2

Linear Algebra and Preliminaries

In this chapter, we review some basic results in numerical linear algebra, which are repeatedly used in later chapters. Among others, the QR decomposition and the singular value decomposition (SVD) are the most valuable tools in the areas of signal processing and system identification.

2.1 Vectors and Matrices

Let $\mathbb{R}$ be the set of real numbers, $\mathbb{R}^n$ the set of n-dimensional real vectors, and $\mathbb{R}^{m\times n}$ the set of $m \times n$ real matrices. The lower case letters $x,\, y,\, \cdots$ denote vectors, and capital letters $A,\, B,\, C,\, \cdots\, ;\, X,\, Y,\, Z,\, \cdots$ denote matrices. Transpositions of a vector x and a matrix A are denoted by x^{T} and A^{T}, respectively. The determinant of a square matrix A is denoted by $|A|$, or $\det(A)$, and the trace by $\mathrm{trace}(A)$.

The $n \times n$ identity matrix is denoted by I_n. If there is no confusion, we simply write I, deleting the subscript denoting the dimension. The inverse of a square matrix A is denoted by A^{-1}. We also use $A^{-\mathrm{T}}$ to denote $(A^{-1})^{\mathrm{T}} = (A^{\mathrm{T}})^{-1}$. A matrix satisfying $A^{\mathrm{T}} = A$ is called a symmetric matrix. If a matrix $A \in \mathbb{R}^{m\times n}$ with $m \geq n$ satisfies $A^{\mathrm{T}}A = I_n$, it is called an orthogonal matrix. Thus, for an orthogonal matrix $A = [a_1\, a_2\, \cdots\, a_n],\ a_i \in \mathbb{R}^m,\ i = 1,\, \cdots\, n$, we have $a_i^{\mathrm{T}} a_j = \delta_{ij}, i,\ j = 1,\, \cdots,\, n$, where δ_{ij} is the Kronecker delta defined by

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

For vectors $x,\ y \in \mathbb{R}^n$, the inner product is defined by

$$(x,\, y) = x^{\mathrm{T}} y = \sum_{i=1}^{n} x_i y_i = y^{\mathrm{T}} x$$

Also, for $A \in \mathbb{R}^{n\times n}$ and $x \in \mathbb{R}^n$, we define the quadratic form

$$x^{\mathrm{T}}Ax = (x,\ Ax) = \sum_{i,j=1}^{n} a_{ij}x_i x_j \tag{2.1}$$

Define $\bar{A} = (A + A^{\mathrm{T}})/2$. Then, we have $x^{\mathrm{T}}\bar{A}x = x^{\mathrm{T}}Ax$. Thus it is assumed without loss of generality that A is symmetric in defining a quadratic form.

If $x^{\mathrm{T}}Ax > 0,\ x \neq 0$, then A is positive definite, and is written as $A > 0$. If $x^{\mathrm{T}}Ax \geq 0$ holds, A is called nonnegative definite, and is written as $A \geq 0$. Moreover, if $A - B > 0$ (or ≥ 0) holds, then we simply write $A > B$ (or $A \geq B$).

The basic facts for real vectors and matrices mentioned above can carry over to complex vectors and matrices. Let $\mathbb{C}$ be the set of complex numbers, $\mathbb{C}^n$ the set of n-dimensional complex vectors, and $\mathbb{C}^{m\times n}$ the set of $m \times n$ complex matrices. The complex conjugate of $\lambda \in \mathbb{C}$ is denoted by $\bar{\lambda}$, and similarly the complex conjugate transpose of $A = (a_{ij}) \in \mathbb{C}^{m\times n}$ is denoted by $A^{\mathrm{H}} = (\bar{a}_{ji})$. We say that $A \in \mathbb{C}^{n\times n}$ is Hermitian if $A^{\mathrm{H}} = A$, and unitary if $A^{\mathrm{H}}A = I_n$.

The inner product of $x,\ y \in \mathbb{C}^n$ is defined by

$$x^{\mathrm{H}}y = \sum_{i=1}^{n} \bar{x}_i y_i = \overline{y^{\mathrm{H}}x}$$

As in the real case, the quadratic form $x^{\mathrm{H}}Ax,\ x \in \mathbb{C}^n$ is defined for a Hermitian matrix A. We say that A is positive definite if $x^{\mathrm{H}}Ax > 0, x \neq 0$, and nonnegative definite if $x^{\mathrm{H}}Ax \geq 0$; being positive (nonnegative) definite is written as $A > 0$ ($A \geq 0$).

The characteristic polynomial for $A \in \mathbb{R}^{n\times n}$ is defined by

$$\varphi_A(z) := \det(zI - A) = z^n + \alpha_1 z^{n-1} + \cdots + \alpha_{n-1}z + \alpha_n \tag{2.2}$$

The n roots of $\varphi_A(z) = 0$ are called the eigenvalues of A. The set of eigenvalues of A, denoted by $\lambda(A)$, is called the spectrum of A. The ith eigenvalue is described by $\lambda_i(A)$. Since $\varphi_A(z)$ has real coefficients, if $\lambda \in \mathbb{C}$ is an eigenvalue, so is $\bar{\lambda} \in \mathbb{C}$. If $\lambda \in \lambda(A)$, there exists a vector $v \in \mathbb{C}^n$ satisfying

$$Av = \lambda v, \qquad v \neq 0$$

In this case, $v \in \mathbb{C}^n$ is called an eigenvector corresponding to the eigenvalue λ. It may be noted that, since the eigenvalues are complex, the corresponding eigenvectors are also complex.

Let the characteristic polynomial of $A \in \mathbb{R}^{n\times n}$ be given by (2.2). Then the following matrix polynomial equation holds:

$$\varphi_A(A) := A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1}A + \alpha_n I = 0 \tag{2.3}$$

where the right-hand side is the zero matrix of size $n \times n$. This result is known as the Cayley-Hamilton theorem.

We see that the eigenvalues $\lambda_i,\ i = 1, \cdots, n$ of a symmetric nonnegative definite matrix $A \in \mathbb{R}^{n\times n}$ are nonnegative. Thus by means of an orthogonal matrix T, we can transform A into a diagonal form, *i.e.*,

$$T^{-1}AT = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} = \mathrm{diag}(\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n)$$

Define $\mu_i = \sqrt{\lambda_i}$, $i = 1,\ \cdots,\ n$. Then we have

$$A = T \begin{bmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_n \end{bmatrix} \begin{bmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_n \end{bmatrix} T^{-1}$$

Also, let B be given by

$$B = \begin{bmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_n \end{bmatrix} T^{-1}$$

Then it follows that $A = B^{\mathrm{T}}B$, so that B is called a square root matrix of A, and is written as $\sqrt{A}$ or $A^{1/2}$. For any orthogonal matrix Q, we see that $B_1 = QB$ satisfies $A = B_1^{\mathrm{T}}B_1$ so that B_1 is also a square root matrix of A, showing that a square root matrix is not unique.

Suppose that $A = (a_{ij}) \in \mathbb{R}^{m\times n}$. Then, $A(p : q, r : s)$ denotes the submatrix of A formed by p, $p + 1, \cdots$, q rows and r, $r + 1, \cdots$, s columns, *e.g.*,

$$A(2 : 4; 3 : 6) = \begin{bmatrix} a_{23} & a_{24} & a_{25} & a_{26} \\ a_{33} & a_{34} & a_{35} & a_{36} \\ a_{43} & a_{44} & a_{45} & a_{46} \end{bmatrix}$$

In particular, $A(p : q, :)$ means the submatrix formed by p, $p + 1, \cdots$, q rows, and similarly $A(:, r : s)$ the submatrix formed by r, $r + 1, \cdots$, s columns. Also, $A(i, :)$ and $A(:, j)$ respectively represent the ith row and jth column of A.

2.2 Subspaces and Linear Independence

In the following, we assume that scalars are real; but all the results are extended to complex scalars.

Consider a set $\mathcal{V}$ which is not void. For $x, y \in \mathcal{V}$, and for a scalar $\alpha \in \mathbb{R}$, the sum $x + y$ and product αx are defined. Suppose that the set $\mathcal{V}$ satisfies the axiom of "linear space" with respect to the addition and product defined above. Then $\mathcal{V}$ is called a linear space over $\mathbb{R}$. The set of n-dimensional vectors $\mathbb{R}^n$ and $\mathbb{C}^n$ are linear spaces over $\mathbb{R}$ and $\mathbb{C}$, respectively.

Suppose that $\mathcal{W}$ be a subset of a linear space $\mathcal{V}$. For any $w_1,\ w_2 \in \mathcal{W}$ and α_1, $\alpha_2 \in \mathbb{R}$, if $\alpha_1 w_1 + \alpha_2 w_2 \in \mathcal{W}$ holds, then $\mathcal{W}$ is called a subspace of $\mathcal{V}$, and this fact is simply expressed as $\mathcal{W} \subset \mathcal{V}$.

For a set of vectors $\{x_1,\ \cdots,\ x_n\}$ in $\mathbb{R}^m$, if there exist scalars $\alpha_1,\ \cdots,\ \alpha_n$ with $\alpha_i \neq 0$ for at least an i such that

$$\sum_{j=1}^{n} \alpha_j x_j = 0$$

holds, then $\{x_1,\ \cdots,\ x_n\}$ are called linearly dependent. Conversely, if we have

$$\sum_{j=1}^{n} \alpha_j x_j = 0 \quad \Rightarrow \quad \alpha_1 = \cdots = \alpha_n = 0$$

then $\{x_1,\ \cdots,\ x_n\}$ are called linearly independent.

