These are the slides of the lecture

Pattern Recognition
Winter term 2020/21
Friedrich-Alexander University of Erlangen-Nuremberg.

Erlangen, January 8, 2021
Prof. Dr.-Ing. Andreas Maier

Pattern Recognition (PR)

Prof. Dr.-Ing. Andreas Maier
Pattern Recognition Lab (CS 5), Friedrich-Alexander-Universität Erlangen-Nürnberg
Winter Term 2020/21

Duality in Optimization

The Primal Problem

- Consider the *primal optimization problem*:

$$\begin{array}{lll} \text{minimize} & f_0(\boldsymbol{x}) & \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, & i = 1, 2, \ldots, m \\ & h_i(\boldsymbol{x}) = 0, & i = 1, 2, \ldots, p \end{array}$$

  with variable $\boldsymbol{x} \in \mathbb{R}^n$.

- The function $f_0(\boldsymbol{x})$ is not required to be convex.

The Lagrangian

Lagrangian

The *Lagrangian L* of the aforementioned problem is defined as

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

The Lagrangian

Lagrangian

The *Lagrangian L* of the aforementioned problem is defined as

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

- λ_i is the Lagrange multipliers associated with the i-th inequality constraint $f_i(\boldsymbol{x}) \leq 0$.

The Lagrangian

Lagrangian

The *Lagrangian L* of the aforementioned problem is defined as

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

- λ_i is the Lagrange multipliers associated with the i-th inequality constraint $f_i(\boldsymbol{x}) \leq 0$.
- ν_i is the Lagrange multiplier associated with the i-th equality constraint $h_i(\boldsymbol{x}) = 0$.

The Lagrangian

Lagrangian

The *Lagrangian L* of the aforementioned problem is defined as

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

- λ_i is the Lagrange multipliers associated with the i-th inequality constraint $f_i(\boldsymbol{x}) \leq 0$.
- ν_i is the Lagrange multiplier associated with the i-th equality constraint $h_i(\boldsymbol{x}) = 0$.
- The vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\nu}$ are called *Lagrange multiplier vectors* or simply *dual variables*.

Langrange Dual Function

Lagrange dual function

The *Lagrange dual function* is defined as the infimum of the Lagrangian over $\boldsymbol{x}$

$$g(\boldsymbol{\lambda}, \nu) \quad = \quad \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu})$$

Langrange Dual Function

Lagrange dual function

The *Lagrange dual function* is defined as the infimum of the Lagrangian over $\boldsymbol{x}$

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

Langrange Dual Function

Lagrange dual function

The *Lagrange dual function* is defined as the infimum of the Lagrangian over $\boldsymbol{x}$

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

Note:

- The Lagrange dual function is a pointwise affine function in the dual variables.

Langrange Dual Function

Lagrange dual function

The *Lagrange dual function* is defined as the infimum of the Lagrangian over $\boldsymbol{x}$

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

Note:

- The Lagrange dual function is a pointwise affine function in the dual variables.
- The Lagrange dual function is concave (even if the original problem is not convex).

