Duality

V. Leclère (ENPC)

April 12th, 2024

Why should I bother to learn this stuff?

- Duality allows a second representation of the same convex problem, giving sometimes some interesting insights (e.g. principle of virtual forces in mechanics)
- Duality is a good way of obtaining lower bounds
- Duality is a powerful tool for decomposition methods
- $\Longrightarrow$ fundamental both for studying optimization (continuous and operations research)
- $\Longrightarrow$ useful in other fields like mechanics and machine learning

Contents

Min-Max duality

♡

Consider the following problem

$$\underset{x\in\mathcal{X}}{\text{Min}} \quad \sup_{y\in\mathcal{Y}} \qquad \Phi(x,y)$$

where, for the moment, $\mathcal{X}$ and $\mathcal{Y}$ are arbitrary sets, and Φ an arbitrary function.
By definition the dual of this problem is

$$\underset{y\in\mathcal{Y}}{\text{Max}} \quad \inf_{x\in\mathcal{X}} \qquad \Phi(x,y)$$

and we have weak duality, that is

$$\sup_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} \Phi(x,y) \leq \inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \Phi(x,y)$$

♣ Exercise: Prove this result. Getting equality is more tricky...

Dual representation of some characteristic functions

Recall that, if $X \subset \mathbb{R}^n$

$$\mathbb{I}_X(x) = \begin{cases} 0 & \text{if } x \in X \\ +\infty & \text{otherwise} \end{cases}$$

and if X is an assertion,

$$\mathbb{I}_X = \begin{cases} 0 & \text{if } X \\ +\infty & \text{otherwise} \end{cases}$$

Note that

$$\mathbb{I}_{g(x)=0} = \sup_{\lambda \in \mathbb{R}^{n_E}} \lambda^\top g(x)$$

and

$$\mathbb{I}_{h(x)\leq 0} = \sup_{\mu \in \mathbb{R}^{n_I}_+} \mu^\top h(x)$$

Dual representation of some characteristic functions

Recall that, if $X \subset \mathbb{R}^n$

$$\mathbb{I}_X(x) = \begin{cases} 0 & \text{if } x \in X \\ +\infty & \text{otherwise} \end{cases}$$

and if X is an assertion,

$$\mathbb{I}_X = \begin{cases} 0 & \text{if } X \\ +\infty & \text{otherwise} \end{cases}$$

Note that

$$\mathbb{I}_{g(x)=0} = \sup_{\lambda \in \mathbb{R}^{n_E}} \lambda^\top g(x)$$

and

$$\mathbb{I}_{h(x)\leq 0} = \sup_{\mu \in \mathbb{R}_+^{n_I}} \mu^\top h(x)$$

From constrained to min-sup formulation

$$
\begin{aligned}
\underset{x \in \mathbb{R}^n}{\text{Min}} \quad & f(x) & (P)\\
\text{s.t.} \quad & g_i(x) = 0 & \forall i \in [n_E]\\
& h_j(x) \leq 0 & \forall j \in [n_I]
\end{aligned}
$$

Is equivalent to

$$\underset{x \in \mathbb{R}^n}{\text{Min}} \quad f(x) + \mathbb{I}_{g(x)=0} + \mathbb{I}_{h(x)\leq 0}$$

or

$$\underset{x \in \mathbb{R}^n}{\text{Min}} \quad f(x) + \sup_{\lambda \in \mathbb{R}^{n_E}} \lambda^\top g(x) + \sup_{\mu \in \mathbb{R}_+^{n_I}} \mu^\top h(x)$$

which is usually written

$$\underset{x \in \mathbb{R}^n}{\text{Min}} \quad \sup_{\lambda, \mu \geq 0} \quad \underbrace{f(x) + \lambda^\top g(x) + \mu^\top h(x)}_{:=\mathcal{L}(x;\lambda,\mu)}$$

From constrained to min-sup formulation ♡

$$
\begin{aligned}
\underset{x\in\mathbb{R}^n}{\text{Min}} \quad & f(x) && (P)\\
\text{s.t.} \quad & g_i(x) = 0 && \forall i \in [n_E]\\
& h_j(x) \leq 0 && \forall j \in [n_I]
\end{aligned}
$$

