Modern Optimization

Jean-Luc Bouchot

School of Mathematics and Statistics
Beijing Institute of Technology
jlbouchot@bit.edu.cn

Spring 2021

Outline

Proposition 2.1 (Cauchy Schwarz)

Remember the Cauchy-Schwarz inequality: let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$*, we have*

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \mathbf{v}\|.$$

This allows, if needed, to define an angle between the two vectors (assuming non zero vectors):

$$\cos(\alpha) = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\|\mathbf{v}\|}.$$

Definition 2.1 (Convex sets)

Let $C \subseteq \mathbb{R}^d$*. We say that* C *is* **convex** *if for any two points* $x, y \in C$*, for any* $\lambda \in [0, 1]$*,*

$$\lambda x + (1 - \lambda) y \in C.$$

Let's doodle some examples.

Remark 2.1

Let $\mathcal{I}$ be a countable index set and $\{C_i\}_{i\in\mathcal{I}}$ a family of convex sets. Then

$$\cap_{i\in\mathcal{I}} C_i$$

is convex.

Definition 2.2 (Graph-Epigraph)

Let $f : \mathrm{dom}(f) \to \mathbb{R}$.

- *The* **graph** *of* f *is the set of points* $\{(x, f(x)), x \in \mathrm{dom}(f)\} \subset \mathbb{R}^{d+1}$.
- *The* **epigraph** *of* f *is the set of points above the graph:*

$$\mathrm{epi}(f) = \{(x, y), x \in \mathrm{dom}(f), y \geq f(x)\}.$$

Let's see what this looks like...

Definition 2.3 (Convex function)

Let $f : \mathrm{dom}(f) \to \mathbb{R}$*. We say that* f *is* **convex** *if:*

1. $\mathrm{dom}(f)$ *is convex, and*
2. *for all* $x, y \in \mathrm{dom}(f)$*, for all* $\lambda \in [0, 1]$*, we have*

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

Example 2.1

Let $f(x) = a^T x + c$ for a given vector $a \in \mathbb{R}^d$ and $c \in \mathbb{R}$. f is convex.

Remark 2.2

The previous example is very specific in that it is convex with, in fact, equality instead of inequality.
If a function has a strict inequality, we will talk about strict convexity (albeit limiting to $\lambda \in (0,1)$). We will come back to this in the near future.

Example 2.2

Let $Q \in \mathbb{R}^{d\times d}$ be a positive definite matrix and define $f(x) = x^T Q x$. f is a convex function.

Proposition 2.2

Let $f : \mathrm{dom}(f) \to \mathbb{R}$*.* f *is convex* $\Leftrightarrow \mathrm{epi}(f)$ *is convex.*

Proposition 2.3 (Jensen's inequality)

Let $f : \mathrm{dom}(f) \to \mathbb{R}$ *be a convex function. Let* $x_1, \cdots, x_n$ *be* n *points in* $\mathrm{dom}(f) \subset \mathbb{R}^d$ *and let* $\lambda_1, \cdots, \lambda_n$ *be* n *nonnegative numbers such that* $\sum \lambda_i = 1$*. Then*

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i).$$

Proposition 2.4

Let f be a convex function on the open set $\mathrm{dom}(f)$. Then f is continuous.

Definition 2.4 (Lipschitz continuity)

A function $f : \mathrm{dom}(f) \to \mathbb{R}$ *is said to be* **Lipschitz continuous** *with Lipschitz constant* L *(sometimes expressed as* L*-Lipschitz or, if context is clear, simply* f *is Lipschitz) if*

$$\|f(x) - f(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathrm{dom}(f).$$

Theorem 2.1

Let $f : \mathrm{dom}(f) \to \mathbb{R}^m$ be a differentiable function and let $X \subset \mathrm{dom}(f)$ be an open convex set. Let $L \in \mathbb{R}^+$. The following statements are equivalent:

1. *f is L-Lipschitz.*
2. *The differentials of f are bounded by L, i.e.*

$$\|Df(x)\| \leq L, \quad \forall x \in X,$$

where D denotes the differential operator (or Jacobian), defined as the unique operator A such that for all y in a neighbourhood of x, we have

$$f(y) = f(x) + A(y-x) + r(y-x),$$

with

$$\lim_{v \to 0} \frac{\|r(v)\|}{\|v\|} = 0.$$

Example 2.3 (in-class)

Consider all the assumptions from the previous theorem except that X is just convex. What can be said about the conclusion?

