Optimization Models

EECS 127 / EECS 227AT

Laurent El Ghaoui

EECS department
UC Berkeley

Fall 2018

LECTURE 16

Convexity

The Future is convex.

P. Jackel, A. Kawai

Outline

Introduction

- The key feature that renders an optimization problem "nice" is a property called *convexity*, which is introduced in this lecture.
- In this lecture, we characterize convex *sets* and convex *functions*; in the next one, we define the class of convex optimization *problems* as those where a convex objective function is minimized over a convex set.
- Problems that can be modeled in this convexity framework are typically amenable to an efficient numerical solution.
- Further, for certain types of convex models having particular structure, such as linear, convex quadratic, or convex conic, specialized algorithms are available that are so efficient as to provide the user with a reliable "technology" for modeling and solving practical problems.

Convex Sets

Combinations and hulls

- Given a set of points (vectors) in $\mathbb{R}^n$:

$$\mathcal{P} = \{x^{(1)}, \ldots, x^{(m)}\},$$

the *linear hull* (subspace) generated by these points is the set of all possible *linear combinations* of the points:

$$x = \lambda_1 x^{(1)} + \cdots + \lambda_m x^{(m)}, \quad \text{for } \lambda_i \in \mathbb{R},\ i = 1, \ldots, m.$$

- The *affine hull*, $\operatorname{aff}\mathcal{P}$, of $\mathcal{P}$ is the set generated by taking all possible linear combinations of the points in $\mathcal{P}$, under the restriction that the coefficients λ_i sum up to one, that is $\sum_{i=1}^m \lambda_i = 1$. $\operatorname{aff}\mathcal{P}$ is the smallest affine set containing $\mathcal{P}$.
- A *convex combination* of the points is a special type of linear combination, in which the coefficients λ_i are restricted to be nonnegative and to sum up to one, that is

$$\lambda_i \geq 0 \text{ for all } i, \text{ and } \sum_{i=1}^m \lambda_i = 1.$$

Convex Sets

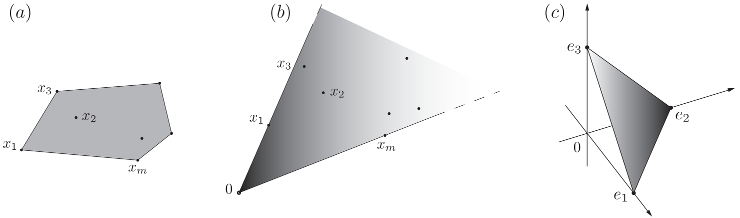
Combinations and hulls

- Intuitively, a convex combination is a weighted average of the points, with weights given by the λ_i coefficients. The set of all possible convex combination is called the *convex hull* of the point set:

$$\mathrm{co}(x^{(1)}, \dots, x^{(m)}) = \left\{ x = \sum_{i=1}^{m} \lambda_i x^{(i)} : \ \lambda_i \geq 0, i = 1, \dots, m; \sum_{i=1}^{m} \lambda_i = 1 \right\}.$$

- Similarly, the *conic hull* of a set of points is defined as

$$\mathrm{conic}(x^{(1)}, \dots, x^{(m)}) = \left\{ x = \sum_{i=1}^{m} \lambda_i x^{(i)} : \ \lambda_i \geq 0, i = 1, \dots, m \right\}.$$



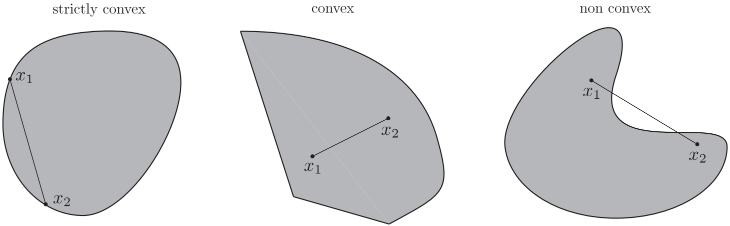
Convex Sets

Convexity

- A subset $C \subseteq \mathbb{R}^n$ is said to be *convex* if it contains the line segment between any two points in it:

$$x_1, x_2 \in C,\ \lambda \in [0,\ 1] \quad \Rightarrow \quad \lambda x_1 + (1-\lambda)x_2 \in C.$$

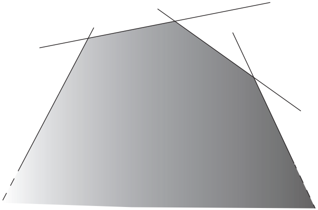
- Subspaces and affine sets, such as lines and hyperplanes are obviously convex, as they contain the entire line passing through any two points. Half-spaces are also convex.
- A set C is a *cone* if $x \in C$, then $\alpha x \in C$, for every $\alpha \geq 0$. A set C is said to be a *convex cone* if it is convex and it is a cone. The conic hull of a set is a convex cone.



