

Preclass 04: Convex Optimization

[SCS4049] Machine Learning and Data Science

Seongsik Park (s.park@dgu.edu)

School of AI Convergence & Department of Artificial Intelligence, Dongguk University

Optimization problem in standard form

$$\text{minimize } f_0(x) \quad (1)$$

$$\text{subject to } f_i(x) \leq 0, \quad i = 1, 2, \dots, m \quad (2)$$

$$h_i(x) = 0, \quad i = 1, 2, \dots, p \quad (3)$$

- $x \in \mathcal{R}^n$ is the optimization variable
- $f_0 : \mathcal{R}^n \rightarrow \mathcal{R}$ is the objective or cost function
- $f_i : \mathcal{R}^n \rightarrow \mathcal{R}, i = 1, 2, \dots, m$ are the inequality constraint functions
- $h_i : \mathcal{R}^n \rightarrow \mathcal{R}$ are the equality constraint functions

Convex optimization problem

Standard form convex optimization problem

$$\text{minimize } f_0(x) \quad (4)$$

$$\text{subject to } f_i(x) \leq 0, \quad i = 1, 2, \dots, m \quad (5)$$

$$a_i^T x = b_i, \quad i = 1, 2, \dots, p \quad (6)$$

- $f_0, f_1, \dots, f_m$ are convex
- equality constraints are affine

Often written as

$$\text{minimize } f_0(x) \quad (7)$$

$$\text{subject to } f_i(x) \leq 0, \quad i = 1, 2, \dots, m \quad (8)$$

$$Ax = b \quad (9)$$

Important property: feasible set of a convex optimization problem is convex

Dual problem and KKT conditions

standard form problem

$$\text{minimize } f_0(x) \quad (10)$$

$$\text{subject to } f_i(x) \leq 0, \quad i = 1, 2, \dots, m \quad (11)$$

$$h_i(x) = 0, \quad i = 1, 2, \dots, p \quad (12)$$

variable $x \in \mathcal{R}^n$, domain $\mathcal{D}$, optimal value p^*

Lagrangian: $L : \mathcal{R}^n \times \mathcal{R}^m \times \mathcal{R}^p \rightarrow \mathcal{R}$ with $\text{dom } L = \mathcal{D} \times \mathcal{R}^m \times \mathcal{R}^p$

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \quad (13)$$

- weighted sum of objective and constraint functions
- λ_i is Lagrange multiplier associated with $f_i(x) \leq 0$
- ν_i is Lagrange multiplier associated with $h_i(x) = 0$

Lagrange dual function

Lagrange dual function: $g : \mathcal{R}^m \times \mathcal{R}^p \rightarrow \mathcal{R}$

$$g(\lambda, \nu) = \min_{x \in \mathcal{D}} L(x, \lambda, \nu) \quad (14)$$

$$= \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right) \quad (15)$$

g is concave, can be $-\infty$ for some λ, ν

lower bound property: if $\lambda \geq 0$, then $g(\lambda, \nu) \leq p^*$

proof: if $\tilde{x}$ is feasible and $\lambda \geq 0$, then

$$f_0(\tilde{x}) \geq L(\tilde{x}, \lambda, \nu) \geq \min_{x \in \mathcal{D}} L(x, \lambda, \nu) = g(\lambda, \nu) \quad (16)$$

minimizing over all feasible $\tilde{x}$ gives $p^* \geq g(\lambda, \nu)$

The dual problem

Lagrange dual problem

$$\text{maximize } g(\lambda, \nu) \quad (17)$$

$$\text{subject to } \lambda \geq 0 \quad (18)$$

- finds best lower bound on p^* , obtained from Lagrange dual function
- a convex optimization problem; optimal value denoted d^*
- λ, ν are dual feasible if $\lambda \geq 0, (\lambda, \nu) \in \text{dom } g$
- often simplified by making implicit constraint $(\lambda, \nu) \in \text{dom } g$ explicit

Weak and strong duality

weak duality: $d^* \leq p^*$

- always holds (for convex and nonconvex problems)
- can be used to find nontrivial lower bounds for difficult problems

strong duality: $d^* = p^*$

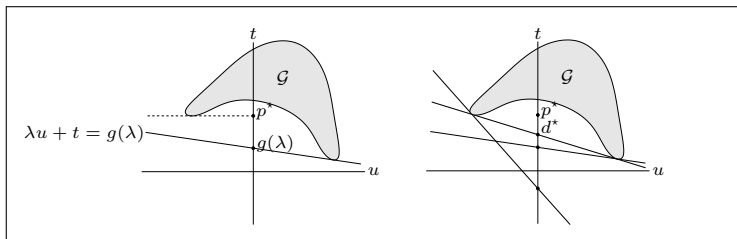
- does not hold in general
- holds for convex problems
- conditions that guarantee strong duality in convex problems are called constraint qualifications

Geometric interpretation

for simplicity, consider problem with one constraint $f_1(x) \leq 0$

interpretation of dual function

$$g(\lambda) = \min_{(u,t) \in \mathcal{G}} (t + \lambda u) \quad \text{where } \mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\} \quad (19)$$



- $\lambda u + t = g(\lambda)$ is supporting hyperplane to $\mathcal{G}$
- hyperplane intersects t -axis at $t = g(\lambda)$

Karush-Kuhn-Tucker (KKT) conditions

the following four conditions are called KKT conditions (for a problem with differentiable f_i, h_i)

1. primal constraints: $f_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p$
2. dual constraints: $\lambda \geq 0$
3. complementary slackness: $\lambda_i f_i(x) = 0, i = 1, \dots, m$
4. gradient of Lagrangian with respect to x vanishes:

$$\nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) + \sum_{i=1}^p \nu_i \nabla h_i(x) = 0 \quad (20)$$

if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions

Reference and further reading

- “Chap 7 | Sparse Kernel Machines” of C. Bishop, Pattern Recognition and Machine Learning
- “Chap 5 | Support Vector Machines” of A. Geron, Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow
- “Chap 4 | Convex Optimization Problems”, “Chap 5 | Duality” of S. Boyd, Convex Optimization