Introduction to duality theory

Kostas Margellos

University of Oxford

February 12, 2020

References

Convex Optimization & Duality Theory:

Boyd & Vandenberghe (2004)
Convex Optimization, *Cambridge University Press*.

Bertsekas (2009)
Convex Optimization Theory, *Athena Scientific*.

Rockafellar (1970)
Convex Analysis, *Princeton, NJ: Princeton University Press*.

Notation: Optimization program

A more common problem format:

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & f_0(x) \\ \text{subject to:} \quad & f_i(x) \leq 0 \quad i = 1, \dots, m \\ & h_i(x) = 0 \quad i = 1, \dots, p \end{aligned}$$

- **Objective function** $f_0 : \mathcal{X} \to \mathbb{R}$
- **Domain** $\mathcal{X} \subseteq \mathbb{R}^n$ of the objective function, from which the decision variable $x := (x_1; x_2; \dots; x_n)$ must be chosen.
- **Inequality constraint functions** $f_i : \mathbb{R}^n \to \mathbb{R}$, for $i = 1, \dots, m$
- **Equality constraint functions** $h_i : \mathbb{R}^n \to \mathbb{R}$, for $i = 1, \dots, p$

$\Rightarrow$ *Maximization* fit the framework with a change of sign.

Notation: Convex optimization program

- Standard form optimization problem:

$$
\begin{aligned}
\min_{x \in \mathcal{X}} \quad & f_0(x) \\
\text{subject to:} \quad & f_i(x) \leq 0 \quad i = 1, \ldots, m \\
& h_i(x) = 0 \quad i = 1, \ldots, p
\end{aligned}
$$

  - *Primal* decision variables x
  - For convex programs:
    - $\mathcal{X}$, f_0, f_i's convex, and $h_i(x) = a_i^\top x$ all affine
    - local = global optimum

Optimization programs and duality

Consider the following optimization program (no convexity assumption yet)

$$
\begin{aligned}
\min_{x \in \mathcal{X}} \quad & f_0(x) \\
\text{subject to:} \quad & f_i(x) \leq 0 \quad i = 1, \ldots, m \\
& h_i(x) = 0 \quad i = 1, \ldots, p
\end{aligned}
$$

- Assume we are interested in the optimal value p^* of (SDP)
- Can we construct a lower bound for p^*, i.e. $d^* \leq p^*$, by solving another problem?
- This problem, called *dual*, might sometimes be easier to solve

To do this we first need some machinery – Duality Theory

The Lagrangian function

Recall our standard form (primal) optimization problem:

$$(\mathcal{P}): \quad \begin{array}{rll} \min\limits_{x \in \mathcal{X}} & f_0(x) & \\ \text{subject to:} & f_i(x) \leq 0 & i = 1 \dots m \\ & h_i(x) = 0 & i = 1 \dots p \end{array}$$

with (primal) decision variable x, domain $\mathcal{X}$ and optimal value p^*.

Lagrangian Function: $L : \mathcal{X} \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

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

- λ_i : inequality Lagrange multiplier for $f_i(x) \leq 0$.
- ν_i : equality Lagrange multiplier for $h_i(x) = 0$.
- Lagrangian: weighted sum of the objective and constraint functions.

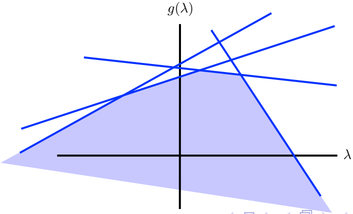
Lagrange dual function

The **dual function** $g : \mathbb{R}^m \times \mathbb{R}^p$ is

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

The dual function $g(\lambda, \nu)$ is always a **concave** function.

- $g(\lambda, \nu)$ is the pointwise infimum of affine functions
  Do you recall pointwise maximum?



Lagrange dual function

The **dual function** $g : \mathbb{R}^m \times \mathbb{R}^p$ is

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

The dual function generates lower bounds for the primal optimal value, i.e. $g(\lambda, \nu) \leq p^*$ for $\lambda \geq 0$:
Proof:
For any primal feasible solution $\bar{x}$: $\sum_{i=1}^{m} \lambda_i f_i(\bar{x}) + \sum_{i=1}^{p} \nu_i h_i(\bar{x}) \leq 0$

$$
\begin{aligned}
&g(\lambda, \nu) = \inf_{x \in \mathcal{X}} L(x, \lambda, \nu) \leq L(\bar{x}, \lambda, \nu) \leq f_0(\bar{x}) \text{ for all } \bar{x} \\
&g(\lambda, \nu) \leq \inf_{x \in \mathcal{X}} f_0(x) \leq p^*
\end{aligned}
$$