Is equivalent to

$$\underset{x\in\mathbb{R}^n}{\text{Min}} \quad f(x) + \mathbb{I}_{g(x)=0} + \mathbb{I}_{h(x)\leq 0}$$

or

$$\underset{x\in\mathbb{R}^n}{\text{Min}} \quad f(x) + \sup_{\lambda\in\mathbb{R}^{n_E}} \lambda^\top g(x) + \sup_{\mu\in\mathbb{R}_+^{n_I}} \mu^\top h(x)$$

which is usually written

$$\underset{x\in\mathbb{R}^n}{\text{Min}} \quad \sup_{\lambda,\mu\geq 0} \quad \underbrace{f(x) + \lambda^\top g(x) + \mu^\top h(x)}_{:=\mathcal{L}(x;\lambda,\mu)}$$

Lagrangian duality

♡

To a (primal) problem (no convexity or regularity assumptions here)

$$
\begin{aligned}
(P) \quad \underset{x\in\mathbb{R}^n}{\text{Min}} \quad & f(x) \\
\text{s.t.} \quad & g_i(x) = 0 && \forall i \in [n_E] \\
& h_j(x) \leq 0 && \forall j \in [n_I]
\end{aligned}
$$

we associate the Lagrangian

$$
\mathcal{L}(x;\lambda,\mu) := f(x) + \lambda^\top g(x) + \mu^\top h(x)
$$

such that

$$
(P) \qquad \underset{x\in\mathbb{R}^n}{\text{Min}} \quad \sup_{\lambda,\mu\geq 0} \quad \mathcal{L}(x;\lambda,\mu)
$$

The dual problem is defined as

$$
(D) \qquad \underset{\lambda,\mu\geq 0}{\text{Max}} \quad \inf_{x\in\mathbb{R}^n} \quad \mathcal{L}(x;\lambda,\mu)
$$

Lagrangian duality

To a (primal) problem (no convexity or regularity assumptions here)

$$
\begin{aligned}
(P) \quad & \underset{x\in\mathbb{R}^n}{\text{Min}} \quad f(x) & \\
& \text{s.t.} \quad g_i(x) = 0 & \forall i \in [n_E] \\
& \qquad\;\; h_j(x) \leq 0 & \forall j \in [n_I]
\end{aligned}
$$

we associate the Lagrangian

$$\mathcal{L}(x;\lambda,\mu) := f(x) + \lambda^\top g(x) + \mu^\top h(x)$$

such that

$$(P) \qquad \underset{x\in\mathbb{R}^n}{\text{Min}} \quad \sup_{\lambda,\mu\geq 0} \quad \mathcal{L}(x;\lambda,\mu)$$

The dual problem is defined as

$$(D) \qquad \underset{\lambda,\mu\geq 0}{\text{Max}} \quad \inf_{x\in\mathbb{R}^n} \quad \mathcal{L}(x;\lambda,\mu)$$

Weak duality

By the min-max duality, we easily see that

$$\mathrm{val}(D) \leq \mathrm{val}(P).$$

Further any admissible dual multipliers $\lambda \in \mathbb{R}^{n_E}$ $\mu \in \mathbb{R}^{n_I}_+$ yields a lower bound:

$$g(\lambda, \mu) := \inf_{x \in \mathbb{R}^n} \mathcal{L}(x; \lambda, \mu) \leq \mathrm{val}(D) \leq \mathrm{val}(P)$$

Obviously, any admissible solution $x \in \mathbb{R}^n$ (i.e. such that $g(x) = 0$ and $h(x) \leq 0$), yields an upper bound

$$\mathrm{val}(P) \leq f(x) = \sup_{\lambda, \mu \geq 0} \mathcal{L}(x; \lambda, \mu)$$

Weak duality

By the min-max duality, we easily see that

$$\mathrm{val}(D) \leq \mathrm{val}(P).$$

Further any admissible dual multipliers $\lambda \in \mathbb{R}^{n_E}$ $\mu \in \mathbb{R}_+^{n_I}$ yields a lower bound:

$$g(\lambda, \mu) := \inf_{x \in \mathbb{R}^n} \mathcal{L}(x; \lambda, \mu) \leq \mathrm{val}(D) \leq \mathrm{val}(P)$$

