2 Review of convexity

2.1 Convex sets

Definition 2.1. A set $C \subset \mathbb{R}^n$ is convex if for any $x, y \in C$ and $\lambda \in [0,1]$, $\lambda x + (1-\lambda)y \in C$.

Proposition 2.1. *If C is a convex set, and $A : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map, then $A(C)$ is convex. If $(C_j)_{j\in J}$ is a collection of convex sets, then $C = \cap_{j\in J} C_j$ is convex.*

Examples:
A *halfspace* is a set of the form $\{x \in \mathbb{R}^n : \langle a, x\rangle \leq b\}$ where $a \neq 0$,

$$\langle a, x\rangle = a^T x = \sum_{i=1}^n a_i x_i$$

is the Euclidean inner product.
An intersection of (any number of) halfspaces is a convex set. If C is an intersection of a *finite* number of halfspaces it is called a *convex polyhedron.*
It turns out that any *closed* convex set can be written as an intersection of halfspaces! This can be proved using the following fundamental fact about convex sets.

Theorem 2.1 (Separating hyperplane theorem)**.** *Let $C \subset \mathbb{R}^n$ be a convex set, and let $y \notin C$. Then there is $a \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}$ such that*

$$\langle a, x\rangle \leq b \ \forall x \in C \qquad \text{and} \qquad \langle a, y\rangle \geq b.$$

Proof. We give the proof when C is closed. The general case is left as an exercise. If C is closed we can define the projection map on C, namely $p_C(y) := \min\{\|y - x\|_2 : x \in C\}$ is well defined and satisfies $\langle y - p_C(y), x - p_C(y)\rangle \leq 0$ for any $x \in C$. Let $a = y - p_C(y)$ and $b = \langle a, p_C(y)\rangle + \frac{1}{2}\|a\|_2^2$. Note that $\langle a, y\rangle - b = \|a\|_2^2 - \frac{1}{2}\|a\|_2^2 > 0$. Also for any $x \in C$ we have $\langle a, x\rangle - b = \langle y - p_C(y), x - p_C(y)\rangle - \frac{1}{2}\|a\|_2^2 < 0$ which is what we wanted. □

EXERCISE: Use theorem above to prove that if C is a closed convex set, then C is equal to the intersection of halfspaces that contain it.

Supporting hyperplane: The result above can be used to prove the existence of *supporting hyperplanes.* If C is a closed convex set, and $y \in C \setminus \textbf{int}\, C$, then there is a hyperplane that supports C at y, i.e., there is $a \in \mathbb{R}^n \setminus \{0\}$ and $b \in \mathbb{R}^n$ such that $\langle a, y\rangle = b$ and $\langle a, x\rangle \leq b$ for all $x \in C$.

2.2 Convex functions

Let $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$.

Definition 2.2. A function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex for any $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y). \tag{1}$$

The domain of f is the set of x where $f(x)$ is finite: $\mathbf{dom}(f) = \{x \in \mathbb{R}^n : f(x) < \infty\}$. Note that $\mathbf{dom}(f)$ is a convex set by (1).

The indicator function of a convex set $C \subset \mathbb{R}^n$ is

$$I_C(x) = \begin{cases} 0 & \text{if } x \in C \\ +\infty & \text{else.} \end{cases}$$

If $f : C \to \mathbb{R}$ is a convex function defined on a convex set C, then we can always think of f as an (extended-valued) convex function $\mathbb{R}^n \to \bar{\mathbb{R}}$, by considering $f + I_C$ which takes the value $f(x)$ for $x \in C$, and $+\infty$ otherwise.

The *epigraph* of a convex function f is defined as

$$\mathbf{epi}(f) = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq t\}.$$

It is easy to see that f is convex if, and only if, $\mathbf{epi}(f)$ is a convex set.

