Lecture 11 - Duality in Convex Optimization

October 30, 2024

Happy Halloween - Part Two!

Typos c/o ChatGPT

Recall (Convex) Duality Framework

$$\begin{aligned} \text{minimize}_{x \in X} \ & f_0(x) \\ \text{subject to} \ & f_i(x) \leq 0, \quad i = 1, \dots, m, \\ & h_j(x) = 0, \quad j = 1, \dots, s. \end{aligned}$$

With λ_i, ν_j denoting Lagrange multipliers for g_i and $h_j(x) = 0$, respectively,

Lagrangian is:

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

With $g(\lambda, \nu) := \inf_{x \in X} \mathcal{L}(x, \lambda, \nu)$, the dual problem becomes:

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

No sign constraints on ν!

QPs and QCQPs

Quadratic Programs

A **Quadratic Program (QP)** is an optimization problem of the form:

$$\min \frac{1}{2}x^\intercal P x + c^\intercal x$$
$$A_1 x = b_1$$
$$A_2 x \leq b_2$$

where $P = P^\intercal$.

QPs and QCQPs

Quadratic Programs

A **Quadratic Program (QP)** is an optimization problem of the form:

$$\begin{aligned} \min\ & \frac{1}{2}x^\intercal P x + c^\intercal x \\ & A_1 x = b_1 \\ & A_2 x \leq b_2 \end{aligned}$$

where $P = P^\intercal$.

Quadratically Constrained Quadratic Programs

A **Quadratically Constrainted Quadratic Program (QCQP)** is a problem:

$$\begin{aligned} \min\ & \frac{1}{2}x^\intercal P_0 x + c^\intercal x \\ & x^\intercal P_i x + q_i^\intercal x + b_i \leq 0, i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

where $Q_i, i = 0, \ldots, m$ are **symmetric** matrices.

Convex if $P \succeq 0, P_i \succeq 0$. Gurobi can now handle **non-convex** QCQPs!

Two Problems to Warm Up

QP with Inequality Constraint

$$\underset{x}{\text{minimize}} \ \frac{1}{2}x^\intercal Qx + c^\intercal x$$
$$Ax \leq b$$

where $Q \succ 0$ is a **positive definite** matrix.

QCQP

$$\begin{aligned} \text{minimize} \ & \frac{1}{2}x^\intercal P_0 x + q_0^\intercal x + r_0 \\ \text{subject to} \ & \frac{1}{2}x^\intercal P_i x + q_i^\intercal x + r_i \leq 0, \quad i = 1, \ldots, m, \end{aligned}$$

where $P_0 \succ 0$ and $P_i \succeq 0$

- **What is the Lagrangian? What is the dual? Does Slater Condition hold?**

A Non-Convex QCQP

A Special Non-Convex QCQP

For $A = A^\intercal$ and $A \not\succeq 0$, consider:

$$\begin{aligned} \text{minimize } & x^\intercal A x + 2b^\intercal x \\ & x^\intercal x \leq 1 \end{aligned}$$

A Non-Convex QCQP

A Special Non-Convex QCQP

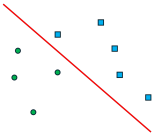
For $A = A^\intercal$ and $A \not\succeq 0$, consider:

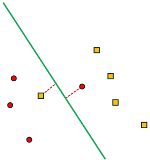
$$\begin{aligned} &\text{minimize } x^\intercal A x + 2b^\intercal x \\ &\qquad\quad x^\intercal x \leq 1 \end{aligned}$$

- Slater condition trivially satisfied!
- We actually have **zero duality gap**, $p^* = d^*$!
- A more general result: strong duality for any quadratic optimization problem with two constraints $\ell \leq x^\intercal P x \leq u$ if P and A are simultaneously diagonalizable

Regularized Support Vector Machines (SVM)

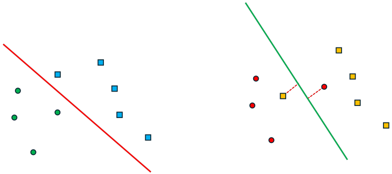
- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.





Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.

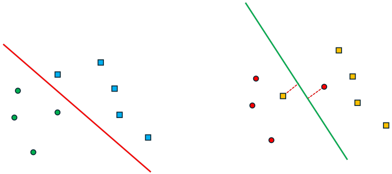


- Separable by hyperplane $H(w, b) = \{x : w^\intercal x + b \leq 0\}$, where $0 \neq w \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$\text{if and only if } \begin{cases} w^\intercal x_i + b \geq 0 & y_i = +1 \\ w^\intercal x_i + b \leq 0 & y_i = -1 \end{cases}$$

Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



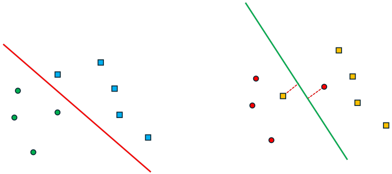
- Separable by hyperplane $H(w, b) = \{x : w^\intercal x + b \leq 0\}$, where $0 \neq w \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$\text{if and only if } \begin{cases} w^\intercal x_i + b \geq 0 & y_i = +1 \\ w^\intercal x_i + b \leq 0 & y_i = -1 \end{cases} \quad \Leftrightarrow \quad y_i(w^\intercal x_i + b) \geq 0, \; i = 1, \ldots, m.$$

- *How to solve this problem?*

Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



- Separable by hyperplane $H(w, b) = \{x : w^\intercal x + b \leq 0\}$, where $0 \neq w \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$\text{if and only if } \begin{cases} w^\intercal x_i + b \geq 0 & y_i = +1 \\ w^\intercal x_i + b \leq 0 & y_i = -1 \end{cases} \quad \Leftrightarrow \quad y_i(w^\intercal x_i + b) \geq 0, \ \ i = 1, \ldots, m.$$

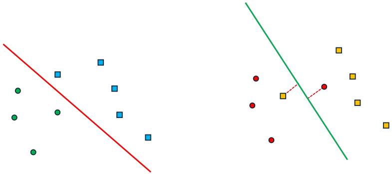
- *How to solve this problem?* This is **an LP**!

