MS-E2122 - Nonlinear Optimization
Lecture VIII

Fernando Dias

Department of Mathematics and Systems Analysis

Aalto University
School of Science

October 19, 2023

Outline of this lecture

Outline of this lecture

The relaxation theorem

A relaxation is an alternative version of a given problem that presents exploitable properties.

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $S \subseteq \mathbb{R}^n$. Consider the following problem:

$$(P): \text{ min. } \{f(x) : x \in S\}$$

The relaxation theorem

A relaxation is an alternative version of a given problem that presents exploitable properties.

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $S \subseteq \mathbb{R}^n$. Consider the following problem:

$$(P): \text{ min. } \{f(x) : x \in S\}$$

A relaxation of P can be stated as:

$$(P_R): \text{ min. } \{f_R(x) : x \in S_R\}$$

where $f_R : \mathbb{R}^n \to \mathbb{R}$ and $S_R \subseteq \mathbb{R}^n$ and that the following holds.

The relaxation theorem

A relaxation is an alternative version of a given problem that presents exploitable properties.

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $S \subseteq \mathbb{R}^n$. Consider the following problem:

$$(P): \ \text{min.} \ \{f(x) : x \in S\}$$

A relaxation of P can be stated as:

$$(P_R): \ \text{min.} \ \{f_R(x) : x \in S_R\}$$

where $f_R : \mathbb{R}^n \to \mathbb{R}$ and $S_R \subseteq \mathbb{R}^n$ and that the following holds.

Definition 1 (Relaxation)

P_R is a relaxation of P if and only if:

1. $f_R(x) \leq f(x)$, for all $x \in S$;
2. $S \subseteq S_R$.

The relaxation theorem

Theorem 2 presents two fundamental results for using relaxations.

Theorem 2 (Relaxation theorem)

Let us define

$$(P): \text{ min. } \{f(x) : x \in S\} \quad \text{and} \quad (P_R): \text{min. } \{f_R(x) : x \in S_R\}$$

If P_R is a relaxation of P, then the following hold:

1. *if P_R is infeasible, so is P;*
2. *if $\overline{x}_R$ is an optimal solution to P_R such that $\overline{x}_R \in S$ and $f_R(\overline{x}_R) = f(\overline{x}_R)$, then $\overline{x}_R$ is optimal to P as well.*

The relaxation theorem

Theorem 2 presents two fundamental results for using relaxations.

Theorem 2 (Relaxation theorem)

Let us define

$$(P): \text{ min. } \{f(x) : x \in S\} \quad \text{and} \quad (P_R): \text{min. } \{f_R(x) : x \in S_R\}$$

If P_R is a relaxation of P, then the following hold:

1. *if P_R is infeasible, so is P;*
2. *if $\overline{x}_R$ is an optimal solution to P_R such that $\overline{x}_R \in S$ and $f_R(\overline{x}_R) = f(\overline{x}_R)$, then $\overline{x}_R$ is optimal to P as well.*

Proof.

Result 1 follows since $S \subseteq S_R$. To show Result 2, notice that $f(\overline{x}_R) = f_R(\overline{x}_R) \leq f_R(x) \leq f(x)$ for all $x \in S$. □

Lagrangian relaxation

Lagrangian duality is the body of theory supporting the use of Lagrangian relaxations to solve (primal) problems.

Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $X \subseteq \mathbb{R}^n$ be an open set. Define

$$
\begin{aligned}
(P): \quad \text{min.} \quad & f(x) \\
\text{subject to:} \quad & g(x) \leq 0 \\
& h(x) = 0 \\
& x \in X.
\end{aligned}
$$

Lagrangian relaxation

Lagrangian duality is the body of theory supporting the use of Lagrangian relaxations to solve (primal) problems.

Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^l$ and $X \subseteq \mathbb{R}^n$ be an open set. Define

$$\begin{aligned} (P):\ \text{min. } & f(x) \\ \text{subject to: } & g(x) \leq 0 \\ & h(x) = 0 \\ & x \in X. \end{aligned}$$

For a given set of dual variables $(u, v) \in \mathbb{R}^{m+l}$ with $u \geq 0$, the Lagrangian relaxation (or Lagrangian dual function) of P is

$$(D):\ \theta(u, v) = \inf_{x \in X} \ \phi(x, u, v);$$

$\phi(x, u, v) = f(x) + u^\top g(x) + v^\top h(x)$ is the Lagrangian function.

(Weak) Lagrangian duality

Theorem 3 (Weak Lagrangian duality)

Let x be a feasible solution to P, and let (u, v) with $u \geq 0$ be a feasible solution to D. Then $\theta(u, v) \leq f(x)$.

