CS 2601 Linear and Convex Optimization

5. Convex optimization problems (part 1)

Bo Jiang

John Hopcroft Center for Computer Science
Shanghai Jiao Tong University

Fall 2022

Outline

- Convex optimization problems
- Linear program

Optimization problems in standard form

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s. t.} \quad & g_i(\boldsymbol{x}) \leq 0, \quad i = 1, 2, \ldots, m \\ & h_i(\boldsymbol{x}) = 0, \quad i = 1, 2, \ldots, k \end{aligned} \tag{P}$$

- $\boldsymbol{x} \in \mathbb{R}^n$ optimization/decision variable
- $f : \operatorname{dom} f \subset \mathbb{R}^n \to \mathbb{R}$ objective function
- $g_i : \operatorname{dom} g_i \subset \mathbb{R}^n \to \mathbb{R}$ are inequality constraint functions
- $h_i : \operatorname{dom} h_i \subset \mathbb{R}^n \to \mathbb{R}$ are equality constraint functions
- The domain of the problem (P) is

$$D = \operatorname{dom} f \cap \left(\bigcap_{i=1}^{m} \operatorname{dom} g_i \right) \cap \left(\bigcap_{i=1}^{k} \operatorname{dom} h_i \right)$$

- The feasible set is

$$X = \{\boldsymbol{x} \in D : g_i(\boldsymbol{x}) \leq 0, 1 \leq i \leq m; h_i(\boldsymbol{x}) = 0, 1 \leq i \leq k\}$$

The problem (P) is called feasible if $X \neq \emptyset$.

Optimal value

The optimal value of the optimization problem (P) is

$$f^* = \inf_{\boldsymbol{x} \in X} f(\boldsymbol{x})$$

We allow f^* to take the extended values $\pm\infty$.

- $f^* = \infty$ if (P) is infeasible, i.e. $X = \emptyset$
  - We use the standard convention $\sup \emptyset = -\infty$ and $\inf \emptyset = \infty$.
- $f^* = -\infty$ if (P) is unbounded below
  - There exists a sequence $\boldsymbol{x}_i \in X$ s.t. $f(\boldsymbol{x}_i) \to -\infty$ as $i \to \infty$.
- $\boldsymbol{x}^*$ is an optimal point of (P) or solves (P), if $\boldsymbol{x}^* \in X$ and $f^* = f(\boldsymbol{x}^*)$, i.e. $\boldsymbol{x}^*$ is feasible and attains the optimal value.
  - Recall f^* is not always attainable, e.g. $f(x) = e^x$, $f^* = 0$.
- $\boldsymbol{x}_0$ is called ϵ-suboptimal if $\boldsymbol{x}_0 \in X$ and $f(\boldsymbol{x}_0) \leq f^* + \epsilon$.
- $\boldsymbol{x}^*$ is called locally optimal if it solves (P) with the additional constraint $\|\boldsymbol{x} - \boldsymbol{x}^*\| \leq \delta$ for some $\delta > 0$.

Convex optimization problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) \le 0, \quad i = 1, 2, \ldots, m \\ & h_i(\boldsymbol{x}) = 0, \quad i = 1, 2, \ldots, k \end{aligned}$$

Affine function's domain is always the entire space

The above is called a convex optimization problem[1] if

1. f, g_i are convex functions
2. h_i are affine functions, i.e. $h_i(\boldsymbol{x}) = \boldsymbol{a}_i^T \boldsymbol{x} - b_i$

The domain of the optimization problem is

$$D = \operatorname{dom} f \cap \left(\bigcap_{i=1}^{m} \operatorname{dom} g_i \right)$$

Sublevel set

hyperplane

Feasible set $X = \{\boldsymbol{x} \in D : g_i(\boldsymbol{x}) \le 0, 1 \le i \le m; h_i(\boldsymbol{x}) = 0, 1 \le i \le k\}$

Note. Both D and X are convex sets (why?)

