Duality

Pontus Giselsson

Today's lecture

- Fenchel weak and strong duality
- necessary and sufficient optimality conditions
- Lagrange weak and strong duality
- KKT conditions

Why duality?

- sometimes it is easier to solve dual than primal problem
- useful if primal solution can be obtained from dual

Fenchel duality

- consider composite optimization problem

$$\begin{array}{ll}\text{minimize} & f(x) + g(y)\\ \text{subject to} & Lx = y\end{array}$$

  with f, g proper, closed, and convex, L linear operator
- will call this primal problem
- equivalent formulation with indicator functions:

$$\text{minimize} \quad f(x) + g(y) + \iota(Lx = y)$$

  where the indicator function is defined as

$$\iota(Lx = y) = \begin{cases} 0 & \text{if } Lx = y \\ \infty & \text{else} \end{cases}$$

Reformulation

- consider

$$h(x,y) = \sup_{\mu} \{\langle \mu, Lx - y\rangle\}$$

- what is the value of h if $Lx = y$? 0
- what is the value of h is $Lx \neq y$? ∞
- what is h? $h(x,y) = \iota(Lx = y)$ and problem can be written:

$$\inf_{x,y} \left\{ f(x) + g(y) + \sup_{\mu}\{\langle \mu, Lx - y\rangle\} \right\}$$

or

$$p^\star := \inf_{x,y} \sup_{\mu} \{f(x) + g(y) + \langle \mu, Lx - y\rangle\}$$

where $p^\star$ is the primal optimal value

Weak duality

- let

$$\mathcal{L}(x,y,\mu) := f(x) + g(y) + \langle \mu, Lx - y\rangle$$

- then

$$p^\star = \inf_{x,y}\sup_{\mu}\{f(x) + g(y) + \langle \mu, Lx - y\rangle\} = \inf_{x,y}\sup_{\mu}\mathcal{L}(x,y,\mu)$$

- what happens if we swap inf-sup (replace ? by $\leq$ or $\geq$)?

$$p^\star = \inf_{x,y}\sup_{\mu}\mathcal{L}(x,y,\mu)\ ?\ \sup_{\mu}\inf_{x,y}\mathcal{L}(x,y,\mu) =: d^\star$$

- it should be $p^\star \geq d^\star$, i.e.:

$$\inf_{x,y}\sup_{\mu}\mathcal{L}(x,y,\mu) \geq \sup_{\mu}\inf_{x,y}\mathcal{L}(x,y,\mu)$$

Weak duality

- we claim $d^\star \leq p^\star$, i.e.:

$$\sup_\mu \inf_{x,y} \mathcal{L}(x,y,\mu) \leq \inf_{x,y} \sup_\mu \mathcal{L}(x,y,\mu)$$

- proof when $\sup_\mu$ attained: let $\psi(\mu) := \inf_{x,y} \mathcal{L}(x,y,\mu)$ then

$$\psi(\mu) \leq \mathcal{L}(x,y,\mu) \qquad \text{for all } x, y, \mu$$

- let $\mu^\star$ maximize $\psi(\mu)$, then

$$\begin{aligned} & \sup_\mu \psi(\mu) = \psi(\mu^\star) \leq \mathcal{L}(x,y,\mu^\star) \leq \sup_\mu \mathcal{L}(x,y,\mu) && \text{for all } x, y \\ \iff \quad & \sup_\mu \inf_{x,y} \mathcal{L}(x,y,\mu) \leq \sup_\mu \mathcal{L}(x,y,\mu) && \text{for all } x, y \\ \iff \quad & \sup_\mu \inf_{x,y} \mathcal{L}(x,y,\mu) \leq \inf_{x,y} \sup_\mu \mathcal{L}(x,y,\mu) \end{aligned}$$

Weak duality comments

- weak duality

$$\sup_{\mu} \inf_{x,y} \mathcal{L}(x,y,\mu) \leq \inf_{x,y} \sup_{\mu} \mathcal{L}(x,y,\mu)$$

