

Convex Optimization Overview

Justin Pyron

July 28, 2021

This document primarily summarizes key elements of Stephen Boyd's *Convex Optimization*.

Contents

1	Convex Sets	2
2	Convex Functions	3
3	Convex Optimization Problems	7
3.1	Slack Variables	7
3.2	Epigraph Form	7
4	Duality	9
4.1	Lagrangian and Dual Function	9
4.2	KKT Conditions	10
5	Algorithms	13

1 Convex Sets

Definition: A set C is *convex* if for any $x_1, x_2 \in C$ and any $\theta \in [0, 1]$

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

Definition: A *hyperplane* is a set of the form

$$\{x : a^T x = b\}$$

where $a \in \mathbb{R}^n, a \neq 0$, and $b \in \mathbb{R}$.

Definition: A *halfspace* is a set of the form

$$\{x : a^T x \leq b\}$$

where $a \in \mathbb{R}^n, a \neq 0$, and $b \in \mathbb{R}$.

Fact: The intersection of two convex sets is convex, i.e. if C_1, C_2 are convex, then $C_1 \cap C_2$ is convex.

Proof: Let $x_1, x_2 \in C_1 \cap C_2$, and let $\theta \in [0, 1]$. Since $x_1, x_2 \in C_1$ and C_1 is convex, $\theta x_1 + (1 - \theta)x_2 \in C_1$. Similarly, $\theta x_1 + (1 - \theta)x_2 \in C_2$. Thus, $\theta x_1 + (1 - \theta)x_2 \in C_1 \cap C_2$.

Fact: Hyperplanes are convex sets.

Proof: Let $H = \{x : a^T x = b\}$ be a hyperplane. Let $\theta \in [0, 1]$ and $x_1, x_2 \in H$, meaning $a^T x_1 = a^T x_2 = b$. Then

$$a^T (\theta x_1 + (1 - \theta)x_2) = \theta a^T x_1 + (1 - \theta)a^T x_2 = \theta b + (1 - \theta)b = b$$

Thus, $\theta x_1 + (1 - \theta)x_2 \in H$.

Fact: Halfspaces are convex sets.

Proof: Let $H = \{x : a^T x \leq b\}$ be a halfspace. Let $\theta \in [0, 1]$ and $x_1, x_2 \in H$, meaning $a^T x_1 \leq b$ and $a^T x_2 \leq b$. Then

$$a^T (\theta x_1 + (1 - \theta)x_2) = \theta a^T x_1 + (1 - \theta)a^T x_2 \leq \theta b + (1 - \theta)b = b$$

Thus, $\theta x_1 + (1 - \theta)x_2 \in H$.

Fact: Hyperplanes have a nice geometric interpretation. They can be visualized as an *offset vector* x_0 plus all vectors orthogonal to a *normal vector* a . Let

$$H = \{x : a^T x = b\}$$

be a hyperplane. Let $x_0 \in H$ be *any* vector contained in H . Since $a^T x_0 = b$,

$$H = \{x : a^T x - b = 0\} = \{x : a^T x - a^T x_0 = 0\} = \{x : a^T (x - x_0) = 0\}$$

So, if y is any vector orthogonal to a , then $x_0 + y \in H$ since $a^T ((x_0 + y) - x_0) = a^T y = 0$.

2 Convex Functions

Definition: A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is *convex* if the domain of f is a convex set and if for all x, y in the domain of f and $\theta \in [0, 1]$

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

Definition: A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is *concave* if $-f$ is convex.

Fact: (First-order conditions) Suppose f is differentiable. Then f is convex if and only if the domain of f is a convex set and

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

for all x, y in the domain of f .

Proof:

We'll first prove the result for the case $n = 1$ where $f : \mathbb{R} \rightarrow \mathbb{R}$.

