UNIVERSIDAD DEL PACÍFICO
Departamento Académico de Economía
Matemáticas III (130233) - Sección A
Segundo Semestre 2017
Profesor Diego Winkelried

2 | Matrix algebra

2.1 Basic concepts

- **Matrix.** A matrix of dimension $m \times n$ is a rectangular array of numbers, with m rows and n columns:

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

Often, matrices are denoted in reference to their entries, $\boldsymbol{A} = [a_{ij}]_{m\times n}$.

When $n = 1$, $\boldsymbol{A}$ is a *column vector* in $\mathbb{R}^m$. If $m = 1$, $\boldsymbol{A}$ is a *row vector* in $\mathbb{R}^n$. When $m = n = 1$, $\boldsymbol{A}$ is a scalar. If $m = n$, $\boldsymbol{A}$ is a *square* matrix. In such cases, the entries a_{ii} ($i = 1, 2, \ldots, n$) conform the *diagonal* of $\boldsymbol{A}$.

- **Addition.** The sum of two matrices *of the same dimension* results in matrix whose entries are the sum of the corresponding entries in the summing matrices. That is, for $\boldsymbol{A} = [a_{ij}]_{m\times n}$ and $\boldsymbol{B} = [b_{ij}]_{m\times n}$, then $\boldsymbol{C} = \boldsymbol{A} + \boldsymbol{B} = [a_{ij} + b_{ij}]_{m\times n}$. Explicitly,

$$\boldsymbol{C} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix}.$$

- **Product by a scalar.** For $\boldsymbol{A} = [a_{ij}]_{m\times n}$ and $\lambda \in \mathbb{R}$, then $\lambda\boldsymbol{A} = [\lambda a_{ij}]_{m\times n}$. That is to say, the product of a matix by a scalar gives a matrix whose entries are those of the original matrix, times the scalar.

- **Basic properties.** If $\boldsymbol{0}$ denotes a null $m \times n$ matrix (all its entries are equal to zero), then for $\boldsymbol{A}_{m\times n}$, $\boldsymbol{B}_{m\times n}$, $\boldsymbol{C}_{m\times n}$, $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}$,

  (1) $\boldsymbol{A} + \boldsymbol{0} = \boldsymbol{A}$.

  (2) $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$.

  (3) $(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C})$.

  (4) $\alpha(\boldsymbol{A} + \boldsymbol{B}) = \alpha\boldsymbol{A} + \alpha\boldsymbol{B}$.

  (5) $(\alpha + \beta)\boldsymbol{A} = \alpha\boldsymbol{A} + \beta\boldsymbol{A}$.

  (6) $\boldsymbol{A} + (-\boldsymbol{A}) = \boldsymbol{0}$.

- **Matrix multiplication.** Let $\boldsymbol{A} = [a_{ij}]_{m\times n}$ and $\boldsymbol{B} = [b_{ij}]_{n\times p}$ be two matrices. The (i, j) entry of the product $\boldsymbol{C} = \boldsymbol{AB}$ is

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}.$$

c_{ij} is the inner product of two vectors: the i-th *row* of $\boldsymbol{A}$ and the j-th *column* of $\boldsymbol{B}$. Clearly, the number of rows of $\boldsymbol{A}$ must be the same as the number of columns of $\boldsymbol{B}$ (i.e., the matrices should be "conformable with the matrix product"). Otherwise, the matrices cannot be multiplied.

- **Properties of matrix multiplication.** Assume conformability when necessary,
  (1) $\boldsymbol{AB}$ is not necessarily equal to $\boldsymbol{BA}$ (in general, matrices do not commute).
  (2) $\boldsymbol{AB} = \boldsymbol{0}$ does not imply, necessarily, that $\boldsymbol{A} = \boldsymbol{0}$ or $\boldsymbol{B} = \boldsymbol{0}$.
  (3) $\boldsymbol{AB} = \boldsymbol{AC}$ does not imply, necessarily, that $\boldsymbol{B} = \boldsymbol{C}$.
  (4) $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$.
  (5) $\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$.
  (6) $(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$.
  (7) $(\boldsymbol{A} + \boldsymbol{B})(\boldsymbol{C} + \boldsymbol{D}) = \boldsymbol{AC} + \boldsymbol{AD} + \boldsymbol{BC} + \boldsymbol{BD}$.

- **Identity matrix.** The matrix "1". An identity matrix of dimension n, denoted by $\boldsymbol{I}_n$, is a square matrix whose diagonal entries are all equal to one, whereas all of its nondiagonal entries are equal to zero.

  For every $m \times n$ matrix $\boldsymbol{A}$, it is true that $\boldsymbol{A} \cdot \boldsymbol{I}_n = \boldsymbol{I}_m \cdot \boldsymbol{A} = \boldsymbol{A}$.

- **Powers.** For a *square* matrix: $\boldsymbol{A}^2 = \boldsymbol{AA}$, $\boldsymbol{A}^3 = \boldsymbol{AAA}$ and, in general,

$$\boldsymbol{A}^k = \underbrace{\boldsymbol{AA} \cdots \boldsymbol{A}}_{k \text{ times}} .$$

- **Transposition.** If $\boldsymbol{A} = [a_{ij}]_{m \times n}$, its transpose is defined as $\boldsymbol{A}' = [a_{ji}]_{n \times m}$. That is, the transpose of a matrix is obtained by interchanging rows and columns such that the i-th row of $\boldsymbol{A}$ becomes the i-the column of $\boldsymbol{A}'$, and the j-the column of $\boldsymbol{A}$ becomes the j-th row of $\boldsymbol{A}'$.

- **Properties of transposition.** Assume conformability when necessary,
  (1) $(\boldsymbol{A}')' = \boldsymbol{A}$.
  (2) $(\boldsymbol{A} + \boldsymbol{B})' = \boldsymbol{A}' + \boldsymbol{B}'$.
  (3) $(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$.
  (4) $(\lambda \boldsymbol{A})' = \lambda \boldsymbol{A}'$, where $\lambda \in \mathbb{R}$.

- **Trace.** The trace of a *square* matrix $\boldsymbol{A} = [a_{ij}]_{n \times n}$ is the sum of the entries along the diagonal:

$$\text{tr}(\boldsymbol{A}) = a_{11} + a_{22} + \ldots + a_{nn} = \sum_{i=1}^{n} a_{ii} .$$

- **Properties of the trace.** It can be verified that
  (1) $\text{tr}(\boldsymbol{A} + \boldsymbol{B} + \boldsymbol{C}) = \text{tr}(\boldsymbol{A}) + \text{tr}(\boldsymbol{B}) + \text{tr}(\boldsymbol{C})$.
  (2) $\text{tr}(\lambda \boldsymbol{A}) = \lambda \text{tr}(\boldsymbol{A})$, where $\lambda \in \mathbb{R}$.
  (3) $\text{tr}(\boldsymbol{A}) = \text{tr}(\boldsymbol{A}')$.
  (4) Under the trace, the matrix product is cyclically commutative: $\text{tr}(\boldsymbol{ABC}) = \text{tr}(\boldsymbol{BCA}) = \text{tr}(\boldsymbol{CAB})$.

- **Determinants.** The determinant of a *square* matrix $\boldsymbol{A}$, denoted as $|\boldsymbol{A}|$ or $\det(\boldsymbol{A})$, is a uniquely defined scalar associated to this matrix. It gives information about a matrix that corresponds to a linear transformation of a vector space. There are many techniques to compute the determinants that are not considered in these notes (the Sarrus' scheme or reduction to echelon form, for instance).

- **Determinant of a** 2×2 **matrix.**

$$\text{If} \quad \boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad \text{then} \qquad |\boldsymbol{A}| = a_{11}a_{22} - a_{12}a_{21} \,.$$

- **Minors.** The ij-th minor of a matrix $\boldsymbol{A}$, M_{ij}, is the determinant of the submatrix rendered after removing the i-th row and j-th column from $\boldsymbol{A}$.

- **Cofactor.** The ij-th cofactor (or the cofactor of a_{ij}) is simply $C_{ij} = (-1)^{i+j} M_{ij}$.

- **Determinant of a** 3×3 **matrix.** (One of many techniques) Let

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}.$$

  Consider the following minors

$$M_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \qquad M_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} \qquad \text{and} \qquad M_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

  Then,

$$|\boldsymbol{A}| = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13} = a_{11}M_{11} - a_{12}M_{12} + a_{13}M_{13} \,.$$

  The last expression corresponds to an *expansion by the first row*.

- **Determinant of an** $n \times n$ **matrix.** The above result can be readily generalized:

$$|\boldsymbol{A}| = \sum_{j=1}^{n} a_{ij}C_{ij} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij}$$

  for *any* row i.

- **Properties of the determinant.** Assume conformability when necessary,
  (1) $|\boldsymbol{A} + \boldsymbol{B}| \neq |\boldsymbol{A}| + |\boldsymbol{B}|$ (it is not a linear operator).
  (2) $|\lambda \boldsymbol{A}| = \lambda^n |\boldsymbol{A}|$, where $\lambda \in \mathbb{R}$.
  (3) $|\boldsymbol{A}| = |\boldsymbol{A}'|$.
  (4) $|\boldsymbol{AB}| = |\boldsymbol{A}| \cdot |\boldsymbol{B}|$.
  (5) If $|\boldsymbol{A}|$ is triangular, i.e. $a_{ij} = 0$ for $i > j$ (or for $i < j$), then $|\boldsymbol{A}| = \prod_{i=1}^{n} a_{ii}$.
  (6) If the rows (or columns) of $\boldsymbol{A}$ form a linearly independent set, then $|\boldsymbol{A}| \neq 0$. On the contrary, If the rows (or columns) of $\boldsymbol{A}$ form a linearly dependent set, then $|\boldsymbol{A}| = 0$.

