Lagrange Duality

Yuanming Shi

ShanghaiTech University

Outline

Lagrangian

- Consider an optimization problem in standard form (not necessarily convex)

primal problem
$$\begin{array}{lll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) & \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0 & i = 1, \cdots, m \\ & h_i(\boldsymbol{x}) = 0 & i = 1, \cdots, p \end{array}$$

with variable $\boldsymbol{x} \in \mathbb{R}^n$, domain $\mathcal{D}$, and optimal value $p^\star$

- The *Lagrangian* is a function $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, with dom $L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$, defined as

dual variable

primal variable
$$L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{x})$$

where λ_i is the Lagrange multiplier associated with $f_i(\boldsymbol{x}) \leq 0$ and ν_i is the Lagrange multiplier associated with $h_i(\boldsymbol{x}) = 0$.

Outline

Lagrange Dual Function I

- The *Lagrange dual function* is defined as the infimum of the Lagrangian over $\boldsymbol{x}$: $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$,

$$
\begin{aligned}
g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\
&= \inf_{\boldsymbol{x} \in \mathcal{D}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i h_i(\boldsymbol{x}) \right)
\end{aligned}
$$

- Observe that:
  - the infimum is unconstrained (as opposed to the original constrained minimization problem)
  - g is concave regardless of original problem (infimum of affine functions)
  - g can be $-\infty$ for some $\boldsymbol{\lambda}, \boldsymbol{\nu}$

Lagrange Dual Function II

- **Lower bound property:** if $\boldsymbol{\lambda} \succeq \mathbf{0}$, then $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \le p^\star$.

$$L(\tilde{x}; \lambda, \nu) = f_0(\tilde{x}) + \sum_{i=1}^{m} \underbrace{\lambda_i f_i(\tilde{x})}_{\le 0} + \sum_{i=1}^{p} \nu_i \underbrace{h_i(\tilde{x})}_{= 0}$$

Proof.

Suppose $\tilde{\boldsymbol{x}}$ is feasible and $\boldsymbol{\lambda} \succeq \mathbf{0}$. Then,

$$f_0(\tilde{\boldsymbol{x}}) \ge L(\tilde{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \ge \inf_{\boldsymbol{x} \in \mathcal{D}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = g(\boldsymbol{\lambda}, \boldsymbol{\nu})$$

Now choose minimizer of $f_0(\tilde{\boldsymbol{x}})$ over all feasible $\tilde{\boldsymbol{x}}$ to get $p^\star \ge g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. $\square$

- We could try to find the best lower bound by maximizing $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. This is in fact the dual problem.

Outline

Dual Problem

- The *Lagrange dual problem* is defined as

$$\begin{array}{ll} \underset{\boldsymbol{\lambda}, \boldsymbol{\nu}}{\text{maximize}} & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{subject to} & \boldsymbol{\lambda} \succeq \mathbf{0} \end{array}$$

minimize $-g(\lambda, \nu)$

convex

- This problem finds the best lower bound on $p^\star$ obtained from the dual function
- It is a convex optimization (maximization of a concave function and linear constraints)
- The optimal value is denoted $d^\star$
- $\boldsymbol{\lambda}, \boldsymbol{\nu}$ are dual feasible if $\boldsymbol{\lambda} \succeq \mathbf{0}$ and $(\boldsymbol{\lambda}, \boldsymbol{\nu}) \in \operatorname{dom} g$ (the latter implicit constraints can be made explicit in problem formulation)

Example: Least-Norm Solution of Linear Equations I

- Consider the problem

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

- The Lagrangian is

convex in x

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

- To find the dual function, we need to solve an unconstrained minimization of the Lagrangian. We set the gradient equal to zero

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\nu}) = 2\boldsymbol{x} + \boldsymbol{A}^T\boldsymbol{\nu} = \boldsymbol{0} \Longrightarrow \boldsymbol{x} = -\frac{1}{2}\boldsymbol{A}^T\boldsymbol{\nu}$$

globally optimal solution

Example: Least-Norm Solution of Linear Equations II

and we plug the solution in L to obtain g:

concave

dual function.