- $g(\lambda, \nu)$ might be $-\infty$; Non-trivial if $\operatorname{dom} g := \{\lambda, \nu \mid g(\lambda, \nu) > -\infty\}$

The dual problem

Every $\nu \in \mathbb{R}^p$, $\lambda \geq 0$ produces a lower bound for p^* using the dual function.
Which is the best?

$$(\mathcal{D}): \quad \begin{array}{cc} \max\limits_{\lambda,\nu} & g(\lambda,\nu) \\ \text{subject to:} & \lambda \geq 0 \end{array}$$

- Problem $(\mathcal{D})$ is **convex**, even if $(\mathcal{P})$ is not.
- Problem $(\mathcal{D})$ has optimal value $d^* \leq p^*$.
- The point (λ,ν) is **dual feasible** if $\lambda \geq 0$ and $(\lambda,\nu) \in \operatorname{dom} g$.
- Often impose the constraint $(\lambda,\nu) \in \operatorname{dom} g$ explicitly in $(\mathcal{D})$.

Example : Dual of LPs

$$(\mathcal{P}) : \quad \begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x \\ \text{subject to:} \quad & Ax = b \\ & Cx \leq d \end{aligned}$$

The **dual function** is

$$g(\lambda, \nu) = \min_{x \in \mathbb{R}^n} \left[c^\top x + \nu^\top (Ax - b) + \lambda^\top (Cx - d) \right]$$

$$= \min_{x \in \mathbb{R}^n} \left[(A^\top \nu + C^\top \lambda + c)^\top x - b^\top \nu - d^\top \lambda \right]$$

$$= \begin{cases} -b^\top \nu - d^\top \lambda & \text{if } A^\top \nu + C^\top \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Example : Dual of LPs – (cont'd)

$$(\mathcal{P}) : \quad \begin{aligned} \min_{x \in \mathbb{R}^n} \quad & c^\top x \\ \text{subject to:} \quad & Ax = b \\ & Cx \le d \end{aligned}$$

The **dual problem** is

$$(\mathcal{D}) : \quad \begin{aligned} \max_{\lambda, \nu} \quad & -b^\top \nu - d^\top \lambda \\ \text{subject to:} \quad & A^\top \nu + C^\top \lambda + c = 0 \\ & \lambda \ge 0 \end{aligned}$$

- Lower bound property:
  $-b^\top \nu - d^\top \lambda \le p^*$ whenever $\lambda \ge 0$.
- The dual of a linear program is also a linear program.

Example : Dual of a mixed-integer linear program (MILP)

$$(\mathcal{P}): \quad \begin{aligned} \min_{x \in \mathcal{X}} \quad & c^\top x \\ \text{subject to:} \quad & Ax \leq b \\ & \mathcal{X} = \{-1, 1\}^n \end{aligned}$$

The **dual function** is

$$g(\lambda) = \min_{x_i \in \{-1,1\}} \left[c^\top x + \lambda^\top (Ax - b) \right]$$

$$= -\|A^\top \lambda + c\|_1 - b^\top \lambda$$

The **dual problem** is

$$(\mathcal{D}): \quad \begin{aligned} \max_{\lambda} \quad & -\|A^\top \lambda + c\|_1 - b^\top \lambda \\ \text{subject to:} \quad & \lambda \geq 0 \end{aligned}$$

The dual of a mixed-integer linear program is a linear program!

Weak and strong duality

Weak Duality

- It is **always** true that $d^* \leq p^*$.
- Sometimes the dual is much easier to solve than the primal (or vice-versa).
- Example: The dual of an MILP (difficult to solve) is a standard LP (easy to solve).

Strong Duality

- It is **sometimes** true that $d^* = p^*$.
- Strong duality usually holds for convex problems.
- Strong duality usually does not hold for non-convex problems.
- Can impose conditions on convex problems to guarantee that $d^* = p^*$.

Strong duality for convex problems

An optimization problem with f_0 and all f_i convex:

$$(\mathcal{P}): \quad \begin{array}{rll} \min & f_0(x) & \\ \text{subject to:} & f_i(x) \leq 0 & i = 1 \dots m \\ & Ax = b & A \in \mathbb{R}^{p \times n} \end{array}$$

Slater Condition

If there is at least one **strictly feasible point**, i.e.

$$\left\{ x \;\middle|\; Ax = b,\; f_i(x) < 0,\; \forall i \in \{1, \dots, m\} \right\} \neq \emptyset$$

Then $p^* = d^*$.