- **Rank.** The rank of an $m \times n$ matrix $\boldsymbol{A}$, denoted by $\rho(\boldsymbol{A})$, is the number of linearly independent columns (or rows). Obviously, $\rho(\boldsymbol{A}) \leq \min\{m, n\}$. When $\rho(\boldsymbol{A}) = m$ we say that $\boldsymbol{A}$ has (or is) *full row rank*, whereas if $\rho(\boldsymbol{A}) = n$ we say that $\boldsymbol{A}$ has (or is) *full column rank*. If $\boldsymbol{A}$ is square and $\rho(\boldsymbol{A}) = n$, we simply say that $\boldsymbol{A}$ has (or is) *full rank*.

- **Properties of the rank.** Assume conformability when necessary,
  (1) $\rho(\boldsymbol{A}) = \rho(\boldsymbol{A}')$.
  (2) $\rho(\boldsymbol{AB}) \leq \min\{\rho(\boldsymbol{A}), \rho(\boldsymbol{B})\}$.
  (3) As a corollary, $\rho(\boldsymbol{A}) = \rho(\boldsymbol{A}'\boldsymbol{A})$.
  (4) Also, if $\boldsymbol{A}$ is square and full rank, then $|\boldsymbol{A}| \neq 0$. On the contrary, if it is not full rank, then $|\boldsymbol{A}| = 0$.

- **Inverse.** The inverse of a *square* $n \times n$ matrix $\boldsymbol{A}$ is another $n \times n$ matrix (unique), denoted as $\boldsymbol{A}^{-1}$, such that

  $$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n \,.$$

  The inverse does not always exist. When $\boldsymbol{A}^{-1}$ does not exist we say that $\boldsymbol{A}$ is *singular*, whereas if the inverse exists, then $\boldsymbol{A}$ is *nonsingular*.

  A nonsingular matrix is necessarily full rank, $\rho(\boldsymbol{A}) = n$. This means that no vector $\boldsymbol{c} \neq \boldsymbol{0}$ such $\boldsymbol{A}\boldsymbol{c} = \boldsymbol{0}$ exists. If such a vector existed, then $\rho(\boldsymbol{A}) < n$ and $\boldsymbol{A}$ would be singular.

- **Properties of the inverse.** For two nonsingular matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of the same dimension,
  (1) $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$.
  (2) $(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$.
  (3) $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.
  (4) $(\boldsymbol{A}+\boldsymbol{B})^{-1} = \boldsymbol{A}^{-1}(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}\boldsymbol{B}^{-1}$.
  (5) $|\,\boldsymbol{A}^{-1}\,| = 1/|\,\boldsymbol{A}\,|$ (the reciprocal of $|\,\boldsymbol{A}\,|$).

- **Computation with cofactors**. Let $\boldsymbol{C}$ be the *cofactor matrix*, whose typical entry is C_{ij}. The inverse satisfies

  $$\boldsymbol{A}^{-1} = \frac{\boldsymbol{C}'}{|\,\boldsymbol{A}\,|} = \frac{\operatorname{adj}(\boldsymbol{A})}{|\,\boldsymbol{A}\,|}\,,$$

  where $\operatorname{adj}(\boldsymbol{A}) = \boldsymbol{C}'$ (the transpose of the cofactor matrix) is the *adjoint* (or *adjugate*) of $\boldsymbol{A}$.

  The adjoint matrix *always* exists. Therefore, a sufficient condition for nonsingularity (i.e., the existence of the inverse) is that $|\,\boldsymbol{A}\,| = 0$ (alternatively, $\rho(\boldsymbol{A}) = n$).

- **The** 2×2 **case**.

  $$\text{If} \quad \boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} a_{22} & -a_{21} \\ -a_{21} & a_{11} \end{bmatrix} \quad \text{so} \quad \boldsymbol{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}.$$

- **Summary of (non) singularity**. Various matrix concepts and results are related to the notion of linear dependence (amongst the columns or rows of a matrix) and the existence of a solution to a linear system. Given a square $n \times n$ matrix $\boldsymbol{A}$, the following are equivalent:
  (1) $|\,\boldsymbol{A}\,| \neq 0$,
  $\boldsymbol{A}$ is full rank (or has full rank), $\rho(\boldsymbol{A}) = n$,
  The rows and columns of $\boldsymbol{A}$ are linearly independent,
  $\boldsymbol{A}$ is nonsingular, its inverse $\boldsymbol{A}^{-1}$ exists and is unique,
  The linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ is uniquely solved by $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{0} = \boldsymbol{0}$.
  (2) $|\,\boldsymbol{A}\,| = 0$,
  $\boldsymbol{A}$ is not of full rank, $\rho(\boldsymbol{A}) < n$,
  The rows and columns of $\boldsymbol{A}$ are linearly dependent,
  $\boldsymbol{A}$ is singular, its inverse $\boldsymbol{A}^{-1}$ does not exist,
  The linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ can be solved by infinitely many vectors $\boldsymbol{x} \neq \boldsymbol{0}$ (if the system is consistent), or it does not have a solution (if the system is inconsistent).

Check from other sources

- Row reduction into echelon form. The analysis of rank with this technique.
- Row reduction into *reduced* echelon form. The computation of the matrix inverse using this technique.
- The analysis of linear systems ($\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$) with the reduction techniques (i.e., the Gauss-Jordan elimination method).

2.2 Special square matrices

Symmetric	: $\boldsymbol{A} = \boldsymbol{A}'$.
Antisymmetric	: $\boldsymbol{A} = -\boldsymbol{A}'$.
Diagonal	: $\boldsymbol{A} = [a_{ij}]_{n\times n}$ such that $a_{ij} = 0$ for all $i \neq j$.
Lower triangular	: $\boldsymbol{A} = [a_{ij}]_{n\times n}$ such that $a_{ij} = 0$ for all $i < j$.
Upper triangular	: $\boldsymbol{A} = [a_{ij}]_{n\times n}$ such that $a_{ij} = 0$ for all $i > j$.
Idempotent	: $\boldsymbol{A}^2 = \boldsymbol{A}$.
Involutory	: $\boldsymbol{A}^2 = \boldsymbol{I}_n$.
Nilpotent	: $\boldsymbol{A}^k = \boldsymbol{0}$ (nilpotent of degree k).
Normal	: $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{A}'\boldsymbol{A}$.
Orthogonal	: The columns (or rows) are orthonormal, $\boldsymbol{a}_i{}'\boldsymbol{a}_j = 0$ for $i \neq j$ y $\boldsymbol{a}_i{}'\boldsymbol{a}_i = 1$. This implies that $\boldsymbol{A}\boldsymbol{A}' = \boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}_n$. It follows that $\boldsymbol{A}^{-1} = \boldsymbol{A}'$.
Sequentially stable	: $\lim_{k\to\infty} \boldsymbol{A}^k = \boldsymbol{0}$ (convergent).

2.3 Eigenvalues and eigenvectors

- **Definition.** Let $\boldsymbol{A}$ be an $n \times n$ *square* matrix. The scalar λ is an eigenvalue of $\boldsymbol{A}$, associated to the eigenvector $\boldsymbol{v} \in \mathbb{R}^n$, which is different from the null vector ($\boldsymbol{v} \neq \boldsymbol{0}$), if

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}\,.$$

- **Computation of the eigenvalues.** From the definition, we obtain that λ and $\boldsymbol{v}$ are such that $(\boldsymbol{A} - \lambda\boldsymbol{I}_n)\boldsymbol{v} = \boldsymbol{0}$. If $\boldsymbol{A} - \lambda\boldsymbol{I}_n$ were a nonsingular matrix, then the only possible solution to this system would be $\boldsymbol{v} = \boldsymbol{0}$, but this contradicts the requirement of a non-null eigenvector, $\boldsymbol{v} \neq \boldsymbol{0}$. Therefore, λ is to be chosen such that matrix $\boldsymbol{A} - \lambda\boldsymbol{I}_n$ is singular.

  Concretely, define the *characteristic polynomial* of matrix $\boldsymbol{A}$ as

$$P(\lambda) = |\,\boldsymbol{A} - \lambda\boldsymbol{I}_n\,|\,.$$

  The eigenvalues of $\boldsymbol{A}$, thus, solve the *characteristic equation* $P(\lambda) = 0$.

- **How many eigenvalues?** Explicitly, the characteristic polynomial of $\boldsymbol{A}$ is

$$P(\lambda) = \begin{vmatrix} a_{11}-\lambda & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-\lambda \end{vmatrix}\,.$$

  This is a polynomial in λ of degree n. Hence, the characteristic equation $P(\lambda) = 0$ is solved by n values of λ, the roots of $P(\lambda)$. Then, an $n \times n$ matrix $\boldsymbol{A}$ has n eigenvalues and also n eigenvectors.

- **The case $n = 2$.** The characteristic polynomial is quadratic in this case,

$$P(\lambda) = \begin{vmatrix} a_{11}-\lambda & a_{12} \\ a_{21} & a_{22}-\lambda \end{vmatrix} = (a_{11}-\lambda)(a_{22}-\lambda) - a_{12}a_{21} = \lambda^2 - (a_{11}+a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21}\,.$$

  In the $n = 2$ case, we can obtain the more compact expression

$$P(\lambda) = \lambda^2 - \mathrm{tr}(\boldsymbol{A})\lambda + \det(\boldsymbol{A})\,.$$

  Therefore, $P(\lambda) = 0$ is solved by

$$\lambda_{1,2} = \frac{1}{2}\left(\mathrm{tr}(\boldsymbol{A}) \pm \sqrt{\mathrm{tr}(\boldsymbol{A})^2 - 4\det(\boldsymbol{A})}\right)\,.$$

- **Computation of the eigenvectors.** If λ_i is an eigenvalue of $\boldsymbol{A}$ associated to the eigenvector $\boldsymbol{v}_i$, then

$$(\boldsymbol{A} - \lambda_i \boldsymbol{I}_n)\boldsymbol{v}_i = \boldsymbol{0}.$$

We have a linear system of n equations and n unknowns (the entries of $\boldsymbol{v}_i$). This system does not have a unique solution since, by construction, $\boldsymbol{A} - \lambda_i \boldsymbol{I}_n$ is singular (the system has redundant equations). Thus, in principle there are infinitely many eigenvectors $\boldsymbol{v}_i$, all of them are collinear. Strictly, what we have here is a subspace $\mathcal{E}(\boldsymbol{v}) = \{\boldsymbol{v} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}\}$ and the "eigenvector" refers to the basis $\boldsymbol{v}_i$ of this subspace. Often, this basis is *normalized* so the eigenvector is a unit vector, $\|\boldsymbol{v}_i\| = 1$.

