Lecture 9: Quadratic Optimization
KKT Optimality Conditions

Oct 20, 2025

Quick Announcements

- Regular class this Friday
- My office hours this week: Wednesday, 3:15-4:15pm (same Google cal link)
- Monday (Oct 27) - midterm review with the CAs
- Agenda for today
  - Duality in Quadratic Optimization
  - A tiny bit of Saddle Theory
  - KKT Optimality Conditions
  - Fenchel duality

Last Time: Convex Duality Framework

$$\begin{aligned} \text{minimize}_{x \in X} \;& f_0(x) \\ \text{subject to} \;& f_i(x) \le 0, \quad i = 1, \ldots, m, \\ & h_j(x) = 0, \quad j = 1, \ldots, s \\ \text{variable} \;& x \in \mathbb{R}^n \end{aligned}$$

- With λ_i, ν_j denoting Lagrange multipliers for g_i, h_j, respectively, Lagrangian is:

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

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

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

- For a **convex optimization problem** (f_0, f_i convex, h_j affine), strong duality holds if the **Slater condition** holds: $\exists x \in \mathrm{rel\,int}(X)$ such that $f_i(x) < 0$ for $i = 1, \ldots, m$

QPs and QCQPs

Quadratic Programs

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

$$\min \frac{1}{2}x^\mathsf{T}Qx + c^\mathsf{T}x$$

$$A_1x = b_1$$

$$A_2x \leq b_2$$

where $Q = Q^\mathsf{T}$.

QPs and QCQPs

Quadratic Programs

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

$$\begin{aligned} \min\ & \frac{1}{2}x^\mathsf{T}Qx + c^\mathsf{T}x \\ & A_1x = b_1 \\ & A_2x \leq b_2 \end{aligned}$$

where $Q = Q^\mathsf{T}$.

Quadratically Constrained Quadratic Programs

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

$$\begin{aligned} \min\ & \frac{1}{2}x^\mathsf{T}Q_0x + c^\mathsf{T}x \\ & x^\mathsf{T}Q_ix + q_i^\mathsf{T}x + b_i \leq 0, i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

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

Convex if $Q_0 \succeq 0, Q_i \succeq 0$. Gurobi can now handle **non-convex** QCQPs!

One Problem to Warm Up

Convex QCQP

$$\begin{aligned} \text{minimize } & \frac{1}{2}x^T Q_0 x + q_0^T x + r_0 \\ \text{subject to } & \frac{1}{2}x^T Q_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m, \end{aligned}$$

where $Q_0 \succ 0$ and $Q_i \succeq 0$

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

Quadratic Programs - Preliminaries

Unconstrained Quadratic Program

For $Q = Q^\mathsf{T}$, consider the following unconstrained problem:

$$\min f(x) := \frac{1}{2}x^\mathsf{T}Qx + q^\mathsf{T}x$$

- **What is the optimal value $p^\star$?**

Quadratic Programs - Preliminaries

Unconstrained Quadratic Program

For $Q = Q^{\mathsf{T}}$, consider the following unconstrained problem:

$$\min\ f(x) := \frac{1}{2}x^{\mathsf{T}}Qx + q^{\mathsf{T}}x$$

- **What is the optimal value $p^{\star}$?**

$$\nabla_x f(x) = 0 \quad \Leftrightarrow \quad Qx = -q$$

$$p^{\star} = \begin{cases} -\frac{1}{2}q^{\mathsf{T}}Q^{\dagger}q & \text{if } Q \succeq 0 \text{ and } q \in \mathcal{R}(Q) \\ -\infty & \text{otherwise.} \end{cases}$$

- For Q with singular value decomposition $Q = U\Sigma V^{\mathsf{T}}$, $Q^{\dagger} := V\Sigma^{-1}U^{\mathsf{T}}$

Convex QCQP

QCQP

$$\begin{aligned} \text{minimize } & \frac{1}{2}x^T Q_0 x + q_0^T x + r_0 \\ \text{subject to } & \frac{1}{2}x^T Q_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m, \end{aligned}$$

where $Q_0 \succ 0$ and $Q_i \succeq 0$

- The Lagrangian is:

Convex QCQP

QCQP

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} x^{\mathsf{T}} Q_0 x + q_0^{\mathsf{T}} x + r_0 \\ \text{subject to} \quad & \frac{1}{2} x^{\mathsf{T}} Q_i x + q_i^{\mathsf{T}} x + r_i \leq 0, \quad i = 1, \ldots, m, \end{aligned}$$

where $Q_0 \succ 0$ and $Q_i \succeq 0$

- The Lagrangian is:

$$\mathcal{L}(x, \lambda) = \frac{1}{2} x^{\mathsf{T}} Q(\lambda) x + q(\lambda)^{\mathsf{T}} x + r(\lambda),$$

where $Q(\lambda) = Q_0 + \sum_{i=1}^m \lambda_i Q_i, \quad q(\lambda) = q_0 + \sum_{i=1}^m \lambda_i q_i, \quad r(\lambda) = r_0 + \sum_{i=1}^m \lambda_i r_i$

Convex QCQP

QCQP

$$\begin{aligned} \text{minimize } & \frac{1}{2}x^\mathsf{T} Q_0 x + q_0^\mathsf{T} x + r_0 \\ \text{subject to } & \frac{1}{2}x^\mathsf{T} Q_i x + q_i^\mathsf{T} x + r_i \le 0, \quad i = 1, \dots, m, \end{aligned}$$

where $Q_0 \succ 0$ and $Q_i \succeq 0$

- The Lagrangian is:
$$\mathcal{L}(x, \lambda) = \frac{1}{2}x^\mathsf{T} Q(\lambda) x + q(\lambda)^\mathsf{T} x + r(\lambda),$$
where $Q(\lambda) = Q_0 + \sum_{i=1}^m \lambda_i Q_i, \quad q(\lambda) = q_0 + \sum_{i=1}^m \lambda_i q_i, \quad r(\lambda) = r_0 + \sum_{i=1}^m \lambda_i r_i$
- Because $\lambda \ge 0$, we have $Q(\lambda) \succ 0$ and therefore:
$$g(\lambda) = \inf_x L(x, \lambda) = -\frac{1}{2} q(\lambda)^\mathsf{T} Q(\lambda)^{-1} q(\lambda) + r(\lambda).$$
- We can express the dual problem as:
$$\max_{\lambda \ge 0} -\frac{1}{2} q(\lambda)^\mathsf{T} Q(\lambda)^{-1} q(\lambda) + r(\lambda)$$

Convex QCQP

QCQP

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} x^T Q_0 x + q_0^T x + r_0 \\ \text{subject to} \quad & \frac{1}{2} x^T Q_i x + q_i^T x + r_i \leq 0, \quad i = 1, \dots, m, \end{aligned}$$

where $Q_0 \succ 0$ and $Q_i \succeq 0$

- The Lagrangian is:
$$\mathcal{L}(x, \lambda) = \frac{1}{2} x^T Q(\lambda) x + q(\lambda)^T x + r(\lambda),$$
where $Q(\lambda) = Q_0 + \sum_{i=1}^m \lambda_i Q_i, \quad q(\lambda) = q_0 + \sum_{i=1}^m \lambda_i q_i, \quad r(\lambda) = r_0 + \sum_{i=1}^m \lambda_i r_i$
- Because $\lambda \geq 0$, we have $Q(\lambda) \succ 0$ and therefore:
$$g(\lambda) = \inf_x L(x, \lambda) = -\frac{1}{2} q(\lambda)^T Q(\lambda)^{-1} q(\lambda) + r(\lambda).$$
- We can express the dual problem as:
$$\max_{\lambda \geq 0} - \frac{1}{2} q(\lambda)^T Q(\lambda)^{-1} q(\lambda) + r(\lambda)$$
- Slater condition holds if there exists an x with
$$\frac{1}{2} x^T Q_i x + q_i^T x + r_i < 0, \quad i = 1, \dots, m.$$

