II. Unconstrained Optimization

We will begin our discussion of solving general optimization problems by considering the **unconstrained** case. Our template problem is

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

where $f : \mathbb{R}^N \rightarrow \mathbb{R}$. We say that such a problem is unconstrained because any possible $\boldsymbol{x} \in \mathbb{R}^N$ is allowable. We will consider the case where only a subset of $\mathbb{R}^N$ is allowable once we have an idea of how to solve problems in this simpler setting.

Our primary interest will be in developing efficient procedures that are guaranteed to solve (1). However, we will see that this might not always be possible. Without placing any restrictions on the kind of function f we are trying to minimize, we cannot say much of anything – even if we restrict ourselves to continuous functions, there are some pathological functions f for which it will be difficult if not impossible to find the minimum in any efficient manner.

We will see that an important class of functions for which we generally *can* solve (1) using efficient algorithms is the set of **convex** functions. As we will see below, convex functions satisfy some desirable properties that make them much easier to work with.

Convex sets

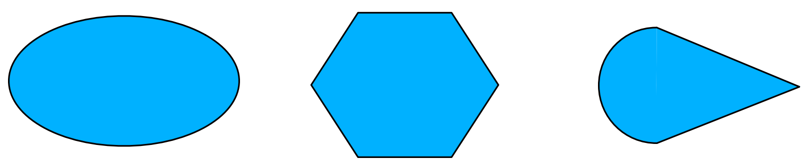
Before giving a formal definition of convex *functions* we will first introduce some of the mathematical fundamentals of convex *sets*.

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

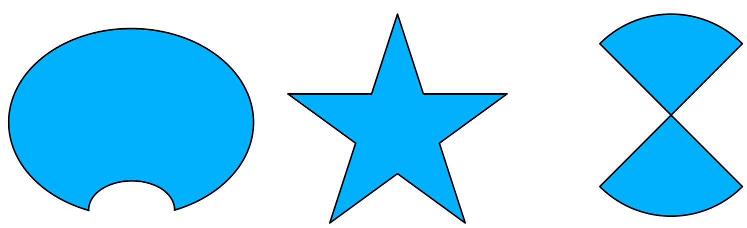
$$\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C} \quad \Rightarrow \quad (1-\theta)\boldsymbol{x} + \theta\boldsymbol{y} \in \mathcal{C} \quad \text{for all } \theta \in [0,1].$$

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

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.
- 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

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

  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 $\boldsymbol{P}$.

- The set

$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1^2 - 2x_1 - x_2 + 1 \le 0\}$$

  is convex. (Sketch it!)

- The set

$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1^2 - 2x_1 - x_2 + 1 \ge 0\}$$

  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

$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1 - x_2 \le -1 \ \text{ and } \ x_1 + x_2 \le -1\}$$

  is convex, while

$$\{\boldsymbol{x} \in \mathbb{R}^2 \ : \ x_1 - x_2 \le -1 \ \text{ or } \ x_1 + x_2 \le -1\}$$

  is not convex.

Convex functions

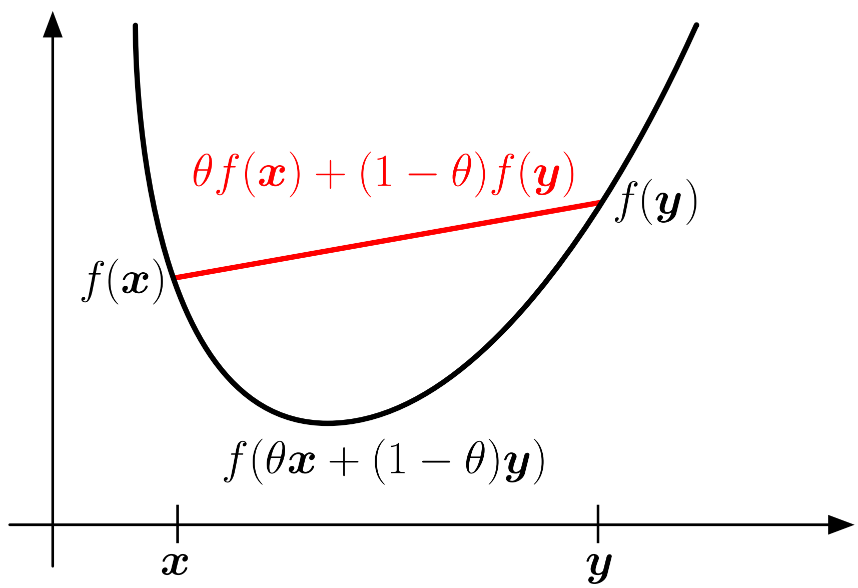
Convex *sets* are a fundamental concept in optimization. An equally important (and closely related) notion is that of convex *functions*.