Operations that preserve convexity

Intersection

- If $C_1, \ldots, C_m$ are convex sets, then their intersection $C = \bigcap_{i=1,\ldots,m} C_i$ is also a convex set.
- The intersection rule actually holds for possibly infinite families of convex sets: if $C(\alpha)$, $\alpha \in \mathcal{A} \subseteq \mathbb{R}^q$, is a family of convex sets, parameterized by α, then the set $C = \bigcap_{\alpha \in \mathcal{A}} C_\alpha$ is convex.
- Example: An halfspace $\mathcal{H} = \{x \in \mathbb{R}^n : \; c^\top x \leq d\}, \quad c \neq 0$ is a convex set. The intersection of m halfspaces $\mathcal{H}_i$, $i = 1, \ldots, m$, is a convex set called a *polyhedron*.



Examples

Second-order cone

The second-order cone in $\mathbb{R}^{n+1}$:

$$\mathcal{K}_n = \{(x,t),\, x \in \mathbb{R}^n,\, t \in \mathbb{R} :\ \|x\|_2 \le t\}.$$

is convex, since it is the intersection of half-spaces:

$$\mathcal{K}_n = \bigcap_{u \,:\, \|u\|_2 \le 1} \left\{(x,t),\, x \in \mathbb{R}^n,\, t \in \mathbb{R} :\ u^\top x \le t\right\}.$$

Here, we have used the representation of $\|\cdot\|_2$ based on the Cauchy-Schwarz inequality:

$$\|x\|_2 = \max_{u \,:\, \|u\|_2 \le 1}\ u^\top x,$$

which implies that

$$\|x\|_2 \le t \Longleftrightarrow u^\top x \le t \text{ for every } u \text{ such that } \|u\|_2 \le 1.$$

Examples

Set of positive semi-definite matrices

Recall that a symmetric matrix $X \in \mathbb{S}^n$ is positive-semidefinite if and only if

$$\forall\, u \in \mathbb{R}^n \ : \ u^\top X u \geq 0.$$

(From the spectral theorem, this is equivalent to the fact that every eigenvalue of X is non-negative.)

The set of symmetric, positive-semidefinite matrices, $\mathbb{S}^n_+$, is the intersection of (an infinite number of) half-spaces in $\mathbb{S}^n$:

$$\mathbb{S}^n_+ = \bigcap_{u \in \mathbb{R}^n} \left\{ X \in \mathbb{S}^n \ : \ u^\top X u \geq 0 \right\}.$$

Hence, $\mathbb{S}^n_+$ is convex. In fact, it is a convex cone, since multiplying a PSD matrix by a positive number results in a PSD matrix.

Operations that preserve convexity

Affine transformation

- If a map $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine, and $C \subset \mathbb{R}^n$ is convex, then the image set

$$f(C) = \{f(x) : \ x \in C\}$$

is convex.

- This fact is easily verified: any affine map has a matrix representation

$$f(x) = Ax + b.$$

Then, for any $y^{(1)}, y^{(2)} \in f(C)$ there exist $x^{(1)}, x^{(2)}$ in C such that $y^{(1)} = Ax^{(1)} + b$, $y^{(2)} = Ax^{(2)} + b$. Hence, for $\lambda \in [0, 1]$, we have that

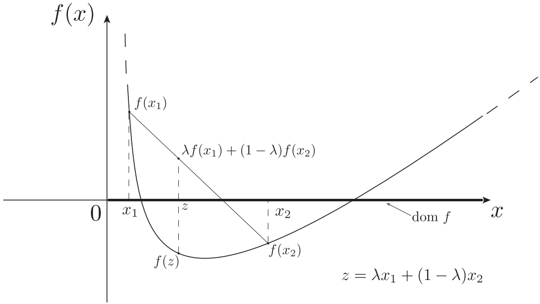
$$\lambda y^{(1)} + (1-\lambda)y^{(2)} = A(\lambda x^{(1)} + (1-\lambda)x^{(2)}) + b = f(x),$$

where $x = \lambda x^{(1)} + (1-\lambda)x^{(2)} \in C$.

