Lecture 11

Convex optimization

Emil Gustavsson
Fraunhofer-Chalmers Centre
Dec 1, 2017

CHALMERS | GÖTEBORGS UNIVERSITET

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

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

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

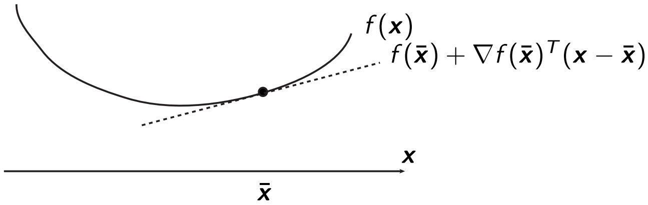
$$\left.\begin{aligned} \boldsymbol{x}^1, \boldsymbol{x}^2 &\in S, \\ \lambda &\in (0,1) \end{aligned}\right\} \Longrightarrow f\left(\lambda \boldsymbol{x}^1 + (1-\lambda)\boldsymbol{x}^2\right) \leq \lambda f(\boldsymbol{x}^1) + (1-\lambda) f(\boldsymbol{x}^2).$$

Convex function, differentiable case

- $\text{dom}(f) = S$ open convex set, f differentiable in S. Then,

$$f \text{ convex} \iff f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x}), \quad \forall\, x, \bar{x} \in S$$

- Illustration in 1-D (when $S \subseteq \mathbb{R}$), for some fixed $\bar{x} \in S$



A **convex optimization problem** is

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

$f : \mathbb{R}^n \to \mathbb{R}$ 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 a **convex** function,
- g_i **convex** functions, $i = 1, \ldots, m$,
- h_j **affine** functions, $j = 1, \ldots, k$. Why not convex h_j's (e.g., $x^2 = 1$)

Local minimum = global minimum

Consider convex optimization problem

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

x^* local minimum of CP $\implies$ x^* 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!

- Most algorithms assume smoothness of objective f. E.g.,

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

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 $\boldsymbol{p} \in \mathbb{R}^n$ is called a **subgradient** of f at $\bar{\boldsymbol{x}} \in S$ if

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

- Set of all subgradients to f at $\bar{\boldsymbol{x}}$ called **subdifferential** of f at $\bar{\boldsymbol{x}}$ as

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

Lemma

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

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

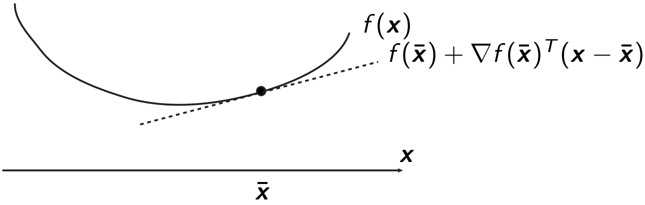


Figure: When f is differentiable, $\partial f(\bar{\mathbf{x}}) = \{\nabla f(\bar{\mathbf{x}})\}$

Subgradient, non-differentiable functions

When f is not differentiable at $\bar{\boldsymbol{x}}$, there could be multiple subgradients

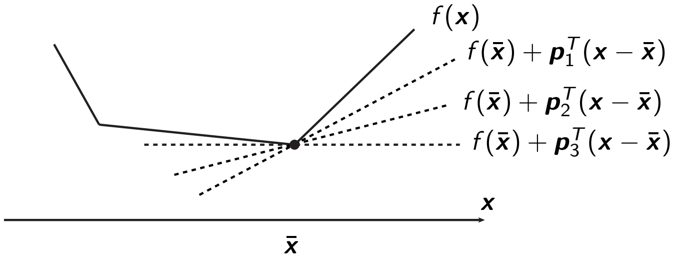


Figure: Example of three subgradients, $\boldsymbol{p}_1$, $\boldsymbol{p}_2$, $\boldsymbol{p}_3$ of f at the point $\bar{\boldsymbol{x}}$.

Existence of subgradient

- Do subgradients exist after all? Typically yes

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

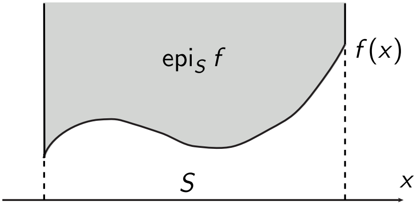
$$f(\boldsymbol{x}) \geq f(\bar{\boldsymbol{x}}) + \boldsymbol{p}^T(\boldsymbol{x} - \bar{\boldsymbol{x}}), \quad \text{for any } \boldsymbol{x} \in S.$$

- Statement holds for all $\bar{x} \in \operatorname{int} S$, but not always at the boundary
- Can show theorem geometrically using 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

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

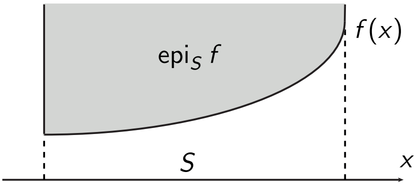


The **graph** of function f (all points $(\boldsymbol{x}, f(\boldsymbol{x}))$) is in the boundary of $\mathrm{epi}_S\, f$.

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 $\text{epi}_S\, f$ is a convex set.**



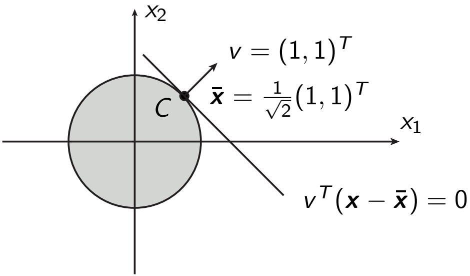
Proof: We show it on blackboard.

