11. Duality

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

Primal problem

we consider the standard form optimization problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & g_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_j(x) = 0, \quad j = 1, \ldots, p \end{array} \tag{11.1}$$

with variable $x \in \mathbb{R}^n$ 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 (11.1) 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, called the dual problem
- dual problem is always a convex problem (even when the primal is not)
- dual optimal value provides a lower bound on the primal optimal value
- dual problems may have a particular structure that makes 'easier' to solve
- in some cases we can recover a primal solution from a dual solution

Lagrangian

the *Lagrangian* $L : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^m \to \mathbb{R}$ associated with problem (11.1) is

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

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

Dual problem

Lagrange dual function: $\phi : \mathbb{R}^m \times \mathbb{R}^p \longrightarrow \mathbb{R}$ (using $\min$ instead of $\inf$):

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

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

Lower bound on the optimal value: for $\mu \geq 0$, λ, we have $\phi(\mu, \lambda) \leq p^\star$

Proof: for feasible $\tilde{x}$ and $\mu_i \geq 0$:

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

since the above holds for any feasible $\tilde{x}$, we have$\phi(\mu, \lambda) \leq p^\star$

Dual problem

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

- gives best lower bound on $p^\star$
- a convex optimization problem; optimal value denoted by $d^\star$
- often simplified by making implicit constraint $(\mu, \lambda) \in \operatorname{dom} \phi$ explicit
- μ, λ are dual feasible if $\mu \geq 0$ and $(\mu, \lambda) \in \operatorname{dom} \phi$
- $d^\star = -\infty$ if problem is infeasible; $d^\star = +\infty$ if unbounded above

Weak duality

$$d^\star \leq p^\star$$

- the above property is called *weak duality*
- can be used to find nontrivial lower bounds for difficult problems
- $p^\star - d^\star$ is called the *optimal duality gap*
- if primal is unbounded below ($p^\star = -\infty$), then the dual is infeasible ($d^\star = -\infty$)
- if dual is unbounded above ($d^\star = \infty$), then the primal is infeasible ($p^\star = \infty$)

Example

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

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

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

with respect to x: $\nabla_x L(x, \mu) = 2x - \mu = 0$ so $x = \frac{1}{2}\mu$

- 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

$$
\begin{array}{ll}
\text{minimize} & x_1 x_2 \\
\text{subject to} & (1/2)x^T x \le 1 \\
& x \ge 0
\end{array}
$$

- solution is $x^\star = 0$, with the primal optimal value $p^\star = f(x^\star) = 0$
- with $\mu = (\mu_1, \bar{\mu})$, $\mu_1 \in \mathbb{R}$ and $\bar{\mu} \in \mathbb{R}^2$, the Lagrangian is:

$$
\begin{aligned}
L(x, \mu) &= (1/2)x^T \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} x + \mu^T \begin{bmatrix} x^T x/2 - 1 \\ -x \end{bmatrix} \\
&= (1/2)x^T \begin{bmatrix} \mu_1 & 1 \\ 1 & \mu_1 \end{bmatrix} x - \mu_1 - \bar{\mu}^T x
\end{aligned}
$$

Hessian is positive semidefinite only if $\mu_1 \ge 1$

- if $\mu_1 < 1$ or $\bar{\mu} \neq 0$, L unbounded below, so we exclude such μ from the dual
- for μ with $\mu_1 \geq 1$ and $\bar{\mu} = 0$, the dual function is:

$$\phi(\mu) = \min_{x \in \mathbb{R}^2} \quad (1/2)x^T \begin{bmatrix} \mu_1 & 1 \\ 1 & \mu_1 \end{bmatrix} x - \mu_1 = -\mu_1$$

  any μ with $\mu_1 \geq 1$ and $\bar{\mu} = 0$ is feasible (*i.e.*, $\mu \geq 0$)

- the dual problem is:

$$\begin{array}{ll} \text{maximize} & -\mu_1 \\ \text{subject to} & \mu_1 \geq 1 \\ & \bar{\mu} = 0 \end{array}$$

  solution is $\mu^\star = (1, 0, 0)$, with $d^\star = \phi(\mu^\star) = -1$

- since $p^\star = 0 > -1 = d^\star$, the duality gap is $f(x^\star) - \phi(\mu^\star) = 1$

Example

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

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

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

- minimizing we see the dual take value

$$\min_x L(x, \lambda) = -\infty$$

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

Form of dual problem

- the dual depends on the particular way in which the primal is represented
- reformulating the primal problem can be useful when the dual is difficult to derive, or uninteresting
- it is often not possible to find a closed form expression for the dual problem

Common reformulations

- introduce new variables and equality constraints
- make explicit constraints implicit or vice versa
- transform objective or constraint functions

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:

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

- the dual function is

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

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

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

  with domain $\mathrm{dom}\,\phi = \{\mu \mid \mu_2 > \mu_1\}$

- hence, the dual problem is

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

Outline

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

Strong duality

strong duality holds if $d^\star = p^\star$

- does not hold in general
- guaranteed to hold if the problem is convex under *Slater's condition*

