# Chapter 3

# Matrices

## 3.1 Matrices

**Definition 3.1 (Matrix)** *A matrix $A$ is a rectangular array of $m \times n$ real numbers $\{a_{ij}\}$ written as*

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

*The array has $m$ rows and $n$ columns.*

If $A$ and $B$ are both $m \times n$ matrices, then $C = A + B$ is an $m \times n$ matrix of real numbers $c_{ij} = a_{ij} + b_{ij}$. If $A$ is an $m \times n$ matrix and $\alpha \in \Re^1$, then $\alpha A$ is an $m \times n$ matrix with elements $\alpha a_{ij}$. Also, $\alpha A = A\alpha$, and $\alpha(A + B) = \alpha A + \alpha B$.

A matrix $A$ can be used to define a function $A : \Re^n \rightarrow \Re^m$.

**Definition 3.2 (Linear function associated with a matrix)** *Given a vector $x \in \Re^n$ with coordinates $(\alpha_1, \ldots, \alpha_n)$ relative to a fixed basis, define $u = Ax$ to be the vector in $\Re^m$ with coordinates given for $i = 1, \ldots, m$,*

$$\beta_i = \sum_{j=1}^{n} a_{ij}\alpha_j = (A_i, x) \tag{3.1}$$

*where $A_i$ is the vector consisting of the $i$th* row *of $A$. The vector $u$ has coordinates $(\beta_1, \ldots, \beta_m)$.*

**Theorem 3.1** *For all $x, y \in \Re^n$, scalars $\alpha$ and $\beta$, and any two $n \times m$ matrices $A, B$, the function (3.1) has the following two properties:*

$$\begin{aligned} A(\alpha x + \beta y) &= \alpha A x + \beta A y \\ (\alpha A + \beta B)(x) &= \alpha A x + \beta B y \end{aligned} \tag{3.2}$$

*In light of (3.2) we will call $A$ a* linear *function, and if $m = n$, $A$ is a* linear transformation*.*

*Proof:* Straightforward application of the definition.

**Definition 3.3** *A square matrix has the same number of rows and columns. It transforms from $\Re^n \to \Re^n$, but is not necessarily onto.*

**Connection between matrices and linear transformations.** In light of (3.1), every square matrix corresponds to a linear transformation from $\Re^n \to \Re^n$. This justifies using the same symbol for both a linear transformation and for its corresponding matrix. The matrix representation of a linear transformation depends on the basis.

*Example*. In $\Re^3$, consider the linear transformation defined as follows. For the canonical basis $\{e_1, e_2, e_3\}$,

$$Ae_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \; Ae_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \; Ae_3 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

Relative to this basis, the matrix of $A$ is

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 1 & 0 \\ 3 & 1 & -1 \end{pmatrix}$$

If we change to a different basis, say

$$x_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \; x_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \; x_3 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

then the matrix $A$ changes. For this example, since $x_1 = e_1 + e_2 + e_3$, $x_2 = e_1 - e_2$ and $x_3 = e_1 + e_2 - 2e_3$, we can compute

$$Ax_1 = A(e_1 + e_2 + e_3) = Ae_1 + Ae_2 + Ae_3 = \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix} = a_1$$

and this is the first column of $A$ relative to this basis. The full matrix is

$$\begin{pmatrix} 2 & 1 & -1 \\ 3 & 1 & 3 \\ 3 & 2 & 2 \end{pmatrix}$$

While the matrix representation of a linear transformation depends on the basis chosen, the two special linear transformations $0$ and $I$ are always identified with the null matrix and the identity matrix, respectively.

**Definition 3.4 (Matrix Multiplication)** *Given two matrices $A : m \times n$ and $B : n \times p$, the matrix product $C = AB$ (in this order) is an $m \times p$ matrix with typical elements*

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

*The matrix product is defined only when the number of columns of $A$ equals the number of rows of $B$. The matrix product is connected to (3.1) using*

**Theorem 3.2** *Suppose $A : m \times n$ and $B : n \times p$, and $C = AB$. Then $C$ defines a linear function with domain $\Re^p$ and range $\Re^m$. For any $x \in \Re^p$, we have*

$$Cx = (AB)x = A(Bx)$$

This theorem shows that *matrix multiplication is the same as function composition*, so $Cx$ is the same as applying the function $A$ to the vector $Bx$.

**Definition 3.5 (Transpose of a matrix)** *The transpose $A'$ of a $n \times m$ matrix $A = (a_{ij})$ is an $m \times n$ array whose entries are given by $(a_{ji})$.*

**Definition 3.6 (Rank of a matrix)** *An $m \times n$ matrix $A$ is a transformation from $\Re^n \to \Re^m$. The* rank *$\rho(A)$ of $A$ is the dimension of the vector subspace $\{u | u = Ax, x \in \Re^n\} \subset \Re^m$.*

**Theorem 3.3** *The rank of $A$ is equal to the number of linearly independent columns of $A$.*

**Theorem 3.4** *$\rho(A) = \rho(A')$, or the number of linearly independent columns of $A$ is the same of the number of linearly independent rows.*

**Definition 3.7** *The $n \times n$ matrix $A$ is nonsingular if $\rho(A) = n$. If $\rho(A) < n$, then $A$ is singular.*