- **The case $n = 2$.** Consider the eigenvalue λ_1. The purpose is to obtain $\boldsymbol{v}_1$. By definition,

$$\begin{bmatrix} a_{11} - \lambda_1 & a_{12} \\ a_{21} & a_{22} - \lambda_1 \end{bmatrix} \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad \rightarrow \quad \begin{array}{r} (a_{11} - \lambda_1)v_{11} + a_{12}v_{12} = 0 \\ a_{21}v_{11} + (a_{22} - \lambda_1)v_{12} = 0 \end{array}.$$

We know that $(a_{11} - \lambda_1)(a_{22} - \lambda_1) = a_{12}a_{21}$. Upon multiplying the second equation by a_{12},

$$a_{12}a_{21}v_{11} + a_{12}(a_{22} - \lambda_1)v_{12} = 0 \quad \rightarrow \quad (a_{22} - \lambda_1)[\,(a_{11} - \lambda_1)v_{11} + a_{12}v_{12}\,] = 0\,,$$

from which we obtain

$$v_{11} = -\left(\frac{a_{12}}{a_{11} - \lambda_1}\right) v_{12} \equiv \beta v_{12}\,.$$

Note that this is clearly the very same expression we would obtain if we worked with the first equation. Both equations provide the same information about v_{11} and v_{12}, because they are linearly dependent. This is a manifestation that one of the equations in the system is redundant.

Define the vector subspace $\mathcal{E} = \{\boldsymbol{x} \in \mathbb{R}^2 \mid x_1 = \beta x_2\}$. The conclusion is that any vector $\boldsymbol{v} \in \mathcal{E}$ is an eigenvector of $\boldsymbol{A}$ associated to the eigenvalue λ_1. That is, $\boldsymbol{v}_1 = (\beta t, t)'$ for any $t \in \mathbb{R}$.

If we normalize this vector, we choose t such that $\|\boldsymbol{v}_1\| = 1$, i.e. $\beta^2 t^2 + t^2 = 1$, and obtain

$$\boldsymbol{v}_1 = \left(\sqrt{\frac{\beta^2}{1+\beta^2}}\,,\sqrt{\frac{1}{1+\beta^2}}\right)'.$$

The same procedure, using the eigenvalue λ_2, determines $\boldsymbol{v}_2$.

- **Matrix diagonalization.** We know that $\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$ for $i = 1, 2, \ldots, n$.

Define an $n \times n$ matrix $\boldsymbol{P}$ whose columns are the eigenvectors of $\boldsymbol{A}$,

$$\boldsymbol{P} = [\,\boldsymbol{v}_1 \quad \boldsymbol{v}_2 \quad \cdots \quad \boldsymbol{v}_n\,]\,.$$

Thus,

$$\boldsymbol{A}\boldsymbol{P} = [\,\boldsymbol{A}\boldsymbol{v}_1 \quad \boldsymbol{A}\boldsymbol{v}_2 \quad \cdots \quad \boldsymbol{A}\boldsymbol{v}_n\,]\,.$$

On the other hand, define an $n \times n$ *diagonal* matrix $\boldsymbol{D}$ that contains the eigenvalues of $\boldsymbol{A}$ along the diagonal,

$$\boldsymbol{D} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

It is easy to verify that

$$\boldsymbol{P}\boldsymbol{D} = [\,\lambda_1\boldsymbol{v}_1 \quad \lambda_2\boldsymbol{v}_2 \quad \cdots \quad \lambda_n\boldsymbol{v}_n\,]\,.$$

Since $\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i$, the i-th column of $\boldsymbol{AP}$ is equal to the i-th column of $\boldsymbol{PD}$. Since this happens for all columns, it follows that

$$\boldsymbol{AP} = \boldsymbol{PD}\,.$$

Solving for $\boldsymbol{A}$ gives the so-called *spectral decomposition*

$$\boldsymbol{A} = \boldsymbol{PDP}^{-1}\,.$$

Similarly, solving for $\boldsymbol{D}$ gives the result of *diagonalizing* $\boldsymbol{A}$,

$$\boldsymbol{D} = \boldsymbol{P}^{-1}\boldsymbol{AP}\,.$$

- **Similarity.** Matrices $\boldsymbol{A}$ and $\boldsymbol{D}$ are *similar*. This means that they have the same properties, as we show below. What happens is that the columns of $\boldsymbol{D}$ are columns of $\boldsymbol{A}$ but expressed in a *different basis* in $\mathbb{R}^n$.

- **The case $n = 2$.** We have shown that for a generic 2×2 matrix $\boldsymbol{A}$, the i-th eigenvector, associated to the i-th eigenvalue λ_i ($i = 1,2$), takes the form $\boldsymbol{v}_i = (\beta_i t, t)'$, where

$$\beta_i = -\left(\frac{a_{12}}{a_{11} - \lambda_i}\right) = -\left(\frac{a_{22} - \lambda_i}{a_{21}}\right)\,.$$

With no loss of generality, we set $t = 1$ to check whether the matrices

$$\boldsymbol{P} = \begin{bmatrix} \beta_1 & \beta_2 \\ 1 & 1 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{P}^{-1} = \frac{1}{\beta_1 - \beta_2}\begin{bmatrix} 1 & -\beta_2 \\ -1 & \beta_1 \end{bmatrix}$$

diagonalize $\boldsymbol{A}$. Note that

$$\begin{aligned}\boldsymbol{P}^{-1}\boldsymbol{AP} &= \frac{1}{\beta_1 - \beta_2}\begin{bmatrix} 1 & -\beta_2 \\ -1 & \beta_1 \end{bmatrix}\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} \beta_1 & \beta_2 \\ 1 & 1 \end{bmatrix} \\ &= \frac{1}{\beta_1 - \beta_2}\begin{bmatrix} a_{11} - \beta_2 a_{21} & a_{12} - \beta_2 a_{22} \\ -(a_{11} - \beta_1 a_{21}) & -(a_{12} - \beta_1 a_{22}) \end{bmatrix}\begin{bmatrix} \beta_1 & \beta_2 \\ 1 & 1 \end{bmatrix}.\end{aligned}$$

The expressions are quite involved but can be simplified considerably. Firstly, a property shown below but can be easily deduced from the characteristic polynomial is that $a_{11} + a_{22} = \lambda_1 + \lambda_2$. Secondly, using also the definitions of β_1 and β_2, for $j \neq i$,

$$\begin{aligned} a_{11} - \beta_i a_{21} &= a_{11} - a_{21}\left(-\frac{a_{22} - \lambda_i}{a_{21}}\right) = a_{11} + a_{22} - \lambda_i = \lambda_j\,, \\ a_{12} - \beta_i a_{22} &= a_{12} - a_{22}\left(-\frac{a_{12}}{a_{11} - \lambda_2}\right) = a_{12}\left(\frac{a_{11} - \lambda_i + a_{22}}{a_{11} - \lambda_i}\right) = \left(\frac{a_{12}}{a_{11} - \lambda_i}\right)\lambda_j = -\beta_i \lambda_j\,. \end{aligned}$$

Then,

$$\boldsymbol{P}^{-1}\boldsymbol{AP} = \frac{1}{\beta_1 - \beta_2}\begin{bmatrix} \lambda_1 & -\beta_2\lambda_1 \\ -\lambda_2 & \beta_1\lambda_2 \end{bmatrix}\begin{bmatrix} \beta_1 & \beta_2 \\ 1 & 1 \end{bmatrix} = \frac{1}{\beta_1 - \beta_2}\begin{bmatrix} \lambda_1(\beta_1 - \beta_2) & \lambda_1(\beta_2 - \beta_2) \\ \beta_1(\lambda_2 - \lambda_2) & \lambda_2(\beta_1 - \beta_2) \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix},$$

which equals $\boldsymbol{D}$, as expected.

- **When is A (not) diagonalizable?** Even though the equality $\boldsymbol{AP} = \boldsymbol{PD}$ applies generally, it is often understood that in order for $\boldsymbol{D}$ to exist, $\boldsymbol{P}$ must nonsingular. Otherwise, $\boldsymbol{A}$ is *not diagonalizable*.

  Define the vector subspace $\mathcal{E}(\lambda) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{Ax} = \lambda\boldsymbol{x}\}$, for a given λ, such that if $\boldsymbol{v} \in \mathcal{E}(\lambda)$, then $\boldsymbol{v}$ is an eigenvector of $\boldsymbol{A}$ associated with the eigenvalue λ. We know that $\boldsymbol{P}$ will be nonsingular if *all n eigenvalues form a linearly independent set*. Two possible cases emerge.

1. **Different eigenvalues**. The case where all eigenvalues of $\boldsymbol{A}$ are different is trivial, since each eigenvector belongs to a *different* vector subspace in $\mathbb{R}^n$ and, therefore, is linearly independence from the rest. Moreover, in this case $\dim(\mathcal{E}_\lambda) = 1$, and the eigenvalue is a basis of $\mathcal{E}_\lambda$.
2. **Repeated eigenvalues**. Suppose that λ appears r times as a root of the characteristic equation of $\boldsymbol{A}$. We say that the *algebraic multiplicity* of λ is r. The dimension of the subspace $\mathcal{E}_\lambda$ is called the *geometric multiplicity*, and it is always the case that $\dim(\mathcal{E}_\lambda) \leq r$.

