Lagrangian Duality and Convex Optimization

David S. Rosenberg

Bloomberg ML EDU

October 11, 2017

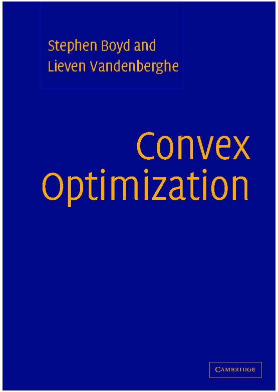
Introduction

Why Convex Optimization?

- Historically:
  - **Linear programs** (linear objectives & constraints) were the focus
  - **Nonlinear programs**: some easy, some hard
- By early 2000s:
  - Main distinction is between **convex** and **non-convex** problems
  - Convex problems are the ones we know how to solve efficiently
  - Mostly batch methods until... around 2010? (earlier if you were into neural nets)
- By 2010 +- few years, most people understood the
  - optimization / estimation / approximation error tradeoffs
  - accepted that **stochatic methods** were often faster to get good results
    - (especially on big data sets)
- These days, nobody's scared of non-convex problems - SGD seems to work well enough on problems of interest (i.e. neural networks).

Your Reference for Convex Optimization

- Boyd and Vandenberghe (2004)
  - Very clearly written, but has a ton of detail for a first pass.
  - See the Extreme Abridgement of Boyd and Vandenberghe.



Notation from Boyd and Vandenberghe

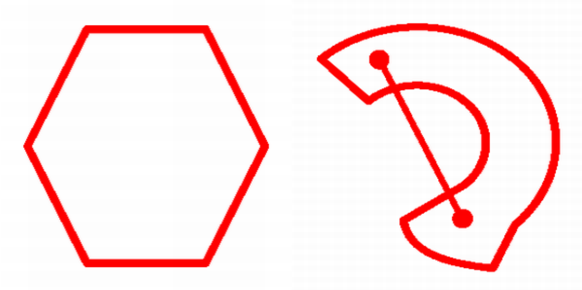
- $f : \mathbf{R}^p \to \mathbf{R}^q$ to mean that f maps from some *subset* of $\mathbf{R}^p$
    - namely $\textbf{dom}\, f \subset \mathbf{R}^p$, where $\textbf{dom}\, f$ is the domain of f

Convex Sets and Functions

Definition

A set C is **convex** if for any $x_1, x_2 \in C$ and any θ with $0 \leqslant \theta \leqslant 1$ we have

$$\theta x_1 + (1-\theta)x_2 \in C.$$

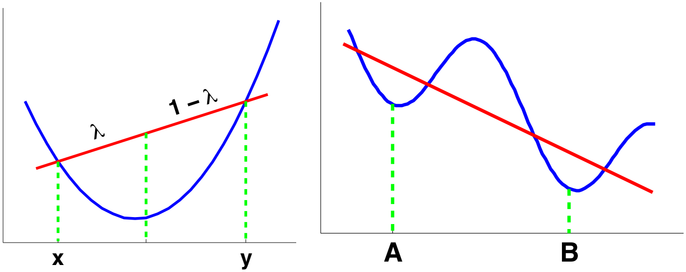


KPM Fig. 7.4

Definition

A function $f : \mathbf{R}^n \to \mathbf{R}$ is **convex** if $\mathbf{dom}\ f$ is a convex set and if for all $x, y \in \mathbf{dom}\ f$, and $0 \leqslant \theta \leqslant 1$, we have

$$f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta) f(y).$$



KPM Fig. 7.5

Examples of Convex Functions on $\mathbf{R}$

Examples

- $x \mapsto ax + b$ is both convex and concave on $\mathbf{R}$ for all $a, b \in \mathbf{R}$.
- $x \mapsto |x|^p$ for $p \geqslant 1$ is convex on $\mathbf{R}$
- $x \mapsto e^{ax}$ is convex on $\mathbf{R}$ for all $a \in \mathbf{R}$
- Every norm on $\mathbf{R}^n$ is convex (e.g. $\|x\|_1$ and $\|x\|_2$)
- Max: $(x_1, \ldots, x_n) \mapsto \max\{x_1 \ldots, x_n\}$ is convex on $\mathbf{R}^n$

Definition

A function f is **strictly convex** if the line segment connecting any two points on the graph of f lies **strictly** above the graph (excluding the endpoints).

