Chapter V
Optimality and duality

Skills to acquire

- Laagrange Duality
- Conjugate function and duality
- Algorithms on dual problems

We consider a general convex optimization problem

$$\min_{x\in\mathbb{R}^n} f(x) \tag{V.1}$$

Duality is associated to a particular formulation of the optimization problem, so that for instance making change of variables results in a different duality.

1 Lagrange Duality

We consider the convex optimization problem:

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & f(\boldsymbol{x}) \\ \text{subject to} & A\,\boldsymbol{x} = y, \\ & g(\boldsymbol{x}) \leq 0. \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{p\times n}$, and $g : \mathbb{R}^n \to \mathbb{R}^q$. This problem will be referred to as the *primal problem*. We write the associated Lagrangian.

$$\mathcal{L}(\boldsymbol{x}, \lambda, \beta) \;=\; f(\boldsymbol{x}) + \lambda^T(A\boldsymbol{x} - \boldsymbol{y}) + \mu^T(g(\boldsymbol{x}))$$

The dual objective function $q : \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R} \cup \{-\infty\}$ is defined to be

$$q(\lambda, \mu) \overset{\text{def.}}{=} \min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \lambda, \mu). \tag{V.2}$$

The dual problem is given by

$$q^\star = \max_{(\lambda,\mu)\in\mathrm{dom}(q)} q(\lambda, \mu). \tag{V.3}$$

Theorem 1.1 (convexity of the dual problem)**.** $\mathrm{dom}(q)$ *is a convex set and* q *is a concave function over* $\mathrm{dom}(q)$*.*

The following proposition is the so-called weak duality, which assert that values of the dual problems always lower bounds values of the primal one

Theorem 1.2 (Weak duality)**.** *Consider the primal problem its dual problem. Then*

$$q^* \leq f^*,$$

where q^**,* f^* *are the optimal dual and primal values respectively.*

Example 1.1.

$$\begin{aligned} \underset{x_1,\, x_2}{\text{minimize}} \quad & x_1^2 - 3x_2^2 \\ \text{subject to} \quad & x_1 = x_2^3 \end{aligned} \tag{V.4}$$

It is not difficult to show that the optimal solutions of the problem are $(1, 1)$*,* $(-1, -1)$ *with an optimal value of* $f^* = -2$*. The Lagrangian function is*

$$\mathcal{L}(x_1, x_2, \mu) = x_1^2 + \mu x_1 - 3x_2^2 - \mu x_2^3.$$

Obviously, for any $\mu \in \mathbb{R}$*,*

$$\min_{x_1, x_2} \mathcal{L}(x1, x_2, \mu) = -\infty,$$

and hence the dual optimal value is $q^* = -\infty$*, which is an extremely poor lower bound on the primal optimal value* $f^* = -2$*.*

The following fundamental theorem gives a sufficient condition (so-called qualification of the constraints) such that one actually has equality.

Theorem 1.3 (Strong duality)**.** *If the primal problem is convex and* $\exists x_0 \in \mathbb{R}^N, \quad Ax_0 = y \quad \text{and} \quad g(x_0) < 0$*, then* $q^\star = f^\star$ *(strong duality holds). Furthermore,* $x^\star$ *and* $(u^\star, v^\star)$ *are solutions of the dual problem verifying strong duality iff and only if*

$$Ax^\star = y, \quad g(x^\star) \leqslant 0, \quad u^\star \geqslant 0 \tag{V.5}$$

$$0 \in \partial f(x^\star) + A^* u^\star + \sum_i v_i^\star \partial g_i(x^\star) \tag{V.6}$$

$$\forall\, i, \quad u_i^\star g_i(x^\star) = 0 \tag{V.7}$$

2 Conjugate function and duality

To simplify and accelerate computation involving Lagrange duality, we introduce the *Legendre-Fenchel transform* which plays a similar role for convex function as the Fourier transform for signal or images.