   Each eigenvalue must be associated to a linearly independent eigenvector for diagonalization. If $\dim(\mathcal{E}_\lambda) < r$, then the subspace $\mathcal{E}_\lambda$ is spanned by less than r vectors and we are not able to obtain the required r eigenvectors. At least one of the eigenvectors associated to λ will be a linear combination of the basis, making $\boldsymbol{P}$ nonsingular.

   On the other hand, if $\dim(\mathcal{E}_\lambda) = r$, then the r vectors that form the basis of $\mathcal{E}_\lambda$ are the linearly independent eigenvectors we are looking for.

   A useful result is the following:

   $$\rho(\boldsymbol{A} - \lambda \boldsymbol{I}_n) = n - \dim(\mathcal{E}_\lambda) \qquad \text{or} \qquad \dim(\mathcal{E}_\lambda) = n - \rho(\boldsymbol{A} - \lambda \boldsymbol{I}_n)\,,$$

   and so $\boldsymbol{A}$ will be diagonalizable if $r = n - \rho(\boldsymbol{A} - \lambda \boldsymbol{I}_n)$, and not diagonalizable if $r > n - \rho(\boldsymbol{A} - \lambda \boldsymbol{I}_n)$.

In what follows, unless stated otherwise, we shall assume that $\boldsymbol{P}$ is nonsingular, and so that the eigenvalues of $\boldsymbol{A}$ are linearly independent.

- **Properties and special cases.**

(1) The eigenvalues of $\boldsymbol{A}'$ are equal to the eigenvalues of $\boldsymbol{A}$.

Proof. The characteristic polynomial of $\boldsymbol{A}$ is $P(\lambda) = |\,\boldsymbol{A} - \lambda \boldsymbol{I}_n\,|$. That of $\boldsymbol{A}'$ is

$$Q(\theta) = |\,\boldsymbol{A}' - \theta \boldsymbol{I}_n\,| = |\,\boldsymbol{A}' - \theta \boldsymbol{I}_n{}'\,| = |\,(\boldsymbol{A} - \theta \boldsymbol{I}_n)'\,| = |\,\boldsymbol{A} - \theta \boldsymbol{I}_n\,| = P(\theta)\,.$$

Both polynomials are identical and so are their roots, which are the eigenvalues.

Alternative proof. Consider the spectral decomposition $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$. Upon transposing both sides of this equality, $\boldsymbol{A}' = (\boldsymbol{P}')^{-1}\boldsymbol{D}'\boldsymbol{P}' = \boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^{-1}$, where $\boldsymbol{Q} = (\boldsymbol{P}')^{-1}$. This is the spectral decomposition of $\boldsymbol{A}'$ and features exactly the same diagonal matrix $\boldsymbol{D}$.

(2) The eigenvalues of $\boldsymbol{A}^{-1}$ are the reciprocal of the eigenvalues of $\boldsymbol{A}$.

Proof. Manipulate the definition:

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v} \quad \to \quad \boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{A}^{-1}\boldsymbol{v} \quad \to \quad \boldsymbol{v} = \lambda\boldsymbol{A}^{-1}\boldsymbol{v} \quad \to \quad \boldsymbol{A}^{-1}\boldsymbol{v} = \left(\frac{1}{\lambda}\right)\boldsymbol{v}\,.$$

Therefore, if λ is an eigenvalue of $\boldsymbol{A}$, λ^{-1} is an eigenvalue of $\boldsymbol{A}^{-1}$. Note that the associated eigenvector is the same in both cases.

Alternative proof. Consider the spectral decomposition $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$. Upon inverting both sides of the equality, $\boldsymbol{A}^{-1} = (\boldsymbol{P}^{-1})^{-1}\boldsymbol{D}^{-1}\boldsymbol{P}^{-1} = \boldsymbol{P}\boldsymbol{D}^{-1}\boldsymbol{P}^{-1}$. The diagonal matrix in the spectral decomposition of $\boldsymbol{A}^{-1}$ is $\boldsymbol{D}^{-1}$ and contains $1/\lambda_i$ along the diagonal. Note that matrix $\boldsymbol{P}$ is the same in both cases.

(3) The eigenvalues of $\boldsymbol{A}^k$ are the eigenvalues of $\boldsymbol{A}$ to the k-th power.

Proof. Manipulate the definition:

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v} \quad \to \quad \boldsymbol{A}\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{A}\boldsymbol{v} \quad \to \quad \boldsymbol{A}^2\boldsymbol{v} = \lambda(\lambda\boldsymbol{v}) \quad \to \quad \boldsymbol{A}^2\boldsymbol{v} = \lambda^2\boldsymbol{v}\,.$$

$$\boldsymbol{A}^2\boldsymbol{v} = \lambda^2\boldsymbol{v} \quad \to \quad \boldsymbol{A}\boldsymbol{A}^2\boldsymbol{v} = \lambda^2\boldsymbol{A}\boldsymbol{v} \quad \to \quad \boldsymbol{A}^3\boldsymbol{v} = \lambda^2(\lambda\boldsymbol{v}) \quad \to \quad \boldsymbol{A}^3\boldsymbol{v} = \lambda^3\boldsymbol{v}\,.$$

Repeating this procedure k times, we conclude that $\boldsymbol{A}^k\boldsymbol{v} = \lambda^k\boldsymbol{v}$. Note that the associated eigenvector is the same in both cases.

Alternative proof. Consider the spectral decomposition $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$. Upon premultiplying by $\boldsymbol{A}$ we get $\boldsymbol{A}^2 = \boldsymbol{A}\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1} = \boldsymbol{P}\boldsymbol{D}^2\boldsymbol{P}^{-1}$. Premultiplying by $\boldsymbol{A}$ again gives $\boldsymbol{A}^3 = \boldsymbol{A}\boldsymbol{A}^2 = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}\boldsymbol{P}\boldsymbol{D}^2\boldsymbol{P}^{-1} = \boldsymbol{P}\boldsymbol{D}^3\boldsymbol{P}^{-1}$. Repeating this procedure k times, we obtain $\boldsymbol{A}^k = \boldsymbol{P}\boldsymbol{D}^k\boldsymbol{P}^{-1}$. The diagonal matrix in the spectral decomposition of $\boldsymbol{A}^k$ is $\boldsymbol{D}^k$ which contains λ_i^k along the diagonal. Matrix $\boldsymbol{P}$ is the same in both cases.

(4) The eigenvalues of an idempotent matrix $\boldsymbol{A}^2 = \boldsymbol{A}$ are equal to either 0 or 1.

Proof. Manipulate the definition:

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v} \quad \rightarrow \quad \boldsymbol{A}^2\boldsymbol{v} = \lambda^2\boldsymbol{v} \quad \rightarrow \quad (\boldsymbol{A}^2 = \boldsymbol{A}) \quad \rightarrow \quad \boldsymbol{A}\boldsymbol{v} = \lambda^2\boldsymbol{v} \quad \rightarrow \quad \lambda\boldsymbol{v} = \lambda^2\boldsymbol{v}\,.$$

Then, it follows that $\lambda = \lambda^2$ which only occurs for $\lambda = \{0, 1\}$.

(5) The eigenvalues of an involutory matrix $\boldsymbol{A}^2 = \boldsymbol{I}_n$ are equal to either 1 or -1.

Proof. Manipulate the definition:

$$\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v} \quad \rightarrow \quad \boldsymbol{A}^2\boldsymbol{v} = \lambda^2\boldsymbol{v} \quad \rightarrow \quad (\boldsymbol{A}^2 = \boldsymbol{I}_n) \quad \rightarrow \quad \boldsymbol{v} = \lambda^2\boldsymbol{v}\,.$$

It follows $(1 - \lambda^2)\boldsymbol{v} = \boldsymbol{0}$. Since $\boldsymbol{v} \neq \boldsymbol{0}$, it must be case the case that $\lambda^2 = 1$ or $\lambda = \pm 1$.

(6) **(Optional)** Let $\boldsymbol{A}$ be a orthogonal matrix, $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}' = \boldsymbol{I}_n$. Its determinant is ± 1. If real, its eigenvalues are also equal to ± 1. In general, the modulus of its eigenvalues is equal to 1.

Proof. Note that $\det(\boldsymbol{A}'\boldsymbol{A}) = \det(\boldsymbol{I}_n)$ so $\det(\boldsymbol{A})^2 = 1$. It follows that $\det(\boldsymbol{A}) = \pm 1$.
On the other hand, upon taking modulus to both sides of the definition $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$,

$$\|\boldsymbol{A}\boldsymbol{v}\| = \|\lambda\boldsymbol{v}\| \quad \rightarrow \quad \sqrt{\boldsymbol{v}'\boldsymbol{A}'\boldsymbol{A}\boldsymbol{v}} = \|\lambda\boldsymbol{v}\| \quad \rightarrow \quad \sqrt{\boldsymbol{v}'\boldsymbol{v}} = \|\lambda\boldsymbol{v}\| \quad \rightarrow \quad \|\boldsymbol{v}\| = |\lambda|\,\|\boldsymbol{v}\|\,.$$

It follows that $|\lambda| = 1$, i.e. the modulus of λ is equal to one. Thus, in general, $\lambda = e^{\pm\theta i}$ for some argument θ. If λ is real, then it must be $\lambda = \pm 1$.

(7) $\operatorname{tr}(\boldsymbol{A}) = \displaystyle\sum_{i=1}^{n} \lambda_i$.

Proof. This property is a manifestation that $\boldsymbol{A}$ and $\boldsymbol{D}$ are similar. Upon taking the trace operator to the spectral decomposition, and recalling that under the trace the matrix multiplication is cyclically commutative, $\operatorname{tr}(\boldsymbol{A}) = \operatorname{tr}(\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}) = \operatorname{tr}(\boldsymbol{D}\boldsymbol{P}^{-1}\boldsymbol{P}) = \operatorname{tr}(\boldsymbol{D})$. The trace of $\boldsymbol{A}$ is equal to the trace of $\boldsymbol{D}$, which in turn equals the sum of all eigenvalues of $\boldsymbol{A}$.