- In practice, non-separable. Find hyperplane minimizing total classification errors:

$$\sum_{i=1}^{m} \psi(y_i(w^\intercal x_i + b)), \ \text{where } \psi(t) = 1 \text{ if } t < 0 \text{ and } 0 \text{ otherwise.}$$

Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



- Separable by hyperplane $H(w, b) = \{x : w^\intercal x + b \leq 0\}$, where $0 \neq w \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$\text{if and only if } \begin{cases} w^\intercal x_i + b \geq 0 & y_i = +1 \\ w^\intercal x_i + b \leq 0 & y_i = -1 \end{cases} \quad \Leftrightarrow \quad y_i(w^\intercal x_i + b) \geq 0, \; i = 1, \dots, m.$$

- *How to solve this problem?* This is **an LP**!

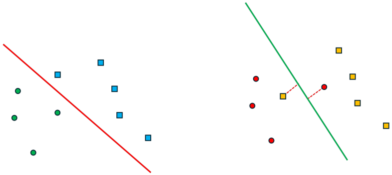
- In practice, non-separable. Find hyperplane minimizing total classification errors:

$$\sum_{i=1}^{m} \psi(y_i(w^\intercal x_i + b)), \text{ where } \psi(t) = 1 \text{ if } t < 0 \text{ and } 0 \text{ otherwise.}$$

- **Hard (MIP) problem!**

Regularized Support Vector Machines (SVM)

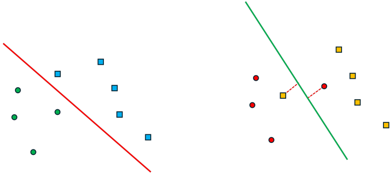
- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



- Separable if and only if $y_i(w^\intercal x_i + b) \geq 0, \; i = 1, \dots, m$.
- Minimize $\sum_{i=1}^{m} \psi(y_i(w^\intercal x_i + b))$, where $\psi(t) = 1$ if $t < 0$ and 0 : **hard MIP!**
- Replace $\psi(t)$ with upper bound $h(t) = (1 - t)_+ = \max(0, 1 - t)$ (**hinge function**)

Regularized Support Vector Machines (SVM)

- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.



- Separable if and only if $y_i(w^\intercal x_i + b) \geq 0, \; i = 1, \ldots, m$.

- Minimize $\sum_{i=1}^m \psi(y_i(w^\intercal x_i + b))$, where $\psi(t) = 1$ if $t < 0$ and 0 : **hard MIP!**

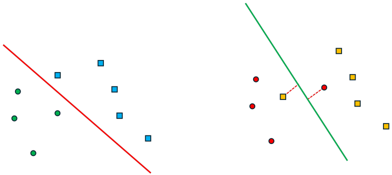
- Replace $\psi(t)$ with upper bound $h(t) = (1 - t)_+ = \max(0, 1 - t)$ (**hinge function**)

- Solve **regularized** version:

$$\min_{w,b} \; C \cdot \sum_{i=1}^m (1 - y_i(w^\intercal x_i + b))_+ + \frac{1}{2}\|w\|_2^2,$$

where parameter $C > 0$ controls trade-off between robustness and performance

Regularized Support Vector Machines (SVM)



- Given m data points $x_i \in \mathbb{R}^n$, each associated with a label $y_i \in \{-1, 1\}$, find a hyperplane that separates, as much as possible, the two classes.

- Solve $\min_{w,b} \; C \cdot \sum_{i=1}^{m} (1 - y_i(w^\top x_i + b))_+ + \frac{1}{2}\|w\|_2^2$

- Can be written as a QP by introducing slack variables:

$$\min_{w,b,v} \; \frac{1}{2}\|w\|_2^2 + C\sum_{i=1}^{m} v_i \quad : \quad v \geq 0, \; y_i(w^\top x_i + b) \geq 1 - v_i, \; i = 1, \ldots, m,$$

or more compactly:

$$\min_{w,b,v} \; \frac{1}{2}\|w\|_2^2 + C \cdot 1^\top v \quad : \quad v \geq 0, \; v + Z^\top w + by \geq 1,$$

where $Z^\top \in \mathbb{R}^{m \times n}$ is the matrix with rows given by $y_i \cdot x_i^\top$

- **What is the Lagrangian? What is the dual? Does Slater Condition hold?**

Regularized Support Vector Machines (SVM)

- Solve

$$\min_{w,b,v} \frac{1}{2}\|w\|_2^2 + C \cdot 1^\mathsf{T} v \quad : \quad v \geq 0,\ v + Z^\mathsf{T} w + by \geq 1,$$

  where $Z^\mathsf{T} \in \mathbb{R}^{m\times n}$ is the matrix with rows given by $y_i \cdot x_i^\mathsf{T}$

- $\mathcal{L}(w, b, \lambda, \mu) = \frac{1}{2}\|w\|_2^2 + C \cdot v^\mathsf{T} 1 + \lambda^\mathsf{T}(1 - v - Z^\mathsf{T} w - by) - \mu^\mathsf{T} v$
- $g(\lambda, \mu) = \min_{w,b} \mathcal{L}(w, b, \lambda, \mu)$
- Taking gradients : $w(\lambda, \mu) = Z\lambda,\ C \cdot 1 = \lambda + \mu,\ \lambda^\mathsf{T} y = 0$
- We obtain

$$g(\lambda, \mu) = \begin{cases} \lambda^\mathsf{T} 1 - \frac{1}{2}\|Z\lambda\|_2^2 & \text{if } \lambda^\mathsf{T} y = 0,\ \lambda + \mu = C \cdot 1, \\ +\infty & \text{otherwise.} \end{cases}$$

- Dual problem

$$d^\star = \max_\lambda \Big\{ \lambda^\mathsf{T} 1 - \frac{1}{2}\lambda^\mathsf{T} Z^\mathsf{T} Z \lambda \quad : \quad 0 \leq \lambda \leq C \cdot 1,\ \lambda^\mathsf{T} y = 0 \Big\}.$$

- Strong duality holds, because the primal problem is a QP
- Dual objective depends only on the **kernel matrix** $K = Z^\mathsf{T} Z \in S_+^m$, and dual problem involves only m variables and $m + 1$ constraints
- Only dependence on the number of dimensions (features) n is through Z; this requires all products $x_i^\mathsf{T} x_j$, $1 \leq i \leq j \leq m$ but still more memory-efficient than solving the primal!

