Convex optimization problems & Duality

Convex Optimization Reading Group

Jaehyeon Shin

January 18, 2024

Table of Contents

Table of Contents

Basic terminology

Optimization problem in standard form

$$\begin{array}{ll} \text{minimize}_{x\in D} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_j(x) = 0, \;\; j = 1, \ldots, p, \end{array}$$

- $f_0(x)$: cost or objective function.
- $f_i(x)$: inequality constraint function.
- $h_j(x)$: equality constraint function.

implicit constraints

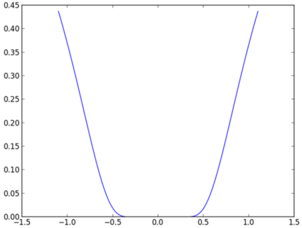
- The constraints $f_i(x) \leq 0, h_j(x) = 0$: explicit constraints
- Domain $D = \text{dom } f_0 \cap \bigcap_{i=1}^m \text{dom } (g_i) \cap \bigcap_{j=1}^p \text{dom } h_j$: implicit constraints

Basic terminology

Optimal value

$$p^\star = \inf\{f_0(x) | f_i(x) \leq 0, i = 1, ..., m, h_j(x) = 0, j = 1, ..., p\} \tag{1}$$

- x is feasible if $x \in \text{dom } f_0$ and satisfies the constraints
- if $f_0(x) = p^\star$, feasible x is optimal.
- If f_0 is strictly convex function, optimal point of f is unique.



Convex optimization problem in standard form

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \dots, m \\
& {a_i}^\top x = b_i, \quad i = 1, \dots, p,
\end{aligned}
$$

- $f_i(x), i = 0, \cdots, m$ are convex functions and equality constraints are linear
- feasible set of a convex optimization problem is convex

Local and global optima in convex optimization problem

Theorem 1

Any local optimal point is global optimal point!

Proof.

- Suppose x is locally optimal but not globally optimal. Then $\exists$ feasible y such that $f_0(y) < f_0(x)$.
- There is $R > 0$ such that x is local minima around the ball with center x and radius R.
- If $z = \theta y + (1-\theta)x$ for $0 \leq \theta \leq R/\|x-y\|_2$ then z is feasible, $\|z-x\|_2 \leq R$ and
- $f_0(z) = f_0((1-\theta)x + \theta y) \leq (1-\theta)f_0(x) + \theta f_0(y) < f_0(x)$: contradiction!

$\square$

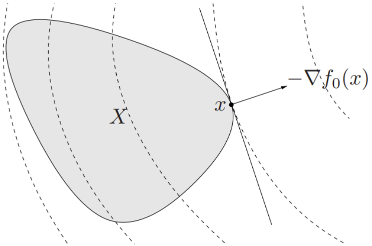
Optimality criterion for differentiable f_0

Theorem 2

x is optimal if and only if

$$\nabla f_0(x)^T(y - x) \geq 0 \text{ for all feasible } y \qquad (2)$$

If $\nabla f_0(x) \neq 0$, it means that $\nabla f_0(x)$ defines a supporting hyperplane to the feasible set at x.



Proof of Optimality Criterion

proof

($\Leftarrow$) If y is feasible and $\nabla f(x)^T(y-x) \geq 0$, by the first-order condition,

$$f_0(y) \geq f_0(x) + \nabla f_0(x)^T(y-x) \geq f_0(x)$$

So x is optimal point.
($\Rightarrow$) Assume that x is optimal.

- For feasible $y \neq x$, points of line segment are feasible by convexity of X. Let $g(t) = f_0(x + t(y-x))$. Then g is locally optimal at $t = 0$.
- Since $g'(t) = \nabla f_0(x + t(y-x))^T(y-x)$ we have $g'(0) = \nabla f_0(x)^T(y-x)$.
- If $g'(0) < 0, \exists t > 0$ such that $f_0(x + t(y-x)) < f_0(x)$, which contradict our assumption. Therefore $g'(0) = \nabla f_0(x)^T(y-x) \geq 0$.

