CS257 Linear and Convex Optimization

Lecture 15

Bo Jiang

John Hopcroft Center for Computer Science
Shanghai Jiao Tong University

December 14, 2020

Recap: Lagrange Dual Function

For a general optimization problem (not necessarily convex),

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

The (Lagrange) dual function is

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x}} \ \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x}} \left(f(\boldsymbol{x}) + \sum_{i=1}^{k} \lambda_i h_i(\boldsymbol{x}) + \sum_{j=1}^{m} \mu_j g_j(\boldsymbol{x}) \right)$$

The variables $\boldsymbol{\lambda} \in \mathbb{R}^k, \boldsymbol{\mu} \in \mathbb{R}^m$ are called dual variables.

Properties

- For any $\boldsymbol{x} \in X$, $\boldsymbol{\lambda}$ and $\boldsymbol{\mu} \ge \boldsymbol{0}$, $f(\boldsymbol{x}) \ge f^* \ge \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \ge \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$
- The dual function is always concave, whether (P) is convex or not.

Lagrange Dual Problem

To find the best lower bound given by the dual function

$$f^* \geq \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

solve the (Lagrange) dual problem associated with the primal problem (P),

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned} \tag{D}$$

The dual problem (D) is always convex, whether or not (P) is convex.

$(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is dual feasible if $\boldsymbol{\mu} \geq \mathbf{0}$ and $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty$.

Note. The domain of a convex function f is $\operatorname{dom} f = \{\boldsymbol{x} : f(\boldsymbol{x}) < +\infty\}$, while the domain of a concave function f is $\operatorname{dom} f = \{\boldsymbol{x} : f(\boldsymbol{x}) > -\infty\}$. Thus the condition $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty$ just means $(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \operatorname{dom} \phi$.

Example

The dual problem of the following LP

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{G}\boldsymbol{x} \leq \boldsymbol{h} \end{aligned}$$

is

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda},\boldsymbol{\mu}) = \begin{cases} -\boldsymbol{\lambda}^T\boldsymbol{b} - \boldsymbol{\mu}^T\boldsymbol{h}, & \text{if } -\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c} \\ -\infty, & \text{otherwise} \end{cases} \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned}$$

$(\boldsymbol{\lambda},\boldsymbol{\mu})$ is dual feasible if $\boldsymbol{\mu} \geq \mathbf{0}$ and $-\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c}$, which just means it is feasible for the dual LP,

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & \psi(\boldsymbol{\lambda},\boldsymbol{\mu}) = -\boldsymbol{\lambda}^T\boldsymbol{b} - \boldsymbol{\mu}^T\boldsymbol{h} \\ \text{s.t.} \quad & -\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c} \\ & \boldsymbol{\mu} \geq \mathbf{0} \end{aligned}$$

Example

The dual problem of the following problem

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

is

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\frac{1}{4}\|\boldsymbol{\mu} - \boldsymbol{A}^T\boldsymbol{\lambda}\|^2 - \boldsymbol{b}^T\boldsymbol{\lambda} \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

$(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is dual feasible if $\boldsymbol{\mu} \geq \boldsymbol{0}$.

Weak and Strong Duality

Denote by f^* and ϕ^* the primal and dual optimal values, i.e.

$$f^* = \inf_{\boldsymbol{x} \in X} f(\boldsymbol{x}), \qquad \phi^* = \sup_{\boldsymbol{\lambda}, \boldsymbol{\mu} : \boldsymbol{\mu} \geq \boldsymbol{0}} \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

Weak duality: $\quad f^* \geq \phi^*$

- always holds.

  Proof. Recall $f^* \geq \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$ for any $\boldsymbol{\lambda}$ and any $\boldsymbol{\mu} \geq \boldsymbol{0}$. Weak duality follows by maximizing over $\boldsymbol{\lambda}$ and $\boldsymbol{\mu} \geq \boldsymbol{0}$.
- $f^* - \phi^*$ is called the (optimal) duality gap of the problem.

Strong duality: $\quad f^* = \phi^*$

- does not hold in general.
- typically holds for convex problems under various conditions known as constraint qualifications, e.g. Slater's condition.
- may also hold for nonconvex problems.
- can solve the dual problem instead if it is easier than the primal.

