Mathematics for AI

Lecture 33

Convex Optimization, Duality examples, Suboptimality Check, KKT conditions

Convex Optimization (Abstract form)

K. N. Toosi
University of Technology

~~Convex~~ Convex Optimization

$$\left.\begin{array}{l}\min f(x) \quad \text{subject to} \quad x \in C \\ \qquad\qquad C \text{ is a convex set.}\end{array}\right\} \text{abstract form}$$

Convex Optimization (Standard form)

K. N. Toosi
University of Technology

Convex Optimization (standard form)

$$\min f(x) \quad \text{subject to} \quad g_i(x) \leq 0 \qquad i = 1, 2, \dots, p$$

$$Ax = b \qquad A \in \mathbb{R}^{q \times n}$$

$$b \in \mathbb{R}^q$$

f convex

g_i convex

for most convex optimization problems strong duality holds. (slater's conditions, etc.)

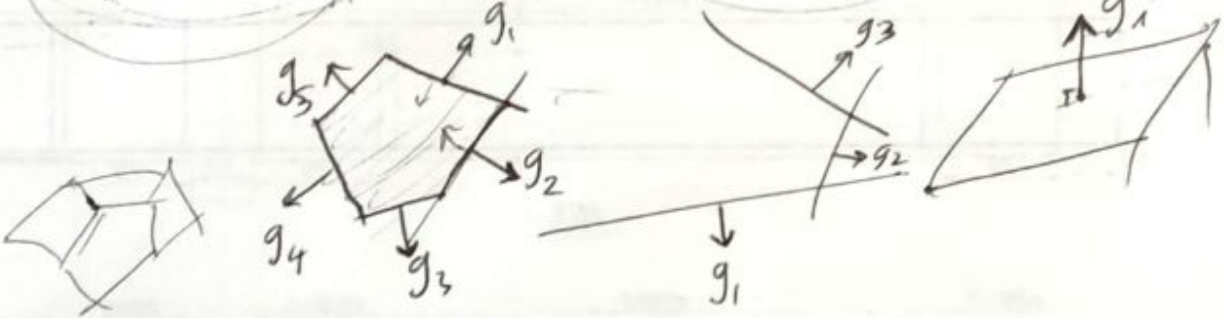
$$z^* = y^*$$

Linear Programming (LP)

Linear Program

$$\min_{x} \; a^T x + b$$
s.t.
$$G x \preceq \vec{c}$$
$$H x = d$$

$$\min_{x} \; a^T x + b$$
$$g_i^T x \leq c_i \qquad i = 1 \cdots q$$
$$h_i^T x = d_i \qquad i = 1 \cdots p$$



Different forms of a Linear Program

K. N. Toosi
University of Technology

$$\min_x \; a^T x + b \quad \text{s.t.} \quad Gx \leq c$$

$$Hx = d \;\equiv\; \begin{cases} Hx \leq d \\ Hx \geq d \end{cases}$$

$$\begin{cases} Hx \leq d \\ (-H)x \leq -d \end{cases}$$

$$\min_x \; a^T x + b \quad \text{s.t.} \quad Hx = d, \;\; x \geq 0$$

$$\min \; a^T x + b \quad \text{s.t.} \quad Hx = d, \;\; Gx \leq c, \;\; x \geq 0$$

$$\max_x \; a^T x + b \quad \text{s.t.} \quad Gx \leq c, \;\; Hx = d$$

Quadratic Programming

$$\min \; x^T H x + g^T x + p$$

$$\begin{cases} Gx \geq c \\ Ax = b \end{cases}$$

H: positive semi-definite

svm

Remember: Dual Problem

Remider: Dual Problem

دوگان

Primal	Dual problem
$y^* = \min_{x} \; f(x)$ subject to $g(x) \leq \vec{0}$, $h(x) = \vec{0}$	$z^* = \max_{\lambda, \mu} \; f^*(\lambda, \mu)$ subject to $\lambda \geq 0$