$$\underline{g(\boldsymbol{\nu})} = L(-\frac{1}{2}\boldsymbol{A}^T\boldsymbol{\nu}, \boldsymbol{\nu}) = \underline{-\frac{1}{4}\boldsymbol{\nu}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{\nu} - \boldsymbol{b}^T\boldsymbol{\nu}}$$

- The function g is, as expected, a concave function of $\boldsymbol{\nu}$.
- From the lower bound property, we have

$$p^\star \geq -\frac{1}{4}\boldsymbol{\nu}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{\nu} - \boldsymbol{b}^T\boldsymbol{\nu} \text{ for all } \boldsymbol{\nu}$$

- The dual problem is the QP

$$\underset{\boldsymbol{\nu}}{\text{maximize}} \quad -\frac{1}{4}\boldsymbol{\nu}^T\boldsymbol{A}\boldsymbol{A}^T\boldsymbol{\nu} - \boldsymbol{b}^T\boldsymbol{\nu}$$

$$\Downarrow$$

cvx opt.

$$\underset{\nu}{\text{minimize}} \quad \frac{1}{4}\nu^T A A^T \nu + b^T \nu$$

Example: Standard Form LP I

- Consider the problem

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{c}^T\boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \quad \boldsymbol{x} \succeq \boldsymbol{0} \end{array} \quad -x \leq 0$$

- The Lagrangian is

$$g(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu)$$

$$\begin{aligned} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) &= \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{\nu}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) \underline{- \boldsymbol{\lambda}^T\boldsymbol{x}} \\ &= (\underbrace{\boldsymbol{c} + \boldsymbol{A}^T\boldsymbol{\nu} - \boldsymbol{\lambda}}_{a})^T\boldsymbol{x} - \boldsymbol{b}^T\boldsymbol{\nu} \end{aligned}$$

- L is a linear function of $\boldsymbol{x}$ and it is unbounded if the term multiplying $\boldsymbol{x}$ is nonzero.

$$a^T x = \sum_i a_i x_i = \begin{cases} 0, & a_i = 0,\ i = 1, \ldots, n \\ -\infty, & \exists a_i \neq 0 \end{cases}$$

Example: Standard Form LP II

- Hence, the dual function is

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\boldsymbol{b}^T \boldsymbol{\nu} & \boldsymbol{c} + \boldsymbol{A}^T \boldsymbol{\nu} - \boldsymbol{\lambda} = \boldsymbol{0} \\ -\infty & \text{otherwise} \end{cases}$$

- The function g is a concave function of $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ as it is linear on an affine domain.
- From the lower bound property, we have

$$p^\star \geq -\boldsymbol{b}^T \boldsymbol{\nu} \quad \text{if } \boldsymbol{c} + \boldsymbol{A}^T \boldsymbol{\nu} \succeq \boldsymbol{0}$$

- The dual problem is the LP

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

$\lambda \geq 0$

Outline

Weak and Strong Duality I

- From the lower bound property, we know that $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^\star$ for feasible $(\boldsymbol{\lambda}, \boldsymbol{\nu})$. In particular, for a $(\boldsymbol{\lambda}, \boldsymbol{\nu})$ that solves the dual problem.

- Hence, **weak duality** always holds (even for nonconvex problems):

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

- The difference $p^\star - d^\star$ is called **duality gap**.

- Solving the dual problem may be used to find nontrivial lower bounds for difficult problems.

- Even more interesting is when equality is achieved in weak duality. This is called **strong duality**:

$$d^\star = p^\star$$

Weak and Strong Duality II

- Strong duality means that the duality gap is zero.
- Strong duality:
  - is very desirable (we can solve a difficult problem by solving the dual)
  - does not hold in general
  - usually holds for convex problems
  - conditions that guarantee strong duality in convex problems are called **constraint qualifications**.

