II. Unconstrained convex optimization

Unconstrained optimization

We will start our discussion about solving convex optimization programs by considering the unconstrained case. Our template problem is

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \;\; f(\boldsymbol{x}), \tag{1}$$

where f is convex. While we state this problem as a search over all of $\mathbb{R}^N$, almost everything we say here can be applied to minimized a convex function over an *open* set.[1]

Before we go too deep into optimization, however, we need to provide a bit more mathematical rigor in terms of how we think about convexity.

Convex sets

In this section, we will be introduced to some of the mathematical fundamentals of convex sets.

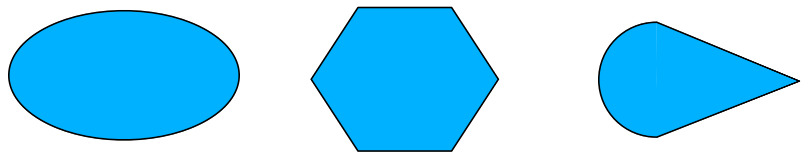
Recall that a set $\mathcal{C} \subset \mathbb{R}^N$ is **convex** if

$$\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C} \quad \Rightarrow \quad (1-\theta)\boldsymbol{x} + \theta\boldsymbol{y} \in \mathcal{C} \quad \text{for all } \theta \in [0,1].$$

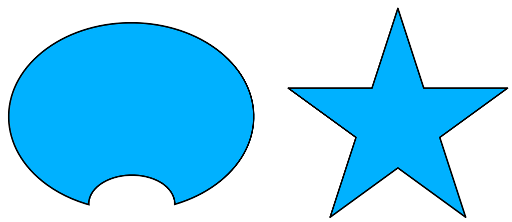
In English, this means that if we travel on a straight line between any two points in $\mathcal{C}$, then we never leave $\mathcal{C}$.

[1] A formal definition of what it means for a set to be open is provided in the technical details at the end of these notes. Informally, an open set is one that doesn't have a boundary. The standard example of such a set is an open interval on the real line, e.g., $(0,1)$. In the context of constrained optimization where the constraint set has a boundary, we must consider the fact that the solution can (and probably is) on this boundary, which complicates the picture considerably.

These sets in $\mathbb{R}^2$ are convex:



These sets are not:



Examples of convex (and nonconvex) sets

- Subspaces. Recall that if $\mathcal{S}$ is a subspace of $\mathbb{R}^N$, then $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S} \Rightarrow a\boldsymbol{x} + b\boldsymbol{y} \in \mathcal{S}$ for all $a, b \in \mathbb{R}$.
  So $\mathcal{S}$ is clearly convex.

- Affine sets. Affine sets are just subspaces that have been offset by the origin:

  $$\{\boldsymbol{x} \in \mathbb{R}^N \ : \ \boldsymbol{x} = \boldsymbol{y} + \boldsymbol{v}, \ \boldsymbol{y} \in \mathcal{T}\}, \quad \mathcal{T} = \text{subspace},$$

  for some fixed vector $\boldsymbol{v}$.

- Bound constraints. Rectangular sets of the form

  $$\mathcal{C} = \{\boldsymbol{x} \in \mathbb{R}^N : \ell_1 \leq x_1 \leq u_1, \ell_2 \leq x_2 \leq u_2, \ldots, \ell_N \leq x_N \leq u_N\}$$

  for some $\ell_1, \ldots, \ell_N, u_1, \ldots, u_N \in \mathbb{R}$ are convex.

- The *simplex* in $\mathbb{R}^N$

$$\{\boldsymbol{x} \in \mathbb{R}^N : x_1 + x_2 + \cdots + x_N \leq 1,\ x_1, x_2, \ldots, x_N \geq 0\}$$

  is convex.

- Any subset of $\mathbb{R}^N$ that can be expressed as a set of linear inequality constraints

$$\{\boldsymbol{x} \in \mathbb{R}^N : \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}\}$$

  is convex. Notice that both rectangular sets and the simplex fall into this category — for the simplex, take

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ -1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & & \cdots & & -1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

  In general, when sets like these are bounded, the result is a polyhedron.

- Norm balls. If $\|\cdot\|$ is a valid norm on $\mathbb{R}^N$, then

$$\mathcal{B}_r = \{\boldsymbol{x} \ : \ \|\boldsymbol{x}\| \leq r\},$$

  is a convex set.

