Mathematical Preliminaries

AA/ME/EE 583: Nonlinear Control Systems

Amir Taghvaei

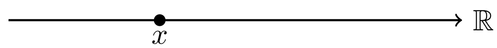
October 25, 2023

Goal: Review the mathematical concepts that is essential for this course.

1 Multi-variable calculus

The goal is to review some definitions for functions of multiple variables. First, we introduce the Euclidean space.

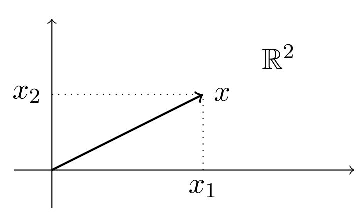
Euclidean spaces: The simplest Euclidean space is the set of all real numbers. It is denoted by the symbol $\mathbb{R}$ and visualized as points on a line.



Stacking n real numbers together gives the n-dimensional Euclidean space. It is denoted by $\mathbb{R}^n$.

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n. \quad \text{where} \quad x_1, x_2, \ldots, x_n \in \mathbb{R}$$

The elements of $\mathbb{R}^n$ are visualized as vectors in an n-dimensional space (In normal condition, you can only visualize up to $n = 3$).



Real-valued functions: It is a function that takes n-variables as input, and outputs a real number. We use the notation $f : \mathbb{R}^n \to \mathbb{R}$ for a real-valued function f. An example is:

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{2}(x_1^2 + x_2^2 + \ldots + x_n^2) \tag{1.1}$$

Since we are lazy, we are going to write $f(x)$ instead of $f(x_1, x_2, \ldots, x_n)$, when we understand from the context that $x = [x_1, x_2, \ldots, x_n]^\top$ (we will use $^\top$ for transpose).

Gradient: It is the derivative of a real-valued function. Specifically, for $f : \mathbb{R}^n \to \mathbb{R}$, the gradient is a n-dimensional vector, denoted by $\nabla f(x)$, and defined as

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$$

where $\frac{\partial f}{\partial x_k}(x)$ is the partial derivative of f with respect to x_k, evaluated at x. For instance, evaluating the gradient for the example (1.1) yields:

$$\nabla f(x) = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x$$

Exercise 1. *Draw a 3d plot of the function $f(x_1, x_2) = x_1 x_2$ and draw the vector-field of its gradient.*

Vector-valued functions: It is a function that takes multiple variables as input and outputs a vector. It is shown by the notation $f : \mathbb{R}^n \to \mathbb{R}^m$. For each $x \in \mathbb{R}^n$, the output $f(x)$ is a m-dimensional vector, where each component is a real-valued function:

$$f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \ldots, \\ f_m(x) \end{bmatrix}$$

For example, $f : \mathbb{R}^3 \to \mathbb{R}^2$ and

$$f(x_1, x_2, x_3) = \begin{bmatrix} x_1 + 2x_2 + \sin(x_3) \\ x_1 - x_2 + 1 \end{bmatrix} \tag{1.2}$$

Jacobian: it is derivative of a vector-valued function. Specifically, for $f : \mathbb{R}^n \to \mathbb{R}^m$, the Jacobian, denoted by $\frac{\partial f}{\partial x}(x)$, is a $m \times n$ matrix defined according to

$$\frac{\partial f}{\partial x}(x) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \frac{\partial f_1}{\partial x_2}(x) & \ldots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \frac{\partial f_m}{\partial x_2}(x) & \ldots & \frac{\partial f_m}{\partial x_n}(x) \end{bmatrix}$$

Example 1. *The Jacobian for the example* (1.2) *is*

$$\frac{\partial f}{\partial x}(x) = \begin{bmatrix} 1 & 2 & \cos(x_3) \\ 1 & -1 & 0 \end{bmatrix}$$

Exercise 2. *Let $f(x) = Ax + b$ where A is a $m \times n$ matrix and $b \in \mathbb{R}^m$. Find $\frac{\partial f}{\partial x}(x)$ in terms of A and b.*

Remark 1. *When $m = 1$, the Jacobian becomes the transpose of the gradient: $\nabla f(x) = \frac{\partial f}{\partial x}(x)^T$*

2 Spectral analysis of matrices

The goal is to review some basic spectral properties of matrices.

