Topics in Applied Optimization

Optimization for ML and Data Sciences

Dr. Pawan Kumar

September 19, 2020

IIIT, Hyderabad

Lagrangian Dual Function

Optimization problem (Standard Form):

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

Lagrangian Dual Function

Optimization problem (Standard Form):

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

Idea: Augment the objective $f_0(x)$ with a weighted sum of the constraint functions.

Lagrangian Dual Function

Optimization problem (Standard Form):

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

Idea: Augment the objective $f_0(x)$ with a weighted sum of the constraint functions.

Lagrangian: Define Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ as

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

with dom $L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$. Here λ_i, ν_i are called Lagrange multipliers. Here λ and ν are called dual variables or Lagrange multiplier vectors.

Lagrange Dual Function

Lagrange Dual Function: Define the Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

Dual Fn

Lagrange fn

- If Lagrangian is unbounded below in x, the dual function takes on $-\infty$.

Lagrange Dual Function

Lagrange Dual Function: Define the Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

- If Lagrangian is unbounded below in x, the dual function takes on $-\infty$.
- The Lagrangian

Lagrange Dual Function

Q $f = \inf\{f_1, f_2, \ldots, f_k\}$ If f_i's are cvx, is f cvx? Ans: No

Proved before $f = \sup(f_1, f_2, \ldots, f_k)$ is cvx if f_i's cvx

Lagrange Dual Function: Define the Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ $= \inf(-f_1, -f_2, \ldots, -f_k)$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

f_2, f_1, f_1 f_2 cvx

$f = \inf(f_1, f_2, \ldots, f_k)$ if f_i's are concave, f is concave

- If Lagrangian is unbounded below in x, the dual function takes on $-\infty$.
- The Lagrangian

$f_0(x) + \bar{\lambda}^T \bar{f} + \bar{\nu}^T \bar{h}$

$Ax + b$

$f_0(x) + [\bar{\lambda}^T | \bar{\nu}^T] \begin{bmatrix} \bar{f} \\ \bar{h} \end{bmatrix}$

$$L(x, \lambda, \nu) = \cancel{\inf_{x \in \mathcal{D}}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

is affine as a function of λ and ν.

$\bar{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{bmatrix}$ $\bar{\nu} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_p \end{bmatrix}$ $\bar{f} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{bmatrix}$ $\bar{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{bmatrix}$

$f_0(x) + A^T d$

Hence Affine $d = [\bar{\lambda}^T | \bar{\nu}^T]^T$

Lagrange Dual Function

Lagrange Dual Function: Define the Lagrange dual function $g : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

$d^* < p^*$

- If Lagrangian is unbounded below in x, the dual function takes on $-\infty$.
- The Lagrangian

$g(\nu, \lambda)$ d^* p^*

dual primal

$$L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(\overset{C_1}{f_0(x)} + \sum_{i=1}^{m} \overset{C_2}{\lambda_i f_i(x)} + \sum_{i=1}^{p} \overset{C_3}{\nu_i h_i(x)} \right)$$

Concave & convex

$\boxed{d^* = p^*}$

is affine as a function of λ and ν.

- Since $g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu)$, i.e., it is infimum of affine functions, $g(\lambda, \nu)$ is concave, **even when original problem is not convex!**

Lower Bounds on Optimal Value

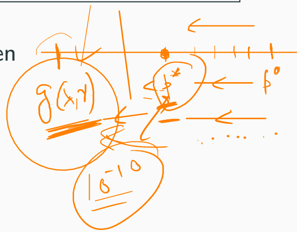
w.r.t primal variable i.e., x

Fact: The dual function yields lower bounds on the optimal value p^*. For any $\lambda \geq 0$ and any ν, we have

$$g(\lambda, \nu) \leq p^* . \quad f(x^*)$$

x^0

Proof: Suppose $\tilde{x}$ is a feasible point, i.e., $f_i(\tilde{x}) \leq 0$ and $h_i(\tilde{x}) = 0$, and $\lambda \geq 0$. then



Lower Bounds on Optimal Value

Fact: The dual function yields lower bounds on the optimal value p^*. For any $\lambda \geq 0$ and any ν, we have

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

$f_0(\tilde{x})$ $\qquad$ $\begin{pmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_m\end{pmatrix} \geq 0$

