Optimality conditions

Bang-Shien Chen*

When it comes to optimality conditions, we often discuss well-known conditions such as the Karush-Kuhn-Tucker (KKT) conditions and Fritz John (FJ) conditions. However, it is helpful to start with a simpler case: unconstrained optimization problems, and derive its necessary and sufficient conditions for optimality.

1 Optimality Conditions for Unconstrained Optimization

In this section, we consider the optimization problem where $X \subseteq \mathbb{R}^n$:

$$\min_x f(x) \; ; \quad x \in X. \tag{1.1}$$

In addition, we often make some basic assumptions:

- f is continuously differentiable (for first-order condition).
- f is twice continuously differentiable (for second-order condition).

The definition of local and global minimum:

- x^* is a *local minimum* if there exists an $\varepsilon > 0$ such that $f(x^*) \leq f(x)$, for all x with $\|x - x^*\| < \varepsilon$.
- x^* is a *global minimum* if $f(x^*) \leq f(x)$, for all $x \in X$.

We say a local or global minimum is *strict* if the corresponding inequality is strict for $x \neq x^*$.

1.1 Necessary Conditions for Optimality

If the objective function is differentiable, we can use gradients and Taylor approximations to compare the function value x^* with the cost of its close neighbors.

Theorem 1.1: First-order Necessary Condition

Let x^* be a local minimum of Problem 1.1, and suppose that f is continuously differentiable at x^*. Then

$$\nabla f(x^*) = 0.$$

Proof. Given some direction $d \in \mathbb{R}^n$ and scalar α, we have the first-order approximation

$$f(x^* + \alpha d) \approx f(x^*) + \nabla f(x^*)^\top((x^* + \alpha d) - x^*).$$

Since x^* is a local minimum, we have the inequality

$$f(x^*) \leq f(x^*) + \nabla f(x^*)^\top(\alpha d).$$

For small $\alpha > 0$ and $\alpha < 0$, we have $d^\top \nabla f(x^*) \geq 0$ for all d and $d^\top \nabla f(x^*) \leq 0$ for all d, respectively. Together, $d^\top \nabla f(x^*) = 0$ for all d, thus $\nabla f(x^*) = 0$. □

* https://dgbshien.com/

Similar to the first-order condition, we could also derive a second-order condition using second-order approximation.

Theorem 1.2: Second-order Necessary Condition

Let x^* be a local minimum of Problem 1.1, and suppose that f is twice continuously differentiable at x^*. Then

$$\nabla^2 f(x^*) \succeq 0.$$

Proof. Given some direction $d \in \mathbb{R}^n$ and scalar α, we have the second-order approximation

$$f(x^* + \alpha d) \approx f(x^*) + \alpha \nabla f(x^*)^\top d + \frac{\alpha^2}{2} d^\top \nabla^2 f(x^*) d.$$

Since x^* is a local minimum, we have the inequality

$$f(x^*) \leq f(x^*) + \alpha \nabla f(x^*)^\top d + \frac{\alpha^2}{2} d^\top \nabla^2 f(x^*) d.$$

By Theorem 1.1 and $\alpha^2 > 0$, we have $d^\top \nabla^2 f(x^*) d \geq 0$ for all d, i.e., $\nabla^2 f(x^*) \succeq 0$. $\square$

1.2 Sufficient Conditions for Optimality

Theorem 1.3: Second-order Sufficient Condition

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice continuously differentiable. Suppose that a vector x^* satisfies the conditions

$$\nabla f(x^*) = 0, \quad \nabla^2 f(x^*) \succ 0.$$

Then x^* is a strict local minimum of f. In particular, there exists scalars $\gamma > 0$ and $\varepsilon > 0$ such that

$$f(x) \geq f(x^*) + \frac{\gamma}{2} \|x - x^*\|^2 \ \forall \ x \text{ with } \|x - x^*\| \leq \varepsilon.$$