Case $n = 1$:

We want to show

$$f \text{ is convex} \iff f(y) \geq f(x) + f'(x)(y - x) \quad \forall x, y \in \text{domain}(f)$$

(\implies direction)

Let $x, y \in \text{domain}(f)$, and let $t \in [0, 1]$. Since f is convex,

$$f(ty + (1 - t)x) \leq tf(y) + (1 - t)f(x)$$

Rearranging and dividing by t gives

$$f(y) \geq f(x) + \frac{f(ty + (1 - t)x) - f(x)}{t} = f(x) + \frac{f(x + t(y - x)) - f(x)}{t}$$

Taking the limit as $t \rightarrow 0$ gives

$$f(y) \geq f(x) + f'(x)(y - x)$$

Note that the above limit corresponds to the derivative of the composition $f \circ g$ at 0, where $g : \mathbb{R} \rightarrow \mathbb{R}$ maps $t \mapsto x + t(y - x)$ (with x, y fixed). The chain rules gives

$$\frac{d}{dt}(f \circ g)(0) = f'(g(t))g'(t)\Big|_{t=0} = f'(x + t(y - x))(y - x)\Big|_{t=0} = f'(x)(y - x)$$

(\impliedby direction)

Let $x, y \in \text{domain}(f)$ and $\theta \in [0, 1]$. Let $z = \theta x + (1 - \theta)y$. Then

$$f(x) \geq f(z) + f'(z)(x - z) \quad \text{and} \quad f(y) \geq f(z) + f'(z)(y - z)$$

Multiplying the first term by θ and the second term by $1 - \theta$ and adding gives

$$\begin{aligned} \theta f(x) + (1 - \theta)f(y) &\geq \theta f(z) + \theta f'(z)(x - z) + (1 - \theta)f(z) + (1 - \theta)f'(z)(y - z) \\ &= f(z) + f'(z)(\theta x + (1 - \theta)y - \theta z - (1 - \theta)z) \\ &= f(z) + f'(z)(\theta x + (1 - \theta)y - z) \\ &= f(z) \\ &= f(\theta x + (1 - \theta)y) \end{aligned}$$

General Case:

Now consider the general case where $f : \mathbb{R}^n \rightarrow \mathbb{R}$. We'll make use of the fact that

$$f \text{ is convex} \iff \text{The restriction of } f \text{ to any line segment is convex}$$

Assume $x, y \in \text{domain}(f)$. Let g be f restricted to the line segment between x, y , i.e. $g : \mathbb{R} \rightarrow \mathbb{R}$ maps $t \mapsto f(x + t(y - x))$. Note that the chain rule gives $g'(t) = \nabla f(x + t(y - x))^T (y - x)$.

(\implies direction)

Assume f is convex, which implies that g is convex. From the one-dimensional proof above, we have that

$$g(1) \geq g(0) + g'(0)(1 - 0)$$

which corresponds to

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

(\impliedby direction)

Assume the inequality holds. Then for any t_1, t_2

$$\begin{aligned} f(x + t_1(y - x)) &\geq f(x + t_2(y - x)) + \nabla f(x + t_2(y - x))^T [(x + t_1(y - x)) - (x + t_2(y - x))] \\ &= f(x + t_2(y - x)) + \nabla f(x + t_2(y - x))^T (y - x)(t_1 - t_2) \end{aligned}$$

which corresponds to

$$g(t_1) \geq g(t_2) + g'(t_2)(t_1 - t_2)$$

From the one-dimensional result, g is thus convex, implying f is convex. ■

This is a profound result since it says that the first-order approximation of a convex function is a global underestimator.

Let's remind ourselves what a first-order approximation is. The derivative is a tool used to approximate how changes in the input of a function will impact changes in the output. As changes in the input become infinitesimally small, the approximation becomes arbitrarily good. A first-order approximation makes the approximation using the derivative.

A first-order approximation for a simple one-dimensional function $g : \mathbb{R} \rightarrow \mathbb{R}$ at point x based on information known at point x_0 would look like $g \approx g(x_0) + g'(x_0)(x - x_0)$. A first-order approximation for a function $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$ at point x based on information known at point x_0 would look like $g \approx g(x_0) + J(x_0)(x - x_0)$, where $J(x_0) \in \mathbb{R}^{m \times n}$ is the derivative, i.e. Jacobian, of g at x_0 .

