Note III — Linear transformations*

Bruno Salcedo†

Fall 2016

1. Matrix Algebra

An $n \times m$ matrix is a two dimensional array of real numbers with n rows and m columns. I will use the notation $A = (A_{ij})$ to represent $n \times m$ matrices, and $\mathbb{R}^{n\times m}$ to denote the the set of such matrices. The element in the ith column and jth row of matrix A is denoted A_{ij}. The ith row and the jth row of A are denoted by $\text{row}_i(A)$ and $\text{col}_j(A)$, respectively. An $n \times m$ matrix is said to be square if $n = m$. Elements of $\mathbb{R}^n$ are treated as column vectors, that is, as $n \times 1$ matrices.

The sum of matrices $A, B \in \mathbb{R}^{n\times m}$ is the matrix $A+B \in \mathbb{R}^{n\times m}$ with $(A+B)_{ij} = A_{ij} + B_{ij}$. The scalar product of $\alpha \in \mathbb{R}$ and $A \in \mathbb{R}^{n\times m}$ is the matrix $\alpha \cdot A \in \mathbb{R}^{n\times m}$ with $(\alpha \cdot A)_{ij} = \alpha A_{ij}$. For any natural numbers n and m, the set $\mathbb{R}^{n\times m}$ endowed with these operations is a vector space.

A $n \times m$ matrix is said to be *square* if $n = m$. The *transpose* of an $n \times m$ matrix A is the $m \times n$ matrix denoted by A^T with $A^\mathsf{T}_{ij} = A_{ji}$. A matrix A is said to be *symmetric* if $A = A^\mathsf{T}$. Note that only square matrices can be symmetric. A square matrix $A \in \mathbb{R}^{n\times n}$ is symmetric if and only if $A_{ij} = A_{ji}$ for $i, j = 1, \ldots, n$.

Given a $n \times m$ matrix A and a $m \times r$ matrix B, the *matrix product* of A and B is the $n \times r$ matrix AB defined by

$$(AB)_{ij} = \sum_{k=1}^{m} A_{ik} B_{kj}.$$

*Lecture notes for Econ 6170, Intermediate Mathematical Economics I, Cornell University.

†Department of Economics, Cornell University · brunosalcedo.com · salcedo@cornell.edu

Note that an equivalent expression is

$$(AB)_{ij} = \mathrm{row}_i^{\mathsf{T}}(A) \cdot \mathrm{col}_j(B)$$

Also note the dot product of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ can be expressed in terms of matrix product as $\mathbf{x} \cdot \mathbf{y} = \mathbf{x}^{\mathsf{T}}\mathbf{y}$.

Matrix product it is associative, it is distributive with respect to matrix addition both from the left and the right, and it is compatible with scalar product. That is, for matrices A, B, C and scalars a we have that $(AB)C = A(BC)$, $A(B + C) = AB + AC$, $(A + B)C = AC + BC$, and $a(AB) = (aA)B = A(aB)$, provided that all all the products are well defined. Unfortunately, matrix product is not commutative. Even in cases when both AB and BA are well defined, it might be the case that $AB \neq BA$.

Example 1.1 Let $A, B \in \mathbb{R}^{2\times 3}$ and $C \in \mathbb{R}3 \times 2$ be the matrices given by

$$A = \begin{pmatrix} 2 & 3 & 5 \\ 1 & 0 & 5 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 1 & 0 \\ 7 & 1 & 0 \end{pmatrix} \qquad C = \begin{pmatrix} 1 & 1 \\ 0 & 2 \\ 3 & 2 \end{pmatrix}$$

We have $\mathrm{row}_2(A) = (1, 0, 5) \notin \mathbb{R}^2$ (why?), $A_{22} = 0$, while A^{T} and $\mathrm{col}_2(A)$ are

$$A^{\mathsf{T}} = \begin{pmatrix} 2 & 1 \\ 3 & 0 \\ 5 & 5 \end{pmatrix} \qquad \mathrm{col}_2(A) = \begin{pmatrix} 3 \\ 0 \end{pmatrix} \in \mathbb{R}^2$$

The sum $A + B$ and scalar product $2 \cdot A$ are given by

$$A + B = \begin{pmatrix} 3 & 4 & 5 \\ 8 & 1 & 5 \end{pmatrix} \qquad 2 \cdot A = \begin{pmatrix} 4 & 6 & 10 \\ 2 & 0 & 10 \end{pmatrix}$$