- Ellipsoids. An ellipsoid is a set of the form

$$\mathcal{E} = \{\boldsymbol{x} \ : \ (\boldsymbol{x} - \boldsymbol{x}_0)^{\mathrm{T}} \boldsymbol{P}^{-1} (\boldsymbol{x} - \boldsymbol{x}_0) \leq r\},$$

  for a symmetric positive-definite matrix $\boldsymbol{P}$. Geometrically, the ellipsoid is centered at $\boldsymbol{x}_0$, its axes are oriented with the eigenvectors of $\boldsymbol{P}$, and the relative widths along these axes are proportional to the eigenvalues of $\boldsymbol{P}$.

- A single point $\{\boldsymbol{x}\}$ is convex.
- The empty set is convex.
- The set
$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1^2 - 2x_1 - x_2 + 1 \leq 0\}$$
is convex. (Sketch it!)

- The set
$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1^2 - 2x_1 - x_2 + 1 \geq 0\}$$
is **not** convex.
- The set
$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1^2 - 2x_1 - x_2 + 1 = 0\}$$
is certainly not convex.
- Sets defined by linear equality constraints where only some of the constraints have to hold are in general not convex. For example
$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1 - x_2 \leq -1 \ \text{ and } \ x_1 + x_2 \leq -1\}$$
is convex, while
$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1 - x_2 \leq -1 \ \text{ or } \ x_1 + x_2 \leq -1\}$$
is not convex.

Cones

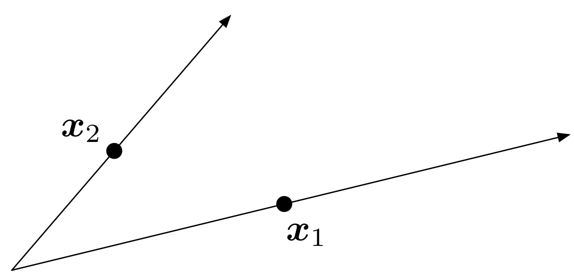
A **cone** is a set $\mathcal{C}$ such that

$$\boldsymbol{x} \in \mathcal{C} \quad \Rightarrow \quad \theta \boldsymbol{x} \in \mathcal{C} \text{ for all } \theta \geq 0.$$

Convex cones are sets which are both convex and a cone. $\mathcal{C}$ is a convex cone if

$$\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{C} \quad \Rightarrow \quad \theta_1 \boldsymbol{x}_1 + \theta_2 \boldsymbol{x}_2 \in \mathcal{C} \text{ for all } \theta_1, \theta_2 \geq 0.$$

Given an $\boldsymbol{x}_1, \boldsymbol{x}_2$, the set of all linear combinations with positive weights makes a wedge. For practice, sketch the region below that consists of all such combinations of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$:



We will mostly be interested in **proper cones**, which in addition to being convex, are closed, have a non-empty interior[2] ("solid"), and do not contain entire lines ("pointed").

Examples of convex cones

Non-negative orthant. The set of non-negative vectors,

$$\mathbb{R}^N_+ = \{\boldsymbol{x} \in \mathbb{R}^N \; : \; x_n \geq 0, \text{ for } n = 1, \ldots, N\},$$

is a proper cone.

[2]See Technical Details for precise definition.

Positive semi-definite cone. The set of $N \times N$ symmetric matrices with non-negative eigenvalues, $\mathbb{S}_+^N$, is a proper cone.

Non-negative polynomials. Vectors of coefficients of non-negative polynomials on $[0,1]$,

$$\{\boldsymbol{x} \in \mathbb{R}^N \; : \; x_1 + x_2 t + x_3 t^2 + \cdots + x_N t^{N-1} \geq 0 \text{ for all } 0 \leq t \leq 1\},$$

form a proper cone. Notice that it is not necessary that all the $x_n \geq 0$; for example $t - t^2$ $(x_1 = 0, x_2 = 1, x_3 = -1)$ is non-negative on $[0,1]$.

Norm cones. The subset of $\mathbb{R}^{N+1}$ defined by

$$\{(\boldsymbol{x}, t), \; \boldsymbol{x} \in \mathbb{R}^N, \; t \in \mathbb{R} \; : \; \|\boldsymbol{x}\| \leq t\}$$

is a proper cone for any valid norm $\|\cdot\|$ and $t > 0$. We have seen this already for the Euclidean norm with $N = 2$, but this holds for arbitrary norms and dimensions.