**Definition 3.8 (Inverse)** *The inverse of a square matrix $A$ is a square matrix of the same size $A^{-1}$ such that $AA^{-1} = A^{-1}A = I$.*

**Theorem 3.5** *$A^{-1}$ exists and is unique if and only if $A$ is nonsingular.*

**Definition 3.9 (Trace)** *trace$(A) =$ tr$(A) = \sum a_{ii}$.*

From this definition, it is easy to show the following:

$$\begin{aligned} \text{tr}(A+B) &= \text{tr}(A) + \text{tr}(B), \text{ trace is } \textit{additive} \\ \text{tr}(ABC) &= \text{tr}(BCA) = \text{tr}(CAB), \text{ trace is } \textit{cyclic} \end{aligned}$$

and, if $B$ is nonsingular,

$$\begin{aligned} \text{tr}(A) &= \text{tr}(ABB^{-1}) \\ &= \text{tr}(BAB^{-1}) \\ &= \text{tr}(B^{-1}AB) \end{aligned}$$

**Definition 3.10 (Symmetric)** *$A$ is symmetric if $a_{ij} = a_{ji},$ for all $i, j$.*

**Definition 3.11 (Diagonal)** *$A$ is diagonal if $a_{ij} = 0, i \neq j$.*

**Definition 3.12 (Determinant)** *The determinant $\det(A)$ is given by*

$$\begin{aligned} \det(A) &= \sum (-1)^{f(i_1,\ldots,i_m)} a_{1i_1} a_{2i_2}, \ldots, a_{mi_m} \\ &= \sum (-1)^{f(i_1,\ldots,i_m)} a_{i_1 1} a_{i_2 2}, \ldots, a_{i_m m} \end{aligned}$$

*where the sum is over all permutations $(i_1, \ldots, i_m)$ of $(1, \ldots, m)$, and $f(i_1, \ldots, i_m)$ is the number of transpositions needed to change $(1, 2, \ldots, n)$ into $(i_1, \ldots, i_m)$. The determinant of a diagonal matrix is the product of its diagonal elements.*

The determinant is a polynomial of degree $n$ for an $n \times n$ matrix.

**Theorem 3.6** *If $A$ is singular, then $\det(A) = 0$.*

## 3.2 Eigenvectors and Eigenvalues

**Definition 3.13 (Eigenvectors and eigenvalues)** *An* eigenvector *of a square matrix $A$ is any nonzero vector $x$ such that $Ax = \lambda x, \lambda \in \Re$. $\lambda$ is an* eigenvalue *of $A$.*

From the definition, if $x$ is an eigenvector of $A$, then $Ax = \lambda x = \lambda I x$ and

$$Ax - \lambda I x \quad = \quad 0 \tag{3.3}$$

**Theorem 3.7** *$\lambda$ is an eigenvalue of $A$ if and only if (3.3) is satisfied. Equivalently, $\lambda$ is an eigenvalue if it is a solution to*

$$\det(A - \lambda I) \quad = \quad 0$$

This last equation provides a prescriptions for finding eigenvalues, as a solution to $\det(A - \lambda I) = 0$. The determinant of an $n \times n$ matrix is a polynomial of degree $n$, showing that the number of eigenvalues must be $n$. The eigenvalues need not be unique or nonzero.

In addition, for any nonzero scalar $c$, $A(cx) = cAx = (c\lambda)x$, so that $cx$ is also an eigenvector with associated eigenvalue $c\lambda$. To resolve this particular source of indeterminacy, we will always require that all eigenvectors will be normalized to have unit length, $\parallel x \parallel = 1$. However, some indeterminacy in the eigenvectors still remains, as shown in the next theorem.

**Theorem 3.8** *If $x_1$ and $x_2$ are eigenvectors with the same eigenvalue, then any non-zero linear combination of $x_1$ and $x_2$ is also an eigenvector with the same eigenvalue.*

*Proof.* If $Ax_i = \lambda x_i$ for $i = 1, 2$, then $A(\alpha_1 x_1 + \alpha_2 x_2) = \alpha_1 A x_1 + \alpha_2 A x_2 = \alpha_1 \lambda x_1 + \alpha_2 \lambda x_2 = \lambda(\alpha_1 x_1 + \alpha_2 x_2)$ as required.

According to Theorem 3.8, the set of vectors corresponding to the same eigenvalue form a vector subspace. The dimension of this subspace can be as large as the *multiplicity* of the eigenvalue $\lambda$, but the dimension of this subspace can be less. For example, the matrix

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix}$$

has $\det(A-\lambda I) = (1-\lambda)^3$, so it has one eigenvalue equal to one with multiplicity three. The equations $Ax = 1x$ have only solutions of the form $x = (a, 0, 0)'$ for any $a$, and so they form a vector subspace of dimension one, less than the multiplicity of $\lambda$.

We can always find an orthonormal basis for this subspace, so the eigenvectors corresponding to the fixed eigenvalue can be taken to be orthogonal.

Eigenvalues can be real or complex. But if $A$ is symmetric, then the eigenvalues must be real.

