Chapter 5
Linear Equations: Basic Theory and Practice

In this chapter and the next, we are interested in the linear algebraic equation

$$A\bar{X} = \bar{b}, \quad (5-1)$$

where $A$ is an $m \times n$ matrix, $\bar{X}$ is an $n \times 1$ vector to be solved for, and $\bar{b}$ is an $m \times 1$ vector. Relative to the material usually found in most engineering mathematics books, our presentation is fairly comprehensive. In addition, we discuss functionality in MatLab that can be applied to this problem.

**Row-Echelon Form (Hermite Normal Form)**

With respect to given bases on vector spaces $U$ and $V$, suppose $m \times n$ matrix $A$ represents the linear transformation $T : U \rightarrow V$. As discussed in Chapter 3, a change of basis on space $V$ can be effected by using an $m \times m$, nonsingular matrix $Q$ to relate the "original" and "new" bases, and this produces a new matrix $A' = Q^{-1}A$ that represents $T$ with respect to the "original" basis on $U$ and the "new" basis on $V$. The logical question to ask here is: How simple can we make $A'$ by changing only the basis on $V$? An answer to this question is given by the following theorem.

**Theorem 5.1**

Given any $m \times n$ matrix $A$ of rank $\rho$, there exists a nonsingular, $m \times m$ matrix $Q$ such that $A' \equiv Q^{-1}A$ has the form described by the three attributes given below.

1) There is at least one non-zero element in each of the first $\rho$ rows of $A'$, and the elements in all remaining rows (after $\rho$) are zero.

2) The first non-zero element appearing in row $i$ ($i \leq \rho$) is a 1 appearing in column $\kappa_i$, where $\kappa_1 < \kappa_2 < \ldots < \kappa_\rho$.

3) In column $\kappa_i$, $1 \leq i \leq \rho$, the only non-zero element is the 1 in row $i$.

The matrix $A'$ is uniquely determined by $A$. 
Matrix $A'$ described by 1) through 3) is said to be in row-echelon form (also known as Hermite Normal Form). It is important to realize that performing the product $Q^{-1}A$ is tantamount to performing row operations on $A$, our next topic.

**Elementary Row Operations and Elementary Matrices**

We define three types of elementary row operations on an $m \times n$ matrix $A$.

**Type I:** Multiply a row by a non-zero scalar.

**Type II:** Add a multiple of one row to another row.

**Type III:** Interchange two rows.

The product $A' = Q^{-1}A$, in row echelon form, can be obtained by performing (usually multiple) elementary row operations on $A$.

Each of the elementary row operations can be accomplished by multiplying $m \times n$ matrix $A$ on the left by an elementary matrix obtained by performing the elementary row operation on an $m \times m$ identity matrix. For example, the second row of $A$ can be multiplied by the scalar $c$ if matrix $A$ is multiplied on the left by the $m \times m$ elementary matrix

$$
\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
0 & c & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1
\end{bmatrix}
$$

And, if to the matrix $A$, we add to the first row the constant $k$ times the 3rd row we get the same result as if we were to multiply $A$ on the left by the elementary matrix

$$
\begin{bmatrix}
1 & 0 & k & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1
\end{bmatrix}
$$
These *elementary matrices* are non-singular. Also, their inverses are elementary matrices.

As stated earlier, given an \( m \times n \) matrix \( A \) there exists a nonsingular \( m \times m \) matrix \( Q \) such that \( A' = Q^{-1}A \) is in row-echelon form. \( A' \) can be generated by applying elementary row operations to \( A \). Apply these same row operation to the identity matrix to obtain \( Q^{-1} \).

