Lecture 11

Convex optimization

Kin Cheong Sou
Department of Mathematical Sciences
Chalmers University of Technology and Göteborg University
December 4, 2014

CHALMERS | GÖTEBORGS UNIVERSITET

Announcement

If you have the receipt of the textbook, you can replace your current copy with a corrected copy at Cremona before Dec 12!

A set $S \subseteq \mathbb{R}^n$ is a **convex set** if

$$\left.\begin{aligned} x^1, x^2 &\in S, \\ \lambda &\in (0,1) \end{aligned}\right\} \Longrightarrow \lambda x^1 + (1-\lambda)x^2 \in S$$

A function $f : \mathbb{R}^n \to \mathbb{R}$ is a **convex function** on the convex set S if

$$\left.\begin{aligned} x^1, x^2 &\in S, \\ \lambda &\in (0,1) \end{aligned}\right\} \Longrightarrow f\left(\lambda x^1 + (1-\lambda)x^2\right) \leq \lambda f(x^1) + (1-\lambda) f(x^2).$$

Convex optimization

A **convex optimization problem** is

$$\begin{aligned} f^* = \text{infimum} \quad & f(x), \\ \text{subject to} \quad & x \in S, \end{aligned}$$

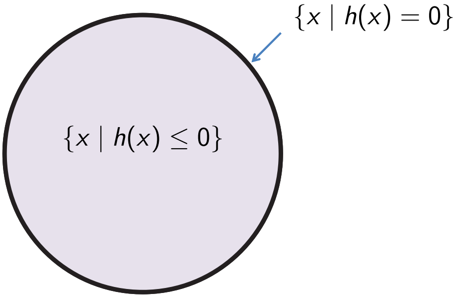
$f : \mathbb{R}^n \to \mathbb{R}$ is a **convex function on S** and $S \subseteq \mathbb{R}^n$ is a **convex set**.

A typical problem:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_j(x) = 0, \quad j = 1, \ldots, k \end{aligned}$$

- f is a **convex** function,
- g_i are **convex** functions, $i = 1, \ldots, m$,
- h_j are **affine** functions, $j = 1, \ldots, k$. Why not just convex h_j's?

- For convex $h(x)$, the set $\{x \mid h(x) = 0\}$ need not be convex!
- Consider $h(x) : \mathbb{R}^2 \to \mathbb{R}$, $h(x) = \|x\|_2^2 - 1$.
- The set $\{x \mid h(x) = 0\} = \{x \in \mathbb{R}^2 \mid \|x\|_2 = 1\}$ is the edge of a circle not including inside. Clearly not convex!



- The only convex equality constraints are linear constraints.

Local minimum = global minimum

Consider convex optimization problem

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x), \\ \text{subject to} \quad & x \in S, \end{aligned} \tag{CP}$$

> If x^* is a local minimum of convex optimization problem (CP), then x^* is also a global minimum of (CP).

Proof: Assume x^* is local but not global minimum.

- If x^* is not global minimum, then there exists $y \in S : f(y) < f(x^*)$.
- For any $0 < \theta < 1$, define $z(\theta) = \theta x^* + (1 - \theta)y$. $z(\theta) \in S$ and $f(z(\theta)) < f(x^*)$ by convexity of S and f.
- For $\theta \ll 1$, $f(z(\theta)) \geq f(x^*)$, as x^* is local minimum. Contradiction!

Algorithmic implication

- Most algorithms for constrained optimization problems find only KKT points.
- Examples include gradient projection method, penalty method, interior point method, etc (see lecture 12 and lecture 13).
- Without additional assumptions KKT points need not be global minima.
- With convexity and Slater's constraint qualification, KKT points are global minima (see Theorem 5.49, Corollary 5.51 in text).

For convex problem with convex objective f and convex feasible set S:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x), \\ \text{subject to} \quad & x \in S, \end{aligned}$$

- Most algorithms assume some smoothness of f. For example,

  $$\text{gradient descent method:} \quad x^{k+1} \leftarrow x^k - \alpha_k \nabla f(x^k)$$

  requires that f is differentiable.