(8) $|\,\boldsymbol{A}\,| = \displaystyle\prod_{i=1}^{n} \lambda_i$.

Proof. Another manifestation that $\boldsymbol{A}$ and $\boldsymbol{D}$ are similar. Upon taking the determinants to the spectral decomposition, and recalling that $|\,\boldsymbol{P}\,| \cdot |\,\boldsymbol{P}^{-1}\,| = 1$, $|\,\boldsymbol{A}\,| = |\,\boldsymbol{P}\,| \cdot |\,\boldsymbol{D}\,| \cdot |\,\boldsymbol{P}^{-1}\,| = |\,\boldsymbol{D}\,|$. The determinant of $\boldsymbol{A}$ is equal to the determinant of $\boldsymbol{D}$. Since $\boldsymbol{D}$ is diagonal, its determinant is the product of the diagonal entries, which in turn equals the product of all eigenvalues of $\boldsymbol{A}$.

(9) The rank of $\boldsymbol{A}$ is equal to the number of its nonzero eigenvalues.

Proof. Yet another manifestation that $\boldsymbol{A}$ and $\boldsymbol{D}$ are similar. From the spectral decomposition of $\boldsymbol{A}$, $\rho(\boldsymbol{A}) = \min\{\rho(\boldsymbol{P}\boldsymbol{D}), \rho(\boldsymbol{P}^{-1})\} = \rho(\boldsymbol{P}\boldsymbol{D}) = \min\{\rho(\boldsymbol{P}), \rho(\boldsymbol{D})\} = \rho(\boldsymbol{D})$. These equalities follow from the properties of the rank, and from the fact that $\boldsymbol{P}$ is full rank (nonsingular). The rank of $\boldsymbol{A}$ is the same as the rank of $\boldsymbol{D}$. Now, since $\boldsymbol{D}$ is diagonal, its rank is equal to the number of nonzero columns, which in turn is equal to the number of nonzero eigenvalues of $\boldsymbol{A}$.

(10) If $\boldsymbol{A}$ is singular, at least one of its eigenvalues is equal to zero.

Proof. If $\boldsymbol{A}$ is singular, then $|\,\boldsymbol{A}\,| = 0$. From property (8) we have that $\prod_{i=1}^{n} \lambda_i = 0$ which happens if at least one $\lambda_i = 0$.

Alternative proof. If $\boldsymbol{A}$ is singular, then it has less than full rank. From property (9), the rank is equal to the number of $\lambda_i \neq 0$. So, if any $\lambda_i = 0$, necessarily $\boldsymbol{A}$ will not be full rank.

(11) If $\boldsymbol{A}$ is nonsingular, then all of its eigenvalues are different from zero.

Proof. If $\boldsymbol{A}$ is nonsingular, then $|\boldsymbol{A}| \neq 0$. From property (8), $\prod_{i=1}^{n} \lambda_i \neq 0$ which happens only if $\lambda_i \neq 0$ for all i.

Alternative proof. If $\boldsymbol{A}$ is nonsingular, then it has full rank. From property (9), the rank is equal to the number of $\lambda_i \neq 0$. $\rho(\boldsymbol{A}) = n$ implies necessarily that $\lambda_i \neq 0$ for all i.

(12) If $\boldsymbol{A}$ is idempotent ($\boldsymbol{A}^2 = \boldsymbol{A}$), then $\rho(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A})$.

Proof. An implication of properties (4) and (9), since $\mathrm{tr}(\boldsymbol{A}) =$ Number of $\lambda_i = 1$ ($\lambda_i \neq 0$).

(13) Suppose that $\boldsymbol{A}$ is symmetric $\boldsymbol{A} = \boldsymbol{A}'$ and that all of its eigenvalues are different. Then, its eigenvectors are mutually orthogonal. An implication is that $\boldsymbol{P}$ is an orthogonal matrix and, thus, the spectral decomposition simplifies to $\boldsymbol{A} = \boldsymbol{PDP}'$ or $\boldsymbol{D} = \boldsymbol{P}'\boldsymbol{AP}$.

Proof. By definition, $\boldsymbol{Av}_i = \lambda_i \boldsymbol{v}_i$. After premultiplying this equality by $\boldsymbol{v}_j'$, where $i \neq j$, we get $Q_1 = \boldsymbol{v}_j'\boldsymbol{Av}_i = \lambda_i(\boldsymbol{v}_j'\boldsymbol{v}_i)$. Analogously, upon premultiplying $\boldsymbol{Av}_j = \lambda_j \boldsymbol{v}_j$ by $\boldsymbol{v}_i'$ we get $Q_2 = \boldsymbol{v}_i'\boldsymbol{Av}_j = \lambda_j(\boldsymbol{v}_i'\boldsymbol{v}_j)$.

It follows that $Q_1' = \boldsymbol{v}_i'\boldsymbol{A}'\boldsymbol{v}_j = \boldsymbol{v}_i'\boldsymbol{Av}_j = Q_2$, since $\boldsymbol{A}$ is symmetric. Moreover, Q_1 is a scalar, so $Q_1' = Q_1$, so we finally establish that $Q_1 = Q_2$. Therefore, $\lambda_i(\boldsymbol{v}_j'\boldsymbol{v}_i) = \lambda_j(\boldsymbol{v}_j'\boldsymbol{v}_i)$; but since $\lambda_i \neq \lambda_j$, the only way to enforce the equality is that $\boldsymbol{v}_j'\boldsymbol{v}_i = 0$.

2.4 Quadratic forms and definite matrices

- **Quadratic form.** Any *polynominal* function $Q : \mathbb{R}^n \to \mathbb{R}$ of n variables, which is homogenous of degree 2.

- **Matrix representation.** For $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{x} \neq \boldsymbol{0}$, and an $n \times n$ matrix $\boldsymbol{A} = [a_{ij}]$,

$$Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{Ax} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j$$

is a quadratic form. Clearly, $Q(\boldsymbol{x})$ is a scalar. It is homogenous of degree 2 since $Q(\lambda\boldsymbol{x}) = \lambda^2 Q(\boldsymbol{x})$ for any $\lambda \in \mathbb{R}$.

- **Symmetry.** Without loss of generality, it could be assumed that the coefficient matrix $\boldsymbol{A}$ of a quadratic form is symmetric. Indeed, $Q(\boldsymbol{x})$ can be defined as

$$Q(\boldsymbol{x}) = \boldsymbol{x}'\left(\frac{\boldsymbol{A}+\boldsymbol{A}'}{2}\right)\boldsymbol{x} = \frac{\boldsymbol{x}'\boldsymbol{Ax}}{2} + \frac{\boldsymbol{x}'\boldsymbol{A}'\boldsymbol{x}}{2} = \frac{\boldsymbol{x}'\boldsymbol{Ax}}{2} + \left(\frac{\boldsymbol{x}'\boldsymbol{Ax}}{2}\right)' = \boldsymbol{x}'\boldsymbol{Ax}\,.$$

The penultimate equality uses the fact that the transpose of a scalar is the scalar itself. The matrix $(\boldsymbol{A}+\boldsymbol{A}')/2$ is symmetric by construction, and renders the same value of $Q(\boldsymbol{x})$ than the matrix $\boldsymbol{A}$.

- **Diagonalization.** A quadratic form can be written as a weighted sum of squares. From the spectral decomposition of $\boldsymbol{A}$ we have that $\boldsymbol{A} = \boldsymbol{PDP}'$, where $\boldsymbol{D}$ is diagonal. Let $\boldsymbol{z} = \boldsymbol{P}'\boldsymbol{x}$. Since $\boldsymbol{x}$ is arbitrary and $\boldsymbol{P}$ is nonsingular, it turns out that $\boldsymbol{z}$ is arbitrary. Then,

$$Q(\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{Ax} = (\boldsymbol{x}'\boldsymbol{P})\boldsymbol{D}(\boldsymbol{P}'\boldsymbol{x}) = \boldsymbol{z}'\boldsymbol{Dz} = \sum_{i=1}^{n} \lambda_i z_i^2\,.$$

- **Definite matrices.** Let $\boldsymbol{x}$ be an *arbitrary* vector in $\mathbb{R}^n$.
  - $\boldsymbol{A}$ is *positive definite* ($\boldsymbol{A} > \boldsymbol{0}$) if $\boldsymbol{x}'\boldsymbol{Ax} > 0$.
  - $\boldsymbol{A}$ is *positive semidefinite* or *nonnegative definite* ($\boldsymbol{A} \geq \boldsymbol{0}$) if $\boldsymbol{x}'\boldsymbol{Ax} \geq 0$.

- $\boldsymbol{A}$ is *negative definite* ($\boldsymbol{A} \prec \mathbf{0}$) if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} < 0$.
- $\boldsymbol{A}$ is *negative semidefinite* or *nonpositive definite* ($\boldsymbol{A} \preceq \mathbf{0}$) if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \leq 0$.
- $\boldsymbol{A}$ is *indefinite* if $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0$ for some vectors $\boldsymbol{x}$, and $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \leq 0$ for others.

Important. The above inequalities occur for the quadratic form as a whole, and not for the individual entries in matrix $\boldsymbol{A}$. For instance, a positive definite matrix can contain negative entries.

- **Usage.** How "big" a matrix is relative to another. In the scalar case we have that if and only if $A > B$, then $x'Ax = x^2A > x'Bx = x^2B$. Also, the sign of $x'Ax = x^2A$ is that of A. In a multivariate world, $\boldsymbol{A} \succ \boldsymbol{B}$ (say, "$\boldsymbol{A}$ is larger than $\boldsymbol{B}$") if and only if $\boldsymbol{x}'(\boldsymbol{A} - \boldsymbol{B})\boldsymbol{x}$ is positive definite, and the sign of $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is tightly related to the definition of $\boldsymbol{A}$.

