INTRODUCTION TO LINEAR ALGEBRA

Hiroshi SUZUKI*
International Christian University

September 30, 2010

1 Introduction

2 System of Linear Equations

A *matrix* (or an $m \times n$ *matrix*) is an $m \times n$ rectangular array of numbers.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots\cdots\cdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Definition 2.1 An arbitrary system of m linear equations in n unknowns can be written as

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \cdots\cdots\cdots & & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{cases}$$

where $x_1, x_2, \ldots, x_n$ are the unknowns.

A sequence of numbers $s_1, s_2, \ldots, s_n$ is called a *solution* of the system if $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$ is a solution of every equation in the system, i.e., every equation is satisfied when we substitute $x_1 = s_1, x_2 = s_2, \ldots, x_n = s_n$. The set of all solutions of the system is called its *solution set* or the *general solution* of the system.

A system of equations that has no solutions is said to be *inconsistent*; if there is at least one solution of the system, it is called *consistent*.

The *augmented matrix* A or extended coefficient matrix, and the *coefficient matrix* C of this system are defined as follows.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix},$$

$$C = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots\cdots\cdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

Definition 2.2 The following are called *elementary row operations.*

1. $[i; c]$: Multiply row i through by a nonzero constant c.
2. $[i, j]$: Interchange rows i and j.
3. $[i, j; c]$: Add c times row j to row i.

Definition 2.3 An $m \times n$ matrix A is in *reduced row-echelon form* if the following conditions hold:

1. If a row does not consist entirely of zeros, then the first nonzero number in the row (from the left) is a 1. We call this a *leading* 1.
2. If there are any rows that consist entirely of zeros, then they are grouped together at the bottom of the matrix.
3. In any successive rows that do not consist entirely of zeros, the leading 1 in the lower row occurs farther to the right than the leading 1 in the higher row.
4. Each column that contains a leading 1 has zeros everywhere else in that column.

If the conditions 1, 2 and 3 are satisfied A is said to be in *row-echelon form.*

Theorem 2.1 (Gauss-Jordan Elimination) *Every matrix can be transformed into a reduced row-echelon form by applying elementary row operations successively finitely many times.*

Theorem 2.2 *Let A be the augmented matrix of a system of linear equations in n unknowns $x_1, x_2, \ldots, x_n$ and B the reduced echelon form of A. Suppose the leading 1's of B are at columns $i_1, i_2, \ldots, i_r$ and rows $j_1, j_2, \ldots j_{n-r}$ are those colums without leading 1's. Then the following hold.*

(i) *The system of linear equations is inconsistent if and only if $i_r = n + 1$, i.e., a leading 1 is in the last column.*

(ii) *The system of linear equation has a unique solution if and only if it is consistent and $r = n$, i.e., $i_1 = 1, i_2 = 2, \ldots, i_r = i_n = n$.*

(iii) *If the system is consistent, the solutions are expressed with $n - r$ parameters $t_1 = x_{j_1}, t_2 = x_{j_2}, \ldots, t_{n-r} = x_{j_{n-r}}$.*

*E-mail:hsuzuki@icu.ac.jp

The number r in the previous theorem is called the *rank* of B (and A).

3 Matrices and Matrix Operations

Definition 3.1 A *matrix* is an $m \times n$ rectangular array of numbers. The numbers in the array are called the *entries* in the matrix.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots\cdots\cdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

It is called an $m \times n$ matrix, a matrix with m rows and n columns, it is also denoted by $A = [a_{ij}]$. Two matrices are defined to be *equal* if they have the same size and their corresponding entries are equal. The entry $a_{i,j}$ in the i-th row j-th column of a matrix A is denoted by $(A)_{i,j}$.

