Some linear algebra

Recall the convention that, for us, all vectors are column vectors.

1. Symmetric matrices

Let $\boldsymbol{A}$ be a real $n \times n$ matrix. Recall that a complex number λ is an *eigenvalue* of $\boldsymbol{A}$ if there exists a real and nonzero vector $\boldsymbol{x}$—called an eigenvector for λ—such that $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$. Whenever $\boldsymbol{x}$ is an eigenvector for λ, so is $a\boldsymbol{x}$ for every real number a.

The *characteristic polynomial* $\chi_{\boldsymbol{A}}$ of matrix $\boldsymbol{A}$ is the function

$$\chi_{\boldsymbol{A}}(\lambda) := \det(\lambda \boldsymbol{I} - \boldsymbol{A}),$$

defined for all complex numbers λ, where $\boldsymbol{I}$ denotes the $n \times n$ identity matrix. It is not hard to see that a complex number λ is an eigenvalue of $\boldsymbol{A}$ if and only if $\chi_{\boldsymbol{A}}(\lambda) = 0$. We see by direct computation that $\chi_{\boldsymbol{A}}$ is an nth-order polynomial. Therefore, $\boldsymbol{A}$ has precisely n eigenvalues, thanks to the fundamental theorem of algebra. We can write them as $\lambda_1, \ldots, \lambda_n$, or sometimes more precisely as $\lambda_1(\boldsymbol{A}), \ldots, \lambda_n(\boldsymbol{A})$.

1. The spectral theorem. The following important theorem is the starting point of our discussion. It might help to recall that vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbf{R}^n$ are *orthonormal* if $\boldsymbol{x}_i'\boldsymbol{x}_j = 0$ when $i \neq j$ and $\boldsymbol{x}_i'\boldsymbol{x}_i = \|\boldsymbol{x}_i\|^2 = 1$.

Theorem 1. *If $\boldsymbol{A}$ is a real and symmetric $n \times n$ matrix, then $\lambda_1, \ldots, \lambda_n$ are real numbers. Moreover, there exist n orthonormal eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ that correspond respectively to $\lambda_1, \ldots, \lambda_n$.*

I will not prove this result, as it requires developing a good deal of elementary linear algebra that we will not need. Instead, let me state and prove a result that is central for us.

Theorem 2 (The spectral theorem). *Let $\boldsymbol{A}$ denote a symmetric $n \times n$ matrix with real eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding orthonormal eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$. Define $\boldsymbol{D} := \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ to be the diagonal matrix of the λ_i's and $\boldsymbol{P}$ to be the matrix whose columns are $\boldsymbol{v}_1$ though $\boldsymbol{v}_n$ respectively; that is,*

$$\boldsymbol{D} := \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix}, \quad \boldsymbol{P} := (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n).$$

Then $\boldsymbol{P}$ is orthogonal $[\boldsymbol{P}' = \boldsymbol{P}^{-1}]$ *and $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'$.*

Proof. $\boldsymbol{P}$ is orthogonal because the orthonormality of the $\boldsymbol{v}_i$'s implies that

$$\boldsymbol{P}'\boldsymbol{P} = \begin{pmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_n \end{pmatrix}' (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n) = \boldsymbol{I}.$$

Furthermore, because $\boldsymbol{A}\boldsymbol{v}_j = \lambda_j \boldsymbol{v}_j$, it follows that $\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{D}$, which is another way to say that $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$. □

Recall that the *trace* of an $n \times n$ matrix $\boldsymbol{A}$ is the sum $A_{1,1} + \cdots + A_{n,n}$ of its diagonal entries.

Corollary 3. *If A is a real and symmetric $n \times n$ matrix with real eigenvalues $\lambda_1, \ldots, \lambda_n$, then*

$$\mathrm{tr}(\boldsymbol{A}) = \lambda_1 + \cdots + \lambda_n \quad \textit{and} \quad \det(\boldsymbol{A}) = \lambda_1 \times \cdots \times \lambda_n.$$

Proof. Write $\boldsymbol{A}$, in spectral form, as $\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$. Since the determinant of $\boldsymbol{P}^{-1}$ is the reciprocal of that of $\boldsymbol{A}$, it follows that $\det(\boldsymbol{A}) = \det(\boldsymbol{D})$, which is clearly $\lambda_1 \times \cdots \times \lambda_n$. In order to compute the trace of $\boldsymbol{A}$ we compute directly also:

$$\begin{aligned}\mathrm{tr}(\boldsymbol{A}) &= \sum_{i=1}^n \sum_{j=1}^n P_{i,j} \left(\boldsymbol{D}\boldsymbol{P}^{-1}\right)_{i,j} = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n P_{i,j} D_{i,k} P^{-1}_{j,k} \\ &= \sum_{i=1}^n \sum_{k=1}^n \left(\boldsymbol{P}\boldsymbol{P}^{-1}\right)_{i,k} D_{i,k} = \sum_{i=1}^n D_{i,i} = \mathrm{tr}(\boldsymbol{D}),\end{aligned}$$