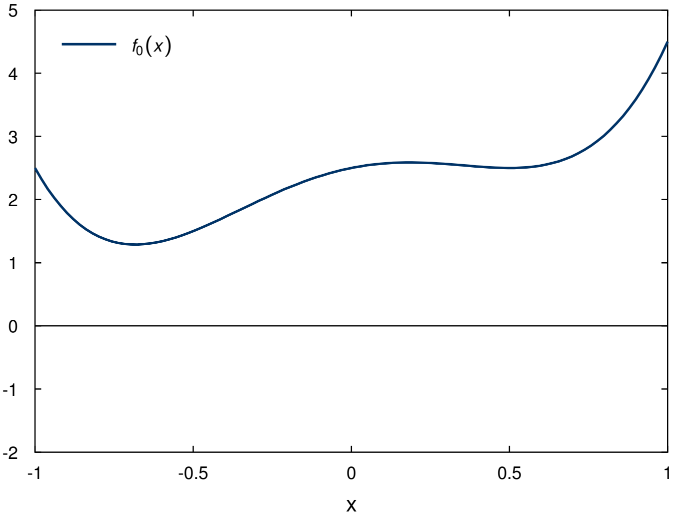
Optimal Value and Lower Bound

Lemma

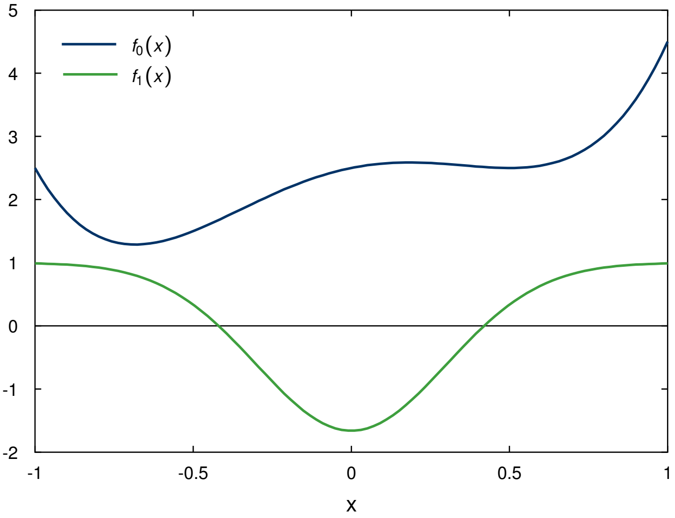
Let p^* be the optimal value of the optimization problem.
For any $\boldsymbol{\lambda} \succeq 0$ and any $\boldsymbol{\nu}$ the following bound is valid:

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

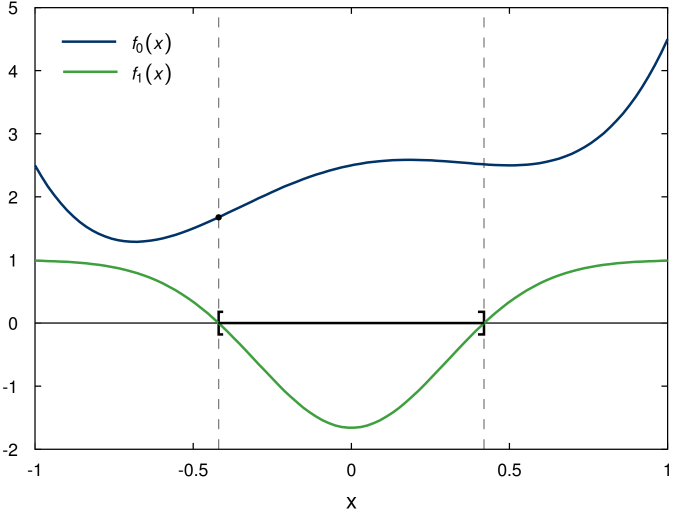
Optimal Value and Lower Bound (cont.)



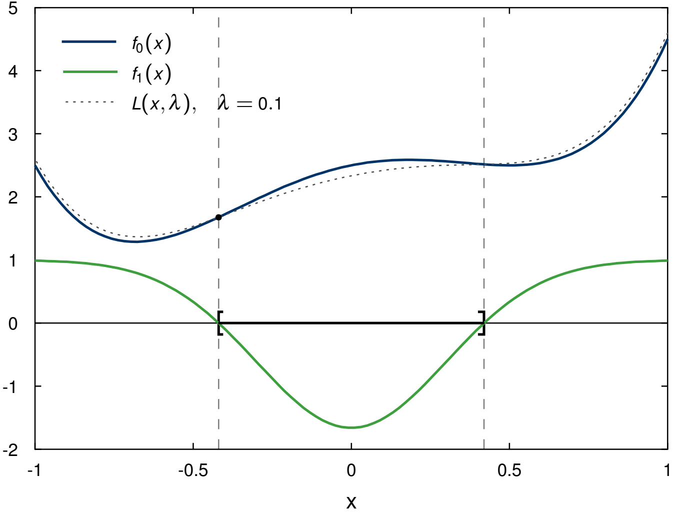
Optimal Value and Lower Bound (cont.)



