Online Learning

— Dual Norms (Prerequisites for Mirror Descent Algorithms)

Joseph Chuang-Chieh Lin

Department of Computer Science & Information Engineering,
Tamkang University

Spring 2023

Credits for the resource

The slides are based on the lectures of Prof. Luca Trevisan:
https://lucatrevisan.github.io/40391/index.html

the lectures of Prof. Shipra Agrawal:
https://ieor8100.github.io/mab/

the lectures of Prof. Francesco Orabona:
https://parameterfree.com/lecture-notes-on-online-learning/
the monograph: https://arxiv.org/abs/1912.13213

and also Elad Hazan's textbook:
Introduction to Online Convex Optimization, 2nd Edition.

Outline

Outline

Norm on a vector space V

- A norm is a real-valued function $f : V \mapsto \mathbb{R}$ satisfying the following properties:
  - Subadditivity:
    - $\forall \mathbf{x}, \mathbf{y} \in V$, $f(\mathbf{x}+\mathbf{y}) \leq f(\mathbf{x}) + f(\mathbf{y})$.
  - Absolute homogeneity:
    - $\forall$ scalar s and $\forall \mathbf{x} \in V$, $f(s\mathbf{x}) = |s| f(\mathbf{x})$.
  - Positive definiteness:
    - $\forall \mathbf{x} \in V$, if $f(\mathbf{x}) = 0$ then $\mathbf{x} = \mathbf{0}$.

Examples

- L_1-norm $\|\cdot\|_1$.
- L_2-norm $\|\cdot\|_2$.
- Infinity-norm $\|\cdot\|_\infty$.
  - $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$.
  - $\|\mathbf{x}\|_\infty = \max_i x_i, \quad \|\mathbf{x}\|_1 = \sum_i |x_i|, \quad \|\mathbf{x}\|_2 = \sqrt{\sum_i |x_i|^2}$.

Outline

Dual Norm

Dual Norm

The dual norm $\|\cdot\|_*$ of a norm $\|\cdot\|$ is defined as

$$\|\boldsymbol{\theta}\|_* := \max_{\mathbf{x}:\|\mathbf{x}\|\leq 1} \langle \boldsymbol{\theta}, \mathbf{x} \rangle = \max_{\mathbf{x}:\|\mathbf{x}\|\leq 1} \boldsymbol{\theta}^\top \mathbf{x}.$$

- A way to measure "how big" are linear functionals.
- Equivalent definition:

$$\|\boldsymbol{\theta}\|_* = \max_{\mathbf{x}\neq\mathbf{0}} \frac{\langle \boldsymbol{\theta}, \mathbf{x} \rangle}{\|\mathbf{x}\|}.$$

Properties of Dual Norms (1/5)

The Dual Norm of L_2-Norm Is the L_2-Norm

$\|\boldsymbol{\theta}\|_* = \|\boldsymbol{\theta}\|_2$.

- $\|\boldsymbol{\theta}\|_* = \max_{\mathbf{x}:\|\mathbf{x}\|_2 \leq 1} \langle \boldsymbol{\theta}, \mathbf{x} \rangle \leq \|\boldsymbol{\theta}\|_2$ (Cauchy-Schwarz Inequality).
- Set $\mathbf{v} = \frac{\boldsymbol{\theta}}{\|\boldsymbol{\theta}\|}$, then $\max_{\mathbf{x}:\|\mathbf{x}\|_2 \leq 1} \langle \boldsymbol{\theta}, \mathbf{x} \rangle \geq \langle \boldsymbol{\theta}, \mathbf{v} \rangle = \|\boldsymbol{\theta}\|_2$.

Properties of Dual Norms (2/5)

Dual Norm of a Matrix Norm

Let $\boldsymbol{A}$ be a positive definite matrix.

- Symmetric & all the eigenvalues (or pivots) are positive.
- For real-valued $\boldsymbol{A}$, $\mathbf{x}^\top \boldsymbol{A}\mathbf{x} > 0$ for all $\mathbf{x}$.

