

Lecture 7: Constrained optimization



School of Mathematical Sciences, Xiamen University

1. Convex optimization

- A convex optimization problem (or a convex problem)

$$\min_{\mathbf{x} \in \mathcal{C}} f(\mathbf{x}),$$

where \mathcal{C} is a convex set and f is a convex function.

Convex optimization problems in functional form

$$\begin{array}{ll} \min & f(\mathbf{x}) \\ \text{s.t.} & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m, \\ & h_j(\mathbf{x}) = 0, \quad j = 1, 2, \dots, p, \end{array}$$

where $f, g_1, g_2, \dots, g_m : \mathbb{R}^n \rightarrow \mathbb{R}$ are convex functions and $h_1, h_2, \dots, h_p : \mathbb{R}^n \rightarrow \mathbb{R}$ are affine functions. The convex set \mathcal{C} is

$$\mathcal{C} = \left(\bigcap_{i=1}^m \text{Lev}(g_i, 0) \right) \cap \left(\bigcap_{j=1}^p \{\mathbf{x} : h_j(\mathbf{x}) = 0\} \right).$$

Theorem 1 (local = global in convex optimization)

Let $f : \mathcal{C} \rightarrow \mathbb{R}$ be a (strictly) convex function defined on the convex set \mathcal{C} . Let $\mathbf{x}_\star \in \mathcal{C}$ be a local minimizer of f over \mathcal{C} . Then \mathbf{x}_\star is a (strict) global minimizer of f over \mathcal{C} .

Theorem 2

Let $f : \mathcal{C} \rightarrow \mathbb{R}$ be a convex function defined over the convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Then the set of optimal solutions of the problem $\min\{f(\mathbf{x}) : \mathbf{x} \in \mathcal{C}\}$, which we denote by \mathcal{X}_\star , is convex. If, in addition, f is strictly convex over \mathcal{C} , then there exists at most one optimal solution.

CVX: a Matlab-based convex modeling framework

[CVX](#)[TFOCS](#)[About us](#)[News](#)[CVX Forum](#)

2. Optimization over a convex set

- Let f be a continuously differentiable function over a closed convex set \mathcal{C} . Then $\mathbf{x}_\star \in \mathcal{C}$ is called a *stationary point* of

$$(P) \quad \min f(\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \mathcal{C},$$

if $\nabla f(\mathbf{x}_\star)^\top (\mathbf{x} - \mathbf{x}_\star) \geq 0$ for any $\mathbf{x} \in \mathcal{C}$.

Theorem 3 (stationarity as a necessary optimality condition)

Let f be a continuously differentiable function over a closed convex set \mathcal{C} , and let \mathbf{x}_\star be a local minimizer of (P). Then \mathbf{x}_\star is a stationary point of (P).

Theorem 4

Let f be a continuously differentiable convex function over a closed convex set $\mathcal{C} \subseteq \mathbb{R}^n$. Then $\mathbf{x}_\star \in \mathcal{C}$ is a stationary point of (P) if and only if \mathbf{x}_\star is an optimal solution of (P).

2.1 The gradient projection method

- The projection

$$\pi_{\mathcal{C}}(\mathbf{x}) = \operatorname{argmin}_{\mathbf{y} \in \mathcal{C}} \|\mathbf{y} - \mathbf{x}\|_2^2.$$

Theorem 5

Let \mathcal{C} be a nonempty closed convex set. Then for any $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$,

$$\|\pi_{\mathcal{C}}(\mathbf{v}) - \pi_{\mathcal{C}}(\mathbf{w})\|_2^2 \leq (\pi_{\mathcal{C}}(\mathbf{v}) - \pi_{\mathcal{C}}(\mathbf{w}))^\top (\mathbf{v} - \mathbf{w}),$$

$$\|\pi_{\mathcal{C}}(\mathbf{v}) - \pi_{\mathcal{C}}(\mathbf{w})\|_2 \leq \|\mathbf{v} - \mathbf{w}\|_2.$$

Theorem 6

Let f be a continuously differentiable function defined on the nonempty closed convex set \mathcal{C} , and let $s > 0$. Then $\mathbf{x}_\star \in \mathcal{C}$ is a stationary point of (P) if and only if

$$\mathbf{x}_\star = \pi_{\mathcal{C}}(\mathbf{x}_\star - s \nabla f(\mathbf{x}_\star)).$$

- The gradient projection method

$$\mathbf{x}_{k+1} = \pi_{\mathcal{C}}(\mathbf{x}_k - t_k \nabla f(\mathbf{x}_k)),$$

where $t_k > 0$ is obtained by using a line search procedure.