- In particular, the projection of a convex set C onto a subspace is representable by means of a linear map, hence the projected set is convex.

Convex Functions

- The *domain* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set over which the function is well-defined:
$$\mathrm{dom}\, f = \{x \in \mathbb{R}^n : -\infty < f(x) < \infty\}.$$
- A function $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if $\mathrm{dom}\, f$ is a convex set, and for all $x, y \in \mathrm{dom}\, f$ and all $\lambda \in [0,1]$ it holds that
$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y). \tag{1}$$
- We say that a function f is *concave* if $-f$ is convex.



About the domain of a convex function

Convex functions must be $+\infty$ outside their domains, so that (1) remains valid even if x or $y \notin \mathrm{dom}\, f$. The function

$$f(x) = \begin{cases} -\sum_{i=1}^{n} \log x_i & \text{if } x > 0, \\ +\infty & \text{otherwise,} \end{cases}$$

is convex, but the function

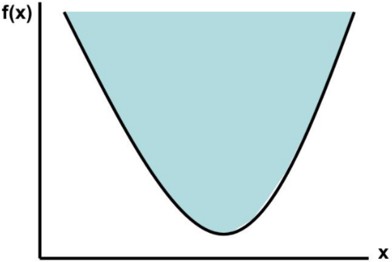
$$f(x) = \begin{cases} -\sum_{i=1}^{n} \log x_i & \text{if } x > 0, \\ -\infty & \text{otherwise,} \end{cases}$$

is not.

Epigraph

Given a function $f : \mathbb{R}^n \to (-\infty, +\infty]$, its *epigraph* (i.e., the set of points lying above the graph of the function) is the set

$$\mathrm{epi} f = \{(x, t), x \in \mathrm{dom}\, f, t \in \mathbb{R} : \; f(x) \leq t\} \,.$$



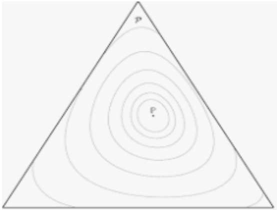
Fact: f is a convex function if and only if epi f is a convex set.

Sublevel sets

- For $\alpha \in \mathbb{R}$, the *α-sublevel set* of f is defined as

$$S_\alpha \doteq \{x \in \mathbb{R}^n : \ f(x) \leq \alpha\}.$$

  It can be easily verified that if f is a convex function, then S_α is a convex set, for any $\alpha \in \mathbb{R}$.
- The converse of the latter two statements are not true in general. For instance, $f(x) = \log(x)$ is not convex (it is actually concave), nevertheless its sublevel sets are the intervals $(0, \mathrm{e}^\alpha]$, which are convex.



The sublevel sets of a "relative entropy function", which measures a form of distance between a discrete probability distribution and a reference one.

Example

Consider the "log-sum-exp" function arising in logistic regression:

$$x \in \mathbb{R}^n \to f(x) = \log \left(\sum_{i=1}^{n} e^{x_i} \right).$$

The epigraph is the set of pairs (x, t) characterized by the inequality $t \geq f(x)$, which can be re-written as

$$\mathrm{epi}\ f = \left\{ (x, t) \in \mathbb{R}^n \times \mathbb{R} \ : \ \sum_{i=1}^{n} e^{x_i - t} \leq 1 \right\},$$

which is convex, due to the convexity of the exponential function.

Operations that preserve convexity

Nonnegative linear combinations

- If $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, m$, are convex functions, then the function

$$f(x) = \sum_{i=1}^{m} \alpha_i f_i(x), \quad \alpha_i \geq 0,\ i = 1, \ldots, m$$

  is also convex over $\cap_i \operatorname{dom} f_i$.