Example

Recall the following pair of primal and dual LP,

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^2} \quad & f(\boldsymbol{x}) = x_1 + 2x_2 \\ \text{s.t.} \quad & 2x_1 + x_2 \geq 2 \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

$$\begin{aligned} \max_{\boldsymbol{\mu}\in\mathbb{R}^3} \quad & \psi(\boldsymbol{\mu}) = 2\mu_1 \\ \text{s.t.} \quad & 2\mu_1 + \mu_2 = 1 \\ & \mu_1 + \mu_3 = 2 \\ & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

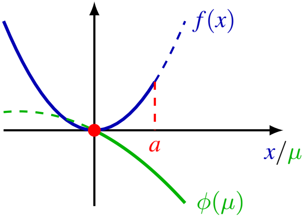
- The primal LP can be solved graphically with $f^* = f(1,0) = 1$.
- The dual is equivalent to

$$\begin{aligned} \max_{\mu_1} \quad & 2\mu_1 \\ \text{s.t.} \quad & 2\mu_1 \leq 1 \\ & \mu_1 \leq 2 \\ & \mu_1 \geq 0 \end{aligned}$$

So $\phi^* = \phi(\frac{1}{2}, 0, \frac{3}{2}) = 1 = f^*$, strong duality holds.

Example

$$\min_{x\in\mathbb{R}} \quad f(x) = x^2$$
$$\text{s.t.} \quad x \le a$$



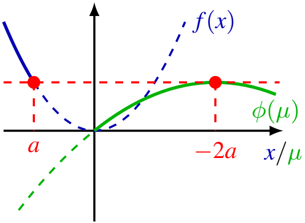
Case 1. $a \ge 0$

The dual function is

$$\phi(\mu) = \inf_x [x^2 + \mu(x-a)] = -\frac{\mu^2}{4} - a\mu$$

The dual problem is

$$\max_{\mu\in\mathbb{R}} \quad \phi(\mu) = -\frac{\mu^2}{4} - a\mu$$
$$\text{s.t.} \quad \mu \ge 0$$



Case 2. $a \le 0$

The primal and dual optimal values are

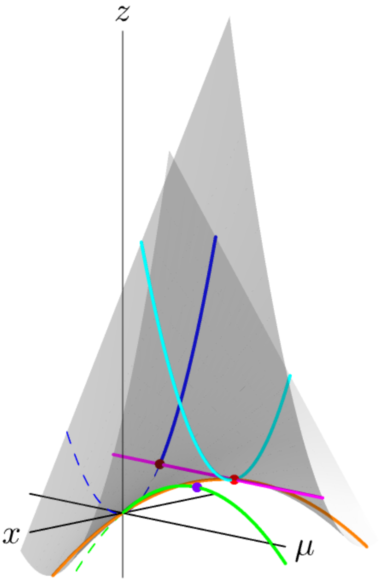
1. If $a \ge 0$, $f^* = f(0) = \phi^* = \phi(0) = 0$
2. If $a \le 0$, $f^* = f(a) = \phi^* = \phi(-2a) = a^2$

Strong duality holds in both cases.

Example (cont'd)

Assume $a < 0$.

- surface: $\mathcal{L}(x, \mu) = x^2 + \mu(x - a)$
- **blue** curve: $x \mapsto (x, 0, f(x))$
- **green** curve: $\mu \mapsto (0, \mu, \phi(\mu))$
- **orange** curve: $\mu \mapsto (x^*(\mu), \mu, \phi(\mu))$.
  Note $x^*(\mu) = \arg\min_x \mathcal{L}(x, \mu) = -\frac{\mu}{2}$.
- **cyan** curve: $x \mapsto (x, \mu^*, \mathcal{L}(x, \mu^*))$
- **magenta** curve: $\mu \mapsto (x^*, \mu, \mathcal{L}(x^*, \mu))$
- **red** dot: (x^*, μ^*, f^*)
- **brown** dot: $(x^*, 0, f^*)$
- **purple** dot: $(0, \mu^*, f^*)$



(x^*, μ^*) is a saddle point of $\mathcal{L}$: for $\mu \geq 0$,

$$\mathcal{L}(x^*, \mu) \leq \mathcal{L}(x^*, \mu^*) \leq \mathcal{L}(x, \mu^*)$$

Example

Consider

$$\min_{x\in\mathbb{R}} \quad f(x) = x^3$$
$$\text{s.t.} \quad x \geq 0$$

The optimal value if $f^* = f(0) = 0$.