**Theorem 3.9** *The eigenvalues of a real, symmetric matrix are real.*

*Proof.* In the proof, we will allow both the eigenvalue and the eigenvector to be complex. Suppose we have eigenvector $x + iy$ with eigenvalue $\lambda_1 + \lambda_2 i$, so

$$A(x + iy) = (\lambda_1 + \lambda_2 i)(x + iy)$$

and thus

$$Ax + Ayi = (\lambda_1 x - \lambda_2 y) + i(\lambda_1 y + \lambda_2 x)$$

Evaluating real and imaginary parts, we get:

$$Ax = \lambda_1 x - \lambda_2 y \tag{3.4}$$

$$Ay = \lambda_1 y + \lambda_2 x \tag{3.5}$$

Multiply (3.4) on the left by $y'$ and (3.5) on the left by $x'$. Since $x'Ay = y'A'x$, we equate the right sides of these modified equations to get

$$\lambda_1 y'x - \lambda_2 y'y = \lambda_1 x'y + \lambda_2 x'x$$

and

$$\lambda_2(x'x + y'y) = 0$$

This last equation holds in general only if $\lambda_2 = 0$ and so the eigenvalue must be real, from which it follows that the eigenvector is real as well.

Here are some more properties of the eigenvalues and eigenvectors of an $n \times n$ real symmetric matrix $A$, all of these are easily demonstrated with the spectral theorem, to be proved shortly.

- $\det(A) = \prod_{i=1}^{n} \lambda_i$.
- If $A$ is nonsingular, then the eigenvalues of $A$ are all nonzero ($\lambda_i \neq 0, i = 1, \ldots, n$) and the eigenvalues of $A^{-1}$ are $\lambda_1^{-1}, \ldots, \lambda_n^{-1}$.

- The eigenvalues of $A'$ are the same as the eigenvalues of $A$, since $0 = \det(A - \lambda I)$ only if $0 = \det(A' - \lambda I)$.
- $\text{tr}(A) = \sum \lambda_i,$ and $\text{tr}(A^r) = \sum \lambda_i^r$. If $A$ is nonsingular, $\text{tr}(A^{-1}) = \sum \lambda_i^{-1}$.

**Definition 3.14 (Block Diagonal Matrix)** *A block diagonal matrix has nonzero diagonal blocks and zero off-diagonal blocks.*

If $A$ is block diagonal, then $\lambda$ is an eigenvalue of $A$ if it is an eigenvalue of one of the blocks.

**Definition 3.15 (Orthogonal Matrix)** *An $n \times n$ matrix $\Gamma$ is orthogonal if $\Gamma'\Gamma = \Gamma\Gamma' = I$.*

**Theorem 3.10** *The product of two orthogonal matrices is orthogonal.*

*Proof.* If $\Gamma_1$ and $\Gamma_2$ are orthogonal matrices, then $(\Gamma_1\Gamma_2)(\Gamma_1\Gamma_2)' = \Gamma_1\Gamma_2\Gamma_2{}'\Gamma_1{}' = I$.

# 3.3 Matrix Decompositions

Working with matrices both in theoretical results and in numerical computations is generally made easier by decomposing the matrix into a product of matrices, each of which is relatively easy to work with, and has some special structure of interest. We pursue several decompositions in this section.

## 3.3.1 Spectral Decomposition

The spectral decomposition allows the representation of any symmetric matrix in terms of an orthogonal matrix and a diagonal matrix of eigenvalues. We begin with a theorem that shows that how to introduce zeroes into a symmetric matrix.

**Theorem 3.11** *Let $A$ by a symmetric matrix and suppose that it has an eigenvalue $\lambda_n$. Then there exists an orthogonal matrix $\Gamma$ such that:*

$$\Gamma' A \Gamma = \begin{pmatrix} \lambda_n & 0' \\ 0 & C \end{pmatrix}$$

*where $C$ is an $(n-1) \times (n-1)$ symmetric matrix such that if $\lambda$ is another eigenvalue of $A$, then $\lambda$ is also an eigenvalue of $C$.*

*Proof.* Let $x_n$ be a normalized (unit length) eigenvector corresponding to $\lambda_n$ and extend $x_n$ so that $\{x_n, u_1, \ldots, u_{n-1}\}$ is an orthonormal basis for $\Re^n$. Let $U = \{u_1, \ldots, u_{n-1}\}$. Then

$$\begin{aligned}
\Gamma' A \Gamma = \begin{pmatrix} x_n{}' \\ U' \end{pmatrix} A(x_n\ U) &= \begin{pmatrix} x_n{}' A x_n & x_n{}' A U \\ U' A x_n & U' A U \end{pmatrix} \\
&= \begin{pmatrix} x_n{}'(\lambda_n x_n) & \lambda_n x_n{}' U \\ U'(\lambda x_n) & U' A U \end{pmatrix} \\
&= \begin{pmatrix} \lambda_n x_n{}' x_n & \lambda x_n{}' U \\ \lambda_n U' x_n & U' A U \end{pmatrix} \\
&= \begin{pmatrix} \lambda_n & 0 \\ 0 & U' A U \end{pmatrix}
\end{aligned}$$

If $\lambda$ is an eigenvalue of $A$, then we can find a corresponding eigenvector $z$. If $\lambda = \lambda_n$, meaning that the eigenvector $\lambda_n$ has multiplicity greater than one, we take $z$ to be orthogonal to $x_n$, as can be justified using Theorem 3.8. In either case, by definition, $Az = \lambda z,$ or $(A - \lambda I)z = 0$, or

$$0 = \Gamma\Gamma'(A - \lambda I)\Gamma\Gamma' z = \Gamma(\Gamma' A \Gamma - \lambda I)(\Gamma' z) = (\Gamma' A \Gamma - \lambda I)(\Gamma' z)$$

Since $\Gamma' A \Gamma$ is symmetric, we get that $\lambda$ is an eigenvalue of $\Gamma' A \Gamma$ with eigenvector $\Gamma' z$ as required.