Then $\|\mathbf{x}\|_{\boldsymbol{A}} := \sqrt{\mathbf{x}^\top \boldsymbol{A}\mathbf{x}}$ is a norm.
The dual norm of it is $\|\mathbf{x}\|_{\boldsymbol{A}^{-1}} = \sqrt{\mathbf{x}^\top \boldsymbol{A}^{-1}\mathbf{x}}$.

- $\|\boldsymbol{\theta}\|_* = \max_{\mathbf{x}:\|\mathbf{x}\|_{\boldsymbol{A}} \le 1} \langle \boldsymbol{\theta}, \mathbf{x} \rangle$

Properties of Dual Norms (2/5)

Dual Norm of a Matrix Norm

Let $\boldsymbol{A}$ be a positive definite matrix.

- Symmetric & all the eigenvalues (or pivots) are positive.
- For real-valued $\boldsymbol{A}$, $\mathbf{x}^\top \boldsymbol{A}\mathbf{x} > 0$ for all $\mathbf{x}$.

Then $\|\mathbf{x}\|_{\boldsymbol{A}} := \sqrt{\mathbf{x}^\top \boldsymbol{A}\mathbf{x}}$ is a norm.
The dual norm of it is $\|\mathbf{x}\|_{\boldsymbol{A}^{-1}} = \sqrt{\mathbf{x}^\top \boldsymbol{A}^{-1}\mathbf{x}}$.

- $\|\boldsymbol{\theta}\|_* = \max_{\mathbf{x}:\|\mathbf{x}\|_{\boldsymbol{A}}\leq 1} \langle \boldsymbol{\theta}, \mathbf{x} \rangle = \max_{\mathbf{x}:\mathbf{x}^\top \boldsymbol{A}\mathbf{x}\leq 1} \boldsymbol{\theta}^\top \mathbf{x}$

Properties of Dual Norms (2/5)

Dual Norm of a Matrix Norm

Let $\boldsymbol{A}$ be a positive definite matrix.

- Symmetric & all the eigenvalues (or pivots) are positive.
- For real-valued $\boldsymbol{A}$, $\mathbf{x}^\top \boldsymbol{A}\mathbf{x} > 0$ for all $\mathbf{x}$.

Then $\|\mathbf{x}\|_{\boldsymbol{A}} := \sqrt{\mathbf{x}^\top \boldsymbol{A}\mathbf{x}}$ is a norm.
The dual norm of it is $\|\mathbf{x}\|_{\boldsymbol{A}^{-1}} = \sqrt{\mathbf{x}^\top \boldsymbol{A}^{-1}\mathbf{x}}$.

- $\|\boldsymbol{\theta}\|_* = \max_{\mathbf{x}:\|\mathbf{x}\|_{\boldsymbol{A}}\le 1} \langle \boldsymbol{\theta}, \mathbf{x}\rangle = \max_{\mathbf{x}:\mathbf{x}^\top \boldsymbol{A}\mathbf{x}\le 1} \boldsymbol{\theta}^\top \mathbf{x}$
  $= \max_{\mathbf{y}:\mathbf{y}^\top\mathbf{y}\le 1} \boldsymbol{\theta}^\top \boldsymbol{A}^{-1/2}\mathbf{y} =$

Properties of Dual Norms (2/5)

Dual Norm of a Matrix Norm

Let $\boldsymbol{A}$ be a positive definite matrix.

- Symmetric & all the eigenvalues (or pivots) are positive.
- For real-valued $\boldsymbol{A}$, $\mathbf{x}^\top \boldsymbol{A}\mathbf{x} > 0$ for all $\mathbf{x}$.

Then $\|\mathbf{x}\|_{\boldsymbol{A}} := \sqrt{\mathbf{x}^\top \boldsymbol{A}\mathbf{x}}$ is a norm.
The dual norm of it is $\|\mathbf{x}\|_{\boldsymbol{A}^{-1}} = \sqrt{\mathbf{x}^\top \boldsymbol{A}^{-1}\mathbf{x}}$.