Every proper cone $\mathcal{K}$ defines a **partial ordering** or **generalized inequality**. We write

$$\boldsymbol{x} \preceq_{\mathcal{K}} \boldsymbol{y} \quad \text{when} \quad \boldsymbol{y} - \boldsymbol{x} \in \mathcal{K}.$$

For example, for vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$, we say

$$\boldsymbol{x} \preceq_{\mathbb{R}_+^N} \boldsymbol{y} \quad \text{when} \quad x_n \leq y_n \text{ for all } n = 1, \ldots, N.$$

For symmetric matrices $\boldsymbol{X}, \boldsymbol{Y}$, we say

$$\boldsymbol{X} \preceq_{\mathbb{S}_+^N} \boldsymbol{Y} \quad \text{when} \quad \boldsymbol{Y} - \boldsymbol{X} \text{ has non-negative eigenvalues.}$$

We will typically just use $\preceq$ when the context makes it clear. In fact, for $\mathbb{R}_+^N$ we will just write $\boldsymbol{x} \leq \boldsymbol{y}$ to denote $\boldsymbol{x} \preceq_{\mathbb{R}_+^N} \boldsymbol{y}$, as we have already done several times.

Partial orderings obey share of the properties of the standard $\leq$ on the real line. For example:

$$\boldsymbol{x} \preceq \boldsymbol{y}, \;\; \boldsymbol{u} \preceq \boldsymbol{v} \quad \Rightarrow \quad \boldsymbol{x} + \boldsymbol{u} \preceq \boldsymbol{y} + \boldsymbol{v}.$$

But other properties do not hold; for example, it is not necessary that either $\boldsymbol{x} \preceq \boldsymbol{y}$ or $\boldsymbol{y} \preceq \boldsymbol{x}$. For an extensive list of properties of partial orderings (most of which will make perfect sense on sight) can be found in [BV04, Chapter 2.4].

Convex functions

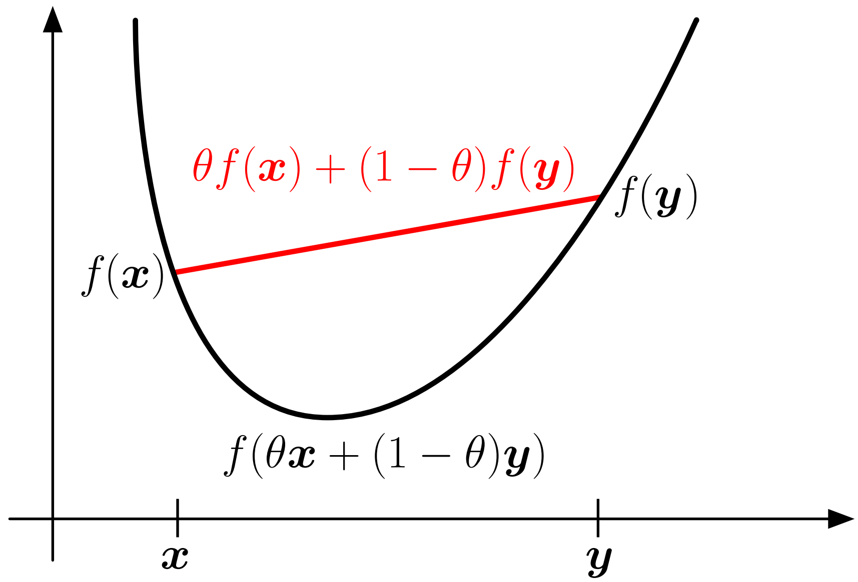
Convex *sets* are a fundamental concept in optimization. An equally important (and closely related) notion is that of convex *functions*.

To define this rigorously, we must sometimes be specific about the subset of $\mathbb{R}^N$ where a function can be applied. Specifically, the **domain** $\operatorname{dom} f$ of a function $f : \mathbb{R}^N \to \mathbb{R}^M$ is the subset of $\mathbb{R}^N$ where f is well-defined. We then say that a function f is **convex** if $\operatorname{dom} f$ is a convex set, and

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \;\leq\; \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \operatorname{dom} f$ and $0 \leq \theta \leq 1$.

