1 Matrix Theory

We collect in this chapter useful background material on matrix theory and linear algebra. The emphasis is on results that are needed for future developments. Among other concepts, we review symmetric and nonnegative-definite matrices, range spaces and nullspaces, as well as several matrix decompositions, including the spectral decomposition, the triangular decomposition, the QR decomposition, and the singular value decomposition (SVD). We also discuss vector and matrix norms, Kronecker products, Schur complements, and the useful Rayleigh–Ritz characterization of the eigenvalues of symmetric matrices.

1.1 SYMMETRIC MATRICES

Symmetric and nonnegative-definite matrices play a prominent role in data analysis. We review some of their properties in this section. Thus, consider an arbitrary square matrix of size $N \times N$ with real entries, written as $A \in \mathbb{R}^{N \times N}$. The transpose of A is denoted by A^{T} and is obtained by transforming the rows of A into columns of A^{T}. For example,

$$A = \begin{bmatrix} 1 & -1 & 3 \\ -2 & 4 & 5 \\ 0 & 6 & 8 \end{bmatrix} \implies A^{\mathsf{T}} = \begin{bmatrix} 1 & -2 & 0 \\ -1 & 4 & 6 \\ 3 & 5 & 8 \end{bmatrix} \tag{1.1}$$

The matrix A is said to be symmetric if it happens to coincide with its matrix transpose, i.e., if it satisfies

$$A = A^{\mathsf{T}} \qquad (\textbf{symmetry}) \tag{1.2}$$

Real eigenvalues

A useful property of symmetric matrices is that they can only have *real* eigenvalues. To see this, let u represent a column eigenvector of A corresponding to some eigenvalue λ, i.e., u is nonzero and satisfies along with λ the relation:

$$Au = \lambda u \tag{1.3}$$

The eigenvector u may be complex-valued so that, in general, $u \in \mathbb{C}^N$. Let the symbol $*$ denote the operation of complex conjugate transposition, so that u^*

is the row vector that is obtained by transposing u and replacing its entries by their complex conjugate values, e.g.,

$$u \triangleq \begin{bmatrix} 1+j \\ 2 \\ -2+3j \end{bmatrix} \implies u^* = \begin{bmatrix} 1-j & 2 & -2-3j \end{bmatrix} \tag{1.4}$$

where $j \triangleq \sqrt{-1}$. The same complex conjugation operation can be applied to matrices as well so that, e.g.,

$$B = \begin{bmatrix} 1 & j & -2+j \\ 3-j & 1-2j & 0 \end{bmatrix} \implies B^* = \begin{bmatrix} 1 & 3+j \\ -j & 1+2j \\ -2-j & 0 \end{bmatrix} \tag{1.5}$$

Returning to (1.3) and multiplying from the left by the row vector u^* we get

$$u^*Au = \lambda u^*u \;=\; \lambda \|u\|^2 \tag{1.6}$$

where the notation $\|\cdot\|$ denotes the Euclidean norm of its vector argument. Note that the quantity u^*Au is a scalar. Moreover, it is real-valued because it coincides with its complex conjugate value:

$$(u^*Au)^* \;=\; u^*A^*(u^*)^* \;=\; u^*Au \tag{1.7}$$

where in the last step we used the fact that $A^* = A$ since A is real-valued and symmetric. Therefore, u^*Au is real and, from equality (1.6), we conclude that $\lambda\|u\|^2$ must also be real. But since $\|u\|^2$ is real and nonzero, we conclude that the eigenvalue λ must be real too.

One consequence of this conclusion is that we can always find *real-valued* eigenvectors for symmetric matrices. Indeed, if we express u in terms of its real and imaginary vector components, say, as

$$u = p + jq, \quad p, q \in \mathbb{R}^N \tag{1.8}$$

then, using (1.3) and the fact that λ is real, we conclude that it must hold:

$$Ap = \lambda p, \quad Aq = \lambda q \tag{1.9}$$

so that p and q are eigenvectors associated with λ.