Other Important Examples in the Notes

- A **non-convex** QCQP: for $Q = Q^\mathsf{T}$ and $Q \not\succeq 0$, consider:

$$\begin{aligned} \text{minimize } & x^\mathsf{T} Q x + 2c^\mathsf{T} x \\ \text{subject to } & x^\mathsf{T} x \leq 1 \end{aligned}$$

- Regularized Support Vector Machines (SVM)
- Entropy Maximization

Saddle Point Theory

- Optional reading in the notes, but very insightful

Saddle Point Theory

- Optional reading in the notes, but very insightful

Alternative Formulation of Primal and Dual Problems

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

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

Saddle Point Theory

- Optional reading in the notes, but very insightful

Alternative Formulation of Primal and Dual Problems

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

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

- **Weak duality** restatement:

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

- **Strong duality** restatement:

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

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

Saddle Problem and Game Theoretic Interpretation

Min-Max and Max-Min

Consider the pair of problems:

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

Saddle Problem and Game Theoretic Interpretation

Min-Max and Max-Min

Consider the pair of problems:

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

- Game theoretic interpretation : zero-sum game
- y player maximizes, x player minimizes. Difference is who moves first.

Saddle Problem and Game Theoretic Interpretation

Min-Max and Max-Min

Consider the pair of problems:

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

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

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

Saddle Problem and Game Theoretic Interpretation

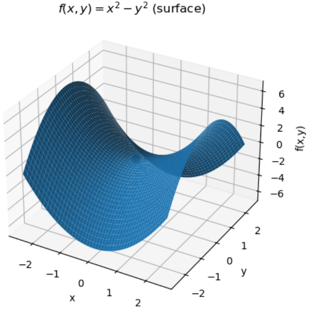
Min-Max and Max-Min

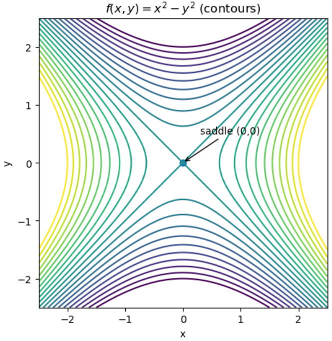
Consider the pair of problems:

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

- When do f, X, Y satisfy the **saddle-point property**, i.e., equality holds:

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





Saddle Problem and Game Theoretic Interpretation

Min-Max and Max-Min

Consider the pair of problems:

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

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 possible: 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 $x \in X$)

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$
\begin{array}{lll}
(\mathcal{P}) & \text{minimize} & f_0(x) \\
& & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& & h_j(x) = 0, \quad j = 1, \ldots, s \\
& & x \in X \\
& \text{variables} & x \in \mathbb{R}^n.
\end{array}
$$

- 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:

$$
\begin{array}{lll}
(\mathcal{P}) & \text{minimize} & f_0(x) \\
& & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& & h_j(x) = 0, \quad j = 1, \ldots, s \\
& & x \in X \\
& \text{variables} & x \in \mathbb{R}^n.
\end{array}
$$

- 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 $x^\star$
  - **sufficient** conditions for the **local optimality** of $x^\star$
- Cannot expect **global optimality** of $x^\star$ without some "global" requirement on f_i, h_j (e.g., convexity)

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$\begin{array}{rll} (\mathcal{P}) \ \text{minimize} & f_0(x) & \\ \lambda_i \rightarrow & f_i(x) \leq 0, & i = 1, \ldots, m \\ \nu_j \rightarrow & h_j(x) = 0, & j = 1, \ldots, s \\ & x \in X & \\ \text{variables} & x \in \mathbb{R}^n. & \end{array}$$

- **If we had strong duality** and $x^\star$ optimal for $(\mathcal{P})$ and $\lambda^\star, \nu^\star$ optimal for $(\mathcal{D})$:

$$f_0(x^\star) = g(\lambda^\star, \nu^\star) = \inf_{x \in X} \left[f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{j=1}^{s} \nu_j^\star h_j(x) \right]$$

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$
\begin{array}{rll}
(\mathcal{P}) \ \text{minimize} & f_0(x) & \\
\lambda_i \rightarrow & f_i(x) \leq 0, & i = 1, \dots, m \\
\nu_j \rightarrow & h_j(x) = 0, & j = 1, \dots, s \\
& x \in X & \\
\text{variables} & x \in \mathbb{R}^n. &
\end{array}
$$

- **If we had strong duality** and $x^\star$ optimal for $(\mathcal{P})$ and $\lambda^\star, \nu^\star$ optimal for $(\mathcal{D})$:

$$
f_0(x^\star) = g(\lambda^\star, \nu^\star) = \inf_{x \in X} \left[f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{j=1}^{s} \nu_j^\star h_j(x) \right]
$$

$$
\leq f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star f_i(x^\star)
$$

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$
\begin{array}{rll}
(\mathcal{P}) \ \text{minimize} & f_0(x) & \\
\lambda_i \rightarrow & f_i(x) \leq 0, & i = 1, \ldots, m \\
\nu_j \rightarrow & h_j(x) = 0, & j = 1, \ldots, s \\
 & x \in X & \\
\text{variables} & x \in \mathbb{R}^n. &
\end{array}
$$

- **If we had strong duality** and $x^\star$ optimal for $(\mathcal{P})$ and $\lambda^\star, \nu^\star$ optimal for $(\mathcal{D})$:

$$
\begin{aligned}
f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_{x \in X} \left[f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{j=1}^{s} \nu_j^\star h_j(x) \right] \\
&\leq f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) \\
&\leq f_0(x^\star)
\end{aligned}
$$

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$\begin{array}{rll} (\mathcal{P}) \text{ minimize} & f_0(x) & \\ \lambda_i \rightarrow & f_i(x) \leq 0, & i = 1, \ldots, m \\ \nu_j \rightarrow & h_j(x) = 0, & j = 1, \ldots, s \\ & x \in X & \\ \text{variables} & x \in \mathbb{R}^n. & \end{array}$$

- **If we had strong duality** and $x^\star$ optimal for $(\mathcal{P})$ and $\lambda^\star, \nu^\star$ optimal for $(\mathcal{D})$:

$$\begin{aligned} f_0(x^\star) = g(\lambda^\star, \nu^\star) &= \inf_{x \in X} \left[f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{j=1}^{s} \nu_j^\star h_j(x) \right] \\ &\leq f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) \\ &\leq f_0(x^\star) \end{aligned}$$

- This implies **complementary slackness**: $\lambda_i^\star \cdot f_i(x^\star) = 0$, or equivalently,

$$\lambda_i^\star > 0 \Rightarrow f_i(x^\star) = 0 \qquad \text{and} \qquad f_i(x^\star) < 0 \Rightarrow \lambda_i^\star = 0$$