We have already talked about convex functions in a loose sense. More formally, a function $f : \mathbb{R}^N \to \mathbb{R}$ is **convex** if

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

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$ 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

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

for all $\boldsymbol{x} \neq \boldsymbol{y} \in \mathbb{R}^N$ and $0 < \theta < 1$. Note that the only real difference here is replacing "$\leq$" with a strict inequality.

Note also that we say that a function 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.

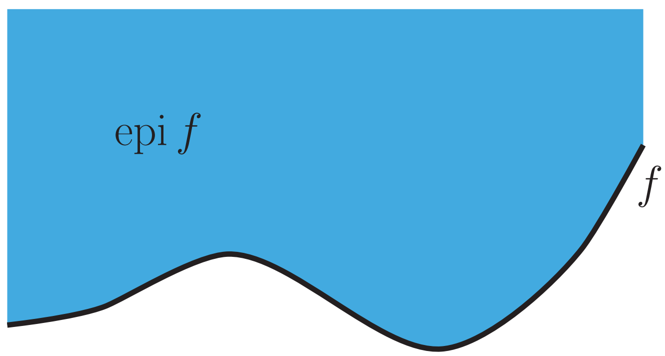
Examples:

- $f(x) = x^2$ 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 $a \in \mathbb{R}$.
- affine functions $f(\boldsymbol{x}) = \langle \boldsymbol{x}, \boldsymbol{a} \rangle + b$ are both convex and concave.
- any valid norm $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ is convex.
- the sum $f_1(\boldsymbol{x}) + f_2(\boldsymbol{x})$ is convex if $f_1(\boldsymbol{x})$ and $f_2(\boldsymbol{x})$ are both convex.

The epigraph

A useful notion that illustrates the connection between convex sets and convex functions 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:

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



It is not hard to show that f is convex if and only if 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.

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 is convex and $\ell(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$ where $\boldsymbol{A} \in \mathbb{R}^{N \times D}$ and $b \in \mathbb{R}^N$, then $f(\ell(\boldsymbol{x}))$ is convex. Note that $\ell(f(\boldsymbol{x}))$ is not necessarily 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 convex functions:** If f_1 and f_2 are convex, then $f(\boldsymbol{x}) = \max(f_1(\boldsymbol{x}), f_2(\boldsymbol{x}))$ is convex.

Equivalent characterizations of convexity

For a convex function f that is **differentiable** (meaning that the gradient $\nabla f(\boldsymbol{x})$ exists for all $\boldsymbol{x} \in \mathbb{R}^N$), there are equivalent (possibly simpler) ways to think about convexity.

