Linear Algebra – A Brief Reminder

Purpose.   The purpose of this document is to quickly refresh (presumably) known notions in linear algebra. It contains a collection of facts related to vectors, matrices, and geometric entities they represent that we will use heavily in our course. Even though this quick reminder may seem redundant or trivial to most of you (I hope), I still suggest at least to skim through it, as it might present less common ways of interpretation of very familiar definitions and properties. And even if you discover nothing new in this document, it will at least be useful to introduce notation.

1   Notation

In our course, we will deal almost exclusively with the field of real numbers, which we denote by $\mathbb{R}$. An n-dimensional Euclidean space will be denoted as $\mathbb{R}^n$, and the space of $m \times n$ matrices as $\mathbb{R}^{m \times n}$. We will try to stick to a consistent typesetting denoting a scalar a with lowercase italic, a vector $\mathbf{a}$ in lowercase bold, and a matrix $\mathbf{A}$ in uppercase bold. Elements of a vector or a matrix will be denoted using subindices as a_i and a_{ij}, respectively. Unless stated otherwise, a vector is a column vector, and we will write $\mathbf{a} = (a_1, \ldots, a_n)^{\mathrm{T}}$ to save space (here, $^{\mathrm{T}}$ is the transpose of the row vector $(a_1, \ldots, a_n)$). In the same way, $\mathbf{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_n)$ will refer to a matrix constructed from n column vectors $\mathbf{a}_i$. We will denote the zero vector by $\mathbf{0}$, and the identity matrix by $\mathbf{I}$.

2   Linear and affine spaces

Given a collection of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbf{R}^n\}$, a new vector $\mathbf{b} = a_1 \mathbf{v}_1 + \cdots + a_m \mathbf{v}_m$, with $a_1, \ldots, a_m \in \mathbb{R}$ some scalars, is called a *linear combination* of the $\mathbf{v}_i$'s. The collection of all linear combinations is called a *linear subspace* of $\mathbb{R}^n$, denoted by

$$\mathcal{L} = \operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_m\} = \{a_1 \mathbf{v}_1 + \cdots + a_m \mathbf{v}_m : a_1, \ldots, a_m \in \mathbb{R}\}.$$

We will say that the $\mathbf{v}_i$'s *span* the linear subspace $\mathcal{L}$.

A vector $\mathbf{u}$ that cannot be described as a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_m$ (i.e., $\mathbf{u} \notin \mathcal{L}$) is said to be *linearly independent* of the latter vectors. The collection of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_m\}$ is linearly independent if no $\mathbf{v}_i$ is a linear combination of the rest of the vectors. In such a case, a vector in $\mathcal{L}$ can be unambiguously defined by the set of m scalars $a_1, \ldots, a_m$; omitting any of these scalars will make the description incomplete. Formally, we say that m is the *dimension* of the subspace, $\dim \mathcal{L} = m$.
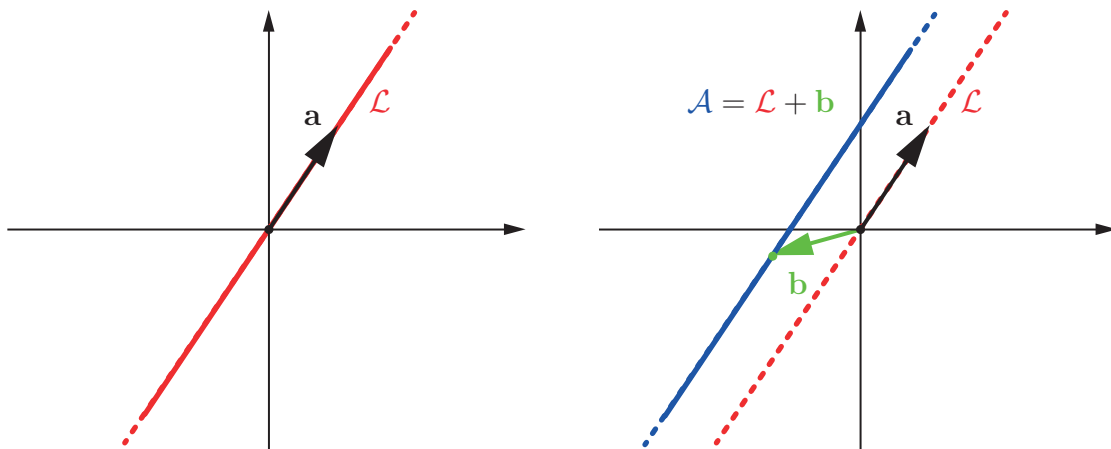
Figure 1: One-dimensional linear (left) and affine (right) subspaces of $\mathbb{R}^2$.