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

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

KKT Conditions

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

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

KKT Conditions

$$0 = \nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \cdot \nabla f_i(x^\star) + \sum_{j=1}^{p} \nu_j^\star \cdot \nabla h_j(x^\star), \qquad (\text{“Stationarity”})$$

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

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

KKT Conditions

$$0 = \nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \cdot \nabla f_i(x^\star) + \sum_{j=1}^{p} \nu_j^\star \cdot \nabla h_j(x^\star), \qquad (\text{“Stationarity”})$$

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

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

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

KKT Conditions

$$0 = \nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \cdot \nabla f_i(x^\star) + \sum_{j=1}^{p} \nu_j^\star \cdot \nabla h_j(x^\star), \qquad (\text{“Stationarity”})$$

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

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

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

$$
\begin{array}{rll}
(\mathcal{P})\ \min_x & f_0(x) & \\
(\lambda_i \rightarrow) & f_i(x) \leq 0, & i = 1, \dots, m \\
(\nu_j \rightarrow) & h_j(x) = 0, & j = 1, \dots, s \\
 & x \in X. &
\end{array}
$$

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

KKT Conditions

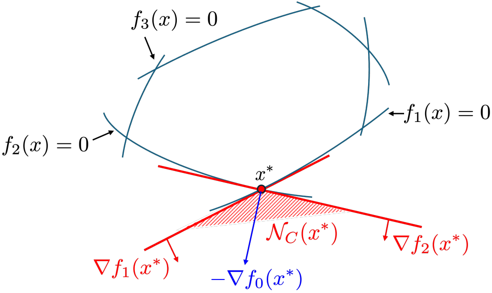
$$
\begin{array}{ll}
0 = \nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \cdot \nabla f_i(x^\star) + \sum_{j=1}^{p} \nu_j^\star \cdot \nabla h_j(x^\star), & (\text{“Stationarity”}) \\
f_i(x^\star) \leq 0,\ i = 1, \dots, m; \quad h_j(x^\star) = 0,\ j = 1, \dots, s, & (\text{“Primal Feasibility”}) \\
\lambda^\star \geq 0 & (\text{“Dual Feasibility”}) \\
\lambda_i^\star f_i(x^\star) = 0, \quad i = 1, \dots, m & (\text{“Complementary Slackness”}).
\end{array}
$$

Geometry Behind KKT Conditions: Inequality Case

KKT Conditions For Case Without Equality Constraints

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

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

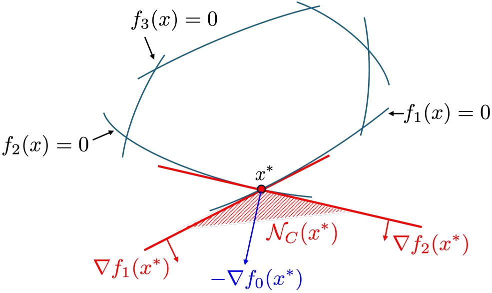


Geometry Behind KKT Conditions: Inequality Case

KKT Conditions For Case Without Equality Constraints

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

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



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

Failure of KKT Conditions

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

KKT Conditions Failing

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

- **Is this a convex optimization problem? What is $p^\star$? What is $x^\star$?**

Failure of KKT Conditions

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

KKT Conditions Failing

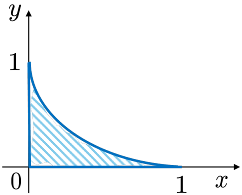
$$\min_{x\in\mathbb{R}} x$$
$$x^3 \geq 0.$$

- **Is this a convex optimization problem? What is $p^\star$? What is $x^\star$?**
- $f_0(x) = x$ and $f_1(x) = -x^3$. Nonconvex because of f_1.
- Feasible set is $(-\infty, 0]$; optimal value is $p^\star = 0$, optimal solution $x^\star = 0$.
- KKT condition fails because $\nabla f_0(x^\star) = 1$ while $\nabla f_1(x^\star) = 0$
- There is no $\lambda \geq 0$ such that $-\nabla f_0(x^\star) = \lambda \nabla f_1(x^\star)$.

Failure of KKT Conditions - More Subtle

KKT Conditions Failing

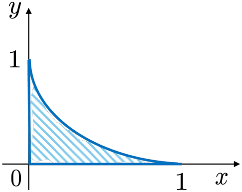
$$\min_{x,y\in\mathbb{R}} \; -x$$
$$y-(1-x)^3 \leq 0$$
$$x, y \geq 0$$



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 $(x^\star, y^\star) := (1,0)$:

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

- No $\lambda_1, \lambda_3 \geq 0$ satisfy $-\nabla f_0(x^\star,y^\star) = \lambda_1 \nabla f_1(x^\star,y^\star) + \lambda_3 \nabla f_3(x^\star,y^\star)$
- 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

Setup: $x^\star$ feasible. Active inequality constraints: $I(x^\star) = \{i \in \{1, \ldots, m\} : f_i(x^\star) = 0\}$.

Constraint Qualification Conditions

Setup: $x^\star$ feasible. Active inequality constraints: $I(x^\star) = \{i \in \{1, \ldots, m\} : f_i(x^\star) = 0\}$.

If one of the following holds, KKT conditions are necessary for $x^\star$ to be optimal:

1. Affine Active Constraints

- all **active** constraints are affine functions

Constraint Qualification Conditions

Setup: $x^\star$ feasible. Active inequality constraints: $I(x^\star) = \{i \in \{1, \ldots, m\} : f_i(x^\star) = 0\}$.

If one of the following holds, KKT conditions are necessary for $x^\star$ to be optimal:

1. Affine Active Constraints

- all **active** constraints are affine functions

2. Slater Conditions

- all functions $\{h_j\}_{j=1}^{s}$ in equality constraints are **affine**
- all functions $\{f_i : i \in I(x)\}$ in **active** inequality constraints are **convex**
- $\exists \bar{x} \in \operatorname{rel\,int}(X) : f_i(\bar{x}) < 0$ for all $i \in I(x^\star)$

Constraint Qualification Conditions

Setup: $x^\star$ feasible. Active inequality constraints: $I(x^\star) = \{i \in \{1, \dots, m\} : f_i(x^\star) = 0\}$.

If one of the following holds, KKT conditions are necessary for $x^\star$ to be optimal:

1. Affine Active Constraints

- all **active** constraints are affine functions

2. Slater Conditions

- all functions $\{h_j\}_{j=1}^{s}$ in equality constraints are **affine**
- all functions $\{f_i : i \in I(x)\}$ in **active** inequality constraints are **convex**
- $\exists \bar{x} \in \mathrm{rel\,int}(X) : \ f_i(\bar{x}) < 0$ for all $i \in I(x^\star)$

3. Regular Point (Linearly Independent Gradients)

- $x^\star$ is a **regular** point: gradients of all active constraints $\{\nabla f_i(x) : i \in I(x^\star)\} \cup \{\nabla h_j(x) : j = 1, \dots, s\}$ are linearly independent

Constraint Qualification Conditions

Setup: $x^\star$ feasible. Active inequality constraints: $I(x^\star) = \{i \in \{1, \ldots, m\} : f_i(x^\star) = 0\}$.
If one of the following holds, KKT conditions are necessary for $x^\star$ to be optimal:

1. Affine Active Constraints

- all **active** constraints are affine functions

2. Slater Conditions

- all functions $\{h_j\}_{j=1}^s$ in equality constraints are **affine**
- all functions $\{f_i : i \in I(x)\}$ in **active** inequality constraints are **convex**
- $\exists \bar{x} \in \operatorname{rel\,int}(X) : \; f_i(\bar{x}) < 0$ for all $i \in I(x^\star)$

3. Regular Point (Linearly Independent Gradients)

- $x^\star$ is a **regular** point: gradients of all active constraints $\{\nabla f_i(x) : i \in I(x^\star)\} \cup \{\nabla h_j(x) : j = 1, \ldots, s\}$ are linearly independent

4. Mangasarian-Fromovitz

- the gradients of equality constraints are linearly independent
- $\exists v \in R^n : v^\mathsf{T} \nabla f_i(x^\star) < 0$ for $i \in I(x^\star)$ and $v^\mathsf{T} \nabla h_j(x^\star) = 0, j = 1, \ldots, s$

Second Order Necessary Conditions

Second Order **Necessary** Optimality Conditions

$x^\star$ feasible for problem $(\mathcal{P})$ and **regular**, $\{f_i\}_{i=1}^m, \{h_j\}_{j=1}^s$ **twice** continuously differentiable in neighborhood of $x^\star$. Let $\mathcal{L}(x;\lambda,\nu)$ denote the Lagrangian function.

Second Order Necessary Conditions

Second Order **Necessary** Optimality Conditions

$x^\star$ feasible for problem $(\mathcal{P})$ and **regular**, $\{f_i\}_{i=1}^m, \{h_j\}_{j=1}^s$ **twice** continuously differentiable in neighborhood of $x^\star$. Let $\mathcal{L}(x; \lambda, \nu)$ denote the Lagrangian function.

If $x^\star$ is locally optimal, then there exist unique $\lambda^\star \geq 0$ and $\nu^\star$ such that:

Second Order Necessary Conditions

Second Order **Necessary** Optimality Conditions

$x^\star$ feasible for problem $(\mathcal{P})$ and **regular**, $\{f_i\}_{i=1}^m, \{h_j\}_{j=1}^s$ **twice** continuously differentiable in neighborhood of $x^\star$. Let $\mathcal{L}(x;\lambda,\nu)$ denote the Lagrangian function.

If $x^\star$ is locally optimal, then there exist unique $\lambda^\star \geq 0$ and $\nu^\star$ such that:

- $(\lambda^\star, \nu^\star)$ certify that $x^\star$ satisfies KKT conditions:

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

$$\nabla_x \mathcal{L}(x^\star;\lambda^\star,\nu^\star) = \nabla f(x^\star) + \sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) + \sum_{j=1}^s \nu_j^\star \nabla h_j(x^\star) = 0.$$

Second Order Necessary Conditions

Second Order **Necessary** Optimality Conditions

$x^\star$ feasible for problem $(\mathcal{P})$ and **regular**, $\{f_i\}_{i=1}^m, \{h_j\}_{j=1}^s$ **twice** continuously differentiable in neighborhood of $x^\star$. Let $\mathcal{L}(x;\lambda,\nu)$ denote the Lagrangian function.

If $x^\star$ is locally optimal, then there exist unique $\lambda^\star \geq 0$ and $\nu^\star$ such that:

- $(\lambda^\star, \nu^\star)$ certify that $x^\star$ satisfies KKT conditions:

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

$$\nabla_x \mathcal{L}(x^\star; \lambda^\star, \nu^\star) = \nabla f(x^\star) + \sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) + \sum_{j=1}^s \nu_j^\star \nabla h_j(x^\star) = 0.$$

- The Hessian $\nabla_x^2 \mathcal{L}(x^\star; \lambda^\star, \nu^\star)$ of $\mathcal{L}$ in x is **positive semidefinite** on the orthogonal complement $M^\star$ to the set of gradients of active constraints at $x^\star$:

$$d^T \nabla_x^2 \mathcal{L}(x^\star; \lambda^\star, \nu^\star)\, d \geq 0 \text{ for any } d \in M^\star$$

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

Second Order **Sufficient** Conditions

Second Order **Sufficient** Local Optimality Conditions

$x^\star$ feasible for problem $(\mathcal{P})$ and **regular**, $\{f_i\}_{i=1}^m, \{h_j\}_{j=1}^s$ **twice** continuously differentiable in neighborhood of $x^\star$. Let $\mathcal{L}(x;\lambda,\nu)$ denote the Lagrangian function.

Assume there exist Lagrange multipliers $\lambda^\star \geq 0$ and $\nu^\star$ such that

Second Order Sufficient Conditions

Second Order **Sufficient** Local Optimality Conditions

$x^\star$ feasible for problem $(\mathcal{P})$ and **regular**, $\{f_i\}_{i=1}^m, \{h_j\}_{j=1}^s$ **twice** continuously differentiable in neighborhood of $x^\star$. Let $\mathcal{L}(x; \lambda, \nu)$ denote the Lagrangian function.

Assume there exist Lagrange multipliers $\lambda^\star \geq 0$ and $\nu^\star$ such that

- $(\lambda^\star, \nu^\star)$ certify that $x^\star$ satisfies KKT conditions;

Second Order Sufficient Conditions

Second Order **Sufficient** Local Optimality Conditions

$x^\star$ feasible for problem $(\mathcal{P})$ and **regular**, $\{f_i\}_{i=1}^m, \{h_j\}_{j=1}^s$ **twice** continuously differentiable in neighborhood of $x^\star$. Let $\mathcal{L}(x; \lambda, \nu)$ denote the Lagrangian function.

Assume there exist Lagrange multipliers $\lambda^\star \geq 0$ and $\nu^\star$ such that

- $(\lambda^\star, \nu^\star)$ certify that $x^\star$ satisfies KKT conditions;
- The Hessian $\nabla_x^2 \mathcal{L}(x^\star; \lambda^\star, \nu^\star)$ of $\mathcal{L}$ in x is **positive definite** on the orthogonal complement M^{**} to the set of gradients of **equality constraints** and **active inequality constraints at $x^\star$ that have positive Lagrange multipliers $\lambda_i^\star$**:

Second Order Sufficient Conditions

Second Order Sufficient Local Optimality Conditions

$x^\star$ feasible for problem $(\mathcal{P})$ and **regular**, $\{f_i\}_{i=1}^m, \{h_j\}_{j=1}^s$ **twice** continuously differentiable in neighborhood of $x^\star$. Let $\mathcal{L}(x; \lambda, \nu)$ denote the Lagrangian function.

Assume there exist Lagrange multipliers $\lambda^\star \geq 0$ and $\nu^\star$ such that

- $(\lambda^\star, \nu^\star)$ certify that $x^\star$ satisfies KKT conditions;
- The Hessian $\nabla_x^2 \mathcal{L}(x^\star; \lambda^\star, \nu^\star)$ of $\mathcal{L}$ in x is **positive definite** on the orthogonal complement M^{**} to the set of gradients of **equality constraints** and **active inequality constraints at $x^\star$ that have positive Lagrange multipliers $\lambda_i^\star$**:

$$d^{\mathsf{T}} \nabla_x^2 \mathcal{L}(x^\star; \lambda^\star, \nu^\star) d > 0 \text{ for any } d \in M^{\star\star}$$

$$\text{where } M^{\star\star} := \{d \mid d^{\mathsf{T}} \nabla f_i(x^\star) = 0, \forall\, i \in I(x^\star) : \lambda_i^\star > 0 \text{ and } d^{\mathsf{T}} \nabla h_j(x^\star) = 0, j = 1, \ldots, s\}.$$

Then $x^\star$ is locally optimal for $(\mathcal{P})$.

