COM S 578X: Optimization for Machine Learning

Lecture Note 4: Duality

Jia (Kevin) Liu

Assistant Professor
Department of Computer Science
Iowa State University, Ames, Iowa, USA

Fall 2019

Outline

In this lecture:

- Lagrange dual problem
- Weak and strong duality
- Geometric interpretation
- Examples in machine learning

The Lagrangian

Standard optimization problem (may or may not be convex):

Convex Optimization if:

f convex.

g_i convex, $\forall i$

h_j affine, $\forall j$

$\Rightarrow$ P is convex opt.

dual var. or Lagrange multipliers.

$$
\begin{array}{lll}
\text{Minimize} & f(\mathbf{x}) & \\
\text{subject to} & g_i(\mathbf{x}) \le 0, & i = 1, \ldots, m \leftarrow u_i \ge 0, \forall i \\
 & h_j(\mathbf{x}) = 0, & j = 1, \ldots, p \leftarrow v_j \in \mathbb{R}, \forall j
\end{array}
$$

$\underline{v} \in \mathbb{R}^p$ $\quad$ $\underline{u} \in \mathbb{R}^m_+$

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

$\to h_j(\underline{x}) \le 0, \ -h_j(\underline{x}) \le 0.$

Lagrangian: L: $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, with $\mathrm{dom}(L) = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$:

$$L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = f(\mathbf{x}) + \sum_{i=1}^{m} u_i g_i(\mathbf{x}) + \sum_{j=1}^{p} v_j h_j(\mathbf{x})$$

- weighted sum of objective and constraint functions
- $u_i \ge 0$ is dual variable (Lagrangian multiplier) associated with $g_i(\mathbf{x}) \le 0$
- $v_j \in \mathbb{R}$ is dual variable (Lagrangian multiplier) associated with $h_j(\mathbf{x}) = 0$

Lagrangian Dual Function

Lagrangian dual function: $\Theta : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$:

$$\Theta(\mathbf{u}, \mathbf{v}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \mathbf{u}, \mathbf{v})$$

affine fn w.r.t. u, v.

$$= \inf_{\mathbf{x} \in \mathcal{D}} \left(f(\overset{\tilde{x}}{\mathbf{x}}) + \underbrace{\sum_{i=1}^{m} \underbrace{\overset{\geq 0}{u_i} \overset{\leq 0}{g_i(\mathbf{x})}}_{\leq 0} + \sum_{j=1}^{p} \underbrace{\overset{uc}{\mathbf{v}_j} \overset{=0}{h_j(\mathbf{x})}}_{=0}}_{\leq 0} \right)$$

$\Theta(\mathbf{u}, \mathbf{v})$ is concave in $\mathbf{u}, \mathbf{v}$, can be $-\infty$ for some $\mathbf{u}, \mathbf{v}$ b/c it's a pt-wise minimum of affine fns w.r.t. u, v.

Lower bound property: If $\mathbf{u} \geq 0$, then $\Theta(\mathbf{u}, \mathbf{v}) \leq p^*$

opt. value of primal prob.

Proof.

If $\tilde{\mathbf{x}} \in \mathcal{D}$ and $\mathbf{u} \geq 0$, then

primal feas. domain

b/c + "≤0"

def. of "inf"

$$f(\tilde{\mathbf{x}}) \geq L(\tilde{\mathbf{x}}, \mathbf{u}, \mathbf{v}) \geq \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = \Theta(\mathbf{u}, \mathbf{v}).$$

def. of Lagrangian dual

Since this holds for any $\tilde{\mathbf{x}} \in \mathcal{D}$, minimizing over all $\tilde{\mathbf{x}}$ yields $p^* \geq \Theta(\mathbf{u}, \mathbf{v})$. □

achieved by x^*

Example: Least-norm solution of linear equations

$$\begin{array}{ll} \text{Minimize} & \mathbf{x}^\top \mathbf{x} \\ \text{subject to} & \underbrace{\mathbf{A}\mathbf{x} = \mathbf{b}}_{\text{convex}} \leftarrow \underline{v} \in \mathbb{R}^n \end{array}$$

Dual function:

- Lagrangian is: $L(\mathbf{x}, \mathbf{v}) = \mathbf{x}^\top \mathbf{x} + \mathbf{v}^\top (\mathbf{A}\mathbf{x} - \mathbf{b})$
- To minimize $L(\mathbf{x}, \mathbf{v})$ over $\mathbf{x}$, set gradient equal to zero $\min_{\underline{x} \in \mathbb{R}^d} \underline{x}^\top \underline{x} + \underline{v}^\top (\underline{A}\underline{x} - \underline{b})$

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{v}) = 2\mathbf{x} + \mathbf{A}^\top \mathbf{v} = 0 \quad \Longrightarrow \quad \mathbf{x} = -\frac{1}{2}\mathbf{A}^\top \mathbf{v}$$

- Plug it in L to obtain Θ:

$$\Theta(\mathbf{v}) = L(-\frac{1}{2}\mathbf{A}^\top \mathbf{v}, \mathbf{v}) = \underbrace{\underbrace{-\frac{1}{4}\mathbf{v}^\top \overset{\text{PSD}}{\mathbf{A}\mathbf{A}^\top} \mathbf{v}}_{\leq 0} - \mathbf{b}^\top \mathbf{v}}_{\leq -\underline{b}^\top \underline{v}},$$

which is clearly a concave function of $\mathbf{v}$.

Lower bound property: $p^* \geq -\frac{1}{4}\mathbf{v}^\top \mathbf{A}\mathbf{A}^\top \mathbf{v} - \mathbf{b}^\top \mathbf{v}$ for all $\mathbf{v}$

$\geq -\underline{b}^\top \underline{v}$ (largest UB).

Example: Linear Programming

$$\begin{aligned}
&\text{Minimize} && \mathbf{c}^\top \mathbf{x} \\
&\text{subject to} && \underbrace{\mathbf{A}\mathbf{x} = \mathbf{b}}_{\underline{v} \in \mathbb{R}^n}, \quad \underbrace{\mathbf{x} \geq \mathbf{0}}_{\underline{u} \in \mathbb{R}^d_+}
\end{aligned}$$

Dual function:

- Lagrangian is:

$$\begin{aligned}
L(\mathbf{x}, \mathbf{u}, \mathbf{v}) &= \underline{\mathbf{c}^\top \mathbf{x}} + \mathbf{v}^\top (\mathbf{A}\mathbf{x} - \mathbf{b}) - \mathbf{u}^\top \mathbf{x} \\
&= -\mathbf{b}^\top \mathbf{v} + \underbrace{(\mathbf{c} + \mathbf{A}^\top \mathbf{v} - \mathbf{u})}_{\underline{w}}{}^\top \mathbf{x}
\end{aligned}$$

$\underline{w}^\top \underline{x}$

$= \sum_{i=1}^{d} w_i x_i$

- L is affine in $\mathbf{x}$, hence

if $w_i > 0$, $x_i \downarrow$

$w_i < 0$, $x_i \uparrow$

$$\Theta(\mathbf{u}, \mathbf{v}) = \inf_{\mathbf{x}} L(\mathbf{x}, \mathbf{u}, \mathbf{v}) = \begin{cases} -\mathbf{b}^\top \mathbf{v} & \underline{\mathbf{A}^\top \mathbf{v} - \mathbf{u} + \mathbf{c} = 0} \\ -\infty & \text{otherwise} \end{cases}$$

Θ is linear on affine domain $\{(\mathbf{u}, \mathbf{v}) : \mathbf{A}^\top \mathbf{v} - \mathbf{u} + \mathbf{c} = \mathbf{0}\}$, hence concave

Lower bound property: $\underline{p^* \geq -\mathbf{b}^\top \mathbf{v} \text{ if } \mathbf{A}^\top \mathbf{v} + \mathbf{c} \geq \mathbf{0}}$

$\underline{u} = \underline{A}^\top \underline{v} + \underline{c} \geq \underline{0}$

Example: Equality Constrained Norm Minimization

$$\begin{aligned} &\text{Minimize} && \|\mathbf{x}\| \\ &\text{subject to} && \mathbf{A}\mathbf{x} = \mathbf{b} \leftarrow \underline{v} \in \mathbb{R}^n \end{aligned}$$

