Lecture 12: KKT Optimality Conditions
Conjugacy and Fenchel Duality

Nov 4, 2024

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

- 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:

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

- 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 $x^\star$
  - **sufficient** conditions for the **local optimality** of $x^\star$
- Cannot expect **global optimality** of $x^\star$ 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 $x^\star$ optimal for $(\mathscr{P})$ and $\lambda^\star, \nu^\star$ optimal for $(\mathscr{D})$:

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

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

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

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

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 $x^\star$ optimal for $(\mathscr{P})$ and $\lambda^\star, \nu^\star$ optimal for $(\mathscr{D})$:

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

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 $x^\star$ optimal for $(\mathscr{P})$ and $\lambda^\star, \nu^\star$ optimal for $(\mathscr{D})$:

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

Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

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

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

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

$$
\lambda_j^\star > 0 \Rightarrow f_j(x_j^\star) = 0 \qquad , \qquad f_j(x_j^\star) < 0 \Rightarrow \lambda_j^\star = 0
$$

Complementary Slackness

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

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

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

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 $x^\star$ optimal for $(\mathscr{P})$ and $\lambda^\star, \nu^\star$ optimal for $(\mathscr{D})$:

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

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 $x^\star$ optimal for $(\mathscr{P})$ and $\lambda^\star, \nu^\star$ optimal for $(\mathscr{D})$:

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

Complementary Slackness

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

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

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

Complementary Slackness

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

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

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

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

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

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

- $x^\star \in X$, $\lambda^\star \in \mathbb{R}^m$ 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:

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

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

KKT Conditions

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

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

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

KKT Conditions

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

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

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

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

KKT Conditions

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

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

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

Karush-Kuhn-Tucker Optimality Conditions

Basic Optimization Problem

We will be concerned with the following optimization problem:

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

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

KKT Conditions

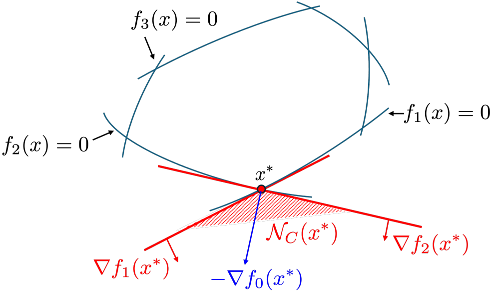
$$\begin{array}{ll} 0 = \nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \cdot \nabla f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star \cdot \nabla h_i(x^\star), & (\text{“Stationarity”}) \\ f_i(x^\star) \leq 0, \ i = 1, \ldots, m; \quad h_i(x^\star) = 0, \ i = 1, \ldots, s, & (\text{“Primal Feasibility”})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

Geometry Behind KKT Conditions: Inequality Case

KKT Conditions For Case Without Equality Constraints

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

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

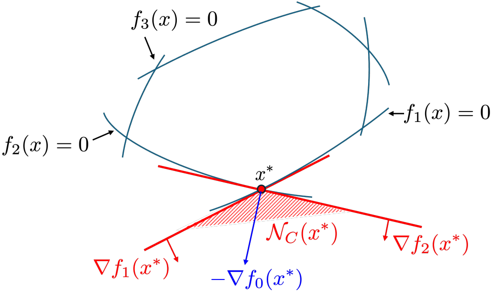


Geometry Behind KKT Conditions: Inequality Case

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

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

Failure of KKT Conditions

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

KKT Conditions Failing

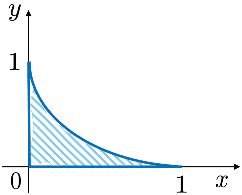
$$\min_{x \in \mathbb{R}} x$$
$$x^3 \geq 0.$$

- $f_0(x) = x$ and $f_1(x) = -x^3$
- Feasible set is $(-\infty, 0]$, the optimal solution is $x^\star = 0$.
- 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 $-\nabla f_0(x^\star) = \lambda \nabla f_1(x^\star)$.
- Note: **not** a convex optimization problem!

