Lagrange Duality

Rishabh Iyer

Lecture 18

Optimization in Machine Learning, UT Dallas

Convex Optimization Problems

Definition

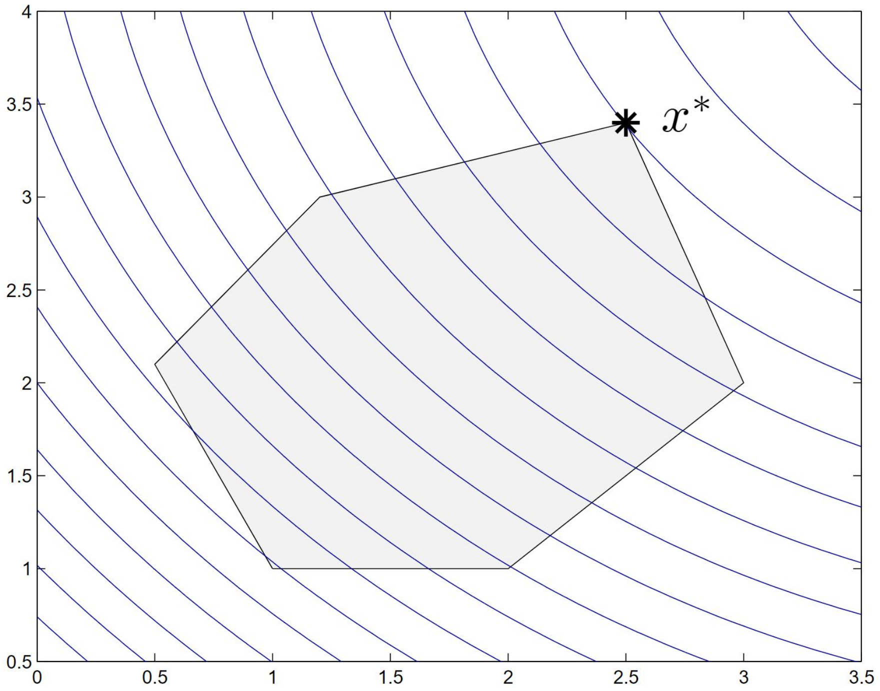
An optimization problem is *convex* if its objective is a convex function, the inequality constraints f_j are convex, and the equality constraints h_j are affine

$$
\begin{aligned}
\underset{x}{\text{minimize}} \ & f_0(x) && \text{(Convex function)} \\
\text{s.t.} \quad & f_i(x) \leq 0 && \text{(Convex sets)} \\
& h_j(x) = 0 && \text{(Affine)}
\end{aligned}
$$

It's nice to be convex

Theorem

If $\hat{x}$ is a local minimizer of a convex optimization problem, it is a global minimizer.



Goals of Lagrange Duality

- Get certificate for optimality of a problem
- Remove constraints
- Reformulate problem

Constructing the dual

- Start with optimization problem:

$$\begin{aligned} \underset{x}{\text{minimize}}\ & f_0(x) \\ \text{s.t.}\ & f_i(x) \le 0,\ i = \{1, \dots, k\} \\ & h_j(x) = 0,\ j = \{1, \dots, l\} \end{aligned}$$

Constructing the dual

- Start with optimization problem:

$$\underset{x}{\text{minimize}}\ f_0(x)$$
$$\text{s.t.}\quad f_i(x) \leq 0,\ i = \{1, \ldots, k\}$$
$$h_j(x) = 0,\ j = \{1, \ldots, l\}$$

- Form *Lagrangian* using Lagrange multipliers $\lambda_i \geq 0$, $\nu_i \in \mathbb{R}$

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{k} \lambda_i f_i(x) + \sum_{j=1}^{l} \nu_j h_j(x)$$

Constructing the dual

- Start with optimization problem:

$$\begin{aligned} \underset{x}{\text{minimize}} \ & f_0(x) \\ \text{s.t.} \ & f_i(x) \leq 0, \ i = \{1, \ldots, k\} \\ & h_j(x) = 0, \ j = \{1, \ldots, l\} \end{aligned}$$