Spectral theorem

A second important property of real symmetric matrices, one whose proof requires a more elaborate argument and is deferred to Appendix 1.A, is that such matrices always have a *full* set of orthonormal eigenvectors. That is, if $A \in \mathbb{R}^{N\times N}$ is symmetric, then there will exist a set of N orthonormal real eigenvectors $u_n \in \mathbb{R}^N$ satisfying

$$Au_n = \lambda_n u_n, \quad \|u_n\|^2 = 1, \quad u_n^{\mathsf{T}}u_m = 0 \;\text{ for } n \neq m \tag{1.10}$$

where all N eigenvalues $\{\lambda_n, n = 1, 2, \ldots, N\}$ are real, and all eigenvectors $\{u_n\}$ have unit norm and are orthogonal to each other. This result is known as the

spectral theorem. For illustration purposes, assume A is 3×3. Then, the above statement asserts that there will exist three real orthonormal vectors $\{u_1, u_2, u_3\}$ and three real eigenvalues $\{\lambda_1, \lambda_2, \lambda_3\}$ such that

$$A \underbrace{\begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}}_{\triangleq U} = \underbrace{\begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_3 \end{bmatrix}}_{\triangleq \Lambda} \tag{1.11}$$

where we are introducing the matrices U and Λ for compactness of notation: U contains real eigenvectors for A and Λ is a diagonal matrix with the corresponding eigenvalues. Then, we can write (1.11) more compactly as

$$AU = U\Lambda \tag{1.12}$$

However, the fact that the columns of U are orthogonal to each other and have unit norms implies that U satisfies the important normalization property:

$$UU^{\mathsf{T}} = I_N \quad \text{and} \quad U^{\mathsf{T}}U = I_N \tag{1.13}$$

That is, the product of U with U^{T} (or U^{T} with U) results in the identity matrix of size $N \times N$ – see Prob. 1.1. We say that U is an *orthogonal* matrix. Using this property and multiplying the matrix equality (1.12) by U^{T} from the right we get

$$A \underbrace{UU^{\mathsf{T}}}_{=I} = U\Lambda U^{\mathsf{T}} \tag{1.14}$$

We therefore conclude that every real symmetric matrix A can be expressed in the following spectral (or eigen-) decomposition form:

$$\boxed{A = U\Lambda U^{\mathsf{T}}} \quad (\textbf{eigen-decomposition}) \tag{1.15a}$$

where, for general dimensions, the $N \times N$ matrices Λ and U are constructed from the eigenvalues and orthonormal eigenvectors of A as follows:

$$\Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_N\} \tag{1.15b}$$

$$U = \begin{bmatrix} u_1 & u_2 & \ldots & u_N \end{bmatrix} \tag{1.15c}$$

Rayleigh–Ritz ratio

There is a useful characterization of the smallest and largest eigenvalues of real symmetric matrices, known as the Rayleigh–Ritz ratio. Specifically, if $A \in \mathbb{R}^{N \times N}$ is symmetric, then it holds that for all vectors $x \in \mathbb{R}^N$:

$$\boxed{\lambda_{\min}\|x\|^2 \leq x^{\mathsf{T}}Ax \leq \lambda_{\max}\|x\|^2} \tag{1.16}$$

as well as

$$\lambda_{\min} = \min_{x \neq 0} \left(\frac{x^{\mathsf{T}} A x}{x^{\mathsf{T}} x} \right) = \min_{\|x\|=1} x^{\mathsf{T}} A x \tag{1.17a}$$

$$\lambda_{\max} = \max_{x \neq 0} \left(\frac{x^{\mathsf{T}} A x}{x^{\mathsf{T}} x} \right) = \max_{\|x\|=1} x^{\mathsf{T}} A x \tag{1.17b}$$

where $\{\lambda_{\min}, \lambda_{\max}\}$ denote the smallest and largest eigenvalues of A. The ratio $x^{\mathsf{T}} A x / x^{\mathsf{T}} x$ is called the Rayleigh–Ritz ratio.