Geometrically, an m-dimensional subspace is an m-dimensional plane passing through the origin (a one-dimensional plane is a line, a two-dimensional plane is the regular plane, and so on). The latter is true, since setting all the a_i's to zero in the definition of $\mathcal{L}$ yields the zero vector. Figure 1 (left) visualizes this fact.

An *affine subspace* can be defined as a linear subspace shifted by a vector $\mathbf{b} \in \mathbb{R}^n$ away from the origin:

$$\mathcal{A} = \mathcal{L} + \mathbf{b} = \{\mathbf{u} + \mathbf{b} : \mathbf{u} \in \mathcal{L}\}.$$

Exercise 1. *Show that any affine subspace can be defined as*

$$\mathcal{A} = \{a_1\mathbf{v}_1 + \cdots + a_m\mathbf{v}_m : a_1 + \cdots + a_m = 1\}.$$

The latter linear combination with the scalars restricted to unit sum is called an affine combination.

3   Vector inner product and norms

Given two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$, we define their *inner* (a.k.s. *scalar* or *dot*) *product* as

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} u_i v_i = \mathbf{u}^{\mathrm{T}}\mathbf{v}.$$

Though the notion of an inner product is more general than this, we will limit our attention exclusively to this particular case (and its matrix equivalent defined in the sequel). This is sometimes called the standard or the Euclidean inner product. The inner product defines or *induces* a vector *norm*

$$\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$$

(it is also convenient to write $\|\mathbf{u}\|^2 = \mathbf{u}^\mathsf{T}\mathbf{u}$), which is called the Euclidean or the *induced* norm.

A more general notion of a norm can be introduced axiomatically. We will say that a non-negative scalar function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}_+$ is a norm if it satisfies the following axioms for any scalar $a \in \mathbb{R}$ and any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$

1. $\|a\mathbf{u}\| = |a|\|\mathbf{u}\|$ (this property is called *absolute homogeneity*);

2. $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$ (this property is called *subadditivity*, and since a norm induces a metric (distance function), the geometric name for it is *triangle inequality*);

3. $\|\mathbf{u}\| = 0$ iff[1] $\mathbf{u} = \mathbf{0}$.

In this course, we will almost exclusively restrict our attention to the following family of norms called the ℓ_p norms:

Definition. *For $1 \leq p < \infty$, the ℓ_p norm of a vector $\mathbf{u} \in \mathbb{R}^n$ is defined as*

$$\|\mathbf{u}\|_p = \left(\sum_{i=1}^n |u_i|^p \right)^{1/p}.$$

Note the sub-index $_p$. The Euclidean norm corresponds to $p = 2$ (hence, the alternative name, the ℓ_2 norm) and, whenever confusion may arise, we will denote it by $\|\cdot\|_2$.

Another important particular case of the ℓ_p family of norms is the ℓ_1 norm

$$\|\mathbf{u}\|_1 = \sum_{i=1}^n |u_i|.$$

As we will see, when used to quantify errors for example in a regression or model fitting task, the ℓ_1 norm (representing mean absolute error) is more robust (i.e., less sensitive to noisy data) than the Euclidean norm (representing mean squared error). The selection of $p = 1$ is the smallest among the ℓ_p norms for which the norm is *convex*. In this course, we will dedicate significant attention to this important notion, as convexity will have profound impact on solvability of optimization problems.