If we continue the process of the last theorem, now working on the matrix $(U' A U)$ rather than $A$ we get:

**Theorem 3.12 (Spectral theorem)** *Let $A$ be a real symmetric $n \times n$ matrix. Then there exists an orthogonal matrix $\Gamma$ and a diagonal matrix $D$ such that $A = \Gamma D \Gamma'$. An alternative statement of the spectral theorem is:*

$$A = \sum_{i=1}^{m} \lambda_i \gamma_i \gamma_i{}'$$

*with $\gamma_i$ the $i$-th column of $\Gamma$, and $\lambda_i$ the $i$-th nonzero diagonal element of $D$, and $m$ is the number of nonzero diagonal elements of $D$.*

An immediate consequence of the spectral theorem is that the columns of $\Gamma$ are the normalized eigenvectors of $A$ and the diagonals of $D$ are the eigenvalues of

$A$, since $A = \Gamma D \Gamma'$ implies that $A\Gamma = \Gamma D$, and the columns of $\Gamma$ are normalized eigenvectors and the diagonals of $D$ are the eigenvalues.
*Proof of spectral theorem.* The proof uses induction on the order of $A$. If $n = 1$, the result holds trivially, and for $n = 2$, Theorem 3.11 is equivalent to the spectral theorem. Consider general $n$. By Theorem 3.11, there is a matrix $\Gamma_n$ such that

$$\Gamma_n A \Gamma_n{}' = \begin{pmatrix} \lambda_n & 0 \\ 0 & C \end{pmatrix}$$

where $C$ is $n-1 \times n-1$. According to the induction argument, there are matrices $\Gamma_{n-1}$ and $D_{n-1}$ with $C = \Gamma_{n-1} D_{n-1} \Gamma_{n-1}{}'$. Let

$$\Gamma = \Gamma_n \begin{pmatrix} 1 & 0 \\ 0 & \Gamma_{n-1} \end{pmatrix}$$

Then it is easily shown that $\Gamma$ is orthogonal and $\Gamma' A \Gamma = D$, a diagonal matrix.

Geometrically, the theorem says that any symmetric transformation can be accomplished by rotating, weighting, and rotating back. Some additional consequences are as follows.

**Theorem 3.13** *The rank of a symmetric matrix is the number of nonzero eigenvalues.*

*Proof.* The result follows since multiplication by a nonsingular matrix does not alter the rank of a matrix (from a homework problem).

Suppose $\nu$ of the eigenvalues of $A$ are zero. Then the columns of $\Gamma$ corresponding to these eigenvalues are an orthonormal basis for $\mathrm{N}(A)$ and the columns of $\Gamma$ corresponding to nonzero eigenvalues are an orthonormal basis for $\mathcal{R}(A)$.

**Definition 3.16** *A symmetric matrix $A$ is positive definite if $x'Ax > 0$ for all $x \neq 0 \in \Re^n$. It is positive semidefinite if $x'Ax \geq 0$.*

This definition corresponds to the definition given for linear transformations, given $\Re^n$and the usual inner product.

**Theorem 3.14** *The eigenvalues of a positive definite matrix are all positive, and of a positive semi-definite matrix are all nonnegative.*

*Proof.* $A$ is positive definite if and only if $x'Ax > 0$ or if and only if $x'\Gamma'\Lambda\Gamma x > 0$, which holds if and only if $y'\Lambda y > 0$ for all $y = \Gamma x$.

**Theorem 3.15** *If $A$ is positive semi-definite, then there exists $B$: $n \times n$ such that $B$ is positive semi-definite, $\rho(A) = \rho(B)$ and $A = BB'$.*

*Proof*. For any diagonal matrix $C$ with nonnegative elements, write $C^{1/2}$ to be the diagonal matrix with elements $(c_{ii}^{1/2})$, so $C^{1/2}C^{1/2} = C$. Then

$$A = \Gamma\Lambda\Gamma' = \Gamma\Lambda^{1/2}\Lambda^{1/2}\Gamma' = \Gamma\Lambda^{1/2}\Gamma'\Gamma\Lambda^{1/2}\Gamma' = BB'.$$

**Theorem 3.16** *If $A$ is positive definite and $A = \Gamma\Lambda\Gamma'$, then $A^{-1} = \Gamma\Lambda^{-1}\Gamma'$.*

**Theorem 3.17** *If $Z$ is $m \times n$, then $Z'Z$ and $ZZ'$ are positive semi-definite.*

Suppose that $A = \Gamma\Lambda\Gamma'$, with the diagonal elements of $\Lambda$ ordered so that $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Let $\Gamma$ have corresponding columns $(\gamma_1, \ldots, \gamma_n)$. Then, for all vectors $xin\Re^n$ such that $\| x \| = 1$, application of the spectral theorem shows that

$$\begin{aligned} \min_x(x'Ax) &= \lambda_1 \\ \max_x(x'Ax) &= \lambda_n \end{aligned}$$

If the $\lambda_1 = \cdots = \lambda_k$, then the minimum is achieved for any vector in $\mathcal{R}(\gamma_1, \ldots, \gamma_k)$. If $\lambda_{n-j} = \cdots = \lambda_n$, then the maximum is achieved for any $x \in \mathcal{R}(\lambda_{n-j}, \ldots, \lambda_n)$.