Supporting hyperplane theorem

Theorem

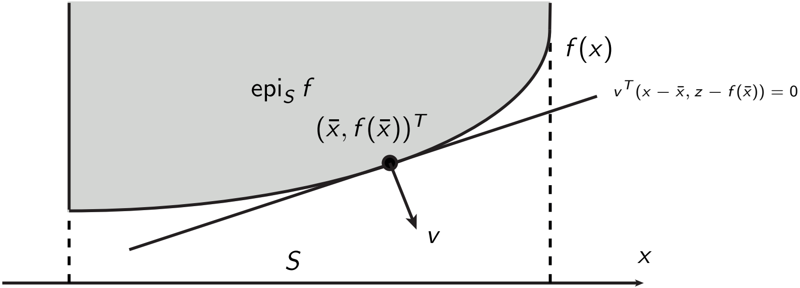
Let $C \subseteq \mathbb{R}^n$ be a nonempty and convex set. Let $\bar{\boldsymbol{x}}$ be a point on the boundary of C. Then there exists a **supporting hyperplane** to C at $\bar{\boldsymbol{x}}$, meaning that there exists $v \neq \mathbf{0}^n$ such that

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



Subgradient by supporting hyperplane, I

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



- v defines subgradient only when hyperplane is "non-vertical"

Subgradient by supporting hyperplane, II

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

$$\forall (x,z) \in \text{epi}_S f, \quad v^T \left(\begin{pmatrix} x \\ z \end{pmatrix} - \begin{pmatrix} \bar{x} \\ f(\bar{x}) \end{pmatrix} \right) \leq 0, \quad \text{some } v \neq \mathbf{0}$$

- Write $v = (u, t) \in \mathbb{R}^{n+1}$, with $u \in \mathbb{R}^n$. 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

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

- If $t = 0$, then $u^T(x - \bar{x}) \leq 0$ for all $x \in S$. Impossible if $\bar{x} \in \text{int } S$.

Subgradient, summary

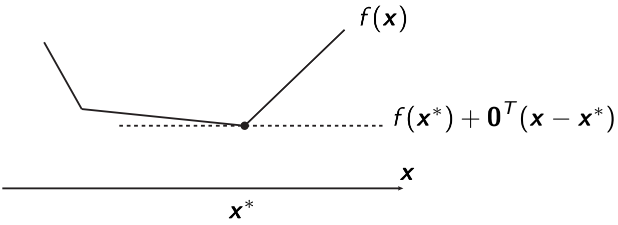
- Now we know subgradient exists for convex f over int S. But...
- How do we find a subgradient when we know at least one exists?
  - We will see one (later in this lecture) when dealing with dual function (in Lagrange duality)
- What is the use of a subgradient? We use it to define...
  - Optimality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

  for convex problems with non-differentiable obj. functions.

Optimality conditions with subgradients

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:

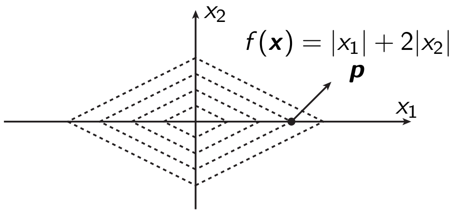
1. f is globally minimized at $\boldsymbol{x}^* \in \mathbb{R}^n$;
2. $\mathbf{0}^n \in \partial f(\boldsymbol{x}^*)$;



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

Subgradient methods

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



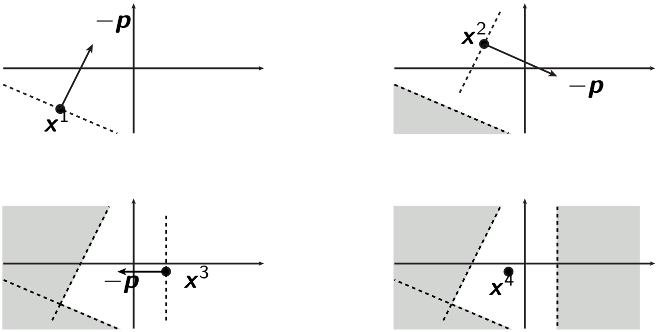
- It however can move us "towards" to the optimal solution

$$-(p^k)^T(x^* - x^k) \geq f(x^k) - f(x^*) \geq 0$$

Subgradient methods

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

$$f(x) \geq f(x^k) + p^T(x - x^k) \implies p^T x \leq p^T x^k \quad \text{(the halfspace containing } x^*\text{)}$$



Subgradient method in unconstrained case

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

Step 1 Find a subgradient $\boldsymbol{p}^k$ to f in $\boldsymbol{x}^k$.

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

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

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

Step 1 Find a subgradient $\boldsymbol{p}^k$ to f in $\boldsymbol{x}^k$.

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

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

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, Lagrange dual problem is naturally suited for subgradient methods.

The Lagrangian dual

Consider the Lagrangian relaxation of the problem to find

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

We first construct the Lagrange function

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

and define the dual function as

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

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

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

- 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.
- Subgradient methods are often utilized to solve the dual problem.

Subgradients for the dual

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

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

  Let the solution set to this problem be

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

- If we take any $\boldsymbol{x} \in X(\boldsymbol{\mu})$, then $\boldsymbol{g}(\boldsymbol{x})$ 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 $\boldsymbol{x} \in X(\boldsymbol{\mu})$, then $\boldsymbol{g}(\boldsymbol{x})$ is a subgradient to q at $\boldsymbol{\mu}$, that is, $\boldsymbol{g}(\boldsymbol{x}) \in \partial q(\boldsymbol{\mu})$.

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

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

Proof: See Proposition 6.20 in text.

Subgradient method for maximizing 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)

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

Let the solution to the problem be $\boldsymbol{x}^k$.

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

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