- Stronger version: Only the nonlinear functions $f_i(x)$ must be strictly satisfiable (non-empty interior).
- Other **constraint qualification** conditions exist.

Duality – A geometric view

Assume one inequality constraint only:

$$\mathcal{G} := \{(u,t) \mid t = f_0(x),\ u = f_1(x),\ x \in \mathcal{X}\}$$

Primal problem:

$$p^* = \min\{t \mid (u,t) \in \mathcal{G},\ \ u \le 0\}$$

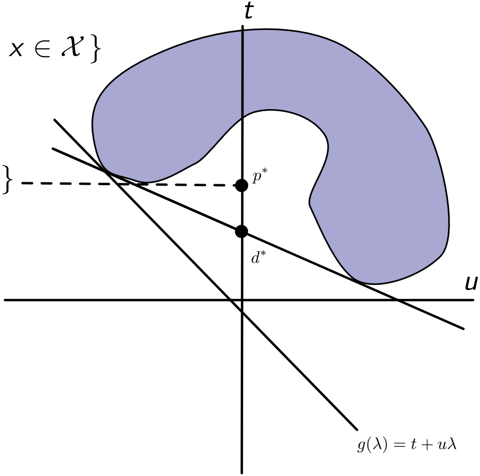
Dual function:

$$g(\lambda) = \min_{(u,t)\in\mathcal{G}} (t + \lambda u)$$

Dual problem:

$$d^* = \max_{\lambda \ge 0} g(\lambda)$$

The quantity $p^* - d^*$ is the **duality gap**.



Primal and dual solution properties

Assume that strong duality holds, with optimal solution x^* and (λ^*, ν^*).

- From strong duality, $d^* = p^* \;\; \Rightarrow \;\; g(\lambda^*, \nu^*) = f_0(x^*)$.
- From the definition of the dual function:

$$f_0(x^*) = g(\lambda^*, \nu^*) = \min_x \left\{ f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x) \right\}$$

$$\leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*) \leq f_0(x^*)$$

[weak duality]

$$\Longrightarrow f_0(x^*) = g(\lambda^*, \nu^*) = f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*)$$

$$\Longrightarrow \left. \begin{array}{l} \lambda_i^* = 0 \text{ for every } f_i(x^*) < 0. \\ f_i(x^*) = 0 \text{ for every } \lambda_i^* > 0. \end{array} \right\} \textbf{Complementary slackness}$$

Karush-Kuhn-Tucker (KKT) optimality conditions

Assume that all f_i and h_i are differentiable (no convexity assumption yet).
Necessary conditions for optimality:

1. Primal Feasibility:

$$f_i(x^*) \leq 0 \quad i = 1, \ldots, m$$
$$h_i(x^*) = 0 \quad i = 1 \ldots, p$$

2. Dual Feasibility:

$$\lambda^* \geq 0$$

3. Complementary Slackness:

$$\lambda_i^* f_i(x^*) = 0 \quad i = 1, \ldots, m$$

4. Stationarity:

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

KKT optimality conditions – Convex Programs

Assume that all f_i and h_i are differentiable and problem is convex:

1) If (x^*, λ^*, ν^*) satisfy the KKT conditions, then
   - they are primal and dual optimal
   - they result in zero duality gap, i.e. $p^* = d^*$

2) If in addition Slater's condition holds, then
   - duality gap is zero and the dual optimum is attained (existence of (λ^*, ν^*) is guaranteed)
   - x^* is optimal **if and only if** there exist (λ^*, ν^*) that, together with x^*, satisfy the KKT conditions

Example : KKT optimality conditions for QPs

Consider a (convex) quadratic program with $Q \succeq 0$:

$$(\mathcal{P}) : \quad \begin{array}{rl} \min\limits_{x \in \mathbb{R}^n} & \frac{1}{2} x^\top Q x + c^\top x \\ \text{subject to:} & Ax = b \\ & x \geq 0 \end{array}$$

The **Lagrangian** is $L(x, \lambda, \nu) = \frac{1}{2} x^\top Q x + c^\top x + \nu^\top (Ax - b) - \lambda^\top x$.

The KKT conditions are:

$$\begin{array}{rll} \nabla_x L(x, \lambda, \nu) = Qx + A^\top \nu - \lambda + c = 0 & & \text{[stationarity]} \\ Ax = b & & \text{[primal feasibility]} \\ x \geq 0 & & \text{[primal feasibility]} \\ \lambda \geq 0 & & \text{[dual feasibility]} \\ x_i \lambda_i = 0, & i = 1 \ldots n & \text{[complementarity]} \end{array}$$