### 3.3.2 Singular Value Decomposition

We now extend the spectral theorem to general $n \times p$ matrices.

**Theorem 3.18 (The singular value decomposition)** *Let $A$ be an $n \times p$ matrix. There there exist orthogonal matrices $U$ and $V$ such that:*

$$V'AU = \begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}$$

*where $\Delta$ = diag$(\delta_1, \ldots, \delta_p)$ are the* singular values *of $A$.*

This is not standard notation in the numerical analysis literature; $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_p)$ is commonly used in place of $\Delta$, but this would be confusing in most statistical applications that reserve $\Sigma$ and $\sigma$ for variances or standard deviations.

Before proving the singular value decomposition, we consider some of its implications, since they may make the proof clearer. The basic result can be rewritten as

$$\begin{aligned} A &= VDU' = (V_1\ V_2)\begin{pmatrix} \Delta & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} U_1{}' \\ U_2{}' \end{pmatrix} \\ &= V_1\Delta U_1{}' + V_2 0 U_2{}'. \end{aligned}$$

where 0 is a matrix of zeroes of appropriate dimension. Thus, the range of $A, \mathcal{R}(A) = \{z|z = Ax = V_1\Delta U_1{}'x,\ x \in \Re^n\} = \mathcal{R}(V_1)$, and the columns of $V_1$ span the same space as the columns of $A$. Since $V$ is an orthonormal basis, the columns of $V_1$ provide an orthonormal basis for $\mathcal{R}(A)$. In other words, *the columns of $V_1$ provide an orthonormal basis for the column space of $A$.*

Similarly, $R(A') = R(U_1)$. *the columns of $U_1$ provide an orthonormal basis for the row space of $A$*, which is the same as the column space of $A'$.

The null space $\text{N}(A) = \{z|Az = 0\} = \mathcal{R}(V_2) = (\mathcal{R}(V_1))^{\perp}$, so the remaining columns of $V$ are an orthonormal basis for the null space of $A$. A similar result using $U$ is found for $A'$.

*Proof of singular value decomposition.* $A'A$ is $p \times p$ and positive semi-definite. We can apply the spectral theorem to write

$$A'A = \Gamma \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \Gamma'$$

or

$$\Gamma' A' A \Gamma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}. \tag{3.6}$$

Now, let $\gamma_i,\ i = 1, \ldots, q \leq p$ be the columns of $\Gamma$ corresponding to nonzero $\lambda_i$, and let $\Gamma = (\Gamma_1, \Gamma_2)$, with $\Gamma_1 = (\gamma_1, \ldots \gamma_q)$. Define $V_i = A\gamma_i/\sqrt{\lambda_i} \in \Re^n, i = 1, 2, \ldots, q$. The $\{V_i\}$ are orthonormal, and can be extended to an orthonormal basis for $\Re^n$ if $n > p$. Denote this orthonormal basis by $V = (V_1, V_2)$. Then:

$$V'A\Gamma = \begin{pmatrix} V_1{}' \\ V_2{}' \end{pmatrix} A(\Gamma_1, \Gamma_2) = \begin{pmatrix} V_1{}'A\Gamma_1 & V_1{}'A\Gamma_2 \\ V_2{}'A\Gamma_1 & V_2{}'A\Gamma_2 \end{pmatrix}.$$

Consider each of the four submatrices separately.

1. By construction, $V_1 = A\Gamma_1 D^{-1/2}$, so ${V_1}'A\Gamma_1 = D^{-1/2}{\Gamma_1}'A'A\Gamma_1 = D^{-1/2}D = D^{1/2}$.

2. Since by equation (3.6),

$$\begin{aligned}
\Gamma' A' A \Gamma &= \begin{pmatrix} {\Gamma_1}' \\ {\Gamma_2}' \end{pmatrix} A'A(\Gamma_1 \ \Gamma_2) \\
&= \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} \\
&= \begin{pmatrix} {\Gamma_1}'A'A\Gamma_1 & {\Gamma_1}'A'A\Gamma_2 \\ {\Gamma_2}'A'A\Gamma_1 & {\Gamma_2}'A'A\Gamma_2 \end{pmatrix}
\end{aligned}$$

   Then by equation(3.6), we must have $A\Gamma_2 = 0,$ and ${V_1}'A\Gamma_2 = {V_2}'A\Gamma_2 = 0$.

3. Finally, $V_2$ is orthogonal to $V_1 = A\Gamma_1 D^{-1/2}$ and ${V_2}'V_1 = {V_2}'A\Gamma_1 D^{-1/2} = 0$ so ${V_2}'A\Gamma_1 = 0$.