Fact: (Second-order conditions) Suppose f is twice-differentiable. Then f is convex if and only if the domain of f is a convex set and its Hessian is positive semi-definite.

This is also a profound result since it provides a fairly simple way to determine if a twice-differentiable function is convex.

Definition: The α -sublevel set of the function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is the set

$$C_\alpha = \{x \in \text{domain}(f) : f(x) \leq \alpha\}$$

Fact: α -sublevel sets of convex functions are convex sets.

Proof:

Let f be a convex function, and let x_1, x_2 be points in the α -sublevel sets of f , and let $\theta \in [0, 1]$. Then

$$f(\theta x_1 + (1 - \theta)x_2) \leq \theta f(x_1) + (1 - \theta)f(x_2) \leq \theta\alpha + (1 - \theta)\alpha = \alpha$$

Definition: The *epigraph* of a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is the set

$$\text{epigraph}(f) = \{(x, t) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq t\}$$

Fact: A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex if and only if its epigraph is a convex set.

Proof:

(\implies direction) Let f be a convex function, $\theta \in [0, 1]$, and $(x_1, t_1), (x_2, t_2) \in \text{epigraph}(f)$. Then

$$f(\theta x_1 + (1 - \theta)x_2) \leq \theta f(x_1) + (1 - \theta)f(x_2) \leq \theta t_1 + (1 - \theta)t_2$$

Thus, $(\theta x_1 + (1 - \theta)x_2, \theta t_1 + (1 - \theta)t_2) \in \text{epigraph}(f)$.

(\impliedby direction) Suppose the epigraph of function f is a convex set. Let $x_1, x_2 \in \text{domain}(f)$, and let $t_1 = f(x_1), t_2 = f(x_2)$. Since $(x_1, t_1), (x_2, t_2) \in \text{epigraph}(f)$ and $\text{epigraph}(f)$ is a convex set,

$$(\theta x_1 + (1 - \theta)x_2, \theta t_1 + (1 - \theta)t_2) \in \text{epigraph}(f),$$

meaning

$$f(\theta x_1 + (1 - \theta)x_2) \leq \theta t_1 + (1 - \theta)t_2 = \theta f(x_1) + (1 - \theta)f(x_2)$$

Thus, f is convex.

Fact: A non-negative weighted sum of convex functions is a convex function.

Proof:

Suppose f_1, \dots, f_k are convex functions, and $\alpha_1, \dots, \alpha_k \geq 0$. Let $f = \sum_{i=1}^k \alpha_i f_i$. Let $\theta \in [0, 1]$ and $x_1, x_2 \in \bigcap_{i=1}^k \text{domain}(f_i)$. Then,

$$f(\theta x_1 + (1 - \theta)x_2) = \sum_{i=1}^k \alpha_i f_i(\theta x_1 + (1 - \theta)x_2) \leq \sum_{i=1}^k \alpha_i \theta f_i(x_1) + \alpha_i (1 - \theta) f_i(x_2) = \theta f(x_1) + (1 - \theta)f(x_2)$$

Fact: The composition of a convex function with an affine function is a convex function. I.e. suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $A \in \mathbb{R}^{n \times m}$, $b \in \mathbb{R}^n$, and function $g : \mathbb{R}^m \rightarrow \mathbb{R}$ is defined by $g(x) = f(Ax + b)$. Then g is convex.

