12. Duality

- Lagrange dual problem
- strong duality
- saddle-point interpretation
- optimality conditions
- example: total variation de-noising

Primal problem

we consider the standard form optimization problem:

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \ldots, m \\ & h_j(\boldsymbol{x}) = 0, \quad j = 1, \ldots, p \end{array} \tag{12.1}$$

with variable $\boldsymbol{x} \in \mathbb{R}^n$ and nonempty domain

$$\mathcal{D} = \operatorname{dom} f \ \cap \ \bigcap_{i=1}^{m} \operatorname{dom} g_i \ \cap \ \bigcap_{j=1}^{p} \operatorname{dom} h_j$$

- problem (12.1) is referred to as the *primal problem*
- we let $p^\star$ denote the the optimal value of the primal problem
- the primal problem is not assumed to be convex unless explicitly stated

Duality

- *duality* provides a technique for transforming the primal problem into another related optimization problem (the dual problem)
- the dual problem is always a convex optimization problem (even when the primal is not)
- dual optimal value provides a lower bound on the optimal objective value of the primal
- certain dual problems may have a particular structure that makes them either solvable analytically, or amenable to certain algorithms that exploit the special structure of the dual
- in some cases we can recover a primal optimal solution from a dual optimal solution

Lagrangian

the *Lagrangian* $L : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}$ associated with problem (12.1) is

$$L(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i g_i(\boldsymbol{x}) + \sum_{j=1}^{p} \lambda_j h_j(\boldsymbol{x})$$

- Lagrangian domain is $\operatorname{dom} L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$
- μ_i is the *Lagrange multiplier* associated with the ith inequality constraint $g_i(\boldsymbol{x}) \leq 0$
- λ_j is the *Lagrange multiplier* associated with the jth equality constraint $h_j(\boldsymbol{x}) = 0$
- the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are called the *Lagrange multiplier vectors* or *dual variables* of problem (12.1)

Dual problem

Lagrange dual function: $\phi : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$:

$$\begin{aligned}\phi(\boldsymbol{\mu}, \boldsymbol{\lambda}) &= \min_{\boldsymbol{x} \in \mathcal{D}} \; L(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \\ &= \min_{\boldsymbol{x} \in \mathcal{D}} \; \left(f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i g_i(\boldsymbol{x}) + \sum_{j=1}^{p} \lambda_j h_j(\boldsymbol{x}) \right)\end{aligned}$$

- can take value $-\infty$ ($\operatorname{dom} \phi = \{(\boldsymbol{\mu}, \boldsymbol{\lambda}) \mid \phi(\boldsymbol{\mu}, \boldsymbol{\lambda}) > -\infty\}$)
- concave function since it is the minimum of affine functions in $(\boldsymbol{\mu}, \boldsymbol{\lambda})$

Lower bound on the optimal value: for $\boldsymbol{\mu} \geq \mathbf{0}$, $\boldsymbol{\lambda}$, we have

$$\phi(\boldsymbol{\mu}, \boldsymbol{\lambda}) \leq p^\star \tag{12.2}$$

Dual problem

$$\begin{array}{ll}\text{maximize} & \phi(\boldsymbol{\mu}, \boldsymbol{\lambda}) \\ \text{subject to} & \boldsymbol{\mu} \geq \mathbf{0}\end{array}$$

dual problem is convex and gives best lower bound for $p^\star$

Proof of lower bound: suppose that $\tilde{\boldsymbol{x}}$ is feasible, then for $\mu_i \geq 0$:

$$L(\tilde{\boldsymbol{x}}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = f(\tilde{\boldsymbol{x}}) + \sum_{i=1}^{m} \mu_i g_i(\tilde{\boldsymbol{x}}) + \sum_{j=1}^{p} \lambda_j h_j(\tilde{\boldsymbol{x}}) \leq f(\tilde{\boldsymbol{x}})$$

where the inequality holds since $\mu_i g_i(\tilde{\boldsymbol{x}}) \leq 0$ and $h_j(\tilde{\boldsymbol{x}}) = 0$; hence,

$$\phi(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \min_{\boldsymbol{x}} \; L(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \leq L(\tilde{\boldsymbol{x}}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \leq f(\tilde{\boldsymbol{x}})$$

since the above holds for any feasible $\tilde{\boldsymbol{x}}$, inequality (12.2) holds

Weak duality

dual problem finds $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ that gives the best lower bound in (12.2):

$$d^\star \leq p^\star \tag{12.3}$$

- the above property is called *weak duality*
- $p^\star - d^\star$ is called the *optimal duality gap*
- if the primal problem is unbounded below ($p^\star = -\infty$), then the dual problem is infeasible ($d^\star = -\infty$)
- if the dual problem is unbounded above ($d^\star = \infty$), then the primal problem is infeasible ($p^\star = \infty$)