Matrices: a $m \times n$ matrix A is a basically a table of numerical entries with m rows and n columns:

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ \vdots & & & \vdots \\ A_{m1} & A_{12} & \dots & A_{mn} \end{bmatrix}$$

We can assign a geometrical meaning to matrices by identifying them with linear maps. Specifically, a $m \times n$ matrix A is identified as a linear map that takes a n-dimensional vector x as input and outputs a m-dimensional vector Ax by matrix vector product rule.

$$x \in \mathbb{R}^n \mapsto Ax \in \mathbb{R}^m$$

Eigenvalues and eigenvectors: Consider a square $n \times n$ matrix A. A (possibly complex) number λ is an eigenvalue of A with (possibly complex) eigenvector $u \in \mathbb{R}^n$ if

$$Au = \lambda u$$

Geometrically, it means that A maps the vector u to a vector with same direction, but a different length that is scaled by λ.

Characteristic equation: The eigenvalue condition implies that there exists a non-zero vector u such that $(A - \lambda I)u = 0$. This only happens when the columns of $A - \lambda I$ are linearly dependent, hence,

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

This is called the characteristic equation of A. The eigenvalues are the (complex) numbers λ that solve the characteristic equation. The eigenvectors are obtained by solving for u in $(A - \lambda I)u = 0$.

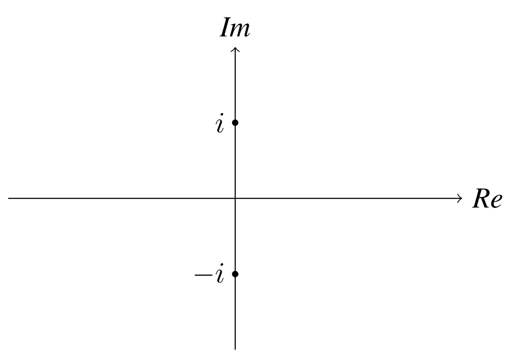
Example 2. *For example, consider*

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

The characteristic equation is

$$\det(A - \lambda I) = \det(\begin{bmatrix} -\lambda & 1 \\ -1 & -\lambda \end{bmatrix}) = \lambda^2 + 1 = 0$$

The solutions of the equation $\lambda = \pm i$ are the two eigenvalues of A. Eigenvalues are visualized on the complex plane.



The eigenvectors are obtained by solving for u in $(A - \lambda I)u = 0$. For this example,

$$\lambda = i \quad \rightarrow \quad \begin{bmatrix} -i & 1 \\ -1 & -i \end{bmatrix} u = 0 \quad \rightarrow \quad u = \begin{bmatrix} 1 \\ i \end{bmatrix}$$
$$\lambda = -i \quad \rightarrow \quad \begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix} u = 0 \quad \rightarrow \quad u = \begin{bmatrix} 1 \\ -i \end{bmatrix}$$

Remark 2. *Note that the eigenvectors are unique up to scaling. If u is an eigenvector, then αu is also an eigenvector for any (complex) number α.*

Remark 3. *If all components of the matrix A are real numbers, then the complex eigenvalues come in conjugate pairs. That is to say, if $\lambda = a + ib$ is an eigenvalue, then its complex conjugate $\lambda^* = a - ib$ is also another eigenvalue.*

Exercise 3. *Find the eigenvalues and eigenvectors for the following matrices:*

$$\text{(a)} \quad \begin{bmatrix} 4 & 3 \\ -3 & 4 \end{bmatrix} \qquad\qquad \text{(b)} \quad \begin{bmatrix} \alpha & 1 \\ 0 & \alpha \end{bmatrix}$$

3 Vector and matrix norms

Norms are used to quantify how "big" an object is. It is one of the most useful tool in a mathematical analysis. The goal of this section is to review definitions of norms for vectors and matrices.

Euclidean norm: It is the most well-known norm. It corresponds to the actual physical length of a vector. In particular, for an n-dimensional vector $x \in \mathbb{R}^n$, the Euclidean norm is denoted by $\|x\|_2$ and defined according to

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{\frac{1}{2}}.$$

p**-norm:** The Euclidean norm is generalized to a family of norms, called p-norms, by changing the number 2 in the definition, to any number $p \in [1, \infty]$:

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}},$$
$$\|x\|_\infty = \max_{i \in \{1,2,\ldots,n\}} |x_i|.$$

Example 3. *Consider the two-dimensional vector $x = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^\top$. Then,*

$$\|x\|_2 = 1, \quad \|x\|_1 = \sqrt{2}, \quad \|x\|_\infty = \frac{1}{\sqrt{2}}.$$

The p-norms (including the Euclidean norm) are the most useful norms on a Euclidean space. However, they are not the only ones. You can come up with any norm as long as it satisfies the following definition.