An $n \times n$ matrix is called a *square matrix.*

Definition 3.2 Let A and B be matrices of the same size and c a scalar. Then the *sum* $A + B$ is the matrix obtained by adding the entries of B to the corresponding entries of A. The the *product* cA is the matrix obtained by multiplying each entry of the matrix A by c. The matrix cA is said to be a it scalar multiple of A.

$$A+B = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2n}+b_{2n} \\ & \cdots\cdots\cdots & & \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \cdots & a_{mn}+b_{mn} \end{bmatrix},$$

$$cA = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & ca_{22} & \cdots & ca_{2n} \\ & \cdots\cdots\cdots & & \\ ca_{m1} & ca_{m2} & \cdots & ca_{mn} \end{bmatrix}$$

Definition 3.3 If $A = (a_{i,j})$ is an $m \times r$ matrix and $B = (b_{k,l})$ is an $r \times n$ matrix, then the *product* $C = AB$ is the $m \times n$ matrix whose (s, t) entry $c_{s,t}$ is defined as follows.

$$c_{s,t} = \sum_{u=1}^{r} a_{s,u}b_{u,t} = a_{s,1}b_{1,t} + a_{s,2}b_{2,t} + \cdots + a_{s,r}b_{r,t}.$$

Proposition 3.1 *Let A be an $m \times r$ matrix and $B = [\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n]$ be an $r \times n$ matrix whose j-th column is $\boldsymbol{b}_j$. If $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_m$ be the rows of A, then $(AB)_{i,j} = \boldsymbol{a}_i\boldsymbol{b}_j$,*

$$AB = A[\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n] = [A\boldsymbol{b}_1, A\boldsymbol{b}_2, \ldots, A\boldsymbol{b}_n] \text{ and}$$

$$AB = \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_m \end{bmatrix} B = \begin{bmatrix} \boldsymbol{a}_1 B \\ \boldsymbol{a}_2 B \\ \vdots \\ \boldsymbol{a}_m B \end{bmatrix}.$$

Definition 3.4 If A is an $m \times n$ matrix, then the *transpose* of A, denoted by A^T, is defined to be the $n \times m$ matrix that results from interchanging the rows and columns of A, that is $(A^T)_{i,j} = A_{j,i}$ $(1 \le i \le n, 1 \le j \le m)$.

Definition 3.5 If A is a square matrix, then the trace of A, denoted by $\mathrm{tr}(A)$, is defined to be the sum of the entries on the main diagonal of A. The trace of A is undefined if A is not a square matrix.

Proposition 3.2 *Let $A = (a_{i,j})$ be an $m \times n$ matrix and $B = (b_{jk})$ an $n \times m$ matrix. Then $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.*

Properties of Arithmetic:

Theorem 3.3 (1.4.1) *Assuming that the sizes of the matrices are such that the indicated operations can be performed, the following rules of matrix arithmetic are valid.*

(a) $A + B = B + A$.

(b) $A + (B + C) = (A + B) + C$.

(c) $A(BC) = (AB)C$.

(d) $A(B + C) = AB + AC$.

(e) $(B + C)A = BA + CA$.

(f) $A(B - C) = AB - AC$.

(g) $(B - C)A = BA - CA$.

(h) $a(B + C) = aB + aC$.

(i) $a(B - C) = aB - aC$.

(j) $(a + b)C = aC + bC$.

(k) $(a - b)C = aC - bC$.

(l) $a(bC) = (ab)C$.

(m) $a(BC) = (aB)C = B(aC)$.

Proof. The proof of (d) is in the textbook. ■

Theorem 3.4 (1.4.9) *If the sizes of the matrices are such that the stated operations can be performed, then:*

(a) $(A^T)^T = A$.

(b) $(A+B)^T = A^T + B^T$ *and* $(A-B)^T = A^T - B^T$.

(c) $(kA)^T = kA^T$, *where k is any scalar.*

(d) $(AB)^T = B^T A^T$.

