Lagrangian Duality Theory

Optimization Methods in Management Science
Master in Management
HEC Lausanne

Dr. Rodrigue Oeuvray

Fall 2019 Semester

Lagrangian Duality Theory of Constrained Optimization

- **Duality** is pervasive in **linear** and **nonlinear** optimization models in a wide variety of engineering and mathematical settings
- In a number of situations, the dual problem have better characteristics than the primal problem
- We have already encountered the concept of duality in linear programming
- We will see this concept in a more general framework not limited to linear problems

Lagrangian Duality Theory

- Let's assume that the optimization problem is given by:

$$\begin{aligned} min_{\mathbf{x} \in X} \quad & f(\mathbf{x}) \\ s.t. \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, k \end{aligned}$$

- Until now, we have made the assumption that $X = \mathbb{R}^n$ but we consider here the general case where X can be different from $\mathbb{R}^n$
- It provides us the flexibility necessary to construct the dual problem

Lagrangian Duality Theory (Cont'd)

The lagrangian of the previous problem is defined by

$$L(\boldsymbol{x}, \boldsymbol{\alpha}) = f(\boldsymbol{x}) + \sum_{i=1}^{k} \alpha_i g_i(\boldsymbol{x}),$$

where $\alpha_i \geq 0, i = 1, \ldots, k$

Steps in the Construction of the Dual Problem

Construction of the Dual Problem

- Define the lagrangian:

$$L(\boldsymbol{x}, \boldsymbol{\alpha}) = f(\boldsymbol{x}) \ + \sum_{i=1}^{k} \alpha_i g_i(\boldsymbol{x}),$$

- Compute the dual function:

$$L^*(\boldsymbol{\alpha}) = min_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\alpha})$$

- Solve the dual problem

$$\begin{aligned} max_{\boldsymbol{\alpha} \in \mathbb{R}^k} \quad & L^*(\boldsymbol{\alpha}) \\ s.t. \quad & \alpha_i \geq 0, \quad i = 1, \ldots, k \end{aligned}$$

Example 1

We consider the following linear problem:

$$\begin{aligned} min_{\boldsymbol{x} \in \mathbb{R}^n} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ s.t. \quad & \boldsymbol{Ax} \geq \boldsymbol{b} \\ & x_i \geq 0 \quad i = 1, \dots, n \end{aligned}$$

where $\boldsymbol{A}$ is a $k \times n$ matrix with $k < n$. We can rearrange the inequality constraints and re-write the problem as:

$$\begin{aligned} min_{\boldsymbol{x} \in \mathbb{R}^n} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ s.t. \quad & \boldsymbol{b} - \boldsymbol{Ax} \leq 0 \\ & x_i \geq 0 \quad i = 1, \dots, n \end{aligned}$$

Example 1 (Cont'd)

The lagrangian is defined by:

$$\begin{aligned} L(\boldsymbol{x}, \alpha) &= \boldsymbol{c}^T \boldsymbol{x} + \alpha^T (\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \\ &= \alpha^T \boldsymbol{b} + (\boldsymbol{c} - \boldsymbol{A}^T \alpha)^T \boldsymbol{x} \end{aligned}$$

To determine the dual function, we need to solve :

$$L^*(\alpha) = min_{x_i \geq 0} L(\boldsymbol{x}, \alpha)$$
$$L^*(\alpha) = min_{x_i \geq 0} \alpha^T \boldsymbol{b} + (\boldsymbol{c} - \boldsymbol{A}^T \alpha)^T \boldsymbol{x}$$

Example 1 (Cont'd)

Indeed, the dual function is given by:

$$L^*(\boldsymbol{\alpha}) = min_{x_i \geq 0} \boldsymbol{\alpha}^T \boldsymbol{b} + (\boldsymbol{c} - \boldsymbol{A}^T \boldsymbol{\alpha})^T \boldsymbol{x}$$
$$= \begin{cases} \boldsymbol{\alpha}^T \boldsymbol{b} \quad \text{if } \boldsymbol{A}^T \boldsymbol{\alpha} \leq \boldsymbol{c} \quad (1) \\ -\infty \quad \text{if } \exists \text{ a column } \boldsymbol{A}_i \text{ such that } \boldsymbol{A}_i^T \boldsymbol{\alpha} > c_i \quad (2) \end{cases}$$