Lagrangian

$$L(x, \lambda, \mu) = f(x) + \lambda^T g(x) + \mu^T h(x)$$

$g(x) \leq 0,\ h(x) = 0,\ \lambda \geq 0 \Rightarrow L(x, \lambda, \mu) \leq f(x)$

$$f^*(\lambda, \mu) = \min_{x} L(x, \lambda, \mu)$$

$z^* \leq y^*$ weak duality / $y^* - z^*$ duality gap

if $z^* = y^* \Rightarrow$ strong duality hold $\Rightarrow$ duality gap $= 0$

Suboptimality check

K. N. Toosi
University of Technology

$z^* \leqslant y^*$ weak duality / $y^* - z^*$ duality gap

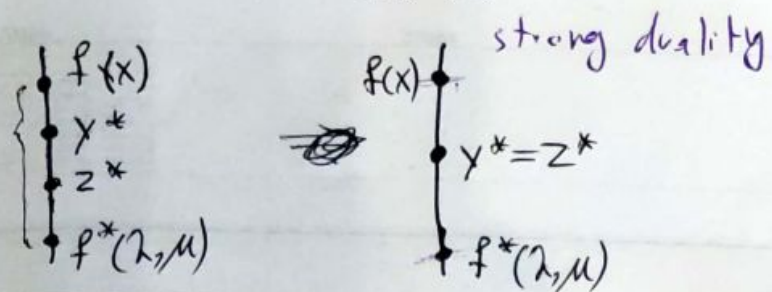
if $z^* = y^* \Rightarrow$ strong duality hold $\Rightarrow$ duality gap $= 0$

ε-suboptimal solution: $\tilde{x}$ $\quad g(\tilde{x}) \leqslant 0,\ h(\tilde{x}) = 0,\quad f(\tilde{x}) - y^* \leqslant \varepsilon$

$x, \mu, \lambda: g(x) \leqslant 0,\ h(x) = 0,\ \lambda \geqslant 0 \qquad f(x) \geqslant f^*(\lambda, \mu)$

stop when $\quad f(x) - y^* \leqslant f(x) - f^*(\lambda, \mu) \leqslant \varepsilon$

stop when $\ f(x) - f^*(\lambda, \mu) \leqslant \varepsilon$



Suboptimality

$z^* \leqslant y^*$ weak duality / $y^* - z^*$ duality gap

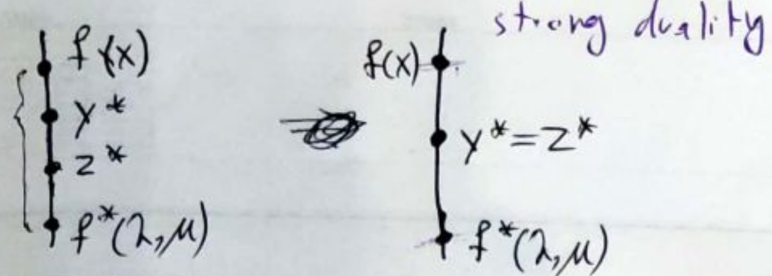
if $z^* = y^* \Rightarrow$ strong duality hold $\Rightarrow$ duality gap $= 0$

ε-suboptimal solution: $\tilde{x}$ $\quad g(\tilde{x}) \leqslant 0,\ h(\tilde{x}) = 0,\quad f(\tilde{x}) - y^* \leqslant \varepsilon$

x, μ, λ: $g(x) \leqslant 0,\ h(x) = 0,\ \lambda \geqslant 0 \qquad f(x) \geqslant f^*(\lambda, \mu)$

stop when $\qquad f(x) - y^* \leqslant f(x) - f^*(\lambda, \mu) \leqslant \varepsilon$

stop when $f(x) - f^*(\lambda, \mu) \leqslant \varepsilon$



Dual of a QP

K. N. Toosi
University of Technology

Dual for QP

MA 33

III

$$\min_x \tfrac{1}{2}x^T H x + g^T x \quad \text{s.t.} \quad Gx \leq c, \quad Ax = b$$

$$L(x, \lambda, \mu) = \tfrac{1}{2}x^T H x + g^T x + \lambda^T (Gx - c) + \mu^T (Ax - b)$$

$$f^*(\lambda, \mu) = \min_x L(x, \lambda, \mu)$$

$v^T x \overset{\nabla}{\Rightarrow} v$

$$L(x, \lambda, \mu) = \tfrac{1}{2}x^T H x + (g + G^T\lambda + A^T\mu)^T x - \lambda^T c - \mu^T b$$

$$\frac{\partial L}{\partial x} = Hx + g + G^T\lambda + A^T\mu = 0 \Rightarrow x = -H^{-1}(g + G^T\lambda + A^T\mu)$$