- This fact easily follows from the definition of convexity, since for any $x, y \in \operatorname{dom} f$ and $\lambda \in [0, 1]$,

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \sum_{i=1}^{m} \alpha_i f_i(\lambda x + (1-\lambda)y) \leq \sum_{i=1}^{m} \alpha_i \left(\lambda f_i(x) + (1-\lambda) f_i(y)\right) \\ &= \lambda f(x) + (1-\lambda) f(y). \end{aligned}$$

- Example: the negative entropy function with values for $x \in \mathbb{R}^n_{++}$

$$f(x) = \sum_{i=1}^{n} x_i \log x_i.$$

Operations that preserve convexity

Affine variable transformation

- Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex, and define

$$g(x) = f(Ax + b), \quad A \in \mathbb{R}^{n,m},\ b \in \mathbb{R}^n.$$

- Then, g is convex over $\mathrm{dom}\, g = \{x : Ax + b \in \mathrm{dom}\, f\}$.

Examples:

- $f(z) = -\log(z)$, is convex over $\mathrm{dom}\, f = \mathbb{R}_{++}$, hence $f(x) = -\log(ax + b)$ is also convex over $ax + b > 0$.
- For any convex function $\mathcal{L} : \mathbb{R} \to \mathbb{R}$, the function

$$(w, b) \in \mathbb{R}^n \times \mathbb{R} \to \sum_{i=1}^{m} \mathcal{L}(w^\top x_i + b),$$

  where $x_1, \ldots, x_m \in \mathbb{R}^n$ are given data points, is convex. (Such functions arise as "loss" functions in machine learning.)

First-order conditions

- If f is differentiable (that is, $\mathrm{dom}\, f$ is open and the gradient exists everywhere on the domain), then f is convex if and only if

$$\forall x, y \in \mathrm{dom}\, f,\ f(y) \geq f(x) + \nabla f(x)^\top (y - x),$$

- **Proof.** Assume that f is convex. Then, the definition implies that for any $\lambda \in (0, 1]$

$$\frac{f(x + \lambda(y - x)) - f(x)}{\lambda} \leq f(y) - f(x),$$

  which, for $\lambda \to 0$ yields $\nabla f(x)^\top (y - x) \leq f(y) - f(x)$.

- Conversely, take any $x, y \in \mathrm{dom}\, f$ and $\lambda \in [0, 1]$, and let $z = \lambda x + (1 - \lambda y)$:

$$f(x) \geq f(z) + \nabla f(z)^\top (x - z), \quad f(y) \geq f(z) + \nabla f(z)^\top (y - z).$$

  Taking a convex combination of these inequalities, we get

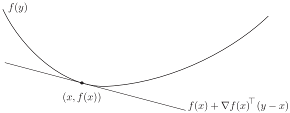
$$\lambda f(x) + (1 - \lambda) f(y) \geq f(z) + \nabla f(z)^\top 0 = f(z),$$

  which concludes the proof.

First-order conditions

Geometric interpretation

$$\forall x, y \in \mathrm{dom}\, f, \;\; f(y) \geq f(x) + \nabla f(x)^\top (y - x),$$



The graph of f is bounded below everywhere by anyone of its tangent hyperplanes.

- The gradient of a convex function at a point $x \in \mathbb{R}^n$ (if it is nonzero) divides the whole space in two halfspaces:

$$\begin{aligned} \mathcal{H}_{++}(x) &= \{y : \nabla f(x)^\top (y - x) > 0\}, \\ \mathcal{H}_{-}(x) &= \{y : \nabla f(x)^\top (y - x) \leq 0\}, \end{aligned}$$

  and any point $y \in \mathcal{H}_{++}(x)$ is such that $f(y) > f(x)$.

- This is a key fact exploited by the so-called "gradient" algorithms for minimizing a convex function.

Second-order conditions

If f is twice differentiable, then f is convex if and only if its Hessian matrix $\nabla^2 f$ is positive semi-definite everywhere on the (open) domain of f, that is if and only if $\nabla^2 f \succeq 0$ for all $x \in \mathrm{dom}\, f$.

Example: a generic quadratic function

$$f(x) = \frac{1}{2} x^\top H x + c^\top x + d$$

has Hessian $\nabla^2 f(x) = H$. Hence f is convex if and only if H is positive semidefinite.