KKT Conditions and Local vs Global Optimality

To summarize...

KKT Conditions and Optimality Notions

- To use the KKT conditions you must first check that one of the constraint qualification conditions holds. Typically, the Slater Conditions might be easiest; the Mangasarian-Fromovitz are the most general from the ones we stated
- If the constraint qualification conditions hold, then:
  - For a **general** optimization problem, the KKT conditions are **necessary** or **sufficient** (depending on which variant you use) for **local optimality** at $x^\star$
  - For a **convex** optimization problem, the KKT conditions are **necessary and sufficient** for **global optimality** at $x^\star$

A Consumer's Constrained Consumption Problem

An Example

Consider a consumer trying to maximize his utility function $u(x)$ by choosing which bundle of goods $x \in \mathbb{R}_n^+$ to purchase. The goods have prices $p > 0$ and the consumer has a budget $B > 0$. The consumer's problem can be stated as:

$$\begin{aligned} \text{maximize } & u(x) \\ \text{such that } & p^\mathsf{T} x \leq B \\ & x \geq 0, \end{aligned}$$

where $u(x)$ is a concave utility function.

- *Write down the first-order KKT conditions and try to interpret them.*
- *Are these conditions necessary for optimality?*
- *Are these conditions sufficient for optimality?*

A Consumer's Constrained Consumption Problem

$$\begin{aligned} \text{minimize} \quad & -u(x) \\ (\lambda \rightarrow) \quad & p^{\mathsf{T}} x \leq B \\ (\mu \rightarrow) \quad & -x \leq 0, \end{aligned}$$

With $\lambda \in \mathbb{R}_+, \mu \in \mathbb{R}_+^n$ denoting the Lagrange multipliers, the Lagrangian becomes:

$$\mathcal{L}(x, \lambda, \mu) = -u(x) + \lambda(p^{\mathsf{T}} x - B) - x^{\mathsf{T}} \mu.$$

A Consumer's Constrained Consumption Problem

$$\begin{aligned} &\text{minimize} \; -u(x) \\ &(\lambda \to) \;\; p^{\mathsf{T}}x \le B \\ &(\mu \to) \;\; -x \le 0, \end{aligned}$$

With $\lambda \in \mathbb{R}_+, \mu \in \mathbb{R}^n_+$ denoting the Lagrange multipliers, the Lagrangian becomes:

$$\mathcal{L}(x, \lambda, \mu) = -u(x) + \lambda(p^{\mathsf{T}}x - B) - x^{\mathsf{T}}\mu.$$

$$0 = -\frac{\partial u}{\partial x_i} + \lambda p_i - \mu_i, \; i = 1, \ldots, n \qquad \text{(“Stationarity”)}$$

$$p^{\mathsf{T}}x \le B, \;\; x \ge 0 \qquad \text{(“Primal Feasibility”)}$$

$$\lambda \ge 0, \; \mu \ge 0 \qquad \text{(“Dual Feasibility”)}$$

$$\lambda \cdot (p^{\mathsf{T}}x - B) = 0 \qquad \text{(“Complementary Slackness” 1)}$$

$$\mu_i \cdot x_i = 0 \qquad \text{(“Complementary Slackness” 2).}$$

A Consumer's Constrained Consumption Problem

$$\begin{aligned} &\text{minimize} \ \ -u(x) \\ &(\lambda \rightarrow) \ \ p^\top x \leq B \\ &(\mu \rightarrow) \ \ -x \leq 0, \end{aligned}$$

With $\lambda \in \mathbb{R}_+, \mu \in \mathbb{R}_+^n$ denoting the Lagrange multipliers, the Lagrangian becomes:

$$\mathcal{L}(x, \lambda, \mu) = -u(x) + \lambda(p^\top x - B) - x^\top \mu.$$

$$0 = -\frac{\partial u}{\partial x_i} + \lambda p_i - \mu_i, \; i = 1, \ldots, n \qquad (\text{“Stationarity”})$$

$$p^\top x \leq B, \quad x \geq 0 \qquad (\text{“Primal Feasibility”})$$

$$\lambda \geq 0, \; \mu \geq 0 \qquad (\text{“Dual Feasibility”})$$

$$\lambda \cdot (p^\top x - B) = 0 \qquad (\text{“Complementary Slackness” 1})$$

$$\mu_i \cdot x_i = 0 \qquad (\text{“Complementary Slackness” 2}).$$

Case 1. If the budget constraint is not binding, $p^\top x < B$

- $\lambda = 0$ and $\mu_i = 0, \forall\, i : x_i > 0$ (complementary slackness)
- For any $x_i > 0$, we must have: $\frac{\partial u}{\partial x_i} = -\mu_i = 0$
- The consumer purchases the unconstrained optimal amount of each good i

A Consumer's Constrained Consumption Problem

$$0 = -\frac{\partial u}{\partial x_i} + \lambda p_i - \mu_i,\ i = 1, \ldots, n \qquad \text{(“Stationarity”)}$$

$$p^{\mathsf{T}} x \leq B, \quad x \geq 0 \qquad \text{(“Primal Feasibility”)}$$

$$\lambda \geq 0,\ \mu \geq 0 \qquad \text{(“Dual Feasibility”)}$$

$$\lambda \cdot (p^{\mathsf{T}} x - B) = 0 \qquad \text{(“Complementary Slackness” 1)}$$

$$\mu_i \cdot x_i = 0 \qquad \text{(“Complementary Slackness” 2).}$$

Case 2.

- $p^{\mathsf{T}} x = B$, then can have $\lambda = 0$ or $\lambda > 0$.
- Case $\lambda > 0$:

$$i : x_i > 0 \quad \Rightarrow \mu_i = 0 \quad \Rightarrow \quad \frac{\partial u}{\partial x_i} = \lambda p_i \quad \Leftrightarrow \quad \frac{\frac{\partial u}{\partial x_i}}{p_i} = \lambda$$

$$i : x_i > 0,\ j : x_j = 0 \quad \Rightarrow \quad \frac{\frac{\partial u}{\partial x_i}}{x_i} = \lambda > \frac{\frac{\partial u}{\partial x_j}}{x_j} = \lambda - \mu_j$$

A Consumer's Constrained Consumption Problem

$$0 = -\frac{\partial u}{\partial x_i} + \lambda p_i - \mu_i, \; i = 1, \dots, n \qquad \text{("Stationarity")}$$

$$p^\mathsf{T} x \le B, \quad x \ge 0 \qquad \text{("Primal Feasibility")}$$

$$\lambda \ge 0, \; \mu \ge 0 \qquad \text{("Dual Feasibility")}$$

$$\lambda \cdot (p^\mathsf{T} x - B) = 0 \qquad \text{("Complementary Slackness" 1)}$$

$$\mu_i \cdot x_i = 0 \qquad \text{("Complementary Slackness" 2).}$$

Case 2.