Proof:

Let $x_1, x_2 \in \text{domain}(g) \subseteq \mathbb{R}^m$, which implies that $Ax_1 + b, Ax_2 + b \in \text{domain}(f)$. Let $\theta \in [0, 1]$. Then

$$\begin{aligned} g(\theta x_1 + (1 - \theta)x_2) &= f(A(\theta x_1 + (1 - \theta)x_2) + b) \\ &= f(\theta(Ax_1 + b) + (1 - \theta)(Ax_2 + b)) \\ &\leq \theta f(Ax_1 + b) + (1 - \theta)f(Ax_2 + b) \\ &= \theta g(x_1) + (1 - \theta)g(x_2) \end{aligned}$$

Fact: The pointwise supremum of a collection convex functions is a convex function. I.e. if \mathcal{A} is some index set, and f_α is convex for all $\alpha \in \mathcal{A}$, then the function g defined by

$$g(x) = \sup_{\alpha \in \mathcal{A}} f_\alpha(x)$$

is convex.

Proof:

Note first the general property that

$$\sup_{\alpha \in \mathcal{A}} [f_\alpha(x_1) + f_\alpha(x_2)] \leq \sup_{\alpha \in \mathcal{A}} f_\alpha(x_1) + \sup_{\alpha \in \mathcal{A}} f_\alpha(x_2)$$

Let $x_1, x_2 \in \text{domain}(g)$ and $\theta \in [0, 1]$. Then

$$\begin{aligned} g(\theta x_1 + (1 - \theta)x_2) &= \sup_{\alpha \in \mathcal{A}} f_{\alpha}(\theta x_1 + (1 - \theta)x_2) \\ &\leq \sup_{\alpha \in \mathcal{A}} [\theta f_{\alpha}(x_1) + (1 - \theta)f_{\alpha}(x_2)] \\ &\leq \theta \sup_{\alpha \in \mathcal{A}} f_{\alpha}(x_1) + (1 - \theta) \sup_{\alpha \in \mathcal{A}} f_{\alpha}(x_2) \\ &= \theta g(x_1) + (1 - \theta)g(x_2) \end{aligned}$$

A similar proof establishes that the pointwise infimum of a collection concave functions is a concave function.

3 Convex Optimization Problems

Definition: A *Convex Optimization Problem* is one of the form

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && a_i^T x = b, \quad i = 1, \dots, p \end{aligned} \tag{1}$$

where $f_i, i = 0, \dots, m$ are convex functions. Note that the objective function is convex, the inequality constraint functions are convex, and the equality constraint functions are affine.

3.1 Slack Variables

A convex optimization problem can be reformulated using *slack variables*. The equivalent formulation of the original problem using *slack variables* is

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && s_i(x) \geq 0, \quad i = 1, \dots, m \\ & && f_i(x) + s_i = 0, \quad i = 1, \dots, m \\ & && a_i^T x = b, \quad i = 1, \dots, p \end{aligned} \tag{2}$$

Note that the formulation that uses slack variable introduces m new variables. Solving the problem formulated with slack variables is sometimes easier than solving the original problem. The two facts below establish equivalence between solutions of problems (1) and (2).

Fact: If x is a solution to problem (1), then (x, s) is a solution to problem (2) where $s_i = -f_i(x), i = 1, \dots, m$.

Proof:

Suppose x is a solution to problem (1). Define s such that $s_i = -f_i(x), i = 1, \dots, m$. Feasibility of x for (1) means $f_i(x) \leq 0, i = 1, \dots, m$, which implies $s_i \geq 0, i = 1, \dots, m$, and thus (x, s) is feasible for (2).

Assume that (x, s) is not a solution for (2). Then there exists a feasible point (\tilde{x}, \tilde{s}) with $\tilde{x} \neq x$ such that $f_0(\tilde{x}) < f_0(x)$. Feasibility of (\tilde{x}, \tilde{s}) for (2) means that $f_i(\tilde{x}) = -\tilde{s}_i \leq 0, i = 1, \dots, m$, which implies feasibility of \tilde{x} for (1). But since $f_0(\tilde{x}) < f_0(x)$, x cannot be a solution to (1), which is a contradiction.

Fact: If (x, s) is a solution to problem (2), then x is a solution to problem (1).

Proof:

Suppose (x, s) is a solution to (2). Assume x is not a solution to (1). Then there exists a feasible $\tilde{x} \neq x$ such that $f_0(\tilde{x}) < f_0(x)$.