Yet another important particular case is the ℓ_∞ norm

$$\|\mathbf{u}\|_\infty = \max_{i=1,\dots,n} |u_i|.$$

The latter can be obtained as the limit of ℓ_p.

Exercise 2. *Show that* $\lim_{p \to \infty} \|\mathbf{u}\|_p = \max_{i=1,\dots,n} |u_i|$.

---

[1]We will henceforth abbreviate "if and only if" as "iff". Two statements related by "iff" are equivalent; for example, if one of the statements is a definition of some object, and the other is its property, the latter property can be used as an alternative definition. We will see many such examples.
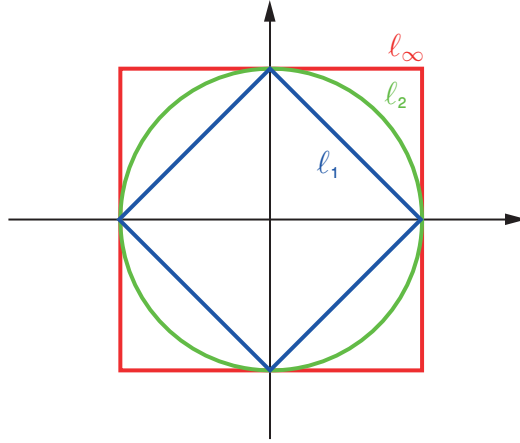
Figure 2: Unit circles with respect to different ℓ_p norms in $\mathbb{R}^2$.

Geometrically, a norm measures the *length* of a vector; a vector of length $\|\mathbf{u}\| = 1$ is said to be a *unit vector* (with respect to[2] that norm). The collection $\{\mathbf{u} : \|\mathbf{u}\| = 1\}$ of all unit vectors is called the *unit circle* (Figure 2). Note that the unit circle is indeed a "circle" only for the Euclidean norm. Similarly, the collection $B_r = \{\mathbf{u} : \|\mathbf{u}\| \leq r\}$ of all vectors with length no bigger than r is called the *ball of radius r* (w.r.t. a given norm). Again, norm balls are round only for the Euclidean norm. From Figure 2 we can deduce that for $p < q$, the ℓ_p unit circle is fully contained in the ℓ_q unit circle. This means that the ℓ_p norm is *bigger* than the ℓ_q norm in the sense that $\|\mathbf{u}\|_q \leq \|\mathbf{b}\|_p$.

Exercise 3. *Show that $\|\mathbf{u}\|_q \leq \|\mathbf{b}\|_p$ for every $p < q$.*

Continuing the geometric interpretation, it is worthwhile mentioning several relations between the inner product and the ℓ_2 norm it induces. The inner product of two (ℓ_2-) unit vectors measures the cosine of the angle between them or, more generally,

$$\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \cos \theta,$$

where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$. Two vectors satisfying $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ are said to be *orthogonal* (if the vectors are unit, they are also said to be *orthonormal*) – the algebraic way of saying "perpendicular".

The following result is doubtlessly the most important inequality in linear algebra (and, perhaps, in mathematics in general):

Theorem (Cauchy-Schwartz inequality). *Let $\| \cdot \|$ be the norm induced by an inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$, Then, for any $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \|\mathbf{v}\|$$

with equality holding iff $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent.

---

[2]We will often abbreviate "with respect to" as "w.r.t."

4   Matrices

A *linear map* from the n-dimensional linear space $\mathbb{R}^n$ to the m-dimensional linear space $\mathbb{R}^m$ is a function mapping $\mathbf{u} \in \mathbb{R}^n$ $\mathbf{v} \in \mathbb{R}^m$ according to

$$v_i = \sum_{j=1}^{n} a_{ij} u_j \quad i = 1, \ldots, m.$$

The latter can be expressed compactly using the matrix-vector product notation $\mathbf{v} = \mathbf{Au}$, where $\mathbf{A}$ is the matrix with the elements a_{ij}. In other words, a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a compact way of expressing a linear map between $\mathbb{R}^m$ and $\mathbb{R}^n$. An matrix is said to be *square* of $m = n$; such a matrix defines an operator mapping $\mathbb{R}^n$ to itself. A *symmetric* matrix is a square matrix $\mathbf{A}$ such that $\mathbf{A}^{\mathrm{T}} = \mathbf{A}$.