Theorem 2.2

Let $f : \mathrm{dom}(f) \to \mathbb{R}$ *with* $\mathrm{dom}(f)$ *open. Assume furthermore that* f *is differentiable on* $\mathrm{dom}(f)$*. Then* f *is convex if and only if*

- $\mathrm{dom}(f)$ *is convex and*
- *the inequality* $f(y) \geq f(x) + \nabla f(x)^T (y - x)$ *holds for all* $x, y \in \mathrm{dom}(f)$*.*

Example 2.4

Use the first order condition to show the convexity of $f : \mathbb{R}^2 \to \mathbb{R}$ defined as $f(x_1, x_2) = x_1^2 + x_2^2$.

Theorem 2.3

Let f be a twice continuously differentiable function. Then f is convex if and only if $\mathrm{dom}(f)$ is convex and for all $x \in \mathrm{dom}(f) \subseteq \mathbb{R}^n$ the Hessian $H_f(x) = \nabla^2 f(x)$ is positive semidefinite.

Lemma 1

A function $f : \mathrm{dom}(f) \subseteq \mathbb{R}^d \to \mathbb{R}$ *is convex if and only* $g_{x,y}(t) := f(x+ty)$ *is (univariate) convex as a function of* f*.*
Note that the domain of g *is dependent on the variables* $x, y \in \mathrm{dom}(f)$*.*

Example 2.5

The negative entropy function f defined as

$$f : \begin{cases} \mathbb{R}_{>0} & \rightarrow & \mathbb{R} \\ x & \mapsto & x\log(x) \end{cases}$$

is a convex function.

Example 2.6

Let f be the function defined as

$$f : \left\{ \begin{array}{ccc} \mathbb{R} \times \mathbb{R}_{>0} & \to & \mathbb{R} \\ (x, y) & \mapsto & \frac{x^2}{y}. \end{array} \right.$$

f is convex.

Definition 2.5 (Convex hull)

For a set C, its convex hull is defined as the set of all convex combinations of points in C, i.e.

$$\mathrm{conv}(C) := \left\{ y = \sum_{i=1}^{k} \alpha_i x_i \text{ for some } k \in \mathbb{N}, \{x_i\}_i \in C^k, \alpha_i \geq 0, \sum \alpha_i = 1 \right\}.$$

Definition 2.6 (Cones)

A set C is called a **cone** *if it is nonnegative homogeneous, i.e.*

$$\forall \alpha \geq 0, x \in C, \alpha x \in C.$$

Definition 2.7 (Convex cones)

A set C is called a **convex cone** *if it is a cone and convex, i.e.*

$$\forall \alpha \geq 0, \beta \geq 0, x, y \in C, \alpha x + \beta y \in C.$$

Example 2.7

Let $\Sigma_s^n := \{x \in \mathbb{R}^n : \|x\|_0 := \#\{i : x_i \neq 0\} \leq s\}$ be the set of s sparse vectors in $\mathbb{R}^n$ with $s \leq n$. Σ_s^n is a cone but not a convex cone.

Exercise 2.1 (in-class)

We work in unit spheres and balls:

- Show that $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ is not convex.
- Let $x \neq y$ be two vectors such that $\|x\|_2 = \|y\|_2 = 1$. Show that $z_\lambda := \lambda x + (1-\lambda)y$ is such that $\|z_\lambda\|_1 = 1$ if and only if $\lambda = 1$ or 0.
- Similarly, consider $\lambda = 1/2$. Show that $\|z_{1/2}\|_2 = 1$ if and only if $x = y$ (which we prevented)
- Show that this is no longer true if we use the ℓ_1 norm.
- What happens when using the ℓ_∞ norm?

Proposition 2.5

Let f be a convex function in two variables x and y and let C denote a non empty set. Then,

$$g(x) := \inf_{y \in C} f(x,y)$$

is convex provided $g(x) > \infty$.
Moreover, its domain is defined as the projection along a coordinate of C:

$$\mathrm{dom}(g) = \{x : \exists y \in C, (x,y) \in \mathrm{dom}(f)\}.$$

Definition 2.8 (Conjugate function)

Let $f : \mathbb{R}^d \to \mathbb{R}$. Its **convex conjugate** *(sometimes called Fenchel conjugate) is defined as the function $f^* : \mathbb{R}^d \to \mathbb{R}$ such that*

$$f^*(y) = \sup_{x \in \mathrm{dom}(f)} \left(y^T x - f(x) \right)$$

Example 2.8

Find the analytical expressions of the convex conjugate of the following functions

1. Linear/affine functions: $f(x) = a^T x + b$
2. Let Q be a positive definite matrix. Define $f(x) = x^T Q x$.
3. $f = \chi_S$ where $S \subset \mathbb{R}^d$ is a subset and χ_S denotes the indicator function:

$$\chi_S(x) = 0, \quad \text{if } x \in S, \qquad \infty, \text{ elsewhere.}$$

Exercise 2.2 (in-class)

Find the Fenchel conjugate of the following functions:

- Maximum function: $f(x) = \max_i(x_i)$.
- Piecewise linear function: assume given $2m$ numbers $a_1 \leq \cdots \leq a_m$ and $b_1, \cdots, b_m$, and define $f(x) = \max_i(a_i x + b_i)$.