Proof. Given some direction $d \in \mathbb{R}^n$ and by the second-order approximation the first condition,

$$\begin{aligned} f(x^* + d) &= f(x^*) + \nabla f(x^*)^\top d + \frac{1}{2} d^\top \nabla^2 f(x^*) d + \mathcal{O}(\|d\|^2) \\ &= f(x^*) + \frac{1}{2} d^\top \nabla^2 f(x^*) d + \mathcal{O}(\|d\|^2). \end{aligned}$$

Since $\nabla^2 f(x^*)$ is real symmetric, we have the spectral decomposition $\nabla^2 f(x^*) = Q \Lambda Q^\top$, where Λ is a diagonal matrix of eigenvalues and Q is orthogonal whose columns are the corresponding eigenvectors. Let $z = Q^\top d$, we have $\|z\| = \|Q^\top d\| = \|d\|$ since Q is orthogonal and

$$d^\top \nabla^2 f(x^*) d = d^\top Q \Lambda Q^\top d = z^\top \Lambda z = \sum_{i=1}^n \lambda_i z_i^2 \geq \lambda_{\min} \sum_{i=1}^n z_i^2 = \lambda_{\min} \|z\|^2 = \lambda_{\min} \|d\|^2.$$

Substituting this inequality back to the second-order approximation, we have

$$\begin{aligned} f(x^* + d) &= f(x^*) + \frac{1}{2} d^\top \nabla^2 f(x^*) d + \mathcal{O}(\|d\|^2) \\ &\geq f(x^*) + \frac{\lambda_{\min}}{2} \|d\|^2 + \mathcal{O}(\|d\|^2) \\ &= f(x^*) + \left(\frac{\lambda_{\min}}{2} + \frac{\mathcal{O}(\|d\|^2)}{\|d\|^2} \right) \|d\|^2 \end{aligned}$$

Therefore, take $\gamma \leq \lambda_{\min} + 2\mathcal{O}(\|d\|^2)/\|d\|^2$ and $\|d\| = \|x - x^*\| \leq \varepsilon$, the proof is done. $\square$

1.3 Existence of Optimal Solutions

In general, a minimum need not exist. For example, $f(x) = x$ and $f(x) = e^x$ have no global minimum. To derive existence, we first review some definitions. Let $X \subseteq \mathbb{R}^n$,

- f is continuous at x if $\lim_{y \to x} f(y) = x$.
- f is continuous on X if f is continuous at every $x \in X$.
- $f : X \to \mathbb{R}$ is lower semi-continuous at $x \in X$ if $f(x) \leq \liminf_{k\to\infty} f(x_k)$ for every sequence $\{x_k\}$ of X converging to x.
- $f : X \to \mathbb{R} \cap \{\pm\infty\}$ (extended valued function) is coercive if $\lim_{x\to\infty} f(x) = \infty$.

Proposition 1.4: Weierstrass' Theorem

Let $X \in \mathbb{R}^n$ be nonempty and let $f : \mathbb{R}^n \to \mathbb{R}$ be lower semi-continuous at all points of X. Suppose that one of the following conditions holds:

1. X is compact (close and bounded).
2. X is closed and f is coercive.
3. There exists a scalar γ such that the level set $\{x \in X \mid f(x) \leq \gamma\}$ is nonempty and compact.

Then there exists a vector $x^* \in X$ such that $f(x^*) = \inf_{x\in X} f(x)$, i.e., x^* is a global minimum.

This can be reduced to the well-known *Weierstrass Extreme Value Theorem*, which states that every continuous function on a nonempty compact set attains its extreme values on that set, including a global minimum. We next give an example of using optimality conditions to prove a well-known inequality.