Recollecting our notion of linear combinations and linear subspaces, observe that the vector $\mathbf{v} = \mathbf{Au}$ is the linear combination of the columns $\mathbf{a}_1, \ldots, \mathbf{a}_n$ of $\mathbf{A}$ with the weights $u_1, \ldots, u_n$. The linear subspace $\mathbf{A}\mathbb{R}^m = \{\mathbf{Au} : \mathbf{u} \in \mathbb{R}^m\}$ is called the *columns space* of $\mathbf{A}$. The space is n-dimensional if the columns are linearly independent; otherwise, if k columns are linearly dependent, the space is $n - k$ dimensional. The latter dimension is called the *column rank* of the matrix. By transposing the matrix, the *row rank* can be defined in the same way. The following result is fundamental in linear algebra:

Theorem. *The column rank and the row rank of a matrix are always equal.*

Exercise 4. *Prove the above theorem.*

Since the row and the column ranks of a matrix are equal, we will simply refer to both as *the rank*, denoting rank $\mathbf{A}$. A square $n \times n$ matrix is *full rank* if its rank is n, and is *rank deficient* otherwise. Full rank is a necessary condition for a square matrix to possess an *inverse* (Reminder: the inverse of a matrix $\mathbf{A}$ is such a matrix $\mathbf{B}$ that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$; when the inverse exists, the matrix is called *invertible* and its inverse is denoted by $\mathbf{A}^{-1}$).

In this course, we will often encounter the *trace* of a square matrix, which is defined as the sum of its diagonal entries,

$$\operatorname{tr} \mathbf{A} = \sum_{i=1}^{n} a_i.$$

The following property of the trace will be particularly useful:

Property. *Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{n \times m}$. Then $\operatorname{tr}(\mathbf{AB}) = \operatorname{tr}(\mathbf{BA})$.*

This is in sharp contrast to the result of the product itself, which is generally not commutative.

In particular, the squared norm $\|\mathbf{u}\|^2 = \mathbf{u}^{\mathrm{T}}\mathbf{u}$ can be written as $\operatorname{tr}(\mathbf{u}^{\mathrm{T}}\mathbf{u}) = \operatorname{tr}(\mathbf{uu}^{\mathrm{T}})$. We will see many cases where such an apparently weird writing is very useful. The above property can be generalized to the product of k matrices $\mathbf{A}_1 \ldots \mathbf{A}_k$ by saying that $\operatorname{tr}(\mathbf{A}_1 \ldots \mathbf{A}_k)$ *is invariant under a cyclic permutation of the factors as long as their product is defined*. For example, $\operatorname{tr}(\mathbf{AB}^{\mathrm{T}}\mathbf{C}) = \operatorname{tr}(\mathbf{CAB}^{\mathrm{T}}) = \operatorname{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{CA})$ (again, as long as the matrix dimensions are such that the products are defined).

5   Matrix inner product and norms

The notion of an inner product can be extended to matrices by thinking of an $m \times n$ matrix as of a long vector $m \times n$ vector containing the matrix elements for example, in the column-stack order. We will denote such a vector as $\text{vec}(\mathbf{A}) = (a_{11}, \ldots, a_{m1}, a_{12}, \ldots, a_{m2}, \ldots, a_{1n}, \ldots, a_{mn})^{\mathrm{T}}$. With such an interpretation, we can define the inner product of two matrices as

$$\langle \mathbf{A}, \mathbf{B} \rangle = \langle \text{vec}(\mathbf{A}), \text{vec}(\mathbf{B}) \rangle = \sum_{i,j} a_{ij} b_{ij}.$$

Exercise 5. *Show that* $\langle \mathbf{A}, \mathbf{B} \rangle = \text{tr}(\mathbf{A}^{\mathrm{T}} \mathbf{B})$.

