Deriving the dual problem

Andersen Ang

Mathématique et recherche opérationnelle
UMONS, Belgium

manshun.ang@umons.ac.be Homepage: angms.science

First draft : March 23, 2020
Last update : March 23, 2020

Primal and dual problem

- Consider the problem in the form

$$(P) \;\; : \;\; \min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{A}\mathbf{x}),$$

  where $\mathbf{x} \in \mathbb{R}^n$ is the primal variable, $f : \mathbb{R}^n \to \bar{\mathbb{R}}$ is proper, close and convex, $g : \mathbb{R}^m \to \bar{\mathbb{R}}$ is proper, close, convex and possibly nonsmooth, and $\mathbf{A} \in \mathbb{R}^{m \times n}$ is a matrix.

- The dual problem[1] is

$$(D) \;\; : \;\; \min_{\boldsymbol{\nu}} f^*(\mathbf{A}^\top \boldsymbol{\nu}) + g^*(-\boldsymbol{\nu}),$$

  where $\boldsymbol{\nu} \in \mathbb{R}^m$ is the dual variable, f^* is the convex conjugate of f.

- Special case : if $m = n$ and $\mathbf{A}$ is identity, we have

$$(P) \;\; : \;\; \min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x}), \quad (D) \;\; : \;\; \min_{\boldsymbol{\nu}} f^*(\boldsymbol{\nu}) + g^*(-\boldsymbol{\nu}).$$

- This document : show how to derive the dual problem.

[1] In fact there should be a negative sign in front of the min, see page 7.

Convex conjugate, recall

- Given a function with $\mathrm{dom} f$ as a set C and taking values on the extended real line,

$$f : C \to \mathbb{R} \cup \{\pm\infty\}.$$

- The convex conjugate, or Legendre - Fenchel conjugate, is defined by the supremum as

$$f^*(\mathbf{y}) = \sup_{\mathbf{x} \in \mathrm{dom} f} \Big\{ \langle \mathbf{x}, \mathbf{y} \rangle - f(\mathbf{x}) \Big\}.$$

- Loosely speaking,

$$f^*(\boldsymbol{\nu}) = \max_{\mathbf{x} \in \mathrm{dom} f} \Big\{ \langle \mathbf{x}, \boldsymbol{\nu} \rangle - f(\mathbf{x}) \Big\}.$$

Deriving the dual : forming the Lagrangian

- Given the problem

$$(P) \; : \; \min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{A}\mathbf{x}),$$

- Transform it to the form

$$(P') \; : \; \min_{\mathbf{x},\mathbf{y}} f(\mathbf{x}) + g(\mathbf{y}) \text{ s.t. } \mathbf{A}\mathbf{x} = \mathbf{y},$$

  where $\mathbf{y} \in \mathbb{R}^m$ is an auxiliary variable. The minimization variables $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ are linked by the constraint $\mathbf{A}\mathbf{x} = \mathbf{y}$.

- Form the associated Lagrangian

$$L(\mathbf{x}, \mathbf{y}, \boldsymbol{\nu}) = f(\mathbf{x}) + g(\mathbf{y}) - \langle \boldsymbol{\nu}, \mathbf{A}\mathbf{x} - \mathbf{y} \rangle,$$

  where $\boldsymbol{\nu} \in \mathbb{R}^m$ is the Lagrangian multiplier.

- Here $\mathbf{x}, \mathbf{y}$ are the primal variables, and $\boldsymbol{\nu}$ is the dual variable. The goal now is to express L as function solely on $\boldsymbol{\nu}$ by removing $\mathbf{x}, \mathbf{y}$ in L.

Deriving the dual : minimizing the Lagrangian on primal variables

- To derive the dual, recall that minimizing the Lagrangian on primal variables gives the dual. Therefore we want to compute

$$\min_{\mathbf{x},\mathbf{y}} L(\mathbf{x},\mathbf{y},\boldsymbol{\nu}).$$

- As $L(\mathbf{x},\mathbf{y},\boldsymbol{\nu}) = f(\mathbf{x}) + g(\mathbf{y}) - \langle \boldsymbol{\nu}, \mathbf{A}\mathbf{x} - \mathbf{y} \rangle$, we have

$$\min_{\mathbf{x},\mathbf{y}} L(\mathbf{x},\mathbf{y},\boldsymbol{\nu}) = \min_{\mathbf{x}} \Big\{ f(\mathbf{x}) - \langle \boldsymbol{\nu}, \mathbf{A}\mathbf{x} \rangle \Big\} + \min_{\mathbf{y}} \Big\{ g(\mathbf{y}) + \langle \boldsymbol{\nu}, \mathbf{y} \rangle \Big\}.$$

  Notice that we have splitted the minimization on $\mathbf{x}$ and $\mathbf{y}$. In fact, this is the reason why we transform (P) to (P') : such splitting unlocks the sum $f + g$ so we can deal with them separately.

