# Extreme Abridgment 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. Most of the text here is copied essentially verbatim from the original.

## 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 $\mathbf{dom}\ f \subset \mathbf{R}^p$, where $\mathbf{dom}\ f$ stands for the domain of the function $f$
- $\mathbf{R}$ are the real numbers
- $\mathbf{R}_+$ are nonnegative reals
- $\mathbf{R}_{++}$ are positive reals
- $a \succeq b$ for $a, b \in \mathbf{R}^d$ means component-wise inequality – i.e. $a_i \geq b_i$ for $i \in \{1, \ldots, d\}$

## 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 expressed 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)

**Definition 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 $\theta$ with $0 \leq \theta \leq 1$ we have

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

Every affine set is also convex.

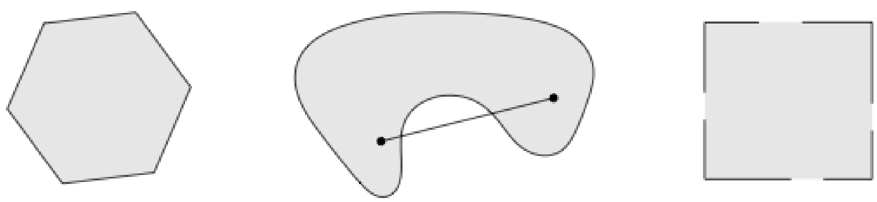

**Figure 2.2** Some simple convex and nonconvex sets. *Left.* The hexagon, which includes its boundary (shown darker), is convex. *Middle.* The kidney shaped set is not convex, since the line segment between the two points in the set shown as dots is not contained in the set. *Right.* The square contains 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 $\theta_i$'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 $x_1, \ldots, x_k$, denoted **aff** $(x_1, \ldots, x_k)$, 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 $x_1, \ldots, x_k$.

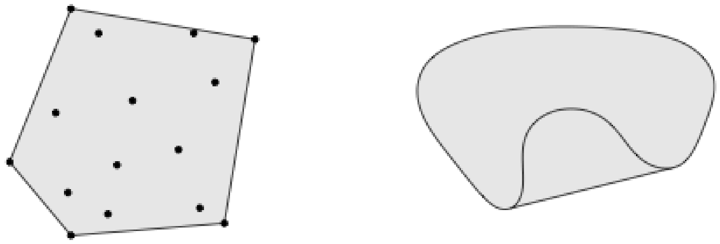

**Figure 2.3** The convex hulls of two sets in $\mathbf{R}^2$. *Left.* The convex hull of a set of fifteen points (shown as dots) is the pentagon (shown shaded). *Right.* The convex hull of the kidney shaped set in figure 2.2 is the shaded set.

# 3 Convex Functions

## 3.1 Definitions (BV 3.1, p. 67)

**Definition 5.** 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 $0 \leq \theta \leq 1$, we have

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

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:

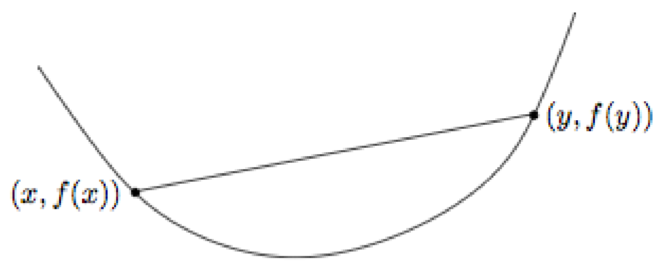


**Figure 3.1** Graph of a convex function. The chord (*i.e.*, line segment) between any two points on the graph lies above the graph.

