Extreme Abridgement of Boyd and Vandenberghe's *Convex Optimization*

Compiled by David Rosenberg

Abstract

Boyd and Vandenberghe's *Convex Optimization* book is very well-written and a pleasure to read. The only potential problem is that, if you read it sequentially, you have to go through almost 300 pages to get through duality theory. It turns out that a well-chosen 10 pages are enough for a self-contained introduction to the topic. Besides a rare extra comment or two, the text here is copied essentially verbatim from the original. My main contribution is deciding what to leave out.

1 Notation

- Use notation $f : \mathbf{R}^p \to \mathbf{R}^q$ to mean that f maps from some *subset* of $\mathbf{R}^p$, namely $\mathrm{dom}\, f \subset \mathbf{R}^p$, where $\mathrm{dom}\, f$ stands for the domain of the function f
- $\mathbf{R}$ are the real numbers
- $\mathbf{R}_+$ are nonnegative reals
- $\mathbf{R}_{++}$ are positive reals

2 Affine and Convex Sets (BV 2.1)

2.1 Affine Sets

Intuitively, an affine set is any point, line, plane, or hyperplane. But let's make this more precise.

Definition 1. A set $C \subseteq \mathbf{R}^n$ is **affine** if the line through any two distinct points in C lies in C. That is, if for any $x_1, x_2 \in C$ and $\theta \in \mathbf{R}$, we have $\theta x_1 + (1-\theta)x_2 \in C$.

Recall that a **subspace** is a subset of a vector space that is closed under sums and scalar multiplication. If C is an affine set and $x_0 \in C$, then the set $V = C - x_0 = \{x - x_0 \mid x \in C\}$ is a subspace. Thus, we can also write an affine set as $C = V + x_0 = \{v + x_0 \mid v \in V\}$, i.e. as a subspace plus an offset. The subspace V associated with the affine set C does not depend on the choice of $x_0 \in C$. Thus we can make the following definition:

Definition 2. The **dimension of an affine set** C is the dimension of the subspace $V = C - x_0$, where x_0 is any element of C.

We note that the solution set of a system of linear equations is an affine set, and every affine set can be expessed as the solution of a system of linear equations [BV Example 2.1, p. 22].

Definition 3. A **hyperplane** in $\mathbf{R}^n$ is a a set of the form

$$\{x | a^T x = b\},$$

for $a \in \mathbf{R}^n, a \neq 0, b \in \mathbf{R}$, and where a is the normal vector to the hyperplane.

Note that a hyperplane in $\mathbf{R}^n$ is an affine set of dimension $n-1$.

2.2 Convex Sets (BV 2.1.4)

A set C is convex if the line segment between any two points in C lies in C. That is, if for any $x_1, x_2 \in C$ and any θ with $0 \leq \theta \leq 1$ we have

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

Every affine set is also convex.

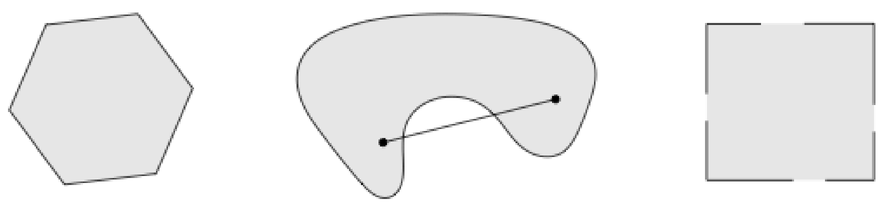

Figure 2.2 Some simple convex and nonconvex sets. *Left.* The hexagon, which includes its boundary (shown darker), is convex. *Middle.* The kidney shaped set is not convex, since the line segment between the two points in the set shown as dots is not contained in the set. *Right.* The square contains some boundary points but not others, and is not convex.

2.3 Spans and Hulls

Given a set of points $x_1, \ldots x_k \in \mathbf{R}^n$, there are various types of linear combinations that we can take:

- A **linear combination** is a point of the form $\theta_1 x_1 + \cdots + \theta_k x_k$, with no constraints on θ_i's. The **span** of $x_1, \ldots, x_k$ is the set of all linear combinations of $x_1, \ldots, x_k$.

- An **affine combination** is a point of the form $\theta_1 x_1 + \cdots + \theta_k x_k$, where $\theta_1 + \cdots + \theta_k = 1$. The **affine hull** of $x_1, \ldots, x_k$ is the set of all affine combinations of $x_1, \ldots, x_k$.