Saddle Point Theory

Primal Problem

$$(\mathscr{P}) \text{ minimize}_x \quad f_0(x) \qquad f_i(x) \leq 0, \quad i = 1, \ldots, m \qquad x \in X. \tag{1}$$

- There is a very insightful way to make the primal and dual look more "symmetric"
- Recall: Lagrangian $\mathcal{L}(x, \lambda)$ and dual objective $g(\lambda) := \inf_{x \in X} \mathcal{L}(x, \lambda)$.
- **Claim:**

$$\sup_{\lambda \geq 0} L(x, \lambda) = \sup_{\lambda \geq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right) =$$

Saddle Point Theory

Primal Problem

$$
\begin{aligned}
(\mathscr{P})\ \text{minimize}_x \quad & f_0(x) \\
& f_i(x) \le 0, \quad i = 1, \ldots, m \\
& x \in X.
\end{aligned} \tag{1}
$$

- There is a very insightful way to make the primal and dual look more "symmetric"
- Recall: Lagrangian $\mathcal{L}(x, \lambda)$ and dual objective $g(\lambda) := \inf_{x \in X} \mathcal{L}(x, \lambda)$.
- **Claim:**

$$
\sup_{\lambda \ge 0} L(x, \lambda) = \sup_{\lambda \ge 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right) = \begin{cases} f_0(x) & \text{if } f_i(x) \le 0,\ i = 1, \ldots, m, \\ \infty & \text{otherwise.} \end{cases}
$$

- So we can express **the optimal values of the primal and dual** as:

$$
p^\star = \inf_{x \in X} \sup_{\lambda \ge 0} L(x, \lambda) \qquad\qquad d^\star = \sup_{\lambda \ge 0} \inf_{x \in X} L(x, \lambda)
$$

Saddle Point Theory

Alternative Formulation of Primal and Dual Problems

We can express **the optimal values of the primal and dual** as:

$$p^\star = \inf_{x \in X} \sup_{\lambda \geq 0} L(x, \lambda) \qquad\qquad d^\star = \sup_{\lambda \geq 0} \inf_{x \in X} L(x, \lambda)$$

- How to restate **weak duality** and **strong duality** in terms of the problems above?
- **Weak duality**:

$$\sup_{\lambda \geq 0} \inf_{x \in X} L(x, \lambda) \leq \inf_{x \in X} \sup_{\lambda \geq 0} L(x, \lambda)$$

- **Strong duality**:

$$\sup_{\lambda \geq 0} \inf_{x \in X} L(x, \lambda) = \inf_{x \in X} \sup_{\lambda \geq 0} L(x, \lambda).$$

- Strong duality holds exactly when we can interchange the order of **min** and **max**

Min-Max and Max-Min Problems

Min-Max and Max-Min

Consider more broadly the pair of problems:

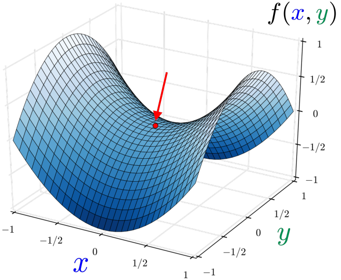
$$\sup_{y \in Y} \inf_{x \in X} f(x, y) \qquad \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

- For any f, Z, W, the **max-min inequality** (i.e., "weak duality") holds:

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) \mathit{leq} \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

- f, Z, W satisfy the **saddle-point property** if equality holds:

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) = \inf_{x \in X} \sup_{y \in Y} f(x, y).$$



Game Theoretic Interpretation

Min-Max and Max-Min

Consider more broadly the pair of problems:

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) \leq \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

- **Zero-sum game** between **player** x and **player** z
  - Player x **pays** player z the amount $f(x, z)$
  - x wants to **minimize** the amount, z wants to **maximize** it

Game Theoretic Interpretation

Min-Max and Max-Min

Consider more broadly the pair of problems:

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) \leq \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

- **Zero-sum game** between **player** x and **player** z
  - Player x **pays** player z the amount $f(x, z)$
  - x wants to **minimize** the amount, z wants to **maximize** it
- Min-max inequality: the player who moves second has an advantage!
  - x moves first and y moves second $\Rightarrow$ larger payment
  - y moves first and x moves second $\Rightarrow$ smaller payment

Game Theoretic Interpretation

Min-Max and Max-Min

Consider more broadly the pair of problems:

$$\sup_{y \in Y} \inf_{x \in X} f(x,y) \leq \inf_{x \in X} \sup_{y \in Y} f(x,y)$$

- **Zero-sum game** between **player** x and **player** z
  - Player x **pays** player z the amount $f(x,z)$
  - x wants to **minimize** the amount, z wants to **maximize** it

- Min-max inequality: the player who moves second has an advantage!
  - x moves first and y moves second $\Rightarrow$ larger payment
  - y moves first and x moves second $\Rightarrow$ smaller payment

- Player moving second has information about first player's move and can use a **strategy**, i.e., make a choice that depends on the first player's choice

$$\text{Left problem: } \inf_{x \in X} f(x,y) \text{ for any given } y \;\Rightarrow x^*(y)$$

$$\text{Right problem: } \inf_{y \in Y} f(x,y) \text{ for any given } x \;\Rightarrow y^*(x)$$

Existence of Saddle Points

Min-Max and Max-Min

Consider more broadly the pair of problems:

$$\sup_{y\in Y}\inf_{x\in X} f(x,y) = \inf_{x\in X}\sup_{y\in Y} f(x,y)$$

Saddle Point: it does not matter who moves first!

Key Q: Under what conditions on f, X, Y does the equality hold?