Example 1.5: Arithmetic-Geometric Mean Inequality

Show the AM-GM inequality

$$\sqrt[n]{x_1 x_2 \cdots x_n} \leq \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

Let $y_i = \ln(x_i)$ for all i, we can rewrite the inequality as

$$e^{y_1} + e^{y_2} + \cdots + e^{y_n} \geq ne^{\frac{y_1+y_2+\cdots+y_n}{n}}.$$

Let $y_1 + y_2 + \cdots y_n = s$ and consider the optimization problem

$$\begin{aligned} \min_{y_i}\ & e^{y_1} + e^{y_2} + \cdots + e^{y_n} \\ \text{s.t. } & y_1 + y_2 + \cdots y_n = s, \end{aligned}$$

we aim to show the optimal value is $ne^{s/n}$. We rewrite an equivalent unconstrained problem

$$\min_{y_i} f(y_1, y_2, \ldots, y_{n-1}) = e^{y_1} + \cdots + e^{y_{n-1}} + e^{s-y_1-\cdots-y_{n-1}}.$$

Note that since f is coercive, by Proposition 1.4, there exists a global minimum. Let $(y_1^*, y_2^*, \ldots, y_{n-1}^*)$ be the global minimum, by Theorem 1.1, we have

$$\frac{\partial f}{\partial y_i} = e^{y_i} + e^{s-y_1-\cdots-y_{n-1}}(-1) = 0 \text{ for } i = 1, \ldots, n-1,$$

which implies $y_i^* = s - y_1 - \cdots - y_{n-1}$ for $i = 1, \ldots, n-1$. The system has only one solution: $y_i = s/n$ for all i, which is also the unique global minimum. Also, $e^{y_1^*} + \cdots + e^{y_n^*} = ne^{s/n}$.

2 Lagrange Duality

In the previous section, we derived optimality conditions for unconstrained problems. We will next show the optimality conditions for constrained problems, such as the Fritz John conditions and Karush-Kuhn-Tucker conditions. However, since it is easier to understand these conditions with a basic knowledge of Lagrange duality, we first give a brief introduction to Lagrange duality.

2.1 Lagrange Dual Problem

Consider the optimization problem

$$\begin{aligned} \min_x \ & f_0(x) \\ \text{s.t.} \ & f_i(x) \leq 0, \ i = 1, \ldots, m \\ & h_i(x) = 0, \ i = 1, \ldots, l. \end{aligned} \tag{2.1}$$

with domain $\mathcal{D} = \bigcap_{i=0}^m \textbf{dom } f_i \ \cap \ \bigcap_{i=1}^p \textbf{dom } h_i$, and denote the optimal value by p^*. The idea of Lagrangian duality is to leverage the constraints in Problem (2.1) into the objective function, and thereby transforms a constrained problem to a unconstrained problem. We define the *Lagrangian* $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l \to \mathbb{R}$ by

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

where λ and ν are called the *Lagrange multipliers* or dual variables. We then define the (Lagrange) *dual function* $g : \mathbb{R}^m \times \mathbb{R}^l \to \mathbb{R}$ by

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

Note that the dual function is the point-wise infimum of a family of affine functions of (λ, ν), which is always concave even if Problem (2.1) is not convex. Now suppose that x^* is an optimal solution, for any $\lambda \geq 0$ and any ν, we have

$$\mathcal{L}(x^*, \lambda, \nu) = f_0(x^*) + \underbrace{\sum_{i=1}^m \lambda_i f_i(x^*)}_{\leq 0} + \underbrace{\sum_{i=1}^l \lambda_i h_i(x^*)}_{=0} \leq f_0(x^*),$$

which implies

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \mathcal{L}(x, \lambda, \nu) \leq \inf_{x \in \mathcal{D}} \mathcal{L}(x^*, \lambda, \nu) \leq f_0(x^*) = p^*. \tag{2.2}$$

That is, the optimal value p^* is an upper bound of the dual function g, which is also the main idea of *weak duality*. Also, the optimal value p^* is a lower bound of the objective function f_0. Figure 2.1 illustrates this property.

Example 2.1

Consider the optimization problem

$$\begin{aligned}\min_x \ & x^3 + 2x^2 - x + 1 \\ \text{s.t. } & x^2 \le 1 \iff 0 \le x \le 1.\end{aligned}$$

The dual function is

$$g(\lambda) = \inf_x \left(x^3 + 2x^2 - x + 1 + \lambda(x^2 - 1)\right).$$

We can observe from Figure 2.1 that the optimal value is the lower bound of the objective function, and also is the upper bound of the dual function.

