The University of Texas at Austin
Department of Electrical and Computer Engineering

EE381K: Convex Optimization — Fall 2019

LECTURE 18

Aryan Mokhtari Monday, November 4, 2019.

Goal: In this lecture, we talk about the Lagrangian, Lagrange multipliers, (Lagrangian) dual function, and dual problem.

1 The Lagrangian

Consider the following optimization problem in the standard form: (not necessarily convex)

$$\begin{aligned} \min: \quad & f_0(\mathbf{x}) \\ \text{s.t.}: \quad & f_i(\mathbf{x}) \leq 0, \quad i = 1, \dots m \\ & h_i(\mathbf{x}) = 0, \quad i = 1, \dots p \end{aligned} \tag{1}$$

with $\mathbf{x} \in \mathbb{R}^n$, domain $\mathcal{D}$, and optimal value p^*.

The Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$ corresponding to Problem (1) is defined as

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

where $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_m] \in \mathbb{R}^m$ and $\boldsymbol{\nu} = [\nu_1, \dots, \nu_p] \in \mathbb{R}^p$.

1. each λ_i corresponds to an inequality constraint
2. each ν_i corresponds to an equality constraint
3. λ and ν are called dual variables or Lagrange multiplier vectors
4. L can be interpreted as weighted sum of objective and constraint functions

2 The (Lagrange) Dual Function

We define the (Lagrange) dual function of Problem (1) is defined as the minimum value of the Lagrangian over the primal variable $\mathbf{x}$, i.e.,

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

Note: The dual function $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ can be interpreted as the pointwise infimum of a family of affine functions of (λ, ν), and therefore, it is a concave function. (Even if Problem (1) is not convex!)

Note: For some choices of (λ, ν) the dual function value could be $-\infty$.

2.1 Lower bound property

The dual function yields lower bounds on the optimal value p^* of the primal problem (when $\lambda \geq \mathbf{0}$). For any $\boldsymbol{\lambda} \geq \mathbf{0}$ and any $\boldsymbol{\nu}$ we have $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \leq p^*$. Why? If $\tilde{\mathbf{x}}$ is feasible and $\boldsymbol{\lambda} \geq \mathbf{0}$ then

$$f_0(\tilde{\mathbf{x}}) \geq L(\tilde{\mathbf{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \geq \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = g(\boldsymbol{\lambda}, \boldsymbol{\nu})$$

by minimizing over all feasible $\tilde{\mathbf{x}}$ we obtain that $p^* \geq g(\boldsymbol{\lambda}, \boldsymbol{\nu})$.

3 Dual problem

The Lagrange dual problem of Problem (1) is defined as

$$\begin{aligned} \max: \quad & g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \\ \text{s.t.}: \quad & \boldsymbol{\lambda} \geq \mathbf{0} \end{aligned}$$

1. Finds best lower bound on p^*, obtained from Lagrange dual function
2. Any feasible point of this problem gives a lower bound for p^*
3. A convex optimization problem; optimal value denoted d^*

4 Examples of dual function and dual problem

Standard form LP: The standard form of LP is given by

$$\begin{aligned} \min: \quad & \mathbf{c}^\top \mathbf{x} \\ \text{s.t.}: \quad & \mathbf{A}\mathbf{x} = \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0} \end{aligned}$$

The Lagrangian in this case is defined as

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) &= \mathbf{c}^\top \mathbf{x} - \boldsymbol{\lambda}^\top \mathbf{x} + \boldsymbol{\nu}^\top (\mathbf{A}\mathbf{x} - \mathbf{b}) \\ &= (\mathbf{c} - \boldsymbol{\lambda} + \mathbf{A}^\top \boldsymbol{\nu})^\top \mathbf{x} - \boldsymbol{\nu}^\top \mathbf{b} \end{aligned}$$