The matrix product AC is the 2×2 square matrix

$$AC = \begin{pmatrix} 2\cdot 1 + 3\cdot 0 + 5\cdot 3 & 2\cdot 1 + 3\cdot 2 + 5\cdot 2 \\ 1\cdot 1 + 0\cdot 0 + 5\cdot 3 & 1\cdot 1 + 0\cdot 2 + 5\cdot 2 \end{pmatrix} = \begin{pmatrix} 17 & 18 \\ 16 & 11 \end{pmatrix}$$

the matrix AC is not symmetric because $(AC)_{12} = 18 \neq 16 = (AC)_{21}$. On the other hand, CA is the 3×3 square matrix

$$CA = \begin{pmatrix} 1 \cdot 2 + 1 \cdot 1 & 1 \cdot 3 + 1 \cdot 0 & 1 \cdot 5 + 1 \cdot 5 \\ 0 \cdot 2 + 2 \cdot 1 & 0 \cdot 3 + 2 \cdot 0 & 0 \cdot 5 + 2 \cdot 5 \\ 3 \cdot 2 + 2 \cdot 1 & 3 \cdot 3 + 2 \cdot 0 & 3 \cdot 5 + 2 \cdot 5 \end{pmatrix} = \begin{pmatrix} 3 & 3 & 10 \\ 2 & 2 & 10 \\ 8 & 9 & 25 \end{pmatrix}$$

Clearly, we have $AC \neq CA$.

The $n \times n$ identity matrix is the matrix $I_n \in \mathbb{R}^{n \times n}$ with $I_{ij} = 1$ if $i = j$, and $I_{ij} = 0$ if $i \neq j$. That is, I_{ij} has ones along the main diagonal, and zeroes elsewhere. For every $A \in \mathbb{R}^{n \times m}$ we have that $I_n \times A = A \times I_m = A$. A square matrix $A \in \mathbb{R}^{n \times n}$ is said to be *invertible* if there exists a matrix $A^{-1} \in \mathbb{R}^{n \times n}$ such that $AA^{-1} = I_n$. When such a matrix exists it is unique, and it is called the *inverse* of A. The inverse of an invertible matrix A is itself invertible, and its inverse is matrix A. That is, if A is invertible, then $A^{-1}A = I_n$.

Let $A, B \in \mathbb{R}^{n \times n}$ be invertible. Then, $A^\intercal$ is also invertible and $(A^\intercal)^{-1} = (A^{-1})^\intercal$. For $\alpha \neq 0$, $\alpha \cdot A$ is invertiblw with $(\alpha \cdot A)^{-1} = (1/\alpha) \cdot A^{-1}$. The product AB is invertible and $(AB)^{-1} = A^{-1}B^{-1}$. Section **??** discusses how to determine whether a matrix is invertible, and how to find its inverse.

2. Linear transformations

Let V and W be linear spaces over $\mathbb{R}$. A function $f : V \to W$ is called a *linear function* if and only if

$$f(\alpha \cdot \mathbf{x} + \mathbf{y}) = \alpha \cdot f(\mathbf{x}) + f(\mathbf{y})$$

for every $\mathbf{x}, \mathbf{y} \in V$ and every scalar α.

Example 2.1 Given a matrix $A \in \mathbb{R}^{n \times m}$ and a vector $\mathbf{b} \in \mathbb{R}^m$, the function $f : \mathbb{R}^n \to \mathbb{R}^m$ given by $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$ is linear if and only if $\mathbf{b} = \mathbf{0}$ (why?)

Example 2.2 Many limit operators $f : \mathcal{C}^\infty([0, 1]) \to \mathbb{R}$ are linear functions. This includes derivatives, integrals, expectations, and limits.

Proposition 1 *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is linear if and only if there exists an $m \times n$ matrix A such that $f(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$*

Proof. ($\Rightarrow$) Suppose $f(\mathbf{x}) = A\mathbf{x}$. Then, using the properties of matrix product, $f(\alpha \cdot \mathbf{x} + \mathbf{y}) = A(\alpha \cdot \mathbf{x} + \mathbf{y}) = \alpha \cdot (A\mathbf{x}) + A\mathbf{y} = \alpha \cdot f(\mathbf{x}) + f(\mathbf{y})$.