Dual function: $\Theta(\mathbf{v}) = \inf_{\mathbf{x}} (\|\mathbf{x}\| - \mathbf{v}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{v}) = \begin{cases} \mathbf{b}^\top \mathbf{v} & \|\underbrace{\mathbf{A}^\top \mathbf{v}}_{y}\|_* \le 1 \\ -\infty & \text{otherwise} \end{cases}$

$= \underline{b}^T \underline{v} + \inf_{\underline{x} \neq 0} \{ \|\underline{x}\| - \underbrace{\underline{v}^T A}_{y} \underline{x} \}$

where $\underbrace{\|\mathbf{v}\|_*}_{l_2} \triangleq \sup_{\|\mathbf{u}\| \le 1} \mathbf{u}^\top \mathbf{v}$ (referred to as dual norm of $\|\cdot\|$)

$\Leftarrow$ ↑ if l_2 $\quad$ $1^\circ \ge 0.$

Proof.

It follows from the fact that $\inf_{\mathbf{x}} (\|\mathbf{x}\| - \mathbf{y}^\top \mathbf{x}) = 0$ if $\|\mathbf{y}\|_* \le 1$ (l_2), $-\infty$ otherwise.

- If $\|\mathbf{y}\|_* \le 1$, then $\|\mathbf{x}\| - \mathbf{y}^\top \mathbf{x} \ge 0$, $\forall \mathbf{x}$, with equality if $\mathbf{x} = \mathbf{0}$
  $\le \|y\|_* \|x\| \le \|x\|$ (see below).
- if $\|\mathbf{y}\|_* > 1$, choose $\mathbf{x} = t\mathbf{u}$, where $\mathbf{u}$ satisfies $\|\mathbf{u}\| \le 1$, $\mathbf{u}^\top \mathbf{y} = \|\mathbf{y}\|_* > 1$:
  $\underline{u}$ is the maximizer of $\|y\|_*$

$$\|\mathbf{x}\| - \mathbf{y}^\top \mathbf{x} = t(\underbrace{\underbrace{\|\mathbf{u}\|}_{\le 1} - \underbrace{\|\mathbf{y}\|_*}_{> 1}}_{< 0}) \to -\infty \quad \text{as } t \to \infty$$

□

Lower bound property: $p^* \ge \mathbf{b}^\top \mathbf{v}$ if $\|\mathbf{A}^\top \mathbf{v}\|_* \le 1$

Note: $\underline{y}^T \underline{x} = \|\underline{x}\| \left(\underline{y}^T \frac{\underline{x}}{\|\underline{x}\|} \right) \le \|\underline{x}\| \cdot \|\underline{y}\|_*$, which follows from the fact that $\left\|\frac{\underline{x}}{\|\underline{x}\|}\right\| = \frac{1}{\|\underline{x}\|} \cdot \|\underline{x}\| = 1$ and the definition of $\|\underline{y}\|_*$.

Example: Two-way Partitioning

symmetric

$$\begin{array}{ll} \text{Minimize} & \mathbf{x}^\top \mathbf{W}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i (\mathbf{W})_{ij} x_j \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \dots, n \end{array}$$

non-convex $\quad \Leftrightarrow x_i \in \{-1, 1\}, \forall i \quad$ $\mathbf{v} \in \mathbb{R}^n$ $\quad$ NP-Hard.

- A non-convex problem; feasible set contains 2^n discrete points
- Interpretation: partition $\{1, \dots, n\}$ into two sets; $(\mathbf{W})_{ij}$ is cost of assigning i, j to the same set; $-(\mathbf{W})_{ij}$ is cost of assigning to different sets

Dual function:

$$\Theta(\mathbf{v}) = \inf_{\mathbf{x}} (\underbrace{\mathbf{x}^\top \mathbf{W}\mathbf{x}}_{\text{orig obj}} + \sum_i \mathbf{v}_i (x_i^2 - 1)) = \inf_{\mathbf{x}} \{ \underbrace{\mathbf{x}^\top (\overbrace{\mathbf{W} + \mathrm{Diag}(\mathbf{v})}^{\text{PSD}}) \mathbf{x}}_{\geq 0} - \mathbf{1}^\top \mathbf{v} \}$$

$$= \begin{cases} -\mathbf{1}^\top \mathbf{v} & \mathbf{W} + \mathrm{Diag}\{\mathbf{v}\} \succeq 0 \\ -\infty & \text{otherwise} \quad \text{indef} \end{cases}$$