**Example**

\[
A = \begin{bmatrix}
4 & 3 & 2 & -1 \\
5 & 4 & 3 & -1 \\
2 & -2 & -1 & 2
\end{bmatrix}, \quad [A : I] = \begin{bmatrix}
4 & 3 & 2 & -1 & | & 1 & 0 & 0 \\
5 & 4 & 3 & -1 & | & 0 & 1 & 0 \\
2 & -2 & -1 & 2 & | & 0 & 0 & 1
\end{bmatrix}
\]

Multiply the 1st row by 1/4; add to 2nd row -5 times 1st row; add to 3rd row -2 time 1st row to obtain

\[
\begin{bmatrix}
1 & 3/4 & 1/2 & -1/4 & | & 1/4 & 0 & 0 \\
0 & 1/4 & 1/2 & 1/4 & | & -5/4 & 1 & 0 \\
0 & -14/4 & -2 & 5/2 & | & -1/2 & 0 & 1
\end{bmatrix}
\]

Multiply 2nd row by 4; add to 1st row -3/4 times 2nd row; add to 3rd row 7/2 times 2nd row to obtain

\[
\begin{bmatrix}
1 & 0 & -1 & -1 | 4 & -3 & 0 \\
0 & 1 & 2 & 1 | -5 & 4 & 0 \\
0 & 0 & 5 & 6 | -18 & 14 & 1
\end{bmatrix}
\]

Multiply 3rd row by 1/5; add to 2nd row -2 times 3rd row; add to 1st row 1 times 3rd row to obtain
We are done! The row-echelon form and the transformation matrix are

\[
A' = \begin{bmatrix} 1 & 0 & 0 & 1/5 \\
0 & 1 & 0 & -7/5 \\
0 & 0 & 1 & 6/5 \end{bmatrix}, \quad Q^{-1} = \begin{bmatrix} 2/5 & -1/5 & 1/5 \\
11/5 & -8/5 & -2/5 \\
-18/5 & 14/5 & 1/5 \end{bmatrix}.
\]

Let's check our work: \( Q^{-1} A = A' \)?

\[
\begin{bmatrix} 2/5 & -1/5 & 1/5 \\
11/5 & -8/5 & -2/5 \\
-18/5 & 14/5 & 1/5 \end{bmatrix} \begin{bmatrix} 4 & 3 & 2 & -1 \\
5 & 4 & 3 & -1 \\
2 & -2 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1/5 \\
0 & 1 & 0 & -7/5 \\
0 & 0 & 1 & 6/5 \end{bmatrix}
\]

YES!♥

If matrix \( A \) is nonsingular, the row echelon form is the identity matrix, and \( Q^{-1} \) is the inverse of \( A \). This method of finding \( A^{-1} \) is one of the easiest available for hand computation, and it is the recommended technique.

**MatLab’s RREF Function**

MatLab can reduce a matrix to row echelon form. The syntax is

\[
R = \text{RREF}(A),
\]

where \( A \) is a user-supplied m×n matrix. MatLab returns the row echelon form in the matrix \( R \). This function can do more than just the row echelon form of a matrix. Consult the MatLab documentation for information on the capabilities of (and full syntax for) the RREF function.
**Linear Algebraic Equations**

Let \( T: U \to V \) be a linear transformation. Let \( \vec{b} \in V \) be any given vector. The problem of finding all \( \vec{X} \in U \) such that \( T(\vec{X}) = \vec{b} \) is a problem of great practical importance.

Of course, once bases have been established for \( U \) and \( V \), we can recast the linear algebraic problem in terms of matrices and coordinate vectors. Suppose \( m \times n \) matrix \( A \) represents transformation \( T \) with respect to some given set of bases. Then, the linear algebraic problem consists of finding all coordinate vectors \( \vec{X} \) such that \( A\vec{X} = \vec{b} \), where \( \vec{b} \) is a coordinate vector representing (with respect to a basis) some vector in \( V \). The meaning of \( \vec{X} \) and \( \vec{b} \) has to be taken from context. If we write \( T(\vec{X}) = \vec{b} \), we mean that \( \vec{X} \in U \) and \( \vec{b} \in V \) are actual vectors. If we write \( A\vec{X} = \vec{b} \), we mean that \( \vec{X} \) and \( \vec{b} \) are “only” coordinates that represent “real” vectors with respect to the underlying bases (and, with respect to the bases, matrix \( A \) represents transformation \( T \)).

In terms of the transformation \( T \), if \( \vec{b} \) is not in the range of \( T \), then a solution to \( T(\vec{X}) = \vec{b} \) does not exist, and the equation is said to be *inconsistent*. If \( \vec{b} \) is in the range of \( T \), then at least one solution exists, and the equation is said to be *consistent*.