- it holds also when maximum in ψ not attained
- it is better to choose last!
- no convexity is assumed in proof $\Rightarrow$ holds also in nonconvex case
- holds for general functions and is called *min-max inequality*
- in our setting this is called *weak duality*
  (left hand side problem is called dual problem)

Fenchel dual problem

- the problem with inf-sup swapped is the Fenchel dual problem:

$$
\begin{aligned}
\sup_{\mu} \inf_{x,y} \mathcal{L}(x,y,\mu) &= \sup_{\mu} \inf_{x,y} \{f(x) + g(y) + \langle \mu, Lx - y\rangle\} \\
&= \sup_{\mu} -\big(\sup_{x,y} \{-f(x) - g(y) + \langle \mu, -Lx + y\rangle\} \big) \\
&= \sup_{\mu} \Big\{ -\Big(\sup_{x} \{\langle x, -L^*\mu\rangle - f(x)\} \\
&\qquad\qquad + \sup_{y} \{\langle y, \mu\rangle - g(y)\} \Big) \Big\} \\
&= \sup_{\mu} \{-f^*(-L^*\mu) - g^*(\mu)\} = d^\star
\end{aligned}
$$

- i.e., primal and dual problems are

$$
p^\star = \inf_{x,y} \sup_{\mu} \mathcal{L}(x,y,\mu) \qquad\qquad d^\star = \sup_{\mu} \inf_{x,y} \mathcal{L}(x,y,\mu)
$$

Strong duality

- when does $p^\star = d^\star$ hold, i.e., when does *strong duality* hold?
- it holds if f, g proper closed convex and

$$\text{ri dom } g \cap \text{ri } L(\text{dom } f) \neq \emptyset$$

- proof: apply "Key result 2"

$$\begin{aligned} p^\star &= \inf_x \{f(x) + g(Lx)\} \\ &= -\sup_x \{\langle 0, x\rangle - f(x) - g(Lx)\} \\ &= -(f + g \circ L)^*(0) \\ &= -\min_\mu \{f^*(-L^*\mu) + g^*(\mu)\} \\ &= \max_\mu \{-f^*(-L^*\mu) - g^*(\mu)\} = d^\star \end{aligned}$$

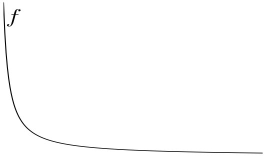
- note by "Key result 2" that dual optimal point attained
- cannot say anything about if primal optimal point attained

Strong duality example

- consider the problem

$$\text{minimize} \quad f(x) + g(x)$$

with $f(x) = 1/x$, dom $f = \{x \mid x > 0\}$ and $g(x) = 0$



- primal optimal $p^\star = 0$ but primal optimal point not attained

Strong duality example, cont'd

- dual problem: $\max_\mu \{-f^*(-\mu) - g^*(\mu)\}$ where

$$f^*(-\mu) = \sup_x \{-\mu x - 1/x + \iota(x > 0)\} = -2\sqrt{\mu} + \iota(\mu \geq 0)$$
$$g^*(\mu) = \sup_x \{\langle \mu, x \rangle - 0\} = \iota(\mu = 0)$$

  (domain encoded with indicator functions)
- dual optimal point: $\mu = 0$, and value: $d^\star = 0$
- in this example:
  - strong duality $d^\star = p^\star$ (assumptions are met)
  - dual optimal point attained
  - primal optimal point not attained
    (should pose problem such that primal optimum attained!)