($\Leftarrow$) Suppose f is linear and let A be the matrix with $\mathrm{col}_j(A) = f(\mathbf{e}^j)$ for $j = 1 \ldots$, where $\{\mathbf{e}^i \mid i = 1, \ldots, n\}$ denote the canonical basis of $\mathbb{R}^n$. Then, for every $\mathbf{x} \in \mathbb{R}^n$ we have

$$f(\mathbf{x}) = f\left(\sum_{i=1}^{n} x_i \mathbf{e}^j\right) = \sum_{i=1}^{n} f(\mathbf{e}^j) x_i = A\mathbf{x}$$

■

Hence, there is a tight link between linear transformations and matrices. Each linear functions from $\mathbb{R}^m$ to $\mathbb{R}^n$ can be identified with a specific $n \times m$ matrix. Given a linear function $f : \mathbb{R}^n \to \mathbb{R}^m$, let M_f denote the (unique) matrix such that $f(\mathbf{x}) = M_f\mathbf{x}$. As the following propositions show, the definitions of matrix inverse and matrix product are useful because they correspond to the composition and inversion of linear transformations.

Proposition 2 *Let $f, g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^m \to \mathbb{R}^r$ be linear functions, and let $\alpha \in \mathbb{R}$ be a scalar. The sum $f + g$, the scalar product $\alpha \cdot f$ and the composition $h \circ g$ are all linear functions. Moreover $M_{f+g} = M_f + M_g$, $M_{\alpha \cdot f} = \alpha \cdot M_f$, and $M_{h \circ g} = M_h M_g$*

Proposition 3 *If a linear function $f : \mathbb{R}^n \to \mathbb{R}^m$ is invertible, then f^{-1} is also a linear function and $M_{f^{-1}} = M_f^{-1}$.*

3. Linear systems I – existence and uniqueness

This section deals with systems of n linear equations over m real variables, that is, systems of the form.

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m & = & b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m & = & b_2, \\ & \vdots & \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m & = & b_n. \end{array} \tag{1}$$

Note that (1) can be written in matrix form as

$$A\mathbf{x} = \mathbf{b}, \tag{2}$$

where $\mathbf{b} = (b_1, b_2, \ldots, b_n)^\mathsf{T} \in \mathbb{R}^n$, and $A \in \mathbb{R}^{m \times n}$ is the matrix given by $A_{ij} = a_{ij}$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

Proposition 4 *Every system of linear equations has either zero, one, or infinitely many solutions.*

Proof. Suppose that (1) has more than one solution, i.e., that there exist $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ such that $A\mathbf{x} = \mathbf{b} = A\mathbf{y}$ and $\mathbf{x} \neq \mathbf{y}$. Fix any $\mu \in \mathbb{R}$, and let $\mathbf{z} = \mu\mathbf{x} + (1-\mu)\mathbf{y}$. Note that $A\mathbf{z} = A(\mu\mathbf{x} + (1-\mu)\mathbf{y}) = \mu(A\mathbf{x}) + (1-\mu)(A\mathbf{y}) = \mu\mathbf{b} + (1-\mu)\mathbf{b} = \mathbf{b}$. That is, $\mathbf{z}$ is also a solution of (1). Since μ was arbitrary, this means that there are infinitely many solutions. ■

3.1. Existence of solutions

The *image* of A is defined to be the image of the linear function $f : \mathbb{R}^m \to \mathbb{R}^n$ given by $f(\mathbf{x}) = A\mathbf{x}$, i.e., $\mathrm{img}(A) := \{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^m\}$. It follows that (1) has a solution if and only $\mathbf{b} \in \mathrm{img}(A)$.

Note that we can write

$$A\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \mathbf{x} = \sum_{j=1}^{n} \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix} x_j = \sum_{i=j}^{n} x_j \operatorname{col}_j(A), \tag{3}$$

where $\operatorname{col}_j(A) = (a_{1j}, a_{2j}, \dots, a_{nj})^\mathsf{T}$ denotes the j-th column of A. Therefore, the image of A coincides with the space generated by the columns of A, i.e., $\operatorname{img}(A) = \operatorname{span}(\{\operatorname{col}_j(A) | j = 1, \dots, m\})$. We know that the span of a set of vectors is a linear space. Hence, the image of A is a linear subspace of $\mathbb{R}^n$. The dimension of the image of A is called the *rank* of A and it is denoted by $\operatorname{rank}(A) = \dim(\operatorname{img}(A))$.