Proof: Suppose $\tilde{x}$ is a feasible point, i.e., $f_i(\tilde{x}) \leq 0$ and $h_i(\tilde{x}) = 0$, and $\lambda \geq 0$. then

$$\sum_{i=1}^{m} \underbrace{\underbrace{\lambda_i}_{\geq 0} \underbrace{f_i(\tilde{x})}_{\leq 0}}_{\leq 0} + \sum_{i=1}^{p} \underbrace{\nu_i h_i(\tilde{x})}_{=0} \leq 0$$

Lower Bounds on Optimal Value

Fact: The dual function yields lower bounds on the optimal value p^*. For any $\lambda \geq 0$ and any ν, we have

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

Proof: Suppose $\tilde{x}$ is a feasible point, i.e., $f_i(\tilde{x}) \leq 0$ and $h_i(\tilde{x}) = 0$, and $\lambda \geq 0$. then

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq 0$$

≤ 0

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

≤ 0

$$\implies g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \underset{\circledast}{\leq} f_0(\tilde{x}), \quad \text{for any feasible } \tilde{x}$$

$$\implies g(\lambda, \nu) \leq \inf_{x \in \mathcal{D}} f_0(x) = p^*$$

Examples: Least Squares Solution of Linear Equation

ν : equality

λ : inequality

Problem-1: Consider the following optimization problem

$$\begin{aligned} &\text{minimize} \quad x^T x \\ &\text{subject to} \quad Ax = b, \quad A \in \mathbb{R}^{p \times n} \end{aligned}$$

$$g(\nu) = \inf_x L(x, \nu)$$

- Find the Lagrangian function
- Find the dual function
- Check whether the dual function is concave
- Check whether the dual function is lower bound to p^*

[Page 218, Byod]

Answer: On chalkboard!

Scratch Space

Is this convex? Yes, by 2nd derivative test $\boxed{\nabla_x^2 L = I \succeq 0}$

$\min_x (x^T x + \nu^T A x)$

① Lagrangian is

$L(x,\nu) = \underbrace{x^T x + \nu^T (Ax - b)}$

$\mathrm{dom}(L) = \mathbb{R}^n \times \mathbb{R}^p$

② $g(\nu) = \inf_x \underbrace{L(x,\nu)} = \inf_x \left(x^T x + \nu^T (Ax - b)\right)$

$= -\nu^T b + \inf_x \left(\underbrace{x^T x + \nu^T A x}\right)$

- quadratic form
- diff.

$\boxed{\nabla_x L(x,\nu) = 0}$

$\Rightarrow 2x + A^T\nu = 0 \Rightarrow x = -\frac{1}{2}A^T\nu \rightarrow$ ①

$\Rightarrow \underbrace{g(\nu)} = L\left(-\frac{1}{2}A^T\nu, \nu\right) = \boxed{\left(-\frac{1}{2}A^T\nu\right)^T\left(-\frac{1}{2}A^T\nu\right) - \nu^T b} \leq p^*$

$L(x,\nu) = x^T x + \nu^T (Ax - b)$

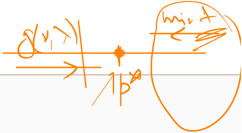
$= x^T x + \nu^T A x - \nu^T b$

① $\Rightarrow x^T x = -\frac{1}{2}x^T A^T \nu$

$\Rightarrow x^T A^T \nu = -2x^T x$

$\boxed{-x^T x - \nu^T b}$

Scratch Space



$g(\nu) = -\frac{1}{4}\nu^T \underbrace{AA^T}_{x^T P x}\nu - \underbrace{b^T\nu}_{b^T x}$ ← · quadratic fn

· diff.

by 2nd deriv. test

$\boxed{\nabla^2_\nu g(\nu) = -\frac{1}{4}AA^T \preceq 0}$ $\Rightarrow$ $g(\nu)$ is concave.

because $AA^T \succeq 0$ for any A

$x^T AA^T x = (Ax)^T(Ax) = \underbrace{\|Ax\|^2}_{\geq 0} \geq 0$

$\forall x$

$\Rightarrow AA^T$ is S.P.S.D.

From the lower bound prop.