Existence of Saddle Points

Min-Max and Max-Min

Consider more broadly the pair of problems:

$$\sup_{y \in Y} \inf_{x \in X} f(x, y) \; = \; \inf_{x \in X} \sup_{y \in Y} f(x, y)$$

Saddle Point: it does not matter who moves first!

Key Q: Under what conditions on f, X, Y does the equality hold?

Sion-Kakutani Theorem

Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be convex and compact subsets and let $f : X \times Y \to \mathbb{R}$ be a continuous function that is convex in $x \in X$ for any fixed $y \in Y$ and is concave in $y \in Y$ for any fixed $x \in X$. Then,

$$\min_{x \in X} \max_{y \in Y} f(x, y) = \max_{y \in Y} \min_{x \in X} f(x, y).$$

Existence of Saddle Points

Min-Max and Max-Min

Consider more broadly the pair of problems:

$$\sup_{y \in Y} \inf_{x \in X} f(x,y) = \inf_{x \in X} \sup_{y \in Y} f(x,y)$$

Saddle Point: it does not matter who moves first!

Key Q: Under what conditions on f, X, Y does the equality hold?

Sion-Kakutani Theorem

Let $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ be convex and compact subsets and let $f : X \times Y \to \mathbb{R}$ be a continuous function that is convex in $x \in X$ for any fixed $y \in Y$ and is concave in $y \in Y$ for any fixed $x \in X$. Then,

$$\min_{x \in X} \max_{y \in Y} f(x,y) = \max_{y \in Y} \min_{x \in X} f(x,y).$$

Generalizations: Y only needs to be convex (not compact); $f(\cdot, y)$ must be quasi-convex on X and with closed lower level sets (for any $y \in Y$); and $f(x, \cdot)$ must be quasi-concave on Y and with closed upper level sets (for any $x \in X$)

Saddle Points and Optimality in Convex Programming

Primal Problem

$$
\begin{aligned}
(\mathscr{P}) \text{ minimize}_x \quad & f_0(x) \\
& f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& x \in X.
\end{aligned}
$$

Saddle Points and Optimality in Convex Programming

Primal Problem

$$
\begin{aligned}
(\mathscr{P})\ \text{minimize}_x \quad & f_0(x) \\
& f_i(x) \leq 0, \qquad i = 1, \ldots, m \\
& x \in X.
\end{aligned}
$$

Saddle Point Optimality Condition in Convex Programming

Let $\mathcal{L}(x, \lambda)$ be the Lagrangian function and $x^* \in X$. Then:

Saddle Points and Optimality in Convex Programming

Primal Problem

$$
\begin{aligned}
(\mathscr{P})\ \text{minimize}_x \quad & f_0(x) \\
& f_i(x) \leq 0, \qquad i = 1, \dots, m \\
& x \in X.
\end{aligned}
$$

Saddle Point Optimality Condition in Convex Programming

Let $\mathcal{L}(x, \lambda)$ be the Lagrangian function and $x^* \in X$. Then:

(i) A **sufficient condition** for x^* to be optimal is the existence of $\lambda^* \geq 0$ such that (x^*, λ^*) is a saddle point of the Lagrange function $\mathcal{L}(x, \lambda)$:

$$\mathcal{L}(x, \lambda^*) \geq \mathcal{L}(x^*, \lambda^*) \geq \mathcal{L}(x^*, \lambda) \quad \forall x \in X,\ \lambda \geq 0.$$

Saddle Points and Optimality in Convex Programming

Primal Problem

$$
\begin{aligned}
(\mathscr{P}) \text{ minimize}_x \quad & f_0(x) \\
& f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& x \in X.
\end{aligned}
$$

Saddle Point Optimality Condition in Convex Programming

Let $\mathcal{L}(x, \lambda)$ be the Lagrangian function and $x^* \in X$. Then:

(i) A **sufficient condition** for x^* to be optimal is the existence of $\lambda^* \geq 0$ such that (x^*, λ^*) is a saddle point of the Lagrange function $\mathcal{L}(x, \lambda)$:

$$\mathcal{L}(x, \lambda^*) \geq \mathcal{L}(x^*, \lambda^*) \geq \mathcal{L}(x^*, \lambda) \quad \forall x \in X,\ \lambda \geq 0.$$

(ii) If $(\mathscr{P})$ is a convex optimization convex and satisfies the Slater condition, then the above condition is also **necessary** for the optimality of x^*.

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$
\begin{aligned}
(\mathscr{P})\ \min_x \quad & f(x) \\
& f_i(x) \leq 0, \qquad i = 1, \dots, m \\
& h_i(x) = 0, \quad i = 1, \dots, s \\
& x \in X.
\end{aligned}
$$

- Will **not** assume convexity unless explicitly stated...
- **Key Q:** *"We have a feasible x. What are the conditions (necessary, sufficient, necessary and sufficient) for x to be optimal?"*
- What to hope for?

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$
\begin{aligned}
(\mathscr{P})\ \min_x \quad & f(x) \\
& f_i(x) \le 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad i = 1, \ldots, s \\
& x \in X.
\end{aligned}
$$

- Will **not** assume convexity unless explicitly stated...
- **Key Q:** *"We have a feasible x. What are the conditions (necessary, sufficient, necessary and sufficient) for x to be optimal?"*
- What to hope for?
  - **necessary** conditions for the optimality of x^*
  - **sufficient** conditions for the **local optimality** of x^*
- Cannot expect **global optimality** of x^* without some "global" requirement on f, g_i, h_i (e.g., convexity)

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$\begin{array}{rll} (\mathscr{P})\ \min_x & f(x) & \\ (\lambda_j \rightarrow) & f_i(x) \leq 0, & i = 1, \ldots, m \\ (\nu_j \rightarrow) & h_i(x) = 0, & i = 1, \ldots, s \\ & x \in X. & \end{array}$$

- **If we had strong duality**, and x^* optimal for $(\mathscr{P})$ and λ^*, ν^* optimal for $(\mathscr{D})$:

$$f_0(x^*) = g(\lambda^*, \nu^*)$$

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$\begin{array}{rll} (\mathscr{P})\ \min_x & f(x) & \\ (\lambda_j \rightarrow) & f_i(x) \leq 0, & i = 1, \ldots, m \\ (\nu_j \rightarrow) & h_i(x) = 0, & i = 1, \ldots, s \\ & x \in X. & \end{array}$$