- **Identification.** Let $\mathcal{A}_i$ denote the i-th leading principal minor of $\boldsymbol{A}$ for $i = 1, 2, \ldots, n$; i.e., $\mathcal{A}_i$ is the determinant of the submatrix formed by the block that starts in a_{11} and ends in a_{ii}. Then,
  - If $\boldsymbol{A} \succ \mathbf{0}$, then $\mathcal{A}_1 > 0, \mathcal{A}_2 > 0, \mathcal{A}_3 > 0, \ldots, \mathcal{A}_n = |\boldsymbol{A}| > 0$.
  - If $\boldsymbol{A} \succeq \mathbf{0}$, then $\mathcal{A}_1 \geq 0, \mathcal{A}_2 \geq 0, \mathcal{A}_3 \geq 0, \ldots, \mathcal{A}_n = |\boldsymbol{A}| = 0$.
  - If $\boldsymbol{A} \prec \mathbf{0}$, then $\mathcal{A}_i < 0$ for i odd and $\mathcal{A}_i > 0$ for i even.
  - If $\boldsymbol{A} \preceq \mathbf{0}$, then $\mathcal{A}_i \leq 0$ for i odd, $\mathcal{A}_i \geq 0$ for i even and $\mathcal{A}_n = |\boldsymbol{A}| = 0$.

- **Properties.**

  (1) If $\boldsymbol{A}$ is semidefinite (positive or negative), then it is singular.

  Proof. Consider the case $\boldsymbol{A} \succeq \mathbf{0}$. Then, for some vectors $\boldsymbol{x}$, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ whereas for others $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = 0$. In the latter case, since $\boldsymbol{x} \neq \mathbf{0}$, it must be the case that $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$. Therefore, $\boldsymbol{A}$ must have linearly dependent rows, and is singular.

  (2) If $\boldsymbol{A} \succ \mathbf{0}$ [resp., $\boldsymbol{A} \succeq \mathbf{0}$], then $-\boldsymbol{A} \prec \mathbf{0}$ [resp., $-\boldsymbol{A} \preceq \mathbf{0}$].

  Consider the case $\boldsymbol{A} \succ \mathbf{0}$. Since $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$, then $\boldsymbol{x}'(-\boldsymbol{A})\boldsymbol{x} < 0$, which happens if $-\boldsymbol{A} \prec \mathbf{0}$. For semidefinite matrices, the inequalities are not strict.

  (3) If $\boldsymbol{A} \succ \mathbf{0}$ [resp., $\boldsymbol{A} \prec \mathbf{0}$] all of its diagonal entries a_{ii} are positive [resp., negative].
  If $\boldsymbol{A} \succeq \mathbf{0}$ [resp., $\boldsymbol{A} \preceq \mathbf{0}$] all of its diagonal elements a_{ii} are nonnegative [resp., nonpositive].

  Proof. Consider the case $\boldsymbol{A} \succ \mathbf{0}$. For *any* vector $\boldsymbol{x}$, it is true that $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$. If we take $\boldsymbol{x} = \boldsymbol{e}_i$ (the i-th canonical unit vector), then $\boldsymbol{e}_i{}'\boldsymbol{A}\boldsymbol{e}_i = a_{ii} > 0$. The same argument leads us to conclude that $a_{ii} < 0$ when $\boldsymbol{A} \prec \mathbf{0}$. For semidefinite matrices, the inequalities are not strict.

  (4) If $\boldsymbol{A} \succ \mathbf{0}$ [resp., $\boldsymbol{A} \prec \mathbf{0}$] all its eigenvalues are positive [resp., negative].
  If $\boldsymbol{A} \succeq \mathbf{0}$ [resp., $\boldsymbol{A} \preceq \mathbf{0}$] all its eigenvalues are positive [resp., negative] and at least one is zero.

  Proof. Consider the case $\boldsymbol{A} \succ \mathbf{0}$, such that $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$ for any vector $\boldsymbol{x}$. If we take $\boldsymbol{x} = \boldsymbol{v}$, the eigenvector of $\boldsymbol{A}$ associated to the eigenvalue λ, we get $\boldsymbol{v}'\boldsymbol{A}\boldsymbol{v} = \boldsymbol{v}'(\lambda\boldsymbol{v}) = \lambda\|\boldsymbol{v}\|^2$. This expression is positive if and only if $\lambda > 0$. The same argument enables us to conclude that $\lambda < 0$ when $\boldsymbol{A} \prec \mathbf{0}$. By property (1), semidefinite matrices are singular, and so they must contain at least one $\lambda = 0$.

  (5) If all eigenvalues of $\boldsymbol{A}$ are positive [resp., negative], then $\boldsymbol{A} \succ \mathbf{0}$ [resp., $\boldsymbol{A} \prec \mathbf{0}$].
  If all eigenvalues of $\boldsymbol{A}$ are nonnegative [resp., nonpositive], then $\boldsymbol{A} \succeq \mathbf{0}$ [resp., $\boldsymbol{A} \preceq \mathbf{0}$].

  Proof. Consider the case $\boldsymbol{A} \succ \mathbf{0}$. Since $\boldsymbol{A}$ is symmetric, its spectral decomposition is $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'$, where $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_n$. The diagonal matrix $\boldsymbol{D}$, as a premise, contains exclusively positive entries along its diagonal. Therefore, $\boldsymbol{z}'\boldsymbol{D}\boldsymbol{z} > 0$ for any vector $\boldsymbol{z}$ (that is, $\boldsymbol{D} \succ \mathbf{0}$). Now define $\boldsymbol{z} = \boldsymbol{P}'\boldsymbol{x}$. Then, $\boldsymbol{z}'\boldsymbol{D}\boldsymbol{z} = (\boldsymbol{P}'\boldsymbol{x})'\boldsymbol{D}(\boldsymbol{P}'\boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}'\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0$. Note $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{z}$ is arbitrary since $\boldsymbol{z}$ is arbitrary and $\boldsymbol{P}$ is full rank.
  The same argument leads us to conclude that if $\lambda < 0$ (that is, $\boldsymbol{D} \prec \mathbf{0}$), then $\boldsymbol{A} \prec \mathbf{0}$. For semidefinite matrices, the inequalities are not strict, since at least one $\lambda = 0$.

(6) If $\boldsymbol{A} \succ \mathbf{0}$ then $|\boldsymbol{A}| > 0$ and $\mathrm{tr}(\boldsymbol{A}) > 0$.
If $\boldsymbol{A} \succeq \mathbf{0}$ then $|\boldsymbol{A}| = 0$ and $\mathrm{tr}(\boldsymbol{A}) \geq 0$.
If $\boldsymbol{A} \prec \mathbf{0}$ then $|\boldsymbol{A}| < 0$ for n odd, $|\boldsymbol{A}| > 0$ for n even and $\mathrm{tr}(\boldsymbol{A}) > 0$.
If $\boldsymbol{A} \preceq \mathbf{0}$ then $|\boldsymbol{A}| = 0$ and $\mathrm{tr}(\boldsymbol{A}) \leq 0$.

Proof. These properties follow from properties previously proved. From (1), we know that semidefinite matrices are singular and, hence, $|\boldsymbol{A}| = 0$. Moreover, for the null matrix $\mathbf{0}$, which is semidefinite, we have that $\mathrm{tr}(\boldsymbol{A}) = 0$.
On the other hand, from (3) we know that the n eigenvalues of $\boldsymbol{A} \succ \mathbf{0}$ are positive. Its determinant is then the product of n positive numbers and, therefore, $|\boldsymbol{A}| > 0$. The trace is the sum of n positive numbers and, therefore, $\mathrm{tr}(\boldsymbol{A}) > 0$. Similarly, the n eigenvalues of matrix $\boldsymbol{A} \prec \mathbf{0}$ are negative. The determinant is the product of n negative numbers and, therefore, $|\boldsymbol{A}| > 0$ if n is even and $|\boldsymbol{A}| < 0$ if n is odd. The trace is the sum of n negative number and, thus, $\mathrm{tr}(\boldsymbol{A}) < 0$.

(7) $\boldsymbol{A} \succ \mathbf{0}$ [resp., $\boldsymbol{A} \prec \mathbf{0}$] if and only if $\boldsymbol{A}^{-1} \succ \mathbf{0}$ [resp., $\boldsymbol{A}^{-1} \prec \mathbf{0}$].

Proof. Consider the case $\boldsymbol{A} \succ \mathbf{0}$. We have that

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0 \quad \rightarrow \quad \boldsymbol{x}'\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{x} > 0 \quad \rightarrow \quad (\boldsymbol{A}\boldsymbol{x})'\boldsymbol{A}^{-1}(\boldsymbol{A}\boldsymbol{x}) > 0 \quad \rightarrow \quad \boldsymbol{z}'\boldsymbol{A}^{-1}\boldsymbol{z} > 0\,,$$

so we conclude that $\boldsymbol{A}^{-1} \succ \mathbf{0}$. The converse (start with $\boldsymbol{A}^{-1} \succ \mathbf{0}$ and conclude $\boldsymbol{A} \succ \mathbf{0}$) is also trivially true. The same analysis holds to establish the results for negative definite matrices.

(8) Let $\boldsymbol{A} = \boldsymbol{B}'\boldsymbol{B}$, where $\boldsymbol{B}$ is an $m \times n$ matrix. If $\boldsymbol{B}$ is full rank [resp., is not], then $\boldsymbol{A} \succ \mathbf{0}$ [resp., $\boldsymbol{A} \succeq \mathbf{0}$].

Proof. Note that $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{B}'\boldsymbol{B}\boldsymbol{x} = (\boldsymbol{B}\boldsymbol{x})'(\boldsymbol{B}\boldsymbol{x}) = \|\boldsymbol{B}\boldsymbol{x}\|^2$. If $\boldsymbol{B}$ is not full rank, then $\boldsymbol{B}\boldsymbol{x} = \mathbf{0}$ for some vectors $\boldsymbol{x}$. In these cases, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \|\mathbf{0}\|^2 = 0$. If $\boldsymbol{B}\boldsymbol{x} \neq \mathbf{0}$, then $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \|\boldsymbol{B}\boldsymbol{x}\|^2 > 0$. If $\boldsymbol{B}$ is full rank, it is always the case that $\boldsymbol{B}\boldsymbol{x} \neq \mathbf{0}$.