Optimality criterion for unconstrained problem

Theorem 3

For unconstrained problem, x is optimal if and only if $x \in \mathbf{dom}\ f_0$ and $\nabla f_0(x) = 0$.

proof

- Suppose x is optimal, which means that $x \in \mathbf{dom}\ f_0$, and for all feasible y we have $\nabla f_0(x)^T(y - x) \geq 0$.
- All y sufficiently close to x are feasible. Take $y = x - t\nabla f_0(x)$ for sufficiently small $t > 0$. Then $\nabla f_0(x)^T(y - x) = -t\left\|\nabla f_0(x)\right\|_2^2 \geq 0$. Thus $\nabla f_0(x) = 0$.

assumption: dom f_0 is an open set if f_0 is differentiable

Table of Contents

Lagrangian

standard form problem (need not be convex)

$$
\begin{array}{ll}
\text{minimize}_{x \in D} & f_0(x) \\
\text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\
 & h_j(x) = 0, \;\; j = 1, \dots, p,
\end{array}
$$

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

Definition 1

Lagrangian $\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{r} \nu_j h_j(x)$.

- $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$
- λ_i and ν_j are Lagrange multiplier associated inequality and equality constraints.

Lagrangian dual function

Definition 2

Lagrange dual function: $g(\lambda, \nu) = \inf_{x \in D} L(x, \lambda, \nu)$

- concave (even if the problem is not convex) - infimum of affine functions of (λ, ν).

Theorem 4 (Lower bound property)

If $\lambda \succeq 0$ *, then* $g(\lambda, \nu) \leq p^\star$

Proof.

if $\tilde{x}$ is feasible and $\lambda \succeq 0$, then

$$\begin{aligned} f_0(\tilde{x}) &\geq L(\tilde{x}, \lambda, \nu) \qquad\qquad \because \lambda \succeq 0 \\ &\geq \inf_{x \in D} L(x, \lambda, \nu) = g(\lambda, \nu) \end{aligned}$$

□

Example: Standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array}$$

- Lagrangian
  $\mathcal{L}(x, \lambda . \nu) = c^T x + \nu^T (Ax - b) - \lambda^T x = -b^T \nu + (c + A^T \nu - \lambda)^T x$
- $\mathcal{L}$ is affine in x, so
  $g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \textit{otherwise} \end{cases}$

- Lower bound property : $p^\star \geq -b^T \nu$ if $A^T \nu + c \succeq 0$

Lagrange dual and conjugate function

Consider the following problem.

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & Ax \preceq b \\
& Cx = d
\end{aligned}
$$

Dual function is given as

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_{x \in dom f_0} \{ f_0 + (A^T\lambda + C^T\nu)^T x - b^T\lambda - d^T\nu \} \\
&= -\sup_{x \in dom f_0} \{ -f_0 - (A^T\lambda + C^T\nu)^T x \} - b^T\lambda - d^T\nu \\
&= -f_0{}^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu
\end{aligned}
$$

- simplifies derivation of dual function, if conjugate of f_0 is known
- conjugate $f^*(y) = \sup_{x \in dom f}(y^T x - f(x))$

The dual problem

$$
\begin{array}{ll}
\text{maximize} & g(\lambda, \nu) \\
\text{subject to} & \lambda \succeq 0
\end{array}
$$

- finds best lower bound on $p^\star$, obtained from Lagrange dual function.
- a convex optimization problem; optimal value denoted $d^\star$
- λ, ν are dual feasible if $\lambda \succeq 0, (\lambda, \nu) \in \text{dom } g$

Properties: weak and strong duality

Theorem 5 (Weak duality)

For all convex and non-convex problems, the following always holds.

$$d^* = \sup_{\lambda \geq 0, \mu} g(\lambda, \mu) \leq \inf_{x \in D} f_0(x) = p^\star \tag{3}$$

$p^\star - d^\star \geq 0$ *is called the optimal duality gap.*

Definition 3 (Strong duality)

If

$$d^* = \sup_{\lambda \geq 0, \mu} g(\lambda, \mu) = \inf_{x \in D} f_0(x) = p^\star \tag{4}$$

then strong duality holds.