- A **convex combination** is a point of the form $\theta_1 x_1 + \cdots + \theta_k x_k$, where $\theta_1 + \cdots + \theta_k = 1$ and $\theta_i \geq 0$ for all i. The **convex hull** of $x_1, \ldots, x_k$ is the set of all convex combinations of $x_1, \ldots, x_k$.

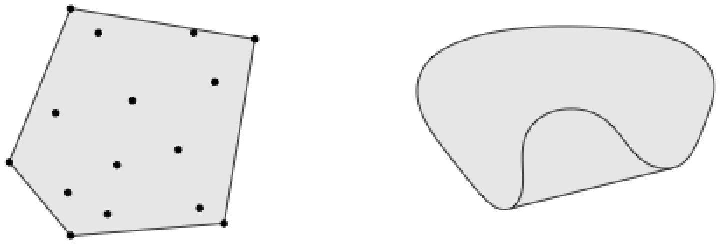

Figure 2.3 The convex hulls of two sets in $\mathbf{R}^2$. *Left.* The convex hull of a set of fifteen points (shown as dots) is the pentagon (shown shaded). *Right.* The convex hull of the kidney shaped set in figure 2.2 is the shaded set.

3 Convex Functions

3.1 Definitions (BV 3.1, p. 67)

Definition 4. A function $f : \mathbf{R}^n \to \mathbf{R}$ is **convex** if dom f is a convex set and if for all $x, y \in$ dom f, and $0 \le \theta \le 1$, we have

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

A function f is **concave** if $-f$ is convex.

Geometrically, a function is convex if the line segment connecting any two points on the graph of f lies above the graph:

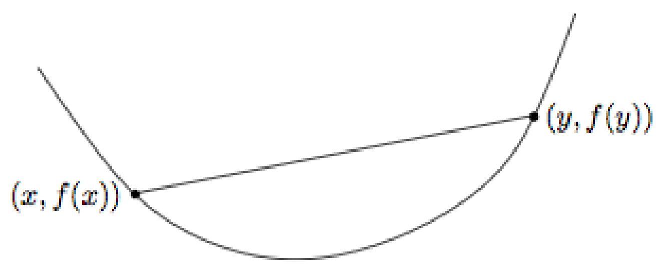

Figure 3.1 Graph of a convex function. The chord (*i.e.*, line segment) between any two points on the graph lies above the graph.

Definition 5. A function f is **strictly convex** if when we additionally restrict $x \neq y$ and $0 < \theta < 1$, then we get strict inequality:

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

Definition 6. A function f is **strongly convex** if $\exists \mu > 0$ such that

$$x \mapsto f(x) - \mu\|x\|^2$$

is convex. The largest possible μ is called the **strong convexity constant**.

3.1.1 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

strongly convex: **there exists a unique global** minimum

3.1.2 First-order conditions (BV 3.1.3)

The following characterization of convex functions is possibly "obvious from the picture", but we highlight it here because later it forms the basis for the definition of the "subgradient", which generalizes the gradient to nondifferentiable functions.

Suppose $f : \mathbf{R}^n \to \mathbf{R}$ is differentiable (*i.e.* dom f is open and ∇f exists at each point indom f). Then f is convex if and only if dom f is convex and

$$f(y) \geq f(x) + \nabla f(x)^T (y - x)$$

holds for all $x, y \in$ dom f. In other words, for a convex differentiable function, the linear approximation to f at x is a *global underestimator* of f:

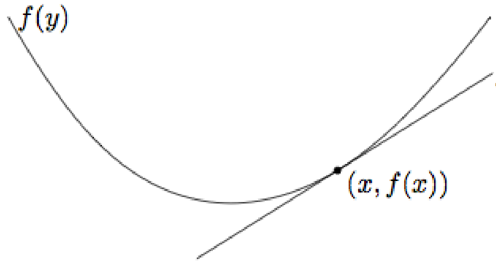

Figure 3.2 If f is convex and differentiable, then $f(x)+\nabla f(x)^T(y-x) \leq f(y)$ for all $x,\ y \in \mathbf{dom}\, f$.

The inequality shows that from *local information* about a convex function (i.e. its value and derivative at a point) we can derive *global information* (i.e. a global underestimator of it). **This is perhaps the most important property of convex funtions**. For example, the inequality shows that if $\nabla f(x) = 0$, then for all $y \in$ dom f, $f(y) \geq f(x)$, i.e. x is a global minimizer of f.