Suppose \( \vec{X}_1 \) and \( \vec{X}_2 \) are solutions of \( T(\vec{X}) = \vec{b} \). Then \( T(\vec{X}_1 - \vec{X}_2) = \vec{0} \), so the difference \( \vec{X}_1 - \vec{X}_2 \) of two solutions is in the kernel of \( T \) (we write \( \vec{X}_1 - \vec{X}_2 \in \text{K}(T) \)). Conversely, let \( \vec{Z} \in \text{K}(T) \), and suppose that \( \vec{X}_0 \) is *any* particular solution of \( T(\vec{X}) = \vec{b} \). Then we have \( T(\vec{X}_0 + \vec{Z}) = T(\vec{X}_0) + T(\vec{Z}) = \vec{b} + \vec{0} = \vec{b} \), so \( \vec{X}_0 + \vec{Z} \) is also a solution. Since \( \text{K}(T) \) is a subspace of \( U \), all solutions of \( T(\vec{X}) = \vec{b} \) form a translation of the subspace \( \text{K}(T) \). That is, the solution set for equation \( T(\vec{X}) = \vec{b} \) is a linear variety.

Of course, these ideas have counterparts when dealing with matrices and coordinate vectors. Given a coordinate vector \( \vec{X} \), the product \( A\vec{X} \) is nothing more than a linear combination of the columns of matrix \( A \). Hence, for a solution of \( A\vec{X} = \vec{b} \) to exist, it is necessary and sufficient
for \( \vec{b} \) to be represented as a linear combination of the columns of \( A \); that is \( \vec{b} \in \text{Span}(\vec{A}_1, \vec{A}_2, \ldots, \vec{A}_n) \), where \( \vec{A}_1, \vec{A}_2, \ldots, \vec{A}_n \) are the \( n \) columns of the \( m \times n \) matrix \( A \) (we write \( A = [\vec{A}_1 | \cdots | \vec{A}_n] \)).

Likewise, let \( \vec{X}_1 \) and \( \vec{X}_2 \) satisfy \( A\vec{X} = \vec{b} \); then \( A(\vec{X}_1 - \vec{X}_2) = \vec{0} \), so that \( \vec{X}_1 - \vec{X}_2 \in \text{K}(A) \). Conversely, let \( \vec{X}_0 \) be any particular solution of \( A\vec{X} = \vec{b} \); for all \( \vec{Z} \in \text{K}(A) \) (\( i.e., A\vec{Z} = \vec{0} \)), we have \( \vec{X}_0 + \vec{Z} \) a solution of \( A\vec{X} = \vec{b} \). Hence, the solution set is a translation of \( \text{K}(A) \).

**Example**

\[
A = \begin{bmatrix} 1 & -1 \\ 3 & -3 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}
\]

Observe that \( \vec{X}_0 = [2 \ 1]^T \) is a particular solution of \( A\vec{X} = \vec{b} \), and \( \text{K}(A) = \alpha[1 \ 1]^T \), where \( \alpha \in \mathbb{R} \). Hence, the set of all solutions of \( A\vec{X} = \vec{b} \) is

\[
\begin{bmatrix} 2 \\ 1 \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 + \alpha \\ 1 + \alpha \end{bmatrix}.
\]

Let \( A \) be an \( m \times n \) matrix. The system \( A\vec{X} = \vec{b} \) is said to be *overdetermined* if \( m > n \). Often, overdetermined algebraic systems occur when we formulate too many constraints when modeling a physical system. Sometimes, due to measurement errors, this may lead to an algebraic system that has no solution. The system is said to be *underdetermined* if \( m < n \). When modeling a physical system, an underdetermined equation may result if we have incomplete knowledge about the interrelationship between system variables.

**Solution Existence: A Necessary and Sufficient Condition**

A necessary and sufficient condition for \( A\vec{X} = \vec{b} \) to have a solution can be stated in terms of the ranks of matrices.
Theorem 5.2

Given

\[ A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \bar{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}, \] (5-5)

define the *augmented* system

\[ [A : \bar{b}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}, \] (5-6)

The system \( A\hat{X} = \bar{b} \) has a solution if and only if

\[ \text{Rank} [A : \bar{b}] = \text{Rank}(A). \] (5-7)

When a solution(s) exist, the collection of solutions can be expressed in terms of \( n - r \), where \( r = \text{Rank}(A) \), independent parameters.