All the linear combinations of vectors $\{w_1,\ \cdots,\ w_p\}$ in $\mathbb{R}^m$ form a subspace of $\mathbb{R}^n$, which is written as

$$\mathcal{W} = \mathrm{span}\{w_1,\ \cdots,\ w_p\} = \left\{ \sum_{j=1}^{p} \alpha_j w_j \;\middle|\; \alpha_1,\ \cdots,\ \alpha_p \in \mathbb{R} \right\}$$

If $\{w_1,\ \cdots,\ w_p\}$ are linearly independent, they are called a basis of the space $\mathcal{W}$.

Suppose that $\mathcal{V}$ is a subspace of $\mathbb{R}^m$. Then there exists a basis $\{v_1,\ \cdots,\ v_d\}$ in $\mathcal{V}$ such that

$$\mathcal{V} = \mathrm{span}\{v_1,\ \cdots,\ v_d\}$$

Hence, any $x \in \mathcal{V}$ can be expressed as a linear combination of the form

$$x = \sum_{j=1}^{d} \beta_j v_j, \qquad \beta_1,\ \cdots,\ \beta_d \in \mathbb{R}$$

where $\beta_1,\ \cdots,\ \beta_d$ are components of x with respect to the basis $\{v_1,\ \cdots,\ v_d\}$. Choice of basis is not unique, but the number of the elements of any basis is unique. The number is called the dimension of $\mathcal{V}$, which is denoted by $\dim(\mathcal{V})$.

For a matrix $A \in \mathbb{R}^{m \times n}$, the image of A is defined by

$$\mathrm{Im}(A) = \{y \in \mathbb{R}^m \mid y = Ax,\ x \in \mathbb{R}^n\} = A\mathbb{R}^n$$

This is a subspace of $\mathbb{R}^m$, and is also called the range of A. If $A = [a_1\ \cdots\ a_n]$, then we have $\mathrm{Im}(A) = \mathrm{span}\{a_1,\ \cdots,\ a_n\}$. Moreover, the set of vectors mapped to zero are called the kernel of A, which is written as

$$\mathrm{Ker}(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$$

This is also called the null space of A, a subspace of $\mathbb{R}^n$.

The rank of $A \in \mathbb{R}^{m\times n}$ is defined by $\dim(\mathrm{Im}\ A)$ and is expressed as $\mathrm{rank}(A)$. We see that $\mathrm{rank}(A) = r$ if and only if the maximum number of independent vectors among the column vectors $a_1,\ \cdots,\ a_n$ of A is r. This is also equal to the number of independent vectors in row vectors $\tilde{a}_1^{\mathrm{T}},\ \cdots,\ \tilde{a}_m^{\mathrm{T}}$ of A. Thus it follows that $\mathrm{rank}(A) = \mathrm{rank}(A^{\mathrm{T}})$.

It can be shown that for $A \in \mathbb{R}^{m\times n}$,

$$\dim(\mathrm{Im}\ A) + \dim(\mathrm{Ker}\ A) = n \tag{2.4}$$

Hence, if $m = n$ holds, the following are equivalent:

(i) A : nonsingular (ii) $\mathrm{Ker}(A) = \{0\}$ (iii) $\mathrm{rank}(A) = n$

Suppose that $x,\ y \in \mathbb{R}^n$. If $x^{\mathrm{T}}y = 0$, or if the vectors are mutually orthogonal, we write $x \perp y$. If $y^{\mathrm{T}}x = 0$ holds for all $x \in \mathcal{V} \subset \mathbb{R}^n$, we say that y is orthogonal to $\mathcal{V}$, which is written as $y \perp \mathcal{V}$. The set of $y \in \mathbb{R}^n$ satisfying $y \perp \mathcal{V}$ is called the orthogonal complement, which is expressed as

$$\mathcal{V}^{\perp} = \{y \in \mathbb{R}^n \mid y^{\mathrm{T}}x = 0,\ \forall\, x \in \mathcal{V}\}$$

The orthogonal complement $\mathcal{V}^{\perp}$ is a subspace whether or not $\mathcal{V}$ is a subspace.

Let $\mathcal{V},\ \mathcal{W} \subset \mathbb{R}^n$ be two subspaces. If $v^{\mathrm{T}}w = 0$ holds for any $v \in \mathcal{V}$ and $w \in \mathcal{W}$, then we say that $\mathcal{V}$ and $\mathcal{W}$ are orthogonal, so that we write $\mathcal{V} \perp \mathcal{W}$. Also, the vector sum of $\mathcal{V}$ and $\mathcal{W}$ is defined by

$$\mathcal{V} \vee \mathcal{W} = \{v + w \mid\ v \in \mathcal{V},\ w \in \mathcal{W}\}$$

It may be noted that this is not the union $\mathcal{V} \cup \mathcal{W}$ of the two subspaces. Moreover, if $\mathcal{V} \cap \mathcal{W} = \{0\}$ holds, the vector sum is called the direct sum, and is written as $\mathcal{V} + \mathcal{W}$. Also, if $\mathcal{V} \perp \mathcal{W}$ holds, then it is called the direct orthogonal sum, and is expressed as $\mathcal{V} \oplus \mathcal{W}$.

For a subspace $\mathcal{V} \subset \mathbb{R}^n$, we have a unique decomposition

$$\mathbb{R}^n = \mathcal{V} \oplus \mathcal{V}^{\perp} \tag{2.5}$$

This implies that $x \in \mathbb{R}^n$ has a unique decomposition $x = v + w$, $v \in \mathcal{V},\ w \in \mathcal{V}^{\perp}$.

Let $\mathcal{V} \subset \mathbb{R}^n$ be a subspace, and $A \in \mathbb{R}^{n\times n}$ a linear transform. Then, if

$$x \in \mathcal{V}\ \Rightarrow\ Ax \in \mathcal{V} \quad (A\mathcal{V} \subset \mathcal{V})$$

holds, $\mathcal{V}$ is called an A-invariant subspace. The spaces spanned by eigenvectors, and $\mathrm{Im}(A)$, $\mathrm{Ker}(A)$ are all important A-invariant subspaces of $\mathbb{R}^n$.

2.3 Norms of Vectors and Matrices

Definition 2.1. *A vector norm* $(\|\cdot\|)$ *has the following properties.*

(i) $||x|| \geq 0; \quad ||x|| = 0 \;\Leftrightarrow\; x = 0$
(ii) $||\lambda x|| = |\lambda| \, ||x||, \quad \lambda$ *: scalar*
(iii) $||x + y|| \leq ||x|| + ||y||$ *(triangular inequality)* □

For a vector $x = (x_1, \cdots, x_n)^{\mathrm{T}} \in \mathbb{R}^n$, the 2-norm (or Euclidean norm) is defined by

$$||x||_2 = \left(|x_1|^2 + \cdots + |x_n|^2\right)^{1/2}$$

and the infinity-norm is defined by

$$||x||_\infty = \max(|x_1|, \cdots, |x_n|)$$

Since $A \in \mathbb{R}^{m\times n}$ can be viewed as a vector in $\mathbb{R}^{mn}$, the definition of a matrix norm should be compatible with that of the vector norm. The most popular matrix norms are the Frobenius norm and the 2-norm. The former norm is given by

$$||A||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2} = \sqrt{\mathrm{trace}(A^{\mathrm{T}}A)} \tag{2.6}$$

The latter is called an operator norm, which is defined by

$$||A||_2 = \sup_{x\neq 0} \frac{||Ax||_2}{||x||_2} \tag{2.7}$$

We have the following inequalities for the above two norms:

$$||Ax||_2 \leq ||A||_2 \, ||x||_2, \quad ||AB||_\alpha \leq ||A||_\alpha \, ||B||_\alpha, \quad \alpha = 2, \ F$$

If Q is orthogonal, *i.e.*, $Q^{\mathrm{T}}Q = I$, we have $||Qx||_2^2 = x^{\mathrm{T}}Q^{\mathrm{T}}Qx = ||x||_2^2$. Moreover, it follows that $||QA||_\alpha = ||A||_\alpha$ for $\alpha = 2, F$. Thus we see that the 2-norm and Frobenius norm are invariant under orthogonal transforms. We often write the 2-norm of x as $||x||$, suppressing the subscript.

For a complex vector $x \in \mathbb{C}^n$, and a complex matrix $A \in \mathbb{C}^{m\times n}$, their norms are defined similarly to the real cases.

Lemma 2.1. *For $A \in \mathbb{R}^{n\times n}$, the spectral radius is defined by*

$$\rho(A) = \max\{|\lambda_i(A)| \mid i = 1, \cdots, n\} \tag{2.8}$$

Then, $\rho(A) \leq ||A||_\alpha$ holds.
Proof. Clearly, there exists an eigenvalue λ for which $|\lambda| = \rho(A)$. Let $Ax = \lambda x$, $x \neq 0$. Let $X := [x \;\; x \;\; \cdots \;\; x] \in \mathbb{C}^{n\times n}$, and consider $AX = \lambda X$. Then, for any matrix norm $||\cdot||_\alpha$, we have

$$|\lambda| \cdot ||X||_\alpha = ||\lambda X||_\alpha = ||AX||_\alpha \leq ||A||_\alpha \cdot ||X||_\alpha, \qquad ||X||_\alpha \neq 0$$

and hence $|\lambda| = \rho(A) \leq ||A||_\alpha$. □

More precisely, the above result holds for many matrix norms [73].

