Indian Institute of Technology Kanpur
CS774 Optimization Techniques

Scribe: Pratik Mishra
Instructor: Purushottam Kar
Date: August 18, 2016

LECTURE
4

Duality

1 Introduction

In Mathematical Optimization theory, duality or the duality principle is the principle that optimization problems may be viewed from either of two perspectives, the primal problem or the dual problem. The solution to the dual problem provides a lower bound to the solution of the primal (minimization) problem Boyd (2004). This document briefly introduces Four major dual problems namely the Lagrangian dual problem, the Wolfe dual problem, the Fenchel dual problem and the Pontryagin dual problem but focusses its attention on the frequently used Fenchel dual and Lagrangian dual.

2 The standard Optimization problem

Definition 4.1. The standard Optimization problem can be stated as the primal problem in the following way :

$$\underset{x \in D}{min}\ F_0(x)$$
$$\text{such that } F_i(x) \leq 0 \quad \text{i=1,...,m.}$$

where D is the Domain of the function F_0.

Example 4.1. Linear Programming (LP)

$$\underset{\mathbf{x} \in \mathbb{R}^d}{min}\ \mathbf{a}^T\mathbf{x}$$
$$\text{such that } \mathbf{b_i}^T\mathbf{x} \leq \mathbf{c_i},\ \ i = 1, ..., n.$$

3 The Dual Problem

Following are Four dual problems in Mathematical Optimization :

3.1 Fenchel dual problem

In mathematics and mathematical optimization, the convex conjugate of a function is a generalization of the Legendre transformation. It is also known as Legendre-Fenchel transformation or Fenchel transformation (after Adrien-Marie Legendre and Werner Fenchel). It is used to transform an optimization problem into its corresponding dual problem, which can often be simpler to solve.

Definition 4.2. Let X be a real vector space and X^* be the dual space of X.

Definition 4.3. *Let* $f : X \mapsto \mathbb{R} \quad (X = \mathbb{R}^d)$
$f^* : X^* \mapsto \mathbb{R}$ is called the fenchel dual of f where
$f^*(\mathbf{x}^*) := sup\{\langle \mathbf{x}^*, \mathbf{x}\rangle - f(\mathbf{x}) | \mathbf{x} \in X\}$

This definition can be interpreted as an encoding of the convex hull of the function's epigraph in terms of its supporting hyperplanes.Nielsen

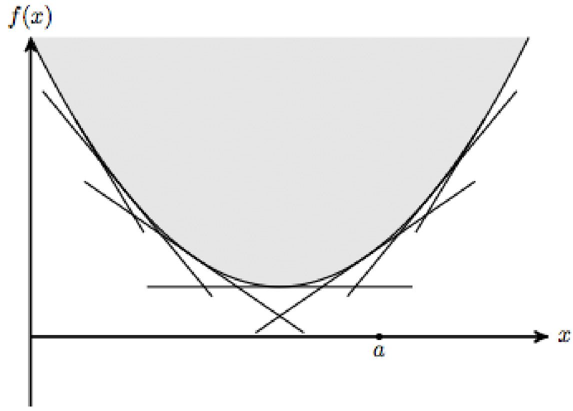

Figure 1: Deciphering function from tangent lines
Courtesy: http://math.stackexchange.com/questions/223235/please-explain-the-intuition-behind-the-dual-problem-in-optimization

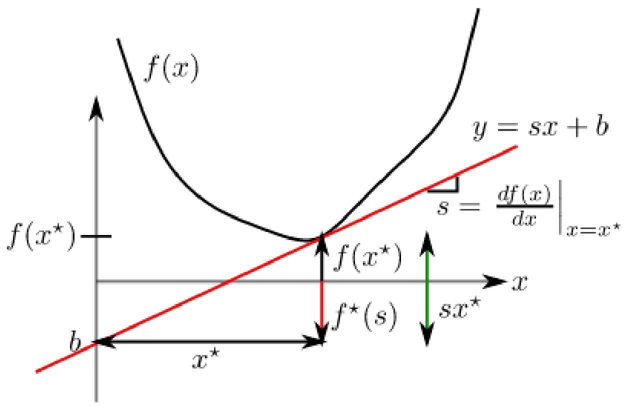

Figure 2: Graphical interpretation of Fenchel Dual
Courtesy: http://www.onmyphd.com/?p=legendre.fenchel.transform