Consequences for optimization:

- **convex:** if there is a local minimum, then it is a **global** minimum
- **strictly convex:** if there is a local minimum, then it is the **unique global** minumum

The General Optimization Problem

General Optimization Problem: Standard Form

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

where $x \in \mathbf{R}^n$ are the **optimization variables** and f_0 is the **objective function.**

Assume **domain** $\mathcal{D} = \bigcap_{i=0}^m \mathbf{dom}\, f_i \cap \bigcap_{i=1}^p \mathbf{dom}\, h_i$ is nonempty.

General Optimization Problem: More Terminology

- The set of points satisfying the constraints is called the **feasible set**.
- A point x in the feasible set is called a **feasible point.**
- If x is feasible and $f_i(x) = 0$,
  - then we say the inequality constraint $f_i(x) \leqslant 0$ is **active** at x.
- The **optimal value** p^* of the problem is defined as

$$p^* = \inf\{f_0(x) \mid x \text{ satisfies all constraints}\}.$$

- x^* is an **optimal point** (or a solution to the problem) if x^* is feasible and $f(x^*) = p^*$.

Do We Need Equality Constraints?

- Note that
$$h(x) = 0 \iff (h(x) \geqslant 0 \text{ AND } h(x) \leqslant 0)$$
- Consider an equality-constrained problem:
$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & h(x) = 0 \end{array}$$
- Can be rewritten as
$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & h(x) \leqslant 0 \\ & -h(x) \leqslant 0. \end{array}$$
- For simplicity, we'll drop equality contraints from this presentation.

Lagrangian Duality: Convexity not required

The Lagrangian

The general [inequality-constrained] optimization problem is:

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leqslant 0, \;\; i = 1, \ldots, m \end{array}$$

Definition

The **Lagrangian** for this optimization problem is

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

- λ_i's are called **Lagrange multipliers** (also called the **dual variables**).

The Lagrangian Encodes the Objective and Constraints

- Supremum over Lagrangian gives back encoding of objective and constraints:

$$
\begin{aligned}
\sup_{\lambda \succeq 0} L(x,\lambda) &= \sup_{\lambda \succeq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right) \\
&= \begin{cases} f_0(x) & \text{when } f_i(x) \leqslant 0 \text{ all } i \\ \infty & \text{otherwise.} \end{cases}
\end{aligned}
$$

- Equivalent **primal form** of optimization problem:

$$
p^* = \inf_x \sup_{\lambda \succeq 0} L(x,\lambda)
$$

The Primal and the Dual

- Original optimization problem in **primal form:**

$$p^* = \inf_x \sup_{\lambda \succeq 0} L(x, \lambda)$$

- Get the **Lagrangian dual problem** by "swapping the inf and the sup":

$$d^* = \sup_{\lambda \succeq 0} \inf_x L(x, \lambda)$$

- We will show **weak duality**: $p^* \geqslant d^*$ for any optimization problem.

Weak Max-Min Inequality

Theorem

For ***any*** $f : W \times Z \to \mathbf{R}$, *we have*

$$\sup_{z \in Z} \inf_{w \in W} f(w,z) \leqslant \inf_{w \in W} \sup_{z \in Z} f(w,z).$$

Proof: For any $w_0 \in W$ and $z_0 \in Z$, we clearly have

$$\inf_{w \in W} f(w,z_0) \leqslant f(w_0,z_0) \leqslant \sup_{z \in Z} f(w_0,z).$$

Since $\inf_{w \in W} f(w,z_0) \leqslant \sup_{z \in Z} f(w_0,z)$ for all w_0 and z_0, we must also have

$$\sup_{z_0 \in Z} \inf_{w \in W} f(w,z_0) \leqslant \inf_{w_0 \in W} \sup_{z \in Z} f(w_0,z).$$

Weak Duality

- For any optimization problem (**not just convex**), weak max-min inequality implies **weak duality**:

$$
\begin{aligned}
p^* &= \inf_x \sup_{\lambda \succeq 0} \left[f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \right] \\
&\geqslant \sup_{\lambda \succeq 0, \nu} \inf_x \left[f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \right] = d^*
\end{aligned}
$$

- The difference $p^* - d^*$ is called the **duality gap**.
- For *convex* problems, we often have **strong duality**: $p^* = d^*$.