Lemma 7

Suppose that $f \in C_L^{1,1}(\mathcal{C})$, where \mathcal{C} is a nonempty closed convex set. Then for any $\mathbf{x} \in \mathcal{C}$ and $t \in (0, 2/L)$ the following inequality holds:

$$f(\mathbf{x}) - f(\pi_{\mathcal{C}}(\mathbf{x} - t \nabla f(\mathbf{x}))) \geq (1/t - L/2) \|\mathbf{x} - \pi_{\mathcal{C}}(\mathbf{x} - t \nabla f(\mathbf{x}))\|_2^2$$

- Define the gradient mapping $G_M(\mathbf{x}) = M[\mathbf{x} - \pi_{\mathcal{C}}(\mathbf{x} - \nabla f(\mathbf{x})/M)]$

Lemma 8

Let f be a continuously differentiable function defined on a nonempty closed convex set \mathcal{C} . Suppose that $L_1 \geq L_2 > 0$. Then for any $\mathbf{x} \in \mathbb{R}^n$,

$$\|G_{L_1}(\mathbf{x})\|_2 \geq \|G_{L_2}(\mathbf{x})\|_2, \quad \|G_{L_1}(\mathbf{x})\|_2/L_1 \leq \|G_{L_2}(\mathbf{x})\|_2/L_2.$$

- **constant stepsize:**

$$t_k = \bar{t} \in \left(0, \frac{2}{L}\right).$$

- **backtracking:** $s > 0$, $\alpha \in (0, 1)$, $\beta \in (0, 1)$.

First, set $t_k = s$. Then, while

$$f(\mathbf{x}_k) - f(\pi_{\mathcal{C}}(\mathbf{x}_k - t_k \nabla f(\mathbf{x}_k))) < \alpha t_k \|G_{1/t_k}(\mathbf{x}_k)\|_2^2,$$

set $t_k \leftarrow \beta t_k$. In other words, $t_k = s\beta^{i_k}$, where i_k is the smallest nonnegative integer satisfying (the sufficient decrease condition)

$$f(\mathbf{x}_k) - f(\pi_{\mathcal{C}}(\mathbf{x}_k - s\beta^{i_k} \nabla f(\mathbf{x}_k))) \geq \alpha s \beta^{i_k} \|G_{1/(s\beta^{i_k})}(\mathbf{x}_k)\|_2^2.$$

If $f \in C_L^{1,1}(\mathcal{C})$, then the backtracking procedure ends when t_k is smaller than or equal to $2(1 - \alpha)/L$. The chosen stepsize t_k satisfies

$$t_k \geq \min \left\{ s, \frac{2(1 - \alpha)\beta}{L} \right\}.$$

Theorem 9 (convergence of the gradient projection method)

Let $f \in C_L^{1,1}(\mathcal{C})$ and \mathcal{C} be a nonempty closed convex set. Let $\{\mathbf{x}_k\}$ be the sequence generated by the gradient projection method for solving (P) with either a constant stepsize $\bar{t} \in (0, 2/L)$ or with a stepsize chosen by the backtracking procedure with parameters $s > 0$, $\alpha \in (0, 1)$, $\beta \in (0, 1)$. Assume that f is bounded below. Then we have the following:

- (a) The sequence $\{f(\mathbf{x}_k)\}$ is nonincreasing. In addition, for any $k > 0$, $f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k)$ unless \mathbf{x}_k is a stationary point of (P).
- (b) $G_d(\mathbf{x}_k) \rightarrow \mathbf{0}$ as $k \rightarrow \infty$, and

$$\min_{k=0,1,\dots,n} \|G_d(\mathbf{x}_k)\|_2 \leq \sqrt{\frac{f(\mathbf{x}_0) - f_\star}{M(n+1)}},$$

where $f_\star = \lim_{k \rightarrow \infty} f(\mathbf{x}_k)$, and

$$d = \begin{cases} 1/\bar{t}, & \text{constant stepsize,} \\ 1/s, & \text{backtracking.} \end{cases} \quad M = \begin{cases} \bar{t}(1 - \bar{t}L/2), & \text{constant stepsize,} \\ \alpha \min\{s, 2(1 - \alpha)\beta/L\} & \text{backtracking.} \end{cases}$$

Theorem 10

Let $f \in C_L^{1,1}(\mathcal{C})$ be convex and \mathcal{C} be a nonempty closed convex set. Let $\{\mathbf{x}_k\}$ be the sequence generated by the gradient projection method for solving (P) with a constant stepsize $\bar{t} \in (0, 1/L]$. Assume that the set of optimal solutions, denoted by \mathcal{X}_\star , is nonempty, and let f_\star be the optimal value of (P). Then we have the following:

(a) for any $k \geq 0$ and $\mathbf{x}_\star \in \mathcal{X}_\star$,

$$2\bar{t}(f(\mathbf{x}_{k+1}) - f(\mathbf{x}_\star)) \leq \|\mathbf{x}_k - \mathbf{x}_\star\|_2^2 - \|\mathbf{x}_{k+1} - \mathbf{x}_\star\|_2^2,$$

which implies

$$\|\mathbf{x}_{k+1} - \mathbf{x}_\star\|_2 \leq \|\mathbf{x}_k - \mathbf{x}_\star\|_2, \quad (\text{Fejér monotonicity})$$

(b) for any $n \geq 0$,

$$f(\mathbf{x}_n) - f_\star \leq \frac{\|\mathbf{x}_0 - \mathbf{x}_\star\|_2^2}{2\bar{t}n},$$

(c) the sequence $\{\mathbf{x}_k\}$ converges to an optimal solution.

3. Karush–Kuhn–Tucker conditions

Theorem 11 (KKT conditions for constrained problems)

Let \mathbf{x}_\star be a local minimizer of

$$\min f(\mathbf{x}), \quad \text{s.t.} \quad g_i(\mathbf{x}) \leq 0, \quad h_j(\mathbf{x}) = 0, \quad i = 1 : m, \quad j = 1 : p,$$

where f, g_i, h_j are continuously differentiable functions over \mathbb{R}^n . Suppose that the gradients of the active constraints and the equality constraints

$$\{\nabla g_i(\mathbf{x}_\star) : i \in I(\mathbf{x}_\star)\} \cup \{\nabla h_j(\mathbf{x}_\star) : j = 1 : p\}$$

are linearly independent (where $I(\mathbf{x}_\star) = \{i : g_i(\mathbf{x}_\star) = 0\}$). Then there exist multipliers $\lambda_i \geq 0$ and $\mu_j \in \mathbb{R}$ such that $\lambda_i g_i(\mathbf{x}_\star) = 0$, $i = 1 : m$,

$$\nabla f(\mathbf{x}_\star) + \sum_{i=1}^m \lambda_i \nabla g_i(\mathbf{x}_\star) + \sum_{j=1}^p \mu_j \nabla h_j(\mathbf{x}_\star) = \mathbf{0}.$$

Theorem 12 (sufficiency of KKT conditions for convex problems)

Let \mathbf{x}_\star be a local minimizer of

$$\min f(\mathbf{x}), \quad \text{s.t.} \quad g_i(\mathbf{x}) \leq 0, \quad h_j(\mathbf{x}) = 0, \quad i = 1 : m, \quad j = 1 : p,$$

where f, g_i are continuously differentiable convex functions over \mathbb{R}^n and h_j are affine functions. Suppose that there exist multipliers $\lambda_i \geq 0$ and $\mu_j \in \mathbb{R}$ such that

$$\lambda_i g_i(\mathbf{x}_\star) = 0, \quad i = 1 : m,$$

$$\nabla f(\mathbf{x}_\star) + \sum_{i=1}^m \lambda_i \nabla g_i(\mathbf{x}_\star) + \sum_{j=1}^p \mu_j \nabla h_j(\mathbf{x}_\star) = \mathbf{0}.$$

Then \mathbf{x}_\star is an optimal solution.

Theorem 13 (necessity of KKT conditions under Slater's condition)

Let \mathbf{x}_\star be a local minimizer of $\min f(\mathbf{x})$ such that

$$g_i(\mathbf{x}) \leq 0, \quad h_j(\mathbf{x}) \leq 0, \quad s_k(\mathbf{x}) = 0, \quad i = 1 : m, \quad j = 1 : p, \quad k = 1 : q,$$

where f, g_i are continuously differentiable convex functions over \mathbb{R}^n , and h_j, s_k are affine functions. Suppose that there exist $\hat{\mathbf{x}}$ such that

$$g_i(\hat{\mathbf{x}}) < 0, \quad h_j(\hat{\mathbf{x}}) \leq 0, \quad s_k(\hat{\mathbf{x}}) = 0, \quad i = 1 : m, \quad j = 1 : p, \quad k = 1 : q.$$

Then there exist multipliers $\lambda_i \geq 0$, $\eta_j \geq 0$, and $\mu_k \in \mathbb{R}$ such that

$$\lambda_i g_i(\mathbf{x}_\star) = 0, \quad i = 1 : m, \quad \eta_j h_j(\mathbf{x}_\star) = 0, \quad j = 1 : p,$$

$$\nabla f(\mathbf{x}_\star) + \sum_{i=1}^m \lambda_i \nabla g_i(\mathbf{x}_\star) + \sum_{j=1}^p \eta_j \nabla h_j(\mathbf{x}_\star) + \sum_{k=1}^q \mu_k \nabla s_k(\mathbf{x}_\star) = \mathbf{0}.$$

Then \mathbf{x}_\star is an optimal solution.