(9) If $\boldsymbol{A} \succ \mathbf{0}$, then there exists a matrix $\boldsymbol{B} \succ \mathbf{0}$ such that $\boldsymbol{B}^2 = \boldsymbol{A}$. If $\boldsymbol{A} \succeq \mathbf{0}$, then $\boldsymbol{B} \succeq \mathbf{0}$. Matrix $\boldsymbol{B}$ is often known as the *square root of* $\boldsymbol{A}$ and is often denoted as $\boldsymbol{A}^{1/2}$.

Proof. Let $\boldsymbol{D}^{1/2}$ be a diagonal matrix whose entries are the (positive) square roots of the corresponding entries of $\boldsymbol{D}$. From the spectral decomposition, $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}' = \boldsymbol{P}\boldsymbol{D}^{1/2}\boldsymbol{D}^{1/2}\boldsymbol{P}' = \boldsymbol{P}\boldsymbol{D}^{1/2}\boldsymbol{P}'\boldsymbol{P}\boldsymbol{D}^{1/2}\boldsymbol{P}' = \boldsymbol{B}\boldsymbol{B} = \boldsymbol{B}^2$, where $\boldsymbol{B} = \boldsymbol{P}\boldsymbol{D}^{1/2}\boldsymbol{P}'$. Note that $\boldsymbol{B}$ is symmetric so that $\boldsymbol{A} = \boldsymbol{B}^2 = \boldsymbol{B}'\boldsymbol{B}$. The distinction between $\boldsymbol{B} \succ \mathbf{0}$ and $\boldsymbol{B} \succeq \mathbf{0}$ depends on whether $\boldsymbol{D}$ has (at least) a zero entry along its diagonal.

(10) $\boldsymbol{B}'\boldsymbol{A}\boldsymbol{B}$, where $\boldsymbol{B}$ is a full rank square matrix, is defined as $\boldsymbol{A}$.

Proof. Note that $\boldsymbol{x}'\boldsymbol{B}'\boldsymbol{A}\boldsymbol{B}\boldsymbol{x} = \boldsymbol{z}'\boldsymbol{A}\boldsymbol{z}$ where $\boldsymbol{z} = \boldsymbol{B}\boldsymbol{x}$. Since $\boldsymbol{B}$ is full rank and $\boldsymbol{x} \neq \mathbf{0}$ is arbitrary, then $\boldsymbol{z} \neq \mathbf{0}$ is arbitrary. The sign of $\boldsymbol{x}'\boldsymbol{B}'\boldsymbol{A}\boldsymbol{B}\boldsymbol{x}$ is then the sign of $\boldsymbol{z}'\boldsymbol{A}\boldsymbol{z}$, and depends on the definition of $\boldsymbol{A}$.

2.5 Partitioned matrices (Optional)

- **Definition and use.** To partition a matrix is to divide or segment it into lower dimension submatrices or blocks, in order to ease its manipulation.

  For instance, an $m \times n$ matrix $\boldsymbol{A}$ can be written, for $r < m$ and $s < n$, as

$$\boldsymbol{A} = \begin{bmatrix} \underset{r\times s}{\boldsymbol{A}_{11}} & \underset{r\times(n-s)}{\boldsymbol{A}_{12}} \\ \underset{(m-r)\times s}{\boldsymbol{A}_{21}} & \underset{(m-r)\times(n-s)}{\boldsymbol{A}_{22}} \end{bmatrix}, \quad \text{or as} \quad \boldsymbol{A} = \begin{bmatrix} \underset{r\times n}{\boldsymbol{A}_{11}} \\ \underset{(m-r)\times n}{\boldsymbol{A}_{21}} \end{bmatrix}, \quad \text{or as} \quad \boldsymbol{A} = \begin{bmatrix} \underset{m\times s}{\boldsymbol{A}_{11}} & \underset{m\times(n-s)}{\boldsymbol{A}_{12}} \end{bmatrix}.$$

  In the first case, the matrix is *square by blocks* or *block square* (note that $\boldsymbol{A}$ itself is not necessarily square), whereas in the second and third cases the matrix is a *vector by blocks* or *block vector*.

- **Equality.** For two matrices to be equal $\boldsymbol{A}$ y $\boldsymbol{B}$, they must be partitioned in the same way, and the corresponding blocks should be equal:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} \qquad \text{y} \qquad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} \end{bmatrix}.$$

Then, $\boldsymbol{A} = \boldsymbol{B}$ if and only of $\boldsymbol{A}_{11} = \boldsymbol{B}_{11}$, $\boldsymbol{A}_{12} = \boldsymbol{B}_{12}$, $\boldsymbol{A}_{21} = \boldsymbol{B}_{21}$ and $\boldsymbol{A}_{22} = \boldsymbol{B}_{22}$.

- **Addition.** In order to sum two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, they must be partitioned in the same way:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \end{bmatrix} \quad \text{y} \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} \end{bmatrix} \qquad \rightarrow \qquad \boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} \boldsymbol{A}_{11} + \boldsymbol{B}_{11} & \boldsymbol{A}_{12} + \boldsymbol{B}_{12} \end{bmatrix}.$$

- **Transposition.** It implies the transposition of the block matrix, as well as of the individual blocks:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \boldsymbol{A}_{13} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \boldsymbol{A}_{23} \end{bmatrix} \qquad \rightarrow \qquad \boldsymbol{A}' = \begin{bmatrix} \boldsymbol{A}_{11}' & \boldsymbol{A}_{21}' \\ \boldsymbol{A}_{12}' & \boldsymbol{A}_{22}' \\ \boldsymbol{A}_{13}' & \boldsymbol{A}_{23}' \end{bmatrix}.$$

- **Multiplication.** Two block matrices can be multiplied if they are conformable with matrix multiplication, and if they are partitioned in such way that the block structure of the columns of the premultiplying matrix coincides to the block structure of the rows of the postmultiplying matrix:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} & \boldsymbol{A}_{13} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} & \boldsymbol{A}_{23} \end{bmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_{11} \\ \boldsymbol{B}_{21} \\ \boldsymbol{B}_{22} \end{bmatrix} \quad \rightarrow \quad \boldsymbol{AB} = \begin{bmatrix} \boldsymbol{A}_{11}\boldsymbol{B}_{11} + \boldsymbol{A}_{12}\boldsymbol{B}_{21} + \boldsymbol{A}_{13}\boldsymbol{B}_{31} \\ \boldsymbol{A}_{21}\boldsymbol{B}_{11} + \boldsymbol{A}_{22}\boldsymbol{B}_{21} + \boldsymbol{A}_{23}\boldsymbol{B}_{31} \end{bmatrix}.$$

Note that in the manipulation of partitioned matrices, the blocks are treated as scalars. We just need to be careful and respect matrix conformability when it is required.

- **Inverse.** A *nonsingular* partitioned matrix can be inverted if it is block square.

  1. We establish a preliminary result for a block triangular matrix:

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{M}_1 & \boldsymbol{M}_2 \\ \boldsymbol{0} & \boldsymbol{M}_4 \end{bmatrix} \qquad \rightarrow \qquad \boldsymbol{M}^{-1} = \begin{bmatrix} \boldsymbol{M}_1^{-1} & -\boldsymbol{M}_1^{-1}\boldsymbol{M}_2\boldsymbol{M}_4^{-1} \\ \boldsymbol{0} & \boldsymbol{M}_4^{-1} \end{bmatrix}. \tag{*}$$

  Check:

$$\begin{bmatrix} \boldsymbol{M}_1 & \boldsymbol{M}_2 \\ \boldsymbol{0} & \boldsymbol{M}_4 \end{bmatrix}\begin{bmatrix} \boldsymbol{M}_1^{-1} & -\boldsymbol{M}_1^{-1}\boldsymbol{M}_2\boldsymbol{M}_4^{-1} \\ \boldsymbol{0} & \boldsymbol{M}_4^{-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{M}_1\boldsymbol{M}_1^{-1} & \boldsymbol{M}_2\boldsymbol{M}_4^{-1} - \boldsymbol{M}_1\boldsymbol{M}_1^{-1}\boldsymbol{M}_2\boldsymbol{M}_4^{-1} \\ \boldsymbol{0} & \boldsymbol{M}_4\boldsymbol{M}_4^{-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I}_{n_1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{n_2} \end{bmatrix}.$$

  2. To compute $\boldsymbol{A}^{-1}$ it is convenient to evoke two ancillary matrices such that $\bar{\boldsymbol{C}}\boldsymbol{A} = \bar{\boldsymbol{B}}$. Then, $\boldsymbol{A}^{-1} = \bar{\boldsymbol{B}}^{-1}\bar{\boldsymbol{C}}$ which comes on handy as long as it is easy to compute $\bar{\boldsymbol{B}}^{-1}$. Then, we propose

$$\bar{\boldsymbol{C}}\boldsymbol{A} = \bar{\boldsymbol{B}} \quad \rightarrow \quad \begin{bmatrix} \boldsymbol{I}_{n_1} & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{I}_{n_2} \end{bmatrix}\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix}.$$

  We have that $\boldsymbol{C}\boldsymbol{A}_{11} + \boldsymbol{A}_{21} = \boldsymbol{0}$ so $\boldsymbol{C} = -\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}$. Moreover, $\boldsymbol{B} = \boldsymbol{C}\boldsymbol{A}_{12} + \boldsymbol{A}_{22} = \boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}$.