**Proof:** As discussed previously, a solution exists if and only if \( \bar{b} \) can be expressed as a linear combination of the columns of \( A \). Hence, using \( \bar{b} \) as an additional column cannot increase rank, and (5-7) must hold. Now, \( \text{nullity}(A) = \dim(K(A)) = n - r \), so there are linearly independent \( \hat{Z}_1, \hat{Z}_2, \ldots, \hat{Z}_{n-r} \) such that \( A\hat{Z}_k = \bar{0}, \; 1 \leq k \leq n-r \). The *complete solution* of \( A\hat{X} = \bar{b} \) can be expressed as

\[ \hat{X} = \hat{X}_0 + \sum_{k=1}^{n-r} \alpha_k \hat{Z}_k, \] (5-8)
where $\alpha_k$, $1 \leq k \leq n-r$, are independent parameters that can be assigned arbitrary values, and $\bar{X}_0$ is any particular solution.

The complete solution can be found by using elementary operations to reduce $[A \; \bar{b}]$ to row echelon form.

**Example**

$$
A = \begin{bmatrix}
4 & 3 & 2 & -1 \\
5 & 4 & 3 & -1 \\
-2 & -2 & -1 & 2 \\
11 & 6 & 4 & 1
\end{bmatrix}, \quad \bar{b} = \begin{bmatrix}
4 \\
4 \\
-3 \\
11
\end{bmatrix}, \quad [A \; \bar{b}] = \begin{bmatrix}
4 & 3 & 2 & -1 & 4 \\
5 & 4 & 3 & -1 & 4 \\
-2 & -2 & -1 & 2 & -3 \\
11 & 6 & 4 & 1 & 11
\end{bmatrix}
$$

The row echelon form is

$$
[A' \; \bar{b}'] = \begin{bmatrix}
1 & 0 & 0 & 1 & 1 \\
0 & 1 & 0 & -3 & 2 \\
0 & 0 & 1 & 2 & -3 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}.
$$

Hence, an equivalent set of equations is

$$
x_1 + x_4 = 1 \\
x_2 - 3x_4 = 2 \\
x_3 + 2x_4 = -3
$$

$\iff$

$$
x_1 = 1 - x_4 \\
x_2 = 2 + 3x_4 \\
x_3 = -3 - 2x_4
$$

were $x_4$ can be thought of as an independent parameter that can be assigned an arbitrary value. This solution set can be expressed as
Here, the vector $[1 \ 2 \ -3 \ 0]^T$ is a particular solution and $[-1 \ 3 \ -2 \ 1]^T$ is in the one-dimensional kernel of $A$. ♥

The method illustrated by the last example leads to an inconsistent result when the system of equations has no solution.

**Example**

\[
A = \begin{bmatrix} 1 & 4 & 5 \\ 2 & 5 & 7 \\ 3 & 6 & 9 \end{bmatrix}, \quad \vec{b} = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}
\]

MatLab gives $\text{rank}(A) = 2$ and $\text{rank}[A \mid \vec{b}] = 3$, which implies that the system is inconsistent. The augmented matrix is

\[
\begin{bmatrix} A \mid \vec{b} \end{bmatrix} = \begin{bmatrix} 1 & 4 & 5 & 1 \\ 2 & 5 & 7 & -2 \\ 3 & 6 & 9 & 1 \end{bmatrix},
\]

MatLab's `rref` function yields

\[
\text{rref} \left( \begin{bmatrix} A \mid \vec{b} \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.
\]

The last row confirms our initial analysis: the system is inconsistent since $0 \neq 1$. ♥
As we have argued, the system $A\mathbf{X} = \mathbf{b}$ has a solution if and only if $\mathbf{b} \in \text{R}(A) = \{\text{span of columns of } A\}$. But, the whole coordinate space (the set of all coordinate vectors used to represent vectors in $V$) for $V$ can be expressed as $\text{R}(A) \oplus \text{R}(A)^\perp$. Hence, the system $A\mathbf{X} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is orthogonal to all vectors in $\text{R}(A)^\perp = \{\text{span of columns of } A\}^\perp$ (Equivalently, we can state that $A\mathbf{X} = \mathbf{b}$ has a solution if and only if $\mathbf{b}$ is orthogonal to any basis that spans $\text{R}(A)^\perp$). A basis that spans $\text{R}(A)^\perp$ can be obtained by considering the adjoint of $A$. But first, we must consider the concept of a linear functional.