Setting $\Delta = D^{-1/2}$, and $\Gamma = U$, and substituting the results into the statement of the theorem proves the result.

In the proof, $U_1$ is any orthonormal basis for the space spanned by the eigenvectors with nonzero eigenvalues of $A'A$. Once $U_1$ is chosen, $V_1$ is determined. $V_2$ and $U_2$ are any orthonormal basis completions and are thus orthonormal basis for N$(A)$ and $N(A')$, respectively. Also, the spectral theorem gives the decomposition of $A' = V'\Delta U$.

**Definition 3.17 (Singular value factorization)** *The singular value factorization of $A = V_1 \Delta {U_1}'$, where $V_1$ is $n \times q$, $\Delta$ is $p \times p$ and $U_1$ is $p \times q$. If $A$ has full column rank, then the spectral factorization is $V_1 \Delta U'$. The columns of $V_1$ are the left singular vectors (eigenvectors of $AA'$) and the columns of $U$ are the right singular vectors (eigenvectors of $A'A$).*

*Reference*. Mandel (1982), *American Statistician*, 15-24, discusses the use of the singular value decomposition in regression analysis.

### 3.3.3 QR Factorization

The next factorization expresses any $n \times p$ matrix $A$ as a function of an orthogonal matrix $Q$ and an upper triangular matrix $R$.

**Theorem 3.19 (QR factorization)** *Let $A$ have linearly independent columns. Then, up to sign changes, $A$ can be written uniquely in the form $A = Q_1R$, where $Q_1$ has orthonormal columns and $R$ is upper triangular.*

*Proof* (Homework problem)

**Definition 3.18 (QR decomposition with column reordering)** *Suppose $A = (A_1, A_2)$ is an $n \times p$ rank $r \leq p$ matrix, so that $A_1$ is $n \times r$ rank $r$. Any $A$ can be put in this form by reordering the columns so that the first $r$ are linearly independent. Then the $QR$ decomposition of $A$ is given by:*

$$A = (Q_1\ Q_2)\begin{pmatrix} R & 0 \\ 0 & 0 \end{pmatrix}$$

*where $Q_2$ is any basis for $N(A)$.*

*Reference*. Ansley, C. F. (1985), "Quick Proofs of some regression theorems via the QR algorithm", *American Statistician*, 39, 55-59. (A copy of this article can be obtained from www.jstor.org.)

### 3.3.4 Projections

Since matrices are linear transformations,

**Definition 3.19** *$P$ is a projection (matrix) if $P = P^2$.*

**Theorem 3.20** *If $P$ is a projection such that $P = P'$, then*

1. *The eigenvalues of P are 0 or 1*
2. *$\rho(P) =$ tr$(P)$*
3. *$P$ is an orthogonal projection on $M$ and along $M^{\perp}$.*

*Proof*. If $\lambda$ is an eigenvalue of $P$, then for some $x \neq 0$, $Px = \lambda x \Rightarrow P^2x = \lambda Px = \lambda^2 x = Px = \lambda x$ or $\lambda(1-\lambda) = 0$ and $\lambda = 0$ or $1$. Part 2 follows immediately since $\mathcal{R}(P)$ = number of nonzero eigenvalues = tr$(P)$. The proof of part 3 is similar to the proof for transformations given earlier.

**Theorem 3.21** *Symmetric projection matrices are positive semi-definite.*

*Proof.* Immediate, since all the eigenvalues are either zero or one. The only projection that is positive definite, and has all its eigenvalues equal to one, is the identity projection, $I$.

**Theorem 3.22** *Let $X$ by an $n \times p$ matrix of rank $p$. Then $P = X(X'X)^{-1}X'$ is the orthogonal projection on $\mathcal{R}(X)$ (along $\mathcal{R}(X)^{\perp}$). If X is $n \times 1$, then $P = xx'/(x'x)$.*

*Proof.* First, $(X'X)^{-1}$ exists (why?). Since

$$(X(X'X)^{-1}X')^2 = X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X'$$

and $(X(X'X)^{-1}X')' = X(X'X)^{-1}X'$, $P$ is idempotent and symmetric, so $P$ is a projection on some space. We need only show that if $z \in \mathcal{R}(X)$, then $Pz = z$, and if $z \in (\mathcal{R}(X))^{\perp}$, then $Pz = 0$. If $z \in \mathcal{R}(X)$, we can write $z = Xa$ for some vector $a \neq 0$ because $X$ is of full rank, and so the columns of $X$ form a basis for $\mathcal{R}(X)$. Then $Pz = X(X'X)^{-1}X'Xa = Xa = z$ as required. If $z \in (\mathcal{R}(X))^{\perp}$, write $z = Zb$, where the columns of $Z$ are an orthonormal basis for $(\mathcal{R}(A))^{\perp}$, and thus $Pz = PZb = 0$ because $X'Z = 0$ by construction.