which is $\lambda_1 + \cdots + \lambda_n$. □

2. The square-root matrix. Let $\boldsymbol{A}$ continue to denote a real and symmetric $n \times n$ matrix.

Proposition 4. *There exists a complex and symmetric $n \times n$ matrix $\boldsymbol{B}$—called the* square root *of $\boldsymbol{A}$ and written as $\boldsymbol{A}^{1/2}$ or even sometimes as $\sqrt{\boldsymbol{A}}$—such that $\boldsymbol{A} = \boldsymbol{B}^2 := \boldsymbol{B}\boldsymbol{B}$.*

The proof of Proposition 4 is more important than its statement. So let us prove this result.

Proof. Apply the spectral theorem and write $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$. Since $\boldsymbol{D}$ is a diagonal matrix, its square root can be defined unambiguously as the following complex-valued $n \times n$ diagonal matrix:

$$\boldsymbol{D}^{1/2} := \begin{pmatrix} \lambda_1^{1/2} & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2^{1/2} & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3^{1/2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n^{1/2} \end{pmatrix}.$$

Define $\boldsymbol{B} := \boldsymbol{P}\boldsymbol{D}^{1/2}\boldsymbol{P}^{-1}$, and note that

$$\boldsymbol{B}^2 = \boldsymbol{P}\boldsymbol{D}^{1/2}\boldsymbol{P}^{-1}\boldsymbol{P}\boldsymbol{D}^{1/2}\boldsymbol{P}^{-1} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1} = \boldsymbol{A},$$

since $\boldsymbol{P}^{-1}\boldsymbol{P} = \boldsymbol{I}$ and $(\boldsymbol{D}^{1/2})^2 = \boldsymbol{D}$. □

2. Positive-semidefinite matrices

Recall that an $n \times n$ matrix $\boldsymbol{A}$ is *positive semidefinite* if it is symmetric and

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0 \qquad \text{for all } \boldsymbol{x} \in \mathbf{R}^n.$$

Recall that $\boldsymbol{A}$ is *positive definite* if it is symmetric and

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0 \qquad \text{for all nonzero } \boldsymbol{x} \in \mathbf{R}^n.$$

Theorem 5. *A symmetric matrix $\boldsymbol{A}$ is positive semidefinite if and only if all of its eigenvalues are ≥ 0. $\boldsymbol{A}$ is positive definite if and only if all of its eigenvalues are > 0. In the latter case, $\boldsymbol{A}$ is also nonsingular.*

The following is a ready consequence.

Corollary 6. *All of the eigenvalues of a variance-covariance matrix are always ≥ 0.*

Now let us establish the theorem.

Proof of Theorem 5. Suppose $\boldsymbol{A}$ is positive semidefinite, and let λ denote one of its eignenvalues, together with corresponding eigenvector $\boldsymbol{x}$. Since $0 \le \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \lambda\|\boldsymbol{x}\|^2$ and $\|\boldsymbol{x}\| > 0$, it follows that $\lambda \ge 0$. This proves that all of the eigenvalues of $\boldsymbol{A}$ are nonnegative. If $\boldsymbol{A}$ is positive definite, then the same argument shows that all of its eigenvalues are > 0. Because $\det(\boldsymbol{A})$ is the product of all n eigenvalues of $\boldsymbol{A}$ (Corollary 3), it follows that $\det(\boldsymbol{A}) > 0$, whence $\boldsymbol{A}$ is nonsingular.

This proves slightly more than half of the proposition. Now let us suppose that all eigenvalues of $\boldsymbol{A}$ are ≥ 0. We write $\boldsymbol{A}$ in spectral form $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'$, and observe that $\boldsymbol{D}$ is a diagonal matrix of nonnegative numbers. By virute of its construction. $\boldsymbol{A}^{1/2} = \boldsymbol{P}\boldsymbol{D}^{1/2}\boldsymbol{P}'$, and hence for all $\boldsymbol{x} \in \mathbf{R}^n$,

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \left(\boldsymbol{D}^{1/2}\boldsymbol{P}\boldsymbol{x}\right)'\left(\boldsymbol{D}^{1/2}\boldsymbol{P}\boldsymbol{x}\right) = \left\|\boldsymbol{D}^{1/2}\boldsymbol{P}\boldsymbol{x}\right\|^2, \tag{1}$$

which is ≥ 0. Therefore, $\boldsymbol{A}$ is positive semidefinite.

