Lagrangian Duality and Convex Optimization

David Rosenberg

New York University

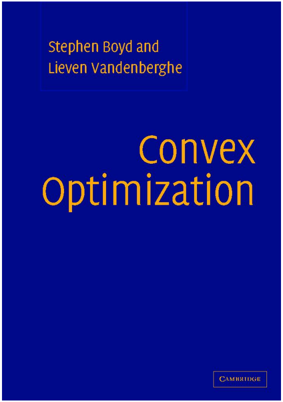
October 29, 2016

Why Convex Optimization?

- Historically:
  - **Linear programs** (linear objectives & constraints) were the focus
  - **Nonlinear programs**: some easy, some hard
- Today:
  - Main distinction is between **convex** and **non-convex** problems
  - Convex problems are the ones we know how to solve efficiently
- Many techniques that are well understood for convex problems are applied to non-convex problems
  - e.g. SGD is routinely applied to 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 my "Extreme Abridgement of Boyd and Vandenberghe".



Notation from Boyd and Vandenberghe

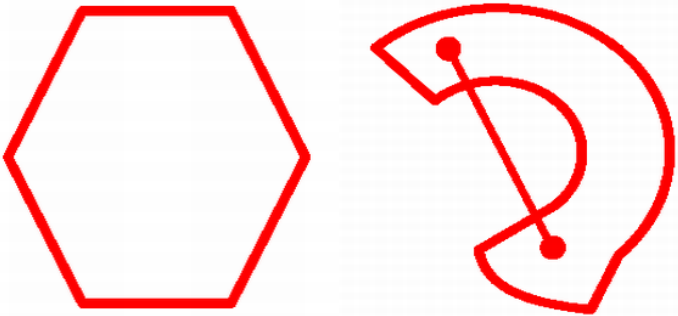
- $f : \mathbf{R}^p \to \mathbf{R}^q$ to mean that f maps from some *subset* of $\mathbf{R}^p$
  - namely $\textbf{dom}\, f \subset \mathbf{R}^p$, where $\textbf{dom}\, f$ is the domain of f

Convex Sets

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

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



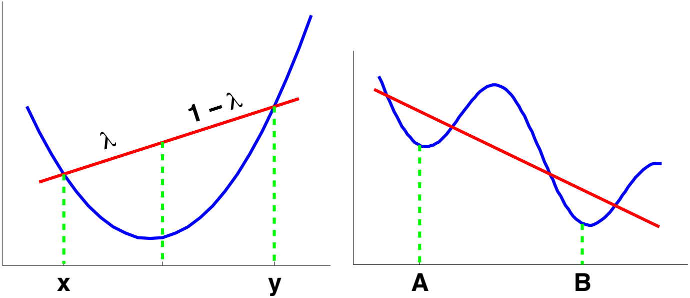
KPM Fig. 7.4

Convex and Concave Functions

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 $a \in \mathbf{R}$

Maximum of Convex Functions is Convex

Theorem

If $f_1, \ldots, f_m : \mathbf{R}^n \to \mathbf{R}$ are convex, then their pointwise maximum

$$f(x) = \max\{f_1(x), \ldots, f_m(x)\}$$

is also convex with domain $\textbf{dom}\, f = \textbf{dom}\, f_1 \cap \cdots \cap \textbf{dom}\, f_m$.

This result extends to sup over arbitrary [infinite] sets of functions.

Proof.

(For $m = 2$.) Fix an $0 \leqslant \theta \leqslant 1$ and $x, y \in \textbf{dom}\, f$. Then

$$\begin{aligned} f(\theta x + (1-\theta) y) &= \max\{f_1(\theta x + (1-\theta) y), f_2(\theta x + (1-\theta) y)\} \\ &\leqslant \max\{\theta f_1(x) + (1-\theta) f_1(y), \theta f_2(x) + (1-\theta) f_2(y)\} \\ &\leqslant \max\{\theta f_1(x), \theta f_2(x)\} + \max\{(1-\theta) f_1(y), (1-\theta) f_2(y)\} \\ &= \theta f(x) + (1-\theta) f(y) \end{aligned}$$

Convex Functions and Optimization

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

General Optimization Problem: Standard Form

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** $\mathcal{D} = \bigcap_{i=0}^{m} \textbf{dom } f_i \cap \bigcap_{i=1}^{p} \textbf{dom } h_i$ 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

$$p^* = \inf\{f_0(x) \mid f_i(x) \leqslant 0, i = 1, \ldots, m,\ h_i(x) = 0,\ i = 1, \ldots, p\}.$$

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

The Lagrangian

Recall the general optimization problem:

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

Definition

The **Lagrangian** for the general optimization problem is

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

- λ_i's and ν's are called **Lagrange multipliers**
- λ and ν also called the **dual variables** .

The Lagrangian Encodes the Objective and Constraints

- Supremum over Lagrangian gives back objective and constraints:

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

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

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

The Primal and the Dual

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

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

- The **Lagrangian dual problem**:

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

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

Weak Max-Min Inequality

Theorem

For ***any*** $f:\mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$, $W \subseteq \mathbf{R}^n$, *or* $Z \subseteq \mathbf{R}^m$, *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 this is true 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**:

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

- 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**:

$$d^* = \sup_{\lambda \succeq 0, \nu} \underbrace{\inf_x L(x, \lambda, \nu)}_{\text{Lagrange dual function}}$$

Definition

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

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

- The dual function may take on the value $-\infty$ (e.g. $f_0(x) = x$).

The Lagrange Dual Problem

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

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

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

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

The Lagrange Dual Problem

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

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

  - (λ,ν) **dual feasible** if $\lambda \succeq 0$ and $g(\lambda,\nu) > -\infty$.
  - (λ^*,ν^*) are **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 Problem: Standard Form

Convex Optimization Problem: Standard Form

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

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

Note: Equality constraints are now linear. Why?

Strong Duality for Convex Problems

- For a convex optimization problems, we **usually** have strong duality, but not always.
  - For example:

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

- 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 $\mathcal{D} \subset \mathbf{R}^n$ is an open set:
  - $\exists x$ such that $Ax = b$ and $f_i(x) < 0$ for $i = 1, \ldots, m$
  - For any affine inequality constraints, $f_i(x) \leqslant 0$ is sufficient
- Otherwise, x must be in the "relative interior" of $\mathcal{D}$
  - See notes, or BV Section 5.2.3, p. 226.

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**":

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

- Lagrange multiplier is zero unless the constraint is active at the optimum.

Complementary Slackness Proof

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

$$
\begin{aligned}
f_0(x^*) &= g(\lambda^*, \nu^*) \\
&= \inf_x \left(f_0(x) + \sum_{l=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right) \\
&\leqslant f_0(x^*) + \sum_{i=1}^{m} \underbrace{\lambda_i^* f_i(x^*)}_{\leqslant 0} + \sum_{i=1}^{p} \underbrace{\nu_i^* h_i(x^*)}_{=0} \\
&\leqslant f_0(x^*).
\end{aligned}
$$

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

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

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