2.4 QR Decomposition

In order to consider the QR decomposition, we shall introduce an orthogonal transform, called the Householder transform (see Figure 2.1).

Lemma 2.2. *Consider two vectors $x \neq y \in \mathbb{R}^n$ with $||x|| = ||y||$. Then there exists a vector $u \in \mathbb{R}^n$ such that*

$$(I - 2uu^{\mathrm{T}})x = y, \qquad ||u|| = 1 \tag{2.9}$$

The vector u is defined uniquely up to signature by

$$u = \pm \frac{x - y}{||x - y||} \tag{2.10}$$

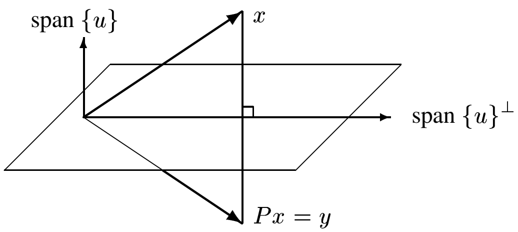

Figure 2.1. Householder transform

Proof. By using (2.10) and the fact that $||x|| = ||y||$ and $x^{\mathrm{T}}y = y^{\mathrm{T}}x$, we compute the left-hand side of (2.9) to get

$$\begin{aligned}(I - 2uu^{\mathrm{T}})x &= x - \frac{2(x-y)(x-y)^{\mathrm{T}}}{(x-y)^{\mathrm{T}}(x-y)}x = x - \frac{2(x-y)(x^{\mathrm{T}}x - y^{\mathrm{T}}x)}{x^{\mathrm{T}}x - x^{\mathrm{T}}y - y^{\mathrm{T}}x + y^{\mathrm{T}}y} \\ &= x - \frac{2(x-y)(x^{\mathrm{T}}x - y^{\mathrm{T}}x)}{2(x^{\mathrm{T}}x - y^{\mathrm{T}}x)} = y\end{aligned}$$

Suppose now that a vector $v \in \mathbb{R}^n$ also satisfies the condition of this lemma. Then we have $||v|| = 1$, and hence

$$y = (I - 2uu^{\mathrm{T}})x = (I - 2vv^{\mathrm{T}})x \;\; \Rightarrow \;\; u(u^{\mathrm{T}}x) = v(v^{\mathrm{T}}x), \quad \forall\, x$$

Putting $x = u$ (or $x = v$) yields $v^{\mathrm{T}}u = \pm 1$. Thus it follows that $v = \pm u$, showing the uniqueness of the vector u up to signature. □

The matrix $P := I - 2uu^{\mathrm{T}}$ of Lemma 2.2, called the Householder transform, is symmetric and satisfies

$$P^2 = (I - 2uu^{\mathrm{T}})^2 = I - 4uu^{\mathrm{T}} + 4u(u^{\mathrm{T}}u)u^{\mathrm{T}} = I$$

Thus it follows that $P^{-1} = P = P^{\mathrm{T}}$, implying that P is an orthogonal transform, and that $Px = y$, $Py = x$ hold.

Let $a,\ b \in \mathbb{R}^n$ with $\|a\| = \|b\|$. We consider a problem of transforming the vector a into the vector b, of which the first element is nonzero, but the other elements are all zeros. More precisely, we wish to find a transform that performs the following reduction

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \longrightarrow b = \begin{bmatrix} b_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \|a\| = \|b\| = |b_1|$$

Since $\|a\| = \|b\|$, we see that $b_1 = \pm\|a\|$. It follows that

$$\tilde{a} := a - b = \begin{bmatrix} a_1 - b_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \qquad \|\tilde{a}\|^2 = 2b_1(b_1 - a_1) = -2b^{\mathrm{T}}\tilde{a}$$

It should be noted that if the sign of b_1 is chosen as the same as that of a_1, there is a possibility that $|a_1 - b_1|$ has a small value, so that a large relative error may arise in the term $u^{\mathrm{T}}u = \tilde{a}^{\mathrm{T}}\tilde{a}$. This difficulty can be simply avoided by choosing the sign of b_1 opposite to that of a_1.

We now define

$$P := I - \frac{2\tilde{a}\tilde{a}^{\mathrm{T}}}{\|\tilde{a}\|^2} = I + \frac{\tilde{a}\tilde{a}^{\mathrm{T}}}{b^{\mathrm{T}}\tilde{a}} = I + \frac{\tilde{a}\tilde{a}^{\mathrm{T}}}{b_1\tilde{a}_1} \tag{2.11}$$

Noting that $a^{\mathrm{T}}a = b_1^2$ and $b^{\mathrm{T}}a = b_1 a_1$, we have

$$Pa = \left[I + \frac{(a-b)(a-b)^{\mathrm{T}}}{b_1\tilde{a}_1}\right] a = a - \frac{(a-b)(a^{\mathrm{T}}a - b^{\mathrm{T}}a)}{b_1(b_1 - a_1)} = b$$

Hence, by knowing $\tilde{a} = a - b$ and b_1, the vector a can be transformed into the vector b with the specified form. It is well known that this method is very efficient and reliable, since only the first component of a is modified in the computation. In the following, $\tilde{a}$ plays the role of the vector u in the Householder transform, though the norm of $\tilde{a}$ is not unity.

Now we introduce the QR decomposition, which is quite useful in numerical linear algebra. We assume that matrices are real, though the QR decomposition is applicable to complex matrices.

Lemma 2.3. *A tall rectangular matrix $A \in \mathbb{R}^{m\times n}$, $m \geq n$ is decomposed into a product of two matrices:*

$$A = QR \tag{2.12}$$

where $Q \in \mathbb{R}^{m\times n}$ is an orthogonal matrix with $Q^{\mathrm{T}}Q = I_n$, and $R \in \mathbb{R}^{n\times n}$ is an upper triangular matrix. The right-hand side of (2.12) is called the QR decomposition of A.

Proof. The decomposition is equivalent to $Q^{\mathrm{T}}A = R$, so that Q^{T} is an orthogonal matrix that transforms a given matrix A into an upper triangular matrix. In the following, we give a method of performing this transform by means of the Householder transforms.

Let $a^{(1)} = A(:,1)$, the first column vector of A. By computing $u^{(1)} := \tilde{a}$ and $b_1^{(1)}$, we perform the following transform:

$$a^{(1)} := \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} \longrightarrow u^{(1)} = \begin{bmatrix} a_{11} - b_1^{(1)} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}, \quad b^{(1)} = \begin{bmatrix} b_1^{(1)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $b_1^{(1)} = \pm\|a^{(1)}\|$. According to (2.11), let

$$P^{(1)} := I + u^{(1)}(u^{(1)})^{\mathrm{T}}/(b^{(1)})^{\mathrm{T}}u^{(1)}$$

and $P^{(1)}A := A^{(1)}$. Then we get

$$P^{(1)}A = A^{(1)} = \begin{bmatrix} b_1^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \cdots & a_{1n}^{(1)} \\ 0 & a_{22}^{(1)} & a_{23}^{(1)} & \cdots & a_{2n}^{(1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & a_{m2}^{(1)} & a_{m3}^{(1)} & \cdots & a_{mn}^{(1)} \end{bmatrix}$$

Thus the first column vector of $A^{(1)}$ is reduced to the vector $b^{(1)}$, where the column vectors $a_2, \cdots, a_n$ are subject to effects of $P^{(1)}$. But, in the transforms that follow, the vector $b^{(1)} := A^{(1)}(:,1)$ is intact, and this becomes the first column of R.

Next we consider the transforms of the second column vector of $A^{(1)}$. We define $a^{(2)}$, $u^{(2)}$ and $b^{(2)}$ as

$$a^{(2)} := \begin{bmatrix} 0 \\ a_{22}^{(1)} \\ a_{32}^{(1)} \\ \vdots \\ a_{m2}^{(1)} \end{bmatrix} \longrightarrow u^{(2)} = \begin{bmatrix} 0 \\ a_{22}^{(1)} - b_2^{(2)} \\ a_{32}^{(1)} \\ \vdots \\ a_{m2}^{(1)} \end{bmatrix}, \quad b^{(2)} = \begin{bmatrix} 0 \\ b_2^{(2)} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

where $b_2^{(2)} = \pm\|a^{(2)}\|$. Let $P^{(2)}$ be defined by

$$P^{(2)} := I + u^{(2)}(u^{(2)})^{\mathrm{T}}/(b^{(2)})^{\mathrm{T}}u^{(2)} = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline 0 & P_{22}^{(2)} & \cdots & P_{2m}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & P_{m2}^{(2)} & \cdots & P_{mm}^{(2)} \end{array}\right]$$

We see that $P^{(2)}$ is an orthogonal matrix, for which all the elements of the first row and column are zero except for $(1,1)$-element. Thus pre-multiplying $A^{(1)}$ by $P^{(2)}$ yields

$$P^{(2)}A^{(1)} = P^{(2)}P^{(1)}A = A^{(2)} = \begin{bmatrix} b_1^{(1)} & a_{12}^{(1)} & a_{13}^{(2)} & \cdots & a_{1n}^{(2)} \\ & b_2^{(2)} & a_{23}^{(2)} & \cdots & a_{2n}^{(2)} \\ & & a_{33}^{(2)} & \cdots & a_{3n}^{(2)} \\ & & \vdots & & \vdots \\ \mathbf{0} & & a_{m3}^{(2)} & \cdots & a_{mn}^{(2)} \end{bmatrix}$$

where we note that the first row and column of $A^{(2)}$ are the same as those of $A^{(1)}$ due to the form of $P^{(2)}$.