If all of the eigenvalues of $\boldsymbol{A}$ are > 0, then (1) tells us that

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \left\|\boldsymbol{D}^{1/2}\boldsymbol{P}\boldsymbol{x}\right\|^2 = \sum_{j=1}^{n}\left(\left[\boldsymbol{D}^{1/2}\boldsymbol{P}\boldsymbol{x}\right]_j\right)^2 = \sum_{j=1}^{n}\lambda_j\left([\boldsymbol{P}\boldsymbol{x}]_j\right)^2, \tag{2}$$

where $\lambda_j > 0$ for all j. Therefore,

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \ge \min_{1\le j\le n}\lambda_j \cdot \sum_{j=1}^{n}\left([\boldsymbol{P}\boldsymbol{x}]_j\right)^2 = \min_{1\le j\le n}\lambda_j \cdot \boldsymbol{x}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{x} = \min_{1\le j\le n}\lambda_j \cdot \|\boldsymbol{x}\|^2.$$

Since $\min_{1\le j\le n}\lambda_j > 0$, it follows that $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for all nonzero $\boldsymbol{x}$. This completes the proof. □

Let us pause and point out a consequence of the proof of this last result.

Corollary 7. *If A is positive semidefinite, then its extremal eigenvalues satisfy*

$$\min_{1\le j\le n}\lambda_j = \min_{\|\boldsymbol{x}\|>0}\frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\|\boldsymbol{x}\|^2}, \qquad \max_{1\le j\le n}\lambda_j = \max_{\|\boldsymbol{x}\|>0}\frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\|\boldsymbol{x}\|^2}.$$

Proof. We saw, during the course of the previous proof, that

$$\min_{1\le j\le n}\lambda_j \cdot \|\boldsymbol{x}\|^2 \le \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \quad \text{for all } \boldsymbol{x} \in \mathbf{R}^n. \tag{3}$$

Optimize over all $\boldsymbol{x}$ to see that

$$\min_{1\le j\le n}\lambda_j \le \min_{\|\boldsymbol{x}\|>0}\frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\|\boldsymbol{x}\|^2}. \tag{4}$$

But $\min_{1\le j\le n}\lambda_j$ is an eigenvalue for $\boldsymbol{A}$; let $\boldsymbol{z}$ denote a corresponding eigenvector in order to see that

$$\min_{1\le j\le n}\lambda_j \le \min_{\|\boldsymbol{x}\|>0}\frac{\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}}{\|\boldsymbol{x}\|^2} \le \frac{\boldsymbol{z}'\boldsymbol{A}\boldsymbol{z}}{\|\boldsymbol{z}\|^2} = \min_{1\le j\le n}\lambda_j.$$

So both inequalities are in fact equalities, and hence follows the formula for the minimum eigenvalue. The one for the maximum eigenvalue is proved similarly. □

Finally, a word about the square root of positive semidefinite matrices:

Proposition 8. *If $\boldsymbol{A}$ is positive semidefinite, then so is $\mathbf{A}^{1/2}$. If $\boldsymbol{A}$ is positive definite, then so is $\mathbf{A}^{1/2}$.*

Proof. We write, in spectral form, $\boldsymbol{A} = \boldsymbol{PDP}'$ and observe [by squaring it] that $\boldsymbol{A}^{1/2} = \boldsymbol{PD}^{1/2}\boldsymbol{P}'$. Note that $\boldsymbol{D}^{1/2}$ is a real diagonal matrix since the eigenvalues of $\boldsymbol{A}$ are ≥ 0. Therefore, we may apply (1) to $\boldsymbol{A}^{1/2}$ [in place of $\boldsymbol{A}$] to see that $\boldsymbol{x}'\boldsymbol{A}^{1/2}\boldsymbol{x} = \|\boldsymbol{D}^{1/4}\boldsymbol{P}\boldsymbol{x}\|^2 \ge 0$ where $\boldsymbol{D}^{1/4}$ denotes the [real] square root of $\boldsymbol{D}^{1/2}$. This proves that if $\boldsymbol{A}$ is positive semidefinite, then so is $\mathbf{A}^{1/2}$. Now suppose there exists a positive definite $\boldsymbol{A}$ whose square root is not positive definite. It would follow that there necessarily exists a nonzero $\boldsymbol{x}\in\mathbf{R}^n$ such that $\boldsymbol{x}'\boldsymbol{A}^{1/2}\boldsymbol{x} = \|\boldsymbol{D}^{1/4}\boldsymbol{P}\boldsymbol{x}\|^2 = 0$. Since $\boldsymbol{D}^{1/4}\boldsymbol{P}\boldsymbol{x} = \mathbf{0}$,

$$\boldsymbol{D}^{1/2}\boldsymbol{P}\boldsymbol{x} = \mathrm{D}^{1/4}\boldsymbol{D}^{1/4}\boldsymbol{P}\boldsymbol{x} = \mathbf{0} \qquad\Rightarrow\qquad \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \left\|\boldsymbol{D}^{1/2}\boldsymbol{P}\boldsymbol{x}\right\|^2 = 0.$$

And this contradicts the assumption that $\boldsymbol{A}$ is positive definite. □

3. The rank of a matrix

Recall that vectors $\boldsymbol{v}_1,\dots,\boldsymbol{v}_k$ are *linearly independent* if

$$c_1\boldsymbol{v}_1 + \cdots + c_k\boldsymbol{v}_k = \mathbf{0} \qquad\Rightarrow\qquad c_1 = \cdots = c_k = 0.$$

For instance, $\boldsymbol{v}_1 := (1\,,0)'$ and $\boldsymbol{v}_2 := (0\,,1)'$ are linearly independent 2-vectors.