Proof of (1.16) and (1.17a)–(1.17b): Consider the eigen-decomposition (1.15a) and introduce the vector $y = U^{\mathsf{T}} x$ for any vector x. Then,

$$x^{\mathsf{T}} A x \;=\; x^{\mathsf{T}} U \Lambda U^{\mathsf{T}} x \;=\; y^{\mathsf{T}} \Lambda y \;=\; \sum_{n=1}^{N} \lambda_n y_n^2 \tag{1.18}$$

with the $\{y_n\}$ denoting the individual entries of y. Now since the squared terms $\{y_n^2\}$ are nonnegative and the $\{\lambda_n\}$ are real, we get

$$\lambda_{\min} \left(\sum_{n=1}^{N} y_n^2 \right) \leq \sum_{n=1}^{N} \lambda_n y_n^2 \leq \lambda_{\max} \left(\sum_{n=1}^{N} y_n^2 \right) \tag{1.19}$$

or, equivalently,

$$\lambda_{\min} \|y\|^2 \leq x^{\mathsf{T}} A x \leq \lambda_{\max} \|y\|^2 \tag{1.20}$$

Using the fact that U is orthogonal and, hence,

$$\|y\|^2 \;=\; y^{\mathsf{T}} y \;=\; x \underbrace{U U^{\mathsf{T}}}_{=I} x \;=\; \|x\|^2 \tag{1.21}$$

we conclude that (1.16) holds. The lower (upper) bound in (1.19) is achieved when x is chosen as the eigenvector $u_{\min}(u_{\max})$ corresponding to $\lambda_{\min}(\lambda_{\max})$.

■

Example 1.1 (**Quadratic curve**) Consider the two-dimensional function

$$g(r, s) = a r^2 + a s^2 + 2 b r s, \quad r, s \in \mathbb{R} \tag{1.22}$$

We would like to determine the largest and smallest values that the function attains on the circle $r^2 + s^2 = 1$. One way to solve the problem is to recognize that $g(r, s)$ can be rewritten as

$$g(r, s) = \underbrace{\begin{bmatrix} r & s \end{bmatrix}}_{\triangleq x^{\mathsf{T}}} \underbrace{\begin{bmatrix} a & b \\ b & a \end{bmatrix}}_{\triangleq A} \underbrace{\begin{bmatrix} r \\ s \end{bmatrix}}_{=x} \;=\; x^{\mathsf{T}} A x \tag{1.23}$$

We therefore want to determine the extreme values of the quadratic form $x^{\mathsf{T}} A x$ under the constraint $\|x\| = 1$. According to (1.17a)–(1.17b), these values correspond to $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$. It can be easily verified that the eigenvalues of A are given by $\lambda(A) = \{a - b, a + b\}$ and, hence,

$$\lambda_{\min}(A) = \min\{a - b, a + b\}, \qquad \lambda_{\max}(A) = \max\{a - b, a + b\} \tag{1.24}$$

1.2 POSITIVE-DEFINITE MATRICES

An $N \times N$ real symmetric matrix A is said to be nonnegative-definite (also called positive semi-definite) if it satisfies the property:

$$v^{\mathsf{T}} A v \geq 0, \qquad \text{for all column vectors } v \in \mathbb{R}^N \tag{1.25}$$

The matrix A is said to be positive-definite if $v^{\mathsf{T}} A v > 0$ for all $v \neq 0$. We denote a positive-definite matrix by writing $A > 0$ and a positive semi-definite matrix by writing $A \geq 0$.

Example 1.2 (**Diagonal matrices**) The notion of positive semi-definiteness is trivial for diagonal matrices. Consider the diagonal matrix

$$A = \text{diag}\{a_1, a_2, a_3\} \in \mathbb{R}^{3\times 3} \tag{1.26}$$

and let

$$v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3 \tag{1.27}$$

denote an arbitrary vector. Then, some simple algebra shows that

$$v^{\mathsf{T}} A v \;=\; a_1 v_1^2 + a_2 v_2^2 + a_3 v_3^2 \tag{1.28}$$

This expression will be nonnegative for *any* v if, and only if, the entries a_n are all nonnegative. This is because if any a_n is negative, say a_2, then we can select a vector v with an entry v_2 that is large enough to result in a negative term $a_2 v_2^2$ that exceeds the contribution of the other two terms in the sum $v^{\mathsf{T}} A v$. Therefore, for a diagonal matrix to be positive semi-definite, it is necessary and sufficient that its diagonal entries be nonnegative. Likewise, a diagonal matrix A is positive-definite if, and only if, its diagonal entries are positive. We cannot extrapolate and say that a general nondiagonal matrix A is positive semi-definite if all its entries are nonnegative; this conclusion is *not* true, as the next example shows.

