Lagrange duality

Mathematical Optimization

Georgiy Kormakov

CMC MSU

15 March 2024

Dual function for norm

Example

Let $\|\cdot\|$ be the norm for $\mathbb{R}^n$, and $\|\cdot\|_*$ be the dual norm. Prove that the conjugate function for $f(x) = \|x\|$ is $f^*(y) = 0$ with $dom\, f^* = B_{\|\cdot\|_*}(0, 1)$.

$\boxed{\|y\|_* > 1:}$

$\boxed{\|y\|_* \leqslant 1:}$

Standart form

$$
\begin{aligned}
\min_x \ & f_0(x) \\
\text{s.t. } & f_i(x) \leqslant 0, \ i = 1, \dots, m \\
& h_j(x) = 0, \ j = 1, \dots, n,
\end{aligned} \tag{1}
$$

Standart form

$$\begin{aligned} \min_x \ & f_0(x) \\ \text{s.t. } & f_i(x) \leqslant 0, \ i = 1, \dots, m \\ & h_j(x) = 0, \ j = 1, \dots, n, \end{aligned} \tag{1}$$

$x \in \mathbb{R}^d$, but the whole domain is $\mathrm{D} = \cap_{i=0}^{m} \mathbf{dom}\ f_i \cap_{j=1}^{n} \mathbf{dom}\ h_j$.

Standart form

$$\begin{aligned} \min_x \ & f_0(x) \\ \text{s.t. } & f_i(x) \leqslant 0, \ i = 1, \ldots, m \\ & h_j(x) = 0, \ j = 1, \ldots, n, \end{aligned} \tag{1}$$

$x \in \mathbb{R}^d$, but the whole domain is $\mathrm{D} = \cap_{i=0}^{m} \mathbf{dom}\ f_i \cap_{j=1}^{n} \mathbf{dom}\ h_j$.

Feasible set

$$F = \{x \in D | f_i(x) \leqslant 0, i = \overline{1, m},\ h_j(x) = 0, \forall j\}$$

Standart form

$$\begin{aligned} &\min_x \ f_0(x) \\ &\text{s.t. } f_i(x) \leqslant 0, \ i = 1, \ldots, m \\ &\qquad h_j(x) = 0, \ j = 1, \ldots, n, \end{aligned} \tag{1}$$

$x \in \mathbb{R}^d$, but the whole domain is $\mathrm{D} = \cap_{i=0}^m \mathbf{dom}\ f_i \cap_{j=1}^n \mathbf{dom}\ h_j$.

Feasible set

$$F = \{x \in D | f_i(x) \leqslant 0, i = \overline{1, m}, \ h_j(x) = 0, \forall j\}$$

Lagrangian

For the optimisation problem 1, the *Lagrangian* is equal to

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{j=1}^n \nu_j h_j(x).$$

Lagrangian Dual Function

Definition

Let's define the *Lagrangian dual function* (or just the dual function) $g : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ as follows:

$$g(\lambda, \nu) = \inf_{x \in \mathrm{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{n} \nu_j h_j(x) \right) = \inf_{x \in \mathrm{D}} L(x, \lambda, \nu) \quad (2)$$

Lagrangian Dual Function

Definition

Let's define the *Lagrangian dual function* (or just the dual function) $g : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}$ as follows:

$$g(\lambda, \nu) = \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{n} \nu_j h_j(x) \right) = \inf_{x \in D} L(x, \lambda, \nu) \quad (2)$$

Property 1

$g(\lambda, \nu)$ is always concave function of (λ, ν)

Lagrangian Dual Function. Properties

Property 2

$g(\lambda, \nu)$ is smaller then optimum p^* of 1 for $\lambda \geq 0$ and any ν:

$$\forall \lambda \geq 0, \nu \in \mathbb{R}^n \quad g(\lambda, \nu) \leqslant p^*$$

Geometrical interpretation of Lagrangian

$$I_-(x) = \begin{cases} 0, & x \leqslant 0 \\ \infty, & \text{else} \end{cases} \qquad I_0(x) = \begin{cases} 0, & x = 0 \\ \infty, & \text{else} \end{cases}$$

Geometrical interpretation of Lagrangian

$$I_-(x) = \begin{cases} 0, & x \leqslant 0 \\ \infty, & \text{else} \end{cases} \qquad I_0(x) = \begin{cases} 0, & x = 0 \\ \infty, & \text{else} \end{cases}$$

$$\min_x \; f_0(x) + \sum_{i=1}^{m} I_-(f_i(x)) + \sum_{j=1}^{n} I_0(h_j(x))$$

Dual problem

Definition

The dual problem for 1 has the following form

$$\max_{\lambda,\nu} g(\lambda,\nu)$$
$$\text{s.t. } \lambda \geq 0. \tag{3}$$

Linear equations with minimal norm

Consider the following optimization problem

$$\begin{aligned} &\min_x \; x^T x \\ &\text{s.t. } \; Ax = b, \end{aligned}$$

where $A \in \mathbb{R}^{p \times d}$, $b \in \mathbb{R}^p$. Formulate the dual problem.

