10-725/36-725: Convex Optimization **Fall 2018**

Lecture 2: August 29, 2018

Lecturer: Ryan Tibshirani *Scribes: Yingjing Lu, Adam Harley, Ruosong Wang*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

2.1 Recap: Convexity and Why

Definition 2.1. *A convex optimization problem is of the form:*

$$\min_{x \in D} f(x) \tag{2.1}$$

Subject to:

$$\begin{aligned} g_i(x) &\leq 0,\ i = 1, \ldots, m \\ h_j(x) &= 0,\ j = 1, \ldots, r \end{aligned} \tag{2.2}$$

Where f and $\{g_i\}_{i=1}^m$ are all convex and $\{h_j\}_{j=1}^r$ are affine. Convex optimization has special property that any local minimizer is a *global minimizer*.

2.2 Convex Sets

Definition 2.2. *A set $C \subseteq \mathbb{R}^n$ is a* convex set *if for any $x, y \in C$, we have $tx + (1-t)y \in C$ for all $t \in [0, 1]$.*

Another way to put in slide is that "line segment joining any two elements lies entirely in set"

Definition 2.3. Convex combination *of $x_1, \ldots, x_k \in \mathbb{R}^n$ is any linear combination of the form*

$$\theta_1 x_1 + \ldots + \theta_k x_k \text{ with } \theta_i \geq 0 \text{ and } \sum_{i=1}^k \theta_1 = 1.$$

Definition 2.4. *For a set of $C \subseteq \mathbb{R}^n$, the* convex hull conv(C) *is the smallest convex set that contains those vectors.*

Notice that a convex hull is always convex.

2.2.1 Convex Set Examples

2.2.1.1 Some simple ones

- Empty set, point, line.

- **Norm ball:** $\{x : \|x\| \leq r\}$, for given norm $\|\cdot\|$, radius r.
- **Hyperplane:** $\{x : a^T x = b\}$ for given a, b.
- **Half space:** $\{x : a^T x \leq b\}$.
- **Affine space:**$\{x : Ax = b\}$ for given matrix A and vector b.
- **Polyhedron:** $\{x : Ax \leq b\}$ ($\leq$ is interpreted component-wise). The set $\{x : Ax \leq b, Cx = d\}$ is also polyhedron.
- **Simplex** $\mathrm{conv}(C)$ for a set $C \subseteq \mathbb{R}^n$ of affinely independent points.

2.2.2 Cones and convexity

Definition 2.5. *A set $C \subseteq \mathbb{R}^n$ is a* cone *if for any $x \in C$ we have $tx \in C$ for all $t \geq 0$.*

Definition 2.6. *A set $C \subseteq \mathbb{R}^n$ is a* convex cone *if for any $x_1, x_2 \in C$ we have $t_1 x_1 + t_2 x_2 \in C$ for all $t_1, t_2 \geq 0$.*

Definition 2.7. Conic combination *of $x_1, \ldots, x_k \in \mathbb{R}^n$ is any linear combination of the form*

$$\theta_1 x_1 + \ldots + \theta_k x_k \text{ with } \theta_i \geq 0.$$

Definition 2.8. *For a set of vectors $x_1, \ldots, x_k \in \mathbb{R}^n$, the* conic hull *is the smallest set that contains all conic combinations of $x_1, \ldots, x_k$.*

2.2.3 Examples of convex cones

- **Norm cone**: $\{(x, t) : \|x\| \leq t\}$ for any norm. The cone is called *second-order cone* when the norm is the ℓ_2 norm.
- **Normal cone**: for any set $C \subseteq \mathbb{R}^n$ and any $x \in C$ define $\mathcal{N}_c(x) = \{g : g^T x \geq g^T y, \text{ for all } y \in C\}$. This is always a convex cone.
- **Positive Semi-definite cone**: $\mathbb{S}^n_+ = \{X \in \mathbb{S}^n : X \succeq 0\}$. Here we use $X \succeq$ to denote X is a positive semidefinite matrix.

2.2.4 Key properties of convex sets

- Separating hyperplane theorem: two disjoint convex sets have a separating hyperplane between them. Formally, if C, D are nonempty convex sets with $C \cap D = \emptyset$, then there exists a, b such that $C \subseteq \{x : A^T x \leq b\}$ and $D \subseteq \{x : A^T x \leq b\}$.