Example 1.3 (**Nondiagonal matrices**) Consider the 2×2 matrix

$$A = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \tag{1.29}$$

This matrix is positive-definite. Indeed, pick any nonzero column vector $v \in \mathbb{R}^2$. Then,

$$\begin{aligned} v^{\mathsf{T}} A v &= \begin{bmatrix} v_1 & v_2 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \\ &= 3v_1^2 + 3v_2^2 - 2v_1 v_2 \\ &= (v_1 - v_2)^2 \;+\; 2v_1^2 \;+\; 2v_2^2 \\ &> 0, \quad \text{for any } v \neq 0 \end{aligned} \tag{1.30}$$

Among the several equivalent characterizations of positive-definite matrices, we note that an $N \times N$ real symmetric matrix A is positive-definite if, and only if, all its N eigenvalues are positive:

$$\boxed{A > 0 \iff \{\lambda_n > 0\}_{n=1}^{N}} \tag{1.31}$$

One proof relies on the use of the eigen-decomposition of A.

Proof of (1.31): We need to prove the statement in both directions. Assume initially that A is positive-definite and let us establish that all its eigenvalues are positive. Let $A = U\Lambda U^{\mathsf{T}}$ denote the spectral decomposition of A. Let also u_n denote the nth column of U corresponding to the eigenvalue λ_n, i.e., $Au_n = \lambda_n u_n$ with $\|u_n\|^2 = 1$. If we multiply this equality from the left by u_n^{T} we get

$$u_n^{\mathsf{T}} A u_n = \lambda_n \|u_n\|^2 = \lambda_n > 0 \tag{1.32}$$

where the last inequality follows from the fact that $u^{\mathsf{T}}Au > 0$ for any nonzero vector u since A is assumed to be positive-definite. Therefore, $A > 0$ implies $\lambda_n > 0$ for $n = 1, 2, \ldots, N$.

Conversely, assume all $\lambda_n > 0$ and let us show that $A > 0$. Multiply the equality $A = U\Lambda U^{\mathsf{T}}$ by any *nonzero* vector v and its transpose, from right and left, to get

$$v^{\mathsf{T}} A v = v^{\mathsf{T}} U \Lambda U^{\mathsf{T}} v \tag{1.33}$$

Now introduce the real diagonal matrix

$$D \triangleq \operatorname{diag}\left\{\sqrt{\lambda_1},\ \sqrt{\lambda_2},\ \ldots,\ \sqrt{\lambda_n}\right\} \tag{1.34}$$

and the vector

$$s \triangleq DU^{\mathsf{T}} v \tag{1.35}$$

The vector s is nonzero. This can be seen as follows. Let $w = U^{\mathsf{T}} v$. Then, the vectors v and w have the same Euclidean norm since

$$\|w\|^2 = w^{\mathsf{T}} w = v^{\mathsf{T}} \underbrace{UU^{\mathsf{T}}}_{=I} v = v^{\mathsf{T}} v = \|v\|^2 \tag{1.36}$$

It follows that the vector w is nonzero since v is nonzero. Now since $s = Dw$ and all entries of D are nonzero, we conclude that $s \neq 0$. Returning to (1.33), we get

$$v^{\mathsf{T}} A v = \|s\|^2 > 0 \tag{1.37}$$

for any nonzero v, which establishes that $A > 0$.

■

In a similar vein, we can show that

$$\boxed{A \geq 0 \iff \{\lambda_n \geq 0\}_{n=1}^{N}} \tag{1.38}$$

Example 1.4 (**Positive-definite matrix**) Consider again the 2×2 matrix from Example 1.3:

$$A = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \tag{1.39}$$

We established in that example from first principles that $A > 0$. Alternatively, we can determine the eigenvalues of A and verify that they are positive. The eigenvalues are the roots of the characteristic equation, $\det(\lambda I - A) = 0$, which leads to the quadratic equation $(\lambda - 3)^2 - 1 = 0$ so that $\lambda_1 = 4 > 0$ and $\lambda_2 = 2 > 0$.