Optimality conditions for composite problems

- objective: state conditions that guarantee that x, y solves:

$$\text{minimize} \quad f(x) + g(y) + \iota(Lx = y)$$

  with f, g proper closed and convex and L a linear operator

- we use (again) the following *constraint qualification*:

$$\text{ri dom} g \cap \text{ri } L(\text{dom} f) \neq \emptyset \qquad \Longleftrightarrow \qquad \text{ri dom}(g \circ L) \cap \text{ri dom } f \neq \emptyset$$

- (note: we assume that the primal optimum attained)

Equivalent formulation

- composite problem:

$$\text{minimize} \quad f(x) + g(y) + \iota(Lx = y)$$

- let

$$\begin{aligned} z &= (x, y), & F(z) &= f(x) + g(y) \\ Kz &= Lx - y, & V &= \{z \mid Kz = 0\} \end{aligned}$$

- the we get the equivalent formulation:

$$\text{minimize} \quad F(z) + \iota_V(z)$$

Translate assumptions

- our assumption:

$$\text{ri dom} g \cap \text{ri } L(\text{dom} f) \neq \emptyset \quad \Longleftrightarrow \quad \text{ri dom } f \cap \text{ri dom}(g \circ L) \neq \emptyset$$

- $z = (x, y)$, $F(z) = f(x) + g(y)$, $Kz = Lx - y$, $\iota_V(z) = \iota_{Kz=0}$
- we have

$$\begin{aligned}
& && \text{ri dom } f \cap \text{ri dom } (g \circ L) \neq \emptyset \\
&\Leftrightarrow && \exists x | (x, x) \in \text{ri dom } f \times \text{ri dom } (g \circ L) \\
&\Leftrightarrow && \exists x | (x, Lx) \in \text{ri dom } f \times \text{ri dom } g \\
&\Leftrightarrow && \exists x | (x, Lx) \in \text{ri dom } F \\
&\Leftrightarrow && \exists z \in V | z \in \text{ri dom } F \\
&\Leftrightarrow && \exists z | z \in \text{dom } \iota_V \cap \text{ri dom } F \\
&\Leftrightarrow && \text{dom } \iota_V \cap \text{ri dom } F \neq \emptyset \\
&\Leftrightarrow && \text{ri dom } \iota_V \cap \text{ri dom } F \neq \emptyset
\end{aligned}$$

  where last step holds since ri dom $\iota_V =$ dom ι_V since V affine
- $\Rightarrow$ can apply subdifferential sum rule to $\partial(F + \iota_V)$!

Fermat's rule

- Fermat's rule (necessary and sufficient for optimal point):

$$0 \in \partial(F(z) + \iota_V(z))$$

- we know that

$$\partial(F(z) + \iota_V(z)) = \partial F(z) + \partial \iota_V(z) = \partial F(z) + N_V(z)$$

Subdifferentials

- the normal cone to linear subspace:

$$N_V(z) = \begin{cases} \mathrm{Im} K^* & \text{if } Kz = 0 \\ \emptyset & \text{else} \end{cases}$$

  i.e. $N_V(z) = K^*\mu$ for some μ

- the adjoint $K^*\mu = (L^*\mu, -\mu)$ since

$$\begin{aligned} \langle Kz, \mu \rangle &= \langle Lx - y, \mu \rangle = \langle x, L^*\mu \rangle - \langle y, \mu \rangle = \langle (x, y), (L^*\mu, -\mu) \rangle \\ &= \langle z, (L^*\mu, -\mu) \rangle \end{aligned}$$

- subdifferential to $F(z) = f(x) + g(y)$ is

$$\partial F(z) = (\partial f(x), \partial g(y))$$

Optimality conditions

- the optimality condition $0 \in \partial F(x) + N_V(z)$ becomes:

$$0 \in \partial F(z) + K^*\mu \qquad \text{and} \qquad Kz = 0(Lx = y)$$

  or equivalently

$$\begin{aligned} 0 &\in \partial f(x) + L^*\mu \\ 0 &\in \partial g(y) - \mu \\ 0 &= Lx - y \end{aligned}$$

- necessary and sufficient under assumptions!