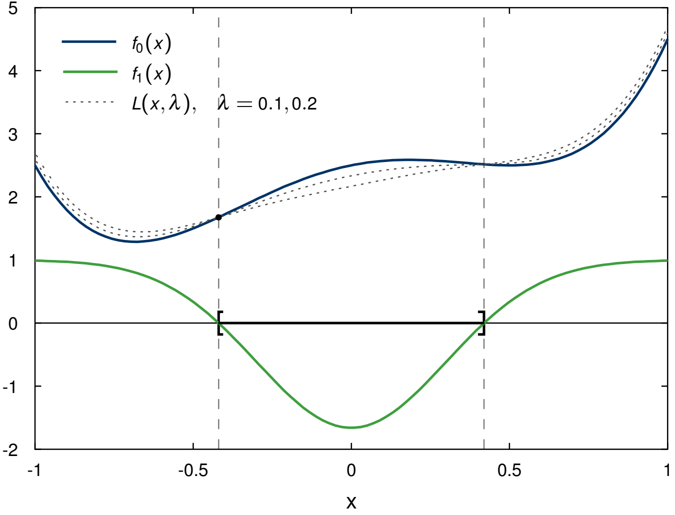
Optimal Value and Lower Bound (cont.)



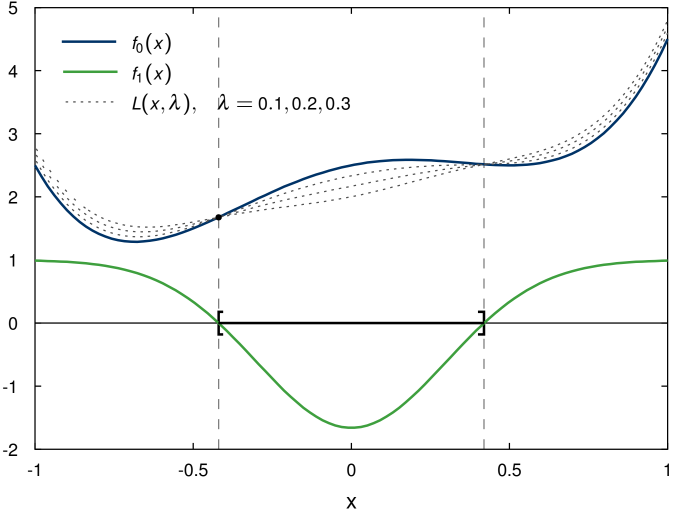
Optimal Value and Lower Bound (cont.)



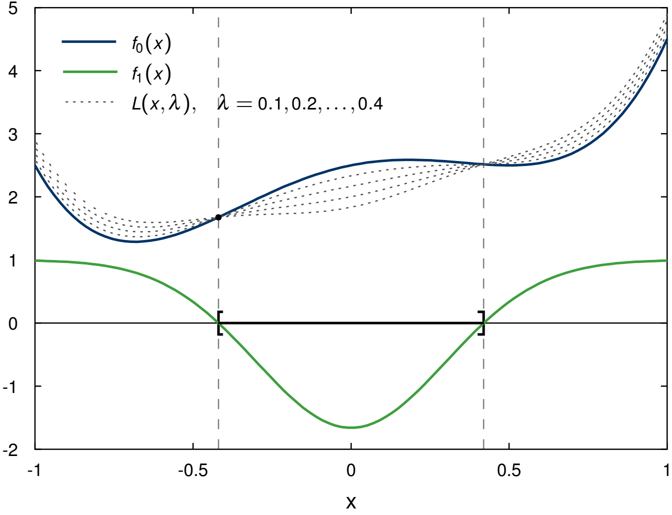
Optimal Value and Lower Bound (cont.)



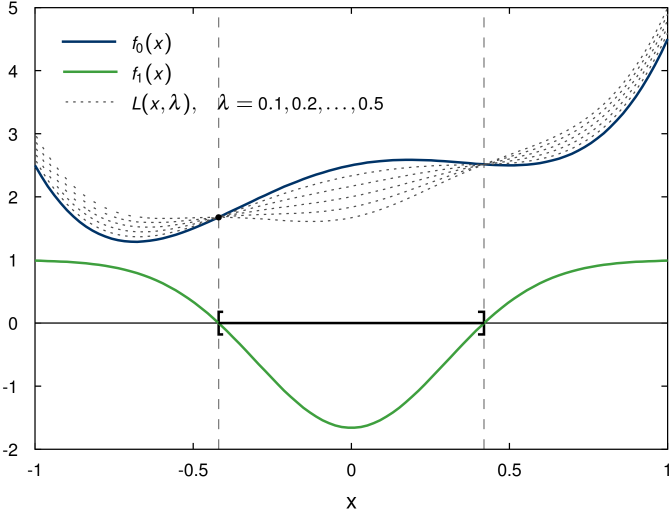
Optimal Value and Lower Bound (cont.)