Hence, the dual function $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \begin{cases} -\mathbf{b}^\top \boldsymbol{\nu} & \text{if } \mathbf{c} - \boldsymbol{\lambda} + \mathbf{A}^\top \boldsymbol{\nu} = \mathbf{0} \\ -\infty & \text{otherwise} \end{cases}$$

And therefore the dual problem is

$$\begin{aligned} \max: \quad & -\mathbf{b}^\top \boldsymbol{\nu} \\ \text{s.t.}: \quad & \mathbf{c} - \boldsymbol{\lambda} + \mathbf{A}^\top \boldsymbol{\nu} = \mathbf{0} \\ & \boldsymbol{\lambda} \geq \mathbf{0}, \end{aligned}$$

which can be simplified as

$$\begin{aligned} \max: \quad & -\mathbf{b}^\top \boldsymbol{\nu} \\ \text{s.t.}: \quad & \mathbf{c} + \mathbf{A}^\top \boldsymbol{\nu} \geq \mathbf{0} \end{aligned}$$

This is what we have seen in the first part of the class!

Least-norm solution of linear equations: Consider the following problem

$$\begin{aligned} \min: &\quad \mathbf{x}^\top \mathbf{x} \\ \text{s.t.}: &\quad \mathbf{A}\mathbf{x} = \mathbf{b}. \end{aligned}$$

The Lagrangian in this case is defined as

$$\begin{aligned} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = L(\mathbf{x}, \boldsymbol{\nu}) &= \mathbf{x}^\top \mathbf{x} + \boldsymbol{\nu}^\top(\mathbf{A}\mathbf{x} - \mathbf{b}) \\ &= \mathbf{x}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{A}^\top \boldsymbol{\nu} - \boldsymbol{\nu}^\top \mathbf{b} \end{aligned}$$

which is a quadratic function with respect to $\mathbf{x}$. Hence, by computing the gradient and setting it to zero we can find the optimal solution which is

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = \mathbf{0} \quad \Leftrightarrow \quad 2\mathbf{x} + \mathbf{A}^\top \boldsymbol{\nu} = \mathbf{0} \quad \Leftrightarrow \quad \mathbf{x}^* = -\frac{1}{2}\mathbf{A}^\top \boldsymbol{\nu}$$

and therefore we can write the dual function as

$$\begin{aligned} g(\boldsymbol{\nu}) = \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\nu}) = L(\mathbf{x}^*, \boldsymbol{\nu}) &= \frac{1}{4}\boldsymbol{\nu}^\top \mathbf{A}\mathbf{A}^\top \boldsymbol{\nu} - \frac{1}{2}\boldsymbol{\nu}^\top \mathbf{A}\mathbf{A}^\top \boldsymbol{\nu} - \boldsymbol{\nu}^\top \mathbf{b} \\ &= -\frac{1}{4}\boldsymbol{\nu}^\top \mathbf{A}\mathbf{A}^\top \boldsymbol{\nu} - \mathbf{b}^\top \boldsymbol{\nu} \end{aligned}$$

In this case, the dual problem is unconstrained and defined as

$$\max: -\frac{1}{4}\boldsymbol{\nu}^\top \mathbf{A}\mathbf{A}^\top \boldsymbol{\nu} - \mathbf{b}^\top \boldsymbol{\nu}$$

5 Lagrange dual function and conjugate function

Recall that the definition of the conjugate function f^* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is given by

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

There is a connection between the conjugate function and Lagrangian dual function. To see this connection consider the following optimization problems:

5.1 Linear constraints

Consider the following optimization problem with linear constraints:

$$\begin{aligned} \min: &\quad f_0(\mathbf{x}) \\ \text{s.t.}: &\quad \mathbf{A}\mathbf{x} \leq \mathbf{b}, \qquad \mathbf{C}\mathbf{x} = \mathbf{d}. \end{aligned}$$