- For convex problem, we can relax the differentiability assumption because of the **subgradient method**, to be detailed:

  $$x^{k+1} \leftarrow x^k - \alpha_k p^k,$$

  where p^k is a **subgradient** of f at x^k. But what is a subgradient?

Subgradient

Definition
Let $S \subseteq \mathbb{R}^n$ be a nonempty convex set and let $f : S \to \mathbb{R}$ be a convex function. Then $p \in \mathbb{R}^n$ is called a **subgradient** of f at $\bar{x} \in S$ if

$$f(x) \geq f(\bar{x}) + p^T(x - \bar{x}), \quad \text{for any } x \in S.$$

- We define the set of all subgradients to f at $\bar{x}$ as the **subdifferential** of f at $\bar{x}$ as

$$\partial f(\bar{x}) = \left\{ p \in \mathbb{R}^n \mid f(x) \geq f(\bar{x}) + p^T(x - \bar{x}), \text{ for all } x \in S. \right\}$$

Subgradient, differentiable functions

Lemma

Let S be a nonempty convex set and $f : S \to \mathbb{R}$ a convex function. Suppose that at $\bar{x} \in \operatorname{int} S$, function f is differentiable, meaning that $\nabla f(\bar{x})$ exists. Then

$$\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$$

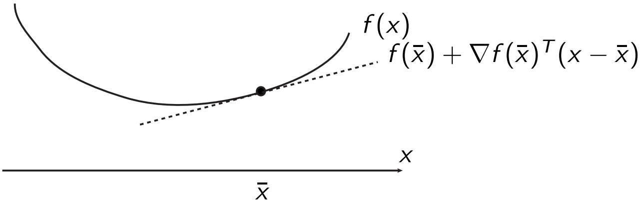


Figure : When f is differentiable, $\partial f(\bar{x}) = \{\nabla f(\bar{x})\}$

Subgradient, non-differentiable functions

When f is not differentiable at $\bar{x}$, $\partial f(\bar{x})$ may not be a singleton.

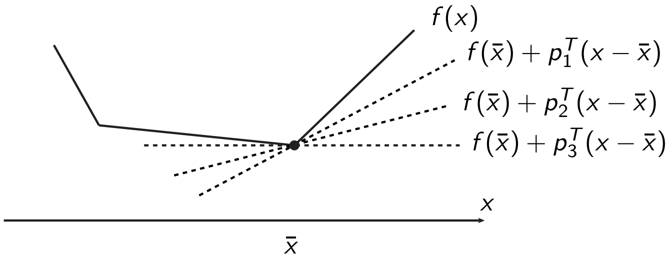


Figure : Example of three subgradients, p_1, p_2, p_3 of f at the point $\bar{x}$.

Existence of subgradient

- Now we know what a subgradient is, but does it exist after all?
- Except possibly at the boundary of $\mathrm{dom}(f)$ subgradient always exists

Theorem
Let $S \subseteq \mathbb{R}^n$ be a convex set and $f : S \to \mathbb{R}$ be a convex function. For each $\bar{x} \in \operatorname{int} S$, there always exists a vector $p \in \mathbb{R}^n$ such that

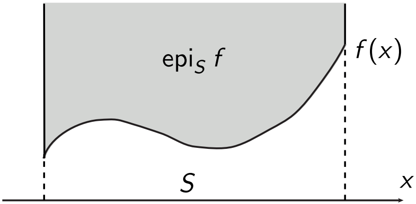
$$f(x) \geq f(\bar{x}) + p^T(x - \bar{x}), \quad \text{for any } x \in S.$$

- The statement holds for all $\bar{x} \in \operatorname{int} S$, but at the boundary of S something strange might happen... when f is not continuous.
- Why is the theorem true? We show it via a geometric approach. We need two concepts: epigraph and supporting hyperplane theorem.

Epigraph

Let $S \subseteq \mathbb{R}^n$ and $f : S \to \mathbb{R}$. The **epigraph** of f with respect to S is

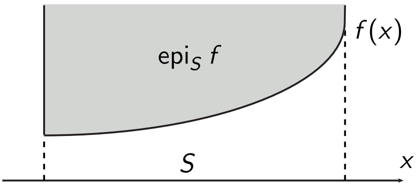
$$\text{epi}_S\, f := \{(x, \alpha) \in S \times \mathbb{R} \mid f(x) \leq \alpha\}, \quad \text{epi}_S\, f \subseteq \mathbb{R}^{n+1}$$



The **graph** of function f (all points $(x, f(x))$) is in the boundary of $\text{epi}_S\, f$.

