22\cdot 201 & 22\cdot 202 & 22\cdot 203 \end{pmatrix} &. \end{aligned} \tag{12}$$

Similarly, if a set of k vectors $\{\mathbf{u}_1,\, \mathbf{u}_2,\, \ldots\,, \mathbf{u}_k\}$ is linear independent, and so is the set $\{\mathbf{v}_1,\, \mathbf{v}_2,\, \ldots\,, \mathbf{v}_k\}$, then the following outer product (see its definition before Eq. (9a)) produces a rank-k matrix:

$$\begin{pmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \\ | & | & & | \end{pmatrix} \begin{pmatrix} \text{——} & \mathbf{v}_1^T & \text{——} \\ \text{——} & \mathbf{v}_2^T & \text{——} \\ & \cdots & \\ \text{——} & \mathbf{v}_k^T & \text{——} \end{pmatrix} \;=\; \mathbf{u}_1\,\mathbf{v}_1^T \;+\; \mathbf{u}_2\,\mathbf{v}_2^T \;+\; \ldots \;+\; \mathbf{u}_k\,\mathbf{v}_k^T. \tag{13}$$

Multiplication by a diagonal matrix

Let Σ be an $n \times n$ diagonal matrix; that is, its only nonzero entries are located on its main diagonal. A shorthand notation for this is: $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$. Then:

- Multiplying any $n \times p$ matrix A by Σ on the *left* results in multiplying each *row* of A by its respective σ_i:

$$\Sigma A \equiv \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \ldots & \sigma_n \end{pmatrix} \begin{pmatrix} \text{---} & \mathbf{a}_1^T & \text{---} \\ \text{---} & \mathbf{a}_2^T & \text{---} \\ & \cdots & \\ \text{---} & \mathbf{a}_n^T & \text{---} \end{pmatrix} = \begin{pmatrix} \text{---} & \sigma_1 \mathbf{a}_1^T & \text{---} \\ \text{---} & \sigma_2 \mathbf{a}_2^T & \text{---} \\ & \cdots & \\ \text{---} & \sigma_n \mathbf{a}_n^T & \text{---} \end{pmatrix}. \tag{14}$$

- Multiplying any $m \times n$ matrix B by Σ on the *right* results in multiplying each *column* of B by its respective σ_i:

$$B \Sigma \equiv \begin{pmatrix} | & | & & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \\ | & | & & | \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \ldots & \sigma_n \end{pmatrix} = \begin{pmatrix} | & | & & | \\ \sigma_1 \mathbf{b}_1 & \sigma_2 \mathbf{b}_2 & \cdots & \sigma_n \mathbf{b}_n \\ | & | & & | \end{pmatrix}. \tag{15}$$

You should verify these statements for some 2×2 matrices A and B (say, with entries 1, 2, 3, 4) and a 2×2 matrix $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2)$.