Explanations: when we minimize a linear function of type $a + \boldsymbol{\beta}^T \boldsymbol{x}$ under the constraints that $x_i \geq 0, i = 1, \ldots, n$, then :

- **Case i**: the optimal value is a as soon as $\beta_i \geq 0, \quad \forall i$
- **Case ii**: the optimal value is $-\infty$ as soon as one β_i is < 0
- (1) corresponds to **Case i** and (2) to **Case ii**

Example 1 (Cont'd)

The dual problem is therefore constructed as:

$$\begin{aligned} max_{\boldsymbol{\alpha}} \quad & \boldsymbol{\alpha}^T \boldsymbol{b} \\ s.t. \quad & \boldsymbol{A}^T \boldsymbol{\alpha} \leq \boldsymbol{c} \\ & \alpha_i \geq 0 \quad i = 1, \ldots, k \end{aligned}$$

Example 2

We consider the following quadratic optimization problem:

$$\begin{aligned} min_{\boldsymbol{x} \in \mathbb{R}^n} \quad & \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} \\ s.t. \quad & \boldsymbol{A}\boldsymbol{x} \geq \boldsymbol{b} \end{aligned}$$

where $\boldsymbol{Q}$ is symmetric and positive definte and $\boldsymbol{A}$ is a $k \times n$ matrix with $k < n$. To construct the dual problem, let's rewrite the inequality as $\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x} \leq \mathbf{0}$. The lagrangian is given by:

$$\begin{aligned} L(\boldsymbol{x}, \alpha) &= \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x} + \alpha^T(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \\ &= \alpha^T\boldsymbol{b} + (\boldsymbol{c} - \boldsymbol{A}^T\alpha)^T\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} \end{aligned}$$

Example 2 (Cont'd)

- The dual is defined by $L^*(\alpha) = min_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \alpha)$
- Since this expression is a **convex** quadratic function (as a function of $\boldsymbol{x}$), it follows that the minimizer $\bar{\boldsymbol{x}}$ of this function is given by

$$\mathbf{0} = \nabla_{\boldsymbol{x}} L(\bar{\boldsymbol{x}}, \alpha) = (\boldsymbol{c} - \boldsymbol{A}^T \alpha) + \boldsymbol{Q} \bar{\boldsymbol{x}}$$

- The solution of the previous system is

$$\bar{\boldsymbol{x}} = -\boldsymbol{Q}^{-1} (\boldsymbol{c} - \boldsymbol{A}^T \alpha)$$

Example 2 (Cont'd)

Substituting this value of $\boldsymbol{x}$ into the lagrangian, we obtain:

$$L^*(\boldsymbol{\alpha}) = L(\bar{\boldsymbol{x}}, \boldsymbol{\alpha}) = \boldsymbol{\alpha}^T \boldsymbol{b} - \frac{1}{2}(\boldsymbol{c} - \boldsymbol{A}^T \alpha)^T \boldsymbol{Q}^{-1}(\boldsymbol{c} - \boldsymbol{A}^T \alpha)$$

Therefore, the dual problem is defined to be

$$\begin{aligned} & max_{\boldsymbol{\alpha}} \ \boldsymbol{\alpha}^T \boldsymbol{b} - \frac{1}{2}(\boldsymbol{c} - \boldsymbol{A}^T \alpha)^T \boldsymbol{Q}^{-1}(\boldsymbol{c} - \boldsymbol{A}^T \alpha) \\ & \text{s.t. } \alpha_i \geq 0, \quad i = 1, \ldots, k \end{aligned}$$