**Theorem 3.23** *Define $P = X(X'X)^{-1}X'$ as above. Then $P = Q_1Q_1{}' = V_1V_1{}'$, where $Q_1$ and $V_1$ are as defined in the $QR$ and $SV$ factorizations, respectively.*

**Theorem 3.24** *$p_{ii}$, the $i$-th diagonal element of an orthogonal projection matrix $P$, is bounded, $0 \leq p_{ii} \leq 1$. If $1 \in \mathcal{R}(X)$, then $1/n \leq p_{ii} \leq 1$. If the number of rows of $X$ exactly equal to $x_i$ is $c$, then $1/n \leq p_{ii} \leq 1/c$.*

*Proof*

1. For all $y, y'Py \leq y'y$. Set $y = e_i$, the $i$-th standard basis vector.

2. Suppose $X = (1 \;\; X_0)$. Get an orthonormal basis for $X$ starting with 1: $X^* = (1/\sqrt{n}, X_0^*)$, and $Px = P_1 + P_0^*$. Since $P_1 = 11'/1'1$ has all elements equal to 1/n, $p_{ii} \geq 1/n$.

3. $p_{ii} = \sum p_{ij}p_{ji} = \sum p_{ij}^2$ (by symmetry). But, if $x_i = x_{i^*}, p_{ii^*}^2 = p_{ii}^2$, so $\sum p_{ij}^2 \geq cp_{ii}^2$, from which $p_{ii} \leq 1/c$.

**Theorem 3.25** *Suppose $X = (X_1, X_2)$, where each of $X_1$ and $X_2$ are of full column rank, but not orthogonal. Then, let $P_j$ be the projection on the columns of $X_j$ (= $X_j({X_j}'X_j)^{-1}{X_j}'$), and $P_{12}$ be the projection on the columns of $X$ (= $X(X'X)^{-1}X'$). Let $X_j^* = (I - P_{3-j})X_j,\ j = 1, 2$. Then:*

$$P_{12} = P_1 + P_2^* \text{ and } P_{12} = P_2 + P_1^*$$

*where*

$$\begin{aligned} P_j^* &= X_j^*(X_j^{*T}X_j)^{-1}X_j^* \\ &= (I - P_{3-j}){X_j}'(X_j(I - P_{3-j})X_j)^{-1}{X_j}'(I - P_{3-j}) \end{aligned}$$

This result tells how to compute the projection on$M$ along $N$ if $M$ and $N$ are not orthogonal. The proof is by direct multiplication.

As a final comment about projection matrices, the central object of interest is the subspace $M$. *The projection matrix does not depend on the basis selected for the space*—it is coordinate free. The projection can be computed starting with any convenient basis.

### 3.3.5 Generalized Inverses

Orthogonal projections are easily constructed both theoretically and numerically by finding an orthonormal basis for $\mathcal{R}(X)$, for example using the singular value decomposition or the QR factorization. We have also seen in Theorem 3.22 that projections can be computed when $X$ is of full column rank and so $X'X$ is invertible.

A different approach to getting projections when $X$ is not of full column rank is through the use of *generalized inverses* or *g-inverses*. Although this approach does not seem to have much to offer from the point of view of this course, it is worth some study if only because it is widely used in other approaches to linear models.

**Definition 3.20 (Generalized inverse)** *Suppose $A$ is an $n \times p$ rank $q$ matrix with $q \leq p \leq n$. Then a generalized inverse $A^-$ of $A$ is a matrix that satisfies*

$$AA^-y = y \tag{3.7}$$

*for all $y \in \Re^n$, and so $A^-$ must be a $p \times n$ matrix.*

The matrix $A$ can be viewed as a linear function that transforms a vector $x \in \Re^p$ to a vector $y \in \Re^n$. The linear function $A^-$ maps in the other direction, from $\Re^n \to \Re^p$. Since for any $x \in \Re^p$, $Ax \in \Re^n$, substitute into (3.7) to get

$$\begin{aligned} AA^-(Ax) &= Ax \\ AA^-Ay &= Ax \\ AA^-A &= A \end{aligned} \tag{3.8}$$

This last result is often given as the definition of a generalized inverse.

Multiplying (3.8) on the left by $A^-$, gives $(A^-A)(A^-A) = A^-A$, which can be summarized by the following theorem.

**Theorem 3.26** *$A^-A$ is a projection, but not necessarily an orthogonal projection.*

We will now construct a generalized inverse. We start with the singular value decomposition of $A = V_1 \Delta U_1{}'$, and substitute into (3.8):

$$\begin{aligned} AA^-A &= A \\ (V_1\Delta U_1{}')A^-(V_1\Delta U_1{}') &= V_1\Delta U_1{}' \end{aligned}$$

This equation will be satisfied if we set

$$A^- = U_1 \Delta^{-1} V_1{}'$$

and so we have produced **a** generalized inverse.

The generalized inverse is not unique. As a homework problem, you should produce a generalized inverse starting with the $QR$ factorization $A = Q_1 R_1$.

The next definition gives a specific generalized inverse that satisfies additional properties.

**Definition 3.21 (Moore-Penrose Generalized Inverse)** *A* Moore-Penrose generalized inverse *of $A$ is an $m \times n$ matrix $A^+$ such that:*

1. $(AA^+)' = AA^+$.
2. $(A^+A)' = A^+A$.
3. $AA^+A = A$ *(so $A^+$ is a g-inverse of $A$)*
4. $A^+AA^+ = A^+$ *(so $A$ is a g-inverse of $A^+$)*

From the definition, it is immediate that $AA^+$ and $A^+A$ are orthogonal projection matrices.