The inner product induces the standard Euclidean norm on the space of column-stack representation of matrices,

$$\|\mathbf{A}\| = \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle} = \sqrt{\langle \text{vec}(\mathbf{A}), \text{vec}(\mathbf{B}) \rangle} = \sqrt{\sum_{i,j} a_{ij}^2},$$

known as the *Frobenius norm*. Using the result of the exercise above, we can write

$$\|\mathbf{A}\|_{\mathrm{F}}^2 = \text{tr}(\mathbf{A}^{\mathrm{T}} \mathbf{A}).$$

Note the qualifier $_{\mathrm{F}}$ used to distinguish this norm from other matrix norm that we will define in the sequel.

There exist another "standard" way of defining matrix norms by thinking of an $m \times n$ matrix $\mathbf{A}$ as a linear operator mapping between two normed spaces, say, $\mathbf{A} : (\mathbb{R}^m, \ell_p) \to (\mathbb{R}^n, \ell_q)$. Then, we can define the *operator norm* measuring the maximum change of length (in the ℓ_q sense) of a unit (in the ℓ_p sense) vector in the operator domain:

$$\|\mathbf{A}\|_{p,q} = \max_{\|\mathbf{u}\|_p = 1} \|\mathbf{A}\mathbf{u}\|_q.$$

Exercise 6. *Use the axioms of a norm to show that* $\|\mathbf{A}\|_{p,q}$ *is a norm.*

6   Eigendecomposition

Eigendecomposition (a.k.a. *eigenvalue decomposition* or in some contexts *spectral decomposition*, from the German *eigen* for "self") is doubtlessly the most important and useful forms of matrix factorization. The following discussion will be valid only for square matrices.

Recall that an $n \times n$ matrix $\mathbf{A}$ represents a linear map on $\mathbb{R}^n$. In general, the effect of $\mathbf{A}$ on a vector $\mathbf{u}$ is a new vector $\mathbf{v} = \mathbf{A}\mathbf{u}$, rotated and elongated or shrunk (and, potentially, reflected). However, there exist vectors which are only elongated or shrunk by $\mathbf{A}$. Such vectors are called *eigenvectors*. Formally, an eigenvector of $\mathbf{A}$ is a non-zero vector $\mathbf{u}$ satisfying $\mathbf{A}\mathbf{u} = \lambda \mathbf{u}$, with the scalar λ (called *eigenvalue*) measuring the amount of elongation or

shrinkage of $\mathbf{u}$ (if $\lambda < 0$, the vector is reflected). Note that the scale of an eigenvector has no meaning as it appears on both sides of the equation; for this reason, eigenvectors are always normalized (in the ℓ_2 sense). For reasons not so relevant to our course, the collection of the eigenvalues is called the *spectrum* of a matrix.

For an $n \times n$ matrix $\mathbf{A}$ with n linearly independent eigenvectors we can write the following system

$$\begin{cases} \mathbf{A}\mathbf{u}_1 &= \lambda_1 \mathbf{u}_1 \\ \vdots & \quad \vdots \\ \mathbf{A}\mathbf{u}_n &= \lambda_n \mathbf{u}_n \end{cases} \tag{0}$$

Stacking the eigenvectors into the columns of the $n \times n$ matrix $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$, and defining the diagonal matrix

$$\mathbf{\Lambda} = \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}, \tag{0}$$

we can rewrite the system more compactly as $\mathbf{A}\mathbf{U} = \mathbf{U}\mathbf{\Lambda}$. Independent eigenvectors means that $\mathbf{U}$ is invertible, which leads to $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$. Geometrically, eigendecomposition of a matrix can be interpreted as a change of coordinates into a basis, in which the action of a matrix can be described as elongation or shrinkage only (represented by the diagonal matrix $\mathbf{\Lambda}$).