Observe that $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex if, and only if, its restriction to any one-dimensional line is convex, i.e., for any $x \in \mathbf{dom}(f)$ and any $h \in \mathbb{R}^n$, the function $t \mapsto f(x + th)$ is convex on the interval $\{t \in \mathbb{R} : x + th \in \mathbf{dom}(f)\}$. This reduces the problem of proving convexity of a multivariate function, to that of univariate functions.

EXERCISE: Use this to show that the function $X \in \mathbf{S}^n_{++} \mapsto -\log\det X$ is convex on the set $\mathbf{S}^n_{++}$ of positive definite matrices.

Recall that a univariate differentiable function $f : \mathbb{R} \to \mathbb{R}$ is convex if, and only if, f' is nondecreasing. When f is twice differentiable this is equivalent to $f'' \geq 0$.

The following proposition gives another useful method for proving convexity.

Proposition 2.2. *Let $(f_j)_{j \in J}$ be a collection of convex functions defined on $\mathbb{R}^n$. Then $f(x) = \sup_{j \in J} f_j(x)$ is convex.*

EXERCISE: Use the above proposition to show that $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(x) =$ sum of k largest components of x, is a convex function.

EXERCISE (Pointwise infimum of a jointly convex function is convex): Let $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be a convex function, and let $f(x) = \inf_{y \in \mathbb{R}^m} g(x, y)$. Prove that f is convex.

EXERCISE: Let $f : \mathbb{R} \to \mathbb{R}$ be a function defined on the real line. For $a \neq b$, let $f[a, b] = \frac{f(b) - f(a)}{b - a}$ be the first order finite difference, and for a, b, c distinct, let $f[a, b, c] = (f[a, b] - f[b, c])/(a - c)$ be the second order finite differences. Show that f is convex if, and only if, $f[a, b, c] \geq 0$ for all $a, b, c \in \mathbf{dom}\, f$.

Differentiable functions A function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is *differentiable* at $x \in \mathbf{int}\,\mathbf{dom}(f)$ if there is a vector, denoted $\nabla f(x)$ and called the *gradient* of f at x, s.t.

$$f(x + h) = f(x) + \langle \nabla f(x), h \rangle + o(\|h\|), \qquad (h \to 0).$$

Note that $\nabla f(x) = (\partial f/\partial x_1(x), \ldots, \partial f/\partial x_n(x))$.

Proposition 2.3. *If $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex and differentiable at x, then for any $y \in \mathbb{R}^n$*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle . \tag{2}$$

Conversely, if f is differentiable everywhere, and (2) holds for all $x, y \in \mathbf{dom}(f)$ then f is convex.

Proof. Let $h = y - x$. Convexity tells us that $f(x) + t\langle \nabla f(x), h\rangle + o(t) = f(x+th) \leq (1-t)f(x) + tf(y)$. Rearranging, and dividing by t gives us $f(y) \geq f(x) + \langle \nabla f(x), h\rangle + o(1)$ which yields the desired inequality in the limit $t \to 0$.

For the converse, let $x, y \in \mathbf{dom}(f)$ and let $t \in [0,1]$. We show that $f(z) \leq (1-t)f(x) + tf(y)$ where $z = (1-t)x + ty$. We have both inequalities:

$$\begin{aligned} f(x) &\geq f(z) + \langle \nabla f(z), x - z\rangle \\ f(y) &\geq f(z) + \langle \nabla f(z), y - z\rangle . \end{aligned}$$

Taking the $(1-t, t)$ linear combination of both inequalities, and using the fact that $x - z = t(x-y)$ and $y - z = (1-t)(y-x)$ we get $(1-t)f(x) + tf(y) \geq f(z)$ as desired. ☐

Second derivatives We say that $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is twice differentiable at x if

$$f(x+h) = f(x) + \langle \nabla f(x), h\rangle + \frac{1}{2}\left\langle h, \nabla^2 f(x) h\right\rangle + o(\|h\|^2), \qquad (h \to 0).$$

for some $n \times n$ symmetric matrix $\nabla^2 f(x)$, called the *Hessian* of f at x. Note that

$$[\nabla^2 f(x)]_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

Recall that a $n \times n$ symmetric matrix A is *positive semidefinite* if $\langle x, Ax\rangle \geq 0$ for all $x \in \mathbb{R}^n$; equivalently, if all the eigenvalues of A are nonnegative. A matrix is *positive definite* if $\langle x, Ax\rangle > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$, or equivalently if all the eigenvalues of A are positive.