**Definition 6.** 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 7.** A function $f$ is **strongly convex** if $\exists \mu > 0$ such that

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

is convex. The largest possible $\mu$ 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** minimum

**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.* $\mathbf{dom}\ f$ is open and $\nabla f$ exists at each point in$\mathbf{dom}\ f$). Then $f$ is convex if and only if $\mathbf{dom}\ f$ is convex and

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

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

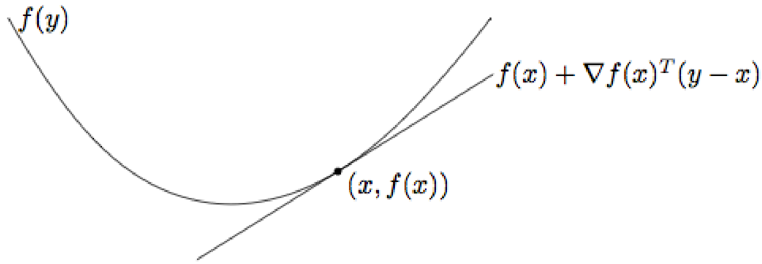


**Figure 3.2** If $f$ is convex and differentiable, then $f(x)+\nabla f(x)^T(y-x) \leq f(y)$ for all $x,\ y \in \mathbf{dom}\, f$.

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 functions**. For example, the inequality shows that if $\nabla f(x) = 0$, then for all $y \in \mathbf{dom}\ f$, $f(y) \geq f(x)$, 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]: $(x_1, \ldots, x_n) \mapsto \log(e^{x_1} + \cdots + e^{x_n})$ 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 constant. 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

$$g(x) = f(Ax + b),$$

with $\mathbf{dom}\ g = \{x \mid Ax + b \in \mathbf{dom}\ f\}$. 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.

### 3.2.3 Simple Composition Rules

- If $g$ is convex then $\exp g(x)$ is convex.
- If $g$ is convex and nonnegative and $p \geq 1$ then $g(x)^p$ is convex.
- If $g$ is concave and positive then $\log g(x)$ is concave
- If $g$ is concave and positive then $1/g(x)$ is convex.

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

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

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

### 3.2.4 Maximum of convex functions is convex (BV Section 3.2.3, p. 80)

*Note: Below we use this to prove that the Lagrangian dual function is concave.*

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 $\mathbf{dom}\ f = \mathbf{dom}\ f_1 \cap \cdots \cap \mathbf{dom}\ f_m$.

This result extends to the supremum over arbitrary sets of functions (including uncountably infinite sets).

# 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:

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

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 **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,

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

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**. If $x$ is feasible and $f_i(x) = 0$, then we say the $i$th inequality constraint $f_i(x) \leq 0$ is **active** at $x$.

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 points 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:

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

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:

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

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 8.** *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:

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

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

**Definition 9.** 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 $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$. We refer to the $\lambda_i$ as the **Lagrange multiplier** associated with the $i$th inequality constraint and $\nu_i$ as the Lagrange multiplier associated with the $i$th equality constraint. The vectors $\lambda$ and $\nu$ are called the **dual variables** or **Lagrange multiplier vectors**.

### 5.1.1 Max-min characterization of weak and strong duality (BV Section 5.4.1)

Note that

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

In words, when $x$ is in the feasible set, we get back the objective function: $\sup_{\lambda\succeq 0} L(x,\lambda) = f_0(x)$. Otherwise, we get $\infty$. Proof: Suppose $x$ violates an inequality constraint, say $f_i(x) > 0$. Then $\sup_{\lambda\succeq 0} L(x,\lambda) = \infty$, which we can see by taking $\lambda_j = 0$ for $j \neq i$, taking all $\nu_i = 0$, and sending $\lambda_i \to \infty$. We can make a similar argument for an equality constraint violation. If $x$ is feasible, then $f_i(x) \le 0$ and $h_i(x) = 0$ for all $i$, and thus the supremum is achieved by taking $\lambda = 0$, which yields $\sup_{\lambda\succeq 0,\nu} L(x,\lambda) = f_0(x)$.

It should now be clear that we can write the original optimization problem as

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

In this context, this optimization problem is called the **primal problem**. We get the **Lagrange dual problem** by swapping the inf and the sup:

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

where $d^*$ is the optimal value of the Lagrange dual problem.