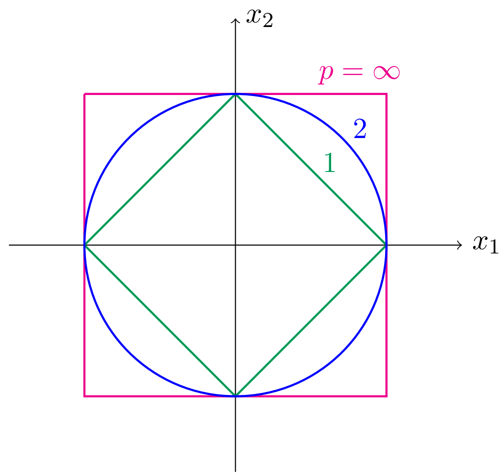
Definition: In general, any function $x \mapsto \|x\| \in \mathbb{R}$ with the following three properties is a valid norm:

1. *(positivity)* $\|x\| \geq 0$ for all $x \in \mathbb{R}^n$, and $\|x\| = 0$ if and only if $x = 0$.
2. $\|\alpha x\| = |\alpha|\|x\|$ for all $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^n$.
3. *(triangle inequality)* $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in \mathbb{R}^n$.

Unit balls: It is helpful to visualize any given norm $\|\cdot\|$ by looking at all vectors x that has norm $\|x\|$ smaller than 1; that is to say we like to look at the subset

$$B_1(0) := \{x \in \mathbb{R}^n;\ \|x\| \leq 1\}.$$

This object is called the unit ball around 0 with respect to the norm $\|\cdot\|$. The 2-dimensional unit ball for p-norms, with $p = 1, 2, \infty$, is depicted below.



Matrix norms: Recall that we can identify a $m \times n$ matrix A with linear function that takes a vector $x \in \mathbb{R}^n$ and outputs $Ax \in \mathbb{R}^m$. The norm of a matrix is then defined as how "big" this function is. In particular, for any given vector norm $\|\cdot\|$, we define the (induced) matrix norm $\|A\|$ as the largest ratio between $\|Ax\|$ and $\|x\|$:

$$\|A\| := \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

For example, if we take the vector p-norm $\|\cdot\|_p$, then the matrix p-norm of A is

$$\|A\|_p := \max_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

It is not easy to obtain a formula for the norm of the matrix, except three special cases:

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^\top A)} = (\text{maximum singular value of } A)$$

$$\|A\|_1 = \max_j \sum_{i=1}^n |A_{ij}|$$

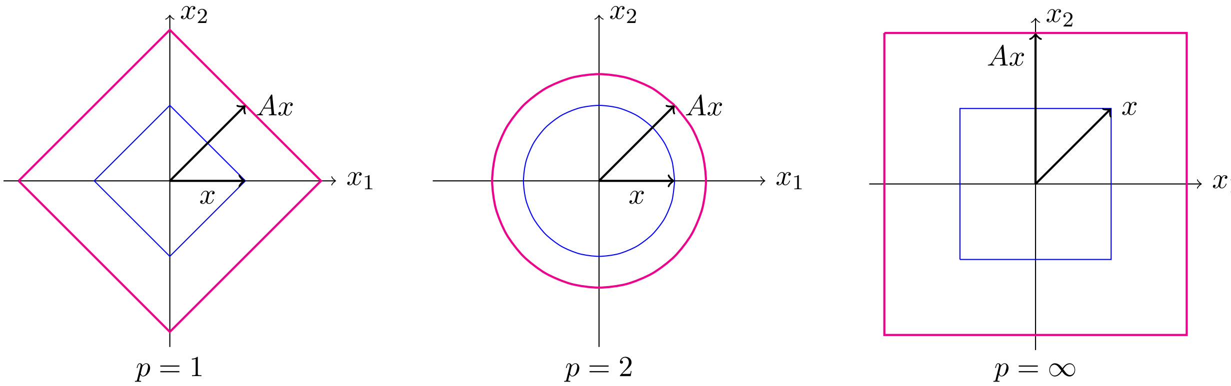
$$\|A\|_\infty = \max_i \sum_{j=1}^n |A_{ij}$$

Visually, the matrix norm corresponds to how large unit balls become after transformation with A.

Example 4. *Consider the* 2×2 *matrix* $A = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$. *Its* p*-norm, for* $p = 1, 2, \infty$ *is given by*

$$\|A\|_1 = 2, \quad \|A\|_2 = \sqrt{2}, \quad \|A\|_\infty = 2$$

The figures below show a vector x *that is enlarged maximally by* A *in these three norms.*



4 Scalar linear differential equations

Simplest differential equation: Assume $x(t) \in \mathbb{R}$ satisfies the linear differential equation:

$$\dot{x}(t) = ax(t), \quad x(0) = x_0 \tag{4.1}$$

The solution to this equation is given by $x(t) = e^{at}x_0$. This is verified by plugging it back to the equation and checking the initial condition.