Slater's Constraint Qualification I

- Slater's constraint qualification is a very simple condition that is satisfied in most cases and ensures strong duality for convex problems.

- Strong duality hold for a convex problem

$$\begin{array}{lll} \underset{\boldsymbol{x}}{\text{minimize}} & f_0(\boldsymbol{x}) & \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0 & i = 1, \cdots, m \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} & \end{array}$$

  if it is strictly feasible, i.e.,

$$\exists\, \boldsymbol{x} \in \operatorname{int} \mathcal{D} : \quad f_i(\boldsymbol{x}) < 0 \quad i = 1, \cdots, m, \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

- There exist many other types of constraint qualifications.

Example: Inequality Form LP

- Consider the problem

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

- The dual problem is

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

- From Slater's condition: $p^\star = d^\star$ if $\boldsymbol{A}\tilde{\boldsymbol{x}} \prec \boldsymbol{b}$ for some $\tilde{\boldsymbol{x}}$.
- In this case, in fact, $p^\star = d^\star$ except when primal and dual are infeasible.

Example: Convex QP

- Consider the problem (assume $\boldsymbol{P} \succeq \mathbf{0}$)

$$\begin{array}{ll} \underset{\boldsymbol{x}}{\text{minimize}} & \boldsymbol{x}^T \boldsymbol{P} \boldsymbol{x} \\ \text{subject to} & \boldsymbol{A}\boldsymbol{x} \preceq \boldsymbol{b} \end{array}$$

- The dual problem is

$$\begin{array}{ll} \underset{\boldsymbol{\lambda}}{\text{maximize}} & -\frac{1}{4}\boldsymbol{\lambda}^T \boldsymbol{A}\boldsymbol{P}^{-1}\boldsymbol{A}^T\boldsymbol{\lambda} - \boldsymbol{b}^T\boldsymbol{\lambda} \\ \text{subject to} & \boldsymbol{\lambda} \succeq \mathbf{0} \end{array}$$

- From Slater's condition: $p^\star = d^\star$ if $\boldsymbol{A}\tilde{\boldsymbol{x}} \prec \boldsymbol{b}$ for some $\tilde{\boldsymbol{x}}$.
- In this case, in fact, $p^\star = d^\star$ always.

Complementary Slackness

- Assume strong duality holds, $\boldsymbol{x}^\star$ is primal optimal and $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ is dual optimal. Then

$\rightarrow$ def.

$$
\begin{aligned}
f_0(\boldsymbol{x}^\star) \underset{\text{strong duality}}{=} g(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) &= \inf_{\boldsymbol{x}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^\star f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^\star h_i(\boldsymbol{x}) \right) \\
\text{① "="} \quad &\le f_0(\boldsymbol{x}^\star) + \underbrace{\sum_{i=1}^{m} \lambda_i^\star f_i(\boldsymbol{x}^\star)}_{\le 0} + \underbrace{\sum_{i=1}^{p} \nu_i^\star h_i(\boldsymbol{x}^\star)}_{=0} \\
\text{② "="} \quad &\le f_0(\boldsymbol{x}^\star)
\end{aligned}
$$

$=0$

- Hence, the two inequalities must hold with equality. Implications:
  - ① $\Rightarrow$ $\boldsymbol{x}^\star$ minimizes $L(\boldsymbol{x}, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ Δ
  - ② $\Rightarrow$ $\lambda_i^\star f_i(\boldsymbol{x}^\star) = 0$ for $i = 1, \cdots, m$; this is called **complementary slackness**:

$$
\lambda_i^\star > 0 \Longrightarrow f_i(\boldsymbol{x}^\star) = 0, \quad f_i(\boldsymbol{x}^\star) < 0 \Longrightarrow \lambda_i^\star = 0
$$

Outline

Karush-Kuhn-Tucker (KKT) Conditions

KKT conditions (for differentiable f_i, h_i):