Dual of a QP (simple case: no equality constraints)

K. N. Toosi
University of Technology

حالت ساده ، قیود تساوی نداریم =>

$$f^*(\lambda) = \frac{1}{2}(g+G^T\lambda)^T H^{-1}(g+G^T\lambda)$$

$$-(g+G^T\lambda)^T H^{-1}(g+G^T\lambda) - \lambda^T c$$

$$= -\frac{1}{2}(g+G^T\lambda)^T H^{-1}(g+G^T\lambda)$$

$$f^*(\lambda) = -\frac{1}{2}\lambda^T G H^{-1} G^T \lambda - g^T H^{-1} G^T \lambda - \frac{1}{2} g^T H^{-1} g - c^T \lambda$$

dual problem

$$\max_{\lambda} \; -\frac{1}{2}\lambda^T \underbrace{G H^{-1} G^T}_{PD} \lambda - (g^T H^{-1} G^T + c^T)\lambda - \frac{1}{2} g^T H^{-1} g$$

$$\text{s.t.} \quad \lambda \geq 0$$

Dual of a QP (General Case)

$$\min \frac{1}{2} x^T H x + g^T x \quad \text{s.t.} \quad Gx \leqslant c \, , \quad Ax = b$$

$$L(x, \lambda, \mu) = \frac{1}{2} x^T H x + g^T x + \lambda^T (Gx - c) + \mu^T (Ax - b)$$

$$= \frac{1}{2} x^T H x + (g + G^T \lambda + A^T \mu)^T x - \lambda^T c - \mu^T b$$

$$\frac{\partial L}{\partial x} = Hx + (g + G^T \lambda + A^T \mu) = 0$$

$$\Rightarrow x^* = -H^{-1} (g + G^T \lambda + A^T \mu)$$

Dual of a QP (General Case)

$$\frac{\partial L}{\partial x} = Hx + (g + G^T\lambda + A^T\mu) = 0$$

$$\Rightarrow x^* = -H^{-1}(g + G^T\lambda + A^T\mu)$$

$$f^*(\lambda,\mu) = \min_x L(x,\lambda,\mu) = L(x^*,\lambda,\mu)$$

$$= \frac{1}{2}(g+G^T\lambda+A^T\mu)^T H^{-1} H H^{-1}(g+G^T\lambda+A^T\mu)$$

$$- (g+G^T\lambda+A^T\mu)^T H^{-1}(g+G^T\lambda+A^T\mu) - \lambda^T c - \mu^T b$$

$$= -\frac{1}{2}(g+G^T\lambda+A^T\mu)^T H^{-1}(g+G^T\lambda+A^T\mu) - \lambda^T c - \mu^T b$$

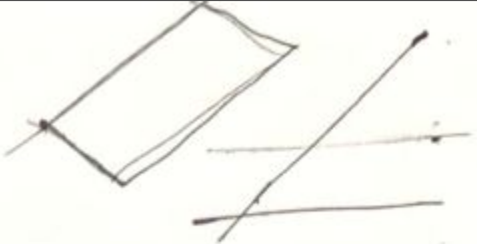
$$= -\frac{1}{2}\begin{bmatrix}\lambda\\ \mu\end{bmatrix}^T \underbrace{\begin{bmatrix}G^T\\ A^T\end{bmatrix}^T \underbrace{H^{-1}}_{PD} \begin{bmatrix}G^T\\ A^T\end{bmatrix}}_{\substack{PD\\ \text{or}\\ (PSD)}} \begin{bmatrix}\lambda\\ \mu\end{bmatrix} - \begin{bmatrix}c^T & b^T\end{bmatrix}\begin{bmatrix}\lambda\\ \mu\end{bmatrix} - \frac{1}{2}g^T H^{-1} g$$