- $\|\boldsymbol{\theta}\|_* = \max_{\mathbf{x}:\|\mathbf{x}\|_{\boldsymbol{A}}\le 1} \langle \boldsymbol{\theta}, \mathbf{x}\rangle = \max_{\mathbf{x}:\mathbf{x}^\top \boldsymbol{A}\mathbf{x}\le 1} \boldsymbol{\theta}^\top \mathbf{x}$
  $= \max_{\mathbf{y}:\mathbf{y}^\top \mathbf{y}\le 1} \boldsymbol{\theta}^\top \boldsymbol{A}^{-1/2}\mathbf{y} = \max_{\|\mathbf{y}\|_2\le 1} (\boldsymbol{A}^{-1/2}\boldsymbol{\theta})^\top \mathbf{y}$

Properties of Dual Norms (2/5)

Dual Norm of a Matrix Norm

Let $\boldsymbol{A}$ be a positive definite matrix.

- Symmetric & all the eigenvalues (or pivots) are positive.
- For real-valued $\boldsymbol{A}$, $\mathbf{x}^\top \boldsymbol{A}\mathbf{x} > 0$ for all $\mathbf{x}$.

Then $\|\mathbf{x}\|_{\boldsymbol{A}} := \sqrt{\mathbf{x}^\top \boldsymbol{A}\mathbf{x}}$ is a norm.
The dual norm of it is $\|\mathbf{x}\|_{\boldsymbol{A}^{-1}} = \sqrt{\mathbf{x}^\top \boldsymbol{A}^{-1}\mathbf{x}}$.

- $\|\boldsymbol{\theta}\|_* = \max_{\mathbf{x}:\|\mathbf{x}\|_{\boldsymbol{A}}\leq 1}\langle \boldsymbol{\theta}, \mathbf{x}\rangle = \max_{\mathbf{x}:\mathbf{x}^\top \boldsymbol{A}\mathbf{x}\leq 1} \boldsymbol{\theta}^\top \mathbf{x}$
  $= \max_{\mathbf{y}:\mathbf{y}^\top \mathbf{y}\leq 1} \boldsymbol{\theta}^\top \boldsymbol{A}^{-1/2}\mathbf{y} = \max_{\|\mathbf{y}\|_2\leq 1}(\boldsymbol{A}^{-1/2}\boldsymbol{\theta})^\top \mathbf{y}$
  $= \|\boldsymbol{A}^{-1/2}\boldsymbol{\theta}\|_2 =$

Properties of Dual Norms (2/5)

Dual Norm of a Matrix Norm

Let $\boldsymbol{A}$ be a positive definite matrix.

- Symmetric & all the eigenvalues (or pivots) are positive.
- For real-valued $\boldsymbol{A}$, $\mathbf{x}^\top \boldsymbol{A}\mathbf{x} > 0$ for all $\mathbf{x}$.

Then $\|\mathbf{x}\|_{\boldsymbol{A}} := \sqrt{\mathbf{x}^\top \boldsymbol{A}\mathbf{x}}$ is a norm.
The dual norm of it is $\|\mathbf{x}\|_{\boldsymbol{A}^{-1}} = \sqrt{\mathbf{x}^\top \boldsymbol{A}^{-1}\mathbf{x}}$.

- $\|\boldsymbol{\theta}\|_* = \max_{\mathbf{x}:\|\mathbf{x}\|_{\boldsymbol{A}}\le 1} \langle \boldsymbol{\theta}, \mathbf{x}\rangle = \max_{\mathbf{x}:\mathbf{x}^\top \boldsymbol{A}\mathbf{x}\le 1} \boldsymbol{\theta}^\top \mathbf{x}$
  $= \max_{\mathbf{y}:\mathbf{y}^\top\mathbf{y}\le 1} \boldsymbol{\theta}^\top \boldsymbol{A}^{-1/2}\mathbf{y} = \max_{\|\mathbf{y}\|_2\le 1} (\boldsymbol{A}^{-1/2}\boldsymbol{\theta})^\top \mathbf{y}$
  $= \|\boldsymbol{A}^{-1/2}\boldsymbol{\theta}\|_2 = \sqrt{\boldsymbol{\theta}^\top \boldsymbol{A}^{-1}\boldsymbol{\theta}}$.
  - Use the trick of change of variable $\mathbf{y} = \boldsymbol{A}^{1/2}\mathbf{x}$.
  - The dual norm of the L_2-norm is the L_2-norm,

Properties of Dual Norms (3/5)

Cauchy-Schwarz Inequality

$\langle \boldsymbol{\theta}, \mathbf{x} \rangle \leq \|\boldsymbol{\theta}\|_* \|\mathbf{x}\|$.