The *column rank* of a matrix $\boldsymbol{A}$ is the maximum number of linearly independent column vectors of $\boldsymbol{A}$. The *row rank* of a matrix $\boldsymbol{A}$ is the maximum number of linearly independent row vectors of $\boldsymbol{A}$. We can interpret these definitions geometrically as follows: First, suppose $\boldsymbol{A}$ is $m\times n$ and define $\mathcal{C}(\boldsymbol{A})$ denote the linear space of all vectors of the form $c_1\boldsymbol{v}_1 + \cdots + c_n\boldsymbol{v}_n$, where $\boldsymbol{v}_1,\dots,\boldsymbol{v}_n$ are the column vectors of $\boldsymbol{A}$ and $c_1,\dots,c_n$ are real numbers. We call $\mathcal{C}(\boldsymbol{A})$ the *column space* of $\boldsymbol{A}$.

We can define the *row space* $\mathcal{R}(\boldsymbol{A})$, of $\boldsymbol{A}$ similarly, or simply define $\mathcal{R}(\boldsymbol{A}) := \mathcal{C}(\boldsymbol{A}')$.

Lemma 9. *For every $m \times n$ matrix $\boldsymbol{A}$,*

$$\mathscr{C}(\boldsymbol{A}) = \{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in \mathbf{R}^n\}, \quad \mathscr{R}(\boldsymbol{A}) := \{\boldsymbol{x}'\boldsymbol{A} : \boldsymbol{x} \in \mathbf{R}^m\}.$$

We can think of an $m \times n$ matrix $\boldsymbol{A}$ as a mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$; namely, we can think of matrix $\boldsymbol{A}$ also as the function $\boldsymbol{f}_{\boldsymbol{A}}(\boldsymbol{x}) := \boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x}$. In this way we see that $\mathscr{C}(\boldsymbol{A})$ is also the "range" of the function $\boldsymbol{f}_{\boldsymbol{A}}$.

Proof. Let us write the columns of $\boldsymbol{A}$ as $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n$. Note that $\boldsymbol{y} \in \mathscr{C}(\boldsymbol{A})$ if and only if there exist $c_1, \ldots, c_n$ such that $\boldsymbol{y} = c_1\boldsymbol{a}_1 + \cdots + c_n\boldsymbol{a}_n = \boldsymbol{A}\boldsymbol{c}$, where $\boldsymbol{c} := (c_1, \ldots, c_n)'$. This shows that $\mathscr{C}(\boldsymbol{A})$ is the collection of all vectors of the form $\boldsymbol{A}\boldsymbol{x}$, for $\boldsymbol{x} \in \mathbf{R}^n$. The second assertion [about $\mathscr{R}(\boldsymbol{A})$] follows from the definition of $\mathscr{R}(\boldsymbol{A}$ equalling $\mathscr{C}(\boldsymbol{A}')$ and the already-proven first assertion. □

It then follows, from the definition of dimension, that

$$\text{column rank of } \boldsymbol{A} = \dim \mathscr{C}(\boldsymbol{A}), \qquad \text{row rank of } \boldsymbol{A} = \dim \mathscr{R}(\boldsymbol{A}).$$

Proposition 10. *Given any matrix $\boldsymbol{A}$, its row rank and column rank are the same. We write their common value as* $\text{rank}(\boldsymbol{A})$.

Proof. Suppose $\boldsymbol{A}$ is $m \times n$ and its column rank is r. Let $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_r$ denote a basis for $\mathscr{C}(\boldsymbol{A})$ and consider the matrix $m \times r$ matrix $\boldsymbol{B} := (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_r)$. Write $\boldsymbol{A}$, columnwise , as $\boldsymbol{A} := (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$. For every $1 \le j \le n$, there exists $c_{1,j}, \ldots, c_{r,j}$ such that $\boldsymbol{a}_j = c_{1,j}\boldsymbol{b}_1 + \cdots + c_{r,j}\boldsymbol{b}_r$. Let $\boldsymbol{C} := (c_{i,j})$ be the resulting $r \times n$ matrix, and note that $\boldsymbol{A} = \boldsymbol{B}\boldsymbol{C}$. Because $A_{i,j} = \sum_{k=1}^{r} B_{i,k}C_{k,j}$;, every row of $\boldsymbol{A}$ is a linear combination of the rows of $\boldsymbol{C}$. In other words, $\mathscr{R}(\boldsymbol{A}) \subseteq \mathscr{R}(\boldsymbol{C})$ and hence the row rank of $\boldsymbol{A}$ is $\le \dim \mathscr{R}(\boldsymbol{C}) = r =$ the column rank of $\boldsymbol{A}$. Apply this fact to $\boldsymbol{A}'$ to see that also the row rank of $\boldsymbol{A}'$ is $\le$ the column rank of $\boldsymbol{A}'$; equivalently that the column rank of $\boldsymbol{A}$ is $\le$ the row rank of $\boldsymbol{A}$. □