Obviously, any admissible solution $x \in \mathbb{R}^n$ (i.e. such that $g(x) = 0$ and $h(x) \leq 0$), yields an upper bound

$$\mathrm{val}(P) \leq f(x) = \sup_{\lambda, \mu \geq 0} \mathcal{L}(x; \lambda, \mu)$$

Weak duality

By the min-max duality, we easily see that

$$\mathrm{val}(D) \leq \mathrm{val}(P).$$

Further any admissible dual multipliers $\lambda \in \mathbb{R}^{n_E}$ $\mu \in \mathbb{R}_+^{n_I}$ yields a lower bound:

$$g(\lambda, \mu) := \inf_{x \in \mathbb{R}^n} \mathcal{L}(x; \lambda, \mu) \leq \mathrm{val}(D) \leq \mathrm{val}(P)$$

Obviously, any admissible solution $x \in \mathbb{R}^n$ (i.e. such that $g(x) = 0$ and $h(x) \leq 0$), yields an upper bound

$$\mathrm{val}(P) \leq f(x) = \sup_{\lambda, \mu \geq 0} \mathcal{L}(x; \lambda, \mu)$$

Contents

Min-Max duality

Recall the generic primal problem of the form

$$p^\star := \underset{x \in \mathcal{X}}{\mathrm{Min}} \quad \sup_{y \in \mathcal{Y}} \qquad \Phi(x, y)$$

with associated dual

$$d^\star := \underset{y \in \mathcal{Y}}{\mathrm{Max}} \quad \inf_{x \in \mathcal{X}} \qquad \Phi(x, y).$$

Recall that the duality gap $p^\star - d^\star \geq 0$.

We say that we have strong duality if $d^\star = p^\star$.

Min-Max duality

Recall the generic primal problem of the form

$$p^\star := \underset{x \in \mathcal{X}}{\mathrm{Min}} \quad \sup_{y \in \mathcal{Y}} \qquad \Phi(x, y)$$

with associated dual

$$d^\star := \underset{y \in \mathcal{Y}}{\mathrm{Max}} \quad \inf_{x \in \mathcal{X}} \qquad \Phi(x, y).$$

Recall that the duality gap $p^\star - d^\star \geq 0$.
We say that we have strong duality if $d^\star = p^\star$.

Saddle point

Definition

Let $\Phi : \mathcal{X} \times \mathcal{Y} \to \overline{\mathbb{R}}$ be any function. $(x^\sharp, y^\sharp)$ is a (local) saddle point of Φ on $\mathcal{X} \times \mathcal{Y}$ if

- $x^\sharp$ is a (local) minimum of $x \mapsto \Phi(x, y^\sharp)$.
- $y^\sharp$ is a (local) maximum of $y \mapsto \Phi(x^\sharp, y)$.

If there exists a global Saddle Point $(x^\sharp, y^\sharp)$ of Φ, then there is strong duality, $x^\sharp$ is an optimal primal solution and $y^\sharp$ an optimal dual solution, i.e.

$$p^\star = d^\star = \Phi(x^\sharp, y^\sharp).$$

Saddle point

Definition

Let $\Phi : \mathcal{X} \times \mathcal{Y} \to \overline{\mathbb{R}}$ be any function. $(x^\sharp, y^\sharp)$ is a (local) saddle point of Φ on $\mathcal{X} \times \mathcal{Y}$ if

- $x^\sharp$ is a (local) minimum of $x \mapsto \Phi(x, y^\sharp)$.
- $y^\sharp$ is a (local) maximum of $y \mapsto \Phi(x^\sharp, y)$.

If there exists a global Saddle Point $(x^\sharp, y^\sharp)$ of Φ, then there is strong duality, $x^\sharp$ is an optimal primal solution and $y^\sharp$ an optimal dual solution, i.e.

$$p^\star = d^\star = \Phi(x^\sharp, y^\sharp).$$

Sufficient conditions for saddle point

Theorem (Von Neumann Minimax Theorem)