1. primal feasibility:

$$f_i(\boldsymbol{x}) \leq 0,\ i = 1, \cdots, m,\ h_i(\boldsymbol{x}) = 0,\ i = 1, \cdots, p$$

2. dual feasibility: $\boldsymbol{\lambda} \succeq \mathbf{0}$

3. complementary slackness: $\lambda_i^\star f_i(\boldsymbol{x}^\star) = 0$ for $i = 1, \cdots, m$

4. zero gradient of Lagrangian with respect to $\boldsymbol{x}$:

A

$$\nabla f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i \nabla f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i \nabla h_i(\boldsymbol{x}) = \mathbf{0}$$

KKT condition

primal dual

- We already known that if strong duality holds and $\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}$ are optimal, then they must satisfy the KKT conditions.
- What about the opposite statement?
- If $\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}$ satisfy the KKT conditions for a convex problem, then they are optimal.

$$L(x;\lambda,\nu) = f_0(x) + \sum_{i=1}^{m} \underbrace{\lambda_i f_i(x)}_{① =0} + \sum_{i=1}^{p} \underbrace{\nu_i h_i(x)}_{=0}$$

Proof.

①

From complementary slackness, $f_0(\boldsymbol{x}) = L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$ and, from 4th KKT condition and convexity, $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$. Hence, $f_0(\boldsymbol{x}) = g(\boldsymbol{\lambda}, \boldsymbol{\nu})$. □

$\nabla_x L(x;\lambda,\nu) = 0$ def. $g(\lambda,\nu) := \inf_x L(x;\lambda,\nu)$

Theorem

If a problem is convex and Slater's condition is satisfied, then x is optimal if and only if there exists $\boldsymbol{\lambda}, \boldsymbol{\nu}$ that satisfy the KKT conditions.

cvx_begin

convex opt.

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, n \\ & Ax = b \end{aligned}$$

cvx_end

$\Downarrow$ CVX step 1: transformation

conic optimization **primal - dual problem**

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax + s = b \\ & (x, s) \in \mathbb{R}^n \times \underline{K} \end{aligned}$$

convex cone

$$\begin{aligned} \text{maximize} \quad & -b^T y \\ \text{subject to} \quad & -A^T y + v = c \\ & (v, y) \in \{0\}^n \times \underline{K^*} \end{aligned}$$

dual cone of K

proof:

$$\begin{aligned} \underset{(x,s) \in D}{\text{minimize}} \quad & c^T x \\ \text{subject to} \quad & Ax + s = b \end{aligned}$$

$$g(y) = \inf_{(x,s)\in D} c^T x + \langle y, Ax+s-b\rangle$$

$$= -b^T y + \inf_{(x,s)\in D} [\langle c, x\rangle + \langle y, Ax+s\rangle]$$

$$= -b^T y + \underbrace{\inf_{x\in R^n} \langle c+A^T y, x\rangle}_{} + \underbrace{\inf_{s\in K} \langle y, s\rangle}_{}$$

① $$\inf_{x\in R^n} \langle c+A^T y, x\rangle = \begin{cases} 0, & c+A^T y = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

② $$\inf_{s\in K} \underbrace{\langle y, s\rangle}_{} = \begin{cases} 0, & \text{if } y \in K^* \\ -\infty, & \text{otherwise} \end{cases}$$

Convex cone K : for all $x \in K$,

$\lambda x \in K$, $\forall \lambda > 0$

Dual cone K^* : $K^* = \{ z \in \mathbb{R}^n ; \langle z, x \rangle \geq 0, \forall x \in K \}$.