### Linear Functional

Let $U$ denote a vector space over a scalar field $\mathcal{F}$. A mapping

$$\phi : U \rightarrow \mathcal{F}$$

is termed a linear functional if

$$\phi(\alpha \mathbf{X}_1 + \beta \mathbf{X}_2) = \alpha \phi(\mathbf{X}_1) + \beta \phi(\mathbf{X}_2)$$

for all $\mathbf{X}_1, \mathbf{X}_2 \in U$ and every $\alpha, \beta \in \mathcal{F}$. That is, a linear functional is a linear mapping from $U$ into the scalar field $\mathcal{F}$.

### Example

For arbitrary but fixed $\mathbf{X} \in U$, we can define the functional $\phi_\mathbf{X} : U \rightarrow \mathcal{F}$ as

$$\phi_\mathbf{X}(\mathbf{X}_1) = \langle \mathbf{X}_1, \mathbf{X} \rangle, \quad \mathbf{X}_1 \in U.$$  \hfill (5-11)

Clearly, as defined by (5-11), $\phi_\mathbf{X}$ is linear:
\[ \phi_X (\alpha \tilde{X}_1 + \beta \tilde{X}_2) = \langle \alpha \tilde{X}_1 + \beta \tilde{X}_2, \tilde{X} \rangle = \alpha \langle \tilde{X}_1, \tilde{X} \rangle + \beta \langle \tilde{X}_2, \tilde{X} \rangle 
= \alpha \phi_X (\tilde{X}_1) + \beta \phi_X (\tilde{X}_2) \] 

for all \( \tilde{X}_1, \tilde{X}_2 \) in \( U \) and \( \alpha, \beta \) in \( \mathcal{F} \).♥

This last example shows how elementary linear functionals can be. However, according to Theorem 5.3, all linear functionals mapping \( U \to \mathcal{F} \) must be representable in the form given by (5-11).

**Theorem 5.3**

Let \( \phi : U \to \mathcal{F} \) be a linear functional on \( n \)-dimensional vector space \( U \). For every such \( \phi \), there exists a unique \( \tilde{X}_0 \in U \) (the vector \( \tilde{X}_0 \) depends on the particular \( \phi \)) for which

\[ \phi(\tilde{X}) = \langle \tilde{X}, \tilde{X}_0 \rangle \] 

(5-13)

for all \( \tilde{X} \in U \). That is, each linear functional can be represented as an inner product with a functional-dependent vector \( \tilde{X}_0 \).

**Proof:**

First we show the existence of an \( \tilde{X}_0 \) with the stated property. Let \( \tilde{\alpha}_1, \tilde{\alpha}_2, \ldots, \tilde{\alpha}_n \) be an orthonormal basis of space \( U \). Define

\[ \tilde{X}_0 = \bar{\phi}(\tilde{\alpha}_1)\tilde{\alpha}_1 + \bar{\phi}(\tilde{\alpha}_2)\tilde{\alpha}_2 + \cdots + \bar{\phi}(\tilde{\alpha}_n)\tilde{\alpha}_n \] 

(5-14)

(\( \bar{\phi} \) denotes the complex conjugate of \( \phi \)). Then for each \( i, 1 \leq i \leq n \), we have

\[ \langle \tilde{\alpha}_i, \tilde{X}_0 \rangle = \langle \tilde{\alpha}_i, \bar{\phi}(\tilde{\alpha}_1)\tilde{\alpha}_1 + \bar{\phi}(\tilde{\alpha}_2)\tilde{\alpha}_2 + \cdots + \bar{\phi}(\tilde{\alpha}_n)\tilde{\alpha}_n \rangle = \phi(\tilde{\alpha}_i) \] 

