10-425/625: Introduction to Convex Optimization **(Fall 2023)**

Lecture 4: Convex Functions

Instructor:[1] *Matt Gormley* *September 8, 2023*

4.1 Sets

Background:

Note that in last lecture, I *incorrectly* claimed on the chalkboard that a boundary point was a point $x \in S$; the notes were correct though: a boundary point is not necessarily in S, it is just a point $x \in \mathbb{R}^n$.

Definition 4.1 (Boundary Point)**.** *We say that a* boundary point *satisfies the property that are points both in S and not in S that are arbitrarily close. That is,* $x \in \mathbb{R}^n$ *is a boundary point of* S *if for all* $\epsilon > 0$, $\exists y \in S$ *and* $\exists z \notin S$ *such that*

$$||y - x||_2 \leq \epsilon \text{ and}$$
$$||z - x||_2 \leq \epsilon$$

Definition 4.2 (Boundary)**.** *The* boundary *of a set* $S \subseteq \mathbb{R}^n$ *are all points in* $\mathbb{R}^n$ *that are boundary points.*

The above definition of a boundary has two consequences of note: First, all points in S that are not interior points are boundary points. Second, not all boundary points are in S, e.g. an open set contains none of its boundary points.

[1]These notes were originally written by Siva Balakrishnan for 10-725 Spring 2023 (original version: here) and were edited and adapted for 10-425/625.

4.2 Operations which Preserve Convexity of a Set

There are some important operations which preserve convexity of sets:

1. **Intersection:** The intersections of convex sets is a convex set.
2. **Scaling and Translation:** If C is convex, then
$$aC + b := \{ax + b : x \in C\},$$
is convex for any $a, b \in \mathbb{R}$.
3. **Affine Images and Pre-Images:** Let us define $f(x) = Ax + b$ to be an affine function. If C is a convex set, then, the *affine image*
$$f(C) = \{f(x) : x \in C\}$$
is also a convex set. Also, the *affine pre-image*
$$f^{-1}(C) = \{x : f(x) \in C\},$$
is a convex set.

There are a couple more that are more involved but useful to know (we may not have time to cover this in lecture, in which case we will re-visit it when we need it).

4. **Perspective:** The perspective function $P : \mathbb{R}^d \times \mathbb{R}_{++} \mapsto \mathbb{R}$ (where $\mathbb{R}_{++}$ is the strictly positive reals), is defined as:
$$P(x, t) = x/t.$$
If $C \subseteq \text{dom}(P)$ is a convex set, then its image $P(C)$ is a convex set, and similarly if D is convex then $P^{-1}(D)$ is convex.
5. **Linear-Fractional:** The linear fractional function for a given A, b, c, d is given by:
$$f(x) = \frac{Ax + b}{c^T x + d}.$$
If $C \subseteq \text{dom}(f)$ is a convex set, then its image $f(C)$ is a convex set, and similarly if D is convex then $f^{-1}(D)$ is convex.

Example 4.3. Conditional Probability Set: This is an example of using the linear-fractional image to characterize convexity. Let U, V be random variables over $\{1, \ldots, n\}$ and $\{1, \ldots, m\}$. Let $C \subseteq \mathbb{R}^{nm}$ be a set of joint distributions for U, V, i.e., each $p \in C$ defines joint probabilities

$$p_{ij} = \mathbb{P}(U = i, V = j)$$

Let $D \subseteq \mathbb{R}^{nm}$ contain corresponding *conditional distributions*, i.e., each $q \in D$ defines

$$q_{ij} = \mathbb{P}(U = i | V = j)$$

Assume C is convex. Let's prove that D is convex. Write

$$D = \left\{ q \in \mathbb{R}^{nm} : q_{ij} = \frac{p_{ij}}{\sum_{k=1}^{n} p_{kj}}, \text{ for some } p \in C \right\} = f(C)$$

where f is a linear-fractional function, hence D is convex.

4.3 Convex Functions

Background: (Writing Dot Products) There are three ways of writing the dot product of two vectors $x, y \in \mathbb{R}^n$:

$$x \cdot y = x^T y = \langle x, y \rangle$$

Background: (Gradient and Hessian) Suppose we have a function $f : \mathbb{R}^n \to \mathbb{R}$. The *gradient* of f is a vector $g = \nabla f(x) \in \mathbb{R}^n$ who entries are the first-order partial derivatives of the function, i.e., $g_i = [\nabla f(x)]_i = \frac{\partial f(x)}{\partial x_i}$.