A second useful property of positive-definite matrices is that they have positive determinants. To see this, recall first that for two square matrices A and B it holds that

$$\det(AB) = \det(A)\,\det(B) \tag{1.40}$$

That is, the determinant of the product is equal to the product of the determinants. Now starting with a positive-definite matrix A, and applying the above determinant formula to its eigen-decomposition (1.15a), we get

$$\det A = (\det U)\,(\det \Lambda)\,\left(\det U^{\mathsf{T}}\right) \tag{1.41}$$

But $UU^{\mathsf{T}} = I$ so that

$$(\det U)\,\left(\det U^{\mathsf{T}}\right) = 1 \tag{1.42}$$

and we conclude that

$$\det A = \det \Lambda = \prod_{n=1}^{N} \lambda_n \tag{1.43}$$

This result is actually general and holds for arbitrary square matrices A (the matrices do not need to be symmetric or positive-definite): the determinant of a matrix is always equal to the product of its eigenvalues (counting multiplicities) – see Prob. 1.2. Now, when the matrix A happens to be positive-definite, all its eigenvalues will be positive and, hence,

$$\boxed{A > 0 \implies \det A > 0} \tag{1.44}$$

Note that this statement goes in one direction only; the converse is not true.

1.3 RANGE SPACES AND NULLSPACES

Let A denote an $N \times M$ real matrix without any constraint on the relative sizes of N and M. When $N = M$, we say that A is a square matrix. Otherwise, when $N > M$, we say that A is a "tall" matrix and when $N < M$ we say that A is a "fat" matrix.

Definitions

The *column span* or the *range space* of A is defined as the set of all $N \times 1$ vectors q that can be generated by Ap, for all $M \times 1$ vectors p. We denote the column span of A by

$$\boxed{\mathcal{R}(A) \triangleq \Big\{\text{set of all } q \in \mathbb{R}^N \text{ such that } q = Ap \text{ for some } p \in \mathbb{R}^M\Big\}} \tag{1.45}$$

Likewise, the nullspace of A is the set of all $M \times 1$ vectors p that are annihilated by A, namely, that satisfy $Ap = 0$. We denote the nullspace of A by

$$\boxed{\mathcal{N}(A) \triangleq \Big\{\text{set of all } p \in \mathbb{R}^M \text{ such that } Ap = 0\Big\}} \tag{1.46}$$

The rank of a matrix A is defined as the number of linearly independent columns of A. It can be verified that, for any matrix A, the number of linearly independent columns is also equal to the number of linearly independent rows – see Prob. 1.5. It therefore holds that

$$\boxed{\text{rank}(A) \leq \min\{N, M\}} \tag{1.47}$$

That is, the rank of a matrix never exceeds its smallest dimension. A matrix is said to have *full rank* if

$$\text{rank}(A) = \min\{N, M\} \tag{1.48}$$

Otherwise, the matrix is said to be *rank deficient*.

If A is a square matrix (i.e., $N = M$), then rank deficiency is equivalent to a zero determinant, $\det A = 0$. Indeed, if A is rank deficient, then there exists a nonzero p such that $Ap = 0$. This means that $\lambda = 0$ is an eigenvalue of A so that its determinant must be zero.

Useful relations

One useful property that follows from the definition of range spaces and nullspaces is that any vector $z \in \mathbb{R}^N$ from the nullspace of A^{T} (not A) is orthogonal to any vector $q \in \mathbb{R}^N$ in the range space of A, i.e.,

$$\boxed{z \in \mathcal{N}(A^{\mathsf{T}}), \quad q \in \mathcal{R}(A) \quad \Longrightarrow \quad z^{\mathsf{T}} q = 0} \tag{1.49}$$

Proof of (1.49): Indeed, $z \in \mathcal{N}(A^{\mathsf{T}})$ implies that $A^{\mathsf{T}} z = 0$ or, equivalently, $z^{\mathsf{T}} A = 0$. Now write $q = Ap$ for some p. Then, $z^{\mathsf{T}} q = z^{\mathsf{T}} Ap = 0$, as desired.