(5-15)
That is, $\phi(\bar{X})$ agrees with $\langle \bar{X}, \bar{X}_0 \rangle$ on a basis of space $U$. Now, any vector can be represented in terms of basis $\bar{\alpha}_1, \bar{\alpha}_2, \ldots, \bar{\alpha}_n$. Hence, the equality

$$\phi(\bar{X}) = \langle \bar{X}, \bar{X}_0 \rangle$$

must hold on the whole space $U$, and we have proved the existence of an $\bar{X}_0$ with the desired property.

Next, we show that, given $\phi$, the vector $\bar{X}_0$ is unique. We do this by arriving at a contradiction after assuming that there are two vectors with the stated property. Suppose that $\bar{X}_0$ and $\bar{X}_1$ exist such that

$$\phi(\bar{X}) = \langle \bar{X}, \bar{X}_0 \rangle = \langle \bar{X}, \bar{X}_1 \rangle$$

(5-16)

for all $\bar{X} \in U$. This would require that $\langle \bar{X}, \bar{X}_0 - \bar{X}_1 \rangle = 0$ for all $\bar{X} \in U$, and this means that $\bar{X}_0$ and $\bar{X}_1$ must be the same vector. Hence, uniqueness of $\bar{X}_0$ has been established, and this completes the proof of Theorem 5.3.

**Adjoint Operator**

This subsection discusses the *adjoint operator*, an operator that plays a major role in linear analysis, both in finite and infinite-dimensional spaces.

**Theorem 5.4**

Let $T : U \rightarrow V$ be a linear operator. Then there exists a unique linear operator $T^* : V \rightarrow U$ such that

$$\langle T(\bar{X}), \bar{Y} \rangle = \langle \bar{X}, T^*(\bar{Y}) \rangle$$

(5-17)
for every $\bar{X} \in U$ and $\bar{Y} \in V$. Operator $T^*$ is called the *adjoint* of $T$.

**Proof:**

Let $\bar{Y} \in V$ be arbitrary but fixed; we generate a $T^*$ that maps $\bar{Y}$ back to $U$ such that (5-17) holds. For all $\bar{X} \in U$, we define

$$\phi(\bar{X}) \equiv \langle T(\bar{X}), \bar{Y} \rangle,$$

(5-18)

a linear functional that maps $U$ into $\mathcal{F}$. However, by Theorem 5.3 there exists a unique vector $\bar{X}_Y \in U$ with the property that

$$\phi(\bar{X}) \equiv \langle T(\bar{X}), \bar{Y} \rangle = \langle \bar{X}, \bar{X}_Y \rangle,$$

(5-19)

for all $\bar{X} \in U$. Define $T^* : V \rightarrow U$ by $T^*(\bar{Y}) = \bar{X}_Y$. Then we have

$$\langle T(\bar{X}), \bar{Y} \rangle = \langle \bar{X}, T^*(\bar{Y}) \rangle$$

as claimed. Next, we show that $T^*$ is linear. For any $\bar{Y}_1 \in V$ and $\bar{Y}_2 \in V$, and scalars $\alpha, \beta$ in $\mathcal{F}$, we have

$$\langle \bar{X}, T^*(\alpha \bar{Y}_1 + \beta \bar{Y}_2) \rangle = \langle T(\bar{X}), \alpha \bar{Y}_1 + \beta \bar{Y}_2 \rangle$$

$$= \bar{\alpha} \langle T(\bar{X}), \bar{Y}_1 \rangle + \bar{\beta} \langle T(\bar{X}), \bar{Y}_2 \rangle \quad \text{(remember: $\langle \cdot, \cdot \rangle$ is conjugate linear in its second entry!)}$$

(5-20)

$$= \bar{\alpha} \langle \bar{X}, T^*(\bar{Y}_1) \rangle + \bar{\beta} \langle \bar{X}, T^*(\bar{Y}_2) \rangle$$

$$= \langle \bar{X}, T^*(\alpha \bar{Y}_1) + T^*(\beta \bar{Y}_2) \rangle$$
for all $\vec{X} \in U$. Hence, we have

$$T^*(\alpha \vec{Y}_1 + \beta \vec{Y}_2) = \alpha T^*(\vec{Y}_1) + \beta T^*(\vec{Y}_2),$$