Linear equations with minimal norm

Consider the following optimization problem

$$\min_x x^T x$$
$$\text{s.t. } Ax = b,$$

where $A \in \mathbb{R}^{p \times d}$, $b \in \mathbb{R}^p$. Formulate the dual problem.

$x^* = -\frac{1}{2}A^T\nu \Rightarrow q(\nu) =$

Linear equations with minimal norm

Consider the following optimization problem

$$\min_x \; x^T x$$
$$\text{s.t. } Ax = b,$$

where $A \in \mathbb{R}^{p \times d}$, $b \in \mathbb{R}^p$. Formulate the dual problem.

$x^* = -\frac{1}{2}A^T\nu \Rightarrow q(\nu) = -\frac{1}{4}\nu^T AA^T\nu - \nu^T b$

Linear Programming

$$\min_x \; c^T x$$
$$\text{s.t. } Ax = b,$$
$$x \succeq 0,$$

where $A \in \mathbb{R}^{p \times d}$, $b \in \mathbb{R}^p$.

Linear Programming

$$\begin{aligned} \min_x \ & c^T x \\ \text{s.t. } & Ax = b, \\ & x \succeq 0, \end{aligned}$$

where $A \in \mathbb{R}^{p \times d}$, $b \in \mathbb{R}^p$.

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

Linear Programming

$$\begin{aligned}\min_x \ & c^T x\\ \text{s.t. } & Ax = b,\\ & x \succeq 0,\end{aligned}$$

where $A \in \mathbb{R}^{p\times d}$, $b \in \mathbb{R}^p$.

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

$$\begin{aligned}\max_{\nu,\lambda} \ & -\nu^T b\\ \text{s.t. } & c + A^T\nu - \lambda = 0,\\ & \lambda \succeq 0.\end{aligned}$$

Splitting problem

$$\min_x \; x^T W x$$
$$\text{s.t. } x_j^2 = 1, \quad j = 1, \dots d,$$

where $W \in \mathcal{S}^d$.

Splitting problem

$$\min_x \; x^T W x$$
$$\text{s.t. } x_j^2 = 1, \quad j = 1, \dots d,$$

where $W \in \mathcal{S}^d$.

$$g(\nu) = \begin{cases} -1^T\nu, \; W + \text{diag}(\nu) \succeq 0, \\ -\infty, \; \text{else}, \end{cases}$$

Splitting problem

$$\min_x \; x^T W x$$
$$\text{s.t. } \; x_j^2 = 1, \quad j = 1, \dots d,$$

where $W \in \mathcal{S}^d$.

$$g(\nu) = \begin{cases} -1^T\nu, & W + \mathrm{diag}(\nu) \succeq 0, \\ -\infty, & \text{else}, \end{cases}$$

$$\max_\nu \; -1^T\nu$$
$$\text{s.t. } \; W + \mathrm{diag}(\nu) \succeq 0.$$

Illustration

$$f^*(y) = \sup_{x \in \mathbf{dom}\ f} (y^T x - f(x)).$$

The simpliest problem

$$\min_x \ f(x)$$
$$\text{s.t.}\ \ x = 0$$

Illustration

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} (y^T x - f(x)).$$

The simpliest problem

$$\min_x f(x)$$
$$\text{s.t. } x = 0$$

$$g(\nu) = \inf_x (f(x) + \nu^T x) = -\sup_x(-\nu^T x - f(x)) = -f^*(-\nu).$$

Generalisation

$$\begin{aligned} \min_x \ & f(x) \\ \text{s.t. } & Ax \preceq b, \\ & Cx = d. \end{aligned}$$

Generalisation

$$
\begin{aligned}
\min_x \ & f(x) \\
\text{s.t. } & Ax \preceq b, \\
& Cx = d.
\end{aligned}
$$

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_x (f(x) + \lambda^T (Ax - b) + \nu^T (Cx - d)) \\
&= -\lambda^T b - \nu^T d + \inf_x (f(x) + (A^T \lambda + C^T \nu)^T x) \\
&= -\lambda^T b - \nu^T d - f^*(-A^T \lambda - C^T \nu).
\end{aligned}
$$

System of equations. Again

$$\min_x \|x\|$$
$$\text{s.t. } Ax = b,$$

where $\|\cdot\|$ - any norm.