  If two sets do not intersect, is there always a hyperplane that strictly separates them? Consider the set $\{(x, y) : y \leq 0\}$, and also the epigraph $\{(x, y) : y \geq bx, x \geq 0\}$. The second set will get infinitely close to the first, so you cannot put a hyperplane bewteen them. This means you cannot expect that disjoint sets will be strictly separated.

- Supporting hyperplane theorem: a boundary point of a convex set has a supporting hyperplane passing through it. Formally, if C is a nonempty convex set, and $x_0 \in \mathrm{bd}(C)$, there exists a such that $C \subseteq \{x : a^T x \leq a^T x_0\}$.

2.2.5 Operations preserving convexity

- Intersection
- Scaling and translation
- Affine images and preimages. For any given affine function of the form $f(x) = Ax + b$ and C is a convex set, then $f(C) = \{f(x) : x \in C\}$ is convex. Also, if D is convex, then the inverse (or pre-image) $f^{-1}(D) = \{x : f(x) \in D\}$ is convex.

2.2.6 Example: Linear matrix inequality solution set

Given $A_1, \ldots, A_k, B \in \mathbb{S}^n$, a linear matrix inequality for a variable $x \in \mathbb{R}^k$ looks like

$$x_1 A_1 + \ldots + x_k A_k \succeq B. \tag{2.3}$$

Let's prove that the set C of points x that satisfy the above inequality is convex.

First approach: check that if two points lie in the set, then all in-between points lie in the set. We can check this by seeing that this is true for any v: $v^T \left(B - \sum_{i=1}^{k} (tx_i + (1-t)_i) A_i \right) v \geq 0$.

Second approach: define $\mathbb{S}_+^n = \{Y, Y \succeq 0\}$ which is convex, and define $f(x) = B - \sum x_i A_i$, then

$$f^{-1}(\mathbb{S}_+^N) = \{x : f(x) \in S_+^N\} = \{x : B - \sum x_i A_i \succeq 0\}, \tag{2.4}$$

which is exactly our set.

2.2.7 More operations preserving convexity

- Perspective images and preimages. (There is a relationship here to pinhole cameras!)
- Linear-fractional images and preimages: The perspective map composed with an affine function, like $f(x) = (Ax + b)/(c^T x + d)$, called a linear-fractional function, preserves convexity! So if $C \subseteq \text{dom}(f)$ is contex, then so is $f(C)$, and if D is convex then so is $f^{-1}(D)$.

2.3 Convex functions

Definition 2.9. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if dom$(f) \subset \mathbb{R}^n$ is convex, and*

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \text{ for } \leq t \leq 1 \text{ and } x, y \in \text{dom}(f).$$

In other words, the function always lies below the line segment joining $f(x)$ and $f(y)$. **Concave functions** have the opposite inequality, and f being concave implies $-f$ is convex.

Important modifiers:

- Strictly convex: same definition but with strict inequalities, and the function is strictly below the line segments. The way I think about this is: f has more curvature than a linear function. Linear functions are convex but not strictly convex.
- Strongly convex with parameter $m > 0$: this says that if you subtract a quadratic, it's still convex: $f - m/2 \|x\|^2$ is convex. In other words, f is more convex/curved than a quadratic function.

Note that strong convexity implies strict convexity, which implies convexity. All of this is analogous for concave functions!

2.3.1 Examples of convex functions

- Univariate functions (e^{ax} for any $a \in \mathbb{R}$, x^a for $a \geq 1$, x^a for $a \leq 0$ when $x \in \mathbb{R}_+$; x^a is concave for $0 \leq a \leq 1$ for $x \in \mathbb{R}_+$, $\log x$ is concave over $\mathbb{R}_{++}$.
- affine functions $a^T x + b$ are both convex and concave
- quadratic functions $\frac{1}{2}x^T Qx + b^T x + c$ are convex provided that $Q \succeq 0$ (i.e., Q is positive semidefinite).
- least squares, like $\|y - Ax\|^2$, because if you expand it, it looks like a quadratic function with $Q = A^T A$, and $A^T A$ is always positive semidefinite (since $b^T A^T Ab \geq 0$ since $z^T z = \sum z_i^2$ where $z = Ab$).
- Norms – all of them are convex! (Three things define a norm; briefly, these are: $\|\mathbf{x}\| \geq 0$, $\|\alpha\mathbf{x}\| = |\alpha| \, \|x\|$, and $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$. These also provide convexity.)