Alternative optimality conditions

- optimality conditions from previous slide (rearranged):

$$\begin{aligned} -L^*\mu &\in \partial f(x) \\ \mu &\in \partial g(y) \\ 0 &= Lx - y \end{aligned}$$

- equivalent optimality conditions using conjugate functions:

$$\begin{aligned} x &\in \partial f^*(-L^*\mu) \\ y &\in \partial g^*(\mu) \\ 0 &= Lx - y \end{aligned}$$

- this gives

$$0 = Lx - y \in -(-L)\partial f^*(-L^*\mu) - \partial g^*(\mu)$$

which is Fermat's rule for the dual problem

$$\max_{\mu}\{-f^*(-L^*\mu) - g^*(\mu)\}$$

(under some constraint qualification)

More alternative optimality conditions

- optimality conditions from previous slide:

$$\begin{aligned} -L^*\mu &\in \partial f(x) \\ \mu &\in \partial g(y) \\ 0 &= Lx - y \end{aligned}$$

- other equivalent reformulations using $Lx = y$:

$$\begin{aligned} &x \in \partial f^*(-L^*\mu) \qquad\qquad & x &\in \partial f^*(-L^*\mu) \\ &\mu \in \partial g(Lx) \qquad\qquad & Lx &\in \partial g^*(\mu) \end{aligned}$$

and

$$\begin{aligned} -L^*\mu &\in \partial f(x) \qquad\qquad & -L^*\mu &\in \partial f(x) \\ Lx &\in \partial g^*(\mu) \qquad\qquad & \mu &\in \partial g(Lx) \end{aligned}$$

- recall Lagrangian $\mathcal{L}(x, y, \mu) = f(x) + g(y) + \langle \mu, Lx - y \rangle$, another equivalent condition:

$$0 \in \partial \mathcal{L}(x, y, \mu)$$

Saddle-point condition

- recall $\mathcal{L}(x, y, \mu) = f(x) + g(y) + \langle \mu, Lx - y \rangle$
- computing:

$$0 \in \partial \mathcal{L}(x, y, \mu) \tag{1}$$

  gives

$$\begin{aligned} 0 &\in \partial f(x) + L^* \mu \\ 0 &\in \partial g(y) - \mu \\ 0 &= Lx - y \end{aligned}$$

- (1) is also necessary and sufficient condition (under assumptions)

Solving the primal from the dual

- we are primarily interested in the primal problem (often x)
- is it possible to solve primal from dual?
- sometimes! if we can find x such that any of the following holds:

$$x \in \partial f^*(-L^*\mu) \qquad\qquad x \in \partial f^*(-L^*\mu)$$
$$\mu \in \partial g(Lx) \qquad\qquad Lx \in \partial g^*(\mu)$$

and

$$-L^*\mu \in \partial f(x) \qquad\qquad -L^*\mu \in \partial f(x)$$
$$Lx \in \partial g^*(\mu) \qquad\qquad \mu \in \partial g(Lx)$$

Example

- consider optimality condition

$$x \in \partial f^*(-L^*\mu)$$
$$Lx \in \partial g^*(\mu)$$

- example: f is strongly convex $\Rightarrow$ f^* differentiable $\Rightarrow$

$$x \in \partial f^*(-L^*\mu) \quad \iff \quad x = \nabla f^*(-L^*\mu)$$

  only x that satisfies condition $\Rightarrow$ must be optimal if exists, i.e., if

$$Lx \in \partial g^*(\mu)$$

- (most algorithms that solve dual also output primal solution)

Fenchel duality summary

have used "Key result 2" to (explicitly or implicitly) show strong duality and necessary and sufficient optimality conditions for composite optimization problems under stated assumptions

Lagrange duality

- some might be familiar with Lagrange duality and KKT-conditions
- can derive this from Fenchel duality
- Fenchel duality can also be derived from Lagrange duality

Lagrange duality

- let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^k$, be convex and L be linear
- consider the following convex problem on standard form:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0 \\ & Lx = b \end{array}$$