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

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

- guarantees $d^\star = p^\star$
- implies the dual optimal value is attained at some $(\mu^\star, \lambda^\star)$
- can be weakened by only requiring the non-affine g_i to hold with strict inequality
- there exist many other types of constraint qualifications

Example

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

- solution is $x^\star = (-1/2, 1/2)$ and $p^\star = -1/2$
- minimizing the Lagrangian

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

with respect to x we get the solution

$$\tilde{x} = \big(-1 - \tfrac{\lambda}{2}, -\tfrac{\lambda}{2} \big)$$

- so the dual function is

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

- the dual problem is thus

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

- $\phi(\lambda) \leq p^\star$ 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

Dual of inequality form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq b \end{array}$$

the Lagrangian is

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

the dual function is

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

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

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

strong duality always holds for LPs except when primal or dual are infeasible

Dual of least-norm problem

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

the Lagrangian is

$$L(x, \lambda) = \|x\|^2 + \lambda^T(Ax - b)$$

the Lagrangian is a convex function in x, hence all minimizers satisfy:

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

hence, the dual problem is

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

since there is no inequalities, Slater condition is just primal feasibility ($b \in \operatorname{range} A$)

Dual of strictly convex quadratic program

for $Q \succ 0$, consider

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

the Lagrangian is

$$L(x, \mu) = x^T Q x + \mu^T (Ax - b)$$

since L is convex in x, it is minimized with respect to x if and only if

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

plug in L, we have

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

the dual problem is

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

strong duality always holds for this problem

Outline

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

Optimality conditions

if strong duality holds, $x^\star$ is primal optimal, and $(\mu^\star, \lambda^\star)$ is dual optimal, then:

1. $g_i(x^\star) \leq 0$ for $i = 1, \ldots, m$ and $h_i(x) = 0$ for $i = 1, \ldots, p$
2. $\mu^\star \geq 0$
3. $f(x^\star) = g(\mu^\star, v^\star)$

conversely, these three conditions imply optimality of $x^\star$, $(\mu^\star, \lambda^\star)$, and strong duality

next, we replace condition 3 with two equivalent conditions that are easier to use

Complementary slackness

if strong duality holds and $x^\star$ is primal optimal and $(\mu^\star, \lambda^\star)$ is dual optimal, then

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

holds if and only if the two inequalities hold with equality:

- first inequality: $x^\star$ minimizes $L(x, \mu, \lambda)$ over $x \in \mathcal{D}$
- second inequality: each term in the sum $\sum_{i=1}^m \mu_i^\star g_i(x^\star) = 0$ is nonpositive, so

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

  i.e., $\mu_i > 0 \Rightarrow g_i(x) = 0$ and $g_i(x) < 0 \Rightarrow \mu_i = 0$

  this condition is known as *complementary slackness*

Optimality conditions

if strong duality holds, $x^\star$ is primal optimal, and $(\mu^\star, \lambda^\star)$ is dual optimal, then

$$
\begin{aligned}
& g_i(x^\star) \le 0 \quad i = 1, \ldots, m \\
& h_j(x^\star) = 0 \quad j = 1, \ldots, p \\
& \mu_i^\star g_i(x^\star) = 0, \quad i = 1, \ldots, m \\
& x^\star \in \operatorname*{argmin}_x L(x, \mu^\star, \lambda^\star)
\end{aligned}
$$

conversely, these four conditions imply optimality of $x^\star$, $(\mu^\star, \lambda^\star)$ and strong duality

- functions are not necessarily differentiable
- recover KKT conditions for differentiable functions by replacing 4th condition with

$$
\nabla_x L(x^\star, \mu^\star, \lambda^\star) = \nabla f(x^\star) + \sum_{i=1}^{m} \mu_i^\star \nabla g_i(x^\star) + \sum_{j=1}^{p} \lambda_j^\star \nabla h_j(x^\star) = 0
$$

Optimality conditions for convex problems

Sufficient conditions

- for convex problems, the optimality conditions are sufficient
- *i.e.*, if $x^\star, (\mu^\star, \lambda^\star)$ satisfy opt. cond., then they're optimal with zero duality gap

Necessary and sufficient conditions

if problem is convex and Slater's constraint qualification holds:

- $x^\star$ is optimal iff there exist $\mu^\star, \lambda^\star$, a such that optimality conditions are satisfied
- Slater's condition implies optimal duality gap is zero and dual optimum is attained

Proof of sufficiency

- L is convex in x, so the 1st KKT condition means $x^\star$ minimizes L over x
- we conclude that

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

- so strong duality holds, and thus, $x^\star$ and $(\mu^\star, \lambda^\star)$ are primal and dual optimal

Recovering primal solution from dual

Unique minimizer: suppose $L(x, \mu^\star, \lambda^\star)$ has a unique minimizer $x^\star$:

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

- $x^\star$ 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: suppose $L(x, \mu^\star, \lambda^\star)$ has multiple minimizers

- it is not guaranteed that each of them is primal-optimal
- what is guaranteed is that the primal-optimal $x^\star$ is among minimizers of L

Example

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

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

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

  is convex over x for any $\mu \ge 0$
- a minimizer of L over x must satisfy:

$$\begin{aligned} \frac{\partial L}{\partial x_1} &= 2(x_1+3) + 2\mu x_1 = 0 \Longrightarrow x_1 = -3/(1+\mu) \\ \frac{\partial L}{\partial x_2} &= 2x_2 - \mu = 0 \Longrightarrow x_2 = \mu/2 \end{aligned}$$

- 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

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

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

Example

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

- $a_i, c_i, b \in \mathbb{R}$ are given
- the Lagrangian is

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

  which is also separable in x_i

- the dual function is

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

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

- the dual problem is thus

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

- dual is unconstrained and concave, so optimal solution must satisfy

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

- 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
- optimality conditions
- **example: total variation de-noising**

Signal de-noising

$$y = x + v$$

- $x \in \mathbb{R}^n$ is original signal
- y is measured signal
- $v \in \mathbb{R}^n$ is an unknown noise vector

Total variation de-noising: recover x by solving

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

- $\delta > 0$ is regularization parameter
- r_{tv} is the total variation function ($R \in \mathbb{R}^{(n-1)\times n}$):

$$r_{\text{tv}}(x) = \sum_{i=1}^{n-1} |x_i - x_{i+1}| = \|Rx\|_1, \; R = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & -1 & 1 \end{bmatrix}$$

Dual derivation

- we we have not yet explored how to manage general non-smooth terms
- 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:

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

  where we introduced the variable $z \in \mathbb{R}^{(n-1)}$
- the associated Lagrangian is:

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

- Lagrangian is separable in x and z, the minimization concerning x yields:

$$x^\star = \operatorname*{argmin}_x L(x,z,\lambda) = \operatorname*{argmin}_x \|x-y\|^2 + \lambda^T Rx = y - \tfrac{1}{2}R^T\lambda$$

- substituting this result, we get:

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

- to minimize with respect to z, we must address:

$$\min_z \quad \delta\|z\|_1 - \lambda^T z$$

- considering each component, we realize:

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

- consequently, the dual function becomes:

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

Dual problem

thus, our dual problem becomes:

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

- the constraints form a simple box constraint:

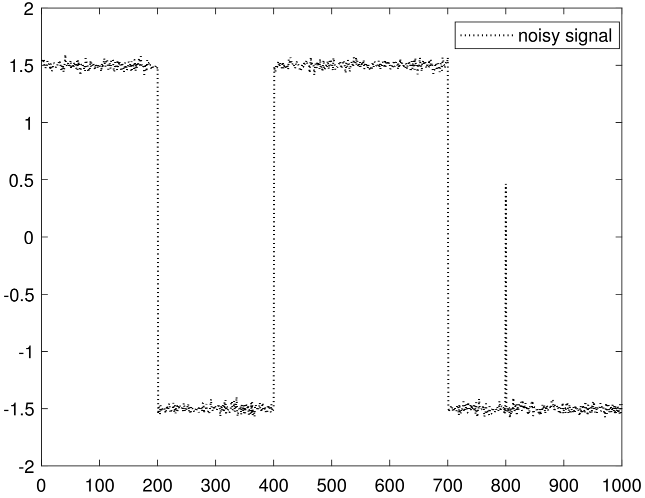
$$C = \{\lambda \in \mathbb{R}^{(n-1)} \mid -\delta \leq \lambda_i \leq \delta,\ i = 1, 2, \ldots, n-1\}$$

- we can solve the problem using the projected gradient descent
- the projection onto C, denoted by $\Pi(\lambda)$, has components:

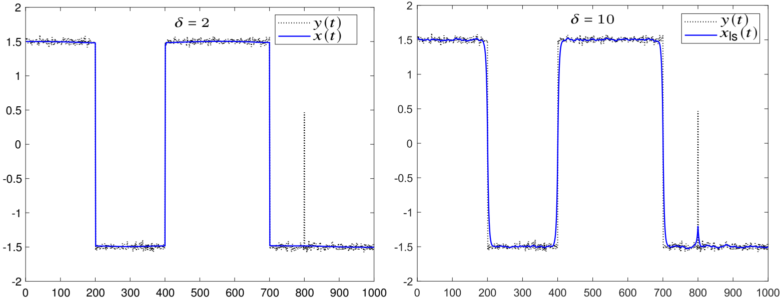
$$\Pi(\lambda)_i = \frac{\delta \lambda_i}{\max\{|\lambda_i|, \delta\}}$$

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

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

- S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004. (chapter 5.1, 5.2, 5.4, and 5.7)

- A. Beck. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with Python and MATLAB.* SIAM, 2023. (chapter 12)