Example 12.1

$$\begin{array}{ll} \text{minimize} & x^2 \\ \text{subject to} & x \geq 1 \end{array}$$

the solution is $x^\star = 1$ with optimal value $p^\star = 1$; the Lagrangian is

$$L(x, \mu) = x^2 + \mu(1 - x)$$

minimizing with respect to x: $\nabla_x L(x, \mu) = 2x - \mu = 0$ so $x = \frac{1}{2}\mu$ and the dual function is

$$\phi(\mu) = \min_x L(x, \mu) = L\big(\tfrac{1}{2}\mu, \mu\big) = (\tfrac{1}{2}\mu)^2 + \mu(1 - \tfrac{1}{2}\mu) = -\tfrac{1}{4}\mu^2 + \mu$$

dual function gives the immediate bound $\phi(\mu) \leq p^\star$ (*e.g.*, $\phi(0) = 0 \leq p^\star$)

the dual problem is

$$\underset{\mu \geq 0}{\text{maximize}} \quad -\tfrac{1}{4}\mu^2 + \mu$$

dual solution is $\mu^\star = 2$ with optimal value $d^\star = 1 = p^\star$

Example 12.2

$$\begin{array}{ll} \text{minimize} & x_1^2 - 3x_2^2 \\ \text{subject to} & x_1 = x_2^3 \end{array}$$

the optimal solutions are $(1, 1)$ and $(-1, -1)$ with $p^\star = -2$; the Lagrangian is

$$L(\boldsymbol{x}, \lambda) = x_1^2 - 3x_2^2 + \lambda(x_1 - x_2^3)$$

we have

$$\min_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda) = -\infty$$

hence, the dual optimal value is $d^\star = -\infty$, which gives a non useful bound on the primal optimal value

Form of dual problem

- the dual depends on the particular way in which the primal is represented
- it is often not possible to find a closed form expression for the dual problem

Example

$$\begin{array}{ll} \text{minimize} & e^x \\ \text{subject to} & x^2 \leq 1 \end{array}$$

the dual function is

$$\phi(\mu) = \min_x e^x + \mu(x^2 - 1)$$

the minimizer is the solution of the nonlinear equation $e^x + 2\mu x = 0$; in this case, the dual problem is

$$\underset{\mu \geq 0}{\text{maximize}} \quad e^x + \mu(x^2 - 1),$$

where x solves $e^x + 2\mu x = 0$

consider the equivalent representation of the previous problem:

$$\begin{array}{ll} \text{minimize} & e^x \\ \text{subject to} & -1 \leq x \leq 1 \end{array}$$

the dual function is

$$\phi(\boldsymbol{\mu}) = \min_x e^x + \mu_1(x-1) - \mu_2(x+1)$$

the minimizer satisfies $e^x + \mu_1 - \mu_2 = 0$, *i.e.*, $x = \log(\mu_2 - \mu_1)$; therefore, the dual function is

$$\begin{aligned} \phi(\boldsymbol{\mu}) &= \mu_2 - \mu_1 + \mu_1(\log(\mu_2 - \mu_1) - 1) - \mu_2(\log(\mu_2 - \mu_1) + 1) \\ &= -(\mu_2 - \mu_1)\log(\mu_2 - \mu_1) - 2\mu_1 \end{aligned}$$

with domain $\operatorname{dom}\phi = \{\boldsymbol{\mu} \mid \mu_2 > \mu_1\}$; hence, the dual problem is

$$\underset{\boldsymbol{\mu} \geq \mathbf{0}}{\text{maximize}} \quad -(\mu_2 - \mu_1)\log(\mu_2 - \mu_1) - 2\mu_1$$

Outline

- Lagrange dual problem
- **strong duality**
- saddle-point interpretation
- optimality conditions
- example: total variation de-noising

Strong duality

strong duality holds if

$$d^\star = p^\star \tag{12.4}$$

- does not hold in general
- guaranteed to hold if the problem is convex (*i.e.*, f, g_i are convex and $h(\boldsymbol{x}) = A\boldsymbol{x} - \boldsymbol{b}$) under *Slater's condition*

Slater's constraint qualification: there exists an $\hat{\boldsymbol{x}} \in \mathcal{D}$ such that

$$g_i(\hat{\boldsymbol{x}}) < 0, \quad i = 1, \ldots, m, \quad A\hat{\boldsymbol{x}} = \boldsymbol{b}$$