[1] $\max\ f(\boldsymbol{x})$ for a concave f is also called a convex optimization problem

Example

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = x_1^2 + x_2^2 \\ \text{s. t.} \quad & g(\boldsymbol{x}) = x_1/(1 + x_2^2) \leq 0 \\ & h(\boldsymbol{x}) = (x_1 + x_2)^2 = 0 \end{aligned}$$

Not an affine function

- f is convex, feasible set $X = \{\boldsymbol{x} : x_1 + x_2 = 0, x_1 \leq 0\}$ is convex
- not a convex problem according to our definition[2]
  - g not convex (check $\nabla^2 g$), h not affine

Equivalent (but not identical) convex problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = x_1^2 + x_2^2 \\ \text{s. t.} \quad & g(\boldsymbol{x}) = x_1 \leq 0 \\ & h(\boldsymbol{x}) = x_1 + x_2 = 0 \end{aligned}$$

[2] It is a convex problem according to the broader definition of “minimizing a convex function over a convex set”. We will use the more stringent definition, as the broader one may hide the complexity of the problem in the description of the feasible set.

Properties

For convex optimization problems,

- The set of solutions (global minima) X_{opt} is convex

$$X_{\text{opt}} = \{\boldsymbol{x}^* \in X : f(\boldsymbol{x}^*) \le f(\boldsymbol{x}), \forall x \in X\}$$

- Any local minimum is a global minimum
  - $\boldsymbol{x}^*$ is a local minimum if it solves the following optimization problem

No practical use here,
Only conceptual

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_i(\boldsymbol{x}) \le 0, \quad i = 1, 2, \dots, m \\ & h_i(\boldsymbol{x}) = 0, \quad i = 1, 2, \dots, k \\ & \|\boldsymbol{x} - \boldsymbol{x}^*\| \le \delta \end{aligned}$$

for some $\delta > 0$

- If f is strictly convex, at most one solution, i.e. $|X_{\text{opt}}| \le 1$

First-order optimality condition

Theorem. For a convex problem whose objective f is differentiable with open domain $\mathrm{dom} f$, a feasible point $\boldsymbol{x}^* \in X$ is optimal iff

$$\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0, \quad \forall \boldsymbol{x} \in X$$

Proof. "$\Leftarrow$". Assume the above condition. By the first-order condition for convexity,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) \geq f(\boldsymbol{x}^*), \quad \forall \boldsymbol{x} \in X$$

"$\Rightarrow$". Assume $\boldsymbol{x}^*$ is optimal. Since X is convex, for $\boldsymbol{x} \in X$, $\alpha \in [0, 1]$,

$$\boldsymbol{x}^* + \alpha(\boldsymbol{x} - \boldsymbol{x}^*) = \alpha \boldsymbol{x} + \bar{\alpha}\boldsymbol{x}^* \in X$$

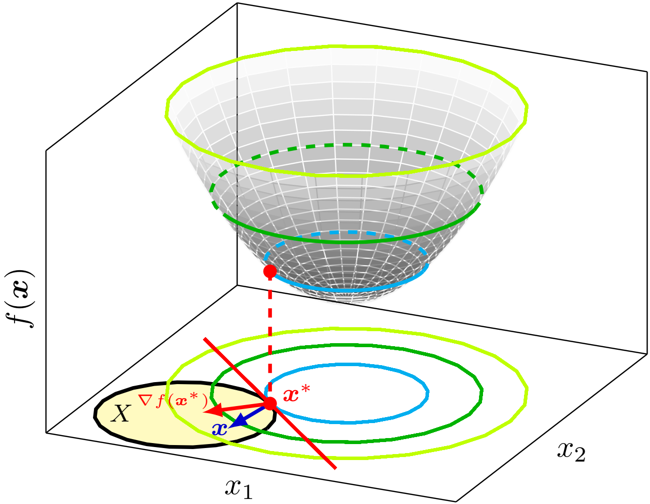
so $\boldsymbol{d} = \boldsymbol{x} - \boldsymbol{x}^*$ is a feasible direction. By slide 13 of $\S 2$,

$$\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) = \nabla f(\boldsymbol{x}^*)^T \boldsymbol{d} \geq 0.$$