**Theorem 3.27** *Every matrix $A$ has a Moore-Penrose generalized inverse.*

*Proof.* It is easy to show directly that the generalized inverse we produced above based on the singular value decomposition satisfies these conditions.

**Theorem 3.28** *The Moore-Penrose generalized inverse is unique.*

*Proof.* Homework.

The following are examples of Moore Penrose G-inverses.

1. If $A$ is an orthogonal projection, $A^+ = A$.
2. If $A$ is nonsingular, $A^+ = A^{-1}$.
3. If $A = \text{diag}(a_i)$, then $A^+$ has diagonal elements $1/a_i$ if $a_i \neq 0$, and equal to 0 if $a_i = 0$.
4. If $\rho(A : m \times n) = m$, then $A^+ = A'(AA')^{-1}$.
5. If $\rho(A : m \times n) = n$, then $A^+ = (A'A)^{-1}A'$.
6. $\rho(A) = \rho(A^+)$
7. For any $A$, $(A^+)' = (A')^+$.
8. If $A$ is symmetric, then $A^+$ is symmetric.
9. $(A^+)^+ = A$
10. For nonsingular matrices, $(AB)^{-1} = B^{-1}A^{-1}$, but if they are not nonsingular, $(AB)^+ \neq B^+A^+$. Here is an example:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \; B = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

and

$$(AB)^+ = \begin{pmatrix} b_1 \\ 0 \end{pmatrix}^+ = (1/b_1, 0)$$

$$B^+A^+ = \frac{(b_1, b_2)}{b_1^2 + b_2^2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} = \left( \frac{b_1}{b_1^2 + b_2^2}, 0 \right)$$

In two important cases, the equality will hold:

(a) For any $A$, $(A'A)^+ = A^+(A')^+$.

(b) If $\rho(A_{m\times n}) = n$ and $\rho(B_{n\times r}) = n$, then $(AB)^+ = B^+A^+$.

.

**Theorem 3.29** *$P = XX^+$ is the orthogonal projection on $\mathcal{R}(X)$.*

*Proof.* We have shown previously that $P$ is an orthogonal projection, so we need only show that it projects on $\mathcal{R}(X)$. First, since $Pz = XX^+z = Xz^*$ is a linear combination of the columns of $X$, we must have that $\mathcal{R}(P) \subset \mathcal{R}(X)$. We need only show that $\rho(X) = \rho(P)$. Since $P$ is an orthogonal projection, $\rho(P) = \text{tr}(P) = \text{tr}(XX^+) = tr(V_1DU_1'V_1D^{-1}U_1') = tr(I_r) = r =$ number of nonzero singular values of $X = \rho(X)$.

There are lots of generalized inverses. The next theorem shows how to get from one to another.

**Theorem 3.30** *Let $X^-$ be any two generalized inverse of an $n\times p$ matrix $X$. Then there exists an $p \times n$ matrix $C$ such that*

$$X^* = X^- + C - X^-XCXX^- \tag{3.9}$$

*is also a generalized inverse of $X$. In addition, for any generalized inverse $X^*$, there is a $C$ so that (3.9) is satisfied.*

*Proof.* Since $A^-$ is a generalized inverse,

$$X(X^- + C - X^-XCXX^-)X = XX^-X + XCX - XX^-XCXX^-X = X$$

and so $X^* = X^- + C + X^-XCXX^-$ is a generalized inverse. Now suppose that $Z$ is any generalized inverse of $X$, and define $C = Z - X^-$. Since $Z$ is a generalized inverse, $XZX = X$, and

$$\begin{aligned}
X^- + C - X^-XCXX^- &= X^- - (Z - X^-) - X^-X(B - X^-)XX^- \\
&= Z - X^-XZXX^- + X^-XX^-XX^- \\
&= Z - X^-XX^- + X^-XX^- \\
&= Z
\end{aligned}$$

and so we have produced a $C$ corresponding to the generalized inverse $Z$.

## 3.4 Solutions to systems of linear equations

Consider the matrix equation $X_{n\times p}\beta_{p\times 1} = y_{n\times 1}$. For a given $X$ and $Y$ does there exist a solution $\beta$ to these equations? Is it unique? If not unique, can we characterize all possible solutions?

1. If $n = p$ and $X$ is nonsingular, the unique solution is $\beta = X^{-1}y$.

2. If $y \in \mathcal{R}(X)$, $y$ can be expressed as a linear combination of the columns of $X$. If $X$ is of full column rank, then the columns of $X$ form a basis for $\mathcal{R}(X)$, and the solution $\beta$ is just the coordinates of $y$ relative to this basis. For any g-inverse $X^-$, we have $XX^-y = y$ for all $y \in R(X)$, and so a solution is given by
$$\beta = X^- y \tag{3.10}$$
If $\mathcal{R}(X) < p$, then the solution is not unique. If $\beta_0$ is *any* solution, for example the solution given by (3.10), then if $z$ is such that $Xz = 0$, then $\beta_0 + z$ is also a solution. By definition $Xz = 0$ if any only if $z \in \mathrm{N}(X)$. The set of solutions is given by $\mathrm{N}(X) + \beta_0$, which is a flat.

3. If $y \notin \mathcal{R}(X)$, then there is no exact solution. This is the usual situation in linear models, and leads to the estimation problem discussed in the next chapter.