- From the equivalent definition of the dual norm of $\boldsymbol{\theta}$.

Properties of Dual Norms (3/5)

Cauchy-Schwarz Inequality

$\langle \boldsymbol{\theta}, \mathbf{x} \rangle \leq \|\boldsymbol{\theta}\|_* \|\mathbf{x}\|$.

- From the equivalent definition of the dual norm of $\boldsymbol{\theta}$.
- For any $\theta, \mathbf{x}$, we have

$$\left\langle \boldsymbol{\theta}, \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\rangle \leq \max_{\mathbf{x}} \left\langle \boldsymbol{\theta}, \frac{\mathbf{x}}{\|\mathbf{x}\|} \right\rangle = \|\boldsymbol{\theta}\|_*.$$

Properties of Dual Norms (4/5)

Dual Norm of L_p-Norm for $1/p + 1/q = 1$

The dual norm of the the L_p-norm is the L_q-norm for $\frac{1}{p} + \frac{1}{q} = 1$, $p, q > 1$.

- We shall show that

$$\max_{\mathbf{x} \in \mathbb{R}^d, \|\mathbf{x}\|_p = 1} \sum_{i=1}^{d} \theta_i x_i = \|\boldsymbol{\theta}\|_q .$$

  - $\|\mathbf{x}\|_p = (\sum_{i=1}^{d} |x_i|^p)^{1/p}$.

Properties of Dual Norms (4/5)

Dual Norm of L_p-Norm for $1/p + 1/q = 1$

The dual norm of the the L_p-norm is the L_q-norm for $\frac{1}{p} + \frac{1}{q} = 1$, $p, q > 1$.

- We shall show that

$$\max_{\mathbf{x} \in \mathbb{R}^d, \|\mathbf{x}\|_p = 1} \sum_{i=1}^{d} \theta_i x_i = \|\boldsymbol{\theta}\|_q.$$

  - $\|\mathbf{x}\|_p = (\sum_{i=1}^{d} |x_i|^p)^{1/p}$.

- Hölder's Inequality:

$$\sum_{i=1}^{d} |x_i y_i| \leq \left(\sum_{i=1}^{d} |x_i|^p \right)^{1/p} \left(\sum_{i=1}^{d} |y_i|^q \right)^{1/q}$$

for $p, q > 1, 1/p + 1/q = 1$.

Properties of Dual Norms (5/5)

Dual Norm of L_1-Norm is the L_∞-Norm

With respect to the L_1-Norm, we have

$$\|\boldsymbol{\theta}\|_* = \|\boldsymbol{\theta}\|_\infty.$$

$$\begin{aligned}
\|\boldsymbol{\theta}\|_* &= \max_{\mathbf{x}:\sum_i |x_i| \leq 1} \langle \boldsymbol{\theta}, \mathbf{x} \rangle \\
&= \max_{\mathbf{x}:\sum_i |x_i| \leq 1} \sum_i \theta_i x_i \\
&= \max_i |\theta_i| \\
&= \|\boldsymbol{\theta}\|_\infty.
\end{aligned}$$

Inner Product & Norm

- Every inner product space induces a norm.
  - $\|\mathbf{x}\| := \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$.