(Weak) Lagrangian duality

Theorem 3 (Weak Lagrangian duality)

Let x be a feasible solution to P, and let (u, v) with $u \geq 0$ be a feasible solution to D. Then $\theta(u, v) \leq f(x)$.

Proof.

From feasibility, $u \geq 0$, $g(x) \leq 0$ and $h(x) = 0$. Thus we have

$$\begin{aligned}\theta(u, v) &= \inf_{x \in X} \left\{ f(x) + u^\top g(x) + v^\top h(x) \right\} \\ &\leq f(x) + u^\top g(x) + v^\top h(x) \leq f(x).\end{aligned}$$

□

(Weak) Lagrangian duality

Theorem 3 (Weak Lagrangian duality)

Let x be a feasible solution to P, and let (u, v) with $u \geq 0$ be a feasible solution to D. Then $\theta(u, v) \leq f(x)$.

Proof.

From feasibility, $u \geq 0$, $g(x) \leq 0$ and $h(x) = 0$. Thus we have

$$\begin{aligned}\theta(u, v) &= \inf_{x \in X} \left\{ f(x) + u^\top g(x) + v^\top h(x) \right\} \\ &\leq f(x) + u^\top g(x) + v^\top h(x) \leq f(x).\end{aligned}$$

□

The Lagrangian dual problem D seeks optimal dual variables (u, v) such that $\theta(u, v)$ is as close as possible to $f(x)$, that is,

$$(D): \sup_{u,v} \{\theta(u, v) : u \geq 0\}.$$

(Weak) Lagrangian duality

Theorem 3 leads to two important corollaries:

Corollary 4 (Weak Lagrangian duality II)

$\sup_{u,v} \{\theta(u,v) : u \geq 0\} \leq \inf_x \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$

(Weak) Lagrangian duality

Theorem 3 leads to two important corollaries:

Corollary 4 (Weak Lagrangian duality II)

$\sup_{u,v} \{\theta(u,v) : u \geq 0\} \leq \inf_x \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$

Proof.

We have $\theta(u,v) \leq f(x)$ for any feasible x and (u,v), implying $\sup_{u,v} \{\theta(u,v) : u \geq 0\} \leq \inf_x \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$. □

(Weak) Lagrangian duality

Theorem 3 leads to two important corollaries:

Corollary 4 (Weak Lagrangian duality II)

$\sup_{u,v} \{\theta(u,v) : u \geq 0\} \leq \inf_x \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$

Proof.

We have $\theta(u,v) \leq f(x)$ for any feasible x and (u,v), implying
$\sup_{u,v} \{\theta(u,v) : u \geq 0\} \leq \inf_x \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$. □

Corollary 5 (one-way strong Lagrangian duality)

If $f(\overline{x}) = \theta(\overline{u}, \overline{v})$, $\overline{u} \geq 0$, *and* $\overline{x} \in \{x \in X : g(x) \leq 0, h(x) = 0\}$, *then* $\overline{x}$ *and* $(\overline{u}, \overline{v})$ *are optimal solutions to* P *and* D, *respectively.*

(Weak) Lagrangian duality

Theorem 3 leads to two important corollaries:

Corollary 4 (Weak Lagrangian duality II)

$\sup_{u,v} \{\theta(u,v) : u \geq 0\} \leq \inf_x \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$

Proof.

We have $\theta(u,v) \leq f(x)$ for any feasible x and (u,v), implying
$\sup_{u,v} \{\theta(u,v) : u \geq 0\} \leq \inf_x \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$. □

Corollary 5 (one-way strong Lagrangian duality)

If $f(\overline{x}) = \theta(\overline{u}, \overline{v})$, $\overline{u} \geq 0$, and $\overline{x} \in \{x \in X : g(x) \leq 0, h(x) = 0\}$, then $\overline{x}$ and $(\overline{u}, \overline{v})$ are optimal solutions to P and D, respectively.

Proof.

Use part 2 of Theorem 2 with D being a Lagrangian relaxation. □

Geometric interpretation of Lagrangian duality

Lagrangian duality has a geometric interpretation that helps understanding when strong duality can hold.

Geometric interpretation of Lagrangian duality

Lagrangian duality has a geometric interpretation that helps understanding when strong duality can hold.

Consider $(P):$ min. $\{f(x) : g(x) \leq 0, x \in X\}$ with one constraint.

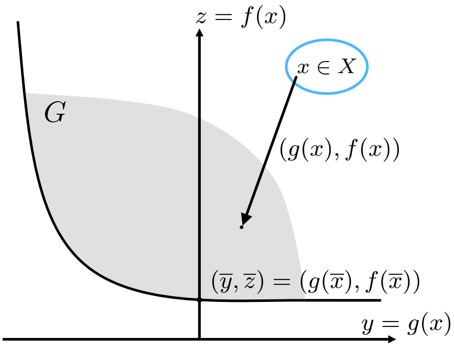
Geometric interpretation of Lagrangian duality

Lagrangian duality has a geometric interpretation that helps understanding when strong duality can hold.