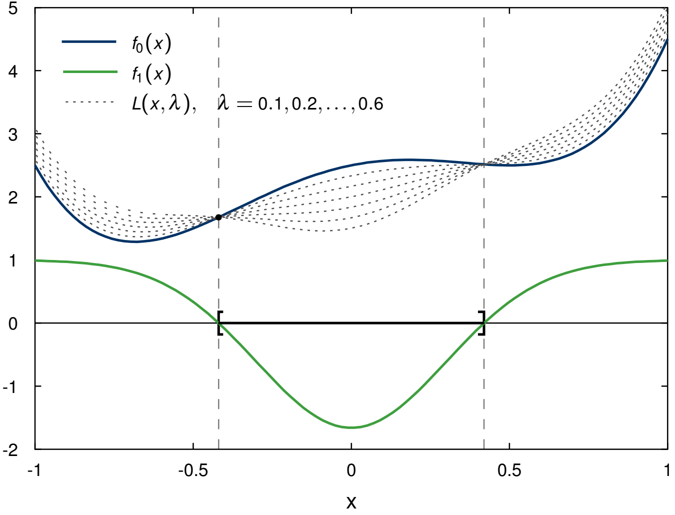
Optimal Value and Lower Bound (cont.)



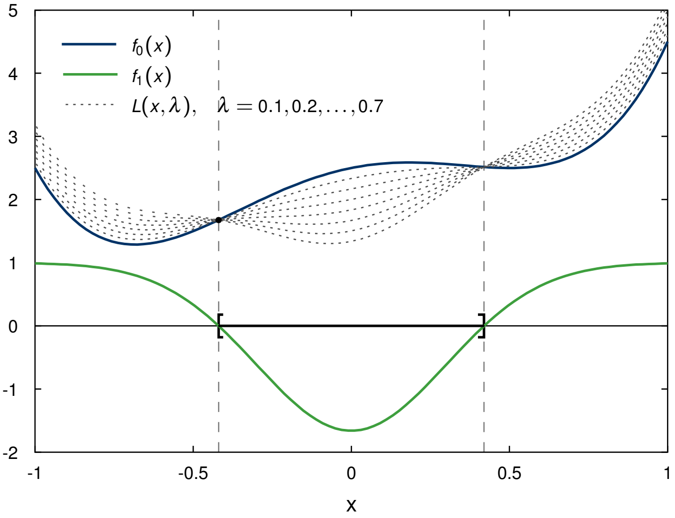
Optimal Value and Lower Bound (cont.)



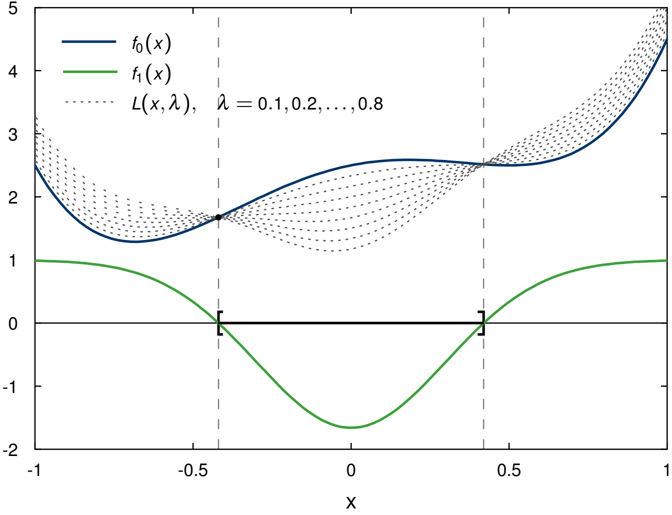
Optimal Value and Lower Bound (cont.)



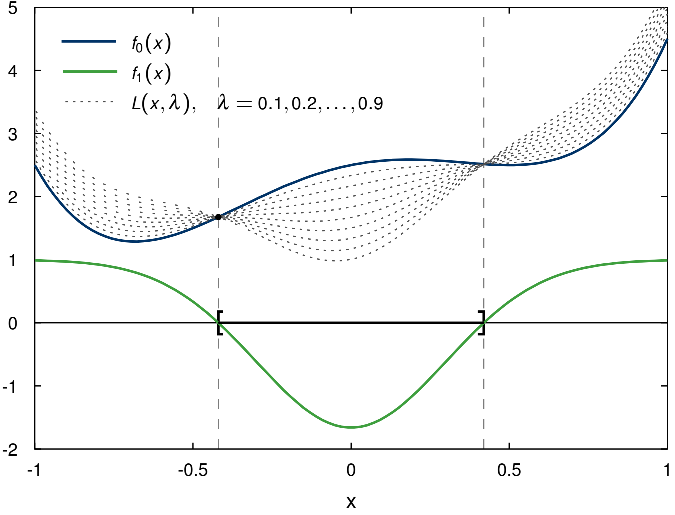
Optimal Value and Lower Bound (cont.)