The dual function is

$$\phi(\mu) = \inf_x [x^3 - \mu x] = -\infty$$

so the dual optimal value is

$$\phi^* = \sup_{\mu\geq 0} \phi(\mu) = -\infty$$

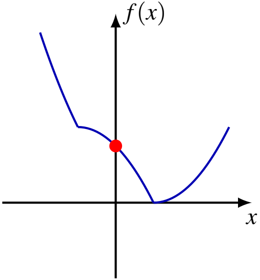
The duality gap is infinite. In particular, strong duality does not hold.

Example

Consider

$$\min_{x\in\mathbb{R}} \quad f(x) = \begin{cases} -x^2 - x + \frac{3}{4}, & |x| \le \frac{1}{2} \\ x^2 - x + \frac{1}{4}, & |x| \ge \frac{1}{2} \end{cases}$$

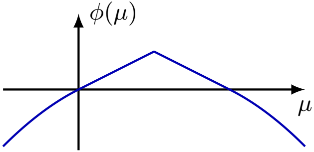
$$\text{s.t.} \quad x \le 0$$



The primal optimal value is $f^* = f(0) = \frac{3}{4}$.

The dual function is

$$\phi(\mu) = \inf_x [f(x) + \mu x] = \begin{cases} \frac{1-|\mu-1|}{2}, & |\mu - 1| \le 1 \\ \frac{1-(\mu-1)^2}{4}, & |\mu - 1| \ge 1 \end{cases}$$



The dual optimal value is $\phi^* = \phi(1) = \frac{1}{2}$.

The duality gap is $f^* - \phi^* = \frac{1}{4}$.

Example (cont'd)

To compute the dual function, note

$$\mathcal{L}(x,\mu) = f(x) + \mu x = \begin{cases} -x^2 + (\mu - 1)x + \frac{3}{4}, & |x| \le \frac{1}{2} \\ x^2 + (\mu - 1)x + \frac{1}{4}, & |x| > \frac{1}{2} \end{cases}$$

Since $y = -x^2 + (\mu - 1)x + \frac{3}{4}$ is a parabola opening down,

$$\phi_1(\mu) = \inf_{|x| \le \frac{1}{2}} \mathcal{L}(x,\mu) = \min\left\{\mathcal{L}(\frac{1}{2},\mu), \mathcal{L}(-\frac{1}{2},\mu)\right\} = \frac{1 - |\mu - 1|}{2}$$

Since $y = x^2 + (\mu - 1)x + \frac{1}{4}$ is a parabola opening up,

$$\phi_2(\mu) = \inf_{|x| \ge \frac{1}{2}} \mathcal{L}(x,\mu) = \begin{cases} \mathcal{L}(\frac{1-\mu}{2},\mu) = \frac{1-(\mu-1)^2}{4}, & |\mu - 1| \ge 1 \\ \min\left\{\mathcal{L}(\frac{1}{2},\mu), \mathcal{L}(-\frac{1}{2},\mu)\right\} = \frac{1-|\mu-1|}{2}, & |\mu - 1| \le 1 \end{cases}$$

Thus

$$\phi(\mu) = \min\{\phi_1(\mu), \phi_2(\mu)\} = \phi_2(\mu)$$

Example (cont'd)

By definition of dual function,

$$\phi(\mu) = \inf_x [f(x) + \mu x] \leq f(x) + \mu x$$

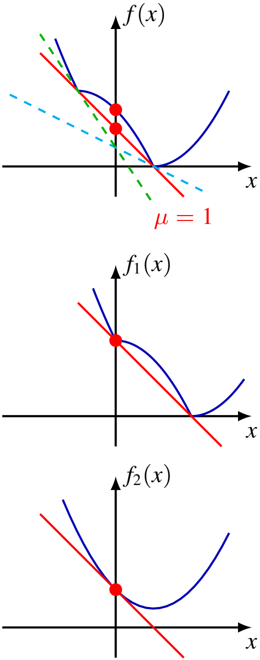
Rearranging,

$$\ell(x) \triangleq -\mu x + \phi(\mu) \leq f(x)$$

Note $\ell(x)$ is a line with slope $-\mu$ and intercept $\phi(\mu)$ that lies below the graph of f.

The dual optimal value ϕ^* is the largest intercept of such lines. We can see pictorially there is a gap.