First-order optimality condition (cont'd)

Or just let d = −nabla f(x)

- If $\boldsymbol{x}^* \in \operatorname{int} X$ (e.g. $X = \mathbb{R}^n$), then $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$ (why?)
- If $\nabla f(\boldsymbol{x}^*) \neq \mathbf{0}$, then $\boldsymbol{x}^* \in \partial X$ and $\nabla f(\boldsymbol{x}^*)^T(\boldsymbol{x} - \boldsymbol{x}^*) = 0$ is a supporting hyperplane of X at $\boldsymbol{x}^*$.



First-order optimality condition (cont'd)

The first-order optimality condition also applies to the general case of minimizing a convex function over a convex set.

Example. Recall the distance of $\boldsymbol{x}_0$ to a convex set C is the optimal value of the problem

$$\begin{array}{ll} \min\limits_{\boldsymbol{x}} & \|\boldsymbol{x}-\boldsymbol{x}_0\| \\ \text{s.t.} & \boldsymbol{x} \in C \end{array} \quad \Longleftrightarrow \quad \begin{array}{ll} \min\limits_{\boldsymbol{x}} & f(\boldsymbol{x}) = \|\boldsymbol{x}-\boldsymbol{x}_0\|^2 \\ \text{s.t.} & \boldsymbol{x} \in C \end{array}$$

By the first-order condition, $\boldsymbol{x}^* \in C$ is optimal iff

$$\langle \nabla f(\boldsymbol{x}^*), \boldsymbol{x}-\boldsymbol{x}^* \rangle = 2\langle \boldsymbol{x}^* - \boldsymbol{x}_0, \boldsymbol{x}-\boldsymbol{x}^* \rangle \geq 0, \quad \forall \boldsymbol{x} \in C$$

which is the condition on slide 30 of §3 for $\boldsymbol{x}^* = \mathcal{P}_C(\boldsymbol{x}_0)$.

Outline

- Convex optimization problems

- Linear program

Linear program

A linear program (LP) is an optimization problem of the form

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{B}\boldsymbol{x} \leq \boldsymbol{d} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

LP is a convex optimization problem.

Standard form

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

Inequality form

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \end{aligned}$$

Informally, two optimization problems are considered equivalent if the solution of one problem can be easily obtained from the solution of the other, and vice versa.

Conversion to standard form

松弛变量

Eliminate inequality constraints by introducing slack variables $\boldsymbol{s}$

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{B}\boldsymbol{x} \le \boldsymbol{d} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \min_{\boldsymbol{x},\boldsymbol{s}} \quad & \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{B}\boldsymbol{x} + \boldsymbol{s} = \boldsymbol{d} \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{s} \ge \boldsymbol{0} \end{aligned}$$

Split variables into positive and negative parts $\boldsymbol{x} = \boldsymbol{x}^+ - \boldsymbol{x}^-$

$$\begin{aligned} \min_{\boldsymbol{x}^+,\boldsymbol{x}^-,\boldsymbol{s}} \quad & \boldsymbol{c}^T\boldsymbol{x}^+ - \boldsymbol{c}^T\boldsymbol{x}^- \\ \text{s.t.} \quad & \boldsymbol{B}\boldsymbol{x}^+ - \boldsymbol{B}\boldsymbol{x}^- + \boldsymbol{s} = \boldsymbol{d} \\ & \boldsymbol{A}\boldsymbol{x}^+ - \boldsymbol{A}\boldsymbol{x}^- = \boldsymbol{b} \\ & \boldsymbol{x}^+ \ge \boldsymbol{0}, \quad \boldsymbol{x}^- \ge \boldsymbol{0}, \quad \boldsymbol{s} \ge \boldsymbol{0} \end{aligned}$$

Conversion to standard form (cont'd)