- the duality gap is zero.
- usually (not always) holds for convex problems.

Strong duality example

- strong duality for a nonconvex program
  - $$\begin{array}{ll} \text{minimize} & \sqrt{x} \\ \text{subject to} & \frac{1}{x} - 1 \leq 0 \end{array}$$ has dual problem $$\begin{array}{ll} \text{maximize} & \frac{3}{2}\lambda - \frac{1}{2}\lambda^3 \\ \text{subject to} & \lambda \geq 0 \end{array}$$ which has optimal value 1 at $\lambda = 1$.
- Strong duality fails for a convex program
  - The two-dimensional domain $D = \{(x, y) : y > 0\} \subseteq R^2$
  - $$\begin{array}{ll} \min\limits_{x \in D} & e^{-x} \\ \text{subject to} & \frac{x^2}{y} \leq 0 \end{array}$$ has dual problem $$\begin{array}{ll} \text{maximize} & 0 \\ \text{subject to} & \lambda \geq 0 \end{array}$$

Slater's condition

Definition 4 (Slater's condition)

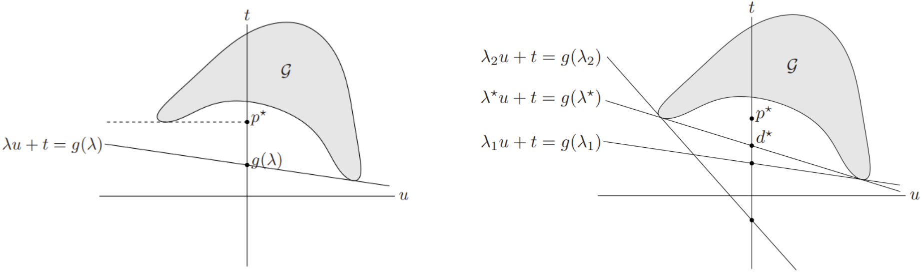
Slater's condition is that there exists a point $\tilde{x}$
$\in K$, such that all inequality constraints defining K are strict at $\tilde{x}$, $i.e. h_i(\tilde{x}) =$
0 for all $i = 1, 2, ..., p$ and for all $j = 1, 2, ..., m$. we have $f_j(\tilde{x}) < 0$.

- It guarantees strong duality for convex optimization problem.
- linear inequalities do not need to hold with strict inequality.

Geometric interpretation of duality

Consider problem with one constraint $f_1(x) \le 0$. Then the dual function is given as

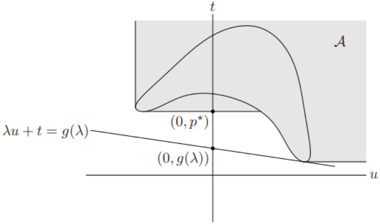
$$g(\lambda) = \inf_{(u,t)\in G} (t + \lambda u), \qquad \mathcal{G} = \{(f_1(x), f_0(x)) | x \in D\} \tag{5}$$



- Given λ, $\lambda u + t = (\lambda, 1)^T(u, t) \ge g(\lambda)$ holds for all $(u, t) \in \mathcal{G}$
- $g(\lambda) = \lambda u + t = (\lambda, 1)^T(u, t)$: non-vertical supporting hyperplane to $\mathcal{G}$
- hyperplane intersects t-axis at $t = g(\lambda)$

Geometric interpretation of duality: epigraph variation

$\mathcal{G}$ is replaced with $\mathcal{A} = \{(u,t) | f_1(x) \leq u, f_0(x) \leq t \text{ for some } x \in \mathcal{D}\}$



Strong duality holds if there is a non-vertical supporting hyperplane to $\mathcal{A}$ at $(0, p^\star)$.