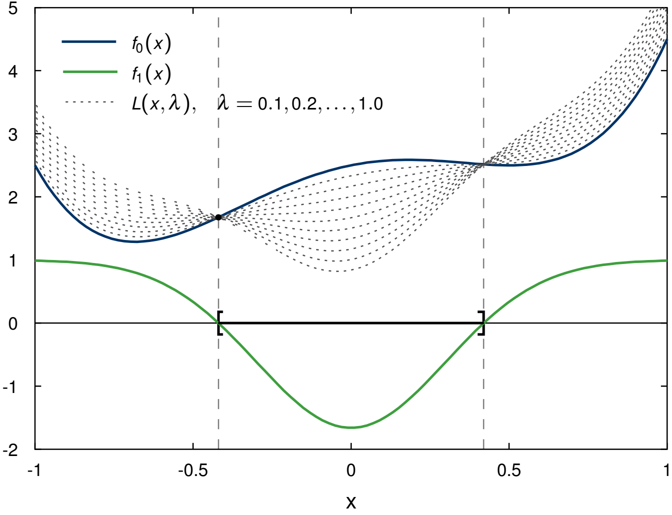
Optimal Value and Lower Bound (cont.)



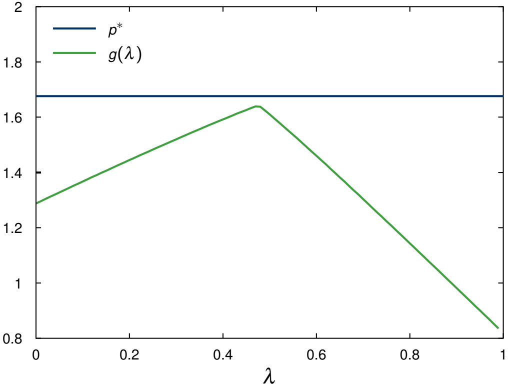
Optimal Value and Lower Bound (cont.)



Optimal Value and Lower Bound (cont.)



Optimal Value and Lower Bound (cont.)



- Neither $f_0(x)$ nor $f_1(x)$ is convex,
- but the dual function $g(\lambda)$ is concave!

Optimal Value and Lower Bound (cont.)

Let $\tilde{\boldsymbol{x}}$ be a feasible point of the optimization problem.

If $\boldsymbol{\lambda} \succeq 0$, we have due to the *m* inequality and *p* equality constraints:

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

Optimal Value and Lower Bound (cont.)

Let $\tilde{\boldsymbol{x}}$ be a feasible point of the optimization problem.

If $\boldsymbol{\lambda} \succeq 0$, we have due to the *m* inequality and *p* equality constraints:

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

Thus we have

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

Optimal Value and Lower Bound (cont.)

Using the definition of the dual function we get:

$$g(\boldsymbol{\lambda},\nu)=\inf_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\nu})\leq L(\tilde{\boldsymbol{x}},\boldsymbol{\lambda},\boldsymbol{\nu})\leq f_0(\tilde{\boldsymbol{x}})$$

Optimal Value and Lower Bound (cont.)

Using the definition of the dual function we get:

$$g(\boldsymbol{\lambda}, \nu) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq L(\tilde{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\tilde{\boldsymbol{x}})$$

The inequality $g(\boldsymbol{\lambda}, \nu) \leq f_0(\tilde{\boldsymbol{x}})$ holds for every feasible point $\tilde{\boldsymbol{x}}$.

Optimal Value and Lower Bound (cont.)

Using the definition of the dual function we get:

$$g(\boldsymbol{\lambda}, \nu) = \inf_{\boldsymbol{x}} L(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq L(\tilde{\boldsymbol{x}}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \leq f_0(\tilde{\boldsymbol{x}})$$

The inequality $g(\boldsymbol{\lambda}, \nu) \leq f_0(\tilde{\boldsymbol{x}})$ holds for every feasible point $\tilde{\boldsymbol{x}}$.