By using the definition of the conjugate function we can write the dual function as

$$\begin{aligned} g(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= \inf_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \\ &= \inf_{\mathbf{x}} \Big(f_0(\mathbf{x}) + \boldsymbol{\lambda}^\top(\mathbf{A}\mathbf{x} - \mathbf{b}) + \boldsymbol{\nu}^\top(\mathbf{C}\mathbf{x} - \mathbf{d}) \Big) \\ &= -\boldsymbol{\lambda}^\top \mathbf{b} - \boldsymbol{\nu}^\top \mathbf{d} + \inf_{\mathbf{x}} \Big(f_0(\mathbf{x}) + (\mathbf{A}^\top \boldsymbol{\lambda} + \mathbf{C}^\top \boldsymbol{\nu})^\top \mathbf{x} \Big) \\ &= -\boldsymbol{\lambda}^\top \mathbf{b} - \boldsymbol{\nu}^\top \mathbf{d} - \sup_{\mathbf{x}} \Big(-f_0(\mathbf{x}) + (-\mathbf{A}^\top \boldsymbol{\lambda} - \mathbf{C}^\top \boldsymbol{\nu})^\top \mathbf{x} \Big) \\ &= -\boldsymbol{\lambda}^\top \mathbf{b} - \boldsymbol{\nu}^\top \mathbf{d} - f_0^*(-\mathbf{A}^\top \boldsymbol{\lambda} - \mathbf{C}^\top \boldsymbol{\nu}) \end{aligned}$$

and the domain of $g(\boldsymbol{\lambda}, \boldsymbol{\nu})$ is defined as $\mathbf{dom} g = \{(\boldsymbol{\lambda}, \boldsymbol{\nu}) | -\mathbf{A}^\top \boldsymbol{\lambda} - \mathbf{C}^\top \boldsymbol{\nu} \in \mathbf{dom} f_0^*\}$.

5.2 Minimum volume covering ellipsoid

Consider an optimization problem with variable $\mathbf{X} \in \mathbf{S}_{++}^n$. Further, recall that for any given positive definite matrix $\mathbf{X}$ the expression $(\mathbf{z}-\mathbf{a})^\top \mathbf{X}(\mathbf{z}-\mathbf{a}) \leq 1$, the set of points $\mathbf{z}$ that satisfy this condition create an ellipsoid where the center of that ellipsoid is $\mathbf{a}$. Therefore, if we focus on an ellipsoid that is centered at the origin its expression can be written as $\mathbf{z}^\top \mathbf{X}\mathbf{z} \leq 1$. Considering these observations, if we have an optimization problem with the following constraints

$$\mathbf{a}_i^\top \mathbf{X}\mathbf{a}_i \leq 1, \quad i = 1, \ldots, m,$$

it means that we want to find a matrix $\mathbf{X}$ such that its corresponding ellipsoid centered at the origin contains (covers) points $\mathbf{a}_1, \ldots, \mathbf{a}_m$.
Now consider the problem that we want to find the ellipsoid centered at the origin that contains points $\mathbf{a}_1, \ldots, \mathbf{a}_m$ and has the minimum possible volume. As the volume of the ellipsoid $(\mathbf{z}-\mathbf{a})^\top \mathbf{X}(\mathbf{z}-\mathbf{a}) \leq 1$ is proportional to $(\det(\mathbf{X}^{-1}))^{1/2}$, our problem of interest can be written as

$$\begin{aligned} \min: \quad & f_0(\mathbf{X}) := \log(\det(\mathbf{X}^{-1}))) \\ \text{s.t.}: \quad & \mathbf{a}_i^\top \mathbf{X}\mathbf{a}_i \leq 1, \quad i = 1, \ldots, m. \end{aligned}$$

The inequality constraints in this problem can be written as

$$\operatorname{tr}(\mathbf{X}\mathbf{a}_i\mathbf{a}_i^\top) \leq 1, \quad i = 1, \ldots, m.$$