- $p^\star = \inf\{t | (0,t) \in \mathcal{A}\}$. If $\lambda \geq 0$, $g(\lambda) = \inf\{(\lambda, 1)^T(u,t) \in \mathcal{A}\}$.
- If infimum is finite, $\lambda u + t \geq g(\lambda)$ defines a non-vertical supporting hyperplane to $\mathcal{A}$. Since $(0, p^\star) \in bd\mathcal{A}$, we have $p^\star = (\lambda, 1)^T(0, p^\star) \geq g(\lambda, \nu)$.
- Strong duality holds if and only if we have equality for some dual feasible (λ, ν), i.e., there exists a non-vertical supporting hyperplane to A at its boundary point $(0, p^*)$.

Optimality conditions

If strong duality holds, then x is primal optimal and (λ, ν) is dual optimal if:

1. $f_i(x) \leq 0$ for i = 1, ..., m and $h_j(x) = 0$ for j = 1, ..., p
2. $\lambda \succeq 0$
3. $f_0(x) = g(\lambda, \nu)$

conversely, these three conditions imply optimality of x, (λ, ν), and strong duality

Complementary slackness

- x* is primal optimal, $(\lambda^\star, \mu^\star)$ is dual optimal. Then $x^\star$ minimize $L(x, \lambda^\star, \mu^\star)$
- $\lambda_i{}^\star > 0 => f_i(x^\star) = 0$ or $f_i(x^\star) < 0 \Rightarrow \lambda_i{}^\star = 0$

Proof.

$f_0(x^\star) = g(\lambda^\star, \nu^\star) \leq \inf_x(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{j=1}^p \nu_j^\star h_j(x))$
$\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{j=1}^p \nu_j^\star h_j(x^\star)$
$\leq f_0(x^\star)$
Thus $\lambda_i^\star f_i(x^\star) = 0$ for i = 1, ..., m

$\square$

Optimality conditions

If strong duality holds, then x is primal optimal and (λ, ν) is dual optimal if:

1. $f_i(x) \leq 0$ for i = 1, ..., m and $h_j(x) = 0$ for j = 1, ..., p
2. $\lambda \succeq 0$
3. $\lambda_i f_i(x) = 0$ for i = 1, ..., m
4. x is minimizer of $\mathcal{L}(\cdot, \lambda, \nu)$

conversely, these four conditions imply optimality of x, (λ, ν), and strong duality if problem is convex and the functions f_i, h_j are differentiable, 4 can written as 4'. the gradient of the Lagrangian with respect to x vanishes :

$\partial_x L(x^\star, \lambda^\star, \nu^\star) = \nabla f_0(x) + \sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^p \nu_i^\star \nabla h_j(x^\star) = 0$
condition 1, 2, 3, 4' : Karush-Kuhn-Tucker (KKT) conditions

KKT conditions

Assume that the functions $f_0, ..., f_m, h_1, ...h_p$ are differentiable. $x^\star$ is primal optimal and $(\lambda^\star, \nu^\star)$ is dual optimal

1. primal feasibility : $f_i(x^\star) \leq 0$ for i = 1, ..., m and $h_i(x^\star) = 0$ i = 1, ..., p.
2. dual feasibility : $\lambda^\star \geq 0$
3. complementary slackness : $\lambda_i f_i(x) = 0$ for i = 1, ..., m
4. stationarity: $\partial_x L(x^\star, \lambda^\star, \nu^\star) = \nabla f_0(x) + \sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^p \nu_i^\star \nabla h_i(x^\star) = 0$

Convex problem with Slater qualification

recall the two implications of Slater's condition for a convex problem

- strong duality : $p^\star = d^\star$
- if optimal value is finite, dual optimum is attained : there exist dual λ, ν

hence, if problem is convex and Slater's constraint qualification holds :

- x is optimal if and only if there exist λ, ν such that 1, 2, 3, 4 satisfied
- if functions are differentiable, condition 4 can be replaced with $4'$