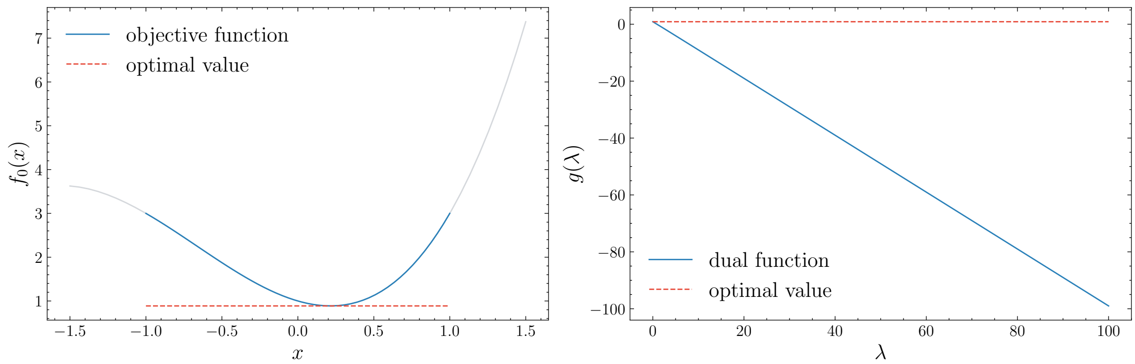

Figure 2.1: Lower and upper bound in Example 2.1.

Since the optimal value is an upper bound of the dual function, the (Lagrange) *dual problem* aims to maximize the dual function:

$$\begin{aligned}\max_{\lambda,\nu} \ & g(\lambda, \nu) \\ \text{s.t. } & \lambda \succeq 0.\end{aligned} \tag{2.3}$$

The dual problem is always convex whether the primal problem is convex or not.

2.2 Weak Duality and Strong Duality

Denote d^* as the optimal value of the dual problem. Since the optimal value of the primal problem p^* is an upper bound of the dual function, we have the *weak duality*:

$$d^* \le p^*.$$

We define $p^* - d^*$ as the duality gap. We say that *strong duality* holds if

$$d^* = p^*.$$

While strong duality does not hold in general, there are some results that establish conditions on the problem under which strong duality holds. These conditions are called *constraint qualifications*. Here we give a simple and widely used constraint qualification in the context of convex optimization, the *Slater's condition*: There exists an $x \in \mathbf{relint}\mathcal{D}$ such that $f_i(x) < 0$ for all $i = 1, \dots, m$ and $h_i(x) = 0$ for all $i = 1, \dots, l$, i.e., there exists a *strictly feasible* point in the relative interior of the feasible set.

2.3 Complementary Slackness

Suppose that the primal optimal value p^* and dual optimal value d^* are attained where x^* is the primal optimal solution and (λ^*, ν^*) is the dual optimal solution. This means that

$$g(\lambda^*, \nu^*) \leq f_0(x^*) + \underbrace{\sum_{i=1}^{m} \lambda_i^* f_i(x^*)}_{\leq 0} + \underbrace{\sum_{i=1}^{l} \nu_i^* h_i(x^*)}_{=0} \leq f_0(x^*).$$

The first inequality is because infimum is a lower bound, and the second inequality is because x^* is feasible such that $f_i(x^*) \leq 0$ for $i = 1, \ldots, m$ and $h(x^*) = 0$ for $i = 1, \ldots, l$. Now suppose that strong duality holds, i.e., $g(\lambda^*, \nu^*) = f_0(x^*)$, we have that the two inequalities in the chain hold equality. This implies the *complementary slackness*:

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \ldots, m. \tag{2.4}$$

We say an inequality constraint is *binding* or *active* if $f_i(x^*) = 0$. An important result of complementary slackness is that the i-th inequality constraint is binding if λ_i^* is not zero, or λ_i^* is zero if the i-th inequality constraint is non-binding.

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

Roughly speaking, only the constraints that are binding at the optimal point have a direct impact on the optimal solution through their Lagrange multipliers, while constraints that are non-binding do not affect the solution, as their Lagrange multipliers are zero. Complementary slackness allows us to identify which constraints are active and further simplify an optimization problem, or identify which constraints are critical and provide insights for decision-making.