- **If we had strong duality**, and x^* optimal for $(\mathscr{P})$ and λ^*, ν^* optimal for $(\mathscr{D})$:

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) \\ &= \inf_{x \in X} \left[f(x) + \sum_{j=1}^{m} \lambda_j^* f_j(x) + \sum_{j=1}^{s} \nu_j^* h_j(x) \right] \end{aligned}$$

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$\begin{array}{rll} (\mathscr{P})\ \min_x & f(x) & \\ (\lambda_j \rightarrow) & f_i(x) \leq 0, & i = 1, \ldots, m \\ (\nu_j \rightarrow) & h_i(x) = 0, & i = 1, \ldots, s \\ & x \in X. & \end{array}$$

- **If we had strong duality**, and x^* optimal for $(\mathscr{P})$ and λ^*, ν^* optimal for $(\mathscr{D})$:

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) \\ &= \inf_{x \in X} \left[f(x) + \sum_{j=1}^{m} \lambda_j^* f_j(x) + \sum_{j=1}^{s} \nu_j^* h_j(x) \right] \\ &\leq f_0(x^*) + \sum_{j=1}^{m} \lambda_j^* g_j(x^*) \end{aligned}$$

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$
\begin{array}{rll}
(\mathscr{P})\ \min_x & f(x) & \\
(\lambda_j \rightarrow) & f_i(x) \leq 0, & i = 1, \ldots, m \\
(\nu_j \rightarrow) & h_i(x) = 0, & i = 1, \ldots, s \\
 & x \in X. &
\end{array}
$$

- **If we had strong duality**, and x^* optimal for $(\mathscr{P})$ and λ^*, ν^* optimal for $(\mathscr{D})$:

$$
\begin{aligned}
f_0(x^*) &= g(\lambda^*, \nu^*) \\
&= \inf_{x \in X} \left[f(x) + \sum_{j=1}^{m} \lambda_j^* f_j(x) + \sum_{j=1}^{s} \nu_j^* h_j(x) \right] \\
&\leq f_0(x^*) + \sum_{j=1}^{m} \lambda_j^* g_j(x^*) \\
&\leq f_0(x^*)
\end{aligned}
$$

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$
\begin{array}{rl}
(\mathscr{P})\ \min_x & f(x) \\
(\lambda_j \rightarrow) & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
(\nu_j \rightarrow) & h_i(x) = 0, \quad i = 1, \ldots, s \\
 & x \in X.
\end{array}
$$

- **If we had strong duality**, and x^* optimal for $(\mathscr{P})$ and λ^*, ν^* optimal for $(\mathscr{D})$:

$$
\begin{aligned}
f_0(x^*) &= g(\lambda^*, \nu^*) \\
&= \inf_{x \in X} \left[f(x) + \sum_{j=1}^{m} \lambda_j^* f_j(x) + \sum_{j=1}^{s} \nu_j^* h_j(x) \right] \\
&\leq f_0(x^*) + \sum_{j=1}^{m} \lambda_j^* g_j(x^*) \\
&\leq f_0(x^*)
\end{aligned}
$$

- This implies **complementary slackness**: $\lambda_j^* \cdot g_j(x_j^*) = 0$, or equivalently,

$$
\lambda_j^* > 0 \Rightarrow g_j(x_j^*) = 0 \qquad , \qquad g_j(x_j^*) < 0 \Rightarrow \lambda_j^* = 0
$$

Complementary Slackness

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$
\begin{array}{rl}
(\mathscr{P})\ \min_x & f_0(x) \\
(\lambda_j \to) & f_j(x) \leq 0, \quad j = 1, \ldots, m \\
(\nu_j \to) & h_j(x) = 0, \quad j = 1, \ldots, s \\
& x \in X.
\end{array}
$$

- **If we had strong duality**, and x^* optimal for $(\mathscr{P})$ and λ^*, ν^* optimal for $(\mathscr{D})$:

$$f_0(x^*) = g(\lambda^*, \nu^*)$$

Complementary Slackness

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$\begin{array}{rl} (\mathscr{P})\ \min_x & f_0(x) \\ (\lambda_j \to) & f_j(x) \le 0, \quad j = 1, \dots, m \\ (\nu_j \to) & h_j(x) = 0, \quad j = 1, \dots, s \\ & x \in X. \end{array}$$

- **If we had strong duality**, and x^* optimal for $(\mathscr{P})$ and λ^*, ν^* optimal for $(\mathscr{D})$:

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) \\ &= \inf_{x \in X} \left[f(x) + \sum_{j=1}^{m} \lambda_j^* f_j(x) + \sum_{j=1}^{s} \nu_j^* h_j(x) \right] \end{aligned}$$

Complementary Slackness

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$\begin{array}{rl} (\mathscr{P})\ \min_x & f_0(x) \\ (\lambda_j \to) & f_j(x) \leq 0, \quad j = 1, \ldots, m \\ (\nu_j \to) & h_j(x) = 0, \quad j = 1, \ldots, s \\ & x \in X. \end{array}$$

- **If we had strong duality**, and x^* optimal for $(\mathscr{P})$ and λ^*, ν^* optimal for $(\mathscr{D})$:

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) \\ &= \inf_{x \in X} \left[f(x) + \sum_{j=1}^{m} \lambda_j^* f_j(x) + \sum_{j=1}^{s} \nu_j^* h_j(x) \right] \\ &\leq f_0(x^*) + \sum_{j=1}^{m} \lambda_j^* g_j(x^*) \end{aligned}$$

Complementary Slackness

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$\begin{array}{rll} (\mathscr{P})\ \min_x & f_0(x) & \\ (\lambda_j \rightarrow) & f_j(x) \leq 0, & j = 1, \ldots, m \\ (\nu_j \rightarrow) & h_j(x) = 0, & j = 1, \ldots, s \\ & x \in X. & \end{array}$$