Proposition 11. *If $\boldsymbol{A}$ is $n \times m$ and $\boldsymbol{B}$ is $m \times k$, then*

$$\text{rank}(\boldsymbol{A}\boldsymbol{B}) \le \min\left(\text{rank}(\boldsymbol{A}), \text{rank}(\boldsymbol{B})\right).$$

Proof. The proof uses an idea that we exploited already in the proof of Proposition 10: Since $(\boldsymbol{A}\boldsymbol{B})_{j,l} = \sum_{\nu=1}^{n} A_{j,\nu}B_{\nu,l}$, the rows of $\boldsymbol{A}\boldsymbol{B}$ are linear combinations of the rows of $\boldsymbol{B}$; that is $\mathscr{R}(\boldsymbol{A}\boldsymbol{B}) \subseteq \mathscr{R}(\boldsymbol{B})$, whence $\text{rank}(\boldsymbol{A}\boldsymbol{B}) \le \text{rank}(\boldsymbol{B})$. Also, $\mathscr{C}(\boldsymbol{A}\boldsymbol{B}) \subseteq \mathscr{C}(\boldsymbol{A})$, whence $\text{rank}(\boldsymbol{A}\boldsymbol{B}) \le \text{rank}(\boldsymbol{A})$. These observations complete the proof. □

Proposition 12. *If A and C are nonsingular, then*

$$\text{rank}(ABC) = \text{rank}(B),$$

provided that the dimensions match up so that ABC makes sense.

Proof. Let $D := ABC$; our goal is to show that $\text{rank}(D) = \text{rank}(B)$.

Two applications of the previous proposition together yield $\text{rank}(D) \leq \text{rank}(AB) \leq \text{rank}(B)$. And since $B = A^{-1}DC^{-1}$, we have also $\text{rank}(B) \leq \text{rank}(A^{-1}D) \leq \text{rank}(D)$. □

Corollary 13. *If A is an $n \times n$ real and symmetric matrix, then* $\text{rank}(A) =$ *the total number of nonzero eigenvalues of A. In particular, A has full rank if and only if A is nonsingular. Finally, $\mathcal{C}(A)$ is the linear space spanned by the eigenvectors of A that correspond to nonzero eigenvalues.*

Proof. We write A, in spectral form, as $A = PDP^{-1}$, and apply the preceding proposition to see that $\text{rank}(A) = \text{rank}(D)$, which is clearly the total number of nonzero eigenvalue of A. Since A is nonsingular if and only if all of its eigenvalues are nonzero, A has full rank if and only if A is nonsingular.

Finally, suppose A has rank $k \leq n$; this is the number of its nonzero eigenvalues $\lambda_1, \ldots, \lambda_k$. Let $v_1, \ldots, v_n$ denote orthonormal eigenvectors such that $v_1, \ldots, v_k$ are eigenvectors that correspond to $\lambda_1, \ldots, \lambda_k$ and $v_{k+1}, \ldots, v_n$ are eigenvectors that correspond to eigenvalues 0 [Gram–Schmidt]. And define $\mathcal{E}$ to be the span of $v_1, \ldots, v_k$; i.e.,

$$\mathcal{E} := \{c_1 v_1 + \cdots + c_k v_k : \ c_1, \ldots, c_k \in \mathbf{R}\}.$$

Our final goal is to prove that $\mathcal{E} = \mathcal{C}(A)$, which we know is equal to the linear space of all vectors of the form Ax.

Clearly, $c_1 v_1 + \cdots + c_k v_k = Ax$, where $x = \sum_{j=1}^{k}(c_j/\lambda_j)v_j$. Therefore, $\mathcal{E} \subseteq \mathcal{C}(A)$. If $k = n$, then this suffices because in that case $v_1, \ldots, v_k$ is a basis for $\mathbf{R}^n$, hence $\mathcal{E} = \mathcal{C}(A) = \mathbf{R}^n$. If $k < n$, then we can write every $x \in \mathbf{R}^n$ as $a_1 v_1 + \cdots + a_n v_n$, so that $\mathbf{Ax} = \sum_{j=1}^{k} a_j \lambda_j v_j \in \mathcal{E}$. Thus, $\mathcal{C}(A) \subseteq \mathcal{E}$ and we are done. □