Lower bound property: $p^* \geq -\mathbf{1}^\top \mathbf{v}$ if $\underbrace{\mathbf{W} + \mathrm{Diag}\{\mathbf{v}\} \succeq 0}_{\text{PSD}} \rightarrow \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top + \begin{bmatrix} v_1 & & \\ & \ddots & \\ & & v_n \end{bmatrix}$

Example: $\mathbf{v} = -\lambda_{\min}(\mathbf{W})\mathbf{1}$ gives non-trivial bound $p^* \geq n\lambda_{\min}(\mathbf{W})$

$$= \mathbf{U} \begin{bmatrix} \lambda_{\max}(\mathbf{W}) + v_1 & & \\ & \ddots & \\ & & \boxed{\lambda_{\min}(\mathbf{W}) + v_n} \end{bmatrix} \mathbf{U}^\top \quad \geq 0$$

Lagrangian Dual and Conjugate Function

$$\begin{array}{ll} \text{Minimize} & f(\mathbf{x}) \\ \text{subject to} & \underbrace{\mathbf{A}\mathbf{x} \le \mathbf{b}^{m}}_{\mathbf{u} \in \mathbb{R}^m_+}, \quad \underbrace{\mathbf{C}\mathbf{x} = \mathbf{d}^{n}}_{\mathbf{v} \in \mathbb{R}^n} \end{array}$$

Dual function:

$$\Theta(\mathbf{u}, \mathbf{v}) = \inf_{\mathbf{x} \in \mathrm{dom}(f)} \left(f(\mathbf{x}) + (\mathbf{A}^\top \mathbf{u} + \mathbf{C}^\top \mathbf{v})^\top \mathbf{x} - \mathbf{b}^\top \mathbf{u} - \mathbf{d}^\top \mathbf{v} \right)$$
$$= -f^*(-\mathbf{A}^\top \mathbf{u} - \mathbf{C}^\top \mathbf{v}) - \mathbf{b}^\top \mathbf{u} - \mathbf{d}^\top \mathbf{v}$$

Legendre–Fenchel conjugate/convex conjugate

- Definition of conjugate function: $f^*(\mathbf{y}) = \sup_{x \in \mathrm{dom}(f)} \left(\mathbf{y}^\top \mathbf{x} - f(\mathbf{x}) \right)$
- Simplifies derivation of dual if conjugate of f is known

Example: Entropy maximization

$$f(\mathbf{x}) = \sum_{i=1}^{n} x_i \log x_i, \quad f^*(\mathbf{y}) = \sum_{i=1}^{n} \exp(y_i - 1)$$

The Lagrangian Dual Problem

$\text{Min}\{-\theta(\underline{u},v)\}$
s.t. $-\underline{u} \leq 0$

$$\left.\begin{array}{ll} \text{Maximize} & \Theta(\mathbf{u},\mathbf{v}) \\ \text{subject to} & \mathbf{u} \geq \mathbf{0} \end{array}\right\} \text{Dual problem}$$

d^*

- Finds largest lower bound on p^*, obtained from Lagrangian dual function
- A convex optimization problem; optimal value denoted d^*
- $\mathbf{u}, \mathbf{v}$ are dual feasible if $\mathbf{u} \geq \mathbf{0}$, $(\mathbf{u}, \mathbf{v}) \in \mathrm{dom}(\Theta)$
- Often simplified by making implicit constraint $\mathbf{u} \geq \mathbf{0}$, $(\mathbf{u}, \mathbf{v}) \in \mathrm{dom}(\Theta)$ explicit

Example: Standard form LP and its dual:

Primal

$$\begin{array}{ll} \text{Minimize} & \mathbf{c}^\top \mathbf{x} \\ \text{subject to} & \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{x} \geq \mathbf{0} \end{array}$$

Dual.

$$\begin{array}{ll} \text{Maximize} & -\mathbf{b}^\top \mathbf{v} \\ \text{subject to} & \mathbf{A}^\top \mathbf{v} + \mathbf{c} \geq \mathbf{0} \end{array}$$

More on this later.