Example 4.2. The fenchel dual of an affine function
$f(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x}\rangle - b, \ \mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}$
is

$$f^*(\mathbf{x}^*) = \begin{cases} b, & \text{if } \mathbf{x} = \mathbf{a} \\ \infty, & \text{otherwise} \end{cases}$$

As Definition 4.3 and Figure 2 try to illustrate, if $f^*(\mathbf{x}^*) = -\infty$ or ∞, then $\mathbf{x}^*$ is not a valid slope of a tangent to f.

Proposition 4.1. Fenchel's Inequality
For any function f and its fenchel dual f^*, Fenchel's inequality (also known as the Fenchel-Young inequality) holds for every $\mathbf{x} \in X$ and $\mathbf{p} \in X^*$:

$\langle \mathbf{p}, \mathbf{x} \rangle \leq f(\mathbf{x}) + f^*(\mathbf{p})$ Wikipedia (a)

Definition 4.4. Let f, f^*, X and X^* as in Definition 4.3. Then,
$f^{**}(\mathbf{x}) := sup\{\langle \mathbf{x}, \mathbf{x}^* \rangle - f^*(\mathbf{x}^*) | \mathbf{x}^* \in X^*\}$ is called the double dual or biconjugate of f. Wikipedia (a)

Claim 4.2. $f = f^{**}$ if and only if f is convex and lower semi-continuous.

Proof. By Fenchel-Moreau Theorem. $\square$

Example 4.3. Consider the function $f(x) = max(1 - x, 0)$.
Then $f^*(x^*) = \begin{cases} x^*, & \text{if } x^* \in [-1, 0] \\ \infty, & \text{otherwise} \end{cases}$

Thus $f^{**}(x) = \underset{x^* \in [-1,0]}{sup} x^*(x - 1) = f(x)$

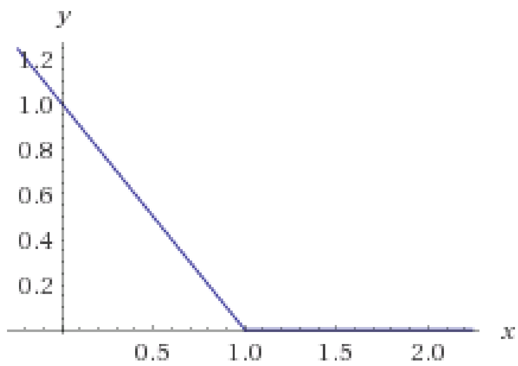

Figure 3: Plot for the function f(x) in example 4.3
Courtesy : http://www.wolframalpha.com/input/?i=f(x)%3Dmax(1-x,0)

Example 4.4. Let $f(x) = \parallel x \parallel$

Then , $f^*(x^*) = \begin{cases} 1, & \text{if } \parallel x \parallel_* \leq 1 \\ \infty, & \text{otherwise} \end{cases}$

Remark 4.1. The fenchel dual is also called Convex Conjugate of f because, $\forall f$, f^* is always convex.

Remark 4.2. Fenchel dual on a non convex function gives a smoothened version of the function that is convex.

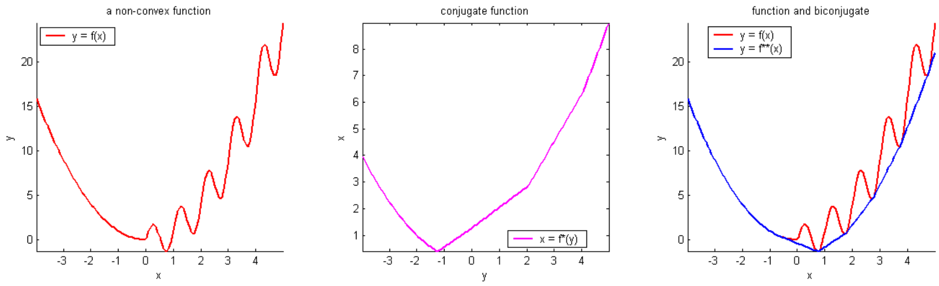

Figure 4: Illustration for Remark 4.2
Courtesy : https://msampler.wordpress.com/2009/07/08/convex-analysis/

3.1.1 Subgradients

In mathematics, the subgradient generalizes the derivative to functions which are not differentiable.