- implies that the dual optimal value is attained at some $(\boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$ and $d^\star = p^\star$ (assuming $d^\star > -\infty$)
- can be weakened if some g_i are affine, by only requiring the non-affine functions to hold with strict inequality

Example 12.3

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 + 2x_1 \\ \text{subject to} & x_1 + x_2 = 0 \end{array}$$

solution is $\boldsymbol{x}^\star = (-1/2, 1/2)$ and $p^\star = -1/2$; the Lagrangian is

$$L(\boldsymbol{x}, \lambda) = x_1^2 + x_2^2 + 2x_1 + \lambda(x_1 + x_2)$$

minimizing the Lagrangian with respect to $\boldsymbol{x}$ we get the solution

$$\tilde{\boldsymbol{x}} = \left(-1 - \tfrac{\lambda}{2}, -\tfrac{\lambda}{2} \right)$$

hence,

$$\begin{aligned} \phi(\lambda) &= L(\tilde{\boldsymbol{x}}, \lambda) \\ &= (-1 - \lambda/2)^2 + (-\lambda/2)^2 + 2(-1 - \lambda/2) + \lambda(-1 - \lambda) \\ &= -\tfrac{\lambda^2}{2} - \lambda - 1 \end{aligned}$$

the dual problem is thus

$$\text{maximize} \quad -\frac{\lambda^2}{2} - \lambda - 1$$

- note that $\phi(\lambda) \leq p^\star$ for any λ. For example,

$$\phi(0) = -1 \leq p^\star = -1/2$$

- the dual problem is solved at $\lambda^\star = -1$ and at the optimal solution, we have

$$\phi(\lambda^\star) = -1/2 = p^\star$$

  hence, strong duality holds
- Slater's conditions is satisfied since the problem is feasible and we only have equality constraint(s)

Dual of inequality form LP

$$\begin{array}{ll} \text{minimize} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & A\boldsymbol{x} \le \boldsymbol{b} \end{array}$$

the Lagrangian is

$$L(\boldsymbol{x}, \boldsymbol{\mu}) = \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{\mu}^T(A\boldsymbol{x} - \boldsymbol{b}) = -\boldsymbol{b}^T\boldsymbol{\mu} + (\boldsymbol{c} + A^T\boldsymbol{\mu})^T\boldsymbol{x}$$

the dual function is

$$\phi(\boldsymbol{\mu}) = -\boldsymbol{b}^T\boldsymbol{\mu} + \min_{\boldsymbol{x}} (\boldsymbol{c} + A^T\boldsymbol{\mu})^T\boldsymbol{x} = \begin{cases} -\boldsymbol{b}^T\boldsymbol{\mu} & \text{if } A^T\boldsymbol{\mu} + \boldsymbol{c} = \boldsymbol{0} \\ -\infty & \text{otherwise} \end{cases}$$

hence, the dual problem (with $\operatorname{dom}\phi$ expressed as constraints) is

$$\begin{array}{ll} \underset{\boldsymbol{\mu} \ge \boldsymbol{0}}{\text{maximize}} & -\boldsymbol{b}^T\boldsymbol{\mu} \\ \text{subject to} & A^T\boldsymbol{\mu} + \boldsymbol{c} = \boldsymbol{0} \end{array}$$

we have $p^\star = d^\star$ if $A\boldsymbol{x} < \boldsymbol{b}$ for some $\boldsymbol{x}$; in fact, strong duality always holds for LPs except when primal or dual are infeasible

Dual of least-norm problem

$$\begin{array}{ll}\text{minimize} & \|\boldsymbol{x}\|^2 \\ \text{subject to} & A\boldsymbol{x} = \boldsymbol{b}\end{array}$$

the Lagrangian is

$$L(\boldsymbol{x}, \boldsymbol{\lambda}) = \|\boldsymbol{x}\|^2 + \boldsymbol{\lambda}^T(A\boldsymbol{x} - \boldsymbol{b})$$

the Lagrangian is a convex function in $\boldsymbol{x}$, hence all minimizers satisfy:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}) = 2\boldsymbol{x} + A^T\boldsymbol{\lambda} = \boldsymbol{0},$$

which gives $\boldsymbol{x}(\boldsymbol{\lambda}) = -\frac{1}{2}A^T\boldsymbol{\lambda}$; hence, the dual problem is

$$\text{maximize} \quad \phi(\boldsymbol{\lambda}) = L(-\tfrac{1}{2}A^T\boldsymbol{\lambda}, \boldsymbol{\lambda}) = -\tfrac{1}{4}\boldsymbol{\lambda}^T A A^T \boldsymbol{\lambda} - \boldsymbol{b}^T\boldsymbol{\lambda}$$

since the primal problem has only linear equality constraints, Slater's condition is simply primal feasibility ($\boldsymbol{b} \in \operatorname{range} A$)