System of equations. Again

$$\min_x \|x\|$$
$$\text{s.t. } Ax = b,$$

where $\|\cdot\|$ - any norm.

$$g(\nu) = -\nu^T b - f^*(-A^T\nu) = \begin{cases} -\nu^T b, & \|A^T\nu\|_* \leqslant 1 \\ -\infty, & \text{else.} \end{cases}$$

System of equations. Again

$$
\min_x \; \|x\| \\
\text{s.t. } Ax = b,
$$

where $\|\cdot\|$ - any norm.

$$
g(\nu) = -\nu^T b - f^*(-A^T\nu) = \begin{cases} -\nu^T b, & \|A^T\nu\|_* \leqslant 1 \\ -\infty, & \text{else.} \end{cases}
$$

$$
\max_\nu \; -\nu^T b \\
\text{s.t. } \|A^T\nu\|_* \leqslant 1.
$$

Max Entropy

$$\begin{aligned} \min_x \; & \sum_{i=1}^{d} x_i \log x_i \\ \text{s.t. } & Ax \preceq b, \\ & 1^T x = 1, \end{aligned}$$

where $\mathbf{dom}\ f = \mathbb{R}^d_{++}$.

Max Entropy

$$
\begin{aligned}
\min_x \ & \sum_{i=1}^{d} x_i \log x_i \\
\text{s.t. } & Ax \preceq b, \\
& 1^T x = 1,
\end{aligned}
$$

where **dom** $f = \mathbb{R}^d_{++}$.

$$
f^*(y) = \sum_{i=1}^{d} e^{y_i - 1}.
$$

Max Entropy

$$\min_x \sum_{i=1}^{d} x_i \log x_i$$
$$\text{s.t. } Ax \preceq b,$$
$$1^T x = 1,$$

where **dom** $f = \mathbb{R}^d_{++}$.

$$f^*(y) = \sum_{i=1}^{d} e^{y_i - 1}.$$

$$g(\lambda, \nu) = -\lambda^T b - \nu - \sum_{i=1}^{d} e^{-a_i^T \lambda - \nu - 1},$$

Max Entropy

$$\min_x \sum_{i=1}^d x_i \log x_i$$
$$\text{s.t. } Ax \preceq b,$$
$$1^T x = 1,$$

where **dom** $f = \mathbb{R}^d_{++}$.

$$f^*(y) = \sum_{i=1}^d e^{y_i - 1}.$$

$$g(\lambda, \nu) = -\lambda^T b - \nu - \sum_{i=1}^d e^{-a_i^T \lambda - \nu - 1},$$

$$\max_{\lambda,\nu} \left[-\lambda^T b - \nu - \sum_{i=1}^d e^{-a_i^T \lambda - \nu - 1} \right] \text{s.t. } \lambda \geq 0.$$

Change of variables

$$\min_x \sum_{i=1}^{N} \|A_i x + b_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2.$$

Change of variables

$$\min_x \sum_{i=1}^{N} \|A_i x + b_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2.$$

$$\begin{aligned} &\min_x \sum_{i=1}^{N} \|y_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2 \\ &\text{s.t. } A_j x + b_j = y_j \ \forall j \end{aligned}$$

Change of variables

$$\min_x \sum_{i=1}^N \|A_i x + b_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2.$$

$$\min_x \sum_{i=1}^N \|y_i\|_2 + \frac{1}{2}\|x - x_0\|_2^2$$
$$\text{s.t. } A_j x + b_j = y_j \ \forall j$$

$$g(\nu_1, \dots, \nu_N) = \begin{cases} \sum_{i=1}^N (A_i x_0 + b_i)^T \nu_j - \frac{1}{2}\|\sum_{j=1}^N A_j^T \nu_i\|_2^2, & \|\nu_j\|_2 \leqslant 1 \ \forall i \\ -\infty, & \text{else} \end{cases}$$

Piece-wise linear

$$\min_{x} \max_{i=1,\dots,m} (a_i^T x + b_i)$$

Piece-wise linear

$$\min_x \max_{i=1,\dots,m} (a_i^T x + b_i)$$

$$\min_x \max_{i=1,\dots,m} y_i$$
$$\text{s.t. } a_j^T x + b_j = y_j \ \forall j.$$

Piece-wise linear

$$\min_x \max_{i=1,\dots,m} (a_i^T x + b_i)$$

$$\begin{aligned} \min_x \max_{i=1,\dots,m} & \; y_i \\ \text{s.t. } & a_j^T x + b_j = y_j \; \forall j. \end{aligned}$$

$$\inf_y (\max_i y_i - \nu^T y) = \begin{cases} 0, & \nu \succeq 0, \; 1^T \nu = 1 \\ -\infty, & \text{else} \end{cases}$$

Piece-wise linear

$$\min_x \max_{i=1,\dots,m} (a_i^T x + b_i)$$

$$\min_x \max_{i=1,\dots,m} y_i$$
$$\text{s.t. } a_j^T x + b_j = y_j \ \forall j.$$

$$\inf_y (\max_i y_i - \nu^T y) = \begin{cases} 0, \ \nu \succeq 0, \ 1^T \nu = 1 \\ -\infty, \quad \text{else} \end{cases}$$

$$g(\nu) = \begin{cases} b^T \nu, \ \sum_{i=1}^m \nu_i a_i = 0, \ \nu \succeq 0, \ 1^T \nu = 1 \\ -\infty, \quad \text{else} \end{cases}$$