Example. LP in inequality form

$$\begin{aligned} \min_{x_1, x_2} \quad & x_1 + 2x_2 \\ \text{s.t.} \quad & x_1 + x_2 \leq 1 \end{aligned}$$

Introduce slack variable s,

$$\begin{aligned} \min_{x_1, x_2, s} \quad & x_1 + 2x_2 \\ \text{s.t.} \quad & x_1 + x_2 + s = 1 \\ & s \geq 0 \end{aligned}$$

Let $x_1 = x_1^+ - x_1^-$, $x_2 = x_2^+ - x_2^-$,

$$\begin{aligned} \min_{x_1^+, x_1^-, x_2^+, x_2^-, s} \quad & x_1^+ - x_1^- + 2x_2^+ - 2x_2^- \\ \text{s.t.} \quad & x_1^+ - x_1^- + x_2^+ - x_2^- + s = 1 \\ & x_1^+ \geq 0, x_1^- \geq 0, x_2^+ \geq 0, x_2^- \geq 0, s \geq 0 \end{aligned}$$

Conversion to inequality form

Example. LP in standard form

$$\begin{aligned} \min_{x_1,x_2,x_3} \quad & x_1 + 3x_2 + 2x_3 \\ \text{s.t.} \quad & x_1 + x_2 + x_3 = 1 \\ & x_1 \geq 0, x_2 \geq 0, x_3 \geq 0 \end{aligned}$$

Method 1. Eliminate x_3 using equality constraint, $x_3 = 1 - x_1 - x_2$,

$$\begin{aligned} \min_{x_1,x_2} \quad & -x_1 + x_2 \quad \text{(we removed a constant 2)} \\ \text{s.t.} \quad & -x_1 \leq 0, -x_2 \leq 0, x_1 + x_2 \leq 1 \end{aligned}$$

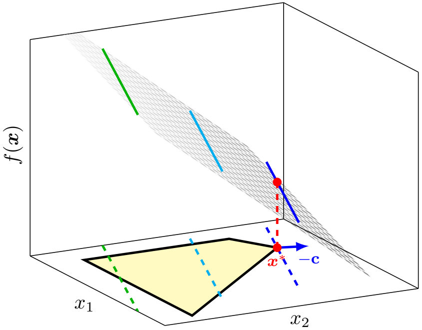
Method 2. Rewrite the equality constraints as two inequality constraints,

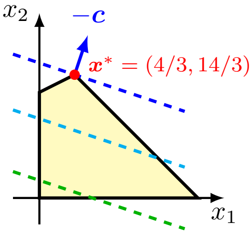
$$\begin{aligned} \min_{x_1,x_2,x_3} \quad & x_1 + 3x_2 + 2x_3 \\ \text{s.t.} \quad & x_1 + x_2 + x_3 \leq 1 \\ & -x_1 - x_2 - x_3 \leq -1 \\ & -x_1 \leq 0, -x_2 \leq 0, -x_3 \leq 0 \end{aligned}$$

Geometry of LP

$$\min_{\boldsymbol{x}} \quad \boldsymbol{c}^T \boldsymbol{x}$$
$$\text{s.t.} \quad \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}$$

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & -x_1 - 3x_2 \\ \text{s.t.} \quad & x_1 + x_2 \le 6 \\ & -x_1 + 2x_2 \le 8 \\ & x_1, x_2 \ge 0 \end{aligned}$$





- optimize linear function over a polyhedron
- optimal solution exists? not always (why?) Empty or infinity
- optimal solution unique? not always (why?)
- when optimal solution exists, there is always a vertex solution

Where is my little point?

Example: Basis pursuit

Let $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ with $\operatorname{rank} \boldsymbol{X} = n < p$. The underdetermined linear system $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$ has infinitely many solutions $\boldsymbol{w}$. 稀疏化表达

We want the sparsest solution, i.e. the representation of $\boldsymbol{y}$ by the smallest number of columns of $\boldsymbol{X}$.