This also give us intuition about why strong duality may hold for nonconvex problem, and why it usually holds for convex problems.

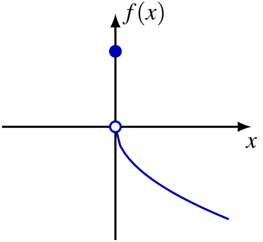


Example

Consider

$$\min_{x \in \mathbb{R}} \quad f(x) = \begin{cases} -\sqrt{x}, & x > 0 \\ 1 & x = 0 \\ +\infty, & x < 0 \end{cases}$$

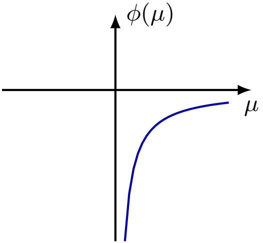
$$\text{s.t.} \quad x \le 0$$



The primal optimal value is $f^* = f(0) = 1$.

The dual function is

$$\phi(\mu) = \inf_x [f(x) + \mu x] = \begin{cases} -\frac{1}{4\mu}, & \mu > 0 \\ -\infty, & \mu \le 0 \end{cases}$$



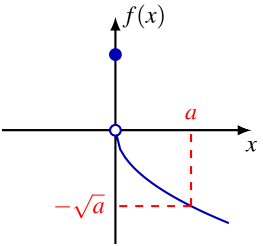
The dual optimal value is $\phi^* = 0$, which is not attainable.

This is a convex problem with nonzero duality gap $f^* - \phi^* = 1$, a nontypical case.

Example

$$\min_{x\in\mathbb{R}} \quad f(x) = \begin{cases} -\sqrt{x}, & x > 0 \\ 1 & x = 0 \\ +\infty, & x < 0 \end{cases}$$

$$\text{s.t.} \quad x \le a$$

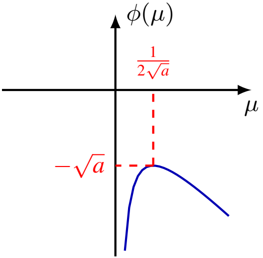


where $a > 0$.

The primal optimal value is $f^* = f(a) = -\sqrt{a}$.

The dual function is

$$\phi(\mu) = \inf_x [f(x) + \mu(x-a)] = \begin{cases} -\frac{1}{4\mu} - a\mu, & \mu > 0 \\ -\infty, & \mu \le 0 \end{cases}$$



The dual optimal value is $\phi^* = \phi(\frac{1}{2\sqrt{a}}) = -\sqrt{a}$

Strong duality holds in this case.

Slater's Condition for Convex Problems

Consider a convex problem,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_j(\boldsymbol{x}) \le 0,\ j = 1, 2, \ldots, m \\ & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0} \end{aligned} \tag{CP}$$

with domain $D = \operatorname{dom} f \cap (\bigcap_{i=1}^{m} \operatorname{dom} g_j)$.

Slater's condition. The above problem is strictly feasible, i.e.

$$\exists \boldsymbol{x} \in \operatorname{int} D^{1} \quad \text{s.t.} \quad g_j(\boldsymbol{x}) < 0 \text{ for } i = 1, 2, \ldots, m, \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

Refined Slater's condition. If some g_j are affine, the requirement $g_j(\boldsymbol{x}) < 0$ can be relaxed to feasibility $g_j(\boldsymbol{x}) \le 0$ for those g_j.

Slater's Theorem. Strong duality holds for (CP) under (refined) Slater's condition. Furthermore, if $\phi^* > -\infty$, it is attained by some $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$.

[1] $\operatorname{int} D$ stands for the interior of D. $\boldsymbol{x} \in \operatorname{int} D$ if there exists $\delta > 0$ s.t. $B(\boldsymbol{x}, \delta) \subset D$. Again we focus on the case $D = \mathbb{R}^n$, so the requirement $\boldsymbol{x} \in \operatorname{int} D$ is always satisfied.

Example

$$\min_{x \in \mathbb{R}} \quad f(x) = \begin{cases} -\sqrt{x}, & x > 0 \\ 1 & x = 0 \\ +\infty, & x < 0 \end{cases}$$

$$\text{s.t.} \quad x \leq a$$

is a convex problem with domain $D = [0, \infty)$. Note $\operatorname{int} D = (0, \infty)$.