- $p^\mathsf{T} x = B$, then can have $\lambda = 0$ or $\lambda > 0$.
- Case $\lambda > 0$:

$$i : x_i > 0 \quad \Rightarrow \mu_i = 0 \quad \Rightarrow \quad \frac{\partial u}{\partial x_i} = \lambda p_i \quad \Leftrightarrow \quad \frac{\frac{\partial u}{\partial x_i}}{p_i} = \lambda$$

$$i : x_i > 0, \; j : x_j = 0 \quad \Rightarrow \quad \frac{\frac{\partial u}{\partial x_i}}{x_i} = \lambda > \frac{\frac{\partial u}{\partial x_j}}{x_j} = \lambda - \mu_j$$

- **Bang-for-the-buck** $\frac{\frac{\partial u}{\partial x_i}}{x_i}$ for all **consumed** goods ($x_i > 0$) must be the same, and larger than for **unconsumed** goods

Fenchel Duality

- Elegant and concise theory of optimization duality

Fenchel Duality

- Elegant and concise theory of optimization duality

Conjugate of a function

Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$. The **conjugate** of f is the function $f^* : \mathbb{R}^n \rightarrow \mathbb{R}$ defined as:

$$f^*(y) = \sup_{x \in \mathrm{dom}(f)} \left\{ y^\mathsf{T} x - f(x) \right\}$$

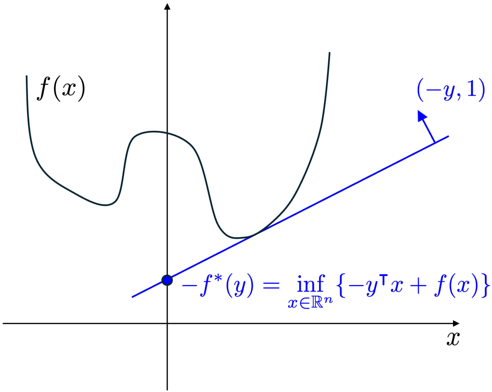
Fenchel Duality

- Elegant and concise theory of optimization duality

Conjugate of a function

Let $f : \mathbb{R}^n \to \mathbb{R}$. The **conjugate** of f is the function $f^* : \mathbb{R}^n \to \mathbb{R}$ defined as:

$$f^*(y) = \sup_{x \in \mathrm{dom}(f)} \{y^\intercal x - f(x)\}$$



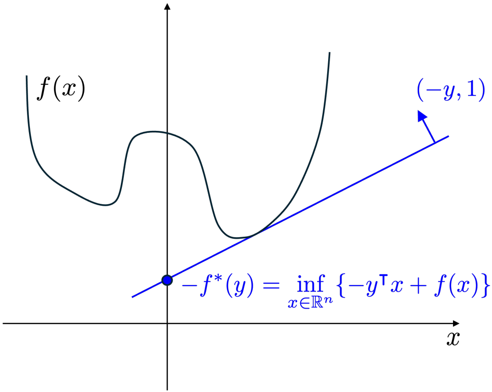
Fenchel Duality

- Elegant and concise theory of optimization duality

Conjugate of a function

Let $f : \mathbb{R}^n \to \mathbb{R}$. The **conjugate** of f is the function $f^* : \mathbb{R}^n \to \mathbb{R}$ defined as:

$$f^*(y) = \sup_{x \in \mathrm{dom}(f)} \{y^\intercal x - f(x)\}$$



- **Is f^* convex or concave?**

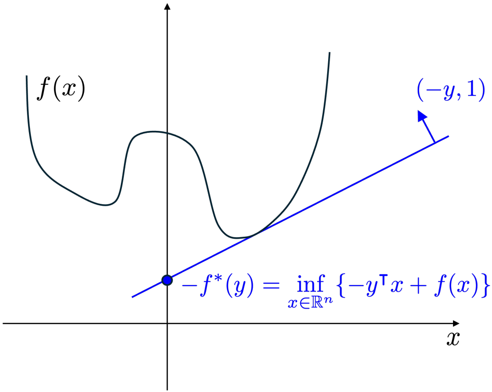
Fenchel Duality

- Elegant and concise theory of optimization duality

Conjugate of a function

Let $f : \mathbb{R}^n \to \mathbb{R}$. The **conjugate** of f is the function $f^* : \mathbb{R}^n \to \mathbb{R}$ defined as:

$$f^*(y) = \sup_{x \in \mathrm{dom}(f)} \{y^\intercal x - f(x)\}$$



- If f convex and epi(f) closed, f^* characterizes f in terms of supporting hyperplanes

Conjugates - Examples

$$f^*(y) = \sup_{x \in \text{dom}(f)} \{y^\mathsf{T} x - f(x)\}$$

The zero function.

For $f(x) = 0$, the conjugate will depend on the relevant domain:

- If $f : \mathbb{R} \to \mathbb{R}$, then
- If $f : \mathbb{R}_+ \to \mathbb{R}$, then
- If $f : [-1, 1] \to \mathbb{R}$, then
- If $f : [0, 1] \to \mathbb{R}$, then

Conjugates - Examples

$$f^*(y) = \sup_{x \in \mathrm{dom}(f)} \{y^\mathsf{T} x - f(x)\}$$

The zero function.

For $f(x) = 0$, the conjugate will depend on the relevant domain:

- If $f : \mathbb{R} \to \mathbb{R}$, then $f^* : \{0\} \to \mathbb{R}$ and $f^*(y) = 0$.
- If $f : \mathbb{R}_+ \to \mathbb{R}$, then
- If $f : [-1, 1] \to \mathbb{R}$, then
- If $f : [0, 1] \to \mathbb{R}$, then

Conjugates - Examples

$$f^*(y) = \sup_{x \in \text{dom}(f)} \{y^T x - f(x)\}$$

The zero function.

For $f(x) = 0$, the conjugate will depend on the relevant domain:

- If $f : \mathbb{R} \to \mathbb{R}$, then $f^* : \{0\} \to \mathbb{R}$ and $f^*(y) = 0$.
- If $f : \mathbb{R}_+ \to \mathbb{R}$, then $f^* : (-\infty, 0] \to \mathbb{R}$ and $f^*(y) = 0$.
- If $f : [-1, 1] \to \mathbb{R}$, then
- If $f : [0, 1] \to \mathbb{R}$, then

Conjugates - Examples

$$f^*(y) = \sup_{x \in \text{dom}(f)} \{y^\mathsf{T} x - f(x)\}$$

The zero function.

For $f(x) = 0$, the conjugate will depend on the relevant domain:

- If $f : \mathbb{R} \to \mathbb{R}$, then $f^* : \{0\} \to \mathbb{R}$ and $f^*(y) = 0$.
- If $f : \mathbb{R}_+ \to \mathbb{R}$, then $f^* : (-\infty, 0] \to \mathbb{R}$ and $f^*(y) = 0$.
- If $f : [-1, 1] \to \mathbb{R}$, then $f^* : \mathbb{R} \to \mathbb{R}$ and $f^*(y) = |y|$.
- If $f : [0, 1] \to \mathbb{R}$, then

Conjugates - Examples

$$f^*(y) = \sup_{x \in \text{dom}(f)} \{y^\mathsf{T} x - f(x)\}$$

The zero function.