Proposition 2.6

The Fenchel/convex conjugate enjoys the following properties:

- f^* *is convex, independently of the convexity (or lack thereof) of* f.
- *Fenchel's inequality holds:*

$$f(x) + f^*(y) \geq x^T y.$$

- *if* f *is convex and its epigraph is closed,* $f^{**} = f$.

Proposition 2.7 (Legendre transform)

If f is convex, differentiable, and $\mathrm{dom}(f) = \mathbb{R}^d$, its Fenchel conjugate is called **Legendre transform** *and*

$$\forall z \in \mathbb{R}^d, f^*(y) = z^T \nabla f(z) - f(z),$$

for $y = \nabla f(z)$.

Definition 2.9 (Convex optimization)

The minimization (optimization) problem

$$\begin{aligned} &\min f_0(x) \\ &\textit{s.t. } f_i(x) \leq 0, i \in \mathcal{I}, \\ &\textit{and } f_i(x) = 0, i \in \mathcal{E}, \end{aligned}$$

is said to be a **convex optimization problem** *if the objective function f_0 is convex and the feasible set is convex.*

Definition 2.10 (Standard form)

A convex optimization problem is said to be in standard form if

- f_i *for* $i \in \mathcal{I}$ *is a convex function and*
- f_i *for* $i \in \mathcal{E}$ *is an affine function.*

Example 2.9

Is the following problem in $\mathbb{R}^2$ convex? If yes, write it in standard form.

$$\begin{aligned} &\min f_0(x) = \|x\|^2 \\ &\text{s.t. } f_1(x) = \frac{x_1}{1+x_2^2} \leq 0, \\ &\text{and } f_2(x) = (x_1+x_2)^2 = 0. \end{aligned}$$

Definition 2.11

Given a function f, a point x^* is called a **local minimizer** *of f if there exists a $R > 0$ such that*

$$f(x^*) \leq f(y)$$

for all feasible y with $\|x^* - y\| \leq R$.

Remark 2.3

It is easy to see that a local optimum point x^* solves the following optimization problem

$$\begin{aligned} &\min f_0(z) \\ \text{s.t. } &f_i(z) \leq 0, i \in \mathcal{I}, \\ &f_i(z) = 0, i \in \mathcal{E}, \\ &\|z - x^*\| \leq R, \end{aligned}$$

for some $R > 0$.

Definition 2.12

A feasible point x^* is said to be a global optimum to a convex (un/constrained) optimization problem if

$$f(x^*) \leq f(y)$$

for all feasible point $y \in C$. (C being the feasible set)

Lemma 2

Let $(\mathcal{P})$ be a convex minimization problem and let x^* be a local optimum solution. Then x^* is globally optimal.

Exercise 2.3

True or false:

1. A convex optimization problem always has a minimizer.
2. If a convex objective function is bounded below, the associated unconstrained minimization problem admits at least one solution.
3. The optimal set (set of globally optimal points) of a convex optimization problem is convex.
4. The optimal set S of an unconstrained convex problem is limited to (at most) a singleton (i.e. either $S = \varnothing$ or $S = \{x^*\}$ for a unique optimal point x^*.)

Theorem 2.4 (First order optimality)

Let $(\mathcal{P})$ be a convex constrained optimization problem and let define C define the feasible set. Assume moreover that f_0 is differentiable. Then the point $x^* \in C$ is optimal if and only if

$$\nabla f_0(x^*)^T(y - x) \geq 0, \qquad \text{for all } y \in C.$$

Example 2.10

Consider the following optimization problem with box constraints:

$$\begin{aligned}
&\min f_0(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T P x + q^T x + r \\
&\text{s.t. } -1 \le x, y, z \le 1,
\end{aligned}$$

with

$$P = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix}, \quad q = \begin{bmatrix} -22 \\ -14.5 \\ 13 \end{bmatrix}, \quad r = 1.$$

Show that $\mathbf{x}^* = [1, 1/2, -1]^T$ solves this problem.

Remark 2.4

Consequences of the previous theorem:

1. if $(\mathcal{P})$ is unconstrained, the first order optimality condition becomes
$$\nabla f(x^*) = 0.$$
2. if the constraints are only linear equality constraints ($Ax = b$, $A \in \mathbb{R}^{p \times d}$), then the first order optimality condition reads
$$\nabla f_0(x^*) + A^T v = 0, \qquad \text{for all } v \in \mathbb{R}^p.$$