- equivalent formulation with indicator functions

$$\text{minimize} \quad f(x) + \iota(g(x) \leq 0) + \iota(Lx = b)$$

Reformulate indicator functions

- the indicator function $\iota(g(x) \leq 0)$ can be modeled as

$$\sup_{\mu \geq 0} \{\langle \mu, g(x) \rangle\} = \left\{ \begin{matrix} 0 & \text{if } g(x) \leq 0 \\ \infty & \text{else} \end{matrix} \right\} = \iota(g(x) \leq 0)$$

- the indicator function $\iota(Lx = b)$ can be modeled as

$$\sup_{\lambda} \{\langle \lambda, Lx - b \rangle\} = \left\{ \begin{matrix} 0 & \text{if } Lx = b \\ \infty & \text{else} \end{matrix} \right\} = \iota(Lx = b)$$

Equivalent formulation of primal problem

- using reformulation of indicator function, we get:

$$\inf_x \{f(x) + \sup_{\mu \geq 0} \langle \mu, g(x) \rangle + \sup_{\lambda} \langle \lambda, Lx - b \rangle\}$$

or

$$\inf_x \sup_{\mu \geq 0, \lambda} \{f(x) + \langle \mu, g(x) \rangle + \langle \lambda, Lx - b \rangle\}$$

- by the min-max inequality, we have

$$\begin{aligned} &\sup_{\lambda, \mu \geq 0} \inf_x \{f(x) + \langle \mu, g(x) \rangle + \langle \lambda, Lx - b \rangle\} \\ &\qquad \leq \inf_x \sup_{\lambda, \mu \geq 0} \{f(x) + \langle \mu, g(x) \rangle + \langle \lambda, Lx - b \rangle\} \end{aligned}$$

- when do we have equality, i.e., strong duality?

Strong duality

- if Slater's condition holds, i.e., if there exists $\bar{x}$ such that

$$g(\bar{x}) < 0 \qquad \text{and} \qquad L\bar{x} = b$$

- then strong duality holds, i.e.,:

$$\sup_{\lambda,\mu\geq 0} \inf_x \{f(x) + \langle \mu, g(x)\rangle + \langle \lambda, Lx - b\rangle\}$$
$$= \inf_x \sup_{\lambda,\mu\geq 0} \{f(x) + \langle \mu, g(x)\rangle + \langle \lambda, Lx - b\rangle\}$$

- can be shown by considering equivalent problem

$$\text{minimize} \quad \underbrace{f(x) + \iota(g(x) \leq y)}_{h_1(x,y)} + \underbrace{\iota(y \leq 0) + \iota(Lx = b)}_{h_2(x,y)}$$

and apply Fenchel strong duality

Lagrange optimality conditions

- the optimality conditions for standard form:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g(x) \leq 0 \\ & Lx = b \end{array}$$

  are called KKT-conditions (Karush-Kuhn-Tucker)

- they are given by

$$\begin{aligned} 0 &\in \partial f(x) + \sum_{i=1}^{k} \mu \partial g(x) + L^* \lambda \\ 0 &= Lx - b \\ 0 &\leq \mu \\ 0 &\geq g(x) \\ 0 &= \mu_i g_i(x) \text{ for all } i = 1, \ldots, k \end{aligned}$$

  (usually stated for differentiable f, g)

Prove KKT conditions

- we will assume, again, Slater's constraint qualification, i.e., $\exists \bar{x}$:

$$g(\bar{x}) < 0 \qquad\qquad L\bar{x} = b$$

- to show KKT-conditions, we formulate problem as:

$$\text{minimize} \quad \underbrace{f(x) + \iota(g(x) \le 0)}_{h_1(x)} + \underbrace{\iota(Lx = b)}_{h_2(x)}$$

and use subdifferential sum rule (and show that it may be used)

Fenchel or Lagrange duality?

- both approaches have their advantages
- Fenchel duality is more suitable for algorithms we will discuss