Definition 4.5. Let f be a convex function. Then, a linear function g is a subgradient of f at $\mathbf{x}$ if , $\forall \mathbf{y}$,
$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle g, \mathbf{y} - \mathbf{x} \rangle$

Definition 4.6. The set of all subgradients of a function f is called its subdifferential $\partial_x(f)$. $\partial_x(f) = \{g | g$ is a subgradient of f at $x\}$.

Remark 4.3. If f is a differentiable function then $\partial_x(f)$ has only 1 member, i.e, the differential.

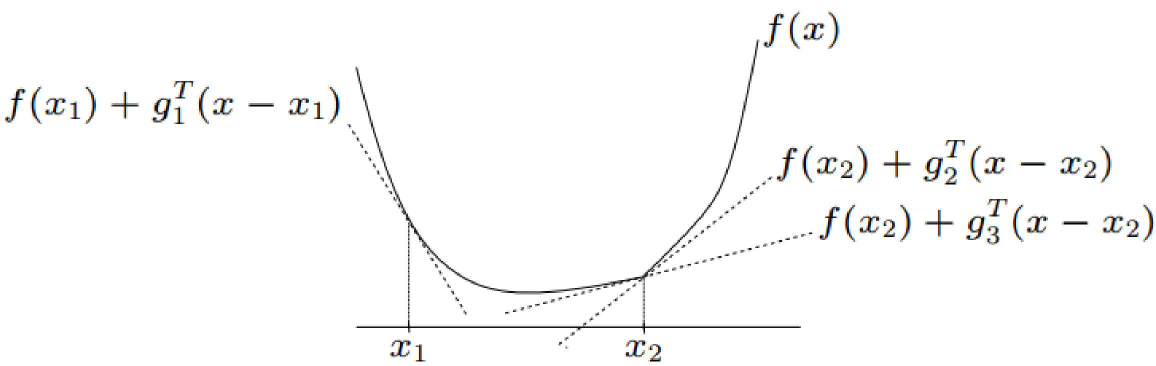

Figure 5: Subgradients of a function
Courtesy : https://optimization.mccormick.northwestern.edu/index.php/Subgradient_optimization

Remark 4.4. For non differentiable functions, the fenchel dual is defined on all directions which are subgradients of f at some point.

3.2 Lagrangian dual problem

The Lagrangian dual problem is obtained by forming the Lagrangian, using nonnegative Lagrange multipliers to add the constraints to the objective function, and then solving for some primal variable values that minimize the Lagrangian. This solution gives the primal variables as

functions of the Lagrange multipliers, which are called dual variables, so that the new problem is to maximize the objective function with respect to the dual variables.Wikipedia (b) It is a method to convert problems into unconstrained problems.
The Primal problem has already been defined in Definition 4.1

Definition 4.7. Let $P := \{\mathbf{x} \in X | F_i(\mathbf{x}) \leq \mathbf{0}\}$ where $F_i(\mathbf{x})$ is as defined in Definition 4.1

Definition 4.8. Let $\mathbf{p}^* = \underset{\mathbf{x} \in P}{min}\ F_0(\mathbf{x})$ where $F_0(\mathbf{x})$ is as defined in Definition 4.1

Definition 4.9. The lagrangian dual function is defined in the following way :

$L(X, \{\lambda_i\}) = inf\{F_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i F_i(\mathbf{x})\}$