Another result of complementary slackness is that since

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

we can conclude that x^* minimizes $\mathcal{L}(x, \lambda^*, \nu^*)$. Therefore, by Theorem 1.1, it follows that

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{l} \nu_i^* \nabla h_i(x^*) = 0. \tag{2.5}$$

Equation (2.5) is also known as the *stationary condition*.

3 Fritz John Conditions

In this section, we state the Fritz John (FJ) necessary and sufficient conditions. These necessary conditions, including the FJ necessary conditions and KKT necessary conditions, basically involves a stationary condition, complementary slackness, primal feasibility, and dual feasibility. The intuition behind these conditions is quite straightforward, so we do not include the proofs here. Please refer to Bazaraa [1] for the detailed proofs.

On the other hand, these conditions rely on different assumptions, which involves continuity, differentiability, and convexity on objective functions, binding/non-binding inequality constraints, and equality constraints. We point out that in most textbooks, it is more common to assume that all functions are continuously differentiable. However, Bazaraa [1] generalizes these assumptions by distinguishing between the regularity requirements for different types of constraints. Thereby, we will focus on clearly stating these assumptions in this and the following section.

3.1 Fritz John Necessary Conditions

The FJ necessary conditions if for local minimums. We need some assumptions:

- Objective function and binding inequality constraints are differentiable.
- Non-binding inequality constraints are continuous.
- Equality constraints are continuously differentiable.

Theorem 3.1: Fritz John Necessary Conditions

Let x be a feasible solution and $I = \{i \mid f_i(x) = 0\}$ be the set of binding constraints. Suppose that f_0, f_i for $i \in I$ are differentiable at x, f_i for $i \notin I$ are continuous at x, and h_i for $i = 1, \ldots, l$ are continuously differentiable at x. If x is a local minimum, then there exists λ_0, λ_i for $i \in I$, and ν_i for $i = 1, \ldots, l$ such that

1. Stationary condition: $\lambda_0 \nabla f_0(x) + \sum_{i \in I} \lambda_i \nabla f_i(x) + \sum_{i=1}^{l} \nu_i \nabla h_i(x) = 0$.
2. Primal feasibility: $f_i(x) \leq 0$ for $i = 1, \ldots, m$ and $h_i(x) = 0$ for $i = 1, \ldots, l$.
3. Dual feasibility: $\lambda_0, \lambda_i \geq 0$ for $i \in I$.
4. Lagrange multipliers are not all zero: $(\lambda_0, \lambda, \nu) \neq 0$.

This can be reduced to a more used FJ necessary conditions [4] with complementary slackness by stronger assumptions:

- Objective function and inequality constraints are differentiable.
- Equality constraints are continuously differentiable.

Corollary 3.2

Let x be a feasible solution. Suppose that f_0, f_i for $i = 1, \ldots, m$ are differentiable at x, and h_i for $i = 1, \ldots, l$ are continuously differentiable at x. If x is a local minimum, then there exists λ_0, λ_i for $i = 1, \ldots, m$, and ν_i for $i = 1, \ldots, l$ such that

1. Stationary condition: $\lambda_0 \nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{l} \nu_i \nabla h_i(x) = 0$.
2. Complementary slackness: $\lambda_i f_i(x) = 0$ for $i = 1, \ldots, m$.
3. Primal feasibility: $f_i(x) \leq 0$ for $i = 1, \ldots, m$ and $h_i(x) = 0$ for $i = 1, \ldots, l$.
4. Dual feasibility: $\lambda_0, \lambda_i \geq 0$ for $i = 1, \ldots, m$.
5. Lagrange multipliers are not all zero: $(\lambda_0, \lambda, \nu) \neq 0$.

We say (x, λ, ν) is a FJ point if it satisfies the FJ necessary conditions.

3.2 FritzJohn Sufficient Conditions

The FJ sufficient conditions is also for local minimums. Again, we need some assumptions:

- Objective function is pseudoconvex.