For $f(x) = 0$, the conjugate will depend on the relevant domain:

- If $f : \mathbb{R} \to \mathbb{R}$, then $f^* : \{0\} \to \mathbb{R}$ and $f^*(y) = 0$.
- If $f : \mathbb{R}_+ \to \mathbb{R}$, then $f^* : (-\infty, 0] \to \mathbb{R}$ and $f^*(y) = 0$.
- If $f : [-1, 1] \to \mathbb{R}$, then $f^* : \mathbb{R} \to \mathbb{R}$ and $f^*(y) = |y|$.
- If $f : [0, 1] \to \mathbb{R}$, then $f^* : \mathbb{R} \to \mathbb{R}$ and $f^*(y) = y^+$.

Conjugates - Examples

$$f^*(y) = \sup_{x \in \text{dom}(f)} \{y^\mathsf{T} x - f(x)\}$$

The zero function.

For $f(x) = 0$, the conjugate will depend on the relevant domain:

- If $f : \mathbb{R} \to \mathbb{R}$, then $f^* : \{0\} \to \mathbb{R}$ and $f^*(y) = 0$.
- If $f : \mathbb{R}_+ \to \mathbb{R}$, then $f^* : (-\infty, 0] \to \mathbb{R}$ and $f^*(y) = 0$.
- If $f : [-1, 1] \to \mathbb{R}$, then $f^* : \mathbb{R} \to \mathbb{R}$ and $f^*(y) = |y|$.
- If $f : [0, 1] \to \mathbb{R}$, then $f^* : \mathbb{R} \to \mathbb{R}$ and $f^*(y) = y^+$.

Affine functions.

For $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = a^\mathsf{T} x + b$, $f^* : \{a\} \to \mathbb{R}$ and $f^*(a) = -b$.

Conjugates - Examples

$$f^*(y) = \sup_{x \in \mathrm{dom}(f)} \{y^\mathsf{T} x - f(x)\}$$

The zero function.

For $f(x) = 0$, the conjugate will depend on the relevant domain:

- If $f : \mathbb{R} \to \mathbb{R}$, then $f^* : \{0\} \to \mathbb{R}$ and $f^*(y) = 0$.
- If $f : \mathbb{R}_+ \to \mathbb{R}$, then $f^* : (-\infty, 0] \to \mathbb{R}$ and $f^*(y) = 0$.
- If $f : [-1, 1] \to \mathbb{R}$, then $f^* : \mathbb{R} \to \mathbb{R}$ and $f^*(y) = |y|$.
- If $f : [0, 1] \to \mathbb{R}$, then $f^* : \mathbb{R} \to \mathbb{R}$ and $f^*(y) = y^+$.

Affine functions.

For $f : \mathbb{R} \to \mathbb{R}$ with $f(x) = a^\mathsf{T} x + b$, $f^* : \{a\} \to \mathbb{R}$ and $f^*(a) = -b$.

What are the conjugates of the following functions?

- $f : (0, \infty), f(x) = -\log x$
- $f : \mathbb{R} \to \mathbb{R}, f(x) = e^x$

Conjugate - Examples

Negative logarithm.

$f : (0, \infty) \to \mathbb{R}$ with $f(x) = -\log x$.

$yx + \log x$ is unbounded above if $y \geq 0$ and reaches its maximum at $x = -1/y$ otherwise. Therefore, $f^* : (-\infty, 0) \to \mathbb{R}$ and $f^*(y) = -\log(-y) - 1$ for $y < 0$.

Conjugate - Examples

Negative logarithm.

$f : (0, \infty) \to \mathbb{R}$ with $f(x) = -\log x$.

$yx + \log x$ is unbounded above if $y \geq 0$ and reaches its maximum at $x = -1/y$ otherwise. Therefore, $f^* : (-\infty, 0) \to \mathbb{R}$ and $f^*(y) = -\log(-y) - 1$ for $y < 0$.

Exponential.

$f : \mathbb{R} \to \mathbb{R}, f(x) = e^x$.

$yx - e^x$ is unbounded if $y < 0$. For $y > 0$, $yx - e^x$ reaches its maximum at $x = \log y$, so we have $f^*(y) = y \log y - y$. For $y = 0$,

$$f^*(y) = \sup_x -e^x = 0.$$

In summary, $f^* : \mathbb{R}_+ \to \mathbb{R}$ and

$$f^*(y) = \begin{cases} y \log y - y & y > 0 \\ 0 & y = 0. \end{cases} \quad (1)$$

Fenchel-Young Inequality

Consider the Fenchel conjugate f^* of a function f:

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \{y^T x - f(x)\}, \quad y \in \mathbb{R}^n.$$

Fenchel-Young Inequality

Consider the Fenchel conjugate f^* of a function f:

$$f^*(y) = \sup_{x \in \mathbb{R}^n} \{y^T x - f(x)\}, \quad y \in \mathbb{R}^n.$$

Fenchel-Young Inequality

$$f^*(y) \geq y^\mathsf{T} x - f(x)$$

- Having access to f^* allows generating lower bounds on $f(x) \geq y^\mathsf{T} x - f^*(y)$

Double Conjugate and Convex Envelope

Consider **the conjugate of the conjugate**, a.k.a. the **double conjugate**, f^{**}:

$$f^{**}(x) = \sup_{y \in \mathbb{R}^n} \{y^T x - f^*(y)\}, \quad x \in \mathbb{R}^n.$$

Double Conjugate and Convex Envelope

Consider **the conjugate of the conjugate**, a.k.a. the **double conjugate**, f^{**}:

$$f^{**}(x) = \sup_{y \in \mathbb{R}^n} \{y^T x - f^*(y)\}, \quad x \in \mathbb{R}^n.$$

Conjugacy Theorem.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be such that epi(f) is closed. Then:

a) $f(x) \geq f^{**}(x)$, *forall* $x \in \mathbb{R}^n$.

b) If f is convex, $f(x) = f^{**}(x)$, $\forall x \in \mathbb{R}^n$.

c) $f^{**}(x)$ is the **convex envelope of** f, i.e., epi(f^{**}) is the smallest closed, convex set containing epi(f).

Double Conjugate and Convex Envelope

Consider **the conjugate of the conjugate**, a.k.a. the **double conjugate**, f^{**}:

$$f^{**}(x) = \sup_{y \in \mathbb{R}^n} \{y^T x - f^*(y)\}, \quad x \in \mathbb{R}^n.$$

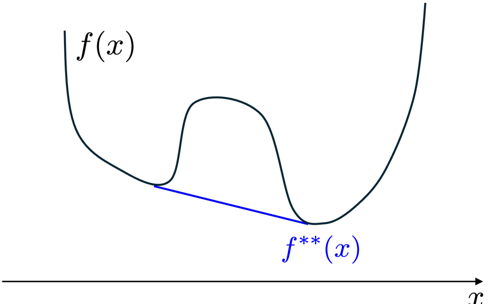
Conjugacy Theorem.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be such that epi(f) is closed. Then:

a) $f(x) \geq f^{**}(x)$, *forall* $x \in \mathbb{R}^n$.

b) If f is convex, $f(x) = f^{**}(x)$, $\forall\, x \in \mathbb{R}^n$.

c) $f^{**}(x)$ is the **convex envelope of** f, i.e., epi(f^{**}) is the smallest closed, convex set containing epi(f).