If the matrix $\mathbf{A}$ is symmetric, it can be shown that its eigenvectors are orthonormal, i.e., $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \mathbf{u}_i^{\mathrm{T}}\mathbf{u}_j = 0$ for every $i \neq j$ and, since the eigenvectors have unit length, $\mathbf{u}_i^{\mathrm{T}}\mathbf{u} = 1$. This can be compactly written as $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}$ or, in other words, $\mathbf{U}^{-1} = \mathbf{U}^{\mathrm{T}}$. Matrices satisfying this property are called *orthonormal* or *unitary*, and we will say that symmetric matrices admit *unitary eigendecomposition* $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}$.

Exercise 7. *Show that a symmetric matrix admits unitary eigendecomposition* $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{T}}$.

Finally, we note two very simple but useful facts about eigendecomposition:

Property. $\operatorname{tr}\mathbf{A} = \operatorname{tr}(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}) = \operatorname{tr}(\mathbf{\Lambda}\mathbf{U}^{-1}\mathbf{U}) = \operatorname{tr}\mathbf{\Lambda} = \sum_{i=1}^{n} \lambda_i$.

Property. $\det\mathbf{A} = \det\mathbf{U}\det\mathbf{\Lambda}\det\mathbf{U}^{-1} = \det\mathbf{U}\det\boldsymbol{\lambda}\dfrac{1}{\det\mathbf{U}} = \prod_{i=1}^{n} \lambda_i$.

In other words, the trace and the determinant of a matrix are given by the sum and the product of its eigenvalues, respectively.

7 Matrix functions

Eigendecomposition is a very convenient way of performing various matrix operations. For example, if we are given the eigendecomposition of $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$, its inverse can be expressed

as $\mathbf{A}^{-1} = (\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{-1})^{-1} = \mathbf{U}\boldsymbol{\Lambda}^{-1}\mathbf{U}^{-1}$; however, since $\boldsymbol{\Lambda}$ is diagonal, $\boldsymbol{\Lambda}^{-1} = \mathrm{diag}\{1/\lambda_1, \ldots, 1/\lambda_n\}$. (This does not suggest, of course, that this is the computationally preferred way to invert matrices, as the eigendecomposition itself is a costly operation).

A similar idea can be applied to the square of a matrix: $\mathbf{A}^2 = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{-1}\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{-1} = \mathbf{U}\boldsymbol{\Lambda}^2\mathbf{U}^{-1}$ and, again, we note that $\boldsymbol{\Lambda}^2 = \mathrm{diag}\{\lambda_1^2, \ldots, \lambda_n^2\}$. By using induction, we can generalize this result to any integer power $p \geq 0$: $\mathbf{A}^p = \mathbf{U}\boldsymbol{\Lambda}^p\mathbf{U}^{-1}$. (Here, if, say, $p = 1000$, the computational advantage of using eigendecomposition might be well justified).

Going one step further, let

$$\varphi(t) = \sum_{i \geq 0} c_i t^i$$

be a polynomial (either of a finite degree or an infinite series). We can apply this function to a square matrix $\mathbf{A}$ as follows:

$$
\begin{aligned}
\varphi(\mathbf{A}) &= \sum_{i \geq 0} c_i \mathbf{A}^i = \sum_{i \geq 0} c_i \mathbf{U}\boldsymbol{\Lambda}^p\mathbf{U}^{-1} \\
&= \mathbf{U}\left(\sum_{i \geq 0} c_i \boldsymbol{\Lambda}^p\right)\mathbf{U}^{-1} = \mathbf{U}\left(\sum_{i \geq 0} \mathrm{diag}\{c_i\lambda_1^i, \ldots, c_i\lambda_n^i\}\right)\mathbf{U}^{-1} \\
&= \mathbf{U}\begin{pmatrix} \sum_{i \geq 0} c_i\lambda_1^i & & \\ & \ddots & \\ & & \sum_{i \geq 0} c_i\lambda_n^i \end{pmatrix}\mathbf{U}^{-1} \\
&= \mathbf{U}\mathrm{diag}\{\varphi(\lambda_1), \ldots, \varphi(\lambda_n)\}\mathbf{U}^{-1}.
\end{aligned}
$$