Dual of an LP

$$\min_{x} \; c^T x$$

$$Gx \leq d$$

$$Ax = b$$



$$L(x,\lambda,\mu) = c^T x + \lambda^T (Gx - d) + \mu^T (Ax - b)$$

$$= (c + G^T \lambda + A^T \mu)^T x - \lambda^T d - \mu^T b$$

$$f^*(\lambda,\mu) = \min_{x} L(x,\lambda,\mu) = \begin{cases} -\infty & \text{if } c + G^T\lambda + A^T\mu \neq 0 \\ -\lambda^T d - \mu^T b & \text{if } c + G^T\lambda + A^T\mu = 0 \end{cases}$$

Dual of an LP

$$f^*(\lambda,\mu) = \min_x L(x,\lambda,\mu) = \begin{cases} -\infty & \text{if } c + G^T\lambda + A^T\mu \neq 0 \\ -\lambda^T d - \mu^T b & \text{if } c + G^T\lambda + A^T\mu = 0 \end{cases}$$

dual problem:

$$\max_{\lambda,\mu} \; f^*(\lambda,\mu) \quad \text{s.t.} \quad \lambda \succcurlyeq 0 \qquad \equiv \qquad \boxed{\begin{array}{ll} \max & -d^T\lambda - b^T\mu \\ \text{s.t.} & \lambda \succcurlyeq 0 \\ & G^T\lambda + A^T\mu + c = 0 \end{array}}$$

$$\begin{array}{ll} \min_x & c^T x \\ & Gx \preccurlyeq d \\ & Ax = b \end{array} \qquad\qquad \begin{array}{ll} \max & -d^T\lambda - b^T\mu \\ & \lambda \succcurlyeq 0 \\ & G^T\lambda + A^T\mu = -c \end{array}$$

Optimal primal and dual solutions

Assume that we have fund an optimal solution x^* for primal, and an optimal solution λ^*, μ^* for the dual problem

$$y^* = f(x^*)$$

$$z^* = f^*(\lambda^*, \mu^*)$$

Optimal primal and dual solutions

$$x^* = \underset{x}{\operatorname{argmin}} \; f(x) \quad \text{s.t.} \quad g(x) \leq 0 \,, \; h(x) = 0$$

$$\lambda^*, \mu^* = \underset{\lambda, \mu}{\operatorname{argmin}} \; f^*(\lambda, \mu) \quad \text{s.t.} \quad \lambda \succeq 0$$

Duality Gap $= 0 \Rightarrow f(x^*) = y^* = z^* = f^*(\lambda^*, \mu^*)$

Optimal primal and dual solutions

λ, μ

Duality Gap $= 0 \Rightarrow f(x^*) = y^* = z^* = f^*(\lambda^*, \mu^*)$

$$
\begin{aligned}
f(x^*) &= f^*(\lambda^*, \mu^*) \\
&= \min_{x} L(x, \lambda^*, \mu^*) \\
&\overset{\text{I}}{\leq} L(x^*, \lambda^*, \mu^*) \\
&= f(x^*) + \lambda^{*T} g(x^*) + \mu^{*T} h(x^*) \\
&= f(x^*) + \lambda^{*T} g(x^*) \\
&= f(x^*) + \sum_{i=1}^{q} \lambda_i^* g_i(x^*) \\
&\overset{\text{II}}{\leq} f(x^*)
\end{aligned}
$$

$h(x^*) = 0$

$\leq$

$g_i(x^*) \leq 0, \ \lambda_i^* \geq 0$

Optimal primal and dual solutions

K. N. Toosi
University of Technology

$$
\begin{aligned}
f(x^*) &= f^*(\lambda^*, \mu^*) \\
&= \min_{x} L(x, \lambda^*, \mu^*) \\
&\overset{\mathrm{I}}{\leq} L(x^*, \lambda^*, \mu^*) \\
&= f(x^*) + \lambda^{*T} g(x^*) + \mu^{*T} h(x^*) \\
&= f(x^*) + \lambda^{*T} g(x^*) \\
&= f(x^*) + \sum_{i=1}^{q} \lambda_i^* g_i(x^*) \\
&\overset{\mathrm{II}}{\leq} f(x^*)
\end{aligned}
$$