This inequality is easier to interpret with a picture. The left-hand side of the inequality above is simply the function f evaluated along a line segment between $\boldsymbol{x}$ and $\boldsymbol{y}$. The right-hand side represents a straight line segment between $f(\boldsymbol{x})$ and $f(\boldsymbol{y})$ as we move along this line segment, which for a convex function must lie above f.



We say that f is **strictly convex** if dom f is convex and

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \;<\; \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$$

for all $\boldsymbol{x} \neq \boldsymbol{y} \in \operatorname{dom} f$ and $0 < \theta < 1$.

Note also that we say that a function is f is **concave** if $-f$ is convex, and similarly for strictly concave functions. We are mostly interested in convex functions, but this is only because we are mostly restricting our attention to *minimization* problems. We justified this because any maximization problem can be converted to a minimization one by multiplying the objective function by -1. Everything that we say about minimizing convex functions also applies maximizing concave ones.

Note that in the definition above, the domain matters. For example,

$$f(x) = x^3$$

is convex if $\operatorname{dom} f = \mathbb{R}_+ = [0, \infty]$ but not if $\operatorname{dom} f = \mathbb{R}$.

It will also sometimes be useful to consider the **extension** of f from $\operatorname{dom} f$ to all of $\mathbb{R}^N$, defined as

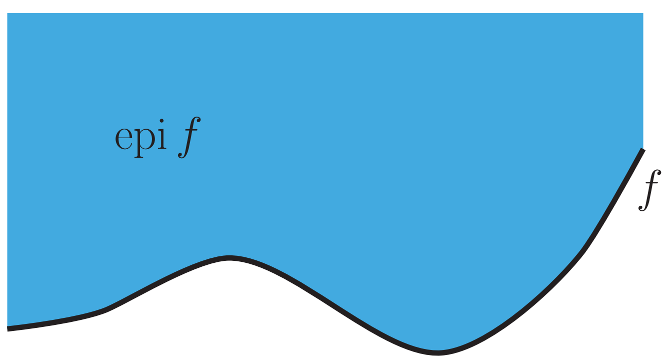
$$\tilde{f}(\boldsymbol{x}) = f(\boldsymbol{x}), \quad \boldsymbol{x} \in \operatorname{dom} f, \quad \tilde{f}(\boldsymbol{x}) = +\infty, \quad \boldsymbol{x} \notin \operatorname{dom} f.$$

If f is convex on $\operatorname{dom} f$, then its extension is also convex on $\mathbb{R}^N$.

The epigraph

A useful notion that we will encounter later in the course is that of the **epigraph** of a function. The epigraph of a function $f : \mathbb{R}^N \to \mathbb{R}$ is the subset of $\mathbb{R}^{N+1}$ created by filling in the space above f:

$$\operatorname{epi} f = \left\{ \begin{bmatrix} \boldsymbol{x} \\ t \end{bmatrix} \in \mathbb{R}^{N+1} \ : \ \boldsymbol{x} \in \operatorname{dom} f, \quad f(\boldsymbol{x}) \le t \right\}.$$



It is not hard to show that f is convex if and only if $\operatorname{epi} f$ is a convex set. This connection should help to illustrate how even though the definitions of a convex set and convex function might initially appear quite different, they actually follow quite naturally from each other.

Examples of convex functions

Here are some standard examples for functions on $\mathbb{R}$:

- $f(x) = x^2$ is (strictly) convex.
- affine functions $f(x) = ax + b$ are both convex and concave for $a, b \in \mathbb{R}$.
- exponentials $f(x) = e^{ax}$ are convex for all $a \in \mathbb{R}$.
- powers x^α are:
  - convex on $\mathbb{R}_+$ for $\alpha \geq 1$,
  - concave for $0 \leq \alpha \leq 1$,
  - convex for $\alpha \leq 0$.
- $|x|^\alpha$ is convex on all of $\mathbb{R}$ for $\alpha \geq 1$.
- logarithms: $\log x$ is concave on $\mathbb{R}_{++} := \{x \in \mathbb{R} : x > 0\}$.
- the entropy function $-x \log x$ is concave on $\mathbb{R}_{++}$.

Here are some standard examples for functions on $\mathbb{R}^N$:

- affine functions $f(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{a} \rangle + b$ are both convex and concave on all of $\mathbb{R}^N$.
- any valid norm $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ is convex on all of $\mathbb{R}^N$.
- if $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ are both convex, then the sum $f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})$ is also convex.