Denoting by $\varphi(\boldsymbol{\Lambda}) = \mathrm{diag}\{\varphi(\lambda_1), \ldots, \varphi(\lambda_n)\}$ the diagonal matrix formed by the element-wise application of the scalar function φ to the eigenvalues of $\mathbf{A}$, we can write compactly

$$\varphi(\mathbf{A}) = \mathbf{U}\varphi(\boldsymbol{\Lambda})\mathbf{U}^{-1}.$$

Finally, since many functions can be described polynomial series, we can generalize the latter definition to a (more or less) arbitrary scalar function φ.

The above procedure is a standard way of constructing a *matrix function* (this term is admittedly confusing, as we will see it assuming another meaning); for example, matrix exponential and logarithm are constructed exactly like this. Note that the construction is sharply different from applying the function φ *element-wise*!

8   Positive definite matrices

Symmetric square matrices define an important family of functions called quadratic forms that we will encounter very often in this course. Formally, a *quadratic form* is a scalar function on $\mathbb{R}^n$ given by

$$\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x} = \sum_{i,j=1}^{n} a_{ij}x_i x_j,$$

where $\mathbf{A}$ is a symmetric $n \times n$ matrix, and $\mathbf{x} \in \mathbb{R}^n$.

Definition. *A symmetric square matrix $\mathbf{A}$ is called* positive definite *(denoted as $\mathbf{A} \succ 0$) iff for every $\mathbf{x} \neq \mathbf{0}$, $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$. The matrix is called* positive semi-definite *(denoted as $\mathbf{A} \succeq 0$) if the inequality is weak.*

Positive (semi-) definite matrices can be equivalently defined through their eigendecomposition:

Property. *Let $\mathbf{A}$ be a symmetric matrix admitting the eigendecomposition $\mathbf{A} = \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T$. Then $\mathbf{A} \succ 0$ iff $\lambda_i > 0$ for $i = 1, \ldots, n$. Similarly, $\mathbf{A} \succeq 0$ iff $\lambda_i \geq 0$.*

In other words, the matrix is positive (semi-) definite if it has positive (non-negative) spectrum. To get a hint why this is true, consider an arbitrary vector $\mathbf{x} \neq \mathbf{0}$, and write

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{x}^T \mathbf{U} \boldsymbol{\Lambda} \mathbf{U}^T \mathbf{x} = (\mathbf{U}^T \mathbf{x})^T \boldsymbol{\Lambda} (\mathbf{U}^T \mathbf{x}).$$

Denoting $\mathbf{y} = \mathbf{U}^T \mathbf{x}$, we have

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{y}^T \boldsymbol{\Lambda} \mathbf{y} = \sum_{i=1}^{n} \lambda_i y_i^2.$$

Since $\mathbf{U}^T$ is full rank, the vector $\mathbf{y}$ is also an arbitrary non-zero vector in $\mathbb{R}^n$ and the only way to make the latter sum always positive is by ensuring that all λ_i are positive. The very same reasoning is also true in the opposite direction.

Geometrically, a quadratic form describes a second-order (hence the name quadratic) surface in $\mathbb{R}^n$, and the eigenvalues of the matrix $\mathbf{A}$ can be interpreted as the surface curvature. Very informally, if a certain eigenvalue λ_i is positive, a small step in the direction of the corresponding eigenvector $\mathbf{u}_i$ rotates the normal to the surface in the same direction. The surface is said to have *positive curvature* in that direction. Similarly, a negative eigenvalue corresponds to the normal rotating in the opposite direction of the step (negative curvature). Finally, if $\lambda_i = 0$, a step in the direction of $\mathbf{u}_i$ leave the normal unchanged (the surface is said to be *flat* in that direction). A quadratic form created by a positive definite matrix represents a positively curved surface in all directions. Such a surface is cup-shaped (if you can imagine an n-dimensional cup) or, formally, is *convex*; in the sequel, we will see the important consequences this property has on optimization problems.