Weak and Strong Duality

Weak duality: $d^* \le p^*$

- Always holds (for convex and non-convex problems)
- Can be used to find non-trivial lower bounds for difficult problems
- For example, solving SDP Semidefinite Programming.

$$\begin{array}{ll} \text{Maximize} & -\mathbf{1}^\top \mathbf{v} \\ \text{subject to} & \mathbf{W} + \mathrm{Diag}\{\mathbf{v}\} \succeq 0 \end{array}$$

yields a lower bound for the two-way partitioning problem

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

- Does not hold in general
- Usually hold for convex problems
- Conditions that guarantee strong duality in convex problems are called constraint qualifications (CQ)

Slater's Constraint Qualification

Strong duality holds for a convex problem

$$
\begin{array}{ll}
\text{Minimize} & f(\mathbf{x}) \\
\text{subject to} & g_i(\mathbf{x}) \le 0, \quad i = 1, \dots, m \\
& \mathbf{A}\mathbf{x} = \mathbf{b}
\end{array}
$$

if $\underline{A}' \underline{x} \le \underline{b}$, then no need to be strict.

if it is strictly feasible, i.e.,

$$\exists \mathbf{x} \in \mathrm{int}\mathcal{D} : \quad g_i(\mathbf{x}) < 0, \quad i = 1, \dots, m, \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$

- Also guarantees that the dual optimum is attained (if $p^* > -\infty$)
- Can be further relaxed: e.g., can replace $\mathrm{int}\mathcal{D}$ with $\mathrm{relint}\mathcal{D}$ (interior relative to affine hull); linear inequalities do not need to hold with strict inequality, ...
- There exist many other types of constraint qualifications (see [BSS, Ch. 5])

Other Well-Known CQs

- LICQ (Linear independence CQ): Gradients of active inequality constraints and gradients of equality constraints are linearly independent at $\mathbf{x}^*$
- MFCQ (Manasarian-Fromovitz CQ): Gradients of equality constraints are LI at $\mathbf{x}^*$, $\exists \mathbf{d} \in \mathbb{R}^n$ such that $\nabla g_i^\top(\mathbf{x}^*)\mathbf{d} < 0$ for active inequality constraints, $\nabla h_j^\top(\mathbf{x}^*)\mathbf{d} = 0$ for equality constraints
- CRCQ (Constant rank CQ): For each subset of gradients of active inequality constraints & gradients of equality constraints, the rank at $\mathbf{x}^*$'s vicinity is constant
- CPLD (Constant positive linear dependence CQ): For each subset of gradients of active inequality constraints & gradients of equality constraints, if it's positive-linear dep. at $\mathbf{x}^*$ then it's positive-linear dependent in $\mathbf{x}^*$'s vicinity
- QNCQ (Quasi-normality CQ): If gradients of active inequality constraints and gradients of equality constraints are positively-linearly dependent at $\mathbf{x}^*$ with duals u_i for inequalities and v_j for equalities, then there is no sequence $\mathbf{x}_k \to \mathbf{x}^*$ such that $v_j \neq 0 \Rightarrow v_j h_j(\mathbf{x}_k) > 0$ and $u_i \neq 0 \Rightarrow u_i g_i(\mathbf{x}_k) < 0$

Geometric Interpretation

$\min f(\mathbf{x}) := p$
$u \geq 0 \rightarrow$ s.t. $g(\mathbf{x}) \leq 0 \leftarrow q$ $\}\ p(q)$