The Lagrange Dual Function

- The **Lagrangian dual problem**:

$$d^* = \sup_{\lambda \succeq 0} \inf_x L(x, \lambda)$$

Definition

The **Lagrange dual function** (or just **dual function**) is

$$g(\lambda) = \inf_x L(x, \lambda) = \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right).$$

- The dual function may take on the value $-\infty$ (e.g. $f_0(x) = x$).
- The dual function is always **concave**
  - since pointwise min of affine functions

The Lagrange Dual Problem: Search for Best Lower Bound

- In terms of Lagrange dual function, we can write weak duality as

$$p^* \geqslant \sup_{\lambda \geqslant 0} g(\lambda) = d^*$$

- So for any λ with $\lambda \geqslant 0$, **Lagrange dual function gives a lower bound on optimal solution**:

$$p^* \geqslant g(\lambda) \text{ for all } \lambda \geqslant 0$$

The Lagrange Dual Problem: Search for Best Lower Bound

- The **Lagrange dual problem** is a search for best lower bound on p^*:

$$\begin{array}{ll} \text{maximize} & g(\lambda) \\ \text{subject to} & \lambda \succeq 0. \end{array}$$

  - λ **dual feasible** if $\lambda \succeq 0$ and $g(\lambda) > -\infty$.
  - λ^* **dual optimal** or **optimal Lagrange multipliers** if they are optimal for the Lagrange dual problem.
- Lagrange dual problem often easier to solve (simpler constraints).
- d^* can be used as stopping criterion for primal optimization.
- Dual can reveal hidden structure in the solution.

Convex Optimization

Convex Optimization Problem: Standard Form

Convex Optimization Problem: Standard Form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leqslant 0, \;\; i = 1, \ldots, m \end{array}$$

where $f_0, \ldots, f_m$ are convex functions.

Strong Duality for Convex Problems

- For a convex optimization problems, we **usually** have strong duality, but not always.
  - e.g. Convex problem without strong duality:

$$
\begin{array}{ll}
\text{minimize} & e^{-x} \\
\text{subject to} & x^2/y \leqslant 0 \\
 & y > 0
\end{array}
$$

- The additional conditions needed are called **constraint qualifications**.

Example from Laurent El Ghaoui's EE 227A: Lecture 8 Notes, Feb 9, 2012

Slater's Constraint Qualifications for Strong Duality

- Sufficient conditions for strong duality in a **convex** problem.
- Roughly: the problem must be **strictly** feasible.
- Qualifications when problem domain[1] $\mathcal{D} \subset \mathbf{R}^n$ is an open set:
    - **Strict feasibility is sufficient**. ($\exists x \; f_i(x) < 0$ for $i = 1, \ldots, m$)
    - For any affine inequality constraints, $f_i(x) \leqslant 0$ is sufficient.
- If $\mathcal{D}$ not open, see notes or BV Section 5.2.3, p. 226.

[1] $\mathcal{D}$ is the set where all functions are defined, NOT the feasible set.

Complementary Slackness

Complementary Slackness

- Consider a general optimization problem (i.e. not necessarily convex).
- If we have **strong duality**, we get an interesting relationship between
  - the optimal Lagrange multiplier λ_i^* and
  - the ith constraint at the optimum: $f_i(x^*)$
- Relationship is called "**complementary slackness**":

$$\lambda_i^* f_i(x^*) = 0$$

- Always have Lagrange multiplier is zero **or** constraint is active at optimum **or** both.

Complementary Slackness "Sandwich Proof"

- Assume strong duality: $p^* = d^*$ in a general optimization problem
- Let x^* be primal optimal and λ^* be dual optimal. Then:

$$
\begin{aligned}
f_0(x^*) &= g(\lambda^*) = \inf_x L(x, \lambda^*) \quad \text{(strong duality and definition)} \\
&\leqslant L(x^*, \lambda^*) \\
&= f_0(x^*) + \sum_{i=1}^m \underbrace{\lambda_i^* f_i(x^*)}_{\leqslant 0} \\
&\leqslant f_0(x^*).
\end{aligned}
$$

Each term in sum $\sum_{i=1} \lambda_i^* f_i(x^*)$ must actually be 0. That is

$$
\boxed{\lambda_i^* f_i(x^*) = 0, \quad i = 1, \ldots, m}.
$$

This condition is known as **complementary slackness**.