(5-21)

and we conclude that $T^*$ is linear.$\heartsuit$

**Matrix Representation of Adjoint**

With respect to given bases on spaces $U$ and $V$, an $m \times n$ matrix $A$ can be found that represents the linear transformation $T : U \to V$. Likewise, with respect to these same bases, a matrix representation for the adjoint transformation $T^* : V \to U$ can be found. The question begs to be asked: is there a *simple* relationship between the matrices that represent $T$ and $T^*$? As shown by the following theorem, the answer is a qualified yes.

**Theorem 5.5**

Let $T : U \to V$ be a linear operator, and let $T^* : V \to U$ be its adjoint. With respect to an *orthonormal basis* $\vec{\alpha}_1, \ldots, \vec{\alpha}_n$ for $U$ and an *orthonormal basis* $\vec{\beta}_1, \ldots, \vec{\beta}_m$ for $V$, let $m \times n$ matrix $A$ represent $T$. Then the conjugate transpose $A^*$ is the matrix representing $T^*$ with respect to the bases $\vec{\alpha}_1, \ldots, \vec{\alpha}_n$ for $U$ and $\vec{\beta}_1, \ldots, \vec{\beta}_m$ for $V$ (interchange the rows and columns, and take the complex conjugate of the elements, of $A$ to obtain $A^*$). Note that $A^*$ is an $n \times m$ matrix.

**Proof:** With respect to the given orthonormal bases, let $n \times m$ matrix $B$ represent $T^*$. We show that $B = A^*$. Notationally, let $\{a_{ij}\} = A$ and $\{b_{ij}\} = B$. Then by Theorem 3.2, we have

$$a_{ij} = \langle T(\vec{\alpha}_j), \vec{\beta}_i \rangle \quad b_{ij} = \langle T^*(\vec{\beta}_j), \vec{\alpha}_i \rangle.$$

(5-22)

However,

$$b_{ij} = \langle T^*(\vec{\beta}_j), \vec{\alpha}_i \rangle = \langle \overline{\vec{\alpha}_i}, T^*(\vec{\beta}_j) \rangle = \overline{\langle T(\vec{\alpha}_i), \vec{\beta}_j \rangle} = \overline{a_{ji}},$$

(5-23)
where over bars denote complex conjugate. Note from (5-23) that \( B = A^* \) as claimed.

**Warning:** Depending on what book you look in, there are two "adjoints", and the first "adjoint" has absolutely nothing to do with the second "adjoint"! First, there is an "adjoint" just as we have defined it in this chapter. Second, in the literature, the transpose of the cofactor matrix is called the "adjoint" (we do not discuss this "adjoint" here). Beware! Do not confuse one usage with the other!

---

**A Useful Dichotomy of Vector Spaces \( U \) and \( V \)**

Both the domain \( U \) and co-domain \( V \) can be partitioned in terms of the range and kernel of \( T \) and \( T^* \). The next three theorems establish this partition; the results are given in terms of the range and kernel of matrices \( A \) and \( A^* \) used to represent \( T \) and \( T^* \), respectively. As always \( R(A) \) and \( K(A) \) denote the range and kernel, respectively, of matrix \( A \).

**Theorem 5.6**

The subspace \( K(A^*) \) is orthogonal to the subspace \( R(A) \).

**Proof:**

Let \( \bar{Y} \in K(A^*) \). Then \( A^* \bar{Y} = \bar{0} \). Also,

\[
\langle A\bar{X}, \bar{Y} \rangle = \langle \bar{X}, A^* \bar{Y} \rangle = \langle \bar{X}, \bar{0} \rangle = 0
\]

for all \( \bar{X} \in U \). But \( R(A) = [A\bar{X} : \bar{X} \in U] \) so \( A\bar{X} \in R(A) \). By (5-24), we conclude that \( K(A^*) \) is orthogonal to \( R(A) \) as claimed.