Proposition 5 (Existence) *Fix an arbitrary matrix $A \in \mathbb{R}^{n\times m}$. If $\operatorname{rank}(A) = m$, then (2) has at least one solution for every $\mathbf{b} \in \mathbb{R}^m$. Otherwise, if $\operatorname{rank}(A) < m$, then there exist some $\mathbf{b} \in \mathbb{R}^m$ for which (2) has no solution.*

Proof. From (3), it follows that $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{b} = \sum_{j=1}^{m} x_j \operatorname{col}_j(A)$. Therefore, $A\mathbf{x} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is a linear combination of the columns of A, i.e., if and only if $\mathbf{b} \in \operatorname{img}(A)$. Therefore, $A\mathbf{x} = \mathbf{b}$ has at least one solution for every $\mathbf{b} \in \mathbb{R}^m$, if and only if $\operatorname{img}(A) = \mathbb{R}^m$. Which is true if and only if $\operatorname{rank}(A) = m$. ■

3.2. Uniqueness of solutions

The *kernel* or *null space* of A is the set of solutions to the homogeneous system $A\mathbf{x} = \mathbf{0}$, i.e., $\ker(A) = \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}$. Note that $\ker(A) \neq \emptyset$ because $A\mathbf{0} = \mathbf{0}$. Also, for all $\mathbf{x}, \mathbf{y} \in \ker(A)$ and $\alpha \in \mathbb{R}$, we have that $A(\alpha\mathbf{x} + \mathbf{y}) = \alpha(A\mathbf{x}) + A\mathbf{y} = \mathbf{0}$ Consequently, the kernel of A is a linear subspace of $\mathbb{R}^m$. The dimension of the kernel of A is called the *nullity* of A, and it is denoted by $\operatorname{nul}(A) = \dim(\ker(A))$.

Proposition 6 (Uniqueness) *Fix an arbitrary matrix $A \in \mathbb{R}^{n\times m}$. If $\operatorname{nul}(A) = 0$, then (2) has at most one solution for every $\mathbf{b} \in \mathbb{R}^m$. Otherwise, if $\operatorname{nul}(A) > 0$, then there exist some $\mathbf{b} \in \mathbb{R}^m$ for which (2) has infinitely many solutions.*

Proof. The only 0-dimensional space is $\{\mathbf{0}\}$. Hence, $\text{nul}(A) = 0$ if and only if $\ker(A) = \{\mathbf{0}\}$. This is equivalent to having $\mathbf{x} - \mathbf{y} = \mathbf{0}$ whenever $A(\mathbf{x} - \mathbf{y}) = \mathbf{0}$. Which is in turn equivalent to having $\mathbf{x} = \mathbf{y}$ whenever $A\mathbf{x} = A\mathbf{y}$. ∎

3.3. Column and row spaces

There is a tight connection between propositions 5 and 6. In order to analyze it we first need an intermediate result regarding the space generated by the rows of A. Note that the ith row of A is exactly the ith column of $A^\top$ Hence, this space is nothing more than the image of $A^\top$ Even though $\text{img}(A) \subseteq \mathbb{R}^n$ and $\text{img}(A^\top) \subseteq \mathbb{R}^m$, the dimension of these spaces coincides. In other words, the column rank and the row rank always coincide.

Proposition 7 *For every matrix $A \in \mathbb{R}^{n\times m}$, $\text{rank}(A) = \text{rank}(A^\top) \leq \min\{n, m\}$.*

Proof. Suppose that $\text{rank}\, A^\top = r$. Since $\text{img}(A^\top) \subseteq \mathbb{R}^m$, we know that $r \leq m$. We will show that $\text{rank}(A) \leq r$. By definition of dimension, we know that $\text{img}(A^\top)$ has a basis $Y = \{\mathbf{y}_1, \dots \mathbf{y}_r\} \subseteq \mathbb{R}^m$ consisting of exactly r vectors. Since $\text{col}(A^\top) \subseteq \text{img}(A^\top) = \text{span}(Y)$, we know that for every $i = 1, \dots, n$ there exists $\lambda_{i1}, \dots, \lambda_{ir} \in \mathbb{R}$ such that $\text{col}\,_i(A^\top) = \sum_{k=1}^r \lambda_{ik}\mathbf{y}_k$.