4. Duality

- The *primal problem*: Consider the general model

$$\begin{aligned} f_{\star} &= \min f(\mathbf{x}) \\ \text{s.t. } \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m, \\ & h_j(\mathbf{x}) = 0, \quad j = 1, 2, \dots, p, \\ & \mathbf{x} \in \mathcal{X}, \end{aligned}$$

where f, g_i, h_j are functions defined on the set $\mathcal{X} \subseteq \mathbb{R}^n$.

- The Lagrangian: $\mathbf{x} \in \mathcal{X}, \boldsymbol{\lambda} \in \mathbb{R}_+^m, \boldsymbol{\mu} \in \mathbb{R}^p$,

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^p \mu_j h_j(\mathbf{x}).$$

- The *dual objective function* $q : \mathbb{R}_+^m \times \mathbb{R}^p \rightarrow \mathbb{R} \cup \{-\infty\}$,

$$q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \min_{\mathbf{x} \in \mathcal{X}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}).$$

- The *dual problem*:

$$\begin{aligned} q_{\star} &= \max q(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t. } & (\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \text{dom}(q), \end{aligned}$$

where $\text{dom}(q) = \{(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathbb{R}_+^m \times \mathbb{R}^p : q(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty\}$.

Theorem 14 (convexity of the dual problem)

The domain $\text{dom}(q)$ of the dual objective function is a convex set, and q is a concave (i.e., $-q$ is convex) function over $\text{dom}(q)$.

Theorem 15 (weak duality theorem)

It holds that

$$q_{\star} \leq f_{\star},$$

where q_{\star} and f_{\star} are the optimal dual and primal values, respectively.

4.1 Strong duality in the convex case

Theorem 16 (convex problems with inequality constraints)

Consider the optimization problem

$$f_{\star} = \min f(\mathbf{x}) \quad \text{s.t.} \quad g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m, \quad \mathbf{x} \in \mathcal{X},$$

where \mathcal{X} is a convex set and f, g_i , are convex functions over \mathcal{X} .

Suppose that there exists $\hat{\mathbf{x}} \in \mathcal{X}$ for which $g_i(\hat{\mathbf{x}}) < 0$ and the optimal value of the primal problem is finite. Then the optimal value of the dual problem

$$q_{\star} = \max\{q(\boldsymbol{\lambda}) : \boldsymbol{\lambda} \in \text{dom}(q)\},$$

where $q(\boldsymbol{\lambda}) = \min_{\mathbf{x} \in \mathcal{X}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$ is attained, and the optimal values of the primal and dual problems are the same:

$$f_{\star} = g_{\star}.$$

Theorem 17

Consider the optimization problem

$$f_{\star} = \min f(\mathbf{x}) \quad \text{s.t.} \quad g_i(\mathbf{x}) \leq 0, \quad h_j(\mathbf{x}) \leq 0, \quad s_k(\mathbf{x}) = 0, \quad \mathbf{x} \in \mathcal{X},$$

where \mathcal{X} is a convex set and $f, g_i, i = 1 : m$, are convex functions over \mathcal{X} . The functions $h_j, s_k, j = 1 : p, k = 1 : q$, are affine functions. Suppose that there exists $\hat{\mathbf{x}} \in \text{int}(\mathcal{X})$ for which $g_i(\hat{\mathbf{x}}) < 0, h_j(\hat{\mathbf{x}}) \leq 0, s_k(\hat{\mathbf{x}}) = 0$. Then if the optimization problem has a finite optimal value, the optimal value of the dual problem

$$q_{\star} = \max\{q(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\mu}) : (\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\mu}) \in \text{dom}(q)\},$$

where $q : \mathbb{R}_+^m \times \mathbb{R}_+^p \times \mathbb{R}^q \rightarrow \mathbb{R} \cup \{-\infty\}$ is given by

$$q(\boldsymbol{\lambda}, \boldsymbol{\eta}, \boldsymbol{\mu}) = \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^p \eta_j h_j(\mathbf{x}) + \sum_{k=1}^q \mu_k s_k(\mathbf{x}),$$

is attained, and $f_{\star} = q_{\star}$.