Define \tilde{s} such that $\tilde{s}_i = -f_i(\tilde{x}), i = 1, \dots, m$. Feasibility of \tilde{x} for (1) means $f_i(\tilde{x}) \leq 0, i = 1, \dots, m$, which implies $\tilde{s}_i \geq 0, i = 1, \dots, m$ and thus feasibility of (\tilde{x}, \tilde{s}) for (2). But since $f_0(\tilde{x}) < f_0(x)$, x cannot be a solution to (2), which is a contradiction.

3.2 Epigraph Form

A convex optimization problem can be reformulated in *epigraph form*. The equivalent formulation of the original problem in *epigraph form* is

$$\begin{aligned} & \text{minimize} && t \\ & \text{subject to} && f_0(x) - t \leq 0 \\ & && f_i(x) \leq 0, \quad i = 1, \dots, m \\ & && a_i^T x \leq b, \quad i = 1, \dots, p \end{aligned} \tag{3}$$

Solving the problem in epigraph form is sometimes easier than solving the original problem. The fact below establishes equivalence between solutions of problems (1) and (3).

Fact: (x, t) is a solution to problem (3) if and only if x is a solution to problem (1) and $t = f_0(x)$.

Proof:

(\implies direction)

Suppose (x, t) is a solution to (3).

Assume $t > f_0(x)$. But then (x, \tilde{t}) , where $\tilde{t} = f_0(x) + \frac{1}{2}(t - f_0(x))$, is feasible for (3) but with lower objective function value, which is a contradiction if (x, t) is a solution. Thus $t = f_0(x)$.

Now assume that x is *not* optimal for (1), i.e. that there exists $\tilde{x} \neq x$ feasible for (1) such that $f_0(\tilde{x}) < f_0(x)$. Let $\tilde{t} = f_0(\tilde{x})$. But then (\tilde{x}, \tilde{t}) is feasible for (3) and

$$\tilde{t} = f_0(\tilde{x}) < f_0(x) = t$$

which implies (x, t) is not optimal for (3), a contradiction.

(\impliedby direction)

Suppose x is a solution to (1) and $t = f_0(x)$. Since x is feasible for (1) and $f_0(x) - t \leq 0$, (x, t) is feasible for (3).

Now assume that (x, t) is *not* a solution to (3), i.e. that there exists $(\tilde{x}, \tilde{t}) \neq (x, t)$ feasible for (3) such that $\tilde{t} < t$. This implies

$$f_0(\tilde{x}) \leq \tilde{t} < t = f_0(x)$$

But since feasibility of (\tilde{x}, \tilde{t}) for (3) implies feasibility of \tilde{x} for (1), this implies x is not an optimal solution for (1), which is a contradiction.

4 Duality

In this section, we will be working with an optimization problem of the form

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0 \quad i = 1, \dots, m \\ & && h_i(x) = 0 \quad i = 1, \dots, p \end{aligned} \tag{4}$$

with optimization variable $x \in \mathbb{R}^n$. We make no assumption about convexity of the above problem, although if convexity holds, several nice properties follow. These properties form a cornerstone for much of convex optimization.

4.1 Lagrangian and Dual Function

Definition: The *Lagrangian* $L : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ associated with problem (4) is the function defined as

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

λ_i is referred to as the *ith Lagrange Multiplier* associated with the *ith* inequality constraint $f_i(x) \leq 0$. ν_i is referred to as the *ith Lagrange Multiplier* associated with the *ith* equality constraint $h_i(x) = 0$.

Definition: The *Lagrange dual function* $g : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ is the function defined as

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \inf_x \left[f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right]$$

Note that for fixed x , the function that maps

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

is an affine (and thus concave) function of (λ, ν) . Since the pointwise infimum of a collection of concave functions is a concave function (see section 2), the Lagrange Dual Function is concave. This is true even when the problem (4) is not convex.