Consider (P) : min. $\{f(x) : g(x) \leq 0, x \in X\}$ with one constraint.

Let $G = \{(y, z) : y = g(x), z = f(x), x \in X\}$ be defined in the (y, z)-plane.

- G is the image of X under the mapping (g, f).
- Solving P consists of finding (y, z) in G with $y \leq 0$ with minimum ordinate z.



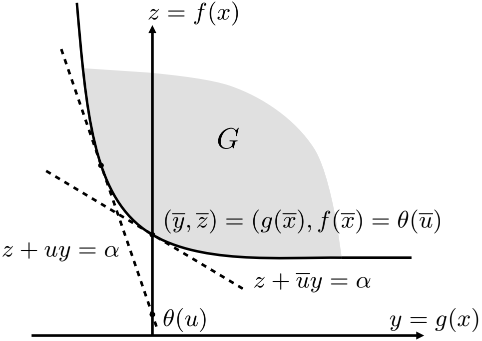
G is formed by all $x \in X$ under mapping $(g(x), f(x))$. $(\overline{y}, \overline{z})$ lowermost point on the z-axis.

Geometric interpretation of Lagrangian duality

Assume that $u \geq 0$ is given. $\theta(u) = \min_x \{f(x) + ug(x) : x \in X\}$ is given by the lowermost (y, z) in G attained at $z + uy = \alpha$.

Geometric interpretation of Lagrangian duality

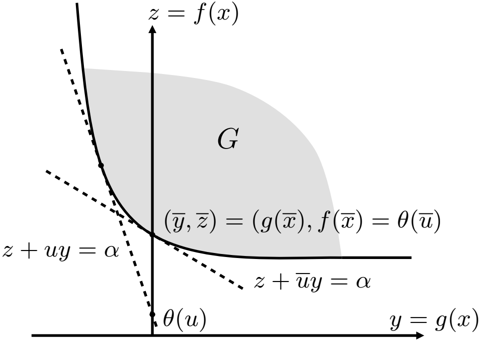
Assume that $u \geq 0$ is given. $\theta(u) = \min_x \{f(x) + ug(x) : x \in X\}$ is given by the lowermost (y, z) in G attained at $z + uy = \alpha$.



Optimal u such that $z = \alpha - uy$ is a supporting hyperplane of G with the uppermost intercept α; $z = \alpha - \overline{u}y$ supports G at $(\overline{y}, \overline{z})$.

Geometric interpretation of Lagrangian duality

Assume that $u \geq 0$ is given. $\theta(u) = \min_x \{f(x) + ug(x) : x \in X\}$ is given by the lowermost (y, z) in G attained at $z + uy = \alpha$.

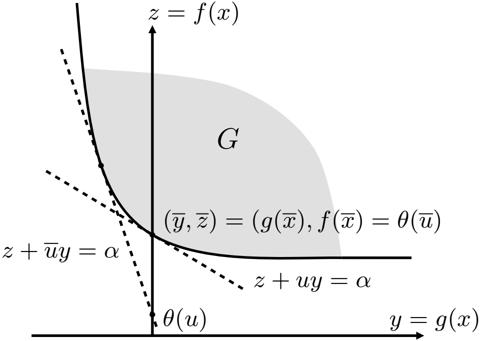


Optimal u such that $z = \alpha - uy$ is a supporting hyperplane of G with the uppermost intercept α; $z = \alpha - \overline{u}y$ supports G at $(\overline{y}, \overline{z})$.

Thus, solving D corresponds to finding the slope $-u$ for which the intercept $\alpha = \theta(u)$ on the z-axis is maximal.

Geometric interpretation of Lagrangian duality

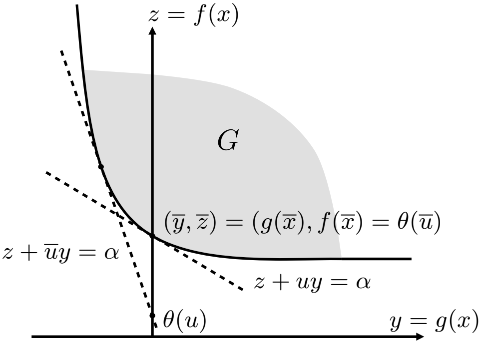
An important analytical tool in this context is the perturbation function $v(y) = \text{min. } \{f(x) : g(x) \leq y, x \in X\}$.