2.1 Definitions

Definition 2.1. *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper function. The conjugate function (or Legendre-Fenchel transformation) is a function $f^* : \mathbb{R}^n \to \overline{\mathbb{R}}$, defined by*

$$f^*(\boldsymbol{y}) = \sup_{\boldsymbol{x} \in \mathrm{dom}(f)} \left\{ \boldsymbol{y}^T \boldsymbol{x} - f(\boldsymbol{x}) \right\}.$$

Proposition 2.1. *f^* is closed and convex (even when f is not).*

The conjugacy operation can be invoked twice resulting in the second conjugate:

$$f^{**}(\boldsymbol{x}) = \sup_{\boldsymbol{x} \in \mathbb{R}^n} \left(\boldsymbol{x}^T \boldsymbol{y} - f^*(y) \right)$$

Theorem 2.4. *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a function. Then*

$$f^{**}(\boldsymbol{x}) \leq f(\boldsymbol{x}), \quad \forall\, \boldsymbol{x} \in \mathbb{R}^n.$$

If $f : \mathbb{R}^n \to (-\infty, +\infty]$ is closed and convex, then $f^{**} = f$.

Remarque 2.1. *f^{**} is the convex envelop of f (i.e. the largest convex function smaller than f).*

Example 2.1 (Indicator function)**.** *$f = \mathbb{1}_S$, where $S \subset \mathbb{R}^n$ is nonempty.*

$$f^*(\boldsymbol{y}) = \sup_{\boldsymbol{x} \in S} \boldsymbol{y}^T \boldsymbol{x}.$$

Example 2.2.

$$f(x) = \alpha x + \beta; \quad g(\boldsymbol{x}) = \|\boldsymbol{x}\|; \quad h(x) = \frac{c}{2} x^2.$$

$$g^*(y) = \begin{cases} \beta, & y = \alpha \\ +\infty, & y \neq \alpha \end{cases}; \quad g^*(\boldsymbol{y}) = \begin{cases} 0, & \|\boldsymbol{y}\|_* \leq 1 \\ \\ +\infty, & \|\boldsymbol{y}\|_* > 1 \end{cases}; \quad h^*(y) = \frac{1}{2c} y^2.$$

Example 2.3 (Quadratic)**.** *$Q \succ 0$*

$$f(\boldsymbol{x}) = \frac{1}{2} \boldsymbol{x}^T Q \boldsymbol{x} + \mathbf{b}^T \boldsymbol{x} + c; \quad f^*(\boldsymbol{y}) = \frac{1}{2} (\boldsymbol{y} - \boldsymbol{b})^T Q^{-1} (\boldsymbol{y} - \boldsymbol{b}) - c$$

2.2 Calculus rules

Theorem 2.5 (Calculus rules)**.**

- *Separable sum: If $g(x_1, x_2) = f_1(x_1) + f_2(x_2)$, then*

$$g^*(y_1, y_2) = f_1^*(y_1) + f_2^*(y_2)$$

- *Scaling ($\alpha > 0$): If $g(x) = \alpha f(x)$, then*

$$g^*(y) = \alpha f^*(y/\alpha),$$

and if $g(\boldsymbol{x}) = \alpha f(\boldsymbol{x}/\alpha)$ then $g^*(\boldsymbol{y}) = \alpha g^*(\boldsymbol{y})$.

- *Summation: If $g(x) = f_1(x) + f_2(x)$, then*

$$g^*(y) = \inf_{z} \{f_1^*(z) + f_2^*(y - z)\}$$

- *Addition to affine function*

$$f(x) = g(x) + a^T x + b; \quad f^*(y) = g^*(y - a) - b$$

- *Infimal convolution*

$$(f \otimes g)(\boldsymbol{x}) \stackrel{\text{def.}}{=} \sup_{\boldsymbol{y}+\boldsymbol{y}'=\boldsymbol{x}} f(\boldsymbol{y}) + g(\boldsymbol{y}').$$

$$(f+g)^* = f \otimes g \quad \textit{and} \quad (f \otimes g)^* = f + g.$$

2.3 Properties

Theorem 2.6 (Fenchel's inequality)**.** *the definition implies that*

$$f(x) + f^*(y) \geqslant x^T y; \quad \forall\, x, y$$

Theorem 2.7 (Conjugate Subgradient Theorem)**.** *Let $f : \mathbb{R}^n \to (-\infty, +\infty]$ be a proper convex function.*

$$\forall\,(\boldsymbol{x}, \boldsymbol{y}), \boldsymbol{x}^T \boldsymbol{y} = f(\boldsymbol{x}) + f^*(\boldsymbol{y}) \quad \Leftrightarrow \quad \boldsymbol{y} \in \partial f(\boldsymbol{x}).$$

If, in addition f is closed, then

$$\boldsymbol{y} \in \partial f(\boldsymbol{x}) \iff \boldsymbol{x} \in \partial f^*(\boldsymbol{y})$$

The conjugate subgradient theorem can be written as the following.