Dual of strictly convex quadratic program

$$\begin{array}{ll} \text{minimize} & \boldsymbol{x}^T Q \boldsymbol{x} \\ \text{subject to} & A\boldsymbol{x} \leq \boldsymbol{b} \end{array}$$

where $Q > 0$; the Lagrangian

$$L(\boldsymbol{x}, \boldsymbol{\mu}) = \boldsymbol{x}^T Q \boldsymbol{x} + \boldsymbol{\mu}^T (A\boldsymbol{x} - \boldsymbol{b})$$

is convex in $\boldsymbol{x}$; hence, it is minimized with respect to $\boldsymbol{x}$ if and only if

$$\nabla_x L(\boldsymbol{x}, \boldsymbol{\mu}) = 2Q\boldsymbol{x} + A^T \boldsymbol{\mu} = \boldsymbol{0} \Longrightarrow \boldsymbol{x} = -\tfrac{1}{2} Q^{-1} A^T \boldsymbol{\mu}$$

plug in L, we have

$$\phi(\boldsymbol{\mu}) = L(-\tfrac{1}{2} Q^{-1} A^T \boldsymbol{\mu}, \boldsymbol{\mu}) = -\tfrac{1}{4} \boldsymbol{\mu}^T A Q^{-1} A^T \boldsymbol{\mu} - \boldsymbol{b}^T \boldsymbol{\mu}$$

hence, the dual problem is

$$\begin{array}{ll} \text{maximize} & -\tfrac{1}{4} \boldsymbol{\mu}^T A Q^{-1} A^T \boldsymbol{\mu} - \boldsymbol{b}^T \boldsymbol{\mu} \\ \text{subject to} & \boldsymbol{\mu} \geq \boldsymbol{0} \end{array}$$

we have $p^\star = d^\star$ if $A\boldsymbol{x} < \boldsymbol{b}$ for some $\boldsymbol{x}$; in fact, strong duality always holds for this problem

Outline

- Lagrange dual problem
- strong duality
- **saddle-point interpretation**
- optimality conditions
- example: total variation de-noising

Max-min characterization

assume that there are no equality constraints; then

$$\max_{\boldsymbol{\mu} \geq \mathbf{0}} L(\boldsymbol{x}, \boldsymbol{\mu}) = \max_{\boldsymbol{\mu} \geq \mathbf{0}} \left(f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i g_i(\boldsymbol{x}) \right) = \begin{cases} f(\boldsymbol{x}) & \text{if } g_i(\boldsymbol{x}) \leq 0, \\ \infty & \text{otherwise} \end{cases}$$

this means that we can write $p^\star$ as

$$p^\star = \min_{\boldsymbol{x}} \max_{\boldsymbol{\mu} \geq \mathbf{0}} L(\boldsymbol{x}, \boldsymbol{\mu})$$

from the definition of the dual function, we have

$$d^\star = \max_{\boldsymbol{\mu} \geq \mathbf{0}} \min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\mu})$$

hence, we can write weak duality as

$$\max_{\boldsymbol{\mu} \geq \mathbf{0}} \min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\mu}) \leq \min_{\boldsymbol{x}} \max_{\boldsymbol{\mu} \geq \mathbf{0}} L(\boldsymbol{x}, \boldsymbol{\mu}) \tag{12.5}$$

Max-min inequality and saddle-point

Max-min inequality: inequality (12.6) does not depend on the property of L; for any function $J : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, we have

$$\max_{\boldsymbol{z} \in \mathcal{Z}} \min_{\boldsymbol{w} \in \mathcal{W}} J(\boldsymbol{w}, \boldsymbol{z}) \leq \min_{\boldsymbol{w} \in \mathcal{W}} \max_{\boldsymbol{z} \in \mathcal{Z}} J(\boldsymbol{w}, \boldsymbol{z})$$

where $\mathcal{W} \subseteq \mathbb{R}^n, \mathcal{Z} \subseteq \mathbb{R}^m$

Saddle-point: a point $(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{z}})$ is called a *saddle-point* for J over $\mathcal{W}, \mathcal{Z}$ if

$$J(\tilde{\boldsymbol{w}}, \boldsymbol{z}) \leq J(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{z}}) \leq J(\boldsymbol{w}, \tilde{\boldsymbol{z}})$$

for all $\boldsymbol{w} \in \mathcal{W}, \boldsymbol{z} \in \mathcal{Z}$; this means that

$$J(\tilde{\boldsymbol{w}}, \tilde{\boldsymbol{z}}) = \min_{\boldsymbol{w} \in \mathcal{W}} J(\boldsymbol{w}, \tilde{\boldsymbol{z}}) = \max_{\boldsymbol{z} \in \mathcal{Z}} J(\tilde{\boldsymbol{w}}, \boldsymbol{z})$$