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = \frac{\mathrm{d}}{\mathrm{d}t}(e^{at}x_0) = ae^{at}x_0 = ax(t).$$

Time-varying equation: Consider the differential equation (4.1), but with a time-varying constant:

$$\dot{x}(t) = a(t)x(t), \quad x(0) = x_0$$

The solution to this differential equation is given by

$$x(t) = e^{\int_0^t a(\tau)\,\mathrm{d}\tau}x_0$$

which can be verified similarly by plugging it back to the differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = \frac{\mathrm{d}}{\mathrm{d}t}(e^{\int_0^t a(\tau)\,\mathrm{d}\tau}x_0) = \frac{\mathrm{d}}{\mathrm{d}t}(\int_0^t a(\tau)\,\mathrm{d}\tau)e^{\int_0^t a(\tau)\,\mathrm{d}\tau}x_0 = a(t)e^{\int_0^t a(\tau)\,\mathrm{d}\tau}x_0 = a(t)x(t).$$

With input: Now consider the differential equation (4.1) with an input:

$$\dot{x}(t) = ax(t) + u(t), \quad x(0) = x_0$$

In order to solve this equation, introduce a new process $z(t) = e^{-at}x(t)$. The time derivative of $z(t)$

$$\dot{z}(t) = \frac{\mathrm{d}}{\mathrm{d}t}(e^{-at}x(t)) = \frac{\mathrm{d}}{\mathrm{d}t}(e^{-at})x(t) + e^{-at}\dot{x}(t) = -ae^{-at}x(t) + e^{-at}(ax(t) + u(t)) = e^{-at}u(t)$$

becomes independent of $z(t)$. Integrating both sides from 0 to t yields

$$z(t) - z(0) = \int_0^t e^{-as}u(s)\,\mathrm{d}s.$$

Transforming the equation back to $x(t)$, using the relationship $z(t) = e^{-at}x(t)$, concludes

$$e^{-at}x(t) - x(0) = \int_0^t e^{-as}u(s)\,\mathrm{d}s,$$

resulting the solution formula

$$x(t) = e^{at}x_0 + \int_0^t e^{a(t-s)}u(s)\,\mathrm{d}s.$$

Exercise 4. *Show that the solution to the differential equation*

$$\dot{x}(t) = a(t)x(t) + u(t), \quad x(0) = x_0$$

is given by

$$x(t) = e^{\int_0^t a(\tau)\,\mathrm{d}\tau}x_0 + \int_0^t e^{\int_s^t a(\tau)\,\mathrm{d}\tau}u(s)\,\mathrm{d}s.$$

5 Symmetric and positive definite matrices

The goal is to review the definitions and properties of symmetric and positive definite matrices. In order to do so, we start with recalling the definition of diagonal and orthogonal matrices.

Diagonal matrix: a $n \times n$ matrix A is diagonal if all the non-diagonal entries are zero. We use the notation $A = \mathrm{diag}(a_1, a_2, \dots, a_n)$ for a diagonal matrix with diagonal entries $a_1, a_2, \dots, a_n$.

$$A = \begin{bmatrix} a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_n \end{bmatrix}$$

Orthogonal matrix: a $n \times n$ real matrix U is orthogonal if

$$UU^\top = U^\top U = I.$$

The definition implies that all the column (or row) vectors of U are orthonormal. In particular, if $u_1, u_2, \dots, u_n$ are the columns (or rows) of U, then $u_i^\top u_j = 0$ if $i \neq j$ and $u_i^\top u_i = 1$.

Example 5. *Consider the two vectors*

$$u_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}, \quad u_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

The vectors are orthonormal because $u_1^\top u_2 = 0$*, and* $u_1^\top u_1 = u_2^\top u_2 = 1$*. The matrix*

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

formed by stacking these two vectors is an orthogonal matrix because

$$\begin{aligned} UU^\top &= \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I \quad \textit{and} \\ U^\top U &= \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I \end{aligned}$$

Symmetric matrices: A $n \times n$ matrix P is symmetric if it is equal to its transpose $P^\top$:

$$P \text{ is symmetric iff } P = P^\top.$$

Spectrum of symmetric matrices: Symmetric matrices have a very user-friendly spectral properties. In particular, a $n \times n$ symmetric matrix P

- has n <u>real</u> eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$,
- has n <u>orthonormal</u> eigenvectors $u_1, u_2, \ldots, u_n$,
- and it can be diagonalized

  $$P = U \Lambda U^\top$$

  where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and U is an orthogonal matrix with $u_1, u_2, \ldots, u_n$ as its columns.