Proposition 2.4. *If $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is convex and twice differentiable at some $x \in \mathbf{dom}(f)$, then $\nabla^2 f(x) \succeq 0$.*

Conversely, if f is twice differentiable for all $x \in \mathbf{dom}(f)$, and $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbf{dom}(f)$, then f is convex.

Proof. For any direction h, let $\phi(t) = f(x+th)$ and note that ϕ is convex. Since f is twice differentiable, ϕ is twice-differentiable at $t = 0$ with $\phi''(0) = \left\langle h, \nabla^2 f(x) h\right\rangle \geq 0$, by convexity of ϕ. This is true for all h, and so $\nabla^2 f(x) \succeq 0$.

Conversely, assume $\nabla^2 f(x) \succeq 0$ for all $x \in \mathbf{dom}(f)$. For $x \in \mathbf{dom}(f)$ and a direction h, let $\phi(t) = f(x+th)$. Since f is twice differentiable everywhere, the same is true for ϕ, and we have $\phi''(t) = \left\langle h, \nabla^2 f(x+th)h\right\rangle \geq 0$ since $\nabla^2 f \succeq 0$. Thus ϕ is convex. This is true for all x, h and so f is convex. ☐

2.3 Some quantitative aspects

Recall that if $\|\cdot\|$ is a norm on $\mathbb{R}^n$, then the dual norm is defined by

$$\|y\|_* = \sup_{\|x\|=1} \langle y, x\rangle .$$

In particular we have the generalized Cauchy-Schwarz inequality

$$\langle x, y\rangle \leq \|x\| \|y\|_*$$

for any $x, y \in \mathbb{R}^n$.

EXERCISE: Show that the dual norm of the Euclidean norm $\|x\|_2 = \sqrt{\langle x, x\rangle}$ is the Euclidean norm. More generally show the dual of the p-norm $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$ is the q norm where $1/p + 1/q = 1$.

L-smoothness We say that a differentiable function $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is *L-smooth* with respect to a norm $\|\cdot\|$, if for any $x, y \in \mathbf{int}\,\mathbf{dom}(f)$,

$$\|\nabla f(x) - \nabla f(y)\|_* \leq L\|x - y\|, \tag{3}$$

where $\|\cdot\|_*$ is the dual norm to $\|\cdot\|$.

(We will sometimes omit the reference to the norm, in which case this means we work with the Euclidean norm.)

The following proposition gives an equivalent characterization of L-smoothness for convex functions.

Proposition 2.5. *(i) If f is L-smooth, then for any $x \in \mathbf{int}\,\mathbf{dom}(f)$ and $y \in \mathbf{dom}(f)$,*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|^2. \tag{4}$$

(ii) Conversely, if f is convex *and differentiable on $\mathbf{dom}(f)$ and (4) holds for all $x, y \in \mathbf{dom}(f)$ then ∇f satisfies the Lipschitz assumption (3).*
(iii) If f is convex and twice differentiable on $\mathbf{dom}(f)$, then L-smoothness is equivalent to

$$\left\langle h, \nabla^2 f(x) h \right\rangle \leq L\|h\|^2$$

for all $x \in \mathbf{dom}(f)$ and $h \in \mathbb{R}^n$.