Repeating this procedure until the nth column, we get an upper triangular matrix $A^{(n)}$ of the form

$$P^{(n)}P^{(n-1)}\cdots P^{(1)}A = A^{(n)} = \begin{bmatrix} R \\ 0 \end{bmatrix} \tag{2.13}$$

Since each component $P^{(j)}$, $j = 1, \cdots, n$ is orthogonal and symmetric, we get

$$A = P^{(1)}P^{(2)}\cdots P^{(n)}A^{(n)} = Q\begin{bmatrix} R \\ 0 \end{bmatrix}$$

where $R \in \mathbb{R}^{n\times n}$ is upper triangular and $Q = [q_1, \cdots, q_n] \in \mathbb{R}^{m\times n}$ is orthogonal. This completes a proof of lemma. □

The QR decomposition is quite useful for computing an orthonormal basis for a set of vectors. In fact, it is a matrix realization of the Gram-Schmidt orthogonalization process. Suppose that $A \in \mathbb{R}^{m\times n}$ and $\operatorname{rank}(A) = n$. Let the QR decomposition of A be given by

$$A = [Q_A \quad Q_A^{\perp}]\begin{bmatrix} R \\ 0 \end{bmatrix} = Q_A R, \qquad Q_A \in \mathbb{R}^{m\times n} \tag{2.14}$$

Since R is nonsingular, we have $\operatorname{Im}(A) = \operatorname{Im}(Q_A)$, *i.e.*, the column vectors of Q_A form an orthonormal basis of $\operatorname{Im}(A)$, and those of $Q_A^{\perp}$ forms an orthonormal basis of the orthogonal complement $(\operatorname{Im} A)^{\perp}$.

It should, however, be noted that if $\operatorname{rank}(A) = r < n$, the QR decomposition does not necessarily gives an orthonormal basis for $\operatorname{Im}(A)$, since some of the diagonal elements of R become zero. For example, consider the following QR decomposition

$$A = [a_1 \;\; a_2 \;\; a_3] = [q_1 \;\; q_2 \;\; q_3]\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

Though we see that $\operatorname{rank}(A) = 2$, it is impossible to span $\operatorname{Im}(A)$ by any two vectors from q_1, q_2, q_3. But, for $\operatorname{rank}(A) = r < n$, it is easy to modify the QR decomposition algorithm so that the r column vectors $(q_1, \cdots, q_r)$ form an orthonormal basis of $\operatorname{Im}(A)$ with column pivoting; see [59].

2.5 Projections and Orthogonal Projections

Definition 2.2. *Suppose that $\mathbb{R}^n$ be given by a direct sum of subspaces $\mathcal{V}$ and $\mathcal{W}$, i.e.,*

$$\mathbb{R}^n = \mathcal{V} + \mathcal{W}, \qquad \mathcal{V} \cap \mathcal{W} = \{0\}$$

Then, $x \in \mathbb{R}^n$ can be uniquely expressed as

$$x = v + w, \qquad v \in \mathcal{V}, \qquad w \in \mathcal{W} \tag{2.15}$$

where v is the projection of x onto $\mathcal{V}$ along $\mathcal{W}$, and w is the projection of x onto $\mathcal{W}$ along $\mathcal{V}$. The uniqueness follows from $\mathcal{V} \cap \mathcal{W} = \{0\}$. □

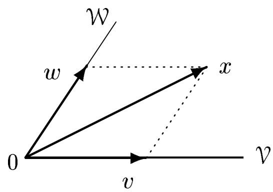

Figure 2.2. Oblique (or parallel) projection

The projection is often called the oblique (or parallel) projection, see Figure 2.2. We write the projection operator that transforms x onto $\mathcal{V}$ along $\mathcal{W}$ as $P^{\mathcal{V}}_{\|\mathcal{W}}$. Then, we have $v = P^{\mathcal{V}}_{\|\mathcal{W}}(x)$ and $w = P^{\mathcal{W}}_{\|\mathcal{V}}(x)$, and hence the unique decomposition of (2.15) is written as

$$x = P^{\mathcal{V}}_{\|\mathcal{W}}(x) + P^{\mathcal{W}}_{\|\mathcal{V}}(x)$$

We show that the projection is a linear operator. For $x,\ y \in \mathbb{R}^n$, we have the following decompositions

$$x = v + w, \qquad y = u + z, \qquad v,\ u \in \mathcal{V}, \qquad w,\ z \in \mathcal{W}$$

Since $x + y = (v + u) + (w + z),\ u + v \in \mathcal{V},\ w + z \in \mathcal{W}$, we see that $v + u$ is the oblique projection of $x + y$ onto $\mathcal{V}$ along $\mathcal{W}$. Hence, we have

$$P^{\mathcal{V}}_{\|\mathcal{W}}(x + y) = v + u = P^{\mathcal{V}}_{\|\mathcal{W}}(x) + P^{\mathcal{V}}_{\|\mathcal{W}}(y)$$

Moreover, for any α, we get $\alpha x = \alpha v + \alpha w,\ \alpha v \in \mathcal{V},\ \alpha w \in \mathcal{W}$, so that αv is the oblique projection of αx onto $\mathcal{V}$ along $\mathcal{W}$, implying that

$$P^{\mathcal{V}}_{\|\mathcal{W}}(\alpha x) = \alpha v = \alpha P^{\mathcal{V}}_{\|\mathcal{W}}(x)$$

From the above, we see that the projection $P^{\mathcal{V}}_{\|\mathcal{W}}$ is a linear operator on $\mathbb{R}^n$, so that it can be expressed as a matrix.

Lemma 2.4. *Suppose that $P \in \mathbb{R}^{n\times n}$ is idempotent, i.e.,*

$$P^2 = P \tag{2.16}$$

Then, we have

$$\mathrm{Ker}(P) = \mathrm{Im}(I_n - P) \tag{2.17}$$

and vice versa.

Proof. Let $x \in \mathrm{Ker}(P)$. Then, since $Px = 0$, we get $x = (I - P)x \in \mathrm{Im}(I - P)$, implying that $\mathrm{Ker}(P) \subset \mathrm{Im}(I-P)$. Also, for any $x \in \mathbb{R}^n$, we see that $P(I-P)x = 0$, showing that $\mathrm{Im}(I - P) \subset \mathrm{Ker}(P)$. This proves (2.17). Conversely, for any $z \in \mathbb{R}^n$, let $x = (I-P)z$. Then, we have $x \in \mathrm{Ker}(P)$, so that $0 = Px = P(I-P)z$ holds for any $z \in \mathbb{R}^n$, implying that $P^2 = P$. □

Corollary 2.1. *Suppose that (2.16) holds. Then, we have*

$$\mathbb{R}^n = \mathrm{Im}(P) + \mathrm{Ker}(P) \tag{2.18}$$

Proof. Since any $x \in \mathbb{R}^n$ can be written as $x = Px + (I - P)x$, we see from (2.17) that

$$\mathbb{R}^n = \mathrm{Im}(P) \vee \mathrm{Im}(I - P) = \mathrm{Im}(P) \vee \mathrm{Ker}(P) \tag{2.19}$$

Now let $x \in \mathrm{Im}(P) \cap \mathrm{Ker}(P)$. Then we have $x = Py$, $y \in \mathbb{R}^n$ and $Px = 0$. From (2.16), we get $0 = Px = P^2y = Py = x$ and hence $\mathrm{Im}(P) \cap \mathrm{Ker}(P) = \{0\}$. Thus the right-hand side of (2.19) is expressed as the direct sum. □

We now provide a necessary and sufficient condition such that P is a matrix that represents an oblique projection.