The *Hessian* of f is a symmetric matrix $H = \nabla^2 f(x) \in \mathbb{R}^{n \times n}$, whose entries are the second-order partial derivatives of the function, i.e., $H_{i,j} = [\nabla^2 f(x)]_{i,j} = \frac{\partial f(x)}{\partial x_i \partial x_j}$.

Background: (Taylor Series Approximation) Given a function $f : \mathbb{R}^n \to \mathbb{R}$, its first-order Taylor series approximation at a given point $y \in \mathbb{R}^n$ is:

$$f(x) \approx T_{1st}(x) = f(y) + \langle \nabla f(y), x - y \rangle$$

The second-order Taylor series approximation involves the gradient and the Hessian $H = \nabla^2 f(x)$:

$$f(x) \approx T_{2nd}(x) = T_{1st}(x) + \frac{1}{2}(x - y)^T H (x - y)$$

There are three characterizations of convexity that you should be familiar with:

1. **No Assumptions (Zeroth-Order):** This is the definition we discussed last time, i.e. f is convex if its domain is a convex set and, for any $x, y \in \mathrm{dom}(f)$,

$$f(\theta x + (1 - \theta) y) \leq \theta f(x) + (1 - \theta) f(y).$$

2. **Differentiable (First-Order):** Suppose our function f has a derivative (at all points in its domain) then, f is convex if its domain is a

convex set and, for any $x, y \in \text{dom}(f)$,

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle.$$

3. **Twice Differentiable (Second-Order):** A function f is convex, if its domain is a convex set and, for any $x \in \text{dom}(f)$,

$$\nabla^2 f(x) \succeq 0.$$

It is also worth noting that there is a definition analogous to (2) above in the case when the function is not differentiable everywhere.

2'. **Non-Smooth:** A function f is convex if its domain is a convex set, and if at every point $x \in \text{dom}(f)$, there exists a vector g_x such that, for any $y \in \text{dom}(f)$,

$$f(y) \geq f(x) + \langle g_x, y - x \rangle.$$

It is worth noting that if f is differentiable at x, then there is only one vector which will satisfy the above definition and it will coincide with the usual gradient, i.e. $g_x = \nabla f(x)$.

Any g_x which satisfies the above property is called a *subgradient* of f at x. The set of all subgradients at a point x is called the *subdifferential* of f at x and it will be denoted as $\partial f(x)$.

Except for some very pathological functions (and only at the boundary of their domain) subgradients always exist. Formally, one can for instance show that a subgradient g_x of a convex function f at x exists if x is in the interior of their domain.

Notational Note: I will often stop adding the qualifiers "for $x, y \in \text{dom}(f)$". One way to make this precise (I, and most textbooks do this implicitly) is to allow f to be whats called an *extended* function, and define it to be ∞ outside its (effective) domain. This won't change any of its convexity properties, and things like the first and zeroth-order characterizations will now make sense for any $x, y \in \mathbb{R}^d$.

4.3.1 Example: A Quadratic Objective

Let us consider the quadratic function

$$f(x) = \frac{1}{2}x^T Q x + a^T x + b$$

where $Q \succeq 0$. Consider its derivatives:

$$\nabla f(x) = Qx + a$$
$$\nabla^2 f(x) = Q$$

Applying definition (3) is easiest, since $\nabla^2 f(x) = Q$ and this is PSD.

Now, let us try to apply definition (2). It is a differentiable function, with gradient $\nabla f(x) = Qx + a$. So we need to verify if,

$$\frac{1}{2}y^T Q y + a^T y + b \overset{?}{\geq} \frac{1}{2}x^T Q x + a^T x + b + \langle Qx + a, y - x \rangle.$$

Re-arranging we obtain that we need to check if,

$$\frac{1}{2}(y - x)^T Q (y - x) \geq 0,$$

which is certainly the case since $Q \succeq 0$.

Finally, let us try to apply definition (1). We see (after cancelling some terms) that we need to verify if for $0 \leq \theta \leq 1$,

$$\frac{1}{2}\left(\theta x + (1-\theta)y\right)^T Q \left(\theta x + (1-\theta)y\right) \overset{?}{\leq} \frac{\theta}{2}x^T Q x + \frac{1-\theta}{2}y^T Q y.$$

Now, use the fact (you should see how you might prove this fact) that, $x^T Q y \leq \frac{1}{2}\left[x^T Q x + y^T Q y\right]$ for PSD Q (this is the matrix analogue of the simple fact that $ab \leq (a^2 + b^2)/2$), to verify that the desired inequality holds.