3.1.3 Examples of Convex Functions (BV 3.1.5)

Functions mapping from $\mathbf{R}$:

- $x \mapsto e^{ax}$ is convex on $\mathbf{R}$ for all $a \in \mathbf{R}$
- $x \mapsto x^a$ is convex on $\mathbf{R}_{++}$ when $a \geq 1$ or $a \leq 0$ and concave for $0 \leq a \leq 1$
- $|x|^p$ for $p \geq 1$ is convex on $\mathbf{R}$
- $\log x$ is concave on $\mathbf{R}^{++}$
- $x \log x$ (either on $\mathbf{R}_{++}$ or on $\mathbf{R}_{+}$ if we define $0 \log 0 = 0$) is convex

Functions mapping from $\mathbf{R}^n$:

- Every norm on $\mathbf{R}^n$ is convex
- Max: $(x_1, \ldots, x_n) \mapsto \max\{x_1 \ldots, x_n\}$ is convex on $\mathbf{R}^n$
- Log-Sum-Exp[1]: $(x_1, \ldots, x_n) \mapsto \log\left(e^{x_1} + \cdots + e^{x_n}\right)$ is convex on $\mathbf{R}^n$.

3.2 Operations the preserve convexity (Section 3.2, p. 79)

3.2.1 Nonnegative weighted sums

If $f_1, \ldots, f_m$ are convex and $w_1, \ldots, w_m \geq 0$, then $f = w_1 f_1 + \cdots + w_m f_m$ is convex. More generally, if $f(x, y)$ is convex in x for each $y \in \mathcal{A}$, and if $w(y) \geq 0$ for each $y \in \mathcal{A}$, then the function

$$g(x) = \int_{\mathcal{A}} w(y) f(x, y)\, dy$$

is convex in x (provided the integral exists).

3.2.2 Composition with an affine mapping

A function $f : \mathbf{R}^n \to \mathbf{R}^m$ is an **affine function** (or **affine mapping)** if it is a sum of a linear function and a constaint. That is, if it has the form $f(x) = Ax + b$, where $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$.

Composition of a convex function with an affine function is convex. More precisely: suppose $f : \mathbf{R}^n \to \mathbf{R},\ A \in \mathbf{R}^{n \times m}$ and $b \in \mathbf{R}^n$. Define $g : \mathbf{R}^m \to \mathbf{R}$ by

$$g(x) = f\left(Ax + b\right),$$

with $\operatorname{dom} g = \{x \mid Ax + b \in \operatorname{dom} f\}$. Then if f is convex, then so is g; if f is concave, so is g. If f is **strictly** convex, and A has linearly independent columns, then g is also strictly convex.

4 Optimization Problems (BV Chapter 4)

4.1 General Optimization Problems (BV Section 4.1.1)

The standard form for an optimization problem is the following:

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

where $x \in \mathbf{R}^n$ are called the **optimization variables**. The function $f_0 : \mathbf{R}^n \to \mathbf{R}$ is the **objective function** (or **cost function**); the inequalities $f_i(x) \leq 0$ are called **inequality constraints** and the corresponding functions $f_i : \mathbf{R}^n \to \mathbf{R}$ are called the

[1] This function can be interpreted as a differentiable (in fact, analytic) approximation to the max function, since

$$\max\{x_1, \ldots, x_n\} \leq \log\left(e^{x_1} + \cdots + e^{x_n}\right) \leq \max\{x_1, \ldots, x_n\} + \log n.$$

Can you prove it? Hint: $\max(a, b) \leq a + b \leq 2 \max(a, b)$.

inequality constraint functions. The equations $h_i(x) = 0$ are called the **equality constraints** and the functions $h_i : \mathbf{R}^n \to \mathbf{R}$ are the **equality constraint functions**. If there are no constraints (i.e. $m = p = 0$), we say the problem is **unconstrained**.

The set of points for which the objective and all constraint functions are defined,

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

is called the **domain of the optimization problem**. A point $x \in \mathcal{D}$ is **feasible** if it satisfies all the equality and inequality constraints. The set of all feasible points is called the **feasible set** or the **constraint set**.

The **optimal value** p^* of the problem is defined as

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

Note that if the problem is infeasible, $p^* = \infty$, since it is the inf of an empty set.

We say that x^* is an **optimal point** (or is a solution to the problem) if x^* is feasible and $f(x^*) = p^*$. The set of optimal pointes is the **optimal set**.

We say that a feasible point x is **locally optimal** if there is an $R > 0$ such that x solves the following optimization problem:

$$\begin{array}{ll} \text{minimize} & f_0(z) \\ \text{subject to} & f_i(z) \leq 0, \quad i = 1, \ldots, m \\ & h_i(z) = 0, \quad i = 1, \ldots p \\ & \|z - x\|_2 \leq R \end{array}$$

with optimization variable z. Roughly speaking, this means x minimizes f_0 over nearby points in the feasible set.