Epigraph, convex case

Theorem
Let $S \subseteq \mathbb{R}^n$ be a nonempty and convex set, and let $f : S \to \mathbb{R}$. Then **f is convex if and only if $\mathrm{epi}_S\, f$ is a convex set.**



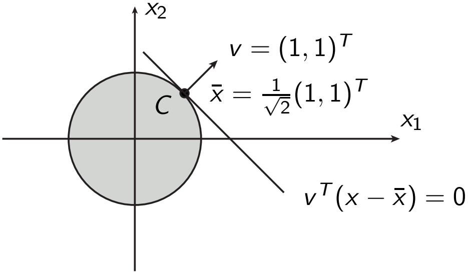
Proof: We show it on blackboard.

Supporting hyperplane theorem

Theorem

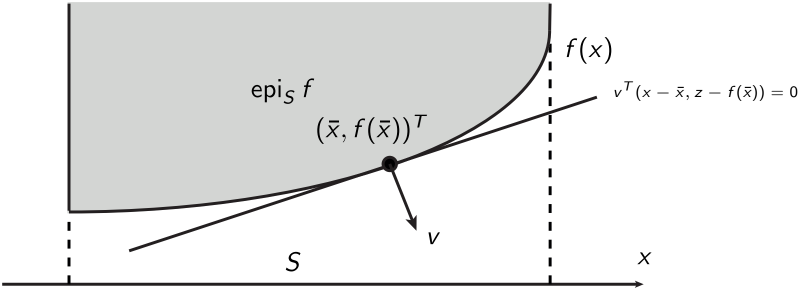
Let $C \subseteq \mathbb{R}^n$ be a nonempty and convex set. Let $\bar{x}$ be a point on the boundary of C. Then there exists a **supporting hyperplane** to C at $\bar{x}$, meaning that there exists $v \neq \mathbf{0}^n$ such that

$$v^T(x - \bar{x}) \leq 0, \quad \text{for all } x \in C.$$



Subgradient by supporting hyperplane, I

- For $\bar{x} \in S$, $(\bar{x}, f(\bar{x}))$ is a point at the boundary of $\text{epi}_S\, f$ (convex).
- Thus, there exists $v : v^T \big(x - \bar{x}, z - f(\bar{x})\big) \leq 0, \quad \forall (x, z) \in \text{epi}_S\, f$.



- Only when the hyperplane is "non-vertical" does v define a subgradient at $\bar{x}$!

Subgradient by supporting hyperplane, II

- At $(\bar{x}, f(\bar{x}))$ apply supporting hyperplane theorem for $\text{epi}_S\, f$ yields

$$v^T\big(x - \bar{x}, z - f(\bar{x})\big) \leq 0, \quad \forall (x, z) \in \text{epi}_S\, f$$

- Write $v = (u, t) \in \mathbb{R}^n \times \mathbb{R}$. For all $(x, z) \in \text{epi}_S\, f$,

  $u^T(x - \bar{x}) + t(z - f(\bar{x})) \leq 0 \implies t \leq 0$ (otherwise LHS $\to \infty$ as $z \to \infty$).

- If $t < 0$ (i.e., hyperplane is non-vertical), replace z with $f(x)$ yields

$$f(x) \geq f(\bar{x}) - \big(\frac{u}{t}\big)^T\big(f(x) - f(\bar{x})\big) \implies -\frac{u}{t} \in \partial f(\bar{x}).$$

- When will $t < 0$? If $t = 0$, then $u^T(x - \bar{x}) \leq 0$ for all $x \in S$. This is impossible when $\bar{x} \in \text{int}\, S$.