Lemma 2.5. *A matrix $P \in \mathbb{R}^{n\times n}$ is the projection matrix onto $\mathrm{Im}(P)$ along $\mathrm{Ker}(P)$ if and only if (2.16) holds.*

Proof. We prove the necessity. Since, for any $x \in \mathbb{R}^n$, $v = Px \in \mathrm{Im}(P)$, we have $P(Px) = Pv = v = Px$ for all x, implying that $P^2 = P$ holds. Conversely, to prove the sufficiency, we define

$$\mathcal{V} := \{v \mid v = Px,\ x \in \mathbb{R}^n\}, \qquad \mathcal{W} := \{w \mid w = (I - P)x,\ x \in \mathbb{R}^n\}$$

Since $\mathcal{V} \cap \mathcal{W} = \{0\}$, Lemma 2.4 implies that $x \in \mathbb{R}^n$ is decomposed uniquely as

$$x = Px + (I - P)x = v + w, \qquad v \in \mathcal{V}, \qquad w \in \mathcal{W}$$

From Definition 2.2, we see that P is the projection matrix onto $\mathcal{V} = \mathrm{Im}(P)$ along $\mathcal{W} = \mathrm{Ker}(P)$. □

Example 2.1. It can be shown that $P \in \mathbb{R}^{n\times n}$ is a projection if and only if P is expressed as

$$P = T\Delta_r T^{-1} \tag{2.20}$$

where T is a nonsingular matrix, and Δ_r is given by

$$\Delta_r = \mathrm{diag}(\underbrace{1, \cdots, 1}_{r}, 0, \ldots, 0) \tag{2.21}$$

In fact, it is obvious that P of (2.20) satisfies $P^2 = P$. Conversely, suppose that $P^2 = P$ holds. Let

$$\mathrm{Im}(P) = \mathrm{span}\{t_1, \cdots, t_r\}, \qquad \mathrm{Ker}(P) = \mathrm{span}\{t_{r+1}, \cdots, t_n\}$$

Noting that $x \in \mathrm{Im}(P) \Leftrightarrow Px = x$ and that $x \in \mathrm{Ker}(P) \Leftrightarrow Px = 0$, we get

$$P[t_1 \ \cdots \ t_r \ t_{r+1} \ \cdots \ t_n] = [t_1 \ \cdots \ t_r \ t_{r+1} \ \cdots \ t_n] \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$$

From Corollary 2.1, $T = [t_1 \ \cdots \ t_n]$ is nonsingular, showing that (2.20) holds.

Thus it follows from (2.20) that if $P^2 = P$, then $\mathrm{rank}(P) = \mathrm{trace}(P)$. □

Definition 2.3. *Suppose that $\mathcal{V} \subset \mathbb{R}^n$. Then, any $x \in \mathbb{R}^n$ can uniquely be decomposed as*

$$x = v + w, \qquad v \in \mathcal{V}, \qquad w \in \mathcal{V}^\perp \tag{2.22}$$

This is a particular case with $\mathcal{W} = \mathcal{V}^\perp$ in Definition 2.2, and v is called the orthogonal projection of x onto $\mathcal{V}$. See Figure 2.3 below. □

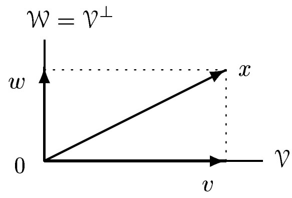

Figure 2.3. Orthogonal projection

For $x, y \in \mathbb{R}^n$, we consider the orthogonal decompositions $x = v_1 + w_1$ and $y = v_2 + w_2$, where $v_1, v_2 \in \mathcal{V}$ and $w_1, w_2 \in \mathcal{V}^\perp$. Let P be the orthogonal projection onto $\mathcal{V}$ along $\mathcal{V}^\perp$. Then, $v_1 = Px$, $v_2 = Py$. Since $v_2 \perp w_1$, $v_1 \perp w_2$,

$$\begin{aligned}(x, Py) &= (v_1 + w_1, v_2) = (v_1, v_2) = (v_1, v_2 + w_2) \\ &= (Px, y) = (x, P^{\mathrm{T}} y)\end{aligned}$$

holds for any x, y, so that we have $P = P^{\mathrm{T}}$. The next lemma provides a necessary and sufficient condition such that P is an orthogonal projection.

Lemma 2.6. *The matrix $P \in \mathbb{R}^{n \times n}$ is the orthogonal projection onto $\mathrm{Im}(P)$ if and only if the following two conditions hold.*

$$(i) \quad P^2 = P \qquad\qquad (ii) \quad P^{\mathrm{T}} = P \tag{2.23}$$

Proof. (Necessity) It is clear from Lemma 2.5 that $P^2 = P$ holds. The fact that $P^{\mathrm{T}} = P$ is already proved above.

(Sufficiency) It follows from Lemma 2.5 that the condition (i) implies that P is the projection matrix onto $\mathrm{Im}(P)$ along $\mathrm{Ker}(P)$. Condition (ii) implies that $\mathrm{Ker}(P) = \mathrm{Ker}(P^{\mathrm{T}}) = (\mathrm{Im}\ P)^{\perp}$. This means that the sufficiency part holds. □

Let $A \in \mathbb{R}^{n\times r}$ with $\mathrm{rank}(A) = r$ and $\mathrm{Im}(A) = \mathcal{A} \subset \mathbb{R}^n$. Let the QR decomposition of A be given by (2.14). Then, it follows that $\mathrm{Im}(Q_A) = \mathcal{A}$. Also, define

$$P_A = Q_A Q_A^{\mathrm{T}} \in \mathbb{R}^{n\times n} \tag{2.24}$$

It is clear that $P_A^{\mathrm{T}} = P_A$ and $P_A^2 = P_A$, so that the conditions (i) and (ii) of Lemma 2.6 are satisfied. Therefore, if we decompose $z \in \mathbb{R}^n$ as

$$z = x + y, \qquad x \in \mathcal{A}, \qquad y \in \mathcal{A}^{\perp} \tag{2.25}$$

then we get $x = P_A z$ and $y = (I - P_A)z$. Hence, P_A and $I - P_A$ are orthogonal projections onto $\mathcal{A}(= \mathrm{Im}\ A)$ and $\mathcal{A}^{\perp}$, respectively.

Lemma 2.7. *Suppose that $\mathcal{A}$ is a subspace of $\mathbb{R}^n$. Then, for any $z \in \mathbb{R}^n$, $P_A z$ is the unique vector satisfying the following*

$$\min_{x\in\mathcal{A}} \|z - x\| = \|z - P_A z\|$$

Proof. If $z \in \mathcal{A}$, then $P_A z = z$. Now suppose that $z \notin \mathcal{A}$. For any $x \in \mathcal{A}$, we have $x - P_A z \in \mathcal{A}$, but $(I - P_A)z$ is orthogonal to $\mathcal{A}$. Thus it follows that $x - P_A z \perp (I - P_A)z$. Hence,

$$\|z - x\|^2 = \|(I - P_A)z - (x - P_A z)\|^2 = \|(I - P_A)z\|^2 + \|x - P_A z\|^2$$

The right-hand side is minimized by $x = P_A z$, which is unique. □

2.6 Singular Value Decomposition

Though the singular value decomposition (SVD) can be applied to complex matrices, it is assumed here that matrices are real.

Lemma 2.8. *Suppose that the rank of $A \in \mathbb{R}^{m\times n}$ is $r \le \min(n, m)$. Then, there exist orthogonal matrices $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ such that*

$$A = U \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} V^{\mathrm{T}}, \qquad \Sigma_+ = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} \tag{2.26}$$

where $U^{\mathrm{T}}U = I_m$, $V^{\mathrm{T}}V = I_n$, and

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_p = 0, \quad p = \min(m,\, n)$$

We say that $\sigma_1, \cdots, \sigma_p$ are the singular values of A, and that (2.26) is the singular value decomposition (SVD).

Proof. Suppose that we know the eigenvalue decomposition of a nonnegative definite matrix. Since $A^{\mathrm{T}}A \in \mathbb{R}^{n\times n}$ is nonnegative definite, it can be diagonalized by an orthogonal transform $V \in \mathbb{R}^{n\times n}$. Let the eigenvalues of $A^{\mathrm{T}}A$ be given by $\lambda_1, \lambda_2, \cdots, \lambda_n$, and let the corresponding eigenvectors be given by $v_1, v_2, \cdots, v_n \in \mathbb{R}^n$. Thus we have $A^{\mathrm{T}}Av_i = \lambda_i v_i$, $i = 1, \cdots, n$. However, since $\mathrm{rank}(A) = r$, we have $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > \lambda_{r+1} = \cdots = \lambda_n = 0$. Define $\sigma_i = \sqrt{\lambda_i}$, $i = 1, \cdots, n$, and $V = [V_r \quad \tilde{V}_r]$, where

$$V_r = [v_1\ v_2\ \cdots\ v_r], \qquad \tilde{V}_r = [v_{r+1}\ v_{r+2}\ \cdots\ v_n]$$

It then follows that $V^{\mathrm{T}}V = I_n$ and that

$$A^{\mathrm{T}}Av_i = \sigma_i^2 v_i, \qquad i = 1, \cdots, r \tag{2.27a}$$

$$A^{\mathrm{T}}Av_i = 0, \qquad i = r+1, \cdots, n \tag{2.27b}$$

Also we define $U_r := AV_r\Sigma_+^{-1} \in \mathbb{R}^{m\times r}$. We see from (2.27a) that $A^{\mathrm{T}}AV_r = V_r\Sigma_+^2$ holds and

$$U_r^{\mathrm{T}}U_r = \Sigma_+^{-1}V_r^{\mathrm{T}}A^{\mathrm{T}}AV_r\Sigma_+^{-1} = \Sigma_+^{-1}V_r^{\mathrm{T}}(V_r\Sigma_+^2)\Sigma_+^{-1} = I_r \tag{2.28}$$

In other words, the column vectors in U_r form a set of orthonormal basis.