Fact: The dual function can be used to establish a lower bound for the value of the objective function at any feasible point. More specifically, if x_0 is a feasible point for problem (4), then for any $\lambda \succeq 0$ and any ν , then

$$g(\lambda, \nu) \leq f_0(x_0)$$

Proof:

Suppose x_0 is a feasible point for problem (4). Then, for any $\lambda \succeq 0$ and any ν ,

$$\sum_{i=1}^m \lambda_i f_i(x_0) + \sum_{i=1}^p \nu_i h_i(x_0) \leq 0$$

since $f_i(x_0) \leq 0, i = 1, \dots, m$ and $h_i(x_0) = 0, i = 1, \dots, p$. It then follows that

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) \leq L(x_0, \lambda, \nu) = f_0(x_0) + \sum_{i=1}^m \lambda_i f_i(x_0) + \sum_{i=1}^p \nu_i h_i(x_0) \leq f_0(x_0)$$

■

From the fact above, the dual function establishes a lower bound for the objective function at feasible points. In particular, if x^* is a solution with corresponding value $p^* = f(x^*)$, then for any $\lambda \succeq 0$ and any ν ,

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

A natural question to ask is how tight the lower bound can be made, which motivates the following definition:

Definition: The *Lagrange dual problem* associated with problem (4) is

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

Since the dual function is concave, the above problem is a convex optimization problem. Suppose that (λ^*, ν^*) is an optimal solution to the Lagrange Dual Problem, with corresponding objective value $d^* = g(\lambda^*, \nu^*)$. An important fact is that

$$d^* \leq p^*$$

Definition: The difference $p^* - d^*$ is called the *duality gap* associated with problem (3). The property that $d^* \leq p^*$ is referred to as *weak duality*. When the duality gap is zero, i.e. $d^* = p^*$, then *strong duality* holds.

Fact: Consider the optimization problem

$$\begin{aligned} & \text{minimize} && f_0(x) \\ & \text{subject to} && f_i(x) \leq 0 \quad i = 1, \dots, m \\ & && Ax = b \end{aligned}$$

If the problem is convex, and there exists $x \in \text{relint}(\mathcal{D})$ (the relative interior of the domain of the problem) such that

$$f_i(x) < 0, i = 1, \dots, m, \quad Ax = b$$

then strong duality holds. This condition is known as *Slater's Condition*. In essence, if the problem is convex and there exists a point that is *strictly feasible*, then strong duality holds.

A critical reason why convex optimization problems are tractable is because strong duality holds for most convex functions that are encountered in practice. For most problems, strong duality allows one to obtain an optimal solution by solving for an input that satisfies the KKT conditions (which will be defined shortly). There exist methods for determining whether a strictly feasible input exists.

4.2 KKT Conditions

Complimentary Slackness

Suppose that x^* is a primal optimal point and (λ^*, ν^*) is a dual optimal point, and that strong duality holds. Then,

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

It follows that the two inequalities in the above expression, in fact, hold with equality. The equality of the second and third lines states that x^* minimizes $L(x, \lambda^*, \nu^*)$ over x . The equality of the third and fourth

lines (along with the fact that $h_i(x^*) = 0, i = 1 \dots, p$ due to feasibility of x^*) states that

$$\sum_{i=1}^m \lambda_i^* f_i(x^*) = 0$$

Primal feasibility of x^* along with dual feasibility of λ^* means that $\lambda_i^* f_i(x^*) \leq 0, i = 1, \dots, m$. So each element in the sum above must be 0. In other words, it must be the case that

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

This condition is known as *complimentary slackness*. If x^* is a primal optimal point and (λ^*, ν^*) is a dual optimal point, complimentary slackness gives the following implications

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

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

KKT Conditions for Non-Convex Problems

We now assume that the functions $f_0, \dots, f_m, h_1, \dots, h_p$ are differentiable.