4.2 Convex Optimization Problems (Section 4.2, p. 136)

4.2.1 Convex optimization problems in standard form (Section 4.2.1)

The **standard form** for a **convex optimization problem** is the following:

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

where $f_0, \ldots, f_m$ are convex functions. Compared with the general standard form, the convex problem has three additional requirements:

- the objective function must be convex
- the inequality constraint functions must be convex
- the equality constraints functions must be affine

We immediately note an important property: the feasible set of a convex optimization problem is convex (see BV p. 137).

4.2.2 Local and global Optima (4.2.2, p. 138)

Fact 7. *A fundamental property of convex optimization problems is that any locally optimal point is also globally optimal.*

5 Duality (BV Chapter 5)

5.1 The Lagrangian (BV Section 5.1.1)

We again consider the general optimization problem in standard form:

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

with variable $x \in \mathbf{R}^n$. We assume its domain $\mathcal{D} = \bigcap_{i=0}^{m} \mathrm{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathrm{dom}\, h_i$ is nonempty and denote the optimal value by p^*. We do not assume the problem is convex.

Definition 8. The **Lagrangian** $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ for the general optimization problem defined above is

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

with dom $L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$. We refer to the λ_i as the **Lagrange multiplier** associated with the ith inequality constraint and ν_i as the Lagrange multiplier associated with the ith equality constraint. The vectors λ and ν are called the **dual variables** or **Lagrange multiplier vectors**.

[WARNING: Beyond this point, it should still be mostly readable, but it's not complete.]

5.1.1 My motivation

Consider the "PRIMAL PROBLEM"

$$f^* = \inf_{x \in \mathbf{R}^d} f(x)$$
$$\text{s.t.} g(x) \leq 0$$

Note that

$$f^* = \inf_{x \in \mathbf{R}^d} \sup_{\lambda \geq 0} \left[f(x) + \lambda g(x) \right]$$

"Proof":

- Say x is s.t. $g(x) > 0$ [i.e. not a feasible point].
- Then $\sup_{\lambda > 0} \left[f(x) + \lambda g(x) \right] = \infty$, so x is not chosen for $\min_{x \in \mathbf{R}^d}$
- If x is s.t $g(x) \leq 0$ [i.e. it is a feasible point], then we get $\lambda = 0$ (or $g(x) = 0$) and so problem reduces to
$$\min_{x \in \{x \mid g(x) \leq 0\}} f(x)$$

5.2 The Lagrange dual function (Section 5.1.2 p. 216)

We define the *Lagrange dual function* (or just *dual function*) $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ as the minimum value of the Lagrangian over x: for $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}^p$,

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

When the Lagrangian is unbounded below in x, the dual function takes on the value $-\infty$. Since the dual function is the pointwise infimum of a family of affine functions of (λ, ν), it is concave, even when the problem (5.1) is not convex.

Theorem 9. *In general, we have the "Saddlepoint Inequality", which is*

$$\min_x \max_y f(x,y) \geq \max_y \min_x f(x,y)$$

Proof. Well, for any x_0, y_0, we have

$$\max_y f(x_0, y) \geq \min_x f(x, y_0)$$

Then since this is true for all x_0 and y_0, we can write

$$\min_{x_0} \max_y f(x_0, y) \geq \max_{y_0} \min_x f(x, y_0)$$

□

So we always have "weak duality":

$$\begin{aligned} f^* &= \inf_{x \in \mathbf{R}^d} \sup_{\lambda \geq 0} [f(x) + \lambda g(x)] \\ &\geq \sup_{\lambda \geq 0} \inf_{x \in \mathbf{R}^d} [f(x) + \lambda g(x)] \end{aligned}$$

In "strong duality", this is an equality. (Strong duality under certain conditions, e.g. everything convex and "Slater's condition".)

- So solving something using the dual, means swapping the sup and the inf.

Definition:

- The Lagrangian is defined as
$$L(x, \lambda) = f(x) + \lambda g(x)$$
- The λ is called a Lagrange multiplier or a dual variable.
- The "dual objective" is defined as
$$h(\lambda) = \inf_{x \in \mathbf{R}^d} [f(x) + \lambda g(x)]$$
Rewriting the above with this notation, AND assuming strong duality, we have
$$\begin{aligned} f^* &= \inf_{x \in \mathbf{R}^d} \sup_{\lambda \geq 0} L(x, \lambda) \\ &= \sup_{\lambda \geq 0} \inf_{x \in \mathbf{R}^d} L(x, \lambda) \\ &= \sup_{\lambda \geq 0} h(\lambda) \end{aligned}$$
That last thing is the "dual problem".

References