Let A be $m \times n$ and define the *null space* [or "kernel"] of A as

$$\mathcal{N}(A) := \{x \in \mathbf{R}^n : \ Ax = \mathbf{0}\}.$$

Note that $\mathcal{N}(A)$ is the linear span of the eigenvectors of A that correspond to eigenvalue 0. The other eigenvectors can be chosen to be orthogonal to these, and hence the preceding proof contains the facts that: (i) Nonzero elements of $\mathcal{N}(\boldsymbol{A})$ are orthogonal to nonzero elements of $\mathcal{C}(\boldsymbol{A})$; and (ii)

$$\dim \mathcal{N}(\boldsymbol{A}) + \text{rank}(\boldsymbol{A}) = n \text{ (= the number of columns of } \boldsymbol{A}). \tag{5}$$

Proposition 14. $\text{rank}(\boldsymbol{A}) = \text{rank}(\boldsymbol{A}'\boldsymbol{A}) = \text{rank}(\boldsymbol{A}\boldsymbol{A}')$ *for every* $m \times n$ *matrix* $\boldsymbol{A}$*.*

Proof. If $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ then $\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, and if $\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$, then $\|\boldsymbol{A}\boldsymbol{x}\|^2 = \boldsymbol{x}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = 0$. In other words, $\mathcal{N}(\boldsymbol{A}) = \mathcal{N}(\boldsymbol{A}'\boldsymbol{A})$. Because $\boldsymbol{A}'\boldsymbol{A}$ and $\boldsymbol{A}$ both have n columnes, it follows from (5) that $\text{rank}(\boldsymbol{A}'\boldsymbol{A}) = \text{rank}(\boldsymbol{A})$. Apply this observation to $\boldsymbol{A}'$ to see that $\text{rank}(\boldsymbol{A}') = \text{rank}(\boldsymbol{A}\boldsymbol{A}')$ as well. The result follows from this and the fact that $\boldsymbol{A}$ and $\boldsymbol{A}'$ have the same rank (Proposition 10). □

4. Projection matrices

A matrix $\boldsymbol{A}$ is said to be a *projection* matrix if: (i) $\boldsymbol{A}$ is symmetric; and (ii) $\boldsymbol{A}$ is "idempotent"; that is, $\boldsymbol{A}^2 = \boldsymbol{A}$.

Note that projection matrices are always positive semidefinite. Indeed, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{A}^2\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{x} = \|\boldsymbol{A}\boldsymbol{x}\|^2 \geq 0$

Proposition 15. *If* $\boldsymbol{A}$ *is an* $n \times n$ *projection matrix, then so is* $\boldsymbol{I} - \boldsymbol{A}$*. Moreover, all eigenvalues of* $\boldsymbol{A}$ *are zeros and ones, and* $\text{rank}(\boldsymbol{A}) =$ *the number of eigenvalues that are equal to one.*

Proof. $(\boldsymbol{I} - \boldsymbol{A})^2 = \boldsymbol{I} - 2\boldsymbol{A} + \boldsymbol{A}^2 = \boldsymbol{I} - \boldsymbol{A}$. Since $\boldsymbol{I} - \boldsymbol{A}$ is symmetric also, it is a projection. If λ is an eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{x}$ is a corresponding eigenvector, then $\lambda\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^2\boldsymbol{x} = \lambda\boldsymbol{A}\boldsymbol{x} = \lambda^2\boldsymbol{x}$. Multiply both sides by $\boldsymbol{x}'$ to see that $\lambda\|\boldsymbol{x}\|^2 = \lambda^2\|\boldsymbol{x}\|^2$. Since $\|\boldsymbol{x}\| > 0$, it follows that $\lambda \in \{0\,,1\}$. The total number of nonzero eigenvalues is then the total number of eigenvalues that are ones. Therefore, the rank of $\boldsymbol{A}$ is the total number of eigenvalues that are one. □

Corollary 16. *If* $\boldsymbol{A}$ *is a projection matrix, then* $\text{rank}(\boldsymbol{A}) = \text{tr}(\boldsymbol{A})$*.*

Proof. Simply recall that $\text{tr}(\boldsymbol{A})$ is the sum of the eigenvalues, which for a projection matrix, is the total number of eigenavalues that are one. □

Why are they called "projection" matrices? Or, perhaps even more importantly, what is a "projection"?

Lemma 17. *Let Ω denote a linear subspace of $\mathbf{R}^n$, and $\boldsymbol{x} \in \mathbf{R}^n$ be fixed. Then there exists a unique element $\boldsymbol{y} \in \Omega$ that is closest to $\boldsymbol{x}$; that is,*

$$\|\boldsymbol{y} - \boldsymbol{x}\| = \min_{\boldsymbol{z} \in \Omega} \|\boldsymbol{z} - \boldsymbol{x}\|.$$

The point $\boldsymbol{y}$ is called the projection *of $\boldsymbol{x}$ onto Ω.*

Proof. Let $k := \dim \Omega$, so that there exists an orthonormal basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$ for Ω. Extend this to a basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n$ for all of $\mathbf{R}^n$ by the Gram–Schmidt method.