Corollary 2.1. *If f is proper closed and convex then*

$$\partial f^*(\boldsymbol{y}) = \operatorname*{argmax}_{\boldsymbol{x}} \{\boldsymbol{y}^T \boldsymbol{x} - f(\boldsymbol{x})\}$$

$$\partial f(\boldsymbol{x}) = \operatorname*{argmax}_{\boldsymbol{y}} \{\boldsymbol{x}^T \boldsymbol{y} - f^*(\boldsymbol{y})\}$$

Theorem 2.8 (Fenchel's duality theorem)**.** *Let $f, g : \mathbb{R}^n \to (-\infty, +\infty]$ be proper convex functions. If* $\operatorname{relint}(\operatorname{dom}(f)) \cap \operatorname{relint}(\operatorname{dom}(g)) \neq \emptyset$*, then*

$$\min_{\boldsymbol{x}} \{f(\boldsymbol{x}) + g(\boldsymbol{x})\} = \max_{\boldsymbol{y}} \{-f^*(\boldsymbol{y}) - g^*(-\boldsymbol{y})\},$$

and the maximum in the right-hand problem is attained whenever it is finite.

Proposition 2.2 (Legendre transform and smoothness)**.** *One has*

$$\nabla f \textit{ is L-Lipschitz} \quad \iff \quad \nabla f^* \textit{ is } \mu\textit{-strongly convex.}$$

Assume f is closed and strongly convex with parameter $\mu > 0$. Then, f^* is defined for all y and is differentiable everywhere, with gradient

$$\nabla f^*(\boldsymbol{y}) = \operatorname*{argmax}_{\boldsymbol{x}} \left(y^T x - f(x)\right).$$

2.4 Fenchel-Rockafellar Duality

Very often the Lagrange dual can be expressed using the conjugate of the function f. We give here a particularly important example, which is often called Fenchel-Rockafellar Duality.

Consider the generic primal problem

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) + g(A\boldsymbol{x})$$

We rewrite the primal as:

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) + g(\boldsymbol{y}), \\ \text{subject to} \quad & A\boldsymbol{x} = \boldsymbol{y}. \end{aligned} \tag{V.8}$$

From Lagrange duality:

$$\inf_{x,y} \left(f(x) + z^T A x + g(y) - z^T y\right) = -f^*(-A^T z) - g^*(z) = q(\boldsymbol{z}).$$

We get the dual problem

$$\max_{\boldsymbol{z}} \; q(\boldsymbol{z}).$$

The next theorem gives primal-dual optimality conditions.

Theorem 2.9 (Fenchel-Rockafellar)**.** *Let $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, $g : \mathbb{R}^m \to \overline{\mathbb{R}}$ two convex functions and $A : \mathbb{R}^n \to \mathbb{R}^m$. If*

$$0 \in \operatorname{relint}(\operatorname{dom}(g) - A \operatorname{dom}(f))$$

the one has the following strong duality

$$\inf_{\boldsymbol{x}} \; (f(\boldsymbol{x}) + g(A\boldsymbol{x})) = \sup_{\boldsymbol{z}} \; - f^*(-A^T \boldsymbol{z}) - g^*(\boldsymbol{z})$$

Furthermore, (x, z) is a pair of optimal primal-dual solutions if and only if

$$-A^T z \in \partial f(x) \quad \textit{and} \quad Ax \in \partial g^*(z) \; (\textit{or } \boldsymbol{z} \in \partial g(Ax)). \tag{V.9}$$

Example 2.4 (Norm regularization)**.**

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) + \|A\boldsymbol{x} - \boldsymbol{b}\|$$