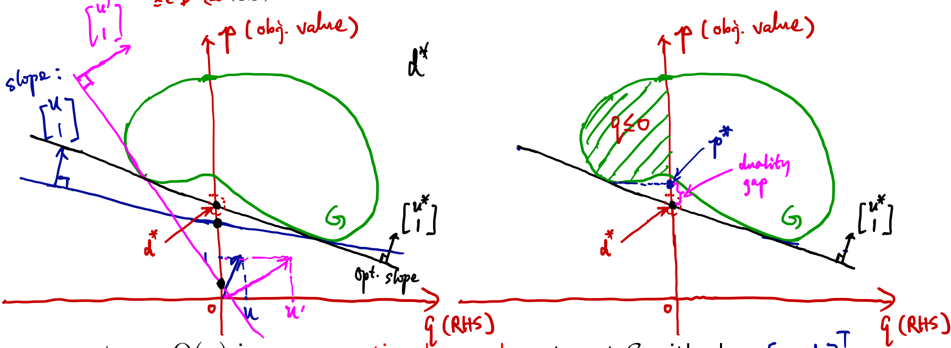
"Perturbed Function": [BSS]

$p^* = p(0)$

For ease of exposition, consider problem with one constraint $g(\mathbf{x}) \leq 0$

Interpretation of dual function

$$\Theta(u) = \inf_{\underset{\mathbf{x}\in\mathcal{D} \Leftrightarrow}{(q,p)\in\mathcal{G}}} (u\overset{g(\mathbf{x})}{q} + \overset{f(\mathbf{x})}{p}), \quad \text{where } \overset{\in\mathbb{R}^2}{\mathcal{G}} = \{(\overset{q}{g(\mathbf{x})}, \overset{p}{f(\mathbf{x})}) | \mathbf{x} \in \mathcal{D}\}$$



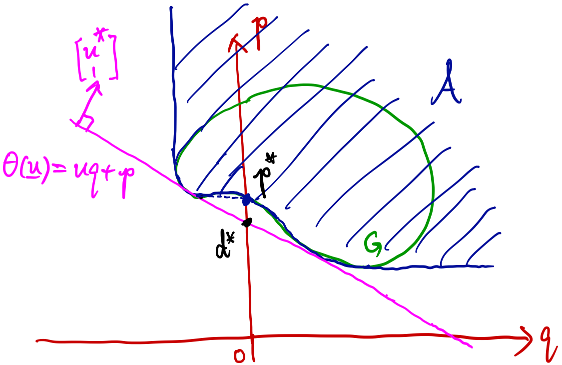
- $uq + p = \Theta(u)$ is an supporting hyperplane to set $\mathcal{G}$ with slope $[u, 1]^T$
- Hyperplane intersects p-axis at $p = \Theta(u)$

Geometric Interpretation

Epigraph variation: Same interpretation if $\mathcal{G}$ is replaced with

$$\mathcal{A} = \{(q,p) | g(\mathbf{x}) \le q, f(\mathbf{x}) \le p \text{ for some } \mathbf{x} \in \mathcal{D}\}$$

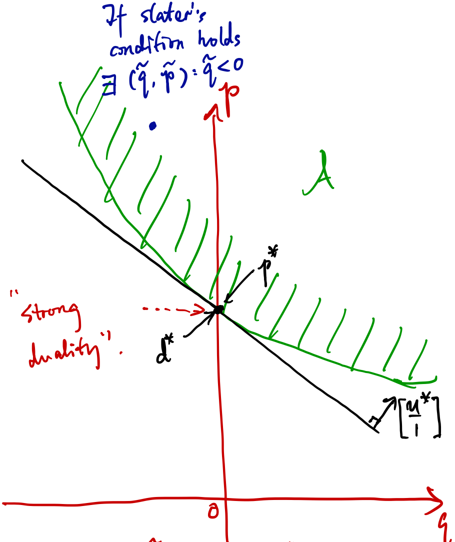
Vector-valued epigraph



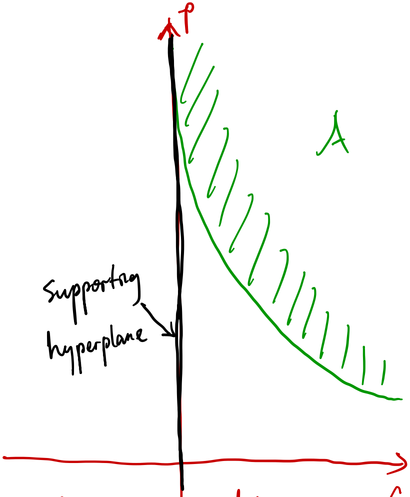
Strong duality

- Holds if there is a non-vertical supporting hyperplane to $\mathcal{A}$ at $(0, p^*)$
- For convex problem, $\mathcal{A}$ is convex, hence has supporting hyperplane at $(0, p^*)$
- Slater's condition: If there exist $(\tilde{q}, \tilde{p}) \in \mathcal{A}$ with $\tilde{q} < 0$, then supporting hyperplanes at $(0, p^*)$ must be non-vertical

Strong Duality Geometric Interpretation



Existence of $\tilde{q} < 0$, "forcing" the supporting hyperplane to be non-vertical.