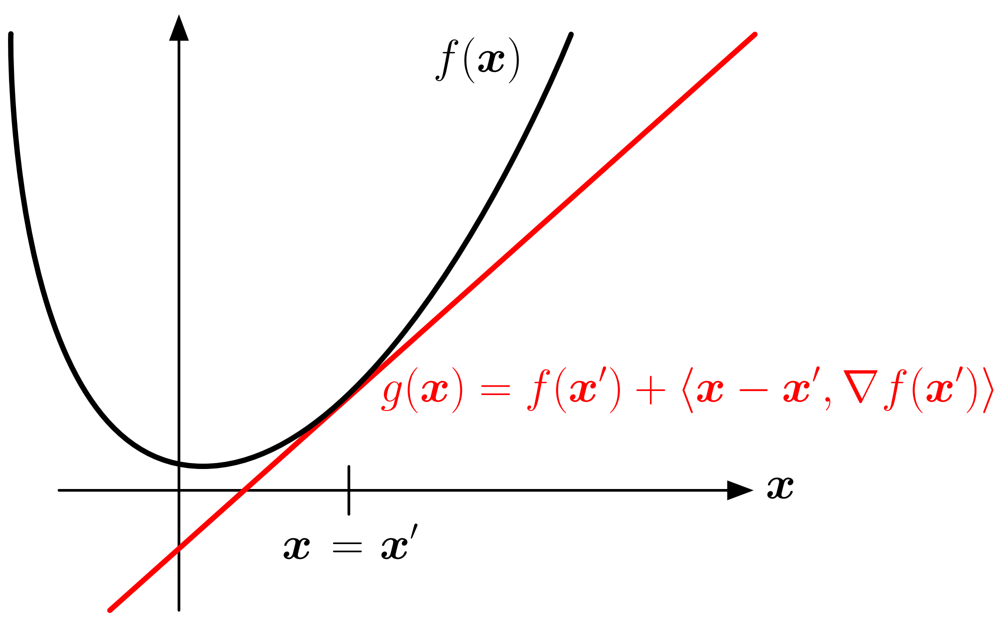
First order conditions for convexity

If f is differentiable, then it is convex if and only if

$$f(\boldsymbol{x}) \;\geq\; f(\boldsymbol{x}') + \langle \boldsymbol{x} - \boldsymbol{x}', \nabla f(\boldsymbol{x}') \rangle \tag{2}$$

for all $\boldsymbol{x}, \boldsymbol{x}' \in \mathbb{R}^N$.

This can perhaps be understood more easily in a picture:



Said again, (2) tells us that the linear approximation of f formed from the tangent line (or plane, or hyperplane, as we move to higher dimensions) will always remain *below* f. This is an incredibly useful fact, and if we never had to worry about functions that were not differentiable, we might actually just take this as the definition of a convex function.

We now prove this result. It is easy to show that if f is convex and differentiable, then we must have (2). Specifically, since f is convex, we have that for any $\theta \in [0,1]$,

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{x}') \leq \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{x}').$$

Rearranging this, we have

$$\begin{aligned} f(\boldsymbol{x}) &\geq \frac{f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{x}') - (1-\theta) f(\boldsymbol{x}')}{\theta} \\ &= f(\boldsymbol{x}') + \frac{f(\boldsymbol{x}' + \theta(\boldsymbol{x} - \boldsymbol{x}')) - f(\boldsymbol{x}')}{\theta}. \end{aligned}$$

The inequality in (2) follows from this by taking the limit as $\theta \to 0$. To see this, recall (from our review of multivariable calculus) that the inner product between the gradient of f evaluated at $\boldsymbol{x}'$ and another vector $\boldsymbol{u}$ is the directional derivative of f in the direction of $\boldsymbol{u}$; setting $\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{x}'$ this is exactly the same as

$$\langle \boldsymbol{x} - \boldsymbol{x}', \nabla f(\boldsymbol{x}') \rangle = \lim_{\theta \to 0} \frac{f(\boldsymbol{x}' + \theta(\boldsymbol{x} - \boldsymbol{x}')) - f(\boldsymbol{x}')}{\theta}.$$