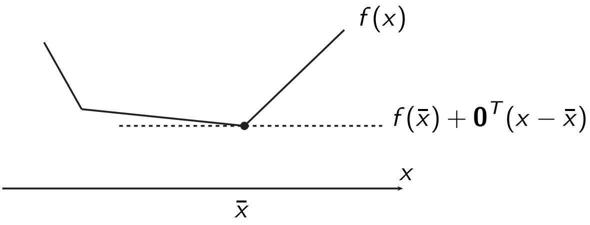
Subgradient, summary

- Now we know subgradient exists for convex f over int S. But...
- How do we find a subgradient when we know at least one exists?
  - We will see one (later in this lecture) when dealing with Lagrangian dual function.
- What is the use of a subgradient? We use it to define...
  - Optimality condition
  - Subgradient method

  for convex optimization problems with non-differentiable objective functions.

Proposition (optimality of a convex function over $\mathbb{R}^n$)
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. The following are equivalent:

1. f is globally minimized at $x^* \in \mathbb{R}^n$;
2. $\mathbf{0}^n \in \partial f(x^*)$;

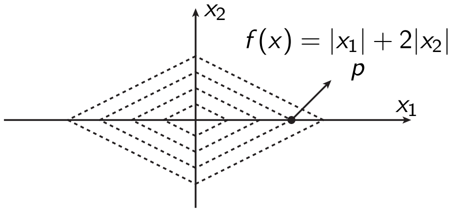


- N.B. A more completed theorem available in text (Proposition 6.19)!

Subgradient methods

We want to define an optimization method using subgradients.

- Analogous to gradient descent method, we move iterate in the negative subgradient direction $-p$ (i.e., $x^{k+1} \leftarrow x^k - \alpha_k p^k$)
- But note: $-p$ need not be a descent direction. (See the figure)

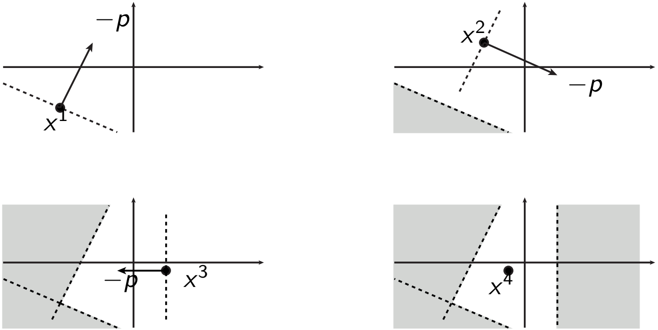


- It however can move us "towards" to the optimal solution.

Subgradient methods

Subgradient direction p defines "cutting plane" for all $x \in S : f(x) \leq f(\bar{x})$,

$$f(x) \geq f(\bar{x}) + p^T(x - \bar{x}) \implies p^T x \leq p^T \bar{x} \quad \text{(i.e., the halfspace containing } x^*\text{)}.$$



Gradually "carve" the optimal solution set out of the feasible set.

Subgradient method, unconstrained

So now we can define a subgradient method

Subgradient method (unconstrained)

Step 0 Initiate x^0, $f^0_{\text{best}} = f(x^0)$. $k = 0$.

Step 1 Find a subgradient p^k to f in x^k.

Step 2 Update $x^{k+1} = x^k - \alpha_k p^k$ $\qquad$ (α_k is the step length in iteration k)

Step 3 Let $f^{k+1}_{\text{best}} = \min\{f^k_{\text{best}}, f(x^{k+1})\}$

Step 4 If some termination criteria is fulfilled, stop. Otherwise, let $k := k + 1$ and go to Step 1.

Subgradient method, constrained

A simple extention if we consider minimizing f over the convex set S.

Subgradient method (nontrivial convex feasible set)

Step 0 Initiate $x^0 \in S$, $f_{\text{best}}^0 = f(x^0)$. $k = 0$.

Step 1 Find a subgradient p^k to f in x^k.

Step 2 Update $x^{k+1} = \text{Proj}_S\left(x^k - \alpha_k p^k\right)$

Step 3 Let $f_{\text{best}}^{k+1} = \min\{f_{\text{best}}^k, f(x^{k+1})\}$

Step 4 If some termination criteria is fulfilled, stop. Otherwise, let $k := k + 1$ and go to Step 1.