Exercise

Prove that

$$\|\boldsymbol{\theta}\| \leq \|\boldsymbol{\theta}\|_*$$

and

$$\|\boldsymbol{\theta}\|^2 \leq \|\boldsymbol{\theta}\|_* \|\boldsymbol{\theta}\|_{**}.$$

Take Home Exercise & Further Study

Dual Norm of a Dual Norm is the Primal Norm

Prove that

$$\|\boldsymbol{\theta}\|_{**} = \|\boldsymbol{\theta}\|.$$

Take Home Exercise & Further Study

Dual Norm of a Dual Norm is the Primal Norm

Prove that

$$\|\boldsymbol{\theta}\|_{**} = \|\boldsymbol{\theta}\|.$$

- *A possible way:*

$$\min_{\boldsymbol{\theta}} \|\boldsymbol{\theta}\| \quad \text{subject to} \quad \boldsymbol{\theta} = \mathbf{x}.$$

Outline

Smoothness

Smooth Functions

- Let $f : V \mapsto \mathbb{R}$ be a function differentiable in an open set $\supseteq V$.
- We say that f is M-smooth w.r.t. $\|\cdot\|$ if $\|\nabla f(\mathbf{x}) - \nabla f(\mathbf{y})\|_* \leq M\|\mathbf{x} - \mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in V$.

Lemma [Bound w.r.t. Linear Approximation]

Let $f : V \mapsto \mathbb{R}$ be an M-smooth function. Then, for any $\mathbf{x}, \mathbf{y} \in V$, we have

$$|f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle| \leq \frac{M}{2}\|\mathbf{y} - \mathbf{x}\|^2.$$

Proof of the Lemma

$$
\begin{aligned}
f(\mathbf{y}) &= f(\mathbf{x}) + \int_0^1 \langle \nabla f(\mathbf{x} + \tau(\mathbf{y} - \mathbf{x})), \mathbf{y} - \mathbf{x} \rangle d\tau \\
&= f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \int_0^1 \langle \nabla f(\mathbf{x} + \tau(\mathbf{y} - \mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle d\tau
\end{aligned}
$$

Then

Proof of the Lemma

$$\begin{aligned} f(\mathbf{y}) &= f(\mathbf{x}) + \int_0^1 \langle \nabla f(\mathbf{x}+\tau(\mathbf{y}-\mathbf{x})), \mathbf{y}-\mathbf{x}\rangle d\tau \\ &= f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle + \int_0^1 \langle \nabla f(\mathbf{x}+\tau(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle d\tau \end{aligned}$$

Then

$$\begin{aligned} |f(\mathbf{y}) - f(\mathbf{x}) - \langle \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle| &= \left| \int_0^1 \langle \nabla f(\mathbf{x}+\tau(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle d\tau \right| \\ &\leq \int_0^1 |\langle \nabla f(\mathbf{x}+\tau(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x}), \mathbf{y}-\mathbf{x}\rangle| \, d\tau \\ &\leq \int_0^1 \|\nabla f(\mathbf{x}+\tau(\mathbf{y}-\mathbf{x})) - \nabla f(\mathbf{x})\|_* \, \|\mathbf{y}-\mathbf{x}\| \, d\tau \\ &\leq \int_0^1 \tau M \|\mathbf{y}-\mathbf{x}\|^2 d\tau = \frac{M}{2}\|\mathbf{y}-\mathbf{x}\|^2. \end{aligned}$$

Another Useful Theorem

Theorem

Let $f : \mathbb{R}^d \mapsto \mathbb{R}$ be M-smooth and bounded from below, then for all $\mathbf{x} \in \mathbb{R}^d$,

$$\|\nabla f(\mathbf{x})\|_*^2 \le 2M \left(f(\mathbf{x}) - \inf_{\mathbf{y} \in \mathbb{R}^d} f(\mathbf{y}) \right).$$

- Proof left as an exercise (refer to the monograph).

Outline

Click for the reference material.

Primal Problem

- Consider a possibly non-convex optimization problem:

Primal Problem

$$p^* := \min_{\mathbf{x} \in \mathcal{D}} F(\mathbf{x})$$

subject to

$$f_i(\mathbf{x}) \leq 0, \ \text{for } i = 1, 2, \ldots, m.$$

- $\mathcal{D} \subseteq \mathbb{R}^n$: the domain of the problem
- $\mathcal{X} \subseteq \mathcal{D}$: the feasible solution space.
- $\mathbf{x} \in \mathbb{R}^n$: primal variable.