Now we choose $\tilde{U}_r \in \mathbb{R}^{m\times(m-r)}$ so that

$$U = [U_r \ \ \tilde{U}_r] \in \mathbb{R}^{m\times m}$$

is an orthogonal matrix, *i.e.*, $U^{\mathrm{T}}U = I_m$. Then it follows that

$$U^{\mathrm{T}}AV = \begin{bmatrix} U_r^{\mathrm{T}} \\ \tilde{U}_r^{\mathrm{T}} \end{bmatrix} A\,[V_r \ \ \tilde{V}_r] = \begin{bmatrix} U_r^{\mathrm{T}}AV_r & U_r^{\mathrm{T}}A\tilde{V}_r \\ \tilde{U}_r^{\mathrm{T}}AV_r & \tilde{U}_r^{\mathrm{T}}A\tilde{V}_r \end{bmatrix}$$

We see from (2.28) that the $(1,1)$-block element of the right-hand side of the above equation is Σ_+. From (2.27b), we get $A\tilde{V}_r = 0$. Thus $(1,2)$- and $(2,2)$-block elements are zero matrices. Also, since $\tilde{U}_r$ and U_r are orthogonal, $\tilde{U}_r^{\mathrm{T}}AV_r\Sigma_+^{-1} = 0$, implying that $(2,1)$-block element is also zero matrix. Thus we have shown that

$$U^{\mathrm{T}}AV = \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} = \Sigma$$

This completes the proof. □

It is clear that (2.26) can be expressed as

$$A = U\Sigma V^{\mathrm{T}} = U_r\Sigma_+V_r^{\mathrm{T}} \tag{2.29}$$

where $U_r \in \mathbb{R}^{m\times r}$ and $V_r \in \mathbb{R}^{n\times r}$. Note that in the following, we often write (2.29) as $A = U\Sigma_+ V^{\mathrm{T}}$, which is called the reduced SVD.

Let $\sigma(A)$ be the set of singular values of A, and $\sigma_i(A)$ ith singular value. As we can see from the above proof, the singular values of A are equal to the positive square roots of the eigenvalues of $A^{\mathrm{T}}A$, *i.e.*, for $A \in \mathbb{R}^{m\times n}$,

$$\sigma_i(A) = \sqrt{\lambda_i(A^{\mathrm{T}}A)}, \qquad i = 1, \cdots, n$$

Also, the column vectors of U, the left singular vectors of A, are the eigenvectors of AA^{T}, and the column vectors of V, the right singular vectors of A, are the eigenvectors of $A^{\mathrm{T}}A$. From (2.29), we have $AV_r = U_r\Sigma_+$ and $A^{\mathrm{T}}U_r = V_r\Sigma_+$, so that the ith right singular vector and the ith left singular vector are related by

$$Av_i = \sigma_i u_i, \qquad A^{\mathrm{T}}u_i = \sigma_i v_i, \qquad i = 1, \cdots, r$$

In the following, $\overline{\sigma}(A)$ and $\underline{\sigma}(A)$ denote the maximum and the minimum singular values, respectively.

Lemma 2.9. *Suppose that* $\mathrm{rank}(A) = r \le \min(m,n)$*. Then, the following properties (i)∼(v) hold.*

(i) Images and kernels of A and A^{T}:

$$\mathrm{Im}(A) = \mathrm{Im}(U_r), \qquad \mathrm{Ker}(A) = \mathrm{Im}(\tilde{V}_r)$$
$$\mathrm{Im}(A^{\mathrm{T}}) = \mathrm{Im}(V_r), \qquad \mathrm{Ker}(A^{\mathrm{T}}) = \mathrm{Im}(\tilde{U}_r)$$

(ii) The dyadic decomposition of A:

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^{\mathrm{T}} \; (= U\Sigma V^{\mathrm{T}})$$

(iii) The Frobenius norm and 2-norm:

$$\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}, \qquad \|A\|_2 = \sigma_1$$

(iv) Equivalence of norms:

$$\|A\|_2 \le \|A\|_F \le \sqrt{p}\,\|A\|_2, \qquad p = \min(m,n)$$

(v) The approximation by a lower rank matrix: Define the matrix A_k by

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^{\mathrm{T}}, \qquad k < r$$

Then, we have $\mathrm{rank}\,(A_k) = k$*, and*

$$\min_{\mathrm{rank}(B)=k} \|A - B\|_2 = \|A - A_k\|_2 = \sigma_{k+1}$$

where $B \in \mathbb{R}^{m\times n}$.

Proof. For a proof, see [59]. We prove only (v). Since

$$A - A_k = U\text{diag}(0, \cdots, 0, \sigma_{k+1}, \cdots, \sigma_p)V^{\mathrm{T}}$$

we have $\|A - A_k\|_2 = \sigma_{k+1}$. Let $B \in \mathbb{R}^{m\times n}$ be a matrix with rank k. Then, it suffices to show that $\|A - B\|_2 \geq \sigma_{k+1}$. Let $\{x_i \in \mathbb{R}^n, i = 1, \cdots, n-k\}$ be orthonormal vectors such that $\text{Ker}(B) = \text{span}\{x_1, \cdots, x_{n-k}\}$. Define also $\mathcal{V}_{k+1} := \text{span}\{v_1, \cdots, v_{k+1}\}$. We see that $\dim \text{Ker}(B) = n - k$ and $\dim(\mathcal{V}_{k+1}) = k + 1$. But $\text{Ker}(B)$ and $\mathcal{V}_{k+1}$ are subspaces of $\mathbb{R}^n$, so that $\text{Ker}(B) \cap \mathcal{V}_{k+1} \neq \{0\}$.

Let $z \in \text{Ker}(B) \cap \mathcal{V}_{k+1} \subset \mathbb{R}^n$ be a vector with $\|z\| = 1$. Then it follows that $Bz = 0$ and

$$Az = \sum_{i=1}^{p} \sigma_i u_i(v_i^{\mathrm{T}} z) = \sum_{i=1}^{k+1} \sigma_i (v_i^{\mathrm{T}} z)u_i$$

Since $(v_i^{\mathrm{T}} z)^2 \leq \|v_i\|^2\|z\|^2 = 1$, we have

$$\|A - B\|_2^2 \geq \|(A - B)z\|^2 = \|Az\|^2 = \sum_{i=1}^{k+1} \sigma_i^2 (v_i^{\mathrm{T}} z)^2 \geq \sigma_{k+1}^2$$

as was to be proved. □

Finding the rank of a matrix is most reliably done by the SVD. Let the SVD of $A \in \mathbb{R}^{m\times n}$ be given by (2.26). Let E be a perturbation to the matrix A, and $\{\tilde{\sigma}_i, i = 1, \cdots, p\}$ be the singular values of the perturbed matrix $A + E$. Then, it follows from [59] that

$$|\sigma_i - \tilde{\sigma}_i| < \|E\|_2 = \overline{\sigma}(E), \qquad i = 1, \cdots, p \tag{2.30}$$

This implies that the singular values are not very sensitive to perturbations.

We now define a matrix B with rank $r - 1$ as

$$B = U\text{diag}(\sigma_1, \cdots, \sigma_{r-1}, 0, \cdots, 0)V^{\mathrm{T}}$$

Then we have $\|A - B\|_2 = \sigma_r$. Thus, for any matrix B satisfying $\|A - B\|_2 < \sigma_r$, the rank of B is greater than or equal to r. Hence, as a "zero threshold," if we can choose a number $\delta < \sigma_r$, we can say that A has numerical rank r. Thus the smallest nonzero singular value plays a significant role in determining numerical rank of matrices.

2.7 Least-Squares Method

In this section, we consider the least-squares problem:

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|, \qquad A \in \mathbb{R}^{m\times n}, \qquad b \in \mathbb{R}^m \tag{2.31}$$

where $m \geq n$. Suppose that $\text{rank}(A) = n$, and let the QR decomposition of A be given by

$$A = Q \begin{bmatrix} R \\ 0 \end{bmatrix}, \qquad Q \in \mathbb{R}^{m \times m}, \qquad R \in \mathbb{R}^{n \times n}$$

Since the 2-norm is invariant under orthogonal transforms, we have

$$||Ax - b||^2 = ||Q^{\mathrm{T}}(Ax - b)||^2 = \left\| \begin{bmatrix} R \\ 0 \end{bmatrix} x - \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \right\|^2, \quad Q^{\mathrm{T}} b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

where $b_1 \in \mathbb{R}^n$, $b_2 \in \mathbb{R}^{m-n}$. Hence, it follows that

$$||Ax - b||^2 = ||Rx - b_1||^2 + ||b_2||^2$$

Since the second term $||b_2||^2$ in the right-hand side is independent of x, the least-squares problem is reduced to

$$Rx = b_1 \quad \Rightarrow \quad \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ 0 & & & r_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}$$

Since R is upper triangular, the solutions $x_n, x_{n-1}, \cdots, x_1$ are recursively computed by back substitution, starting from $x_n = \beta_n / r_{nn}$.

If the rank of $A \in \mathbb{R}^{m \times n}$ is less than n, some of the diagonal elements of R are zeros, so that the solution of the least-squares problem is not unique. But, putting the additional condition that the norm $||x||$ is minimum, we can obtain a unique solution. In the following, we explain a method of finding the minimum norm solution of the least-squares problem by means of the SVD.