Proof. The proof of (d) is in the textbook. ■

Lemma 3.5 *Let I be an $r \times s$ matrix such that $IX = X$ for all $m \times n$ matrix. Then $r = s = m$ and I is a square matrix with 1's on the main diagonal and 0's off the main diagonal.*

4 Inverse of Matrices

In order to solve a system of linear equations expressed in a matrix equation $A\boldsymbol{x} = \boldsymbol{b}$, we wanted to have a matrix B such that $BA = I$ and $AB = I$, where I is a matrix performs as 1.

Definition 4.1 A square matrix with 1's on the main diagonal and 0's off the main diagonal is called an *identity matrix* and is denoted by I, or I_n when it is of size $n \times n$.

If A is a square matrix, and if a matrix B of the same size can be found such that $AB = BA = I$, then A is said to be *invertible* and B is called the inverse of A. If not such matrix B can be found, then A is said to be *singular*.

Theorem 4.1 *Let A be an $n \times n$ square matrix, and $I = I_n$ the identity matrix of size n. Set $C = [A \mid I]$. If the reduced row echelon form of C is of form $[I \mid B]$, then $B = A^{-1}$, otherwise the inverse of A does not exist.*

Proposition 4.2 (1.4.6, 1.4.9, 1.4.10) (a) *If both A and B are invertible matrices. Then AB is also invertible and $(AB)^{-1} = B^{-1}A^{-1}$.*

(b) *If A is an $m \times r$ matrix and B is an $r \times n$ matrix, then $(AB)^T = B^T A^T$.*

(c) *If A is an invertible matrix, then A^T is invertible and $(A^T)^{-1} = (A^{-1})^T$.*

Definition 4.2 An $n \times n$ matrix is called an *elementary matrix* if it can be obtained from the $n \times n$ identity matrix I_n by performing a single elementary row operation.

1. $P(i;c)$: the matrix obtained from I_n by performing $[i;c]$ $(c \neq 0)$.
2. $P(i,j)$: the matrix obtained from I_n by performing $[i,j]$.
3. $P(i,j;c)$: the matrix obtained from I_n by performing $[i,j;c]$.

Proposition 4.3 (1.5.1) *If the elementary matrix E results from performing a certain row operation on I_m and A is an $m \times n$ matrix, then the product EA is the matrix that results when this same row operation is performed on A.*

Proposition 4.4 (1.5.2) *Every elementary matrix is invertible, and the inverse is also an elementary matrix.*

$$P(i;c)^{-1} = P(i;1/c),\ P(i,j)^{-1} = P(i,j),$$

$$\text{and } P(i,j;c)^{-1} = P(i,j;-c).$$

5 Invertibility of a Matrix and its Applications

First goal of this section is to prove the following theorem.

Theorem 5.1 (1.5.3, 1.6.4) *If A is an $n \times n$ matrix, then the following are equivalent.*

(a) *A is invertible.*

(b) *$A\boldsymbol{x} = \boldsymbol{0}$ has only the trivial solution, i.e., $\boldsymbol{x} = \boldsymbol{0}$.*

(c) *The reduced row-echelon form of A is I_n.*

(d) *A is expressible as a product of elementary matrices.*

(e) *$A\boldsymbol{x} = \boldsymbol{b}$ is consistent for every $n \times 1$ matrix $\boldsymbol{b}$.*

(f) *$A\boldsymbol{x} = \boldsymbol{b}$ has exactly one solution for every $n \times 1$ matrix $\boldsymbol{b}$.*

Proposition 5.2 (1.6.3) *Let A and B be square matrices of the same size. Then*

(a) *If $BA = I$ then $AB = I$. In particular, both A and B are invertible and $B = A^{-1}$, $A = B^{-1}$.*

(b) *AB is invertible if and only if both A and B are invertible.*

A System of Linear Equations:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \cdots\cdots\cdots & & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{cases},$$

$$B = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}$$

where $x_1, x_2, \ldots, x_n$ are the unknowns.