Further, the conjugate of the objective function f_0 is

$$f_0^*(\mathbf{Y}) = -n + \log(\det(-\mathbf{Y}^{-1}))$$

Since the constrains are affine, we can use the previous result and show that the dual function is

$$\begin{aligned} g(\boldsymbol{\lambda}) &= \inf_{\mathbf{X}} L(\mathbf{x}, \boldsymbol{\lambda}) \\ &= \inf_{\mathbf{X}} \left(\log(\det(\mathbf{X}^{-1}))) + \sum_{i=1}^m \lambda_i (\operatorname{tr}(\mathbf{X}\mathbf{a}_i\mathbf{a}_i^\top) - 1) \right) \\ &= -\sum_{i=1}^m \lambda_i + \inf_{\mathbf{X}} \left(\log(\det(\mathbf{X}^{-1}))) + \sum_{i=1}^m \operatorname{tr}(\mathbf{X}\lambda_i\mathbf{a}_i\mathbf{a}_i^\top) \right) \\ &= -\boldsymbol{\lambda}^\top \mathbf{1} + \inf_{\mathbf{X}} \left(\log(\det(\mathbf{X}^{-1}))) + \operatorname{tr}\left(\mathbf{X} \left(\sum_{i=1}^m \lambda_i \mathbf{a}_i\mathbf{a}_i^\top \right) \right) \right) \\ &= -\boldsymbol{\lambda}^\top \mathbf{1} - \sup_{\mathbf{X}} \left(-\log(\det(\mathbf{X}^{-1}))) + \operatorname{tr}\left(\mathbf{X} \left(-\sum_{i=1}^m \lambda_i \mathbf{a}_i\mathbf{a}_i^\top \right) \right) \right) \\ &= -\boldsymbol{\lambda}^\top \mathbf{1} - f_0^*\left(-\sum_{i=1}^m \lambda_i \mathbf{a}_i\mathbf{a}_i^\top \right) \\ &= -\boldsymbol{\lambda}^\top \mathbf{1} + n - \log\left(\det\left(\left(\sum_{i=1}^m \lambda_i \mathbf{a}_i\mathbf{a}_i^\top \right)^{-1} \right) \right) \\ &= -\boldsymbol{\lambda}^\top \mathbf{1} + n + \log\left(\det\left(\sum_{i=1}^m \lambda_i \mathbf{a}_i\mathbf{a}_i^\top \right) \right) \end{aligned}$$

where the last two equalities are correct if $\sum_{i=1}^m \lambda_i \mathbf{a}_i \mathbf{a}_i^\top \succ \mathbf{0}$. Hence, the dual function is

$$g(\boldsymbol{\lambda}) = \begin{cases} -\boldsymbol{\lambda}^\top \mathbf{1} + n + \log\left(\det\left(\sum_{i=1}^m \lambda_i \mathbf{a}_i \mathbf{a}_i^\top\right)\right) & \text{if } \sum_{i=1}^m \lambda_i \mathbf{a}_i \mathbf{a}_i^\top \succ \mathbf{0} \\ -\infty & \text{otherwise} \end{cases}$$

Therefore, the dual problem is

$$\begin{aligned} \max: \quad & g(\boldsymbol{\lambda}) = n - \boldsymbol{\lambda}^\top \mathbf{1} + \log\left(\det\left(\sum_{i=1}^m \lambda_i \mathbf{a}_i \mathbf{a}_i^\top\right)\right) \\ \text{s.t.}: \quad & \boldsymbol{\lambda} \geq \mathbf{0}. \end{aligned}$$

and its domain is $\mathbf{dom} g = \{\sum_{i=1}^m \lambda_i \mathbf{a}_i \mathbf{a}_i^\top \succ \mathbf{0}\}$. We can also think of $\sum_{i=1}^m \lambda_i \mathbf{a}_i \mathbf{a}_i^\top \succ \mathbf{0}$ as an implicit constraint.