Lagrangian saddle-point

for any $\boldsymbol{x}, \boldsymbol{\mu}$, we have

$$d^\star = \max_{\boldsymbol{\mu} \geq \mathbf{0}} \min_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\mu}) \leq \min_{\boldsymbol{x}} \max_{\boldsymbol{\mu} \geq \mathbf{0}} L(\boldsymbol{x}, \boldsymbol{\mu}) = p^\star \tag{12.6}$$

- if strong duality holds at optimal primal and dual points $\boldsymbol{x}^\star$ and $\boldsymbol{\mu}^\star$, then they form a saddle-point for the Lagrangian
- the converse is also true: If $(\boldsymbol{x}^\star, \boldsymbol{\mu}^\star)$ is a saddle-point of the Lagrangian, then $\boldsymbol{x}^\star$ is primal optimal, $\boldsymbol{\mu}^\star$ is dual optimal, and the optimal duality gap is zero
- strong duality means that the order of the minimization over $\boldsymbol{x}$ and the maximization over $\boldsymbol{\mu} \geq \mathbf{0}$ can be switched without affecting the result

Outline

- Lagrange dual problem
- strong duality
- saddle-point interpretation
- **optimality conditions**
- example: total variation de-noising

Strong duality implication

suppose that strong duality holds and $\boldsymbol{x}^\star$ is a primal optimal and $(\boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$ is a dual optimal point, then we have

$$\begin{aligned} f(\boldsymbol{x}^\star) = \phi(\boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star) &= \min_{\boldsymbol{x}} \Big(f(\boldsymbol{x}) + \sum_{i=1}^m \mu_i^\star g_i(\boldsymbol{x}) + \sum_{j=1}^p \lambda_j^\star h_j(\boldsymbol{x}) \Big) \\ &\le f(\boldsymbol{x}^\star) + \sum_{i=1}^m \mu_i^\star g_i(\boldsymbol{x}^\star) + \sum_{j=1}^p \lambda_j^\star h_j(\boldsymbol{x}^\star) \\ &\le f(\boldsymbol{x}^\star) \end{aligned}$$

we conclude that the inequalities hold with equality; thus, $\sum_{i=1}^m \mu_i^\star g_i(\boldsymbol{x}^\star) = 0$

Complementary slackness

- since each term in the sum $\sum_{i=1}^m \mu_i^\star g_i(\boldsymbol{x}^\star) = 0$ is nonpositive; we conclude that

$$\mu_i^\star g_i(\boldsymbol{x}^\star) = 0, \quad i = 1, \ldots, m$$

- this condition is known as *complementary slackness*

Optimality conditions

suppose that strong duality holds and $\boldsymbol{x}^\star$ and $(\boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$ are optimal solutions of the primal and dual problems, then

$$\boldsymbol{x}^\star \in \operatorname*{argmin}_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$$
$$\mu_i^\star g_i(\boldsymbol{x}^\star) = 0, \quad i = 1, \ldots, m$$

- first condition implies that $\boldsymbol{x}^\star$ is a minimizer $L(\boldsymbol{x}, \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$
- the Lagrangian $L(\boldsymbol{x}, \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$ can have other minimizers; $\boldsymbol{x}^\star$ is simply a minimizer
- functions are not necessarily differentiable
- for differentiable functions, we recover KKT conditions

KKT conditions

suppose that strong duality holds and $\boldsymbol{x}^\star$ and $(\boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$ are optimal solutions of the primal and dual problems, then

$$\begin{aligned} \nabla_{\boldsymbol{x}} L(\boldsymbol{x}^\star, \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star) &= \mathbf{0} \\ g_i(\boldsymbol{x}^\star) &\leq 0, \quad i = 1, \dots, m \\ h_j(\boldsymbol{x}^\star) &= 0, \quad j = 1, \dots, p \\ \mu_i^\star &\geq 0, \quad i = 1, \dots, m \\ g_i(\boldsymbol{x}^\star)\mu_i^\star &= 0, \quad i = 1, \dots, m \end{aligned}$$

this means that if strong-duality holds, then any pair of primal and dual optimal points must satisfy the KKT conditions

