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

Lecture 2: August 29, 2018

Lecturer: Ryan Tibshirani *Scribes: Adam Harley*

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

Convexity 1: Sets and functions

A convex optimization problem is of the form

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

subject to

$$g_i(x) \leq 0, \text{ for } i = 1, \dots, m \tag{2.2}$$

and

$$h_j(x) = 0, \text{ for } j = 1, \dots, r, \tag{2.3}$$

where f and all g_i are convex, and all h_j are affine. These problems have the special property that a local minimum is a global minimum.

From now on, we will drop the domain notation D (as in $x \in D$), and just deal with x.

2.1 Convex sets

A convex set $C \in \mathbb{R}^n$ is one where a line segment joining any two elements lies entirely inside the set. This can be written

$$x, y \in C \implies tx + (1-t)y \in C \text{ for all } 0 \leq t \leq 1. \tag{2.4}$$

A convex combination is a weighted sum of elements where the weights are non-negative and sum to one.

A convex hull of a set is all convex combinations of elements – it is convex by construction. The convex hull is the smallest convex set that contains the set.

2.1.1 Examples of convex sets

- empty set
- point
- line
- norm ball $\{x : \|x\| \leq r\}$ for a given norm, and a given radius r;

- hyperplanes $\{x : a^T x = b\}$ for a given a, b
- halfspace $\{x : a^T x \leq b\}$
- affine space $\{x : Ax = b\}$ for given A, b
- polyhedron $\{x : Ax \leq b\}$ where the inequality is interpreted component-wise. (A is a matrix, x, b are vectors, so we compare the elements of Ax and b.) Note that polyhedron is the intersection of a finite number of halfspaces (and hyperplanes).
- simplex, which is a special case of a polyhedra, given by $\text{conv}\{x_0, \ldots, x_k\}$, where these vectors are affinely independent, and conv denotes the convex hull of the points. (A set of vectors $x_1, \ldots, x_k$ is **affinely independent** if $\left[\sum_{i=1}^k a_i x_i, \sum_{i=1}^k a_i = 0\right] \iff a_i = 0$ for all i. Affine independence implies linear independence.) The canonical example of a simplex is the **probability simplex**, written $\text{conv}\{e_1, \ldots, e_n\} = \{w : w \geq 0, 1^T w = 1\}$, where the e_i are the standard unit vectors.
- convex cone. First of all, a cone is given by $C \subseteq \mathbb{R}^n$ such that $x \in C \implies tx \in C$ for all $t \geq 0$. Cones are not necessarily convex, since they might be hollow. However, convex cones are convex! These are defined by $x_1, x_2 \in C \implies t_1 x_1 + t_2 x_2 \in C$ for all $t_1, t_2 \geq 0$.
- conic combination of $x_1, \ldots, x_k \in \mathbb{R}^n$, which means any linear combination $\theta_1 x_1 + \ldots + \theta_k x_k$ with $\theta_i \geq 0$ for all i.
- conic hull: this is the union of all conic combinations of a set of points.
- norm cone (a type of convex cone): $\{(x, t) : \|x\| \leq t\}$ for any norm. Under the L_2 norm, this is a second-order cone. This cone is sometimes called the ice cream cone.
- normal cone (a type of convex cone): given any set C and point $x \in C$, we can define the "normal cone to the set" as $\mathcal{N}_C(x) = \{g : g^T x \geq g^T y \text{ for all } y \in C\}$. This can be equivalently written as $\mathcal{N}_C(x) = \{g : g^T(x - y) \geq 0 \text{ for all } y \in C\}$. In other words, if you form any vector $(x - y)$ from elements in the set, the dot product with g will be greater than zero, meaning the vectors are more than 90 degrees apart – this is what makes it "normal to" the set. Note this does not depend on the set C, because you're just forming a cone away from it.
- positive semidefinite cone (a type of convex cone): $\mathbb{S}^n_+ = \{X \in \mathbb{S}^n : X \succeq 0\}$ where $\mathbb{S}^n$ is the set of $n \times n$ symmetric matrices, and $X \succeq 0$ means that X is positive semidefinite (In more detail, $X \succeq 0$ means that the smallest eigenvalue of X is greater than zero, which also means $a^+ X a \geq 0$ for all a. We can also say $A \succeq B \implies A - B \succeq 0$.)

  Let's check to see if semidefinite cones truly are convex. We have $S^n_+ = \{x : x \geq 0\}$. Take $X, Y \in \mathbb{S}^n_+$. To prove convexity, we want $t_1 X + t_2 Y \succeq 0$. Since the matrices X, Y are positive semidefinite, we can use $a^+ X a \geq 0$ for all a to write $a^T(t_1 X + t_2 Y)a = t_1(a^T X a) + t_2(a^T Y a) \geq 0$.

2.1.2 Key properties of convex sets

- Separating hyperplane theorem: two disjoint convex sets have a separating hyperplane between them. In other words, if C, D are nonempty convex sets with $C \bigcap D = \emptyset$, 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 $x_0 \in \text{bd}(C)$, there exists a such that $C \subseteq \{x : a^T x \leq a^T x_0\}$.

2.1.3 Operations preserving convexity

- Intersection
- Scaling and translation
- Affine images and preimages. If you give me an affine function, like $f(x) = Ax + b$ (whereas linear would be $f(x) = Ax$)), and C is convex, then $f(C) = \{f(x) : x \in C$ is convex. Also, if D is convex, then the inverse (or pre-image) $f^{-1}(D) = \{x : f(x) \in D\}$ (whether it exists or not,) is convex. This will be a useful fact later.

2.1.4 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

$$x_1 A_1 + \ldots + x_k A_k \succeq B. \tag{2.5}$$

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

$$f^{-1}(\mathbb{S}^N_+) = \{x : f(x) \in S^N_+\} = \{x : B - \sum x_i A_i \succeq 0\} = \text{ our set!} \tag{2.6}$$

2.1.5 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 $f^{-1}(D)$.

2.2 Convex functions

A convex function f maps from $\mathbb{R}^n$ to $\mathbb{R}$ such that $\text{dom}(f) \subset \mathbb{R}^n$ is convex, and

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y) \text{ for } \leq t \leq 1 \text{ and } x, y \in \text{dom}(f). \tag{2.7}$$

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!

2.2.1 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 $\frac{1}{2}x^T Q x + b^T x + c$ 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 $b^T A^T A b \geq 0$ 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 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.

2.2.2 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 $\text{epi}(f) = \{(x, t) \in dom(f) \times \mathbb{R} : f(x) \leq t\}$ 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 $\iff$ $dom(f)$ 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 $\max_{i=1,\ldots,k}(a_i^T x + b_i)$.

2.2.3 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.

2.2.4 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.