If

- $\mathcal{X}$ *and* $\mathcal{Y}$ *are convex, one of them is compact*
- Φ *is continuous on* $\mathcal{X} \times \mathcal{Y}$
- $\Phi(\cdot, y)$ *is convex for all* $y \in \mathcal{Y}$
- $\Phi(x, \cdot)$ *is concave for all* $x \in \mathcal{X}$

then there exists a saddle point of Φ

$\rightsquigarrow$ *we can exchange "Min" and "Max".*

Slater's conditions for convex optimization ♡

Consider the following convex optimization problem

$$
\begin{aligned}
(P) \qquad \underset{x \in \mathbb{R}^n}{\text{Min}} \quad & f(x) \\
\text{s.t.} \quad & Ax = b \\
& h_j(x) \leq 0 \qquad\qquad \forall j \in [n_I]
\end{aligned}
$$

We say that a point x^s such that $Ax^s = b$, $x^s \in \mathrm{ri}(\mathrm{dom}(f))$[1], and $h_j(x^s) < 0$ for all $j \in [n_I]$, is a Slater's point.

[1] for extended function f

Slater's conditions for convex optimization

Consider the following convex optimization problem

$$
\begin{aligned}
(P) \qquad \underset{x \in \mathbb{R}^n}{\text{Min}} \quad & f(x) \\
\text{s.t.} \quad & Ax = b \\
& h_j(x) \leq 0 \qquad\qquad \forall j \in [n_I]
\end{aligned}
$$

We say that a point x^s such that $Ax^s = b$, $x^s \in \mathrm{ri}(\mathrm{dom}(f))$[1], and $h_j(x^s) < 0$ for all $j \in [n_I]$, is a Slater's point.

Theorem

If (P) is convex (i.e. f and h_j are convex), and there exists a Slater's point then there is strong (Lagrangian) duality.

Further if (P) admits an optimal solution $x^\sharp$ then $\mathcal{L}$ admits a saddle point $(x^\sharp, \lambda^\sharp)$, and $\lambda^\sharp$ is an optimal solution to (D).

[1] for extended function f

Contents

Perturbed problem

We consider the following perturbed problem

$$
\begin{aligned}
v(p,q) = \quad & \underset{x\in\mathbb{R}^n}{\mathrm{Min}} & & f(x) \\
& \text{s.t.} & & g(x) = p \\
& & & h(x) \le q
\end{aligned}
$$

In particular we have $v(0,0) = \mathrm{val}(P)$.

By duality,

$$
v(p,q) \ge d(p,q) = \sup_{\lambda,\mu\ge 0} \inf_x f(x) + \lambda^\top (g(x) - p) + \mu^\top (h(x) - q).
$$

In particular, d is convex as a supremum of convex functions.

Perturbed problem

We consider the following perturbed problem

$$v(p,q) = \quad \begin{aligned} \underset{x\in\mathbb{R}^n}{\text{Min}} \quad & f(x) \\ \text{s.t.} \quad & g(x) = p \\ & h(x) \leq q \end{aligned}$$

In particular we have $v(0,0) = \mathrm{val}(P)$.
By duality,

$$v(p,q) \geq d(p,q) = \sup_{\lambda,\mu\geq 0} \inf_{x} f(x) + \lambda^\top (g(x) - p) + \mu^\top (h(x) - q).$$

In particular, d is convex as a supremum of convex functions.

Perturbed problem

We consider the following perturbed problem

$$v(p,q) = \quad \underset{x\in\mathbb{R}^n}{\text{Min}} \quad f(x)$$
$$\text{s.t.} \quad g(x) = p$$
$$h(x) \leq q$$

In particular we have $v(0,0) = \mathrm{val}(P)$.
By duality,

$$v(p,q) \geq d(p,q) = \sup_{\lambda,\mu\geq 0} \inf_{x} f(x) + \lambda^\top(g(x)-p) + \mu^\top(h(x)-q).$$

In particular, d is convex as a supremum of convex functions.

Marginal interpretation of the dual multiplier

Assume that (P) is convex, and satisfies the Slater's qualification condition. In particular $v(0,0) = d(0,0)$.

Let (λ, μ) be optimal multipliers of (P).