A useful tool for showing that a function $f : \mathbb{R}^N \to \mathbb{R}$ is convex is the fact that f is convex if and only if the function $g_{\boldsymbol{v}} : \mathbb{R} \to \mathbb{R}$,

$$g_{\boldsymbol{v}}(t) = f(\boldsymbol{x} + t\boldsymbol{v}), \quad \operatorname{dom} g = \{t \ : \ \boldsymbol{x} + t\boldsymbol{v} \in \operatorname{dom} f\}$$

is convex for every $\boldsymbol{x} \in \operatorname{dom} f$, $\boldsymbol{v} \in \mathbb{R}^N$.

Example:

Let $f(\boldsymbol{X}) = -\log\det \boldsymbol{X}$ with $\operatorname{dom} f = \mathbb{S}^N_{++}$, where $\mathbb{S}^N_{++}$ denotes the set of symmetric and (strictly) positive definite matrices. For any $\boldsymbol{X} \in \mathbb{S}^N_{++}$, we know that

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}},$$

for some diagonal, positive $\boldsymbol{\Lambda}$, so we can define

$$\boldsymbol{X}^{1/2} = \boldsymbol{U}\boldsymbol{\Lambda}^{1/2}\boldsymbol{U}^{\mathrm{T}}, \quad \text{and} \quad \boldsymbol{X}^{-1/2} = \boldsymbol{U}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}^{\mathrm{T}}.$$

Now consider any $\boldsymbol{V} \in \mathbb{S}^N$ and t such that $\boldsymbol{X} + t\boldsymbol{V} \in Sym^N_{++}$:

$$\begin{aligned}
g_{\boldsymbol{V}}(t) &= -\log\det(\boldsymbol{X} + t\boldsymbol{V}) \\
&= -\log\det(\boldsymbol{X}^{1/2}(\mathbf{I} + t\boldsymbol{X}^{-1/2}\boldsymbol{V}\boldsymbol{X}^{-1/2})\boldsymbol{X}^{1/2}) \\
&= -\log\det \boldsymbol{X} - \log\det(\mathbf{I} + t\boldsymbol{X}^{-1/2}\boldsymbol{V}\boldsymbol{X}^{-1/2}) \\
&= -\log\det \boldsymbol{X} - \sum_{n=1}^{N} \log(1 + \sigma_i t),
\end{aligned}$$

where the σ_i are the eigenvalues of $\boldsymbol{X}^{-1/2}\boldsymbol{V}\boldsymbol{X}^{-1/2}$. The function $-\log(1 + \sigma_i t)$ is convex, so the above is a sum of convex functions, which is convex.

Operations that preserve convexity

There are a number of useful operations that we can perform on a convex function while preserving convexity. Some examples include:

- **Positive weighted sum:** A **positive** weighted sum of convex functions is also convex, i.e., if $f_1, \ldots, f_m$ are convex and $w_1, \ldots, w_m \geq 0$, then $w_1 f_1 + \ldots + w_m f_m$ is also convex.
- **Composition with an affine function:** If $f : \mathbb{R}^N \to \mathbb{R}$ is convex, then $g : \mathbb{R}^D \to \mathbb{R}$ defined by

  $$g(\boldsymbol{x}) = f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}),$$

  where $\boldsymbol{A} \in \mathbb{R}^{N \times D}$ and $b \in \mathbb{R}^N$, is convex.
- **Composition with scalar functions:** Consider the function $f(\boldsymbol{x}) = h(g(\boldsymbol{x}))$, where $g : \mathbb{R}^N \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$.
  - f is convex if g is convex and h is convex and non-decreasing.
    Example: $e^{g(\boldsymbol{x})}$ is convex if g is convex.
  - f is convex if g is concave and h is convex and non-increasing.
    Example: $\frac{1}{g(\boldsymbol{x})}$ is convex if g is concave and positive.
- **Max of concave functions:** If f_1 and f_2 are convex, then $f(\boldsymbol{x}) = \max\left(f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\right)$ is convex.

Technical Details: Basic topology in $\mathbb{R}^N$

Here we provide a brief review of basic topological concepts in $\mathbb{R}^N$. Our discussion will take place using the standard Euclidean distance measure (i.e., ℓ_2 norm), but all of these definitions can be generalized to other metrics. An excellent source for this material is [Rud76].