- Form *Lagrangian* using Lagrange multipliers $\lambda_i \geq 0$, $\nu_i \in \mathbb{R}$

$$\mathcal{L}(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{k} \lambda_i f_i(x) + \sum_{j=1}^{l} \nu_j h_j(x)$$

- Form *dual function*

$$g(\lambda, \nu) = \inf_x \mathcal{L}(x, \lambda, \nu) = \inf_x \left\{ f_0(x) + \sum_{i=1}^{k} \lambda_i f_i(x) + \sum_{j=1}^{l} \nu_j h_j(x) \right\}$$

Remarks

- Original problem is equivalent to

$$\underset{x}{\text{minimize}} \left[\sup_{\lambda \succeq 0, \nu} \mathcal{L}(x, \lambda, \nu) \right]$$

- Dual problem is *switching* the min and max:

$$\underset{\lambda \succeq 0, \nu}{\text{maximize}} \left[\inf_{x} \mathcal{L}(x, \lambda, \nu) \right].$$

One Great Property of Dual

Lemma (Weak Duality)

If $\lambda \succeq 0$*, then*

$$g(\lambda, \nu) \leq f_0(x^*).$$

Proof.

We have

$$\begin{aligned} g(\lambda, \nu) &= \inf_x \mathcal{L}(x, \lambda, \nu) \leq \mathcal{L}(x^*, \lambda, \nu) \\ &= f_0(x^*) + \sum_{i=1}^{k} \lambda_i f_i(x^*) + \sum_{j=1}^{l} \nu_j h_j(x^*) \leq f_0(x^*). \end{aligned}$$

$\square$

The Greatest Property of the Dual

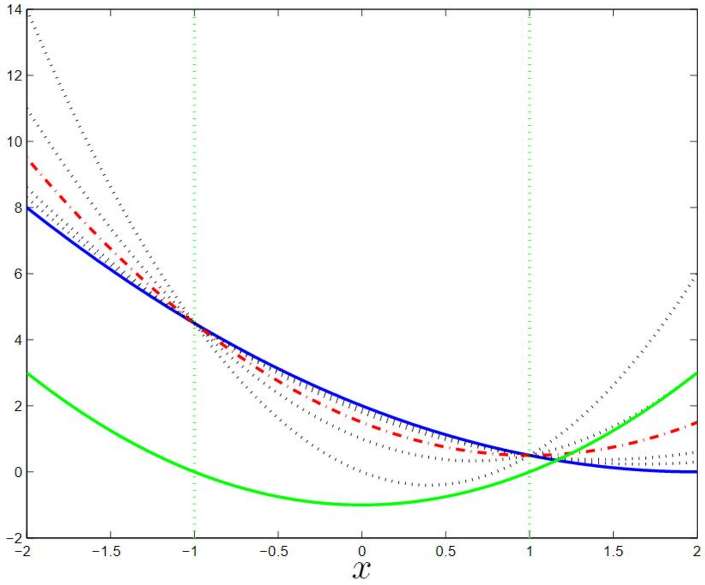
Theorem

For reasonable convex problems,

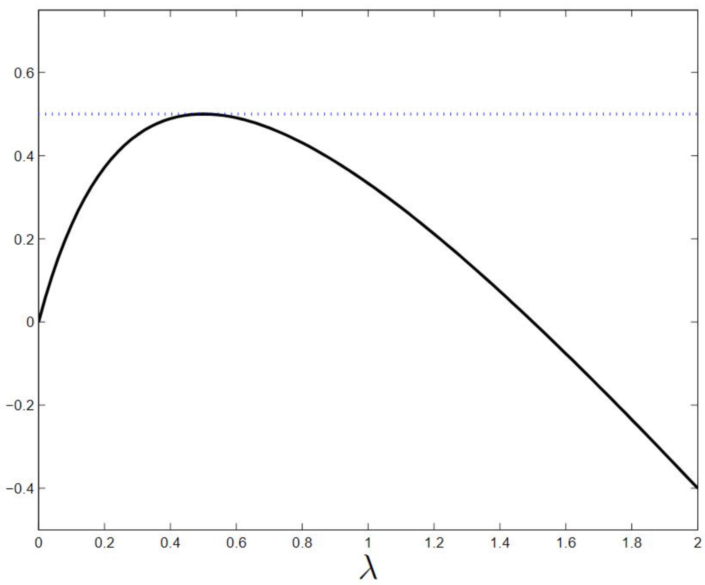
$$\sup_{\lambda \succeq 0, \nu} g(\lambda, \nu) = f_0(x^*)$$

Geometric Look

Minimize $\frac{1}{2}(x - c - 1)^2$ subject to $x^2 \leq c$.