Duality: General Case

Let's assume that the optimization problem is given by:

$$
\begin{aligned}
min_{\mathbf{x} \in X} \quad & f(\mathbf{x}) \\
s.t. \quad & g_i(\mathbf{x}) \leq 0, \quad i \in L \\
& g_i(\mathbf{x}) \geq 0, \quad i \in G \\
& g_i(\mathbf{x}) = 0, \quad i \in E \\
& \mathbf{x} \in X
\end{aligned}
$$

Duality: General Case (Cont'd)

Then the lagrangian takes the form:

$$L(\boldsymbol{x}, \boldsymbol{\alpha}) = f(\boldsymbol{x}) + \sum_{i \in L} \alpha_i g_i(\boldsymbol{x}) + \sum_{i \in G} \alpha_i g_i(\boldsymbol{x}) + \sum_{i \in E} \alpha_i g_i(\boldsymbol{x})$$

and the dual function $L^*(\boldsymbol{\alpha})$ is constructed as

$$L^*(\boldsymbol{\alpha}) = min_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\alpha})$$

Duality: General Case (Cont'd)

The dual problem is then defined to be:

$$
\begin{aligned}
max_{\boldsymbol{\alpha}} \quad & L^*(\boldsymbol{\alpha}) \\
s.t. \quad & \alpha_i \geq 0, \quad i \in L \\
& \alpha_i \leq 0, \quad i \in G \\
& \alpha_i \in \mathbb{R}, \quad i \in E
\end{aligned}
$$

Duality Revisited

We now consider the following case:

$$\begin{aligned} min_{\boldsymbol{x} \in \mathbb{R}^n} \quad & f(\boldsymbol{x}) \\ s.t. \quad & g_i(\boldsymbol{x}) \leq 0, \quad i = 1, \dots, k \\ & h_i(\boldsymbol{x}) = 0, \quad i = 1, \dots, l \end{aligned}$$

We assume that the functions f, g_i, h_i are continuously differentiable. This is the primal problem $\mathcal{P}$. Its lagrangian is defined by

$$L(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = f(\boldsymbol{x}) + \sum_{i=1}^{k} \alpha_i g_i(\boldsymbol{x}) + \sum_{i=1}^{l} \beta_i h_i(\boldsymbol{x})$$

Duality Revisited (Cont'd)

Consider the quantity

$$\theta_{\mathcal{P}} = \max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\beta}} L(\boldsymbol{x}, \boldsymbol{\alpha}, \boldsymbol{\beta})$$

- If $g_i(\boldsymbol{x}) < 0$, then $max_{\alpha_i} \alpha_i g_i(\boldsymbol{x}) = 0 \quad \forall \alpha_i \geq 0$
- If $g_i(\boldsymbol{x}) = 0$, then $\alpha_i g_i(\boldsymbol{x}) = 0 \quad \forall \alpha_i \geq 0$
- If $g_i(\boldsymbol{x}) > 0$, then $max_{\alpha_i \geq 0} \alpha g_i(\boldsymbol{x}) = +\infty$
- If $h_i(\boldsymbol{x}) = 0$, then $\beta_i h_i(\boldsymbol{x}) = 0$
- If $h_i(\boldsymbol{x}) \neq 0$, then $max_{\beta_i} \beta_i h_i(\boldsymbol{x}) = +\infty$

This simply shows that :

$$\theta_{\mathcal{P}}(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}) & \text{if } \boldsymbol{x} \text{ satisfies primal constraints} \\ +\infty & \text{otherwise} \end{cases}$$

Duality Revisited (Cont'd)

- So, if we consider

$$min_{\mathbf{x}} \theta_{\mathcal{P}}(\mathbf{x}) = min_{\mathbf{x}} max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\beta}} L(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta}),$$

  then this is the same problem as the original one but formulated without any constraint

- However, this formulation is not very helpful... The objective function is not even continuous !

- A more interesting approach is based on the **dual** formulation !