We have, for any $x_{p,q}$ admissible for the perturbed problem, that is such that $g(x_{p,q}) = p$, $h(x_{p,q}) \leq q$,

$$\begin{aligned} \mathrm{val}(P) = v(0,0) &= \inf_x f(x) + \lambda^\top g(x) + \mu^\top h(x) \\ &\leq f(x_{p,q}) + \lambda^\top g(x_{p,q}) + \mu^\top h(x_{p,q}) \\ &\leq f(x_{p,q}) + \lambda^\top p + \mu^\top q \end{aligned}$$

In particular we have,

$$v(p,q) = \inf_{x_{p,q}} f(x_{p,q}) \geq v(0,0) - \lambda^\top p - \mu^\top q$$

which reads

$$-(\lambda, \mu) \in \partial v(0,0)$$

Marginal interpretation of the dual multiplier

Assume that (P) is convex, and satisfies the Slater's qualification condition. In particular $v(0,0) = d(0,0)$.
Let (λ, μ) be optimal multipliers of (P).
We have, for any $x_{p,q}$ admissible for the perturbed problem, that is such that $g(x_{p,q}) = p$, $h(x_{p,q}) \leq q$,

$$
\begin{aligned}
\mathrm{val}(P) = v(0,0) &= \inf_{x} f(x) + \lambda^\top g(x) + \mu^\top h(x) \\
&\leq f(x_{p,q}) + \lambda^\top g(x_{p,q}) + \mu^\top h(x_{p,q}) \\
&\leq f(x_{p,q}) + \lambda^\top p + \mu^\top q
\end{aligned}
$$

In particular we have,

$$
v(p,q) = \inf_{x_{p,q}} f(x_{p,q}) \geq v(0,0) - \lambda^\top p - \mu^\top q
$$

which reads

$$
-(\lambda, \mu) \in \partial v(0,0)
$$

Marginal interpretation of the dual multiplier ♡

Assume that (P) is convex, and satisfies the Slater's qualification condition. In particular $v(0,0) = d(0,0)$.
Let (λ, μ) be optimal multipliers of (P).
We have, for any $x_{p,q}$ admissible for the perturbed problem, that is such that $g(x_{p,q}) = p$, $h(x_{p,q}) \leq q$,

$$\begin{aligned}
\mathrm{val}(P) = v(0,0) &= \inf_x f(x) + \lambda^\top g(x) + \mu^\top h(x) \\
&\leq f(x_{p,q}) + \lambda^\top g(x_{p,q}) + \mu^\top h(x_{p,q}) \\
&\leq f(x_{p,q}) + \lambda^\top p + \mu^\top q
\end{aligned}$$

In particular we have,

$$v(p,q) = \inf_{x_{p,q}} f(x_{p,q}) \geq v(0,0) - \lambda^\top p - \mu^\top q$$

which reads

$$-(\lambda, \mu) \in \partial v(0,0)$$

Exercise

♣ Exercise: Consider the following problem, for $b \in \mathbb{R}$,

$$\begin{aligned} \underset{x \in \mathbb{R}}{\text{Min}} \quad & x^2 \\ \text{s.t.} \quad & x \leq b \end{aligned}$$

1. Does there exist an optimal multiplier?
2. Without solving the dual, give the optimal multiplier μ_b.

Contents

KKT conditions

Recall the first order KKT conditions for our problem (P)

$$\begin{aligned}
&\nabla f(x) + A^\top \lambda + \sum_{j=1}^{n_I} \mu_j \nabla h_j(x) = 0 \\
&Ax = b, \quad h(x) \leq 0 \\
&\lambda \in \mathbb{R}^{n_E}, \quad \mu \in \mathbb{R}_+^{n_I} \\
&\mu_j g_j(x) = 0 && \forall j \in [n_I]
\end{aligned}$$

Further, recall that

- the existence of a Slater's point in a convex problem ensures constraints qualifications,
- first order conditions are sufficient for convex problems.

KKT conditions

Recall the first order KKT conditions for our problem (P)

$$
\begin{aligned}
&\nabla f(x) + A^\top \lambda + \sum_{j=1}^{n_I} \mu_j \nabla h_j(x) = 0 \\
&Ax = b, \quad h(x) \leq 0 \\
&\lambda \in \mathbb{R}^{n_E}, \quad \mu \in \mathbb{R}_+^{n_I} \\
&\mu_j g_j(x) = 0 && \forall j \in [n_I]
\end{aligned}
$$

Further, recall that

- the existence of a Slater's point in a convex problem ensures constraints qualifications,
- first order conditions are sufficient for convex problems.