**Theorem 10** (Weak max-min inequality, BV Exercise 5.24, p. 281)**.** *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*

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

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

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

Since this is true for all $w_0$ and $z_0$, we must also have

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

□

In the context of an optimization problem, the weak max-min inequality is called **weak duality**, *which always holds for any optimization problem (**not just convex**):*

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

The gap $p^* - d^*$ is called the **duality gap**. For convex optimization problems, we very often have **strong duality**, which is when we have the equality: $p^* = d^*$**.**

### 5.1.2 The Lagrange Dual Function (BV 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 $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}^p$,

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

This is the inner minimization problem of the Lagrange dual problem discussed above. When the Lagrangian is unbounded below in $x$, the dual function takes on the value $-\infty$. **The dual function is concave even when the optimization problem is not convex**, since the dual function is the pointwise infimum of a family of affine functions of $(\lambda, \nu)$ (a different affine function for each $x \in \mathcal{D}$).

### 5.1.3 The Lagrange Dual Problem

From weak duality, it is clear than for each pair $(\lambda, \nu)$ with $\lambda \geq 0$, the Lagrange dual function $g(\lambda, \nu)$ gives us a lower bound on $p^*$. A search for the best possible lower bound is one motivation for the Lagrange dual problem, which now we can write as

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

In this context, a pair $(\lambda, \nu)$ is called **dual feasible** is $\lambda \succeq 0$ and $g(\lambda, \nu) > -\infty$. We refer to $(\lambda^*, \nu^*)$ as **dual optimal** or **optimal Lagrange multipliers** if they are optimal for the Lagrange dual problem.

The Lagrange dual problem is as convex optimization problem, since the objective is concave and the constraint is convex. This is the case whether or not the primal problem is convex.

## 5.2 Strong duality and Slater's constraint qualification (5.2.3, p. 226)

For a convex optimization problem in standard form, we usually have strong duality, but not always. The additional conditions needed are called **constraint qualifications**. To state these conditions in their full generality, we need some new definitions. So we'll start with some consequences that are easier to state and use:

**Corollary 11.** *For a convex optimization problem in standard form, if the domain*[2] $\mathcal{D} = \bigcap_{i=0}^{m}$ ***dom*** $f_i$ *is open*[3]*, and there exists an* $x \in \mathcal{D}$ *such that* $Ax = b$ *and* $f_i(x) < 0$ *for* $i = 1, \ldots, m$ *(such a point is called* ***strictly feasible****), then strong duality holds. Moreover, if* $d^* > -\infty$*, then the dual optimum is attained – that is, there exists a dual feasible* $(\lambda^*, \nu^*)$ *with* $g(\lambda^*, \nu^*) = d^* = p^*$*. If* $f_1, \ldots, f_k$ *are affine functions, it is sufficient to replace the strict inequality constraints for* $f_1, \ldots, f_k$ *with inequality constraints* $f_i(x) \leq 0$ *for* $i = 1, \ldots, k$*, while the other conditions remain the same.*

[2] Recall the domain of a function **dom** $f_i$ is the set where the function is defined, and so $\mathcal{D}$ is the set where all the functions in the optimization problem are defined. In particular, $\mathcal{D}$ is NOT the feasible set (though it contains the feasible set).

[3] For example, $\mathbf{R}^d$ is an open set, and is the domain we encounter most often.

Simplified version appropriate for SVM:

**Corollary 12.** *For a convex optimization problem in standard form, if the domain of $f_0$ is open, all equality and inequality constraints are linear, and the problem is feasible (i.e. there is some point in the domain that satisfies all the constraints), then we have strong duality and, if $d^* > -\infty$, then the dual optimum is attained.*

For a more general statement, let's define the **affine dimension** of a set $C$ as the dimension of its affine hull. We define the **relative interior** of the set $C$, denoted **relint** $C$, as the interior relative to **aff** $C$:

$$\mathbf{relint}\ C = \{x \in C \mid B(x,r) \cap \mathbf{aff}\ C \subseteq C \text{ for some } r > 0\}\,,.$$

where $B(x,r) = \{x \mid \|y-x\| \le r\}$, is the ball of radius $r$ and center $x$ in the norm $\|\cdot\|$. Also, recall that for a convex optimization problem in standard form, the domain $\mathcal{D}$ is the intersection of the domain of the objective and the inequality constraint functions:

$$\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\ f_i.$$

**Theorem 13.** *For a convex optimization problem, if there exists an $x \in$ relint $\mathcal{D}$ such that $Ax = b$ and $f_i(x) < 0$ for $i = 1, \dots, m$ (such a point is sometimes called strictly feasible), then strong duality holds and, if $d^* > -\infty$, then the dual optimum is attained. If $f_1, \dots, f_k$ are affine functions, it is sufficient to replace the strict inequality constraints for $f_1, \dots, f_k$ with inequality constraints $f_i(x) \le 0$ for $i = 1, \dots, k$, while the other conditions remain the same.*