Consequently, the dual function $g(\boldsymbol{\lambda}, \nu)$ is also smaller or equal to the optimal value p^*:

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

$\square$

The Lagrange Dual Problem

Problem: how to find the best lower bound for the primal problem

The Lagrange dual problem is given by the optimization problem:

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

The Lagrange Dual Problem (cont.)

Optimal duality gap

Let p^* be the optimal value of the primal problem and
d^* the optimal value of the Lagrange dual problem.

The Lagrange Dual Problem (cont.)

Optimal duality gap

Let p^* be the optimal value of the primal problem and d^* the optimal value of the Lagrange dual problem.

- The difference $p^* - d^*$ is the *optimal duality gap*.

The Lagrange Dual Problem (cont.)

Optimal duality gap

Let p^* be the optimal value of the primal problem and d^* the optimal value of the Lagrange dual problem.

- The difference $p^* - d^*$ is the *optimal duality gap*.
- If $p^* = d^*$, the duality gap is zero.
  In this case we talk about *strong duality*.

The Lagrange Dual Problem (cont.)

Optimal duality gap

Let p^* be the optimal value of the primal problem and d^* the optimal value of the Lagrange dual problem.

- The difference $p^* - d^*$ is the *optimal duality gap*.
- If $p^* = d^*$, the duality gap is zero.
  In this case we talk about *strong duality*.
- If $p^* > d^*$, we have *weak duality*.

Slater's Condition

Theorem

Given a *convex* primal optimization problem:

$$\begin{array}{ll} \text{minimize} & f_0(\boldsymbol{x}) \\ \text{subject to} & f_i(\boldsymbol{x}) \leq 0, \quad i = 1, 2, \ldots, m \\ & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

with $f_0, f_1, \ldots, f_m$ being convex.

If there exists an $\boldsymbol{x} \in \text{relint}\left\{\mathscr{D} = \cap_{i=0}^{m} \text{dom}(f_i)\right\}$ with

$$\begin{array}{l} f_i(\boldsymbol{x}) < 0, \quad i = 1, \ldots, m \\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{array}$$

then strong duality holds.

Refinement of Slater's Condition

Theorem

Given a *convex* primal optimization problem.

If the first k constraint functions $f_1, \dots, f_k$ are *affine*, and if there exists an $\boldsymbol{x} \in \operatorname{relint} \mathscr{D}$ with

$$
\begin{aligned}
f_i(\boldsymbol{x}) &\le 0, \quad i = 1, \dots, k && \text{(affine constraints)} \\
f_i(\boldsymbol{x}) &< 0, \quad i = k+1, \dots, m && \text{(convex constraints)} \\
\boldsymbol{A}\boldsymbol{x} &= \boldsymbol{b}
\end{aligned}
$$

then strong duality holds.

Refinement of Slater's Condition

Theorem

Given a *convex* primal optimization problem.

If the first k constraint functions $f_1, \dots, f_k$ are *affine*, and
if there exists an $\boldsymbol{x} \in \operatorname{relint} \mathscr{D}$ with

$$
\begin{aligned}
f_i(\boldsymbol{x}) &\leq 0, \quad i = 1, \dots, k && \text{(affine constraints)} \\
f_i(\boldsymbol{x}) &< 0, \quad i = k+1, \dots, m && \text{(convex constraints)} \\
\boldsymbol{A}\boldsymbol{x} &= \boldsymbol{b}
\end{aligned}
$$

then strong duality holds.

Note: the refined Slater's condition reduces to feasibility when the constraints are all linear equalities and inequalities, and $\operatorname{dom}(f_0)$ is open.

Karush-Kuhn-Tucker Optimality Conditions

Let $\boldsymbol{x}^*$ be a primal and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ dual optimal points with zero duality gap.