- **If we had strong duality**, and x^* optimal for $(\mathscr{P})$ and λ^*, ν^* optimal for $(\mathscr{D})$:

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) \\ &= \inf_{x \in X} \left[f(x) + \sum_{j=1}^{m} \lambda_j^* f_j(x) + \sum_{j=1}^{s} \nu_j^* h_j(x) \right] \\ &\leq f_0(x^*) + \sum_{j=1}^{m} \lambda_j^* g_j(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

Complementary Slackness

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$\begin{array}{rll} (\mathscr{P})\ \min_x & f_0(x) & \\ (\lambda_j \to) & f_j(x) \leq 0, & j = 1, \ldots, m \\ (\nu_j \to) & h_j(x) = 0, & j = 1, \ldots, s \\ & x \in X. & \end{array}$$

- **If we had strong duality**, and x^* optimal for $(\mathscr{P})$ and λ^*, ν^* optimal for $(\mathscr{D})$:

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) \\ &= \inf_{x \in X} \left[f(x) + \sum_{j=1}^{m} \lambda_j^* f_j(x) + \sum_{j=1}^{s} \nu_j^* h_j(x) \right] \\ &\leq f_0(x^*) + \sum_{j=1}^{m} \lambda_j^* g_j(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

- This implies **complementary slackness**: $\lambda_j^* \cdot g_j(x_j^*) = 0$, or equivalently,

$$\lambda_j^* > 0 \Rightarrow g_j(x_j^*) = 0 \qquad , \qquad g_j(x_j^*) < 0 \Rightarrow \lambda_j^* = 0$$

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$
\begin{array}{rll}
(\mathscr{P})\ \min_x & f_0(x) & \\
(\lambda_j \rightarrow) & f_j(x) \leq 0, & j = 1, \ldots, m \\
(\nu_j \rightarrow) & h_j(x) = 0, & j = 1, \ldots, s \\
& x \in X. &
\end{array}
$$

- $x^* \in X$, $\lambda^* \in \mathbb{R}^m$ and ν^* dual variables
- The **Karush-Kuhn-Tucker (KKT) conditions** at x^* are given by:

KKT Conditions

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$
\begin{array}{rll}
(\mathscr{P})\ \min_x & f_0(x) & \\
(\lambda_j \to) & f_j(x) \leq 0, & j = 1, \dots, m \\
(\nu_j \to) & h_j(x) = 0, & j = 1, \dots, s \\
& x \in X. &
\end{array}
$$

- $x^* \in X$, $\lambda^* \in \mathbb{R}^m$ and ν^* dual variables
- The **Karush-Kuhn-Tucker (KKT) conditions** at x^* are given by:

KKT Conditions

$$0 = \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \cdot \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \cdot \nabla h_i(x^*), \qquad (\text{“Stationarity”})$$

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$
\begin{array}{rll}
(\mathscr{P})\ \min_x & f_0(x) & \\
(\lambda_j \to) & f_j(x) \leq 0, & j = 1, \ldots, m \\
(\nu_j \to) & h_j(x) = 0, & j = 1, \ldots, s \\
& x \in X. &
\end{array}
$$

- $x^* \in X$, $\lambda^* \in \mathbb{R}^m$ and ν^* dual variables
- The **Karush-Kuhn-Tucker (KKT) conditions** at x^* are given by:

KKT Conditions

$$0 = \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \cdot \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \cdot \nabla h_i(x^*), \qquad (\text{“Stationarity”})$$

$$f_i(x^*) \leq 0,\ i = 1, \ldots, m; \quad h_i(x^*) = 0,\ i = 1, \ldots, s, \quad (\text{“Primal Feasibility”})$$

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$\begin{array}{rll} (\mathscr{P})\ \min_x & f_0(x) & \\ (\lambda_j \rightarrow) & f_j(x) \leq 0, & j = 1, \ldots, m \\ (\nu_j \rightarrow) & h_j(x) = 0, & j = 1, \ldots, s \\ & x \in X. & \end{array}$$

- $x^* \in X$, $\lambda^* \in \mathbb{R}^m$ and ν^* dual variables
- The **Karush-Kuhn-Tucker (KKT) conditions** at x^* are given by:

KKT Conditions

$$0 = \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \cdot \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \cdot \nabla h_i(x^*), \qquad (\text{“Stationarity”})$$

$$f_i(x^*) \leq 0,\ i = 1, \ldots, m; \quad h_i(x^*) = 0,\ i = 1, \ldots, s, \qquad (\text{“Primal Feasibility”})$$

$$\lambda^* \geq 0 \qquad (\text{“Dual Feasibility”})$$

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$\begin{array}{rll} (\mathscr{P})\ \min_x & f_0(x) & \\ (\lambda_j \to) & f_j(x) \leq 0, & j = 1, \ldots, m \\ (\nu_j \to) & h_j(x) = 0, & j = 1, \ldots, s \\ & x \in X. & \end{array}$$

- $x^* \in X$, $\lambda^* \in \mathbb{R}^m$ and ν^* dual variables
- The **Karush-Kuhn-Tucker (KKT) conditions** at x^* are given by:

KKT Conditions

$$0 = \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \cdot \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \cdot \nabla h_i(x^*), \qquad (\text{“Stationarity”})$$

$$f_i(x^*) \leq 0,\ i = 1, \ldots, m; \quad h_i(x^*) = 0,\ i = 1, \ldots, s, \qquad (\text{“Primal Feasibility”})$$

$$\lambda^* \geq 0 \qquad (\text{“Dual Feasibility”})$$

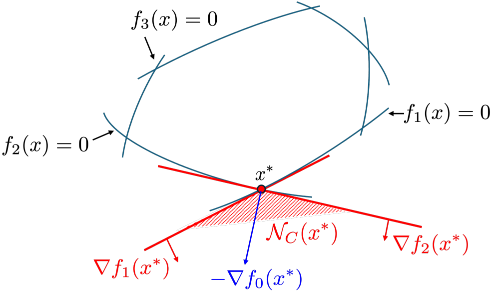
$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \ldots, m \qquad (\text{“Complementary Slackness”}).$$