---

1) $y \in K^* \Rightarrow \langle y, s \rangle \geq 0, \ \forall s \in K$

$$\Downarrow \ \lambda s \in K, \ \forall \lambda > 0$$

$$\langle y, \lambda s \rangle = \lambda \underbrace{\langle y, s \rangle}_{\geq 0}$$

2) $y \notin K^* \Rightarrow \underline{\langle y, s \rangle < 0, \ \exists s \in K}$

$$\Downarrow \ \lambda s \in K, \ \forall \lambda > 0$$

$$\langle y, \lambda s \rangle = \underbrace{\lambda}_{\lambda \to +\infty} \langle y, s \rangle \to -\infty$$

KKT conditions

primal feasible: $Ax + s = b$, $s \in \mathcal{K}$

dual feasible: $A^T y + c = r$, $r = 0$, $y \in \mathcal{K}^*$

complementary slackness: $c^T x + b^T y = 0$

$(y^T s = 0) \Leftrightarrow$ (strong duality)

$\Downarrow$

KKT system (two new variables)

$$\begin{aligned} A^T y + c\tau &= r \\ -Ax + b\tau &= s \\ c^T x + b^T y + \kappa &= 0 \end{aligned}$$

homogenous self-dual embedding system

$$(\underbrace{x, s, \tau}, \underbrace{r, y, \kappa}) \in \underbrace{\mathbb{R}^n \times \mathcal{K} \times \mathbb{R}_+}_{\mathcal{C}} \times \underbrace{\{0\}^n \times \mathcal{K}^* \times \mathbb{R}_+}_{\mathcal{C}^*}$$

$\Downarrow$ Solver: SDPT3, MOSEK, SeDuMi

Any solution of the self-dual embedding (x, s, z, r, y, κ) falls into one of three cases:

1. $z > 0$, $\kappa = 0$. The point

$$(\hat{x}, \hat{y}, \hat{s}) = \left(\frac{x}{z}, \frac{y}{z}, \frac{s}{z}\right) \quad \to \in K$$

satisfies the KKT conditions $\Rightarrow$ a primal-dual solution

2. $z = 0$, $\kappa > 0 \Rightarrow \underline{c^T x + b^T y < 0} \Rightarrow$ either primal or dual infeasible

Theorem: certificates of infeasibility (sect. 5.8)

If strong duality holds, then exactly one of the sets:

① $P = \{(x, s) : Ax + s = b,\ s \in K\}$: encodes primal feasibility

② $D = \{y : A^T y = 0,\ y \in K^*,\ b^T y < 0\}$

is non empty.

$$\boxed{\text{Theorem of strong alternatives: Any dual variable } y \in D \text{ serves as a proof or certificate that the set } P \text{ is empty, i.e., the problem is primal infeasible}}$$

Similarly, exactly one of the following two sets is non-empty:

① $\tilde{P} = \{x : -Ax \in \mathcal{K}, \ \underbrace{c^T x < 0}\}$

② $\tilde{D} = \{y : A^T y = -c, \ y \in \mathcal{K}^*\}$:

dual feasible

claim: Any primal varible $x \in \tilde{P}$ is a certificate of dual infeasibility.

2. $z=0$, $x > 0 \Rightarrow \boxed{c^T x + b^T y < 0} \Rightarrow$ either primal or dual infeasible

1) if $b^T y < 0$, then $\hat{y} = \frac{y}{-b^T y}$ is a certificate of primal infeasibility (i.e., D is nonempty), since

$$A^T \hat{y} = \frac{r}{-b^T y} = 0, \quad \hat{y} \in \mathcal{K}^*, \quad b^T \hat{y} = -1$$

2) if $c^T x < 0$, then $\hat{x} = \frac{x}{-c^T x}$ is a certificate of dual infeasibility (i.e., $\tilde{P}$ is nonempty) since

$$-A\hat{x} = \frac{s}{\underbrace{-c^T x}_{>0}} \in \mathcal{K}, \quad c^T \hat{x} = -1$$

3) $c^T x < 0$, $b^T y < 0 \Rightarrow$ both primal and dual infeasible

strong duality assumption is violated!

3. $\tau = \kappa = 0$, nothing can be concluded. can be avoided.

Reference

Chapter 5 of:

- Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.