Proof. (i) Let $h = y - x$ and $\phi(t) = f(x + th) - (f(x) + t\langle \nabla f(x), h \rangle)$. Then ϕ is differentiable and $\phi'(t) = \langle \nabla f(x + th) - \nabla f(x), h \rangle \leq \|\nabla f(x + th) - \nabla f(x)\|_* \|h\| \leq Lt\|h\|^2$ where we used the Lipschitz assumption (3). Thus it follows that $\phi(1) = \phi(0) + \int_0^1 \phi'(t)dt \leq L/2\|h\|^2$ which gives precisely the desired inequality (4). □

Strong convexity We say that f is m-strongly convex (with respect to the norm $\|\cdot\|$) if for any $x, y \in \mathbf{dom}(f)$, and $t \in [0, 1]$

$$f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) - \frac{m}{2}t(1 - t)\|x - y\|^2. \tag{5}$$

The following proposition gives an equivalent characterization of strong convexity.

Proposition 2.6. *(i) Let $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ be m-strongly convex. If f is differentiable at x, then for any $y \in \mathbf{dom}(f)$ we have*

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle + \frac{m}{2}\|y - x\|^2. \tag{6}$$

(ii) Conversely if f is differentiable on $\mathbf{dom}(f)$ and satisfies the above for all $x, y \in \mathbf{dom}(f)$, then it is m-strongly convex.
(iii) If f is twice differentiable on its domain, then strong convexity is equivalent to $\left\langle h, \nabla^2 f(x) h \right\rangle \geq m\|h\|^2$ for all $x \in \mathbf{dom}(f)$ and $h \in \mathbb{R}^n$.

Proof. (i) Let $h = y - x$. Strong convexity tells us that $f(x) + t\langle \nabla f(x), h \rangle + o(t) = f(x + th) \leq (1 - t)f(x) + tf(y) - (m/2)t(1 - t)\|h\|^2$. Rearranging, and dividing by t gives us $f(y) \geq f(x) + \langle \nabla f(x), h \rangle + (m/2)(1 - t)\|h\|^2 + o(1)$ which yields the desired inequality in the limit $t \to 0$.

(ii) Conversely, assume (6) holds for all $x, y \in \mathbf{dom}(f)$. Then using the same argument as the proof of Proposition 2.3 we get that (5) holds.

(iii) Assume now that f is twice differentiable at x. Let $\phi(t) = f(x+th) - (f(x) + t\langle \nabla f(x), h\rangle + (m/2)t^2\|h\|^2)$. Strong convexity tells us that $\phi(t) \geq 0$. Furthermore $\phi(0) = 0$. Thus, necessarily $\phi''(0) \geq 0$, i.e., $\langle h, \nabla^2 f(x)h \rangle \geq m\|h\|^2$. Conversely, assume $\langle h, \nabla^2 f(z)h \rangle \geq 0$ for all $z \in \mathbf{dom}(f)$, and all h. Then, using the same definition as ϕ above (with $h = y - x$), we have $\phi''(t) = \langle h, \nabla^2 f(x+th)h \rangle - m\|h\|^2 \geq 0$. Thus this means that ϕ' is increasing, and since $\phi(0) = \phi'(0) = 0$, this means that $\phi(1) \geq 0$, i.e., that (6) holds. $\square$

Remark 1. *When considering the Euclidean norm, we see that f is L-smooth if, and only if, $\nabla^2 f(x) \preceq LI$, i.e., $LI - \nabla^2 f(x)$ is positive semidefinite, i.e., all the eigenvalues of f are $\leq L$. Similarly, a function f is strongly convex if, and only if, $\nabla^2 f(x) \succeq mI$, i.e., all the eigenvalues of $\nabla^2 f(x)$ are $\geq m$.*

To summarize, if a function f is L-smooth, and m-strongly convex, then we can find, at any point $x \in \mathbf{int}\,\mathbf{dom}(f)$ quadratic lower and upper bounds on f:

$$\underbrace{f(x) + \langle \nabla f(x), y - x\rangle + \frac{m}{2}\|y - x\|^2}_{\text{strong convexity}} \;\leq\; f(y) \;\leq\; \underbrace{f(x) + \langle \nabla f(x), y - x\rangle + \frac{L}{2}\|y - x\|^2}_{L\text{-smoothness}}$$