Restriction to a line

- A function f is convex if and only if its restriction to *any* line is convex.
- By restriction to a line we mean the function

$$g(t) = f(x_0 + tv)$$

  of scalar variable t, for fixed $x_0 \in \mathbb{R}^n$ and $v \in \mathbb{R}^n$.
- This rule gives a very powerful criterion for proving convexity of certain functions.
- Example: for the log-determinant function $f(X) = -\log\det X$ over $X \succ 0$, it holds that

$$\begin{aligned} g(t) &= -\log\det(X_0 + tV) = -\log\det X_0 \prod_{i=1,\dots,n} (1 + t\lambda_i(Z)), \\ &= -\log\det X_0 + \sum_{i=1}^{n} -\log(1 + t\lambda_i(Z)); \quad Z \doteq X_0^{-1/2} V X_0^{-1/2}. \end{aligned}$$

- The first term in the previous expression is a constant, and the second term is the sum of convex functions, hence $g(t)$ is convex for any $X_0 \in \mathbb{S}^n_{++}$, $V \in \mathbb{S}^n$, thus $-\log\det X$ is convex over the domain $\mathbb{S}^n_{++}$.

Pointwise maximum

If $(f_\alpha)_{\alpha \in \mathcal{A}}$ is a family of convex functions indexed by parameter α, and $\mathcal{A}$ is a set, then the pointwise max function

$$f(x) = \max_{\alpha \in \mathcal{A}} \; f_\alpha(x)$$

is convex over the domain $\{\cap_{\alpha \in \mathcal{A}} \operatorname{dom} f_\alpha\} \cap \{x : \; f(x) < \infty\}$.

Proof: The epigraph of f is the set of pairs (x, t) such that

$$\forall \, \alpha \in \mathcal{A} \; : \; f_\alpha(x) \leq t.$$

hence, the epigraph of f is the intersection of the epigraphs of all the functions involved, therefore f is convex.

Pointwise maximum rule

Example: functions arising in SOCP

The function $f : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, with values

$$f(y,t) = \|y\|_2 - t$$

is convex, since it is the pointwise maximum of linear functions of (y,t):

$$f(y,t) = \max_{u \,:\, \|u\|_2 \le 1} u^\top y - t.$$

Using the rule of affine variable transformation, we obtain that for any matrices A, C, vector b and scalar d, the function

$$x \to \|Ax+b\|_2 - (c^\top x + d)$$

is also convex.

Pointwise maximum rule

Example: sum of k largest elements

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ with values

$$f(x) = \sum_{i=1}^{k} x_{[i]}$$

where $x_{[i]}$ denotes the i-th largest element in x.

We have:

$$f(x) = \max_u \; u^\top x \;\; : \;\; u \in \{0,1\}^n, \;\; \mathbf{1}^\top u = k.$$

For every u, $x \to u^\top x$ is linear, hence f is convex.

Pointwise maximum rule

Example: largest eigenvalue of a symmetric matrix

Consider the function $f : \mathbb{S}^n \to \mathbb{R}$, with values for a given $X = X^\top \in \mathbb{S}^n$ given by

$$f(X) = \lambda_{\max}(X),$$

where $\lambda_{\max}$ denotes the largest eigenvalue.

The function is the pointwise maximum of *linear* functions of X:

$$F(x) = \max_{u \,:\, \|u\|_2 = 1} u^\top X u.$$

Hence, f is convex.

Pointwise maximum rule

Example: norm plus linear

Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ with values

$$f(x) = \|Ax + b\|_2 + c^T x,$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$ are given.

From the Cauchy-Schwartz inequality:

$$f(x) = \max_{u \; \|u\|_2 \le 1} F(x, u), \text{ where } F(x, u) \doteq u^\top (Ax + b) + c^T x.$$

For every u, $F(\cdot, u)$ is convex.

Pointwise maximum rule

Extending the earlier example

Consider the function $f : \mathbb{R}^{n\times p} \to \mathbb{R}$, with values

$$f(X) = \|AX + B\| + \mathrm{trace}\, C^T X$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^{m\times p}$, $c \in \mathbb{R}^{n\times p}$ are given. Here, $\|M\|$ is the largest singular value norm of its matrix argument M, also characterized as

$$\|M\| = \max_{u \,:\, u^\top u = 1} \|Mu\|_2.$$

From the definition of the matrix norm above:

$$f(X) = \max_{u \ \|u\|_2 \leq 1} F(X, u) \doteq \|(AX + B)u\|_2 + \mathrm{trace}\, C^T X.$$

For every u, $F(\cdot, u)$ is convex.