- Binding inequality constraints are strictly pseudoconvex.
- Equality constraints are affine.
- Gradient of equality constraints are linear independent.

Theorem 3.3: Fritz John Sufficient Conditions

Let $I = \{i \mid f_i(x) = 0\}$ be the set of binding constraints and define $S = \{x \mid f_i(x) \leq 0$ for $i \in I, h_i(x) = 0$ for $i = 1, \ldots, l\}$. Suppose that x is a FJ point, h_i for $i = 1, \ldots, l$ are affine, ∇h_i for $i = 1, \ldots, l$ are linear independent. If there exists some neighborhood $N_\varepsilon(x)$ such that f_0 is pseudoconvex on $S \cap N_\varepsilon(x)$ and f_i for $i \in I$ are strictly pseudoconvex on $S \cap N_\varepsilon(x)$, then x is a local minimum.

4 Karush-Kuhn-Tucker Conditions

The FJ conditions provide a general set of necessary conditions for optimality, which does not require the non-negativity of the Lagrange multiplier associated with the objective function, i.e., $\lambda_0 \geq 0$. This makes FJ conditions more general but less restrictive. For example, if $\lambda_0 = 0$, the stationary condition no longer reflects a balance between the objective and the constraints. The main difference between the FJ conditions and the KKT conditions is that the Lagrange multiplier λ_0 cannot be zero, i.e., $\lambda_0 > 0$. Thus, the KKT conditions can be seen as a special case of the FJ conditions, where the presence of regularity in the problem ensures meaningful Lagrange multipliers. We can derive the KKT conditions from the FJ conditions if some constraint qualification holds.

4.1 Karush-Kuhn-Tucker Necessary Conditions

Under appropriate assumptions, including a valid constraint qualification, the KKT necessary conditions hold for a local minimum.

- Objective function and binding inequality constraints are differentiable.
- Non-binding constraints are continuous.
- Equality constraints are continuously differentiable.
- Constraint qualification.

Theorem 4.1: Karush-Kuhn-Tucker Necessary Conditions

Let x be a feasible solution and $I = \{i \mid f_i(x) = 0\}$ be the set of binding constraints. Suppose that f_0, f_i for $i \in I$ are differentiable at x, f_i for $i \notin I$ are continuous at x, and h_i for $i = 1, \ldots, l$ are continuously differentiable at x. Suppose some constraint qualification holds (e.g., $\nabla f_i(x)$ for $i \in I$ and $\nabla h_i(x)$ for $i = 1, \ldots, l$ are linear independent). If x is a local solution, then there exists *unique* λ_i for $i \in I, \nu_i$ for $i = 1, \ldots, l$ such that

1. Stationary condition: $\nabla f_0(x) + \sum_{i \in I} \lambda_i \nabla f_i(x) + \sum_{i=1}^{l} \nu_i \nabla h_i(x) = 0$.
2. Primal feasibility: $f_i(x) \leq 0$ for $i = 1, \ldots, m$ and $h_i(x) = 0$ for $i = 1, \ldots, l$.
3. Dual feasibility: $\lambda_i \geq 0$ for $i \in I$.

Similar to the FJ necessary conditions, this can be reduced to a more used KKT necessary conditions with complementary slackness by stronger assumptions:

- Objective function and inequality constraints are differentiable.
- Equality constraints are continuously differentiable.
- Constraint qualification.

Corollary 4.2

Let x be a feasible solution. Suppose that f_0, f_i for $i = 1, \ldots, m$ are differentiable at x, and h_i for $i = 1, \ldots, l$ are continuously differentiable at x. Suppose some constraint qualification holds (e.g., Slater's condition). If x is a local minimum, then there exists λ_0, λ_i for $i = 1, \ldots, m$, and ν_i for $i = 1, \ldots, l$ such that

1. Stationary condition: $\nabla f_0(x) + \sum_{i \in I} \lambda_i \nabla f_i(x) + \sum_{i=1}^{l} \nu_i \nabla h_i(x) = 0$.
2. Complementary slackness: $\lambda_i f_i(x) = 0$ for $i = 1, \ldots, m$.
3. Primal feasibility: $f_i(x) \leq 0$ for $i = 1, \ldots, m$ and $h_i(x) = 0$ for $i = 1, \ldots, l$.
4. Dual feasibility: $\lambda_i \geq 0$ for $i = 1 \ldots, m$.