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \|\boldsymbol{w}\|_0 \\ \text{s.t.} \quad & \boldsymbol{X}\boldsymbol{w} = \boldsymbol{y} \end{aligned}$$

其实不是范数

where the ℓ_0 "norm" $\|\boldsymbol{w}\|_0 = \sum_{j=1}^{p} \mathbb{1}\{w_j \neq 0\}$ is nonconvex (check!)

The ℓ_1 approximation, called basis pursuit[3], is convex

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \|\boldsymbol{w}\|_1 = \sum_{j=1}^{p} |w_j| \\ \text{s.t.} \quad & \boldsymbol{X}\boldsymbol{w} = \boldsymbol{y} \end{aligned}$$

[3]We are trying to find a small set of "basis" vectors from columns of $\boldsymbol{X}$

Example: Basis pursuit (cont'd)

Basis pursuit

$$\begin{aligned} \min_{\boldsymbol{w}} \quad & \|\boldsymbol{w}\|_1 = \sum_{j=1}^{p} |w_j| \\ \text{s.t.} \quad & \boldsymbol{X}\boldsymbol{w} = \boldsymbol{y} \end{aligned}$$

can be reformulated as an LP by introducing variables t_j, $j = 1, 2, \ldots, p$

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{t}} \quad & \mathbf{1}^T \boldsymbol{t} = \sum_{j=1}^{p} t_j \\ \text{s.t.} \quad & \boldsymbol{X}\boldsymbol{w} = \boldsymbol{y} \\ & t_j \geq |w_j|, \; j = 1, 2, \ldots, p \end{aligned} \quad \Longrightarrow \quad \begin{aligned} \min_{\boldsymbol{w},\boldsymbol{t}} \quad & \mathbf{1}^T \boldsymbol{t} \\ \text{s.t.} \quad & \boldsymbol{X}\boldsymbol{w} = \boldsymbol{y} \\ & -\boldsymbol{t} \leq \boldsymbol{w} \leq \boldsymbol{t} \end{aligned}$$

Note. Another possibility is to let $w_i = w_i^+ - w_i^-$ with $w_i^+, w_i^- \geq 0$, so $|w_i| = w_i^+ + w_i^-$.

Example: Basis pursuit (cont'd)

Example.

$$\min_{w_1,w_2,w_3} \quad |w_1| + |w_2| + |w_3|$$

$$\text{s.t.} \quad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

LP reformulation

$$\min_{w_1,w_2,w_3,t_1,t_2,t_3} \quad t_1 + t_2 + t_3$$

$$\text{s.t.} \quad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$$

$$-t_1 \le w_1 \le t_1$$

$$-t_2 \le w_2 \le t_2$$

$$-t_3 \le w_3 \le t_3$$

Example: Piecewise linear minimization

The unconstrained problem

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) = \max_{i \le i \le m} (\boldsymbol{a}_i^T \boldsymbol{x} + b_i)$$

is convex.

Can be reformulated as an LP.

- transform into epigraph form by introducing variable t,

$$\begin{aligned} \min_{\boldsymbol{x},t} \quad & t \\ \text{s.t.} \quad & t \ge \max_{i \le i \le m} (\boldsymbol{a}_i^T \boldsymbol{x} + b_i) \end{aligned}$$

- equivalent to

$$\begin{aligned} \min_{\boldsymbol{x},t} \quad & t \\ \text{s.t.} \quad & t \ge \boldsymbol{a}_i^T \boldsymbol{x} + b_i, \quad i = 1, 2, \ldots, m \end{aligned}$$

Example: Piecewise linear minimization (cont'd)

Exmaple.

$$\min_{x_1,x_2} \quad \max\{x_1 + 2x_2, 2x_1 - x_2, 3x_1 + x_2\}$$

LP reformulation

$$\begin{aligned} \min_{x_1,x_2,t} \quad & t \\ & x_1 + 2x_2 - t \leq 0 \\ & 2x_1 - x_2 - t \leq 0 \\ & 3x_1 + x_2 - t \leq 0 \end{aligned}$$