Clearly, Slater's condition does NOT hold. (no well-defined intersect pt. $\uparrow \infty$)

Example: Inequality Form of LP

Primal problem:

Assume $\underline{A}$ is full row rank.

$$\begin{aligned} \text{Minimize} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} \leftarrow \underline{u} \geq \underline{0} \end{aligned}$$

Dual function:

$$L(\underline{x}, \underline{u}) = \underbrace{\underline{c}^\top \underline{x}}_{\text{obj}} + \underbrace{\underline{u}^\top}_{\geq 0} \underbrace{(\underline{A}\underline{x} - \underline{b})}_{\leq 0} \leq \underline{c}^\top \underline{x}$$

$$\Theta(\mathbf{u}) = \inf_{\mathbf{x}} \left((\mathbf{c} + \mathbf{A}^\top \mathbf{u})^\top \mathbf{x} - \mathbf{b}^\top \mathbf{u} \right) = \begin{cases} -\mathbf{b}^\top \mathbf{u} & \mathbf{A}^\top \mathbf{u} + \mathbf{c} = \mathbf{0} \\ -\infty & \text{otherwise} \end{cases}$$

Dual problem:

$$\left. \begin{aligned} \max \quad & \theta(\underline{u}) \\ \text{s.t.} \quad & \underline{u} \geq \underline{0} \end{aligned} \right\}$$

$$\boxed{\begin{aligned} \text{Maximize} \quad & -\mathbf{b}^\top \mathbf{u} \\ \text{subject to} \quad & \mathbf{A}^\top \mathbf{u} + \mathbf{c} = \mathbf{0}, \quad \mathbf{u} \geq \mathbf{0} \end{aligned}}$$ LP.

- From Slater's condition: $p^* = d^*$ if $\mathbf{A}\tilde{\mathbf{x}} < \mathbf{b}$ for some $\tilde{\mathbf{x}}$
- In fact, $p^* = d^*$ except when primal and dual are infeasible (LICQ)

Example: Quadratic Program

positive definite.

Primal problem ($\mathbf{P} \in \mathbb{S}^n_{++}$):

$$\begin{aligned} &\text{Minimize} \quad \mathbf{x}^\top \mathbf{P} \mathbf{x} \\ &\text{subject to} \quad \mathbf{A}\mathbf{x} < \mathbf{b} \leftarrow \underline{u} \geq \underline{0} \end{aligned}$$

Dual function:

$$\Theta(\overset{u}{\cancel{\lambda}}) = \inf_{\mathbf{x}} \Big(\overbrace{\underbrace{\mathbf{x}^\top \mathbf{P}\mathbf{x} + \mathbf{u}^\top(\mathbf{A}\mathbf{x} - \mathbf{b})}_{\text{quadratic fn.}}}^{\text{Lagrangian}} \Big) = -\frac{1}{4}\mathbf{u}^\top \mathbf{A}\mathbf{P}^{-1}\mathbf{A}^\top \mathbf{u} - \mathbf{b}^\top \mathbf{u}$$

Dual problem:

take der, set to 0 $\Rightarrow$ LES: $2\underline{P}\underline{x} + \underline{A}^T\underline{u} = \underline{0} \Rightarrow \underline{x} = -\frac{1}{2}\underline{P}^{-1}\underline{A}^T\underline{u}$

invertible

$$\begin{aligned} &\text{Maximize} \quad -\frac{1}{4}\mathbf{u}^\top \mathbf{A}\mathbf{P}^{-1}\mathbf{A}^\top \mathbf{u} - \mathbf{b}^\top \mathbf{u} \\ &\text{subject to} \quad \mathbf{u} \geq \mathbf{0} \end{aligned}$$

- From Slater's condition: $p^* = d^*$ if $\mathbf{A}\tilde{\mathbf{x}} < \mathbf{b}$ for some $\tilde{\mathbf{x}}$
- In fact, $p^* = d^*$ always (slater's condition is automatic).