$h(x^*) = 0$

$g_i(x^*) \leq 0$, $\lambda_i^* \geq 0$

$\Longrightarrow$

I $\quad L(x^*, \lambda^*, \mu^*) = \min_{x} L(x, \lambda^*, \mu^*)$

$\Rightarrow$ find a solution to primal problem having a solution (λ^*, μ^*) to the dual problem

II $\quad \sum_{i=1}^{q} \lambda_i^* g_i(x^*) = 0 \xrightarrow[\lambda_i^* \geq 0,\ g_i(x) \leq 0]{} \lambda_i^* g_i(x^*) = 0$ for all $i = 1 - q$

Optimal primal and dual solutions

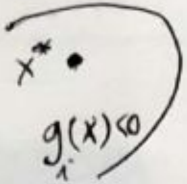
$\leq f(x^*)$

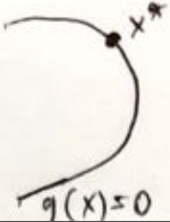
$$\Rightarrow \begin{cases} \text{I} \quad L(x^*, \lambda^*, \mu^*) = \min_x L(x, \lambda^*, \mu^*) \\ \text{II} \quad \sum_{i=1}^{q} \lambda_i^* g_i(x^*) = 0 \end{cases}$$

$\Rightarrow$ find a solution to primal problem having a solution (λ^*, μ^*) to the dual problem

$$\sum_{i=1}^{q} \lambda_i^* g_i(x^*) = 0 \xrightarrow[\substack{\lambda_i^* \geq 0 \\ g_i(x^*) \leq 0}]{} \lambda_i^* g_i(x^*) = 0$$

for all $i = 1 - q$





Find a primal solution from dual solutions

K. N. Toosi
University of Technology

$$L(x^*, \lambda^*, \mu^*) = \min_x L(x, \lambda^*, \mu^*)$$

$\Rightarrow$ find a solution to primal problem having a solution (λ^*, μ^*) to the dual problem

Necessary conditions for optimality

K. N. Toosi
University of Technology

$$\begin{cases} g(x^*) \geq 0 \\ h(x^*) = 0 \\ \lambda^* \geq 0 \\ \lambda_i^* \, g_i(x^*) = 0 \\ x^* = \arg\min L(x, \lambda^*, \mu^*) \end{cases}$$

Necessary conditions for optimality - differentiable case

for optimal x^*, λ^*, μ^* MA33 VI

$$g(x^*) \leq 0$$

$$h(x^*) = 0$$

$$\lambda^* \geq 0$$

$$\lambda_i \, g_i(x^*) = 0$$

$$x^* = \arg\min_x L(x, \lambda^*, \mu^*) = \arg\min_x f(x) + \lambda^{*T} g(x) + \mu^{*T} h(x)$$

$$= \arg\min_x f(x) + \sum \lambda_i^* g_i(x) + \sum \mu_i^* h_i(x)$$

f, g_i
h_i
differentiable

$$\nabla f(x^*) + \sum \lambda_i^* \nabla g_i(x^*) + \sum \mu_i^* \nabla h_i(x^*) = 0$$

Karush-Kuhn-Tucker (KKT) Conditions

- Necessary conditions for optimal x^*, λ^*, μ^*

$$g(x^*) \leqslant 0$$

$$h(x^*) = 0$$

$$\lambda^* \geqslant 0$$

$$\lambda_i^* g_i(x^*) = 0 \quad i = 1, 2, \ldots q$$

$$\nabla f(x^*) + \sum \lambda_i^* \nabla g_i(x^*) + \sum \mu_i^* \nabla h_i(x^*) = 0$$

Karush - Kuhn - Tucker (KKT) conditions

Sometimes it is also sufficient for optimality