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ & \cdots\cdots\cdots & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix},\ \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix},\ B = [A, \boldsymbol{b}].$$

If the product $A\boldsymbol{x}$ of A and $\boldsymbol{x}$ is defined as follows, the linear system above can be written as $A\boldsymbol{x} = \boldsymbol{b}$.

We obtained the reduced echelon form C of B by performing a sequence of elementary row operations.

Let $P_1, P_2, \ldots, P_\ell$ be the corresponding elementary matrices, and $P = P_\ell P_{\ell-1} \cdots P_1$. Then

$$PB = C,\ P[A, \boldsymbol{b}] = [PA, P\boldsymbol{b}] \text{ and } PA\boldsymbol{x} = P\boldsymbol{b}.$$

$$\begin{aligned} &\boldsymbol{x} \text{ is a solution to } PA\boldsymbol{x} = P\boldsymbol{b} \\ \Leftrightarrow\quad &\boldsymbol{x} \text{ is a solution to } A\boldsymbol{x} = \boldsymbol{b}. \end{aligned}$$

We call the number of leading 1's in the reduced echelon form of a matrix the *rank* of the matrix.

Theorem 5.3 *The reduced row echelon form of a matrix is unique.*

Theorem 5.4 *Let A be an $m \times n$ matrix and $A\boldsymbol{x} = \boldsymbol{b}$ is a system of linear equation expressed as a matrix equation. Let $B = [A, \boldsymbol{b}]$ be the augmented matrix and $C = [A_1, \boldsymbol{b}_1]$ is the reduced echelon form of B. Then*

$$\begin{aligned} &\boldsymbol{x} \textit{ is a solution to } A\boldsymbol{x} = \boldsymbol{b} \\ \Leftrightarrow\quad &\boldsymbol{x} \textit{ is a solution to } A_1\boldsymbol{x} = \boldsymbol{b}_1. \end{aligned}$$

Moreover,

(i) *If* $\mathrm{rank}(A) \neq \mathrm{rank}(B)$, *the equation* $A\boldsymbol{x} = \boldsymbol{b}$ *is inconsistent.*

(ii) *If* $\mathrm{rank}(A) = \mathrm{rank}(B) = n$, *the equation* $A\boldsymbol{x} = \boldsymbol{b}$ *has the only one solution.*

(iii) *If* $\mathrm{rank}(A) = \mathrm{rank}(B) < n$, *the equation* $A\boldsymbol{x} = \boldsymbol{b}$ *has solutions with* $n - \mathrm{rank}(A)$ *parameters.*

Theorem 5.5 *Let $A\boldsymbol{x} = \boldsymbol{b}$ be an equation. Suppose $A\boldsymbol{x}_0 = \boldsymbol{b}$. Then*

$$\begin{aligned} &\boldsymbol{x} \textit{ is a solution to } A\boldsymbol{x} = \boldsymbol{b} \\ \Leftrightarrow\quad &\boldsymbol{x} = \boldsymbol{x}_0 + \boldsymbol{y} \textit{ where } \boldsymbol{y} \textit{ is a solution to } A\boldsymbol{x} = \boldsymbol{0}. \end{aligned}$$

$\boldsymbol{x}_0$ *is called a particular solution.*

Proof. $\Rightarrow$: Let $\boldsymbol{y} = \boldsymbol{x}_0 - \boldsymbol{x}$. Then $A\boldsymbol{y} = A(\boldsymbol{x}_0 - \boldsymbol{x}) = A\boldsymbol{x}_0 - A\boldsymbol{x} = \boldsymbol{b} - \boldsymbol{b} = \boldsymbol{0}$. Hence $\boldsymbol{y}$ is a solution to $A\boldsymbol{x} = \boldsymbol{0}$.