Since the columns of $A^\top$ are just the rows of A transposed, this means that we can write

$$A = \begin{pmatrix} \text{col}_1(A^\top)^\top \\ \text{col}_2(A^\top)^\top \\ \vdots \\ \text{col}_n(A^\top)^\top \end{pmatrix} = \sum_{k=1}^r \begin{pmatrix} \lambda_{1k}\mathbf{y}_k^\top \\ \lambda_{2k}\mathbf{y}_k^\top \\ \vdots \\ \lambda_{nk}\mathbf{y}_k^\top \end{pmatrix} = \sum_{k=1}^r \begin{pmatrix} \lambda_{1k}y_{k1} & \lambda_{1k}y_{k2} & \dots & \lambda_{1k}y_{km} \\ \lambda_{2k}y_{k1} & \lambda_{2k}y_{k2} & \dots & \lambda_{2k}y_{km} \\ \vdots & & \ddots & \vdots \\ \lambda_{nk}y_{k1} & \lambda_{nk}y_{k2} & \dots & \lambda_{nk}y_{km} \end{pmatrix}.$$

This implies that for every $j = 1, \dots, m$ we have that

$$\text{col}\, j(A) = \sum_{k=1}^r \begin{pmatrix} \lambda_{1k}y_{kj} \\ \lambda_{2k}y_{kj} \\ \vdots \\ \lambda_{nk}y_{kj} \end{pmatrix} = \sum_{k=1}^r y_{kj} \begin{pmatrix} \lambda_{1k} \\ \lambda_{2k} \\ \vdots \\ \lambda_{nk} \end{pmatrix} = \sum_{k=1}^r y_{jk}\boldsymbol{\lambda}_k,$$

where $\boldsymbol{\lambda}_k = (\lambda_{1k}, \dots, \lambda_{nk})^\top \in \mathbb{R}^n$. Which means that the columns of A can be generated as linear combinations of $r \leq m$ vectors $\boldsymbol{\lambda}_k$, $k = 1, \ldots, r$, and hence the dimension of their span cannot be greater than r. We have thus shown that $\operatorname{rank}(A) \leq \operatorname{rank}(A^\mathsf{T}) \leq m$. To see that $\operatorname{rank}(A^\mathsf{T}) \leq \operatorname{rank}(A) \leq n$, simply note that $(A^\mathsf{T})^\mathsf{T} = A$. ■

3.4. Relation between existence and uniqueness

Our sufficient conditions for existence and uniqueness are expressed in terms of the rank, and the nullity of A, respectively. As it turn out the rank and nullity of a matrix are tightly connected.

Proposition 8 *For every matrix* $A \in \mathbb{R}^{n \times m}$, $\operatorname{rank}(A) + \operatorname{nul}(A) = m$.

Proof. $A\mathbf{x} = \mathbf{0}$ if and only if $\mathbf{x}^\mathsf{T} A^\mathsf{T} = \mathbf{0}^\mathsf{T}$. This is equivalent to having $\mathbf{x}^\mathsf{T} \operatorname{col}_j(A^\mathsf{T}) = 0$ for all $j = 1, \ldots, m$. Hence $\ker(A)$ is the orthogonal complement of $\operatorname{img}(A^\mathsf{T})$. Using problem 3.(c) in Problem Set 4, this implies that $\operatorname{nul}(A) = \operatorname{rank}(A^\mathsf{T})$. The result then follows from Proposition 7. ■

Proposition 8 implies that $\operatorname{nul}(A) = 0$ if and only if $\operatorname{rank}(A) = m$. Therefore, our sufficient conditions for existence and uniqueness coincide! It follows that if $\operatorname{rank}(A) = m$, then (3) has a unique solution for every $\mathbf{b}$. On the other hand, if $\operatorname{rank}(A) < m$, then there exist vectors $\mathbf{b}$ for which (3) has no solution, and there exist vectors $\mathbf{b}$ for which (3) has an infinity of solutions.