We say (x, λ, ν) is a KKT point if it satisfies the KKT necessary conditions.

4.2 Karush-Kuhn-Tucker Sufficient Conditions

In a similar manner to the FJ sufficient conditions, the KKT sufficient conditions ensure global optimality when the problem satisfies specific convexity properties.

- Objective is pseudoconvex.
- Binding inequality constraints are quasiconvex.
- Equality constraints are quasiconvex/quasiconcave.

Theorem 4.3: Karush-Kuhn-Tucker Sufficient Conditions

Suppose that x is a KKT point, and let $I = \{i \mid f_i(x) = 0\}$ be the set of binding constraints. If f_0 is pseudoconvex, f_i for $i \in I$ are quasiconvex, and h_i is quasiconvex if $\nu_i > 0$ and quasiconcave if $\nu_i < 0$, then x is a global minimum.

For convex optimization problems, the KKT conditions are sufficient. Here is an example of solving a convex program with the KKT conditions.

Example 4.4

Consider the convex program

$$\begin{aligned} \min_{x} \ & f_0(x_1, x_2) = (x_1 - 1)^2 + (x_2 - 2)^2 \\ \text{s.t.} \ & f_1(x_1, x_2) = x_1 + 3x_2 - 1 \leq 0 \end{aligned}$$

We first check the Slater's condition. We have a feasible point $(0, 0)$ where $f_1(0, 0) = -1$ is strictly feasible (and both the objective and constraint are differentiable). Then by the

KKT conditions, there exists a unique λ such that stationary condition holds

$$\nabla f_0(x_1, x_2) + \lambda \nabla f_1(x_1, x_2) = 0 \implies 2x_1 - 2 + \lambda = 0, \quad 2x_2 - 4 + 3\lambda = 0,$$

complementary slackness holds

$$\lambda f_1(x_1, x_2) = 0 \implies \lambda(x_1 + 3x_2 - 1) = 0,$$

and primal/dual feasibility holds. We first solve the system

$$\begin{cases} 2x_1 - 2 + \lambda = 0 \\ 2x_2 - 4 + 3\lambda = 0 \\ \lambda(x_1 + 3x_2 - 1) = 0, \end{cases}$$

then check the primal/dual feasibility. Suppose that the constraint is binding, the system becomes

$$\begin{cases} 2x_1 - 2 + \lambda = 0 \\ 2x_2 - 4 + 3\lambda = 0 \\ x_1 + 3x_2 - 1 = 0 \end{cases} \implies (x_1, x_2, \lambda) = \left(\frac{2}{5}, \frac{1}{5}, \frac{6}{5}\right),$$

where (x_1, x_2, λ) indeed satisfies primal and dual feasibility.

5 Constraint Qualifications

The FJ necessary conditions are a general set of conditions that hold for a local minimum of constrained optimization problems. The KKT necessary conditions reduce the FJ conditions by requiring that the Lagrange multiplier associated to the objective function is non-negative, which requires a *constraint qualification* (regularity condition) must hold. We list some constraint qualifications:

- Slater's condition.
- Linearity constraint qualification: f_i for $i = 1, \ldots, m$ and h_i for $i = 1, \ldots, l$ are affine.
- Linear independence constraint qualification: $\nabla f_i(x)$ for $i \in I$ and $\nabla h_i(x)$ for $i = 1, \ldots, l$ are linear independent.

For more information about constraint qualifications, please refer to Bazaraa [1].

References

[1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, 2006.

[2] D. P. Bertsekas. *Nonlinear programming*. Athena Scientific, 1999.

[3] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[4] G. Giorgi. Remarks on fritz john conditions for problems with inequality and equality constraints. *International Journal of Pure and Applied Mathematics*, 71:643–657, 2011.