## 5.3 Optimality Conditions (BV 5.5, p. 241)

### 5.3.1 Complementary slackness (BV 5.5.2, p. 242)

Suppose that the primal and dual optimal values are attained and equal (so, in particular, strong duality holds, but we're not assuming convexity). Let $x^*$ be a primal optimal and $(\lambda^*, \nu^*)$ be a dual optimal point. This means that

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

The first line states that the duality gap is zero, and the second line is the definition of the dual function. The third line follows since the infimum of the Lagrangian over $x$ is less than or equal to its value at $x = x^*$. The last inequality follows from feasibility of $\lambda^*$ and $x^*$ (meaning $\lambda^* \succeq 0$, $f_i(x^*) \le 0$ and $h_i(x^*) = 0$, for all $i$. Thus the inequalities are actually equalities. We can draw two interesting conclusions:

1. Since the third line is an equality, $x^*$ minimizes $L(x, \lambda^*, \nu^*)$ over $x$. (Note: $x^*$ may not be the unique minimizer of $L(x, \lambda^*, \nu^*)$.)

2. Since $\sum_{i=1}^{p} \nu_i^* h_i(x^*) = 0$ and each term in the sum $\sum_{i=1} \lambda_i^* f_i(x^*)$ is $\leq 0$, each must actually be 0. That is
$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \ldots, m.$$
This condition is known as **complementary slackness**, and it holds for any primal $x^*$ and any dual optimal $(\lambda^*, \nu^*)$ when strong duality holds. Roughly speaking, it means the $i$th optimal Lagrange multiplier is zero unless the $i$th constraint is active at the optimum.

### 5.3.2 KKT optimality conditions for convex problems (BV 5.5.3, p. 243)

Consider a standard form convex optimization problem for which $f_0, \ldots, f_m, h_1, \ldots, h_p$ are differentiable (and therefore have open domains). Let $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ be any points that satisfy the following **Karush-Kuhn-Tucker** (KKT) conditions:

1. Primal and dual feasibility: $f_i(\tilde{x}) \leq 0$, $h_i(\tilde{x}) = 0$, $\tilde{\lambda}_i \geq 0$ for all $i$.

2. Complementary slackness: $\tilde{\lambda}_i f_i(\tilde{x}) = 0$ for all $i$.

3. First order condition: $\nabla f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i \nabla h_i(\tilde{x}) = 0$.

Then $\tilde{x}$ and $\left(\tilde{\lambda}, \tilde{\nu}\right)$ are primal and dual optimal, respectively, with zero duality gap. To see this, note that the $\tilde{x}$ is primal feasible, and $L(x, \tilde{\lambda}, \tilde{\nu})$ is convex in $x$, since $\tilde{\lambda}_i \geq 0$. Thus the first order condition implies that $\tilde{x}$ minimizes $L(x, \tilde{\lambda}, \tilde{\nu})$ over $x$. So

$$\begin{aligned} g(\tilde{\lambda}, \tilde{\nu}) &= L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \\ &= f_0(\tilde{x}) + \sum_{i=1}^{m} \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^{p} \tilde{\nu}_i h_i(\tilde{x}) \\ &= f_0(\tilde{x}), \end{aligned}$$

where in the last line we use complementary slackness and $h_i(\tilde{x}) = 0$. Thus $\tilde{x}$ and $\left(\tilde{\lambda}, \tilde{\nu}\right)$ have zero duality gap, and therefore are primal and dual optimal.

In summary, for any *convex* optimization problem with differentiable objective and constraint functions, any points that satisfy the KKT conditions are primal and dual optimal and have zero duality gap.