Geometry Behind KKT Conditions: Inequality Case

KKT Conditions For Case Without Equality Constraints

$$0 = \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \cdot \nabla f_i(x^*) \qquad (\text{"Stationarity"})$$

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \ldots, m \qquad (\text{"Complementary Slackness"}).$$

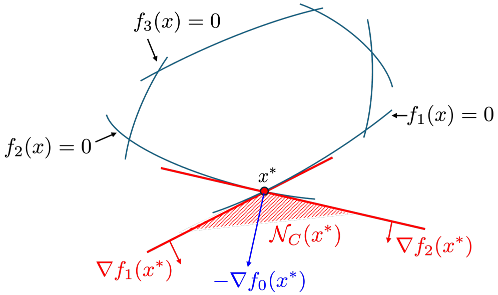


Geometry Behind KKT Conditions: Inequality Case

KKT Conditions For Case Without Equality Constraints

$$0 = \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \cdot \nabla f_i(x^*) \qquad \text{(“Stationarity”)}$$

$$\lambda_i^* f_i(x^*) = 0, \quad i = 1, \ldots, m \qquad \text{(“Complementary Slackness”).}$$



- Consider all **active** constraints at x^*, i.e., $\{i : f_i(x^*) = 0\}$
- **Stationarity:** $-\nabla f_0(x^*)$ is conic combination of gradients $\nabla f_i(x^*)$ of **active constraints**
- (Complementary slackness: only **active** constraints have $\lambda_i > 0$)
- FYI: $\mathcal{N}_C(x^*) := \{\sum_{i=1}^m \lambda_i \nabla f_i(x^*) : \lambda \geq 0\}$ is the **normal cone** at x^*

Failure of KKT Conditions

- In some cases, KKT conditions **are not necessary** at optimality

KKT Conditions Failing

$$\min_{x \in \mathbb{R}} x$$
$$x^3 \geq 0.$$

Failure of KKT Conditions

- In some cases, KKT conditions **are not necessary** at optimality

KKT Conditions Failing

$$\min_{x \in \mathbb{R}} x$$
$$x^3 \geq 0.$$

- $f_0(x) = x$ and $f_1(x) = -x^3$
- Feasible set is $(-\infty, 0]$, the optimal solution is $x^* = 0$.
- KKT condition fails because $\nabla f_0(x^*) = 1$ while $\nabla f_1(x^*) = 0$
- There is no $\lambda \geq 0$ such that $-\nabla f_0(x^*) = \lambda \nabla f_1(x^*)$.
- Note: **not** a convex optimization problem!

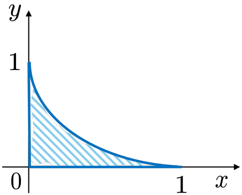
Failure of KKT Conditions - More Subtle

KKT Conditions Failing

$$\min_{x,y\in\mathbb{R}} \; -x$$

$$y-(1-x)^3 \leq 0$$

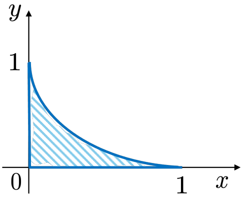
$$x, y \geq 0$$



Failure of KKT Conditions - More Subtle

KKT Conditions Failing

$$\min_{x,y\in\mathbb{R}} \; -x$$
$$y-(1-x)^3 \leq 0$$
$$x, y \geq 0$$



- $f_0(x,y) := -x$, $f_1(x,y) := y-(1-x)^3$, $f_2(x,y) := -x$ and $f_3(x,y) := -y$.
- Gradients of objective and binding constraints f_1 and f_3 at $(x^*, y^*) := (1,0)$:

$$\nabla f_0(x^*,y^*) = \begin{pmatrix} -1 \\ 0 \end{pmatrix}, \quad \nabla f_1(x^*,y^*) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \nabla f_3(x^*,y^*) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

- No $\lambda_1, \lambda_3 \geq 0$ satisfy $-\nabla f_0(x^*,y^*) = \lambda_1 \nabla f_1(x^*,y^*) + \lambda_3 \nabla f_3(x^*,y^*)$
- Reason for failing: the linearization of constraint $f_1 \leq 0$ around $(1,0)$ is $y \leq 0$, which is parallel to the existing constraint $f_3(x,y) := -y \geq 0$

Constraint Qualification Conditions

If the following conditions hold, KKT conditions are necessary at optimal x^*

Affine Active Constraints

- all **active** constraints are affine functions

Slater Conditions

- equality constraints $\{h_i\}_{i=1}^r$ are affine
- **active** inequality constraints $\{f_j\}_{j \in I(x)}$ are convex differentiable functions
- $f_j(x^*) < 0$ for all $j : f_j(x^*) = 0$

Regular Point (Linearly Independent Gradients)

- x^* is a **regular** point: gradients of active constraints $\{\nabla h_i(x)\} \cup \{\nabla f_j(x) : f_j(x) = 0\}$ are linearly independent

Mangasarian-Fromovitz

- the gradients of equality constraints are linearly independent
- there exists $v \in R^n$ such that $v^\intercal \nabla h_i(x^*) = 0$ for all equality constraints and $v^\intercal \nabla f_j(x^*)$ for all active inequality constraints

Second Order Necessary Conditions

Second Order Necessary Optimality Conditions

x^* feasible for Problem ($\mathscr{P}$) and **regular**, $f_0, f_1, \ldots, f_m, h_1, \ldots, h_s$ twice continuously differentiable in neighborhood of x^*. Define the Lagrangian function of the problem:

$$\mathcal{L}(x; \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{s} \mu_j h_j(x)$$

Second Order Necessary Conditions

Second Order **Necessary** Optimality Conditions

x^* feasible for Problem $(\mathscr{P})$ and **regular**, $f_0, f_1, \ldots, f_m, h_1, \ldots, h_s$ twice continuously differentiable in neighborhood of x^*. Define the Lagrangian function of the problem:

$$\mathcal{L}(x; \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{s} \mu_j h_j(x)$$