■

A second useful property is that the matrices $A^{\mathsf{T}} A$ and A^{T} have the *same* range space (i.e., they span the same space). Also, A and $A^{\mathsf{T}} A$ have the same nullspace, i.e.,

$$\boxed{\mathcal{R}(A^{\mathsf{T}}) = \mathcal{R}(A^{\mathsf{T}} A), \qquad \mathcal{N}(A) = \mathcal{N}(A^{\mathsf{T}} A)} \tag{1.50}$$

Proof of (1.50): Consider a vector $q \in \mathcal{R}(A^{\mathsf{T}}A)$, i.e., $q = A^{\mathsf{T}}Ap$ for some p. Define $r = Ap$, then $q = A^{\mathsf{T}}r$. This shows that $q \in \mathcal{R}(A^{\mathsf{T}})$ and we conclude that $\mathcal{R}(A^{\mathsf{T}}A) \subset \mathcal{R}(A^{\mathsf{T}})$. The proof of the converse statement requires more effort.

Consider a vector $q \in \mathcal{R}(A^{\mathsf{T}})$ and let us show by contradiction that $q \in \mathcal{R}(A^{\mathsf{T}}A)$. Assume, to the contrary, that q does not lie in $\mathcal{R}(A^{\mathsf{T}}A)$. This implies by (1.49) that there exists a vector z in the nullspace of $A^{\mathsf{T}}A$ that is not orthogonal to q, i.e., $A^{\mathsf{T}}Az = 0$ and $z^{\mathsf{T}}q \neq 0$. Now, if we multiply the equality $A^{\mathsf{T}}Az = 0$ by z^{T} from the left we obtain that $z^{\mathsf{T}}A^{\mathsf{T}}Az = 0$ or, equivalently, $\|Az\|^2 = 0$. Therefore, Az is necessarily the zero vector, $Az = 0$. But from $q \in \mathcal{R}(A^{\mathsf{T}})$ we have that $q = A^{\mathsf{T}}p$ for some p. Then, it must hold that $z^{\mathsf{T}}q = z^{\mathsf{T}}A^{\mathsf{T}}p = 0$, which contradicts $z^{\mathsf{T}}q \neq 0$. Therefore, we must have $q \in \mathcal{R}(A^{\mathsf{T}}A)$ and we conclude that $\mathcal{R}(A^{\mathsf{T}}) \subset \mathcal{R}(A^{\mathsf{T}}A)$.

The second assertion in (1.50) is more immediate. If $Ap = 0$ then $A^{\mathsf{T}}Ap = 0$ so that $\mathcal{N}(A) \subset \mathcal{N}(A^{\mathsf{T}}A)$. Conversely, if $A^{\mathsf{T}}Ap = 0$ then $p^{\mathsf{T}}A^{\mathsf{T}}Ap = \|Ap\|^2 = 0$ and we must have $Ap = 0$. That is, $\mathcal{N}(A^{\mathsf{T}}A) \subset \mathcal{N}(A)$. Combining both facts we conclude that $\mathcal{N}(A) = \mathcal{N}(A^{\mathsf{T}}A)$.

■

Normal equations

One immediate consequence of result (1.50) is that linear systems of equations of the following form:

$$\boxed{A^{\mathsf{T}}Ax = A^{\mathsf{T}}b} \qquad (\textbf{normal equations}) \tag{1.51}$$

are always consistent, i.e., they always have a solution x for *any* vector b. This is because $A^{\mathsf{T}}b$ belongs to $\mathcal{R}(A^{\mathsf{T}})$ and, therefore, also belongs to $\mathcal{R}(A^{\mathsf{T}}A)$. This type of linear system of equations will appear as normal equations when we study least-squares problems later in Chapter 50 – see (50.25); the reason for the designation "normal equations" will be explained there. We can say more about the solution of such equations. For example, when the coefficient matrix $A^{\mathsf{T}}A$, which is always square regardless of the column and row dimensions of the $N \times M$ matrix A, happens to be invertible, then the normal equations (1.51) will have a unique solution given by