KKT conditions for convex problems

- for convex problems, the KKT conditions are sufficient for optimality
- if f, g_i are convex and h_j are affine, and $\boldsymbol{x}^\star, \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star$ are any points that satisfy the KKT conditions, then $\boldsymbol{x}^\star$ and $(\boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$ are primal and dual optimal, with zero duality gap

Proof: since $L(\boldsymbol{x}, \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$ is convex in $\boldsymbol{x}$; the first KKT condition implies that $\boldsymbol{x}^\star$ minimizes $L(\boldsymbol{x}, \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$ over $\boldsymbol{x}$; we conclude that

$$\begin{aligned} g(\boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star) &= L(\boldsymbol{x}^\star, \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star) \\ &= f(\boldsymbol{x}^\star) + \sum_{i=1}^{m} \mu_i^\star g_i(\boldsymbol{x}^\star) + \sum_{j=1}^{p} \lambda_j^\star h_j(\boldsymbol{x}^\star) = f(\boldsymbol{x}^\star) \end{aligned}$$

hence, strong duality holds, and thus, $\boldsymbol{x}^\star$ and $(\boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$ are primal and dual optimal

Necessary and sufficient conditions

- when a convex problem satisfies Slater's condition, then the KKT conditions provide necessary and sufficient conditions for optimality
- this is because Slater's condition implies that the optimal duality gap is zero and the dual optimum is attained
- hence, $\boldsymbol{x}$ is optimal if and only if there are $(\boldsymbol{\mu}, \boldsymbol{\lambda})$ that, together with $\boldsymbol{x}$, that satisfy the KKT conditions

Recovering primal solution from dual

Unique minimizer: suppose that $L(\boldsymbol{x}, \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$ has a unique minimizer $\boldsymbol{x}^\star$ (*e.g.*, $L(\boldsymbol{x}, \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$ is strictly convex) then $\boldsymbol{x}^\star$ solves

$$\nabla L(\boldsymbol{x}, \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star) = \boldsymbol{0}$$

- $\boldsymbol{x}^\star$ of L is either primal feasible; hence, it is the primal-optimal solution
- or it is not primal feasible and no primal-optimal solution exists

Multiple minimizers

- if $L(\boldsymbol{x}, \boldsymbol{\mu}^\star, \boldsymbol{\lambda}^\star)$ has multiple global minimizers, then it is not guaranteed that any global minimizer of L is a primal-optimal solution
- what is guaranteed is that the primal-optimal solution $\boldsymbol{x}^\star$ is among the global minimizers of L

Example 12.4

$$\begin{array}{ll} \text{minimize} & (x_1+3)^2 + x_2^2 \\ \text{subject to} & x_1^2 \le x_2 \end{array}$$

problem is convex with strictly convex objective; thus, it has a unique solution; the Lagrangian

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

is convex over $\boldsymbol{x}$ for any $\mu \ge 0$; a minimizer of L over $\boldsymbol{x}$ must satisfy:

$$\frac{\partial L}{\partial x_1} = 2(x_1+3) + 2\mu x_1 = 0 \Longrightarrow x_1 = -3/(1+\mu)$$

$$\frac{\partial L}{\partial x_2} = 2x_2 - \mu = 0 \Longrightarrow x_2 = \mu/2$$

hence, the dual function is

$$
\begin{aligned}
\phi(\mu) &= (-3/(1+\mu)+3)^2 + (\mu/2)^2 + \mu((-3/(1+\mu))^2 - \mu/2) \\
&= \frac{9\mu}{1+\mu} - \frac{\mu^2}{4}
\end{aligned}
$$

and the dual problem is

$$
\underset{\mu \geq 0}{\text{maximize}} \quad \frac{9\mu}{1+\mu} - \frac{\mu^2}{4}
$$

The derivative of ϕ is

$$
\phi'(\mu) = \frac{9}{(1+\mu)^2} - \frac{\mu}{2}
$$

solving for $\phi'(\mu) = 0$, we get the unique optimal dual solution $\mu^\star = 2$ and $d^\star = 5$; using this dual solution, the primal solution is

$$
\boldsymbol{x}^\star = (-3/(1+\mu^\star), \mu^\star/2) = (-1, 1)
$$

and the optimal value is $p^\star = 5 = d^\star$

Example 12.5

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\sum_{i=1}^{n}(x_i - c_i)^2 \\ \text{subject to} & \sum_{i=1}^{n} a_i x_i = b \end{array}$$

where $a_i, c_i, b \in \mathbb{R}$ are given

The Lagrangian is

$$\begin{aligned} L(\boldsymbol{x}, \lambda) &= \frac{1}{2}\sum_{i=1}^{n}(x_i - c_i)^2 + \lambda\Big(\sum_{i=1}^{n} a_i x_i - b\Big) \\ &= -b\lambda + \sum_{i=1}^{n}\big(\tfrac{1}{2}(x_i - c_i)^2 + \lambda a_i x_i\big), \end{aligned}$$