The dual problem can be written as

$$\begin{aligned} \underset{\boldsymbol{z}}{\text{maximize}} \quad & -\boldsymbol{b}^T \boldsymbol{z} - f^*(-A^T \boldsymbol{z}) \\ \text{subject to} \quad & \|\boldsymbol{z}\|_* \leqslant 1 \end{aligned} \tag{V.10}$$

3 Algorithms

3.1 Dual Subgradient method

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(\boldsymbol{x}) \\ \text{subject to} \quad & A\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

Dual Problem

$$\max_{\boldsymbol{z}} h(\boldsymbol{z}) = -f^*(-A^T\boldsymbol{z}) - \boldsymbol{b}^T\boldsymbol{z}$$

Subgradient of the Dual

- $\partial h(\boldsymbol{z}) = A\partial f^*(-A^T\boldsymbol{z}) - \boldsymbol{b}$
- $u \in \partial f^*(-A^T\boldsymbol{z}) \iff \boldsymbol{z} \in \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\{ f(\boldsymbol{x}) + \boldsymbol{z}^T A\boldsymbol{x} \right\}$

If f is strictly convex, f^* is differentiable. We get dual gradient ascent.

$$\begin{aligned} x^{k+1} &\in \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\{ f(\boldsymbol{x}) + (\boldsymbol{z}^k)^T A\boldsymbol{x} \right\}, \\ \boldsymbol{z}^{k+1} &= \boldsymbol{z}^k + \beta_k (A\boldsymbol{x}^{k+1} - \boldsymbol{b}) \end{aligned}$$

Dual Decomposition

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \sum_{i=1}^{n} f_i(\boldsymbol{x}_i) \\ \text{subject to} \quad & A\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

Algorithm

$$\begin{aligned} x_i^{k+1} &\in \underset{\boldsymbol{x}_i}{\operatorname{argmin}} \left\{ f_i(\boldsymbol{x}_i) + (\boldsymbol{z}^k)^T A_i\boldsymbol{x}_i \right\}, \\ \boldsymbol{z}^{k+1} &= \boldsymbol{z}^k + \beta_k (\sum_{i=1}^{n} A_i\boldsymbol{x}_i^{k+1} - \boldsymbol{b}) \end{aligned}$$

3.2 Forward-backward on the Dual

Problem 1

$$\begin{aligned} &\text{primal:} \ \min_{\boldsymbol{x}} f(\boldsymbol{x}) + g(\boldsymbol{x}), \\ &\text{dual:} \quad \max_{\boldsymbol{z}} \ -g^*(\boldsymbol{z}) - f^*(-\boldsymbol{z}) \end{aligned}$$

Use Moreau decomposition to simplify DR iteration

$$\begin{aligned} x_{k+1} &= \operatorname{prox}_f(y_k), \\ y_{k+1} &= x_{k+1} + \operatorname{prox}_{g^*}(2x_{k+1} - y_k) \end{aligned}$$

Make change of variables $z_k = x_k - y_k$:

$$\begin{aligned} x_{k+1} &= \operatorname{prox}_f(x_k - z_k), \\ z_{k+1} &= \operatorname{prox}_{g^*}(z_k + 2x_{k+1} - x_k) \end{aligned}$$

Problem 2: constrained convex problem

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & x \in V \end{aligned} \tag{V.11}$$

f closed and convex; V a subspace.

- Douglas–Rachford splitting with $g = \delta_V$

$$\begin{aligned} x_{k+1} &= \text{prox}_f(y_k), \\ y_{k+1} &= y_k + P_V(2x_{k+1} - y_k) - x_{k+1} \end{aligned}$$

- Primal-dual form

$$\begin{aligned} x_{k+1} &= \text{prox}_f(x_k - z_k), \\ z_{k+1} &= P_{V^\perp}(z_k + 2x_{k+1} - x_k) \end{aligned}$$

Problem 3: Equality constraints optimization problem

Consider the "primal" proble

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & Ax = b \end{aligned} \tag{V.12}$$