Lagrangian

- $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$, such that

$$\mathcal{L}(\mathbf{x}, \lambda) := F(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}).$$

Lagrangian

- $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$, such that

$$\mathcal{L}(\mathbf{x}, \lambda) := F(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}).$$

- $\lambda \in \mathbb{R}^m$: Lagrange multipliers.

Lagrangian

- $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$, such that

$$\mathcal{L}(\mathbf{x}, \lambda) := F(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}).$$

- $\lambda \in \mathbb{R}^m$: Lagrange multipliers.
- Observation:

$$\forall \mathbf{x} \in \mathcal{X}, \forall \lambda \in \mathbf{R}_+^m : \quad F(\mathbf{x}) \geq \mathcal{L}(\mathbf{x}, \lambda).$$

Lagrangian

- $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$, such that

$$\mathcal{L}(\mathbf{x}, \lambda) := F(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}).$$

- $\lambda \in \mathbb{R}^m$: Lagrange multipliers.
- Observation:

$$\forall \mathbf{x} \in \mathcal{X}, \forall \lambda \in \mathbf{R}_+^m : \quad F(\mathbf{x}) \geq \mathcal{L}(\mathbf{x}, \lambda).$$

- The purpose of Lagrangian:

$$p^* = \min_{\mathbf{x} \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}_+^m} \mathcal{L}(\mathbf{x}, \lambda).$$

Lagrangian

- $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$, such that
$$\mathcal{L}(\mathbf{x}, \lambda) := F(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}).$$
- $\lambda \in \mathbb{R}^m$: Lagrange multipliers.
- Observation:
$$\forall \mathbf{x} \in \mathcal{X}, \forall \lambda \in \mathbf{R}_+^m : \quad F(\mathbf{x}) \geq \mathcal{L}(\mathbf{x}, \lambda).$$
- The purpose of Lagrangian:
$$p^* = \min_{\mathbf{x} \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}_+^m} \mathcal{L}(\mathbf{x}, \lambda). \Rightarrow \text{Unconstrained!}$$

- Note that for any $\mathbf{z} \in \mathbb{R}^m$,

$$\max_{\lambda \succeq \mathbf{0}} \lambda^\top \mathbf{z} = \begin{cases} 0 & \text{if } \mathbf{z} \preceq \mathbf{0} \\ +\infty & \text{otherwise} \end{cases}$$

Lagrange Dual Function (1/2)

Lagrange Dual Function

$$g(\lambda) := \min_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \lambda).$$

Lagrange Dual Function (1/2)

Lagrange Dual Function

$$g(\lambda) := \min_{\mathbf{x} \in \mathbb{R}^n} \mathcal{L}(\mathbf{x}, \lambda).$$

- $\mathcal{L}(\mathbf{x}, \cdot)$ is affine for every $\mathbf{x}$.
- g is pointwise minimum of affine functions.

⋆ g is concave.

Lagrange Dual Function (2/2)

$$\forall \mathbf{x} \in \mathcal{X}, \forall \lambda \in \mathbf{R}_+^m : \quad F(\mathbf{x}) \geq \min_{\mathbf{x}'} \mathcal{L}(\mathbf{x}', \lambda) = g(\lambda)$$

Minimizing the left-hand side, we have

Lagrange Dual Function (2/2)

$$\forall \mathbf{x} \in \mathcal{X}, \forall \lambda \in \mathbf{R}_+^m : \quad F(\mathbf{x}) \geq \min_{\mathbf{x}'} \mathcal{L}(\mathbf{x}', \lambda) = g(\lambda)$$

Minimizing the left-hand side, we have

$$\forall \lambda \in \mathbf{R}_+^m : \quad p^* \geq g(\lambda).$$

The Lagrange Dual Problem

The best lower bound is by using $p^* \geq d^*$, where

$$d^* = \max_{\lambda \succeq \mathbf{0}} g(\lambda).$$

Lagrange Dual Function (2/2)

$$\forall \mathbf{x} \in \mathcal{X}, \forall \lambda \in \mathbf{R}_+^m : \quad F(\mathbf{x}) \geq \min_{\mathbf{x}'} \mathcal{L}(\mathbf{x}', \lambda) = g(\lambda)$$

Minimizing the left-hand side, we have

$$\forall \lambda \in \mathbf{R}_+^m : \quad p^* \geq g(\lambda).$$

The Lagrange Dual Problem

The best lower bound is by using $p^* \geq d^*$, where

$$d^* = \max_{\lambda \succeq \mathbf{0}} g(\lambda).$$

- $p^* - d^*$: The duality gap.