  The most common ℓ_p norms are $1, 2, \infty$. The ℓ_1 norm is good for inducing sparsity, ℓ_2 is ubiquitous because we use it to measure distances, and ℓ_∞ is useful for reasons we'll see later. (Note that the ℓ_0 "norm" does not satisfy the triangle inequality, so it is not a norm, not convex, and not our friend.) Operator (also called spectral) and trace (also called nuclear) norms are also convex.
- Indicator functions. If C is convex, then the indicator function $I_C(x)$, defined as 0 inside the set and ∞ outside, is also convex.

  Let's check this. We have $f(x) = I_C(x)$. We need to first check the domain: $\text{dom}(f) = C$ is convex by assumption; good. Now, for any convex combination of $x, y \in \text{dom}(f)$, we need $f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$. On the right hand side, $f(x)$ and $f(y)$ are zero (since x, y are in the domain), so we have $tf(x) + (1-t)f(y)) \leq t0 + (1-t)0 = 0$. On the left hand side, we know the term $(tx + (1-t)y)$ lies inside the set (since the domain C is convex), and the indicator function gives 0 for all values in the set, so we have $0 = 0$ and we are done.

2.3.2 Key properties of convex functions

- Epigraph characterization: a function f is convex $\iff$ its epigraph (which is the set of all points above the function) is a convex set. In other words, f is convex if its epigraph $\text{epi}(f) = \{(x,t) \in \mathit{dom}(f) \times \mathbb{R} : f(x) \leq t\}$ is also convex.
- Convex sublevel sets: if f is convex, then its sublevel sets, defined by $\{x \in \text{dom}(f) : f(x) \leq t\}$, are convex, for all $t \in \mathbb{R}$. The converse is not true!
- First-order characterization: if f is differentiable, then f is convex $\iff$ $\text{dom}(f)$ is convex and $f(y) \geq f(x) + \nabla f(x)^T(y - x)$ for all $\mathbf{x}, y \in \text{dom}(f)$. This means that in a differentiable convex function, $\nabla f(x) = 0 \implies x$ minimizes f.
- Second-order characterization: if f is twice differentiable, then f is convex $\iff$ $\mathit{dom}(f)$ is convex, and $\nabla^2 f(x) \succeq 0$ for all $x \in \text{dom}(f)$.

  You might wonder about that inequality. Consider $f(x) = x^4$, which is quadratic, with second derivative zero at zero. So, it's strictly convex, but the second derivative is not strictly positive.
- Jensen's inequality: if f is convex, and X is a random variable on $\text{dom}(f)$, then $f(\mathbb{E}[X]) \leq \mathbb{E}[f(x)]$.
- Log-sum-exp function: $g(x) = \log\left(\sum_{i=1}^k e^{a_i^T x + b_i}\right)$ for fixed a_i, b_i. This is often called the soft max, since it smoothly approximates $\max_{i=1,\ldots,k}(a_i^T x + b_i)$.

2.3.3 Operations preserving convexity

- Nonnegative linear combination: $f_1, \ldots, f_m$ being convex implies $a_1 f_1 + \ldots + a_m f_m$ is also convex, for any $a_i \geq 0$.
- Pointwise maximization: if f_s for $s \in S$ are all convex, then the pointwise max of them is also convex. Note that the functions can be discrete or continuous, and the set S can even be infinite!
- Partial minimization: if $g(x, y)$ is convex in x, y, and C is convex, then $f(x) = min_{y \in C} g(x, y)$ is convex.

2.3.4 Example: distances to a set

Consider the max (or min) distance to a set, under an arbitrary norm, written $f(x) = \max_{y \in C} \|x - y\|$, is convex (whether C is convex or not). This is because the norm is convex, and the pointwise max is also convex. As for the min distance to the set, it is convex as long as the set C is convex.

2.4 Contributions

- **Yingjing Lu**: Section **1.1** to **1.2.3**
- **Adam Harley**: Section **1.2.4** to the end.
- **Ruosong Wang**: Revisions.