  3. Then, using (*),

$$\boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{0} & \boldsymbol{B} \end{bmatrix}^{-1}\begin{bmatrix} \boldsymbol{I}_{n_1} & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{I}_{n_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11}^{-1} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{B}^{-1} \\ \boldsymbol{0} & \boldsymbol{B}^{-1} \end{bmatrix}\begin{bmatrix} \boldsymbol{I}_{n_1} & \boldsymbol{0} \\ \boldsymbol{C} & \boldsymbol{I}_{n_2} \end{bmatrix}.$$

4. Finally,

$$A^{-1}=\begin{bmatrix} A_{11}^{-1}+A_{11}^{-1}A_{12}(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1} \\ -(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} & (A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}\end{bmatrix}.$$

An slightly less cumbersome expression is

$$A^{-1}=\begin{bmatrix} A_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}+\begin{bmatrix} A_{11}^{-1}A_{12} \\ -I_{n_2}\end{bmatrix}(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1}\begin{bmatrix} A_{21}A_{11}^{-1} & -I_{n_1}\end{bmatrix}.$$

It is a good exercise to verify that $AA^{-1}=A^{-1}A=I_n$.

- **Alternative expression for the inverse.** We have implicitly assumed that block A_{11} is nonsingular. This needs not to be true. Following a similar procedure we can reach the following expression, that assumes instead that A_{22} is nonsingular:

$$A^{-1}=\begin{bmatrix} (A_{11}-A_{12}A_{22}^{-1}A_{21})^{-1} & -(A_{11}-A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}(A_{11}-A_{12}A_{22}^{-1}A_{21})^{-1} & A_{22}^{-1}+A_{22}^{-1}A_{21}(A_{11}-A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1}\end{bmatrix}.$$

An slightly less cumbersome expression is

$$A^{-1}=\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A_{22}^{-1}\end{bmatrix}+\begin{bmatrix} -I_{n_1} \\ A_{22}^{-1}A_{21}\end{bmatrix}(A_{11}-A_{12}A_{22}^{-1}A_{21})^{-1}\begin{bmatrix} -I_{n_2} & A_{12}A_{22}^{-1}\end{bmatrix}.$$

Again, it is a good exercise to verify that $AA^{-1}=A^{-1}A=I_n$.

- **Determinant.** We may use the "trick" $\bar{C}A=\bar{B}$ again, since $|\bar{C}|=1$ (which is clearly a triangular matrix) so that $|A|=|\bar{B}|$. It turns out that since $|\bar{B}|$ is block triangular, its determinant is equal to the product of the determinants of the diagonal blocks. We now verify this:

1. Consider the matrices

$$Z_*=\begin{bmatrix} I_{n_1} & \mathbf{0} \\ \mathbf{0} & B\end{bmatrix} \qquad \text{and} \qquad Z_{**}=\begin{bmatrix} A_{11} & \mathbf{0} \\ \mathbf{0} & I_{n_2}\end{bmatrix}.$$

For the first case, we note that the first row of Z_* is the first row of $I_{n_1+n_2}$. Then, we can factorize the first element, which is a leading 1, out the determinant and reduce the dimension of the matrix by one. Then, the determinant of Z_* equals $(-1)^2|Z_{(1)}|=|Z_{(1)}|$, where $Z_{(1)}$ is the $n-1$ square matrix that results from removing the first row and column of Z_*. The first row of $Z_{(1)}$, is also the first row of an identity matrix. Hence, we can proceed analogously and establish that $|Z_{(1)}|=(-1)^2|Z_{(2)}|=|Z_{(2)}|$, where $Z_{(2)}$ is the $n-2$ square matrix that results from removing the first row and column of $Z_{(1)}$. The first row of $Z_{(2)}$ is also the first row of an identity matrix, and the process is repeated. Following the recursion, we get that $|Z_*|=|Z_{(n_1)}|=|B|$.
A similar reasoning leads us to the result $|Z_{**}|=|A_{11}|$.

2. Let $Z_{***}=\begin{bmatrix} I_{n_1} & A_{12} \\ \mathbf{0} & I_{n_2}\end{bmatrix}$. Since Z_{***} is upper triangular with ones along the diagonal, $|Z_{***}|=1$.

3. Finally,

$$|\bar{B}|=\begin{vmatrix} A_{11} & A_{12} \\ \mathbf{0} & B\end{vmatrix}=\begin{vmatrix} I_{n_1} & \mathbf{0} \\ \mathbf{0} & B\end{vmatrix}\begin{vmatrix} I_{n_1} & A_{12} \\ \mathbf{0} & I_{n_2}\end{vmatrix}\begin{vmatrix} A_{11} & \mathbf{0} \\ \mathbf{0} & I_{n_2}\end{vmatrix}=|Z_*|\cdot|Z_{***}|\cdot|Z_{**}|=|A_{11}|\cdot|B|.$$

Then, if A_{11} is nonsingular,

$$|A|=|A_{11}|\cdot|A_{22}-A_{21}A_{11}^{-1}A_{12}|.$$

Alternatively, for A_{22} nonsingular,

$$|A|=|A_{22}|\cdot|A_{11}-A_{12}A_{22}^{-1}A_{21}|.$$

2.6 Kronecker product

- **Definition.** Let $\boldsymbol{A}$ be an $m \times n$ matrix and let $\boldsymbol{B}$ be an $r \times s$ matrix. The Kronecker prodcut between $\boldsymbol{A}$ and $\boldsymbol{B}$, which is denoted as $\boldsymbol{A} \otimes \boldsymbol{B}$, results in an $mr \times ns$ matrix with the following block structure:

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \cdots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \cdots & a_{2n}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \cdots & a_{mn}\boldsymbol{B} \end{bmatrix} . \tag{1}$$

- **Properties of the Kronecker product.**
  (1) Unlike the usual matrix product $\boldsymbol{AB}$, the Kronecker $\boldsymbol{A} \otimes \boldsymbol{B}$ always exists, regardless of the dimensions of $\boldsymbol{A}$ and $\boldsymbol{B}$.
  (2) Like the usual matrix product, the Kronecker product is not commutative: in general, $(\boldsymbol{A} \otimes \boldsymbol{B}) \neq (\boldsymbol{B} \otimes \boldsymbol{A})$.
  (3) $(\boldsymbol{A} + \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{C}$.
  (4) $(\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C})$.
  (5) $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{AC} \otimes \boldsymbol{BD}$ as long as $\boldsymbol{AC}$ y $\boldsymbol{BD}$ exist.
  (6) $(\boldsymbol{A} \otimes \boldsymbol{B})' = \boldsymbol{A}' \otimes \boldsymbol{B}'$.
  (7) $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$ as long as $\boldsymbol{A}$ and $\boldsymbol{B}$ are nonsingular.
  (8) $\text{tr}(\boldsymbol{A} \otimes \boldsymbol{B}) = \text{tr}(\boldsymbol{A})\text{tr}(\boldsymbol{B})$.
  (9) If $\boldsymbol{A}$ is $n \times n$ and $\boldsymbol{B}$ is $m \times m$, $|\,\boldsymbol{A} \otimes \boldsymbol{B}\,| = |\,\boldsymbol{A}\,|^m |\,\boldsymbol{B}\,|^n$.
  (10) The eigenvalues of $\boldsymbol{A} \otimes \boldsymbol{B}$ are formed as the eigenvalues of $\boldsymbol{A}$ times the eigenvalues of $\boldsymbol{B}$.

- **Vectorization (Optional).** If $\boldsymbol{A}$ is an $m \times n$ matrix, its vectorization is a linear operation that renders a $mn \times 1$ vector that stacks vertically the columns of $\boldsymbol{A}$:

$$\text{vec}(\boldsymbol{A}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} & a_{21} & a_{22} & \cdots & a_{2m} & \cdots & a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}' ,$$

where a_{ij} is the (i, j) entry of $\boldsymbol{A}$.

- **Properties of vectorization (Optional).**
  (1) $\text{vec}(\boldsymbol{A} + \boldsymbol{B}) = \text{vec}(\boldsymbol{A}) + \text{vec}(\boldsymbol{B})$.
  (2) $\text{vec}(\lambda\boldsymbol{A}) = \lambda\text{vec}(\boldsymbol{A})$, where λ is a scalar.
  (3) $\text{vec}(\boldsymbol{ABC}) = (\boldsymbol{C}' \otimes \boldsymbol{A})\text{vec}(\boldsymbol{B}) = (\boldsymbol{I}_{n_C} \otimes \boldsymbol{AB})\text{vec}(\boldsymbol{C}) = (\boldsymbol{C}'\boldsymbol{B}' \otimes \boldsymbol{I}_{m_A})\text{vec}(\boldsymbol{A})$, where n_C is the number of columns of $\boldsymbol{C}$ and m_A is the number of rows of $\boldsymbol{A}$.
  (4) $\text{tr}(\boldsymbol{AB}) = \text{vec}(\boldsymbol{A}')'\text{vec}(\boldsymbol{B}) = \text{vec}(\boldsymbol{B}')'\text{vec}(\boldsymbol{A})$.

- **Lyapunov equations (Optional).** The common use of the $\text{vec}(\cdot)$ operator is to manipulate matrix products. An interesting application is to solve the *Lyapunov equations*:

$$\boldsymbol{AX} + \boldsymbol{BX} = \boldsymbol{C} \quad \to \quad (\boldsymbol{I} \otimes \boldsymbol{A})\text{vec}(\boldsymbol{X}) + (\boldsymbol{B}' \otimes \boldsymbol{I})\text{vec}(\boldsymbol{X}) = \text{vec}(\boldsymbol{C}) \quad \to \quad \text{vec}(\boldsymbol{X}) = (\boldsymbol{I} \otimes \boldsymbol{A} + \boldsymbol{B}' \otimes \boldsymbol{I})^{-1}\text{vec}(\boldsymbol{C}) .$$

$$\boldsymbol{X} = \boldsymbol{AXA}' + \boldsymbol{C} \quad \to \quad \text{vec}(\boldsymbol{X}) = (\boldsymbol{A}' \otimes \boldsymbol{A})\text{vec}(\boldsymbol{X}) + \text{vec}(\boldsymbol{C}) \quad \to \quad \text{vec}(\boldsymbol{X}) = (\boldsymbol{I} - \boldsymbol{A}' \otimes \boldsymbol{A})^{-1}\text{vec}(\boldsymbol{C}) .$$