Given a fixed vector $\boldsymbol{x} \in \mathbf{R}^n$, we can write it as $\boldsymbol{x} := c_1\boldsymbol{b}_1 + \cdots + c_n\boldsymbol{b}_n$ for some $c_1, \ldots, c_n \in \mathbf{R}$. Define $\boldsymbol{y} := c_1\boldsymbol{b}_1 + \cdots + c_k\boldsymbol{b}_k$. Clearly, $\boldsymbol{y} \in \Omega$ and $\|\boldsymbol{y} - \boldsymbol{x}\|^2 = \sum_{i=k+1}^n c_i^2$. Any other $\boldsymbol{z} \in \Omega$ can be written as $\boldsymbol{z} = \sum_{i=1}^k d_i\boldsymbol{b}_i$, and hence $\|\boldsymbol{z} - \boldsymbol{x}\|^2 = \sum_{i=1}^k (d_i - c_i)^2 + \sum_{i=k+1}^n c_i^2$, which is strictly greater than $\|\boldsymbol{y} - \boldsymbol{x}\|^2 = \sum_{i=k+1}^n c_i^2$ unless $d_i = c_i$ for all $i = 1, \ldots, k$; i.e., unless $\boldsymbol{z} = \boldsymbol{y}$. □

Usually, we have a k-dimensional linear subspace Ω of $\mathbf{R}^n$ that is the range of some $n \times k$ matrix $\boldsymbol{A}$. That is, $\Omega = \{\boldsymbol{A}\boldsymbol{y} : \boldsymbol{y} \in \mathbf{R}^k\}$. Equivalently, $\Omega = \mathscr{C}(\boldsymbol{A})$. In that case,

$$\min_{\boldsymbol{z} \in \Omega} \|\boldsymbol{z} - \boldsymbol{x}\|^2 = \min_{\boldsymbol{y} \in \mathbf{R}^k} \|\boldsymbol{A}\boldsymbol{y} - \boldsymbol{x}\|^2 = \min_{\boldsymbol{y} \in \mathbf{R}^k} \left[\boldsymbol{y}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{A}'\boldsymbol{x} - \boldsymbol{x}'\boldsymbol{A}\boldsymbol{y} + \boldsymbol{x}'\boldsymbol{x}\right].$$

Because $\boldsymbol{y}'\boldsymbol{A}'\boldsymbol{x}$ is a scalar, the preceding is simplified to

$$\min_{\boldsymbol{z} \in \Omega} \|\boldsymbol{z} - \boldsymbol{x}\|^2 = \min_{\boldsymbol{y} \in \mathbf{R}^k} \left[\boldsymbol{y}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{A}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{x}\right].$$

Suppose that the $k \times k$ positive semidefinite matrix $\boldsymbol{A}'\boldsymbol{A}$ is nonsingular [so that $\boldsymbol{A}'\boldsymbol{A}$ and hence also $(\boldsymbol{A}\boldsymbol{A}')^{-1}$ are both positive definite]. Then, we can relabel variables [$\boldsymbol{\alpha} := \boldsymbol{A}'\boldsymbol{A}\boldsymbol{y}$] to see that

$$\min_{\boldsymbol{z} \in \Omega} \|\boldsymbol{z} - \boldsymbol{x}\|^2 = \min_{\boldsymbol{\alpha} \in \mathbf{R}^k} \left[\boldsymbol{\alpha}'(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{\alpha} - 2\boldsymbol{\alpha}'(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{x}\right].$$

A little arithmetic shows that

$$\begin{aligned}&(\boldsymbol{\alpha} - \boldsymbol{A}'\boldsymbol{x})'(\boldsymbol{A}'\boldsymbol{A})^{-1}(\boldsymbol{\alpha} - \boldsymbol{A}'\boldsymbol{x}) \\ &\qquad = \boldsymbol{\alpha}'(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{\alpha} - 2\boldsymbol{\alpha}'(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{x}.\end{aligned}$$

Consequently,

$$\begin{aligned}&\min_{\boldsymbol{z} \in \Omega} \|\boldsymbol{z} - \boldsymbol{x}\|^2 \\ &= \min_{\boldsymbol{\alpha} \in \mathbf{R}^k} \left[(\boldsymbol{\alpha} - \boldsymbol{A}'\boldsymbol{x})'(\boldsymbol{A}'\boldsymbol{A})^{-1}(\boldsymbol{\alpha} - \boldsymbol{A}'\boldsymbol{x}) - \boldsymbol{x}'\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{x}\right].\end{aligned}$$

The first term in the parentheses is ≥ 0; in fact it is > 0 unless we select $\boldsymbol{\alpha} = \boldsymbol{A}'\boldsymbol{x}$. This proves that the projection of $\boldsymbol{x}$ onto Ω is obtained by setting $\boldsymbol{\alpha} := \boldsymbol{A}'\boldsymbol{x}$, in which case the projection itself is $\boldsymbol{A}\boldsymbol{y} =$

$\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{x}$ and the distance between $\boldsymbol{y}$ and $\boldsymbol{x}$ is the square root of $\|\boldsymbol{x}\|^2 - \boldsymbol{x}'\boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{x}$.