$\Leftarrow$: Suppose $\boldsymbol{y}$ is a solution to $A\boldsymbol{x} = \boldsymbol{0}$. Then $A\boldsymbol{y} = \boldsymbol{0}$. Hence $A(\boldsymbol{x}_0 + \boldsymbol{y}) = A\boldsymbol{x}_0 + A\boldsymbol{y} = \boldsymbol{b}$, and $\boldsymbol{x}_0 + \boldsymbol{y}$ is a solution to $A\boldsymbol{x} = \boldsymbol{b}$. ■

Definition 5.1 If $b_1 = b_2 = \cdots = b_m = 0$, the system is called *homogeneous*. Homogeneous system is always consistent as $x_1 = x_2 = \cdots = x_n = 0$ is a solution. This solution is called the *trivial* solution; if there are other solutions, they are called *nontrivial solutions*.

Theorem 5.6 (1.2.1) *A homogeneous system of linear equations with more unknowns than equations has infinitely many solutions. In particular, if A is an $m \times n$ matrix, then a system of linear equation $A\boldsymbol{x} = \boldsymbol{0}$ has a non-trivial solution.*

Proof. Every homogeneous system of linear equations is consistent. The rank of the augmented matrix is at most the number of rows. Hence the number of parameters to express solutions is $n - \mathrm{rank}(A) \geq n - m > 0$. Thus there are infinitely many solutions, in particular the system has a non-trivial solution. ■

Example 5.1 The matrix below is the augmented matrix of the system of linear equations on the left below.

$$\left\{ \begin{array}{rcr} x_1 + 0x_2 + x_3 + 0x_4 + x_5 + 3x_6 & = & -1 \\ -x_1 + 0x_2 - x_3 + 0x_4 + 0x_5 - 4x_6 & = & -1 \\ 0x_1 + x_2 - 2x_3 + 3x_4 + 0x_5 - x_6 & = & 3 \\ -2x_1 - 2x_2 + 2x_3 - 6x_4 - 2x_5 - 4x_6 & = & -4 \end{array} \right\}$$

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 3 & -1 \\ -1 & 0 & -1 & 0 & 0 & -4 & -1 \\ 0 & 1 & -2 & 3 & 0 & -1 & 3 \\ -2 & -2 & 2 & -6 & -2 & -4 & -4 \end{bmatrix}$$

$$\xrightarrow{[4,1;2]} \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 3 & -1 \\ -1 & 0 & -1 & 0 & 0 & -4 & -1 \\ 0 & 1 & -2 & 3 & 0 & -1 & 3 \\ 0 & -2 & 4 & -6 & 0 & 2 & -6 \end{bmatrix}$$

$$\xrightarrow{[2,1;1]} \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 3 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & -2 \\ 0 & 1 & -2 & 3 & 0 & -1 & 3 \\ 0 & -2 & 4 & -6 & 0 & 2 & -6 \end{bmatrix}$$

$$\xrightarrow{[2,3]} \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 3 & -1 \\ 0 & 1 & -2 & 3 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & -1 & -2 \\ 0 & -2 & 4 & -6 & 0 & 2 & -6 \end{bmatrix}$$

$$\xrightarrow{[4,2;2]} \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 3 & -1 \\ 0 & 1 & -2 & 3 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\xrightarrow{[1,3;-1]} \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 4 & 1 \\ 0 & 1 & -2 & 3 & 0 & -1 & 3 \\ 0 & 0 & 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\left\{ \begin{array}{rcc} x_1 & = & 1 - s - 4u, \\ x_2 & = & 3 + 2s - 3t + u, \\ x_3 & = & s, \\ x_4 & = & t, \\ x_5 & = & -2 + u, \\ x_6 & = & u. \end{array} \right. \quad \text{or}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 0 \\ 0 \\ -2 \\ 0 \end{bmatrix} + s \begin{bmatrix} -1 \\ 2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 0 \\ -3 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + u \begin{bmatrix} -4 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

s, t and u are parameters.

The solution above is the general solution and $x_1 = 1, x_2 = 3, x_3 = 0, x_4 = 0, x_5 = -2, x_6 = 0$ or the 6×1 matrix corresponding to it is a particular solution.