4.3.2 More Examples of Convex Functions

Here are a few examples of convex functions:

1. $\exp(ax)$ is convex for any a over $\mathbb{R}$.
2. $\log x$ is concave on $\mathbb{R}_{++}$.

3. $a^T x + b$ is convex (and concave).
4. The least squares loss $\|Ax - b\|^2$ is convex (for any A, b).
5. Any norm is convex, i.e. $\|x\|$ is a convex function.
6. The spectral norm, and the trace norm of a matrix are convex, i.e. $\|X\|_{\text{op}} = \sigma_1(X)$, $\|X\|_{\text{tr}} = \sum_{i=1}^{d} \sigma_i(X)$ where $\sigma_i(X)$ denotes the i-th singular value of X.
7. **Convex Indicators:** If C is a convex set, then the indicator function (which is defined on the extended reals):

$$I_C(x) = \begin{cases} 0 & x \in C \\ \infty & x \notin C. \end{cases}$$

is convex.

4.4 Operations which Preserve Convexity of a Function

1. **Non-negative Linear Combination:** Suppose $f_1, \ldots, f_m$ are convex, then so is $\sum_{i=1}^{m} a_i f_i$ for any $a_1, \ldots, a_m \geq 0$.
2. **Pointwise Max:** If the collection of functions f_s for $s \in S$ are convex, then so is $g(x) = \sup_{s \in S} f_s(x)$.
3. **Partial Minimization:** If $g(x, y)$ is a convex function in $\mathbb{R}^{n+m}$ where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, and $C \subseteq \mathbb{R}^m$ is a convex set, then $f(x) = \min_{y \in C} g(x, y)$ is a convex function.

   An Example:

   - Suppose C is an arbitrary set, consider $f(x) = \max_{y \in C} \|x - y\|$. Intuitively, f tells us how far x is from the *farthest* point in C. f is convex. To see this, we can view f as a maximum of convex functions $f_y(x) = \|x - y\|$.
   - Let C be a convex set, consider $f(x) = \min_{y \in C} \|x - y\|$. Intuitively, f tells us how far x is from the *closest* point in C. f is convex. We can view this as a partial minimization of the function $g(x, y) = \|x - y\|$ which is a convex function (in (x, y)).

Function compositions:

4. **Affine Composition:** If f is convex then so is $g(x) = f(Ax + b)$.

5. **General Composition:** Suppose that $f = h \circ g$, where $g : \mathbb{R}^d \mapsto \mathbb{R}$, $h : \mathbb{R} \mapsto \mathbb{R}$, $f : \mathbb{R}^d \mapsto \mathbb{R}$. Then one can ask when f is convex. There are many cases to cover (see BV) but we'll simply study one, and try to understand where it comes from: f is convex if h is convex and nondecreasing, g is convex.

   Proof: To see this: imagine everything was twice differentiable, then by the chain rule

   $$f'(x) = h'(g(x))g'(x) \qquad f''(x) = h''(g(x))(g'(x))^2 + h'(g(x))g''(x).$$

   When h is convex and non-decreasing, h'' and h' are positive, and when g is convex, g'' is positive, so f'' is positive. ■

4.5 Example: Support Vector Machines (SVMs)

4.5.1 Data

Suppose we have a dataset $\mathcal{D} = \{\mathbf{x}^{(i)}, y^{(i)}\}_{i=1}^N$ where $\mathbf{x}^{(i)} \in \mathbb{R}^m$ and $y^{(i)} \in \{+1, -1\}$. We wish to learn a linear decision boundary separating the points i labeled $+1$ from those labeled -1.

4.5.2 SVM Mathematical Programs

One way to do this is with a support vector machine (SVM), which finds the linear decision boundary with largest margin. We'll consider two versions:

Hard-margin SVM (Primal)

$$\begin{aligned}
&\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|_2^2 \\
&\text{s.t. } y^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)} + b) \geq 1, \quad \forall i = 1, \ldots, N
\end{aligned}$$

Soft-margin SVM (Primal)

$$\begin{aligned}
\min_{\mathbf{w},b}\ & \frac{1}{2}\|\mathbf{w}\|_2^2 + C\left(\sum_{i=1}^{N} e_i\right) \\
\text{s.t. } & y^{(i)}(\mathbf{w}^T\mathbf{x}^{(i)} + b) \geq 1 - e_i, \quad \forall i = 1, \ldots, N \\
& e_i \geq 0, \quad \forall i = 1, \ldots, N
\end{aligned}$$