The dual problem is

$$\max_{\boldsymbol{z}} \; - b^T\boldsymbol{z} - f^*(-A^T\boldsymbol{z})$$

Dual gradient ascent algorithm

$$\begin{aligned} \hat{x} \; &= \underset{\boldsymbol{x}}{\text{argmin}} \; \left(f(\boldsymbol{x}) + \boldsymbol{z}^T A\boldsymbol{x}\right), \\ z^+ &= z + t(A\hat{x} - b) \end{aligned}$$

- Step one: compute a subgradient $\hat{x} \in \partial f^*(-A^T\boldsymbol{z})$
- Step two: compute a subgradient $b - A\hat{x}$ of $\boldsymbol{b}^T\boldsymbol{z} + f^*(-A^T\boldsymbol{z})$ at $\boldsymbol{z}$
- Step 3: Update $\boldsymbol{z}$: $\boldsymbol{z}^+ = \boldsymbol{z} + t(A\hat{x} - \boldsymbol{b})$.

This algorithm is of interest if calculation of $\hat{x}$ is inexpensive (for example, f is separable).

Problem 4: composite optimization problem

Consider the "primal" problem

$$\min_{\boldsymbol{x}} f(\boldsymbol{x}) + g(A\boldsymbol{x})$$

f_1 and f_2 have simple prox-operators.

The associated dual problem

$$\max_{\boldsymbol{z}} \; - f^*(-A^Tz) - g^*(z)$$

or

$$\min_{\boldsymbol{z}} g^*(z) + f^*(-A^Tz)$$

- proximal gradient update:

$$z^+ = \text{prox}_{tg^*}\left(\boldsymbol{z} + tA\nabla f^*(-A^T\boldsymbol{z})\right)$$

where

$$\nabla f^*(-A^T z) = \underset{\boldsymbol{x}}{\text{argmin}}\ \left(f(x) + z^T Ax\right)$$

Total variation image denoisng

A typical example, which was the one used by Antonin Chambolle to develop this class of method, is the total varation denoising. Consider that images $\boldsymbol{x} \in \mathbb{R}^N$ are represented by arrays X of size (n, n) and gradient vector fields $\nabla\boldsymbol{x} = \boldsymbol{u} \in \mathbb{R}^P$ are represented by arrays U size $(n, n, 2)$, where $P = N \times 2$. The total variation denoising model writes as

$$\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2 + \lambda\|\nabla\boldsymbol{x}\|_{1,2}$$

where $\|\cdot\|_{1,2}$ is the vectorial-ℓ^1 norm (also called $\ell^1 - \ell^2$ norm):

$$\|\boldsymbol{u}\|_{1,2} \overset{\text{def.}}{=} \sum_{i,j}(U_{i,j,1})^2 + (U_{i,j,2})^2.$$

The primal problem corresponds to minimizing $E(x) = f(x) + g(A(x))$ where

$$f(x) = \frac{1}{2}\|x - y\|^2 \quad \text{and} \quad g(u) = \lambda\|u\|_{1,2}.$$

The dual problem corresponds to minimzing $F(u) + G(u)$ where

$$F(u) = \frac{1}{2}\|y - A^*u\|^2 - \frac{1}{2}\|y\|^2 \quad \text{and} \quad G(u) = \iota_{\mathcal{C}}(u) \quad \text{where} \quad \mathcal{C} = \{u\ ;\ \|u\|_{\infty,2} \leqslant \lambda\},$$

and

$$\|u\|_{\infty,2} = \max_{i,j} \|u_{i,j}\|.$$

One can thus solves the ROF problem by computing

$$x^\star = y - A^*u^\star$$

where

$$u^\star \in \underset{\|u\|_{1,2}\leqslant\lambda}{\text{argmin}} \|y - A^*u\|.$$

One can compute explicitly the gradient of F:

$$\nabla F(u) = A(A^*u - y).$$

The proximal operator of G is the orthogonal projection on $\mathcal{C}$, which is obtained as

$$\text{prox}_{\gamma G}(u)_{i,j} = \frac{u_{i,j}}{\max(1, \|u_{i,j}\|/\lambda)}.$$

Note that it does not depends on γ. The gradient step size of the FB should satisfy

$$\gamma < \frac{2}{\|A^*A\|} = \frac{1}{4}.$$