Failure of KKT Conditions - More Subtle

KKT Conditions Failing

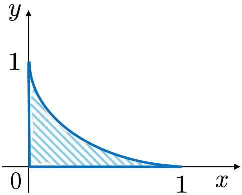
$$\min_{x,y\in\mathbb{R}} \quad -x$$
$$y-(1-x)^3 \leq 0$$
$$x,y \geq 0$$



Failure of KKT Conditions - More Subtle

KKT Conditions Failing

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



- $f_0(x,y) := -x$, $f_1(x,y) := y-(1-x)^3$, $f_2(x,y) := -x$ 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 $\lambda_1, \lambda_3 \ge 0$ 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 $f_1 \le 0$ around $(1,0)$ is $y \le 0$, which is parallel to the existing constraint $f_3(x,y) := -y \ge 0$

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

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

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

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

3. Regular Point (Linearly Independent Gradients)

- $x^\star$ is a **regular** point: gradients of active constraints $\{\nabla h_i(x)\}_{i=1}^s \cup \{\nabla f_j(x) : j \in I(x^\star)\}$ 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

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

3. Regular Point (Linearly Independent Gradients)

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

4. Mangasarian-Fromovitz

- the gradients of equality constraints are linearly independent
- $\exists v \in R^n : v^\intercal \nabla f_j(x^\star) < 0$ for $j \in I(x^\star)$ and $v^\intercal \nabla h_i(x^\star) = 0, i = 1, \ldots, s$

Second Order Necessary Conditions

Second Order **Necessary** Optimality Conditions

$x^\star$ 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 $x^\star$. 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

$x^\star$ 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 $x^\star$. 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 $x^\star$ is locally optimal, then there exist unique $\lambda_i^\star \geq 0$ and $\mu_j^\star$ such that:

Second Order Necessary Conditions

Second Order **Necessary** Optimality Conditions

$x^\star$ 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 $x^\star$. 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 $x^\star$ is locally optimal, then there exist unique $\lambda_i^\star \geq 0$ and $\mu_j^\star$ such that:

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

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

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

Second Order Necessary Conditions

Second Order Necessary Optimality Conditions

$x^\star$ 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 $x^\star$. 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 $x^\star$ is locally optimal, then there exist unique $\lambda_i^\star \geq 0$ and $\mu_j^\star$ such that:

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

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

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

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

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

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 $(\mathscr{P})$ and **regular**, $f_0, f_1, \ldots, f_m, h_1, \ldots, h_s$ twice continuously differentiable in neighborhood of $x^\star$. 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 $\lambda_i^\star \geq 0$ and $\mu_j^\star$ such that

Second Order Sufficient Conditions

Second Order **Sufficient** Local Optimality Conditions

$x^\star$ 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 $x^\star$. 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 $\lambda_i^\star \geq 0$ and $\mu_j^\star$ such that

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

Second Order Sufficient Conditions

Second Order **Sufficient** Local Optimality Conditions

$x^\star$ 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 $x^\star$. 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 $\lambda_i^\star \geq 0$ and $\mu_j^\star$ such that

- $(\lambda^\star, \mu^\star)$ certify that $x^\star$ satisfies KKT conditions;
- The Hessian $\nabla_x^2 \mathcal{L}(x^\star; \lambda^\star, \mu^\star)$ 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 $x^\star$ **associated with positive Lagrange multipliers** $\lambda_i^\star$:

Second Order Sufficient Conditions

Second Order **Sufficient** Local Optimality Conditions

$x^\star$ 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 $x^\star$. 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 $\lambda_i^\star \geq 0$ and $\mu_j^\star$ such that

- $(\lambda^\star, \mu^\star)$ certify that $x^\star$ satisfies KKT conditions;
- The Hessian $\nabla_x^2 \mathcal{L}(x^\star; \lambda^\star, \mu^\star)$ 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 $x^\star$ **associated with positive Lagrange multipliers** $\lambda_i^\star$:

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

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

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

A Consumer's Constrained Consumption Problem

Second Order Sufficient Local Optimality Conditions

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:

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

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

A Consumer's Constrained Consumption Problem

Fenchel Duality

- Elegant and concise theory of optimization duality

Fenchel Duality

- Elegant and concise theory of optimization duality

Conjugate of a function

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

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

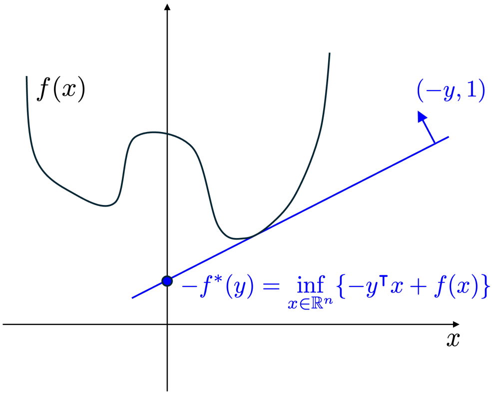
Fenchel Duality

- 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:

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



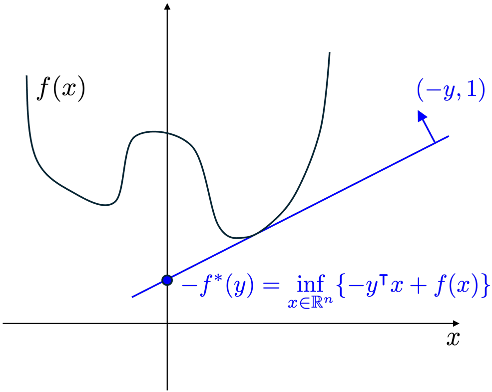
Fenchel Duality

- 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:

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



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

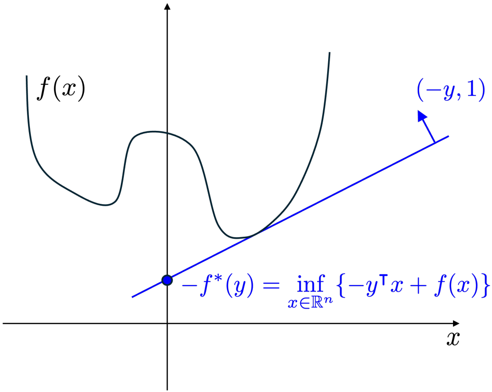
Fenchel Duality

- 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:

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



- f^* convex. When f closed and convex, f^* provides a description of f in terms of supporting hyperplanes!

Conjugates - Examples

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

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

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

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

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

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

Conjugates - Examples

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 $f(x) = a^\intercal x + b$, $f^* : \{a\} \to \mathbb{R}$ 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

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 $\text{epi}(f)$ is closed and let f^{**} be the double-conjugate.

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., $\text{epi}(f^{**})$ is the smallest closed, convex set containing $\text{epi}(f)$.

Double Conjugate and Convex Envelope

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

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

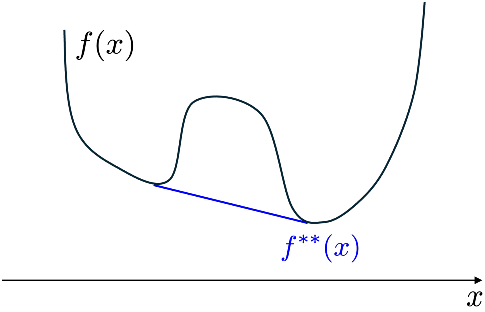
Conjugacy Theorem.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be such that $\mathrm{epi}(f)$ is closed and let f^{**} be the double-conjugate.

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., $\mathrm{epi}(f^{**})$ is the smallest closed, convex set containing $\mathrm{epi}(f)$.



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 and let f^{**} be the double-conjugate.

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 extremely 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:

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

- Assume optimal value is finite, $p^\star$. 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:

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

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

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

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 is finite, $p^\star$. Problem can be converted into:

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

- Can dualize the constraint $z = y$. 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 is finite, $p^\star$. Problem can be converted into:

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

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

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

- *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:

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

- 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:

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

- 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) the relative interiors of their domains intersect, i.e., $\mathrm{rel\,int}(\mathrm{dom}(f_1)) \cap \mathrm{rel\,int}(\mathrm{dom}(f_2) \neq \emptyset$ or
(ii) $\mathrm{dom}(f_i)$ is polyhedral and f_i can be extended to $\mathbb{R}$-valued convex function 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.