$$x = (A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}b \tag{1.52}$$

We explain later in (1.58) that the matrix product $A^{\mathsf{T}}A$ will be invertible when the following two conditions hold: $N \geq M$ *and* A has full rank. In all other cases, the matrix product $A^{\mathsf{T}}A$ will be singular and will, therefore, have a nontrivial nullspace. Let p be any nonzero vector in the nullspace of $A^{\mathsf{T}}A$. We know from (1.50) that this vector also lies in the nullspace of A. Since we know that a solution x always exists for (1.51) then, by adding any such p to x, we obtain another solution. This is because:

$$\begin{aligned} A^{\mathsf{T}}A(x+p) &= A^{\mathsf{T}}Ax + A^{\mathsf{T}}Ap \\ &= A^{\mathsf{T}}Ax + 0 \\ &= A^{\mathsf{T}}Ax \\ &= A^{\mathsf{T}}b \end{aligned} \tag{1.53}$$

Knowing that there exist infinitely many vectors in $\mathcal{N}(A^{\mathsf{T}}A)$, e.g., any scaled multiple of p belongs to the same nullspace, we conclude that when $A^{\mathsf{T}}A$ is singular, there will exist infinitely many solutions to the normal equations (1.51). We therefore find that the normal equations (1.51) either have a unique solution (when $A^{\mathsf{T}}A$ is invertible) or infinitely many solutions (when $A^{\mathsf{T}}A$ is singular).

We can be more explicit about the latter case and verify that, when infinitely many solutions exist, they all differ by a vector in the nullspace of A. Indeed, assume $A^{\mathsf{T}}A$ is singular and let x_1 and x_2 denote two solutions to the normal equations (1.51). Then,

$$A^{\mathsf{T}}Ax_1 = A^{\mathsf{T}}b, \quad A^{\mathsf{T}}Ax_2 = A^{\mathsf{T}}b \tag{1.54}$$

Subtracting these two equalities we find that

$$A^{\mathsf{T}}A(x_1 - x_2) = 0 \tag{1.55}$$

which means that the difference $x_1 - x_2$ belongs to the nullspace of $A^{\mathsf{T}}A$ or, equivalently, to the nullspace of A in view of (1.50), namely,

$$x_1 - x_2 \in \mathcal{N}(A) \tag{1.56}$$

as claimed. We collect the results in the following statement for ease of reference.

Lemma 1.1. (Solution of normal equations) *Consider the normal system of equations $A^{\mathsf{T}}Ax = A^{\mathsf{T}}b$, where $A \in \mathbb{R}^{N\times M}$, $b \in \mathbb{R}^N$, and $x \in \mathbb{R}^M$. The following facts hold:*

(a) *A solution x always exists.*
(b) *The solution x is unique when $A^{\mathsf{T}}A$ is invertible (i.e., when $N \geq M$ and A has full rank). In this case, the solution is given by expression (1.52).*
(c) *There exist infinitely many solutions x when $A^{\mathsf{T}}A$ is singular.*
(d) *Under (c), any two solutions x_1 and x_2 will differ by a vector in $\mathcal{N}(A)$, i.e., (1.56) holds.*

The next result clarifies when the matrix product $A^{\mathsf{T}}A$ is invertible. Note in particular that the matrix $A^{\mathsf{T}}A$ is symmetric and nonnegative-definite; the latter property is because, for any nonzero x, it holds that

$$x^{\mathsf{T}}A^{\mathsf{T}}Ax \;=\; \|Ax\|^2 \;\geq 0 \tag{1.57}$$

Thus, let A be $N \times M$, with $N \geq M$ (i.e., A is a "tall" or square matrix). Then,

$$\boxed{A \text{ has full rank} \iff A^{\mathsf{T}}A \text{ is positive-definite}} \tag{1.58}$$

That is, every tall full rank matrix is such that the square matrix $A^{\mathsf{T}}A$ is invertible (actually, positive-definite).

Proof of (1.58): Assume first that A has full rank. This means that all columns of A are linearly independent, which in turn means that $Ax \neq 0$ for any nonzero x. Consequently, it holds that $\|Ax\|^2 > 0$, which is equivalent to $x^{\mathsf{T}}A^{\mathsf{T}}Ax > 0$ for any