which is also separable in x_i

the dual function is

$$\begin{aligned}\phi(\lambda) &= -b\lambda + \sum_{i=1}^{n} \min_{x_i} \left(\tfrac{1}{2}(x_i - c_i)^2 + \lambda a_i x_i\right) \\ &= -b\lambda - \sum_{i=1}^{n} \left(\tfrac{1}{2}a_i^2\lambda^2 - a_i c_i \lambda\right)\end{aligned}$$

where the minimum is achieved at $x_i = c_i - a_i\lambda$

the dual problem is thus

$$\underset{\lambda}{\text{maximize}} \quad -b\lambda - \sum_{i=1}^{n} \left(\tfrac{1}{2}a_i^2\lambda^2 - a_i c_i \lambda\right)$$

since the dual problem is unconstrained and concave, the optimal solution must satisfy

$$\phi'(\lambda) = -b - \lambda \sum_{i=1}^{n} a_i^2 + \sum_{i=1}^{n} a_i c_i = 0$$

hence,

$$\lambda^\star = -\frac{b - \sum_{i=1}^{n} a_i c_i}{\sum_{i=1}^{n} a_i^2}$$

and we can recover the primal by the formula

$$x_i^\star = c_i - a_i \lambda^\star = c_i + a_i \frac{b - \sum_{i=1}^{n} a_i c_i}{\sum_{i=1}^{n} a_i^2}, \quad i = 1, \ldots, n$$

Outline

- Lagrange dual problem
- strong duality
- saddle-point interpretation
- optimality conditions
- **example: total variation de-noising**

Signal de-noising

- a noisy version of signal $\boldsymbol{x}$ is denoted by $\boldsymbol{y}$, given by

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{v}$$

  where $\boldsymbol{v} \in \mathbb{R}^n$ is an unknown noise vector
- our objective is to retrieve $\boldsymbol{x}$ by solving

$$\text{minimize} \quad \|\boldsymbol{x} - \boldsymbol{y}\|^2 + \rho r_{\text{tv}}(\boldsymbol{x})$$

  where $\rho > 0$ and the regularizer r_{tv} refers to the total variation function defined as:

$$r_{\text{tv}}(\boldsymbol{x}) = \sum_{i=1}^{n-1} |x_i - x_{i+1}| = \|R\boldsymbol{x}\|_1$$

  and R denotes the $(n-1) \times n$ smoothing matrix:

$$R = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times n}$$

- we we have not yet explored how to manage general non-smooth terms in the objective function, by considering the dual problem, we can bypass the non-smooth term r_{tv}

- to derive the dual, we recast the problem as an equivalent constrained one:

$$\begin{array}{ll} \text{minimize} & \|\boldsymbol{x}-\boldsymbol{y}\|^2 + \rho\|\boldsymbol{z}\|_1 \\ \text{subject to} & \boldsymbol{z} = R\boldsymbol{x} \end{array}$$

  introducing the variable $\boldsymbol{z} \in \mathbb{R}^{(n-1)}$

- the associated Lagrangian is:

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\lambda}) &= \|\boldsymbol{x}-\boldsymbol{y}\|^2 + \rho\|\boldsymbol{z}\|_1 + \boldsymbol{\lambda}^T(R\boldsymbol{x} - \boldsymbol{z}) \\ &= \|\boldsymbol{x}-\boldsymbol{y}\|^2 + \boldsymbol{\lambda}^T R\boldsymbol{x} + \rho\|\boldsymbol{z}\|_1 - \boldsymbol{\lambda}^T\boldsymbol{z} \end{aligned}$$

observing that the Lagrangian is separable in terms of $\boldsymbol{x}$ and $\boldsymbol{z}$, the minimization concerning $\boldsymbol{x}$ yields:

$$\begin{aligned}\boldsymbol{x}^\star &= \underset{\boldsymbol{x}}{\operatorname{argmin}}\, L(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\lambda}) \\ &= \underset{\boldsymbol{x}}{\operatorname{argmin}}\, \|\boldsymbol{x} - \boldsymbol{y}\|^2 + \boldsymbol{\lambda}^T R\boldsymbol{x} \\ &= \boldsymbol{y} - \tfrac{1}{2} R^T \boldsymbol{\lambda}\end{aligned}$$

substituting this result, we get:

$$\begin{aligned}L(\boldsymbol{x}^\star, \boldsymbol{z}, \boldsymbol{\lambda}) &= \|\boldsymbol{y} - \tfrac{1}{2}R^T\boldsymbol{\lambda} - \boldsymbol{y}\|^2 + \boldsymbol{\lambda}^T R(\boldsymbol{y} - \tfrac{1}{2}R^T\boldsymbol{\lambda}) + \rho\|\boldsymbol{z}\|_1 - \boldsymbol{\lambda}^T\boldsymbol{z} \\ &= -\tfrac{1}{4}\boldsymbol{\lambda}^T R R^T \boldsymbol{\lambda} + \boldsymbol{\lambda}^T R\boldsymbol{y} + \rho\|\boldsymbol{z}\|_1 - \boldsymbol{\lambda}^T\boldsymbol{z}\end{aligned}$$

to minimize with respect to $\boldsymbol{z}$, we must address:

$$\min_{\boldsymbol{z}} \quad \rho\|\boldsymbol{z}\|_1 - \boldsymbol{\lambda}^T\boldsymbol{z}$$

considering each component, we realize:

$$\min_{z_i} \quad \rho|z_i| - \lambda_i z_i = \begin{cases} 0, & \text{if } |\lambda_i| \leq \rho, \\ -\infty, & \text{otherwise} \end{cases}$$

consequently, the dual function becomes:

$$\phi(\boldsymbol{\lambda}) = \min_{\boldsymbol{x},\boldsymbol{z}} L(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\lambda}) = \begin{cases} -\frac{1}{4}\boldsymbol{\lambda}^T R R^T \boldsymbol{\lambda} + \boldsymbol{\lambda}^T R \boldsymbol{y}, & \text{if } ||\boldsymbol{\lambda}||_\infty \leq \rho, \\ -\infty, & \text{otherwise} \end{cases}$$

thus, our dual problem becomes:

$$\begin{array}{ll} \text{maximize} & -\frac{1}{4}\boldsymbol{\lambda}^T R R^T \boldsymbol{\lambda} + \boldsymbol{\lambda}^T R \boldsymbol{y} \\ \text{subject to} & ||\boldsymbol{\lambda}||_\infty \leq \rho, \end{array}$$

where the constraints form a simple box constraint:

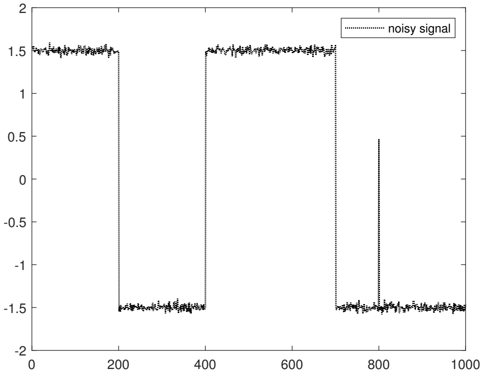
$$\mathcal{C} = \{\boldsymbol{\lambda} \in \mathbb{R}^{(n-1)} \mid -\rho \leq \lambda_i \leq \rho,\, i = 1, 2, \ldots, n-1\}$$

we can solve the problem using the projected gradient descent; the projection onto $\mathcal{C}$, denoted by $\Pi[\boldsymbol{\lambda}]$, has components:

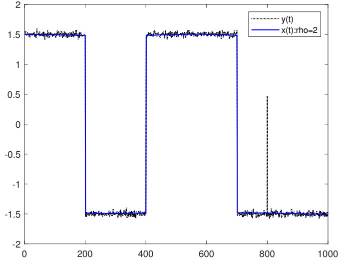
$$\Pi[\boldsymbol{\lambda}]_i = \frac{\rho \lambda_i}{\max\{|\lambda_i|, \rho\}}$$

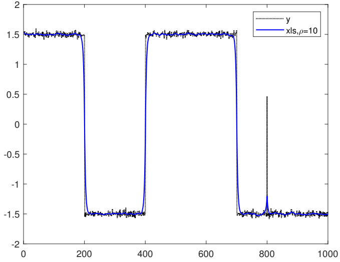
once we get $\boldsymbol{\lambda}^\star$, then $\boldsymbol{x}^\star = \boldsymbol{y} - \frac{1}{2} R^T \boldsymbol{\lambda}^\star$

Numerical example



example: total variation de-noising





the Total Variation (TV) denoising effectively captures jump discontinuities and noise spikes, an outcome not achieved by the least-squares reconstruction

References and further readings

- Stephen Boyd and Lieven Vandenberghe. *Convex Optimization,* Cambridge University Press, 2004, chapter 5.1, 5.2, 5.4, and 5.7.
- Amir Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB,* SIAM, 2014, chapter 12.