Suppose that x^* is a primal optimal point and (λ^*, ν^*) is a dual optimal point, and that strong duality holds. As discussed above, x^* minimizes $L(x, \lambda^*, \nu^*)$ over x , meaning that its gradient with respect to x vanishes at x^* :

$$\nabla L(x^*, \lambda^*, \nu^*) = \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*) = 0$$

This fact, along with complimentary slackness, primal feasibility of x^* , and dual feasibility of (λ^*, ν^*) gives the following set of conditions that hold

$$f_i(x^*) \leq 0, \quad i = 1, \dots, m$$

$$h_i(x^*) = 0, \quad i = 1, \dots, p$$

$$\lambda_i^* \geq 0, \quad i = 1, \dots, m$$

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

$$\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*) = 0$$

This set of conditions is called the *KKT conditions* (*Karush-Kuhn-Tucker conditions*). For any problem with differentiable objective and constraint functions for which strong duality holds, any set of primal and dual optimal points must satisfy the KKT conditions.

KKT Conditions for Convex Problems

We saw above that for a problem with differentiable objective and constraint functions for which strong duality holds that

$$x^* \text{ primal optimal, } (\lambda^*, \nu^*) \text{ dual optimal} \implies \text{KKT conditions hold for } x^*, (\lambda^*, \nu^*)$$

If the optimization problem is convex, a converse holds as well, which is shown below. But first, recall the following fact

Fact: Suppose f is a convex function and $\nabla f(x) = 0$. Then x is a minimal point of f .

Proof:

Since f is convex, the first-order condition

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

holds for all x, y (see section 2). If $\nabla f(x) = 0$, then $f(y) \geq f(x)$ for all y , and thus x is a minimal point.

Fact: Suppose that the problem

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

is convex, i.e. that f_0, \dots, f_m are convex and h_1, \dots, h_p are affine. If primal point x^* and dual point (λ^*, ν^*) satisfy the KKT conditions

$$\begin{aligned} f_i(x^*) &\leq 0, & i = 1, \dots, m \\ h_i(x^*) &= 0, & i = 1, \dots, p \\ \lambda_i^* &\geq 0, & i = 1, \dots, m \\ \lambda_i^* f_i(x^*) &= 0, & i = 1, \dots, m \end{aligned}$$

$$\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*) = 0$$

then x^* is primal optimal, (λ^*, ν^*) is dual optimal, and there is zero duality gap.

Proof:

The third KKT condition that $\lambda_i^* \geq 0, i = 1, \dots, m$, along with the fact that $f_0, \dots, f_m, h_1, \dots, h_p$ are convex, implies that

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

is convex in x , since a non-negative weighted sum of convex functions is a convex function (see section 2). The fifth KKT conditions gives that the gradient of $L(x, \lambda^*, \nu^*)$ with respect to x vanishes at x^* . Since the gradient of a convex function vanishing at a point implies that the point is minimal, it follows that x^* minimizes $L(x, \lambda^*, \nu^*)$. So,

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

where the fourth equality follows from the second and fourth KKT conditions.

5 Algorithms

Unconstrained Optimization

In the unconstrained setting, we solve the problem

$$\text{minimize } f_0(x)$$

where $x \in \mathbb{R}^n$ is the optimization variable.

Equality-Constrained Optimization

In the equality-constrained setting, we solve the problem

$$\begin{aligned} &\text{minimize } f_0(x) \\ &\text{subject to } Ax = b \end{aligned}$$

where $x \in \mathbb{R}^n$ is the optimization variable, $A \in \mathbb{R}^{p \times n}$ is assumed to have rank $p < n$, and function f_0 is convex.

Inequality-Constrained Optimization

We solve the problem

$$\begin{aligned} &\text{minimize } f_0(x) \\ &\text{subject to } f_i(x) \leq 0 \quad i = 1, \dots, m \\ &\quad Ax = b \end{aligned}$$

where $x \in \mathbb{R}^n$ is the optimization variable, $A \in \mathbb{R}^{p \times n}$ is assumed to have rank $p < n$, and all functions $f_i, i = 0, \dots, m$ are convex.