For the primal optimal point $\boldsymbol{x}^*$, the gradient with respect to $\boldsymbol{x}$ of $L(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ is 0:

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

Karush-Kuhn-Tucker Optimality Conditions

The following four conditions are called KKT conditions:

Karush-Kuhn-Tucker Optimality Conditions

The following four conditions are called KKT conditions:

1. Primal constraints:
   - $f_i(\boldsymbol{x}) \leq 0, \quad i = 1, 2, \ldots, m$
   - $h_i(\boldsymbol{x}) = 0, \quad i = 1, 2, \ldots, p$

Karush-Kuhn-Tucker Optimality Conditions

The following four conditions are called KKT conditions:

1. Primal constraints:
   - $f_i(\boldsymbol{x}) \leq 0, \quad i = 1, 2, \ldots, m$
   - $h_i(\boldsymbol{x}) = 0, \quad i = 1, 2, \ldots, p$
2. Dual constraints: $\boldsymbol{\lambda} \succeq 0$

Karush-Kuhn-Tucker Optimality Conditions

The following four conditions are called KKT conditions:

1. Primal constraints:
   - $f_i(\boldsymbol{x}) \leq 0, \quad i = 1, 2, \dots, m$
   - $h_i(\boldsymbol{x}) = 0, \quad i = 1, 2, \dots, p$
2. Dual constraints: $\boldsymbol{\lambda} \succeq 0$
3. Complementary slackness: $\lambda_i f_i(\boldsymbol{x}) = 0$

Karush-Kuhn-Tucker Optimality Conditions

The following four conditions are called KKT conditions:

1. Primal constraints:
   - $f_i(\boldsymbol{x}) \leq 0, \quad i = 1, 2, \dots, m$
   - $h_i(\boldsymbol{x}) = 0, \quad i = 1, 2, \dots, p$
2. Dual constraints: $\boldsymbol{\lambda} \succeq 0$
3. Complementary slackness: $\lambda_i f_i(\boldsymbol{x}) = 0$
4. Gradient of the Lagrangian L is zero:

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

Karush-Kuhn-Tucker Optimality Conditions

The following four conditions are called KKT conditions:

1. Primal constraints:
   - $f_i(\boldsymbol{x}) \leq 0, \quad i = 1, 2, \ldots, m$
   - $h_i(\boldsymbol{x}) = 0, \quad i = 1, 2, \ldots, p$
2. Dual constraints: $\boldsymbol{\lambda} \succeq 0$
3. Complementary slackness: $\lambda_i f_i(\boldsymbol{x}) = 0$
4. Gradient of the Lagrangian L is zero:

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

If strong duality holds and if $\boldsymbol{x}^*$ and $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$ are optimal points, then the KKT conditions hold.

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

$$f_0(\boldsymbol{x}^*)$$

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

$$f_0(\boldsymbol{x}^*) \quad = \quad g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$$

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

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

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

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

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

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

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

$$
\begin{aligned}
f_0(\boldsymbol{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\boldsymbol{x}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}) \right) \\
&\leq f_0(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}^*) + \underbrace{\sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}^*)}_{=0}
\end{aligned}
$$

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

$$
\begin{aligned}
f_0(\boldsymbol{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\boldsymbol{x}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}) \right) \\
&\leq f_0(\boldsymbol{x}^*) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}^*) + \underbrace{\sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}^*)}_{=0}
\end{aligned}
$$

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

$$
\begin{aligned}
f_0(\boldsymbol{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\boldsymbol{x}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}) \right) \\
&\leq f_0(\boldsymbol{x}^*) + \underbrace{\boxed{\sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}^*)}}_{\leq 0} + \underbrace{\sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}^*)}_{=0}
\end{aligned}
$$

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

$$
\begin{aligned}
f_0(\boldsymbol{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\boldsymbol{x}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}) \right) \\
&\leq f_0(\boldsymbol{x}^*) + \underbrace{\sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}^*)}_{\leq 0} + \underbrace{\sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}^*)}_{= 0} \\
&\leq f_0(\boldsymbol{x}^*)
\end{aligned}
$$

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

$$
\begin{aligned}
\boxed{f_0(\boldsymbol{x}^*)} &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\boldsymbol{x}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}) \right) \\
&\leq f_0(\boldsymbol{x}^*) + \underbrace{\sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}^*)}_{\leq 0} + \underbrace{\sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}^*)}_{= 0} \\
&\leq \boxed{f_0(\boldsymbol{x}^*)}
\end{aligned}
$$

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

$$
\begin{aligned}
\boxed{f_0(\boldsymbol{x}^*)} &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\boldsymbol{x}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}) \right) \\
&\le f_0(\boldsymbol{x}^*) + \underbrace{\sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}^*)}_{\le 0} + \underbrace{\sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}^*)}_{=0} \\
&\overset{=}{\cancel{\le}} \boxed{f_0(\boldsymbol{x}^*)}
\end{aligned}
$$