Duality Revisited (Cont'd)

We define $\theta_{\mathcal{D}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ as :

$$\theta_{\mathcal{D}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\alpha}, \boldsymbol{\beta})$$

$\mathcal{D}$ stands for "dual". The dual optimization problem is defined by:

$$max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\beta}} \theta_{\mathcal{D}}(\boldsymbol{\alpha}, \boldsymbol{\beta})$$

It can be easily shown that

$$d^* = max_{\boldsymbol{\alpha} \geq 0, \boldsymbol{\beta}} \theta_{\mathcal{D}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \leq min_{\mathbf{x}} \theta_{\mathcal{P}}(\mathbf{x}) = p^*$$

However, under certains conditions, we can have $d^* = p^*$ so that we can **solve the dual problem in place of the original problem**

Duality Revisited (Cont'd)

Theorem

Suppose that f and the g_i's are convex, and that h_i are affine. Suppose further that there exists some feasible $\mathbf{x}$ so that $g_i(\mathbf{x}) < 0$, $\forall i$. Then there exists $\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$ so that $\mathbf{x}^*$ is the solution to the primal problem $\mathcal{P}$,$\boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$ are the solutions of the dual problem, and $p^* = d^* = L(\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*)$. Moreover, $\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$ satisfy the ***Karush-Kuhn-Tucker (KKT)*** *conditions given by:*

$$
\begin{aligned}
\frac{\partial}{\partial x_i} L(\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) &= 0, \quad i = 1, \ldots, n \\
\frac{\partial}{\partial \beta_i} L(\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*) &= 0, \quad i = 1, \ldots, l \,(\iff h_i(\mathbf{x}) = 0, \quad i = 1, \ldots, l) \\
\alpha_i^* g_i(\mathbf{x}^*) &= 0, \quad i = 1, \ldots, k \text{ (dual complementary cond.)} \\
g_i(\mathbf{x}^*) &\leq 0, \quad i = 1, \ldots, k \\
\alpha_i^* &\geq 0, \quad i = 1 \ldots, k
\end{aligned}
$$

Moreover, if some $\mathbf{x}^*, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^*$ satisfy the KKT conditions, then they are also a solution to the primal and dual problems

Slater's Condition

- In the previous theorem, the existence of some **feasible $\boldsymbol{x}$** so that $g_i(\boldsymbol{x}) < 0, \ \forall i$ is called **Slater's condition**

- It is a condition for **strong duality** to hold for a convex optimization problem. Informally, Slater's condition states that the feasible region must have an interior point

- It is a specific example of a **constraint qualification**. In particular, if Slater's condition holds for the primal problem and has a finite solution, then the **duality gap** is 0

Lagrange Duality With Equality: Geometrical Interpretation

We consider the simplified problem with only one equality:

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h(\boldsymbol{x}) = 0 \end{aligned}$$

Lagrangian function:

$$L(\boldsymbol{x}, \beta) = f(\boldsymbol{x}) + \beta h(\boldsymbol{x})$$

First-order optimality conditions:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \beta^*) = \nabla f(\boldsymbol{x}^*) + \beta^* \nabla h(\boldsymbol{x}^*) = 0$$

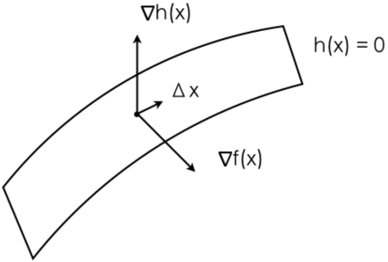
or, equivalently:

$$\nabla f(\boldsymbol{x}^*) = -\beta^* \nabla h(\boldsymbol{x}^*)$$

Concretly, it means that the gradient of f at $\boldsymbol{x}^*$ is a multiple of the gradient of h at the same point. Why is it the case ?