Examples of step size rules:

- **Constant step size**

$$\alpha_k = \alpha$$

- **Square summable but not summable**

$$\sum_{k=0}^{\infty} \alpha_k^2 < \infty, \quad \sum_{k=0}^{\infty} \alpha_k = \infty$$

For example: $\alpha_k = \frac{a}{b+ck}$

Depending on the step size rules, different convergence results can be shown.

Subgradient methods, summary

- Now we know subgradient methods can solve convex optimization problems with non-differentiable objective function. But...
- What are typical problems with non-differentiable objective?
- To use subgradient methods, we need to find subgradients. Are they easy to find?

It turns out, the Lagrangian dual problem is naturally suited for subgradient methods.

The Lagrangian dual

Consider the Lagrangian relaxation of the problem to find

$$\begin{aligned} f^* = \text{infimum} \quad & f(x), \\ \text{subject to } & g(x) \leq \mathbf{0}, \\ & x \in X. \end{aligned}$$

We first construct the Lagrangian function

$$L(x, \boldsymbol{\mu}) = f(x) + \boldsymbol{\mu}^T g(x),$$

and define the dual function as

$$q(\boldsymbol{\mu}) = \underset{x \in X}{\text{infimum}}\, L(x, \boldsymbol{\mu})$$

We then define the Lagrangian dual problem, which is to find

$$q^* = \text{supremum } q(\boldsymbol{\mu}),$$
$$\text{subject to} \quad \boldsymbol{\mu} \geq \mathbf{0}$$

- As we have shown, the dual function q is always concave, so the dual problem is a convex problem.
- q is however not in general differentiable.
- Therefore, subgradient methods are often utilized to solve the dual problem.

But how do we find subgradients to q at a point $\boldsymbol{\mu}$?

- In order to evaluate $q(\boldsymbol{\mu})$, we need to solve the problem

$$q(\boldsymbol{\mu}) = \underset{x \in X}{\text{infimum}}\, L(x, \boldsymbol{\mu}) = \underset{x \in X}{\text{infimum}}\, f(x) + \boldsymbol{\mu}^T g(x).$$

  Let the solution set to this problem be

$$X(\boldsymbol{\mu}) = \underset{x \in X}{\operatorname{argmin}}\, L(x, \boldsymbol{\mu})$$

- If we take any $x \in X(\boldsymbol{\mu})$, then $g(x)$ will be a subgradient. (We will show this soon)
- So when evaluating the dual function q at the point $\boldsymbol{\mu}$, we obtain a subgradient to q.

Subgradients for the dual

Proposition

Assume that X is nonempty and compact. Then the following hold.

a) Let $\boldsymbol{\mu} \in \mathbb{R}^m$. If $x \in X(\boldsymbol{\mu})$, then $g(x)$ is a subgradient to q at $\boldsymbol{\mu}$, that is, $g(x) \in \partial q(\boldsymbol{\mu})$.

b) Let $\boldsymbol{\mu} \in \mathbb{R}^m$. Then

$$\partial q(\boldsymbol{\mu}) = \operatorname{conv} \{ g(x) \mid x \in X(\boldsymbol{\mu}) \} .$$

Proof: See Proposition 6.20 in text.

Subgradient method for the Lagrangian dual problem

Step 0 Initialize $\boldsymbol{\mu}^0$, $q^0_{\text{best}} = q(\boldsymbol{\mu}^0)$, $k := 0$

Step 1 Solve the problem (dual function evaluation)

$$q(\boldsymbol{\mu}) = \underset{x \in X}{\text{infimum}}\, L(x, \boldsymbol{\mu}^k)$$

Let the solution to the problem be x^k.

$g(x^k)$ is then a subgradient to q at $\boldsymbol{\mu}^k$.

Step 3 Update $\boldsymbol{\mu}^{k+1} = \left[\boldsymbol{\mu}^k + \alpha_k g(x^k)\right]_+$ (nonnegative orthant projection)

Step 4 Let $q^{k+1}_{\text{best}} = \max\{q^k_{\text{best}}, q(\boldsymbol{\mu}^{k+1})\}$

Step 4 If some termination criteria is fulfilled, stop. Otherwise, let $k := k + 1$ and go to Step 1.