We next need to show that if (2) holds, then f is convex. To do so, let $\boldsymbol{x} \neq \boldsymbol{y}$ be arbitrary vectors in $\mathbb{R}^N$ and fix $\theta \in [0,1]$. Set $\boldsymbol{z} = \theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}$. From (2) we have

$$f(\boldsymbol{x}) \geq f(\boldsymbol{z}) + \langle \boldsymbol{x} - \boldsymbol{z}, \nabla f(\boldsymbol{z}) \rangle$$

and

$$f(\boldsymbol{y}) \geq f(\boldsymbol{z}) + \langle \boldsymbol{y} - \boldsymbol{z}, \nabla f(\boldsymbol{z}) \rangle$$

If we multiply the first inequality by θ, the second by $1-\theta$, and then add the two, then since $\theta(\boldsymbol{x} - \boldsymbol{z}) + (1-\theta)(\boldsymbol{y} - \boldsymbol{z}) = \mathbf{0}$, we obtain

$$\theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}) \geq f(\boldsymbol{z}) = f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}),$$

which is exactly the definition of a convex function.

Second-order conditions for convexity

Recall that we say that $f : \mathbb{R}^N \to \mathbb{R}$ is **twice differentiable** if the Hessian matrix $\nabla^2 f(\boldsymbol{x})$ exists for every $\boldsymbol{x} \in \mathbb{R}^N$.

> If f is twice differentiable, then it is convex if and only if the Hessian matrix $\nabla^2 f(\boldsymbol{x})$ is positive semidefinite (meaning its eigenvalues are all nonnegative) for all $\boldsymbol{x} \in \mathbb{R}^N$.

Note that for a one-dimensional function $f : \mathbb{R} \to \mathbb{R}$, the above condition just reduces to $f''(x) \geq 0$. You can prove the one-dimensional version relatively easy (although we will not do so here) using the first-order characterization of convexity described above and the definition of the second derivative. You can then prove the general case by considering the function $g(t) = f(\boldsymbol{x} + t\boldsymbol{v})$. To see how, note that if f is convex and twice differentiable, then so is g. Using the chain rule, we have

$$g''(t) = \boldsymbol{v}^{\mathrm{T}} \nabla^2 f(\boldsymbol{x} + t\boldsymbol{v}) \boldsymbol{v}.$$

Since g is convex, the one-dimensional result above tells us that $g''(0) \geq 0$, and hence $\boldsymbol{v}^{\mathrm{T}} \nabla^2 f(\boldsymbol{x}) \boldsymbol{v} \geq 0$. Since this has to hold for any $\boldsymbol{v}$, this means that $\nabla^2 f(\boldsymbol{x})$ is positive semidefinite. The proof that $\nabla^2 f(\boldsymbol{x})$ being positive semidefinite implies convexity follows a similar strategy.

Examples

- **Quadratic functionals:** The function

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{x} + r,$$

  where $\boldsymbol{P}$ is symmetric, has $\nabla^2 f(\boldsymbol{x}) = \boldsymbol{P}$, so $f(\boldsymbol{x})$ is convex if and only if $\boldsymbol{P}$ is positive semidefinite.

- **Least-squares:** The least squares objective function

$$f(\boldsymbol{x}) = \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_2^2,$$

  where $\boldsymbol{A}$ is an arbitrary $M \times N$ matrix, has $\nabla^2 f(\boldsymbol{x}) = 2\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, so $f(\boldsymbol{x})$ is convex for any $\boldsymbol{A}$.

Strict convexity

It is relatively straightforward to show that for f differentiable, *strict* convexity is equivalent to (2) holding with a strict inequality. It is also easy to show that if $\nabla^2 f(\boldsymbol{x})$ is strictly positive definite (all of its eigenvalues are strictly positive) for all $\boldsymbol{x}$, then f is strictly convex. For example, the function $f(x) = x^2$ has $f''(x) = 2$ for all x and is strictly convex.

However, it is *not* the case that f being strictly convex implies that $\nabla^2 f(\boldsymbol{x})$ is positive definite for all $\boldsymbol{x}$. As an example, consider the function $f(x) = x^4$. This function *is* strictly convex, but also has $f''(0) = 0$.