Lagrange Duality With Equality (Cont'd)

- Assume our udpate to $\boldsymbol{x}$ is $\Delta\boldsymbol{x}$, meaning we move from the curent position $\boldsymbol{x}$ to $\boldsymbol{x} + \Delta\boldsymbol{x}$, where is $\Delta\boldsymbol{x}$ is assumed to be very small (in norm)
- $h(\boldsymbol{x}) = 0$ defines a "surface" and its gradient at $\boldsymbol{x}$ is othogonal to it
- In order to stay on this surface, $\Delta\boldsymbol{x}$ must be orthogonal to $\nabla h(\boldsymbol{x})$, thus we get $\Delta\boldsymbol{x}^T \nabla h(\boldsymbol{x}) = 0$
- In order to decrease $f(\boldsymbol{x})$, then $\Delta\boldsymbol{x}^T \nabla f(\boldsymbol{x}) < 0$



Lagrange Duality With Equality (Cont'd)

- Keep doing this, eventually, at the optimality point, we naturally have two gradient vectors being parallel (or anti-parallel, meaning pointing at different directions) becoming the end result when an optimal solution is found
- So, at the optimum $\boldsymbol{x}^*$, $\Delta\boldsymbol{x}^T\nabla f(\boldsymbol{x}^*) = 0$ for **any** $\Delta\boldsymbol{x}$ such that $\Delta\boldsymbol{x}^T\nabla h(\boldsymbol{x}^*) = 0$. So $\nabla f(\boldsymbol{x}^*)$ is perpendicular to all the directions that are perpendicular to $\nabla h(\boldsymbol{x}^*)$
- **We conclude that the gradient of f at the optimum must be a multiple of the gradient of h at the same point**

Lagrange Duality With Inequality

We consider the simplified problem with only one inequality:

$$\begin{aligned} \min \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g(\boldsymbol{x}) \leq 0 \end{aligned}$$

Lagrangian function:

$$L(\boldsymbol{x}, \alpha) = f(\boldsymbol{x}) + \alpha g(\boldsymbol{x}), \quad \alpha \geq 0$$

First-order optimality conditions:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \alpha^*) = \nabla f(\boldsymbol{x}^*) + \alpha^* \nabla g(\boldsymbol{x}^*) = 0, \quad \text{with some } \alpha^* \geq 0$$

or, equivalently:

$$\nabla f(\boldsymbol{x}^*) = -\alpha^* \nabla g(\boldsymbol{x}^*), \quad \alpha^* \geq 0 \;\; \text{(gradients are antiparallel)}$$

Lagrange Duality With Inequality (Cont'd)

First case: the optimum $\boldsymbol{x}^*$ is on the surface of $g(\boldsymbol{x}^*) = 0$

- We can replace the constraint $g(\boldsymbol{x}) \leq 0$ by $g(\boldsymbol{x}) = 0$. We know that the gradient of f should be a multiple of the gradient of g at the optimum
- But why α should be ≥ 0?
- If it is not the case ($\alpha^* < 0$), then $\nabla f(\boldsymbol{x}^*)$ and $\nabla g(\boldsymbol{x}^*)$ are parallel. If we go to the direction $-\nabla f(\boldsymbol{x}^*)$, we will decrease the primal objective value as well as the value of $g(\boldsymbol{x}^*)$ which will become < 0. This is a contradiction with our assumption that $\boldsymbol{x}^*$ lies on the surface of $g(\boldsymbol{x}^*) = 0$

Lagrange Duality With Inequality (Cont'd)

Second case: the optimal solution is at position $\boldsymbol{x}^*$ where the constraint is not active

- In that case, this constraint should be effectively removed from the lagrangian equation, so that we are only optimizing under the primal optimal condition: $\nabla f(\boldsymbol{x}^*) = 0$, thus $\alpha^* = 0$ since $\nabla f(\boldsymbol{x}^*) = -\alpha^* \nabla g(\boldsymbol{x}^*)$

To conclude

- if $g(\boldsymbol{x}^*) = 0$, then $\alpha^* \geq 0$ as in the first case, and if $g(\boldsymbol{x}^*) < 0$, then $\alpha^* = 0$ as in the second case. In both cases, $\alpha^* g(\boldsymbol{x}^*) = 0$. This condition is called the **complementary slackness**