Case with Equality Constraints

- If there are equality constraints such as $h_i(\mathbf{x}) = 0$ for some i's,

$$
\begin{aligned}
h_i(\mathbf{x}) &\le 0, \\
-h_i(\mathbf{x}) &\le 0
\end{aligned}
$$

Case with Equality Constraints

- If there are equality constraints such as $h_i(\mathbf{x}) = 0$ for some i's,

$$h_i(\mathbf{x}) \leq 0,$$
$$-h_i(\mathbf{x}) \leq 0$$

- Then,

$$\begin{aligned}\mathcal{L}(\mathbf{x}, \lambda, \nu^+, \nu^-) &= F(\mathbf{x}) + \sum_i \lambda_i f_i(\mathbf{x}) + \sum_j \nu_j^+ h_j(\mathbf{x}) + \sum_j \nu_j^- (-h_j(\mathbf{x})) \\ &= F(\mathbf{x}) + \sum_i \lambda_i f_i(\mathbf{x}) + \sum_j \nu_j h_j(\mathbf{x}),\end{aligned}$$

where $\nu := \nu^+ + \nu^-$.

Minimax Inequality

Minimax Inequality

For any function f of two vector variables $\mathbf{x}, \mathbf{y}$ and domains $\mathcal{X}$, $\mathcal{Y}$,

$$\max_{\mathbf{y}\in\mathcal{Y}} \min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x}, \mathbf{y}) \leq \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{x}, \mathbf{y}).$$

- Observe that

$$\min_{\mathbf{x}'\in\mathcal{X}} f(\mathbf{x}', \mathbf{y}) \leq$$

$$\max_{\mathbf{y}'\in\mathcal{Y}} f(\mathbf{x}, \mathbf{y}').$$

Minimax Inequality

Minimax Inequality

For any function f of two vector variables $\mathbf{x}, \mathbf{y}$ and domains $\mathcal{X}$, $\mathcal{Y}$,

$$\max_{\mathbf{y}\in\mathcal{Y}} \min_{\mathbf{x}\in\mathcal{X}} f(\mathbf{x},\mathbf{y}) \le \min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} f(\mathbf{x},\mathbf{y}).$$

- Observe that

$$\begin{aligned} \min_{\mathbf{x}'\in\mathcal{X}} f(\mathbf{x}',\mathbf{y}) \le f(\mathbf{x},\mathbf{y}) &\le \\ f(\mathbf{x},\mathbf{y}) &\le \max_{\mathbf{y}'\in\mathcal{Y}} f(\mathbf{x},\mathbf{y}'). \end{aligned}$$

Slater's Condition

Slater's Condition

There exists $\mathbf{x}_0 \in \mathcal{D}$ such that $f_i(\mathbf{x}_0) < 0 \ \forall i$ and $h_j(\mathbf{x}_0) = 0, \ \forall j$.

Slater's Condition

Slater's Condition

There exists $\mathbf{x}_0 \in \mathcal{D}$ such that $f_i(\mathbf{x}_0) < 0\ \forall i$ and $h_j(\mathbf{x}_0) = 0,\ \forall j$.

- A *weak* form of Slater's condition: strictly feasibility is not required whenever f_i is affine.

Slater's Condition

Slater's Condition

There exists $\mathbf{x}_0 \in \mathcal{D}$ such that $f_i(\mathbf{x}_0) < 0 \; \forall i$ and $h_j(\mathbf{x}_0) = 0, \; \forall j$.

- A *weak* form of Slater's condition: strictly feasibility is not required whenever f_i is affine.

Strong Duality

If the primal problem is convex and satisfies the weak Slater's condition, then the strong duality holds, that is,

$$p^* = d^*.$$

Discussions