True function (blue), constraint (green), $\mathcal{L}(x, \lambda)$ for different λ (dotted)



Dual function $g(\lambda)$ (black), primal optimal (dotted blue)

Intuition

Can interpret duality as linear approximation.

Intuition

Can interpret duality as linear approximation.

- $\mathbb{I}_{-}(a) = \infty$ if $a > 0$, 0 otherwise; $\mathbb{I}_0(a) = \infty$ unless $a = 0$. Rewrite problem as

$$\underset{x}{\text{minimize}} \;\; f_0(x) + \sum_{i=1}^{k} \mathbb{I}_{-}(f_i(x)) + \sum_{j=1}^{l} \mathbb{I}_0(h_j(x))$$

Intuition

Can interpret duality as linear approximation.

- $\mathbb{I}_-(a) = \infty$ if $a > 0$, 0 otherwise; $\mathbb{I}_0(a) = \infty$ unless $a = 0$. Rewrite problem as

$$\underset{x}{\text{minimize}} \;\; f_0(x) + \sum_{i=1}^{k} \mathbb{I}_-(f_i(x)) + \sum_{j=1}^{l} \mathbb{I}_0(h_j(x))$$

- Replace $\mathbb{I}(f_i(x))$ with $\lambda_i f_i(x)$; a measure of "displeasure" when $\lambda_i \geq 0$, $f_i(x) > 0$. $\nu_i h_j(x)$ lower bounds $\mathbb{I}_0(h_j(x))$:

$$\underset{x}{\text{minimize}} \;\; f_0(x) + \sum_{i=1}^{k} \lambda_i f_i(x) + \sum_{j=1}^{l} \nu_j h_j(x)$$

Example: Linearly constrained least squares

$$\underset{x}{\text{minimize}} \ \frac{1}{2} \|Ax - b\|^2 \quad \text{s.t.} \quad Bx = d.$$

Form the Lagrangian:

$$\mathcal{L}(x, \nu) = \frac{1}{2} \|Ax - b\|^2 + \nu^T (Bx - d)$$

Take infimum:

$$\nabla_x \mathcal{L}(x, \nu) = A^T Ax - A^T b + B^T \nu \quad \Rightarrow \quad x = (A^T A)^{-1}(A^T b - B^T \nu)$$

Simple unconstrained quadratic problem!

$$\begin{aligned} &\inf_x \mathcal{L}(x, \nu) \\ &= \frac{1}{2} \left\|A(A^T A)^{-1}(A^T b - B^T \nu) - b\right\|^2 + \nu^T B((A^T A)^{-1} A^T b - B^T \nu) - d^T \nu \end{aligned}$$

Example: Quadratically constrained least squares

$$\underset{x}{\text{minimize}} \;\; \frac{1}{2}\|Ax-b\|^2 \quad \text{s.t.} \;\; \frac{1}{2}\|x\|^2 \le c.$$

Form the Lagrangian ($\lambda \ge 0$):

$$\mathcal{L}(x,\lambda) = \frac{1}{2}\|Ax-b\|^2 + \frac{1}{2}\lambda(\|x\|^2 - 2c)$$

Take infimum:

$$\nabla_x \mathcal{L}(x,\nu) = A^T A x - A^T b + \lambda I \quad \Rightarrow \quad x = (A^T A + \lambda I)^{-1} A^T b$$

$$\inf_x \mathcal{L}(x,\lambda) = \frac{1}{2}\left\|A(A^T A + \lambda I)^{-1} A^T b - b\right\|^2 + \frac{\lambda}{2}\left\|(A^T A + \lambda I)^{-1} A^T b\right\|^2 - \lambda c$$

One variable dual problem!

$$g(\lambda) = -\frac{1}{2} b^T A (A^T A + \lambda I)^{-1} A^T b - \lambda c + \frac{1}{2}\|b\|^2 .$$