**Theorem 5.7**

\( R(A) \) is the orthogonal complement of \( K(A^*) \).

**Proof:** As shown by Theorem 5.6, \( K(A^*) \) is orthogonal to \( R(A) \) so \( K(A^*) \subset R(A)^\perp \). Let \( r \) denote
the rank of matrix $A$. Then $r = \text{rank}[A] = \text{rank}[A^*]$ since the number of linear independent columns is equal to the number of independent rows. However, $\text{rank}[A^*] + \text{nullity}[A^*] = m$. So $\text{nullity}[A^*] = m - r$. So $\dim(K(A^*)) = \dim(R(A)^\perp)$; this coupled with $K(A^*) \subseteq R(A)^\perp$ leads to the conclusion that $K(A^*) = R(A)^\perp$, and $K(A^*)$ is the orthogonal complement of $R(A)$ as claimed.

**Theorem 5.8**

$R(A^*)$ is the orthogonal complement of $K(A)$.

**Proof:** Apply Theorem 5.7 to $A^*$ and realize that $A^{**} = A$.

Theorems 5.7 and 5.8 state useful partitions of spaces $U$ and $V$. $R(A^*)$ is the orthogonal complement of $K(A^*)$. 

U can be split into $R(A^*)$ and $K(A)$, pieces that are orthogonal to each other.

$V$ can be split into $R(A)$ and $K(A^*)$, pieces that are orthogonal to each other.

**Figure 5.1:** $R(A^*)$ and $K(A)$ are orthogonal complements. $R(A)$ and $K(A^*)$ are orthogonal complements.
complement of $K(A)$, and $U = R(A^*) \oplus K(A)$. Also, $R(A)$ is the orthogonal complement of $K(A^*)$, and $V = R(A) \oplus K(A^*)$. These very important partitions are illustrated by Figure 5-1.

Example

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad A^* = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \\ 3 & 1 & 1 \end{bmatrix}$$

$A : U = \mathbb{R}^3 \rightarrow V = \mathbb{R}^4$ while $A^* : V = \mathbb{R}^4 \rightarrow U = \mathbb{R}^3$.

$\text{Rank}[A] = 2$, $\text{Nullity}[A] = 1$, $\text{Rank}[A] + \text{Nullity}[A] = \text{dim}[ U ] = 3$

$\text{Rank}[A^*]=2$, $\text{Nullity}[A^*] = 2$, $\text{Rank}[A^*] + \text{Nullity}[A^*] = \text{dim}[V] = 4$

$R(A) = \text{span}\{ [1 \ 0 \ 1 \ 0]^T, \ [2 \ 1 \ 0 \ 1]^T \}$, $K(A) = \text{span}\{ [1 \ 1 \ -1]^T \}$

$R(A^*) = \text{span}\{ [0 \ 1 \ 1]^T, \ [1 \ 0 \ 1]^T \}$, $K(A^*) = \text{span}\{ [-1 \ 1 \ 1 \ 1]^T, \ [0 \ -1 \ 0 \ 1]^T \}$

$R(A) \perp K(A^*)$ and $R(A) \oplus K(A^*) = V = \mathbb{R}^4$

$R(A^*) \perp K(A)$ and $R(A^*) \oplus K(A) = U = \mathbb{R}^3$. ♥

Theorem 5.2 provides necessary and sufficient conditions for the existence of a solution(s) to the linear algebraic equation $A\vec{x} = \vec{b}$. By using Theorems 5.7 and 5.8, an alternative necessary and sufficient condition can be given for the existence question. It is called Fredholm Alternative, and it plays a very important role in the theory of finite dimensional space (like we study in this course). More general versions of the Fredholm Alternative apply in infinite dimensional spaces where they are used to establish the existence of solution(s) to linear functional equations.

**Theorem 5.9 (Fredholm Alternative)**

The linear algebraic equation $A\vec{x} = \vec{b}$ has a solution if and only if $\vec{b} \in K(A^*)^\perp$.

**Proof:**

From Theorem 5.2, we know that $A\vec{x} = \vec{b}$ has a solution if and only if $\vec{b} \in R(A)$. This theorem follows from the realization that $R(A) = K(A^*)^\perp$. ♥