Lemma 2.10. *Suppose that the rank of $A \in \mathbb{R}^{m \times n}$ is $r < \min(m, n)$, and the SVD is given by $A = U \Sigma_+ V^{\mathrm{T}}$, where $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{n \times r}$. Then, there exists a unique solution X satisfying the Moore-Penrose conditions:*

(i) $AXA = A$ (iii) $(AX)^{\mathrm{T}} = AX$

(ii) $XAX = X$ (iv) $(XA)^{\mathrm{T}} = XA$

The unique solution is given by

$$X = V \Sigma_+^{-1} U^{\mathrm{T}} =: A^{\dagger} \tag{2.32}$$

In this case, $X = A^{\dagger}$ is called the Moore-Penrose generalized inverse, or the pseudo-inverse, of A.

Proof. [83] It is easy to see that $A^{\dagger}$ of (2.32) satisfies the above four conditions. To prove the uniqueness, suppose that both X and Y satisfy the conditions. Then, it follows that

$$\begin{aligned} X &= XAX = (XA)^{\mathrm{T}} X = A^{\mathrm{T}} X^{\mathrm{T}} X = A^{\mathrm{T}} Y^{\mathrm{T}} A^{\mathrm{T}} X^{\mathrm{T}} X \\ &= A^{\mathrm{T}} Y^{\mathrm{T}} (XA) X = A^{\mathrm{T}} Y^{\mathrm{T}} X = YAX = (YAY)(AX) \\ &= YY^{\mathrm{T}} A^{\mathrm{T}} X^{\mathrm{T}} A^{\mathrm{T}} = YY^{\mathrm{T}} A^{\mathrm{T}} = YAY = Y \end{aligned}$$

as was to be proved. Note that all four conditions are used in the proof. □

In the above lemma, if $\mathrm{rank}(A) = n$, then $A^\dagger = (A^\mathrm{T} A)^{-1} A^\mathrm{T}$ and $A^\dagger A = I_n$. If $\mathrm{rank}(A) = m$, then we have $A^\dagger = A^\mathrm{T}(AA^\mathrm{T})^{-1}$ and $AA^\dagger = I_m$.

Lemma 2.11. *Suppose that the rank of $A \in \mathbb{R}^{m\times n}$ is $r < n$. Then, a general solution of the least-squares problem*

$$\min_{x\in\mathbb{R}^n} \|Ax - b\|, \qquad A \in \mathbb{R}^{m\times n}, \qquad b \in \mathbb{R}^m$$

is given by

$$x = A^\dagger b + (I_n - A^\dagger A)z, \qquad \forall\, z \in \mathbb{R}^n \tag{2.33}$$

Moreover, $x = A^\dagger b$ is the unique minimum norm solution.

Proof. It follows from Lemma 2.7 that the minimizing vector x should satisfy $Ax = P_A b$, where P_A is the orthogonal projection onto $\mathrm{Im}(A)$, which is given by $UU^\mathrm{T} = AA^\dagger$. Since $A(A^\dagger b) = P_A b$, we see that $x = A^\dagger b$ is a solution of the least-squares problem. We now seek a general solution of the form $x = A^\dagger b + y$, where y is to be determined. Since

$$Ay = A(x - A^\dagger b) = Ax - P_A b = 0$$

we get $y \in \mathrm{Ker}(A)$. By using $A = U\Sigma_+ V^\mathrm{T}$,

$$A^\dagger A = V\Sigma_+^{-1} U^\mathrm{T} U \Sigma_+ V^\mathrm{T} = VV^\mathrm{T}$$

Since $VV^\mathrm{T} = A^\dagger A$ is the orthogonal projection onto $\mathrm{Im}(A^\mathrm{T}) = (\mathrm{Ker}\ A)^\perp$, the orthogonal projection onto $\mathrm{Ker}(A)$ is given by $I_n - VV^\mathrm{T} = I_n - A^\dagger A$. Thus $y \in \mathrm{Ker}(A)$ is expressed as

$$y = (I_n - A^\dagger A)z, \qquad z \in \mathbb{R}^n$$

This proves (2.33). Finally, since $A^\dagger b$ and $(I_n - A^\dagger A)z$ are orthogonal, we get

$$\|x\|^2 = \|A^\dagger b\|^2 + \|(I_n - A^\dagger A)z\|^2 \geq \|A^\dagger b\|^2$$

where the equality holds if and only if $z = 0$. This completes the proof. □

Lemma 2.12. *A general solution of the least-squares problem*

$$\min_{X\in\mathbb{R}^{n\times p}} \|AX - B\|_F, \qquad A \in \mathbb{R}^{m\times n}, \qquad B \in \mathbb{R}^{m\times p}$$

is given by

$$X = A^\dagger B + (I_n - A^\dagger A)Z, \qquad \forall\, Z \in \mathbb{R}^{n\times p} \tag{2.34}$$

Proof. A proof is similar to that of Lemma 2.11. □

The minimum norm solution defined by (2.33) is expressed as

$$x = A^\dagger b = V\Sigma_+^{-1} U^\mathrm{T} b = \sum_{i=1}^{r} \frac{u_i^\mathrm{T} b}{\sigma_i} v_i$$

This indicates that if the singular values are small, then small changes in A and b may result in large changes in the solution x. From Lemma 2.9 (iv), $\|A - A_{r-1}\|_2 = \sigma_r$. Since the smallest singular value σ_r equals the distance from A to a set of matrices with ranks less than $r - 1$, it has the largest effect on the solution x. But, since the singular values are scale dependent, the normalized quantity, called the condition number,

$$\kappa(A) = \|A\|_2 \cdot \|A^\dagger\|_2 = \frac{\sigma_1}{\sigma_r}$$

is used as the sensitivity measure of the minimum norm solution to the data.

By definition, the condition number satisfies $\kappa(A) \geq 1$. If $\kappa(A)$ is very large, then A is called ill-conditioned. If $\kappa(A)$ is not very large, we say that A is well-conditioned. Obviously, the condition number of any orthogonal matrix is one, so that orthogonal matrices are perfectly conditioned.

2.8 Rank of Hankel Matrices

In this section, we consider the rank of Hankel matrices [51]. We assume that the sequence $h_1, h_2, \cdots$ below are real, but results are valid for complex sequences.

Definition 2.4. *Consider the infinite matrix*

$$H = \begin{bmatrix} h_1 & h_2 & h_3 & \cdots \\ h_2 & h_3 & h_4 & \cdots \\ h_3 & h_4 & h_5 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{2.35}$$

where (i, j)-element is given by h_{i+j}. This is called an infinite Hankel matrix, or Hankel operator. It should be noted that H has the same element along anti-diagonals. Also, define the matrix formed by the first k rows and l columns of H by

$$H_{k,l} = \begin{bmatrix} h_1 & h_2 & h_3 & \cdots & h_l \\ h_2 & h_3 & h_4 & \cdots & h_{l+1} \\ h_3 & h_4 & h_5 & \cdots & h_{l+2} \\ \vdots & \vdots & \vdots & & \vdots \\ h_k & h_{k+1} & h_{k+2} & \cdots & h_{k+l-1} \end{bmatrix} \tag{2.36}$$

This is called a finite Hankel matrix. □

Lemma 2.13. *Consider the finite Hankel matrix $H_{n,n}$ of order n. Suppose that the first l row vectors are linearly independent, but the first $l+1$ row vectors are linearly dependent. Then, it follows that* $\det H_{l,l} \neq 0$.

Proof. Let the first $l + 1$ row vectors of $H_{n,n}$ be given by $R_1, R_2, \cdots, R_l, R_{l+1}$. Since, from the assumption, $R_1, \cdots, R_l$ are linearly independent, we see that R_{l+1} is expressed as

$$R_{l+1} = \sum_{k=1}^{l} \alpha_k R_{l-k+1}$$

In particular, we have

$$h_i = \sum_{k=1}^{l} \alpha_k h_{i-k}, \qquad i = l+1, \cdots, l+n \tag{2.37}$$

Then the matrix formed by the first l row vectors $R_1, R_2, \cdots, R_l$ is given by

$$H_{l,n} = \begin{bmatrix} h_1 & h_2 & \cdots & h_n \\ h_2 & h_3 & \cdots & h_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ h_l & h_{l+1} & \cdots & h_{l+n-1} \end{bmatrix} \in \mathbb{R}^{l \times n} \tag{2.38}$$

where the rank of this matrix is l.