$-\frac{1}{4}\nu^T AA^T\nu - b^T\nu \leq \inf\{x^T x \mid Ax = b\}$

Examples: Standard Form LP

Problem-2: Consider an LP in standard form

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

must put in $\leq$ form

f_i

$\boxed{-x} \leq 0 \iff -x_i \leq 0 \quad \forall\, i = 1, n$

- Find the Lagrangian
- Find the dual function
- Check whether the dual function gives lower bound to p^*

Scratch Space

① $\mathcal{L}(x,\lambda,\nu) = c^T x - \sum_{i=1}^{n} \lambda_i x_i + \nu^T (Ax - b)$

$= -\nu^T b + (c + A^T\nu - \lambda)^T x$

② Dual fn is: $g(\lambda,\nu) = \inf_x \mathcal{L}(x,\lambda,\nu)$

$= -\nu^T b + \inf_x \left(\underbrace{(c + A^T\nu - \lambda)^T x}_{} \right)$

Q Which type of fn is this?

A Linear in x.

$\inf_x \left((c + A^T\nu - \lambda)^T x \right) = \begin{cases} -\infty & \text{if } (c + A^T\nu - \lambda) \neq 0 \\ 0 & \text{if } c + A^T\nu - \lambda = 0 \end{cases}$

Scratch Space

affine subset

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

Observe: $g(\lambda, \nu)$ is finite only on a proper affine subset of $\mathbb{R}^m \times \mathbb{R}^p$.

Lower bound prop. $\boxed{g(\lambda, \nu) \le p^*}$

is non-trivial when: $\lambda \succeq 0$, $A^T\nu - \lambda + c = 0$

in which case: $\boxed{-\nu^T b \le p^*}$

Two way Partitioning Problem

observation

· Constraint restrict values of x_i to ± 1.

$[-1, 1, -1, -1, 1, \dots]W$

· Feasible set $S = \{x \in \mathbb{R}^n \mid x_i^2 = 1 \ \forall i = 1, \dots, n\}$

Problem-3:

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \dots, n, \end{array}$$

where $W \in S^n$.

- Find the Lagrangian
- Find the dual function
- Check whether dual is a lower bound for p^*

$|S| = 2^n \leftarrow$ finite feasible set

· Since $|S| < \infty \Rightarrow$ can test $x^T W x$ for all x.

· $|S|$ grows exponentially

~~Why 2-way Partition~~

· Like partition of set of n elements

Scratch Space

$$\{1, \dots, n\} = \{i \mid x_i = -1\} \cup \{i \mid x_i = 1\}$$

w_{ij} : interpreted as cost of i & j in same partition

$-w_{ij}$: " cost of i & j in different partition

<u>Problem</u> : Partition with <u>least cost</u>

$x_i \nu_i x_i$ $\qquad -\sum \nu_i = \mathbf{1}^T\nu$ $\qquad x^T P x$

Lagrange fn $\quad \mathcal{L}(x,\nu) = \underbrace{x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1)}_{\text{quadratic form ?}}$

$= \underbrace{x^T \left(W + \begin{bmatrix} \nu_1 & & 0 \\ & \nu_2 & \\ 0 & & \ddots \; \nu_n \end{bmatrix} \right)}\, x - \mathbf{1}^T \nu$, where $\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$, $\nu = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_n \end{bmatrix}$

Dual Fn $\quad g(\nu) = \inf_x \left(x^T \underbrace{\overbrace{\left(W + \operatorname{diag}(\nu) \right)}^{P}}_{\text{quadratic}} x - \mathbf{1}^T \nu \right)$

- $x^T P x \geq 0$
- P is S.P.D.

If P is not P.D.

If $\quad W + \operatorname{diag}(\nu) \succeq 0$

$\Rightarrow \quad g(\nu) = -\mathbf{1}^T \nu$

& If $W + \mathrm{diag}(\gamma) < 0$ then

$$g(\gamma) = -\infty$$

because if $W + \mathrm{diag}(\gamma) < 0 \Rightarrow \exists\, x$ s.t.

$$x^T (W + \mathrm{diag}(\gamma))\, x < 0$$

$\cancel{x = \beta y}$ $\boxed{y = \beta x}$

As $\beta \to \infty$

$\Rightarrow \quad y^T (W + \mathrm{diag}(\gamma))\, x \longrightarrow -\infty$

Scratch Space

$$\Rightarrow \quad g(\nu) = \begin{cases} -1^T\nu & \text{if } \quad W + \mathrm{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

<u>Lower bound</u>

$$\boxed{-1^T\nu \le p^*}$$