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

$$
\begin{aligned}
\boxed{f_0(\boldsymbol{x}^*)} &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\boldsymbol{x}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}) \right) \\
&\overset{=}{\cancel{\leq}} f_0(\boldsymbol{x}^*) + \underbrace{\sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}^*)}_{\leq 0} + \underbrace{\sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}^*)}_{= 0} \\
&\overset{=}{\cancel{\leq}} \boxed{f_0(\boldsymbol{x}^*)}
\end{aligned}
$$

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

$$
\begin{aligned}
\boxed{f_0(\boldsymbol{x}^*)} &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\boldsymbol{x}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}) \right) \\
&\overset{=}{\cancel{\leq}} f_0(\boldsymbol{x}^*) + \underbrace{\sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}^*)}_{\overset{=}{\cancel{\leq}} 0} + \underbrace{\sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}^*)}_{= 0} \\
&\overset{=}{\cancel{\leq}} \boxed{f_0(\boldsymbol{x}^*)}
\end{aligned}
$$

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness

$$
\begin{aligned}
\boxed{f_0(\boldsymbol{x}^*)} &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\boldsymbol{x}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}) \right) \\
&\overset{=}{\cancel{\leq}} f_0(\boldsymbol{x}^*) + \underbrace{\boxed{\sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}^*)}}_{\overset{=}{\cancel{\leq}}\, 0} + \underbrace{\sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}^*)}_{=0} \\
&\overset{=}{\cancel{\leq}} \boxed{f_0(\boldsymbol{x}^*)}
\end{aligned}
$$

Karush-Kuhn-Tucker Optimality Conditions

Complementary slackness: $\lambda_i^* \cdot f_i(\boldsymbol{x}^*) = 0$

$$
\begin{aligned}
f_0(\boldsymbol{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*) \\
&= \inf_{\boldsymbol{x}} \left(f_0(\boldsymbol{x}) + \sum_{i=1}^{m} \lambda_i^* f_i(\boldsymbol{x}) + \sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}) \right) \\
&\overset{=}{\cancel{\leq}} f_0(\boldsymbol{x}^*) + \underbrace{\sum_{i=1}^{m} \underbrace{\lambda_i^* f_i(\boldsymbol{x}^*)}_{=0}}_{\overset{=}{\cancel{\leq}} 0} + \underbrace{\sum_{i=1}^{p} \nu_i^* h_i(\boldsymbol{x}^*)}_{=0} \\
&\overset{=}{\cancel{\leq}} f_0(\boldsymbol{x}^*)
\end{aligned}
$$

$\square$

Karush-Kuhn-Tucker Optimality Conditions (cont.)

Conclusions (Boyd 2004, Sec. 5.5.3)

- For *any* optimization problem with differentiable objective and constraint functions for which strong duality obtains, any pair of primal and dual optimal points must satisfy the KKT conditions.

Karush-Kuhn-Tucker Optimality Conditions (cont.)

Conclusions (Boyd 2004, Sec. 5.5.3)

- For *any* optimization problem with differentiable objective and constraint functions for which strong duality obtains, any pair of primal and dual optimal points must satisfy the KKT conditions.
- For any *convex* optimization problem with differentiable objective and and constraint functions, any points that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap.

Karush-Kuhn-Tucker Optimality Conditions (cont.)

Conclusions (Boyd 2004, Sec. 5.5.3)

- For *any* optimization problem with differentiable objective and constraint functions for which strong duality obtains, any pair of primal and dual optimal points must satisfy the KKT conditions.
- For any *convex* optimization problem with differentiable objective and and constraint functions, any points that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap.
- If a *convex* optimization problem with differentiable objective and constraint functions satisfies Slater's condition, then the KKT conditions provide necessary and sufficient conditions for optimality.

Lessons Learned

- Formalization of the primal problem using the Lagrangian
- Lagrange dual function
- Duality gap
- Karush-Kuhn-Tucker optimality conditions

Next Time in

Pattern Recognition

Further Readings

- S. Boyd, L. Vandenberghe:
  Convex Optimization,
  Cambridge University Press, 2004.
  ☞ http://www.stanford.edu/~boyd/cvxbook/
- Jorge Nocedal, Stephen Wright:
  Numerical Optimization,
  Springer, New York, 1999.

Comprehensive Questions

- What is the Lagrangian of a constrained objective function?
- What is the Lagrange dual function?
- What is the duality gap?
- What are the Karush-Kuhn-Tucker optimality conditions?