- If $a > 0$, Slater's condition is satisfied, e.g. $\frac{a}{2} \in \operatorname{int} D$ and $\frac{a}{2} < a$, so strong duality must hold.
- If $a = 0$, no point in $\operatorname{int} D$ is feasible. Slater's Theorem is not applicable[2] , and it turns out that strong duality does not hold.

[2] Slater's condition is only a sufficient condition for strong duality. It is not necessary.

Example: Strong Duality for LP

$$\begin{array}{ll} \min_{\boldsymbol{x}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{G}\boldsymbol{x} \le \boldsymbol{h} \end{array} \xrightarrow{\text{dual LP}} \begin{array}{ll} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} & -\boldsymbol{b}^T\boldsymbol{\lambda} - \boldsymbol{h}^T\boldsymbol{\mu} \\ \text{s.t.} & -\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c} \\ & \boldsymbol{\mu} \ge \boldsymbol{0} \end{array}$$

equivalence $\Updownarrow$ $\Updownarrow$ equivalence

$$\begin{array}{ll} \max_{\boldsymbol{x},\boldsymbol{y}} & -\boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{G}\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{h} \\ & \boldsymbol{y} \ge \boldsymbol{0} \end{array} \xleftarrow{\text{dual LP}} \begin{array}{ll} \min_{\boldsymbol{\lambda},\boldsymbol{\mu}} & \boldsymbol{b}^T\boldsymbol{\lambda} + \boldsymbol{h}^T\boldsymbol{\mu} \\ \text{s.t.} & -\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c} \\ & -\boldsymbol{\mu} \le \boldsymbol{0} \end{array}$$

- Essentially, dual of dual is primal.
- By refined Slater's condition, strong duality holds if either the primal or the dual is feasible.
- When both primal and dual are feasible, $f^* = \phi^*$ are finite and they are both attained.

Example: Strong Duality for LP (cont'd)

There are four possibilities

1. Primal feasible, dual feasible, $-\infty < \phi^* = f^* < +\infty$
2. Primal feasible, dual infeasible, $f^* = \phi^* = -\infty$

$$
\begin{aligned}
\min \quad & x_1 - 2x_2 \\
\text{s.t.} \quad & x_1 - x_2 = -1 \\
& x_1, x_2 \geq 0
\end{aligned}
\qquad\qquad
\begin{aligned}
\max \quad & \lambda \\
\text{s.t.} \quad & \lambda + \mu_1 = 1 \\
& -\lambda + \mu_2 = -2 \\
& \mu_1, \mu_2 \geq 0
\end{aligned}
$$

3. Primal infeasible, dual feasible, $f^* = \phi^* = +\infty$
4. Primal infeasible, dual infeasible, $f^* = +\infty, \phi^* = -\infty$

$$
\begin{aligned}
\min \quad & x_1 - 2x_2 \\
\text{s.t.} \quad & x_1 - x_2 \leq 1 \\
& -x_1 + x_2 \leq -2
\end{aligned}
\qquad\qquad
\begin{aligned}
\max \quad & -\mu_1 + 2\mu_2 \\
\text{s.t.} \quad & -\mu_1 + \mu_2 = 1 \\
& \mu_1 - \mu_2 = -2 \\
& \mu_1, \mu_2 \geq 0
\end{aligned}
$$

Note. No duality gap in Case 2 and Case 3, but $f^* - \phi^*$ is undefined.

Example: Dual Formulation of SVM

Recall the primal formulation of SVM,

$$\begin{aligned}
\min_{\boldsymbol{w},b,\boldsymbol{\xi}} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\mathbf{1}^T\boldsymbol{\xi} \\
\text{s. t.} \quad & y_i(\boldsymbol{x}_i^T\boldsymbol{w} + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, n \\
& \boldsymbol{\xi} \geq \mathbf{0}
\end{aligned}$$

where $C > 0$ is a hyperparameter, and $\mathbf{1}$ is the vector of all 1's.

- convex problem with affine constraints.
- always feasible. Indeed, given any $\boldsymbol{w}, b$,

$$\xi_i = [1 - y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)]^+, \quad i = 1, 2, \ldots, n$$

  yields a feasible solution $(\boldsymbol{w}, b, \boldsymbol{\xi})$, where $(x)^+ = \max\{x, 0\}$.
- strong duality holds by refined Slater's condition
- can solve the dual problem instead, which turns out to be useful!