Now consider the column vectors of $H_{l,n}$. It follows from (2.37) that all the column vectors are expressed as a linear combination of the l preceding column vectors. Hence, in particular, the $(l+1)$th column vector is linearly dependent on the first l column vectors. But since the matrix of (2.38) has rank l, the first l column vectors are linearly independent, showing that $\det H_{l,l} \neq 0$. □

Example 2.2. Consider a finite symmetric Hankel matrix

$$H_{n,n} = \begin{bmatrix} h_1 & h_2 & \cdots & h_n \\ h_2 & h_3 & \cdots & h_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ h_n & h_{n+1} & \cdots & h_{2n-1} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Define the anti-diagonal matrix (or the backward identity)

$$J_n = \begin{bmatrix} 0 & & & 1 \\ & & 1 & \\ & \iddots & & \\ & 1 & & \\ 1 & & & 0 \end{bmatrix} \in \mathbb{R}^{n \times n} \tag{2.39}$$

Then it is easy to see that

$$J_n H_{n,n} = \begin{bmatrix} h_n & h_{n+1} & \cdots & h_{2n-1} \\ h_{n-1} & h_n & \cdots & h_{2n-2} \\ \vdots & \vdots & \ddots & \vdots \\ h_1 & h_2 & \cdots & h_n \end{bmatrix} =: T_n$$

where the matrix T_n is called a Toeplitz matrix with $t_{ij} = h_{n-i+j}$, *i.e.*, elements are constant along each diagonal. Also, from $J_n = J_n^{\mathrm{T}} = J_n^{-1}$, we see that $J_n T$ is a Hankel matrix for any Toeplitz matrix $T \in \mathbb{R}^n$. □

Lemma 2.14. *The infinite Hankel matrix of (2.35) has finite rank r if and only if there exist r real numbers $a_1, a_2, \cdots, a_r$ such that*

$$h_i = \sum_{k=1}^{r} a_k h_{i-k}, \qquad i = r+1,\, r+2,\, \cdots \tag{2.40}$$

Moreover, r is the least number with this property.

Proof. Suppose that $\operatorname{rank}(H) = r$ holds. Then the first $r+1$ rows $R_1, R_2, \cdots, R_{r+1}$ are linearly dependent. Hence, there exists an l $(\le r)$ such that $R_1, \cdots, R_l$ are linearly independent, and R_{l+1} is expressed as their linear combination

$$R_{l+1} = \sum_{k=1}^{l} a_k R_{l-k+1}$$

Now consider the row vectors $R_{i+1}, R_{i+2}, \cdots, R_{i+l+1}$, where i is an arbitrary nonnegative integer. From the structure of H, these vectors are obtained by removing the first i elements from $R_1, R_2, \cdots, R_{l+1}$, respectively. Thus we have

$$R_{i+l+1} = \sum_{k=1}^{l} a_k R_{i+l-k+1}, \qquad i = 0,\, 1,\, \cdots \tag{2.41}$$

It therefore follows that any row vector of H below the $(l+1)$th row can be expressed in terms of a linear combination of the l preceding row vectors, and hence in terms of linearly independent first l row vectors. Replacing l by r in (2.41), we have (2.40).

Conversely, suppose that (2.40) holds. Then, all the rows (columns) of H are expressed in terms of linear combinations of the first r rows (columns). Thus all the minors of H whose orders are greater than r are zero, and H has rank r at most. But the rank cannot be smaller than r; otherwise (2.40) is satisfied with a smaller value of r. This contradicts the second condition of the lemma. □

The above result is a basis for the realization theory due to Ho and Kalman [72], to be discussed in Chapter 3, where a matrix version of Lemma 2.14 will be proved.

2.9 Notes and References

- In this chapter, we have presented basic facts related to numerical linear algebra which will be needed in later chapters, including the QR decomposition, the orthogonal and oblique projections, the SVD, the least-squares method, the rank of Hankel matrices. Problems at the end of chapter include some useful formulas and results to be used in this book.
- Main references used are Golub and Van Loan [59], Gantmacher [51], Horn and Johnson [73], and Trefethen and Bau [157]. Earlier papers that have dealt with the issues of numerical linear algebra in system theory are [94] and [122]; see also the preprint book [125].

- For the history of SVD and related numerical methods, see [60, 148, 165]. The early developments in statistics, including the least-squares and the measurement of uncertainties, are covered in [149].

2.10 Problems

2.1 Prove the following by using the SVD, where $A \in \mathbb{R}^{m\times n}$, $B \in \mathbb{R}^{n\times p}$.

(a) $\mathrm{Im}(A) \oplus \mathrm{Ker}(A^{\mathrm{T}}) = \mathbb{R}^m, \quad \mathrm{Im}(A^{\mathrm{T}}) \oplus \mathrm{Ker}(A) = \mathbb{R}^n$

(b) $\mathrm{Ker}(A^{\mathrm{T}}) = (\mathrm{Im}\ A)^{\perp}, \quad \mathrm{Im}(A^{\mathrm{T}}) = (\mathrm{Ker}\ A)^{\perp}$

(c) $\mathrm{Im}(A) = A\mathbb{R}^n = \mathrm{Im}(AA^{\mathrm{T}}), \quad A\,\mathrm{Im}(B) = \mathrm{Im}(AB)$

2.2 Prove the following matrix identities

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix}$$
$$= \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix}$$

where it is assumed that A^{-1} and D^{-1} exist.

2.3 (a) Using the above results, prove the determinant of the block matrix.

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A)\det(D - CA^{-1}B)$$
$$= \det(D)\det(A - BD^{-1}C)$$

(b) Defining $A = I_n$ and $D = I_m$, show that

$$\det(I_m - CB) = \det(I_n - BC)$$

(c) Prove the formulas for the inverses of block matrices

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B\Delta^{-1}CA^{-1} & -A^{-1}B\Delta^{-1} \\ -\Delta^{-1}CA^{-1} & \Delta^{-1} \end{bmatrix}$$
$$= \begin{bmatrix} \Pi^{-1} & -\Pi^{-1}BD^{-1} \\ -D^{-1}C\Pi^{-1} & D^{-1} + D^{-1}C\Pi^{-1}BD^{-1} \end{bmatrix}$$

where $\Delta := D - CA^{-1}B$, $\Pi := A - BD^{-1}C$. For $C = 0$, we get

$$\begin{bmatrix} A & B \\ 0 & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BD^{-1} \\ 0 & D^{-1} \end{bmatrix}$$

(d) Prove the matrix inversion lemma.

$$[A + BD^{-1}C]^{-1} = A^{-1} - A^{-1}B[D + CA^{-1}B]^{-1}CA^{-1}$$

2.4 Show without using the result of Example 2.1 that if P is idempotent ($P^2 = P$), then all the eigenvalues are either zero or one.

2.5 For $P \in \mathbb{R}^{n\times n}$, show that the following statements are equivalent.

(a) $P^2 = P$

(b) $\mathrm{Im}(P) + \mathrm{Im}(I_n - P) = \mathbb{R}^n$

(c) $\mathrm{rank}(P) + \mathrm{rank}(I_n - P) = n$

2.6 Suppose that $Z = [T \quad U] \in \mathbb{R}^{n\times n}$ is nonsingular, where $T \in \mathbb{R}^{n\times r}$, $U \in \mathbb{R}^{n\times(n-r)}$. Let the inverse matrix of Z be given by

$$Z^{-1} = \begin{bmatrix} L \\ V \end{bmatrix}, \quad L \in \mathbb{R}^{r\times n}, \quad V \in \mathbb{R}^{(n-r)\times n}$$

Then it follows that $TL + UV = I_n$ and

$$\begin{bmatrix} L \\ V \end{bmatrix} [T \quad U] = \begin{bmatrix} LT & LU \\ VT & VU \end{bmatrix} = \begin{bmatrix} I_r & 0 \\ 0 & I_{n-r} \end{bmatrix}$$

Show that $P := TL$ is the oblique projection onto $\mathrm{Im}(T)$ along $\mathrm{Ker}(L)$, and that $Q := UV$ is the oblique projection onto $\mathrm{Ker}(L)$ [$= \mathrm{Im}(U)$] along $\mathrm{Im}(T)$ [$= \mathrm{Ker}(V)$].

2.7 In the above problem, define

$$T = \begin{bmatrix} I_r \\ 0 \end{bmatrix}, \qquad U = \begin{bmatrix} -X \\ I_{n-r} \end{bmatrix}$$

Compute the projection $P = TL$ by means of L and V. Show that (2.16) is satisfied if P has the following representation

$$P = \begin{bmatrix} I_r & X \\ 0 & 0 \end{bmatrix}, \qquad X \in \mathbb{R}^{r\times n}$$

2.8 By using (2.29), prove the following.

(a) $V_r V_r^{\mathrm{T}}$: the orthogonal projection from $\mathbb{R}^n$ onto $\mathrm{Im}(A^{\mathrm{T}})$

(b) $\tilde{V}_r \tilde{V}_r^{\mathrm{T}}$: the orthogonal projection from $\mathbb{R}^n$ onto $\mathrm{Ker}(A)$

(c) $U_r U_r^{\mathrm{T}}$: the orthogonal projection from $\mathbb{R}^m$ onto $\mathrm{Im}(A)$

(d) $\tilde{U}_r \tilde{U}_r^{\mathrm{T}}$: the orthogonal projection from $\mathbb{R}^m$ onto $\mathrm{Ker}(A^{\mathrm{T}})$

2.9 For $A \in \mathbb{R}^{m\times n}$, show that $A^{\mathrm{T}}(AA^{\mathrm{T}})^{\dagger} = A^{\dagger}$ and $(A^{\mathrm{T}}A)^{\dagger}A^{\mathrm{T}} = A^{\dagger}$.

2.10 Let $A \in \mathbb{R}^{m\times n}$ with $m \ge n$. By using the SVD, show that there exist an orthogonal matrix $Q \in \mathbb{R}^{m\times n}$ and a nonnegative matrix $\Pi \in \mathbb{R}^{n\times n}$ such that $A = Q\Pi$.

http://www.springer.com/978-1-85233-981-4

Subspace Methods for System Identification
Katayama, T.
2005, XVI, 392 p., Softcover
ISBN: 978-1-85233-981-4