KKT and duality

If (P) is convex and there exists a Slater's point. Then the following assertions are equivalent:

1. $x^\sharp$ is an optimal solution of (P),
2. $(\exists \lambda^\sharp$ such that$)$ $(x^\sharp, \lambda^\sharp)$ is a saddle point of $\mathcal{L}$,
3. $(\exists \lambda^\sharp$ such that$)$ $(x^\sharp, \lambda^\sharp)$ satisfies the KKT conditions.

Recovering KKT conditions from Lagrangian duality

$$
\begin{aligned}
(P) \quad \underset{x \in \mathbb{R}^n}{\text{Min}} \quad & f(x) \\
\text{s.t.} \quad & A(x) = b \\
& h_j(x) \leq 0 \qquad\qquad \forall j \in [n_I]
\end{aligned}
$$

with associated Lagrangian

$$\mathcal{L}(x; \lambda, \mu) := f(x) + \lambda^\top (Ax - b) + \mu^\top h(x)$$

The KKT conditions can be seen as:

1. $\nabla_x \mathcal{L}(x; \lambda, \mu) = 0$ (Lagrangian minimized in x)
2. $g(x) = 0, \; h(x) \leq 0$ (x primal admissible, also obtained as $\nabla_\lambda \mathcal{L} = 0$)
3. $\mu \geq 0$ ((λ, μ) dual admissible)
4. $\mu_j = 0$ or $h_j(x) = 0$, for all $j \in [n_I]$

(complementarity constraint $\rightsquigarrow 2^{n_I}$ possibilities).

Recovering KKT conditions from Lagrangian duality

$$
\begin{aligned}
(P) \quad \underset{x\in\mathbb{R}^n}{\text{Min}} \quad & f(x) \\
\text{s.t.} \quad & A(x) = b \\
& h_j(x) \leq 0 \qquad\qquad \forall j \in [n_I]
\end{aligned}
$$

with associated Lagrangian

$$\mathcal{L}(x;\lambda,\mu) := f(x) + \lambda^\top (Ax - b) + \mu^\top h(x)$$

The KKT conditions can be seen as:

1. $\nabla_x \mathcal{L}(x;\lambda,\mu) = 0$ (Lagrangian minimized in x)
2. $g(x) = 0$, $h(x) \leq 0$ (x primal admissible, also obtained as $\nabla_\lambda \mathcal{L} = 0$)
3. $\mu \geq 0$ ((λ,μ) dual admissible)
4. $\mu_j = 0$ or $h_j(x) = 0$, for all $j \in [n_I]$

(complementarity constraint $\rightsquigarrow 2^{n_I}$ possibilities).

Complementarity condition and marginal value interpretation

Consider a convex problem satisfying Slater's condition.
Recall that $-\mu^\sharp \in \partial v(0)$ where $v(p)$ is the value of the perturbed problem. From this interpretation, we can recover the complementarity condition

$$\mu_j = 0 \qquad \text{or} \qquad h_j(x) = 0$$

Indeed, let x be an optimal solution.

- If constraint j is not saturated at x (i.e $h_j(x) < 0$), we can marginally move the constraint without affecting the optimal solution, and thus the optimal value. In particular, it means that $\mu_j = 0$.
- If $\mu_j \neq 0$, it means that marginally moving the constraint changes the optimal value and thus the optimal solution. In particular, constraint j must be saturated, i.e $h_j(x) = 0$.

Contents

What you have to know

- Weak duality: $\sup \inf \Phi \leq \inf \sup \Phi$
- Definition of the Lagrangian $\mathcal{L}$
- Definition of primal and dual problem

$$\underbrace{\operatorname*{Max}_{\lambda,\mu} \quad \inf_{x} \quad \mathcal{L}(x;\lambda,\mu)}_{\text{Dual}} \leq \underbrace{\operatorname*{Min}_{x} \quad \sup_{\lambda,\mu} \quad \mathcal{L}(x;\lambda,\mu)}_{\text{Primal}}$$

- Marginal interpretation of the optimal multipliers

What you really should know

- A saddle point of $\mathcal{L}$ is a primal-dual optimal pair
- Sufficient condition of strong duality under convexity (Slater's)

What you have to be able to do

- Turn a constrained optimization problem into an unconstrained Min sup problem through the Lagrangian
- Write the dual of a given problem
- Heuristically recover the KKT conditions from the Lagrangian of a problem

What you should be able to do

- Get lower bounds through duality