Example: Dual Formulation of SVM (cont'd)

The Lagrangian is

$$\begin{aligned}\mathcal{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\mu}, \boldsymbol{\alpha}) &= \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\mathbf{1}^T\boldsymbol{\xi} + \sum_{i=1}^{n} \mu_i[1 - \xi_i - y_i(\boldsymbol{x}_i^T\boldsymbol{w} + b)] - \boldsymbol{\alpha}^T\boldsymbol{\xi} \\ &= \frac{1}{2}\|\boldsymbol{w}\|_2^2 - \left(\sum_{i=1}^{n} y_i\mu_i\boldsymbol{x}_i\right)^T \boldsymbol{w} - \boldsymbol{\mu}^T\boldsymbol{y}b + (C\mathbf{1} - \boldsymbol{\mu} - \boldsymbol{\alpha})^T\boldsymbol{\xi} + \mathbf{1}^T\boldsymbol{\mu}\end{aligned}$$

Minimizing over $\boldsymbol{w}, b, \boldsymbol{\xi}$ yields the dual function ($\boldsymbol{w} = \sum_{i=1}^{n} y_i\mu_i\boldsymbol{x}_i$),

$$\phi(\boldsymbol{\mu}, \boldsymbol{\alpha}) = \begin{cases} \mathbf{1}^T\boldsymbol{\mu} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mu_i\mu_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j, & \text{if } \boldsymbol{\mu}^T\boldsymbol{y} = 0, C\mathbf{1} - \boldsymbol{\mu} - \boldsymbol{\alpha} = \mathbf{0} \\ -\infty, & \text{otherwise} \end{cases}$$

The dual problem is

$$\begin{aligned} \max_{\boldsymbol{\mu}, \boldsymbol{\alpha}} \quad & \phi(\boldsymbol{\mu}, \boldsymbol{\alpha}) \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \mathbf{0},\ \boldsymbol{\alpha} \geq \mathbf{0} \end{aligned}$$

Example: Dual Formulation of SVM (cont'd)

Making the constraints explicit, we obtain the equivalent problem,

$$\begin{aligned}
\max_{\boldsymbol{\mu},\boldsymbol{\alpha}} \quad & \mathbf{1}^T\boldsymbol{\mu} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mu_i\mu_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j \\
\mathrm{s.\,t.} \quad & \boldsymbol{\mu}^T\boldsymbol{y} = 0 \\
& \boldsymbol{\mu} + \boldsymbol{\alpha} = C\mathbf{1} \\
& \boldsymbol{\mu} \geq \mathbf{0},\ \boldsymbol{\alpha} \geq \mathbf{0}
\end{aligned}$$

Eliminating $\boldsymbol{\alpha}$, we obtain the following dual formulation of SVM,

$$\begin{aligned}
\max_{\boldsymbol{\mu}} \quad & \mathbf{1}^T\boldsymbol{\mu} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mu_i\mu_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j \\
\mathrm{s.\,t.} \quad & \boldsymbol{\mu}^T\boldsymbol{y} = 0 \\
& \mathbf{0} \leq \boldsymbol{\mu} \leq C\mathbf{1}
\end{aligned}$$

Can be solved by specialized algorithms called Sequential Minimal Optimization (SMO). Also amenable to further generalization using the kernel trick that replaces $\boldsymbol{x}_i^T\boldsymbol{x}_j$ by a kernel (function) $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

Duality Gap

Given primal feasible $\boldsymbol{x}$ and dual feasible $(\boldsymbol{\lambda}, \boldsymbol{\mu})$, the difference

$$f(\boldsymbol{x}) - \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

is called the duality gap associated with $\boldsymbol{x}$ and $(\boldsymbol{\lambda}, \boldsymbol{\mu})$.

Note

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \le \phi^* \le f^* \le f(\boldsymbol{x})$$

If the duality gap is zero, i.e. $f(\boldsymbol{x}) = \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$, then all inequalities become equalities, so $\boldsymbol{x}$ is primal optimal, and $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is dual optimal.

If the gap $f(\boldsymbol{x}) - \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \le \epsilon$, then the dual solution $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ serves as a proof or certificate that $\boldsymbol{x}$ is ϵ-suboptimal,

$$f(\boldsymbol{x}) - f^* \le f(\boldsymbol{x}) - \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \le \epsilon$$

When strong duality holds, this can serve as a stopping criterion in an iterative algorithm, i.e. stop when $f(\boldsymbol{x}) - \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \le \epsilon$ for some $(\boldsymbol{\lambda}, \boldsymbol{\mu})$.