If x^* is locally optimal, then there exist unique $\lambda_i^* \geq 0$ and μ_j^* such that:

Second Order Necessary Conditions

Second Order Necessary Optimality Conditions

x^* feasible for Problem $(\mathscr{P})$ and **regular**, $f_0, f_1, \ldots, f_m, h_1, \ldots, h_s$ twice continuously differentiable in neighborhood of x^*. Define the Lagrangian function of the problem:

$$\mathcal{L}(x; \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{s} \mu_j h_j(x)$$

If x^* is locally optimal, then there exist unique $\lambda_i^* \geq 0$ and μ_j^* such that:

- (λ^*, μ^*) certify that x^* satisfies KKT conditions:

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

$$\nabla_x L(x^*; \lambda^*, \mu^*) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{j=1}^{s} \mu_j^* \nabla h_j(x^*) = 0$$

Second Order Necessary Conditions

Second Order **Necessary** Optimality Conditions

x^* feasible for Problem $(\mathscr{P})$ and **regular**, $f_0, f_1, \ldots, f_m, h_1, \ldots, h_s$ twice continuously differentiable in neighborhood of x^*. Define the Lagrangian function of the problem:

$$\mathcal{L}(x; \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{j=1}^{s} \mu_j h_j(x)$$

If x^* is locally optimal, then there exist unique $\lambda_i^* \geq 0$ and μ_j^* such that:

- (λ^*, μ^*) certify that x^* satisfies KKT conditions:

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

$$\nabla_x L(x^*; \lambda^*, \mu^*) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{j=1}^{s} \mu_j^* \nabla h_j(x^*) = 0$$

- The Hessian $\nabla_x^2 \mathcal{L}(x^*; \lambda^*, \mu^*)$ of $\mathcal{L}$ in x is positive semidefinite on the orthogonal complement M^* to the set of gradients of active constraints at x^*: $d^T \nabla_x^2 L(x^*; \lambda^*, \mu^*) d \geq 0$ for any $d \in M^*$, where:

$$M^* := \{d \mid d^T \nabla f_i(x^*) = 0,\ i : f_i(x^*) = 0,\ d^T \nabla h_j(x^*) = 0,\ j = 1, \ldots, s\}.$$

Second Order Sufficient Conditions

Second Order **Sufficient** Local Optimality Conditions

x^* feasible for Problem ($\mathscr{P}$) and **regular**, $f_0, f_1, \ldots, f_m, h_1, \ldots, h_s$ twice continuously differentiable in neighborhood of x^*. Define the Lagrangian function of the problem:

$$\mathcal{L}(x; \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{k} \mu_j h_j(x)$$

Assume there exist Lagrange multipliers $\lambda_i^* \geq 0$ and μ_j^* such that

Second Order Sufficient Conditions

Second Order Sufficient Local Optimality Conditions

x^* feasible for Problem $(\mathscr{P})$ and **regular**, $f_0, f_1, \ldots, f_m, h_1, \ldots, h_s$ twice continuously differentiable in neighborhood of x^*. Define the Lagrangian function of the problem:

$$\mathcal{L}(x; \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{k} \mu_j h_j(x)$$

Assume there exist Lagrange multipliers $\lambda_i^* \geq 0$ and μ_j^* such that

- (λ^*, μ^*) certify that x^* satisfies KKT conditions

Second Order Sufficient Conditions

Second Order Sufficient Local Optimality Conditions

x^* feasible for Problem $(\mathscr{P})$ and **regular**, $f_0, f_1, \ldots, f_m, h_1, \ldots, h_s$ twice continuously differentiable in neighborhood of x^*. Define the Lagrangian function of the problem:

$$\mathcal{L}(x; \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{k} \mu_j h_j(x)$$

Assume there exist Lagrange multipliers $\lambda_i^* \geq 0$ and μ_j^* such that

- (λ^*, μ^*) certify that x^* satisfies KKT conditions
- The Hessian $\nabla_x^2 \mathcal{L}(x^*; \lambda^*, \mu^*)$ of $\mathcal{L}$ in x is **positive definite** on the orthogonal complement M^{**} to the set of gradients of equality constraints and the active inequality constraints at x^* **that are associated with positive Lagrange multipliers** λ_i^*: $d^\intercal \nabla_x^2 \mathcal{L}(x^*; \lambda^*, \mu^*) d > 0$ for any $d \in M^* *$, where:

$$M^{**} := \{d \mid d^\intercal \nabla f_i(x^*) = 0, \forall\, i : \lambda_i^* > 0 \text{ and } f_i(x^*) < 0,$$
$$d^\intercal \nabla h_j(x^*) = 0, j = 1, \ldots, s\}$$

Second Order Sufficient Conditions

Second Order Sufficient Local Optimality Conditions

x^* feasible for Problem $(\mathscr{P})$ and **regular**, $f_0, f_1, \dots, f_m, h_1, \dots, h_s$ twice continuously differentiable in neighborhood of x^*. Define the Lagrangian function of the problem:

$$\mathcal{L}(x; \lambda, \mu) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{j=1}^{k} \mu_j h_j(x)$$

Assume there exist Lagrange multipliers $\lambda_i^* \geq 0$ and μ_j^* such that

- (λ^*, μ^*) certify that x^* satisfies KKT conditions
- The Hessian $\nabla_x^2 \mathcal{L}(x^*; \lambda^*, \mu^*)$ of $\mathcal{L}$ in x is **positive definite** on the orthogonal complement M^{**} to the set of gradients of equality constraints and the active inequality constraints at x^* **that are associated with positive Lagrange multipliers** λ_i^*: $d^\intercal \nabla_x^2 \mathcal{L}(x^*; \lambda^*, \mu^*) d > 0$ for any $d \in M^* *$, where:

$$M^{**} := \{d \mid d^\intercal \nabla f_i(x^*) = 0, \forall\, i : \lambda_i^* > 0 \text{ and } f_i(x^*) < 0,$$
$$d^\intercal \nabla h_j(x^*) = 0, j = 1, \dots, s\}$$

Then x^* is locally optimal for $(\mathscr{P})$.