- The remaining work is to use conjugate and show this minimization gives the dual function.

Deriving the dual : using conjugate

- Continue from the last page, $\min_{\mathbf{x},\mathbf{y}} L(\mathbf{x},\mathbf{y},\boldsymbol{\nu})$ is equal to

$$\begin{aligned} & \min_{\mathbf{x}} \Big\{ f(\mathbf{x}) - \langle \boldsymbol{\nu}, \mathbf{A}\mathbf{x} \rangle \Big\} + \min_{\mathbf{y}} \Big\{ g(\mathbf{y}) + \langle \boldsymbol{\nu}, \mathbf{y} \rangle \Big\} \\ = \;& \min_{\mathbf{x}} \Big\{ f(\mathbf{x}) - \langle \mathbf{A}^\top \boldsymbol{\nu}, \mathbf{x} \rangle \Big\} + \min_{\mathbf{y}} \Big\{ g(\mathbf{y}) - \langle -\boldsymbol{\nu}, \mathbf{y} \rangle \Big\} \\ = \;& -\max_{\mathbf{x}} \Big\{ \langle \mathbf{A}^\top \boldsymbol{\nu}, \mathbf{x} \rangle - f(\mathbf{x}) \Big\} - \max_{\mathbf{y}} \Big\{ \langle -\boldsymbol{\nu}, \mathbf{y} \rangle - g(\mathbf{y}) \Big\} \\ = \;& -f^*(\mathbf{A}^\top \boldsymbol{\nu}) - g^*(-\nu), \end{aligned}$$

  where the third equality comes from $\min\{f\} = -\max\Big\{ -f(\mathbf{x}) \Big\}$, and the last equality comes from the definition of conjugate.

- Hence we arrive at the dual problem

$$(D') \;:\; \max_{\boldsymbol{\nu}} \Big\{ -f^*(\mathbf{A}^\top \boldsymbol{\nu}) - g^*(-\boldsymbol{\nu}) \Big\}.$$

Deriving the dual : finishing the derivation

- Continue from the last page, we have the dual problem

$$(D') \;:\; \max_{\boldsymbol{\nu}} \Big\{ -f^*(\mathbf{A}^\top \boldsymbol{\nu}) - g^*(-\boldsymbol{\nu}) \Big\}.$$

- Apply $\min\{f\} = -\max\Big\{ -f(\mathbf{x})\Big\}$ again, we have

$$(D) \;:\; -\min_{\boldsymbol{\nu}} \Big\{ f^*(\mathbf{A}^\top \boldsymbol{\nu}) + g^*(-\boldsymbol{\nu}) \Big\}.$$

- If we want the argmin instead of $\min$, the negative sign in front of $\min$ can be ignored and we have

$$(D) \;:\; \operatorname*{argmin}_{\boldsymbol{\nu}} \Big\{ f^*(\mathbf{A}^\top \boldsymbol{\nu}) + g^*(-\boldsymbol{\nu}) \Big\}.$$

Last page - summary

- The primal problem

$$(P) \ : \min_{\mathbf{x}} \left\{ f(\mathbf{x}) + g(\mathbf{A}\mathbf{x}) \right\} \text{ or } \operatorname*{argmin}_{\mathbf{x}} \left\{ f(\mathbf{x}) + g(\mathbf{A}\mathbf{x}) \right\}.$$

- The dual problem

$$(D') \ : \max_{\boldsymbol{\nu}} \left\{ -f^*(\mathbf{A}^\top \boldsymbol{\nu}) - g^*(-\boldsymbol{\nu}) \right\} \text{ or } \operatorname*{argmax}_{\mathbf{x}} \left\{ -f^*(\mathbf{A}^\top \boldsymbol{\nu}) - g^*(-\boldsymbol{\nu}) \right\}.$$

$$(D) \ : -\min_{\boldsymbol{\nu}} \left\{ f^*(\mathbf{A}^\top \boldsymbol{\nu}) + g^*(-\boldsymbol{\nu}) \right\} \text{ or } \operatorname*{argmin}_{\mathbf{x}} \left\{ f^*(\mathbf{A}^\top \boldsymbol{\nu}) + g^*(-\boldsymbol{\nu}) \right\}.$$

End of document