Example 6. *Consider the matrix*

$$P = \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}$$

The eigenvalues are $\lambda_1 = 1$ *and* $\lambda_2 = 7$*, with eigenvectors* $u_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$ *and* $u_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$*, respectively. The matrix is expressed as*

$$P = \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 7 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = U \Lambda U^\top$$

Upper and lower bounds on symmetric matrices: For a symmetric matrix P, let $\lambda_{\min}(P)$ denote the minimum eigenvalue and $\lambda_{\max}(P)$ denote the maximum eigenvalue. Then, the matrix P satisfies the inequalities

$$\lambda_{\min}(P)\|x\|^2 \leq x^\top P x \leq \lambda_{\max}(P)\|x\|^2, \quad \forall x \in \mathbb{R}^n.$$

To prove the lower-bound, express $x^\top P x$ as

$$x^\top P x = x^\top U \Lambda U^\top x = v^\top \Lambda v,$$

where $v = U^\top x$. Because U is orthogonal, $\|v\|_2^2 = v^\top v = x^\top U U^\top x = x^\top x = \|x\|_2^2$. Then, the lower bound is true because

$$v^\top \Lambda v = \sum_{i=1}^{n} \lambda_i v_i^2 \geq \lambda_{\min}(P) \sum_{i=1}^{n} v_i^2 \geq \lambda_{\min}(P)\|v\|_2^2 = \lambda_{\min}(P)\|x\|^2.$$

The proof of the upper-bound is similar and left as an exercise.

Positive-definite matrices: A symmetric matrix P is positive definite if all eigenvalues are positive. We use the notation $P \succ 0$ to say P is positive-definite.

$$P \succ 0 \quad \Leftrightarrow \quad \lambda_{\min}(P) > 0$$

A positive definite matrix $P \succ 0$ has the property that

$$x^\top P x > 0 \quad \forall x \neq 0$$

because

$$x^\top P x \geq \lambda_{\min}(P)\|x\|_2^2 > 0 \quad \forall x \neq 0.$$

The converse is also true. If $x^\top P x > 0$ for all $x \neq 0$, then P is positive-definite because

$$\begin{aligned} &\forall i, \quad \lambda_i = \lambda_i u_i^\top u_i = u_i^\top P u_i > 0 \\ &\Rightarrow \quad \lambda_{\min}(P) > 0 \quad \Rightarrow \quad P \succ 0 \end{aligned}$$

We call a symmetric matrix P positive semi-definite if all the eigenvalues are non-negative, i.e. $\lambda_{\min}(P) \geq 0$. We use the notation $P \succeq 0$. Equivalently, $P \succeq 0$ iff $x^\top P x \geq 0$ for all x. We summarize the definitions as follows

$$\begin{aligned} P \text{ is positive semi-definite} \quad &\Leftrightarrow \quad P \succeq 0 \quad \Leftrightarrow \quad \lambda_{\min}(P) \geq 0 \quad \Leftrightarrow \quad x^\top P x \geq 0 \quad \forall x \\ P \text{ is positive definite} \quad &\Leftrightarrow \quad P \succ 0 \quad \Leftrightarrow \quad \lambda_{\min}(P) > 0 \quad \Leftrightarrow \quad x^\top P x > 0 \quad \forall x \neq 0. \end{aligned}$$

Example 7. *The symmetric matrix* $P = \begin{bmatrix} 4 & 3 \\ 3 & 4 \end{bmatrix}$ *is positive-definite because both eigenvalues* $\lambda_1 = 1$ *and* $\lambda_2 = 7$ *are positive.*

Example 8. *Consider a general real* 2×2 *matrix* $P = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$. *We like to find conditions on* a, b, *and* c *such that* P *is positive-definite. By definition,* P *is positive-definite if and only if* $x^\top P x > 0$ *for all* $x \neq 0$. *Therefore,* P *is positive definite if and only if*

$$x^\top P x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = a x_1^2 + 2b x_1 x_2 + c x_2^2 = a(x_1 + \frac{b}{a} x_2)^2 + (c - \frac{b^2}{a}) x_2^2 > 0, \quad \forall x \neq 0$$

This is true if and only if $a > 0$ *and* $c - \frac{b^2}{a} > 0$*. As a result*

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \succeq 0 \quad \Leftrightarrow \quad a > 0 \quad \textit{and} \quad c - \frac{b^2}{a} > 0$$

Exercise 5. *Show that* $P = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ *is positive definite if and only if* $c > 0$ *and* $a - \frac{b^2}{c} > 0$*.*