Strong Duality and KKT for Convex Problems

Consider a differentiable convex problem and its dual,

$$
\begin{aligned}
\min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\
\text{s.t.} \quad & \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0} \\
& \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \mathbf{0}
\end{aligned}
\qquad\Bigg|\qquad
\begin{aligned}
\max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \\
\text{s.t.} \quad & \boldsymbol{\mu} \geq \mathbf{0}
\end{aligned}
$$

KKT conditions hold at $\boldsymbol{x}^*$ with Lagrange multipliers $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$, i.e.

1. (primal feasibility) $\boldsymbol{h}(\boldsymbol{x}^*) = \mathbf{0}$, $\boldsymbol{g}(\boldsymbol{x}^*) \leq \mathbf{0}$
2. (dual feasibility) $\boldsymbol{\mu}^* \geq \mathbf{0}$
3. (stationarity) $\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}$
4. (complementary slackness) $\mu_j^* g_j(\boldsymbol{x}^*) = 0, j = 1, 2, \ldots, m$

if and only if all the following conditions hold,

1. strong duality holds, i.e. $f^* = \phi^*$
2. $\boldsymbol{x}^*$ is a primal optimal solution, i.e. $f^* = f(\boldsymbol{x}^*)$
3. $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a dual optimal solution, i.e. $\phi^* = \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$

Proof of Necessity

Assume KKT holds at $\boldsymbol{x}^*$ with Lagrange multipliers $\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*$.

- Since $\boldsymbol{\mu}^* \geq \mathbf{0}$, $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\boldsymbol{x}) + (\boldsymbol{\lambda}^*)^T \boldsymbol{h}(\boldsymbol{x}) + (\boldsymbol{\mu}^*)^T \boldsymbol{g}(\boldsymbol{x})$ is convex in $\boldsymbol{x}$.
- The stationarity condition $\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}$ implies $\boldsymbol{x}^*$ is a global minimum of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, i.e.

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$$

- By primal feasibility and complementary slackness,

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\boldsymbol{x}^*) + (\boldsymbol{\lambda}^*)^T \underbrace{\boldsymbol{h}(\boldsymbol{x}^*)}_{=\mathbf{0}} + \underbrace{(\boldsymbol{\mu}^*)^T \boldsymbol{g}(\boldsymbol{x}^*)}_{=\mathbf{0}} = f(\boldsymbol{x}^*)$$

  so

$$f(\boldsymbol{x}^*) = \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$$

- By the discussion on slide 22, $\boldsymbol{x}^*$ is primal optimal, $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is dual optimal and strong duality holds.

Proof of Sufficiency

Assume strong duality holds, $\boldsymbol{x}^*$ is primal optimal, and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is dual optimal. We only need to show the stationarity condition and the complementary slackness condition.

$$
\begin{aligned}
f^* &= \phi^* && \text{(strong duality)} \\
&= \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) && \text{(dual optimality of } (\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)) \\
&= \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) && \text{(definition of dual function)} \\
&\le \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) && \text{(definition of infimum)} \\
&= f(\boldsymbol{x}^*) + (\boldsymbol{\lambda}^*)^T \underbrace{\boldsymbol{h}(\boldsymbol{x}^*)}_{=\mathbf{0}} + (\underbrace{\boldsymbol{\mu}^*}_{\ge \mathbf{0}})^T \underbrace{\boldsymbol{g}(\boldsymbol{x}^*)}_{\le \mathbf{0}} \\
&\le f(\boldsymbol{x}^*) && \text{(primal and dual feasibility of } \boldsymbol{x}^*,\ \boldsymbol{\mu}^*) \\
&= f^* && \text{(primal optimality of } \boldsymbol{x}^*)
\end{aligned}
$$

So both inequality holds with equality. The first implies $\boldsymbol{x}^*$ is a minimum of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, so $\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}$. The second implies $(\boldsymbol{\mu}^*)^T \boldsymbol{g}(\boldsymbol{x}^*) = \mathbf{0}$, so $\mu_j g_j(\boldsymbol{x}^*) = 0$ for $j = 1, 2, \ldots, m$.