Double Conjugate and Convex Envelope

Consider **the conjugate of the conjugate**, a.k.a. the **double conjugate**, f^{**}:

$$f^{**}(x) = \sup_{y \in \mathbb{R}^n} \{y^T x - f^*(y)\}, \quad x \in \mathbb{R}^n.$$

Conjugacy Theorem.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be such that epi(f) is closed. Then:

a) $f(x) \geq f^{**}(x)$, *forall* $x \in \mathbb{R}^n$.

b) If f is convex, $f(x) = f^{**}(x)$, $\forall\, x \in \mathbb{R}^n$.

c) $f^{**}(x)$ is the **convex envelope of** f, i.e., epi(f^{**}) is the smallest closed, convex set containing epi(f).

- The optimal value when minimizing an **arbitrary** f – if finite – equals the optimal value when minimizing the convex envelope of f
- **IF** we had access to f^{**}, we could solve a convex optimization problem to determine the optimal value of any function f
- **Key caveat:** Gaining access to f^{**} is difficult for general f!

Fenchel Duality

Starting Problem.

Consider $f_i : \mathbb{R}^n \to \mathbb{R}$ and $X_i \subseteq \mathbb{R}^n$ for $i = 1, 2$ and the problem:

$$\text{minimize } f_1(x) + f_2(x)$$
$$\text{subject to } x \in X_1 \cap X_2$$

- Assume optimal value $p^\star$ is finite. Problem can be converted into:

Fenchel Duality

Starting Problem.

Consider $f_i : \mathbb{R}^n \to \mathbb{R}$ and $X_i \subseteq \mathbb{R}^n$ for $i = 1, 2$ and the problem:

$$\text{minimize } f_1(x) + f_2(x)$$
$$\text{subject to } x \in X_1 \cap X_2$$

- Assume optimal value $p^\star$ is finite. Problem can be converted into:

$$\text{minimize } f_1(y) + f_2(z)$$
$$\text{subject to } y = z,\ y \in X_1,\ z \in X_2.$$

Fenchel Duality

Starting Problem.

Consider $f_i : \mathbb{R}^n \to \mathbb{R}$ and $X_i \subseteq \mathbb{R}^n$ for $i = 1, 2$ and the problem:

$$\begin{aligned} \text{minimize } & f_1(x) + f_2(x) \\ \text{subject to } & x \in X_1 \cap X_2 \end{aligned}$$

- Assume optimal value $p^\star$ is finite. Problem can be converted into:

$$\begin{aligned} \text{minimize } & f_1(y) + f_2(z) \\ \text{subject to } & y = z,\ y \in X_1,\ z \in X_2. \end{aligned}$$

- Can dualize the constraint $y = z$. For $\lambda \in \mathbb{R}^n$, define the following functions:

Fenchel Duality

Starting Problem.

Consider $f_i : \mathbb{R}^n \to \mathbb{R}$ and $X_i \subseteq \mathbb{R}^n$ for $i = 1, 2$ and the problem:

$$\begin{aligned} \text{minimize } & f_1(x) + f_2(x) \\ \text{subject to } & x \in X_1 \cap X_2 \end{aligned}$$

- Assume optimal value $p^\star$ is finite. Problem can be converted into:

$$\begin{aligned} \text{minimize } & f_1(y) + f_2(z) \\ \text{subject to } & y = z,\ y \in X_1,\ z \in X_2. \end{aligned}$$

- Can dualize the constraint $y = z$. For $\lambda \in \mathbb{R}^n$, define the following functions:

$$\begin{aligned} g(\lambda) &= \inf_{y \in X_1, z \in X_2} \{f_1(y) + f_2(z) + (z - y)^\mathsf{T}\lambda\} \\ &= -\sup_{y \in X_1} \{y^\mathsf{T}\lambda - f_1(y)\} - \sup_{z \in X_2} \{-z^\mathsf{T}\lambda - f_2(z)\} \\ &= -g_1(\lambda) - g_2(-\lambda), \end{aligned}$$

- **What are $g_1(\lambda)$ and $g_2(\lambda)$ here?**

Fenchel Duality

Starting Problem.

Consider $f_i : \mathbb{R}^n \to \mathbb{R}$ and $X_i \subseteq \mathbb{R}^n$ for $i = 1, 2$ and the problem:

$$\begin{aligned} &\text{minimize } f_1(x) + f_2(x) \\ &\text{subject to } x \in X_1 \cap X_2 \end{aligned}$$

- Assume optimal value $p^\star$ is finite. Problem can be converted into:

$$\begin{aligned} &\text{minimize } f_1(y) + f_2(z) \\ &\text{subject to } y = z,\ y \in X_1,\ z \in X_2. \end{aligned}$$

- Can dualize the constraint $y = z$. For $\lambda \in \mathbb{R}^n$, define the following functions:

$$\begin{aligned} g(\lambda) &= \inf_{y \in X_1, z \in X_2} \{f_1(y) + f_2(z) + (z - y)^\mathsf{T}\lambda\} \\ &= -\sup_{y \in X_1} \{y^\mathsf{T}\lambda - f_1(y)\} - \sup_{z \in X_2} \{-z^\mathsf{T}\lambda - f_2(z)\} \\ &= -g_1(\lambda) - g_2(-\lambda), \end{aligned}$$

- **What are $g_1(\lambda)$ and $g_2(\lambda)$ here?**
- $g_i(\lambda)$ is the conjugate of $f_i(x), i = 1, 2$

Fenchel Duality

Starting Problem.

Consider $f_i : \mathbb{R}^n \to \mathbb{R}$ and $X_i \subseteq \mathbb{R}^n$ for $i = 1, 2$ and the problem:

$$\begin{aligned} &\text{minimize } f_1(x) + f_2(x) \\ &\text{subject to } x \in X_1 \cap X_2 \end{aligned}$$

- Dual objective is: $g(\lambda) = -g_1(\lambda) - g_2(-\lambda)$
- The dual problem can be rewritten as:

$$\max_{\lambda \in \mathbb{R}^n} \{-g_1(\lambda) - g_2(-\lambda)\} \qquad \Leftrightarrow \qquad \min_{\lambda \in \mathbb{R}^n} \{g_1(\lambda) + g_2(-\lambda)\}.$$

Fenchel Duality

Starting Problem.

Consider $f_i : \mathbb{R}^n \to \mathbb{R}$ and $X_i \subseteq \mathbb{R}^n$ for $i = 1, 2$ and the problem:

$$\begin{aligned} &\text{minimize } f_1(x) + f_2(x) \\ &\text{subject to } x \in X_1 \cap X_2 \end{aligned}$$

- Dual objective is: $g(\lambda) = -g_1(\lambda) - g_2(-\lambda)$
- The dual problem can be rewritten as:

$$\max_{\lambda \in \mathbb{R}^n} \{-g_1(\lambda) - g_2(-\lambda)\} \qquad \Leftrightarrow \qquad \min_{\lambda \in \mathbb{R}^n} \{g_1(\lambda) + g_2(-\lambda)\}.$$

Fenchel Duality

Suppose f_1 and f_2 are convex and **either**

(i) $\mathrm{rel\,int}(\mathrm{dom}(f_1)) \cap \mathrm{rel\,int}(\mathrm{dom}(f_2) \neq \emptyset$

or

(ii) $\mathrm{dom}(f_i)$ is polyhedral and f_i can be extended to $\mathbb{R}$-valued convex functions over $\mathbb{R}^n$ for $i = 1, 2$.

Then, there exists $\lambda^\star \in \mathbb{R}^n$ such that $p^\star = g(\lambda^\star)$ and strong duality holds.