A recurring theme in this course relates to the convergence of an iterative algorithm. We say that a sequence of vectors $\{\boldsymbol{x}_k,\ k = 1, 2, \ldots\}$ **converges** to $\widehat{\boldsymbol{x}}$ if

$$\|\boldsymbol{x}_k - \widehat{\boldsymbol{x}}\|_2 \to \widehat{\boldsymbol{x}} \quad \text{as} \quad k \to \infty.$$

More precisely, this means that for every $\epsilon > 0$, there exists an n_ϵ such that

$$\|\boldsymbol{x}_k - \widehat{\boldsymbol{x}}\|_2 \leq \epsilon \ \text{ for all } \ k \geq n_\epsilon.$$

It is easy to show that a sequence of vectors converge if and only if their individual components converge point-by-point.

A set $\mathcal{X}$ is **open** if we can draw a small ball around every point in $\mathcal{X}$ which is also entirely contained in $\mathcal{X}$. More precisely, let $\mathcal{B}(\boldsymbol{x}, \epsilon)$ be the set of all points within ϵ of $\boldsymbol{x}$:

$$\mathcal{B}(\boldsymbol{x}, \epsilon) = \{\boldsymbol{y} \in \mathbb{R}^N \ : \ \|\boldsymbol{x} - \boldsymbol{y}\|_2 \leq \epsilon\}.$$

Then $\mathcal{X}$ is open if for every $\boldsymbol{x} \in \mathcal{X}$, there exists an $\epsilon_{\boldsymbol{x}} > 0$ such that $\mathcal{B}(\boldsymbol{x}, \epsilon_{\boldsymbol{x}}) \subset \mathcal{X}$. The standard example here is open intervals of the real line, e.g. $(0, 1)$.

There are many ways to define **closed** sets. The easiest is that a set $\mathcal{X}$ is closed if its complement is open. A more illuminating (and equivalent) definition is that $\mathcal{X}$ is closed if it contains all of its limit points. A vector $\widehat{\boldsymbol{x}}$ is a **limit point** of $\mathcal{X}$ if there exists a sequence of vectors $\{\boldsymbol{x}_k\} \subset \mathcal{X}$ that converge to $\widehat{\boldsymbol{x}}$.

The **closure** of a general set $\mathcal{X}$, denoted $\mathrm{cl}(\mathcal{X})$, is the set of all limit points of $\mathcal{X}$. Note that every $\boldsymbol{x} \in \mathcal{X}$ is trivially a limit point (take the sequence $\boldsymbol{x}_k = \boldsymbol{x}$), so $\mathcal{X} \subset \mathrm{cl}(\mathcal{X})$. By construction, $\mathrm{cl}(\mathcal{X})$ is the smallest closed set that contains $\mathcal{X}$.

Related to the definition of open and closed sets are the technical definitions of boundary and interior. The **interior** of a set $\mathcal{X}$ is the collection of points around which we can place a ball of finite width which remains in the set:

$$\mathrm{int}(\mathcal{X}) = \{\boldsymbol{x} \in \mathcal{X} \;:\; \exists\, \epsilon > 0 \;\text{ such that }\; \mathcal{B}(\boldsymbol{x}, \epsilon) \subset \mathcal{X}\}.$$

The **boundary** of $\mathcal{X}$ is the set of points in $\mathrm{cl}(\mathcal{X})$ that are not in the interior:

$$\mathrm{bd}(\mathcal{X}) = \mathrm{cl}(\mathcal{X}) \setminus \mathrm{int}(\mathcal{X}).$$

Another (equivalent) way of defining this is the set of points that are in both the closure of $\mathcal{X}$ and the closure of its complement $\mathcal{X}^c$. Note that if the set is not closed, there may be boundary points that are not in the set itself.

The set $\mathcal{X}$ is **bounded** if we can find a uniform upper bound on the distance between two points it contains; this upper bound is commonly referred to as the **diameter** of the set:

$$\mathrm{diam}\, \mathcal{X} = \sup_{\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}} \|\boldsymbol{x} - \boldsymbol{y}\|_2.$$

The set $\mathcal{X} \subset \mathbb{R}^N$ is **compact** if it is closed and bounded. A key fact about compact sets is that every sequence has a convergent subsequence — this is known as the Bolzano-Weierstrauss theorem.

References

[BV04] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[Rud76] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1976.