Let $\boldsymbol{P}_\Omega := \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$. It is easy to see that $\boldsymbol{P}_\Omega$ is a projection matrix. The preceding shows that $\boldsymbol{P}_\Omega\boldsymbol{x}$ is the projection of $\boldsymbol{x}$ onto Ω for every $\boldsymbol{x} \in \mathbf{R}^n$. That is, we can think of $\boldsymbol{P}_\Omega$ as the matrix that projects onto Ω. Moreover, the distance between $\boldsymbol{x}$ and the linear subspace Ω [i.e., $\min_{\boldsymbol{z}\in\mathbf{R}^k} \|\boldsymbol{z}-\boldsymbol{x}\|$] is exactly the square root of $\boldsymbol{x}'\boldsymbol{x} - \boldsymbol{x}'\boldsymbol{P}_\Omega\boldsymbol{x} = \boldsymbol{x}'(\boldsymbol{I}-\boldsymbol{P}_\Omega)\boldsymbol{x} = \|(\boldsymbol{I}-\boldsymbol{P}_\Omega)\boldsymbol{x}\|^2$, because $\boldsymbol{I}-\boldsymbol{P}_\Omega$ is a projection matrix. What space does it project into?

Let $\Omega^\perp$ denote the collection of all n-vectors that are perpendicular to every element of Ω. If $\boldsymbol{z} \in \Omega^\perp$, then we can write, for all $\boldsymbol{x} \in \mathbf{R}^n$,

$$\begin{aligned}\|\boldsymbol{z}-\boldsymbol{x}\|^2 &= \|\boldsymbol{z} - (\boldsymbol{I}-\boldsymbol{P}_\Omega)\boldsymbol{x} + \boldsymbol{P}_\Omega\boldsymbol{x}\|^2 \\ &= \|\boldsymbol{z} - (\boldsymbol{I}-\boldsymbol{P}_\Omega)\boldsymbol{x}\|^2 + \|\boldsymbol{P}_\Omega\boldsymbol{x}\|^2 - 2\left\{\boldsymbol{z} - (\boldsymbol{I}-\boldsymbol{P}_\Omega)\boldsymbol{x}\right\}'\boldsymbol{P}_\Omega\boldsymbol{x} \\ &= \|\boldsymbol{z} - (\boldsymbol{I}-\boldsymbol{P}_\Omega)\boldsymbol{x}\|^2 + \|\boldsymbol{P}_\Omega\boldsymbol{x}\|^2,\end{aligned}$$

since $\boldsymbol{z}$ is orthogonal to every element of Ω including $\boldsymbol{P}_\Omega\boldsymbol{x}$, and $\boldsymbol{P}_\Omega = \boldsymbol{P}_\Omega^2$. Take the minimum over all $\boldsymbol{z} \in \Omega^\perp$ to find that $\boldsymbol{I}-\boldsymbol{P}_\Omega$ is the projection onto $\Omega^\perp$. Let us summarize our findings.

Proposition 18. *If $\boldsymbol{A}'\boldsymbol{A}$ is nonsingular [equivalently, has full rank], then $\boldsymbol{P}_{\mathcal{C}(\boldsymbol{A})} := \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$ is the projection onto $\mathcal{C}(\boldsymbol{A})$, $\boldsymbol{I} - \boldsymbol{P}_{\mathcal{C}(\boldsymbol{A})} = \boldsymbol{P}_{\mathcal{C}(\boldsymbol{A})^\perp}$ is the projection onto $\Omega^\perp$, and we have*

$$\boldsymbol{x} = \boldsymbol{P}_{\mathcal{C}(\boldsymbol{A})}\boldsymbol{x} + \boldsymbol{P}_{\mathcal{C}(\boldsymbol{A})^\perp}\boldsymbol{x}, \quad \text{and} \quad \|\boldsymbol{x}\|^2 = \left\|\boldsymbol{P}_{\mathcal{C}(\boldsymbol{A})}\boldsymbol{x}\right\|^2 + \left\|\boldsymbol{P}_{\mathcal{C}(\boldsymbol{A})^\perp}\boldsymbol{x}\right\|^2.$$

The last result is called the "Pythagorean property."