Definition 4.10. The lagrangian dual problem can be formalized in the following way:

$$\mathbf{d}^* = \underset{\lambda_i \geq 0}{max} \quad L(X, \{\lambda_i\})$$

Theorem 4.3. Weak Duality Theorem

Consider the Primal Problem P defined in Definition 4.1 and its solution $\mathbf{p}^*$ defined in Definition 4.8. Also consider its Lagrangian dual problem D and its solution $\mathbf{d}^*$ defined in Definition 4.10. Then :
$\mathbf{p}^* \geq \mathbf{d}^*$

Proof. Let $\boldsymbol{\lambda}$ and $\mathbf{x}$ be the feasible solutions to D and P. Then,
$\mathbf{d}^* = L(X, \boldsymbol{\lambda}) = inf\{f(\mathbf{x}_1) + \boldsymbol{\lambda}^T F_i(\mathbf{x}_1) | \mathbf{x}_1 \in X\} \leq f(\mathbf{x}) + \boldsymbol{\lambda}^T F_i(\mathbf{x}) \leq f(\mathbf{x}) = \mathbf{p}^*$ □

Definition 4.11. Duality Gap is defined as the difference between $\mathbf{p}^*$ and $\mathbf{d}^*$.

Theorem 4.4. Strong Duality Theorem

Let X be a non empty convex set in $\mathbb{R}^n$. Let $F_0 : \mathbb{R}^n \mapsto \mathbb{R}$ and $F : \mathbb{R}^n \mapsto \mathbb{R}^m$ be convex. Suppose there exists an $\hat{\mathbf{x}} \in X$ such that $F(\hat{\mathbf{x}}) < \mathbf{0}$. Then
$p^* = d^*$.

Proof. For proof please refer to http://www.eng.newcastle.edu.au/eecs/cdsc/books/cce/Slides/Duality.pdf □

Corollary 4.5. In case of Strong Duality, Duality gap is Zero.

Remark 4.5. The above theorem is also known as Slater's condition for strong duality.

3.3 Wolfe dual problem

In mathematical optimization, Wolfe duality, named after Philip Wolfe, is a type of dual problem in which the objective function and constraints are all differentiable functions.

Definition 4.12. Wolfe dual problem

Consider the Lagrangian dual problem defined in Definition 4.9.
Provided that the functions $F_0, F_1, F_1, ..., F_m$ are continuously differentiable, the infimum occurs

where the gradient is zero. The problem

$\underset{\mathbf{x},\boldsymbol{\lambda}}{maximize}\ \ F_0(\mathbf{x}) + \sum_{j=1}^{m} \lambda_i F_i(\mathbf{x})$

subject to $\nabla F_0(\mathbf{x}) + \sum_{j=1}^{m} \lambda_i \nabla F_i(\mathbf{x})$ and $\boldsymbol{\lambda} \geq \mathbf{0}$

is called the Wolfe dual problem. cha (2011)

Remark 4.6. The Wolfe dual problem is typically a non convex optimization problem.

3.4 Pontryagin duality

In mathematics, specifically in harmonic analysis and the theory of topological groups, Pontryagin duality explains the general properties of the Fourier transform on locally compact groups, such as $\mathbb{R}$, the circle, or finite cyclic groups. The Pontryagin duality theorem itself states that locally compact groups identify naturally with their bidual.

The subject is named after Lev Semenovich Pontryagin who laid down the foundations for the theory of locally compact abelian groups and their duality during his early mathematical works in 1934. Pontryagin's treatment relied on the group being second-countable and either compact or discrete. This was improved to cover the general locally compact abelian groups by Egbert van Kampen in 1935 and Andr Weil in 1940. Wikipedia (c)

References

Duality in convex optimization. 2011. URL http://wwwhome.math.utwente.nl/~stillgj/conopt/chap3.pdf.

Lieven Boyd, Stephen P.; Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

Frank. Nielsen. *Legendre transformation and information geometry.*

Wikipedia. Convex Conjugate, a. URL https://en.wikipedia.org/wiki/Convex_conjugate.

Wikipedia. Fenchel Dual, b. URL https://en.wikipedia.org/wiki/Duality_(optimization).

Wikipedia. Pontryagin duality, c. URL https://en.wikipedia.org/wiki/Pontryagin_duality.