Example: Support Vector Machine

Given labels $y \in \{-1, 1\}^n$, feature vectors $\mathbf{x}_1, \ldots, \mathbf{x}_m$. Let $\mathbf{X} \triangleq [\mathbf{x}_1, \ldots, \mathbf{x}_m]^\top$
Recall from Lecture 1 that the support vector machine problem:

WTF: $\underline{w}, b$, s.t.
if $y_i = 1$, $\underline{w}^T\underline{x}_i + b \geq 1$
if $y_i = -1$, $\underline{w}^T\underline{x}_i + b \leq -1$
$\Rightarrow y_i(\underline{w}^T\underline{x}_i + b) \geq 1$.

Q: Slater's ?

$$\underset{\mathbf{w}, b, \boldsymbol{\epsilon}}{\text{Minimize}} \quad \frac{1}{2}\mathbf{w}^\top\mathbf{w} + C\sum_{i=1}^{m}\epsilon_i$$

$$\text{subject to} \quad \underbrace{y_i(\mathbf{w}^\top\mathbf{x}_i + b)}_{0} \geq \underbrace{1 - \epsilon_i}_{<0}, \quad i = 1, \ldots, m \leftarrow u_i \geq 0$$

$$\underbrace{\epsilon_i}_{>0} \geq 0, \quad i = 1, \ldots, m \leftarrow v_i \geq 0$$

Guess:
$\underline{w} = \underline{0}$, $b = 0$,
choose ε_i suff. large.

Introducing dual variables $\mathbf{u}, \mathbf{v} \geq \mathbf{0}$ to obtain the Lagrangian:

$$L(\mathbf{w}, b, \boldsymbol{\epsilon}, \mathbf{u}, \mathbf{v}) = \underbrace{\frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{m}\epsilon_i}_{\text{orig. obj.}} + \underbrace{\sum_{i=1}^{n} u_i(1 - \epsilon_i - y_i(\mathbf{w}^\top\mathbf{x}_i + b)) - \sum_{i=1}^{n} v_i\epsilon_i}_{\text{penalty terms.}}$$

Minimizing over $\mathbf{w}, b, \boldsymbol{\epsilon}$ yields the Lagrangian dual function ...

Example: Support Vector Machine

$$\Theta(\mathbf{u},\mathbf{v}) = \begin{cases} \overbrace{-\frac{1}{2}\mathbf{u}^\top \underbrace{\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}}_{\text{PSD}}\mathbf{u}}^{\text{NSD}} + \mathbf{1}^\top\mathbf{u} & \text{if } \mathbf{u} = C\mathbf{1}-\mathbf{v}, \mathbf{u}^\top\mathbf{y}=0 \\ -\infty & \text{otherwise} \end{cases}$$

where $\tilde{\mathbf{X}} \triangleq \mathbf{X}\mathrm{Diag}\{y_1,\ldots,y_m\}$. The dual problem, after eliminating $\mathbf{v}$, becomes:

$$\boxed{\begin{array}{ll} \underset{\boldsymbol{\alpha}}{\text{Minimize}} & \frac{1}{2}\mathbf{u}^\top\tilde{\mathbf{X}}^\top\tilde{\mathbf{X}}\mathbf{u} - \mathbf{1}^\top\mathbf{u} \\ \text{subject to} & \mathbf{u}^\top\mathbf{y} = 0, \quad 0 \le u_i \le C, \ i = 1,\ldots,m. \end{array}}$$

- Slater's condition is satisfied $\Rightarrow$ strong duality
- In ML literature, more common to work with the dual: Quadratic having PSD Hessian with simple bounds, plus a single linear & simple box constraints
- At optimality, we can verify that $\mathbf{w} = \tilde{\mathbf{X}}\mathbf{u}$ (this is not a coincidence! will be proved by KKT conditions in next class)

Next Class

Optimality Conditions