$v(y)$ is the greatest monotone nonincreasing lower envelope of G.

Geometric interpretation of Lagrangian duality

An important analytical tool in this context is the perturbation function $v(y) = \min. \{f(x) : g(x) \leq y, x \in X\}$.

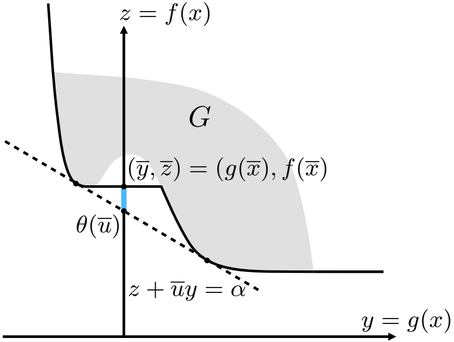


$v(y)$ is the greatest monotone nonincreasing lower envelope of G.

The fact that $v(y) \geq v(0) - \overline{u}y$ for all $y \in \mathbb{R}$ is why $f(\overline{x}) = \theta(\overline{u})$ match in this case, i.e., the optimal solutions of P and D coincide.

Duality gaps

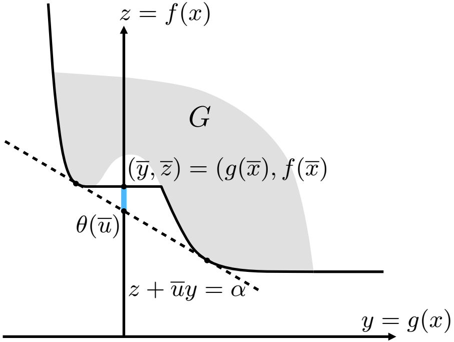
There will be a duality gap if for some $y \in \mathbb{R}$ there is no $\overline{u}$ for which $v(y) \geq v(0) - \overline{u}y$ holds.



$v(y)$ is not convex; the intercept of $z = \alpha - \overline{u}y$ and $(\overline{y}, \overline{z})$ do not match

Duality gaps

There will be a duality gap if for some $y \in \mathbb{R}$ there is no $\overline{u}$ for which $v(y) \geq v(0) - \overline{u}y$ holds.



$v(y)$ is not convex; the intercept of $z = \alpha - \overline{u}y$ and $(\overline{y}, \overline{z})$ do not match

Clearly, the minimum ordinate of (y, z) in G and the maximal z-axis intercept for slope $-u$ cannot match. Thus $f(\overline{x}) > \theta(\overline{u})$.

Duality gaps

Example 1: consider the problem:

$$\begin{aligned} \text{min. } & x_1^2 + x_2^2 \\ & x_1 + x_2 \geq 4 \\ & x_1, x_2 \geq 0. \end{aligned}$$

- The optimal point is $(\overline{x}_1, \overline{x}_2) = (2, 2)$.
- $f(x) = x_1^2 + x_2^2$, $g(x) = -x_1 - x_2 + 4$, $X = \{(x_1, x_2) : x_1, x_2 \geq 0\}$

Duality gaps

Example 1: consider the problem:

$$\begin{aligned} \text{min. } & x_1^2 + x_2^2 \\ & x_1 + x_2 \geq 4 \\ & x_1, x_2 \geq 0. \end{aligned}$$

- The optimal point is $(\overline{x}_1, \overline{x}_2) = (2, 2)$.
- $f(x) = x_1^2 + x_2^2$, $g(x) = -x_1 - x_2 + 4$, $X = \{(x_1, x_2) : x_1, x_2 \geq 0\}$

The Lagrangian dual function is given by:

$$\begin{aligned} \theta(u) &= \inf \left\{x_1^2 + x_2^2 + u(-x_1 - x_2 + 4) : x_1, x_2 \geq 0\right\} \\ &= \inf \left\{x_1^2 - ux_1 : x_1 \geq 0\right\} + \inf \left\{x_2^2 - ux_2 : x_2 \geq 0\right\} + 4u \\ &= \begin{cases} -1/2u^2 + 4u, & \text{if } u \geq 0 \\ -4u, \ \ 1 + 4u^2 & \text{if } u < 0. \end{cases} \end{aligned}$$

Duality gaps

Example 1: consider the problem:

$$\begin{aligned} \text{min. } & x_1^2 + x_2^2 \\ & x_1 + x_2 \geq 4 \\ & x_1, x_2 \geq 0. \end{aligned}$$

- The optimal point is $(\overline{x}_1, \overline{x}_2) = (2, 2)$.
- $f(x) = x_1^2 + x_2^2$,
  $g(x) = -x_1 - x_2 + 4$,
  $X = \{(x_1, x_2) : x_1, x_2 \geq 0\}$

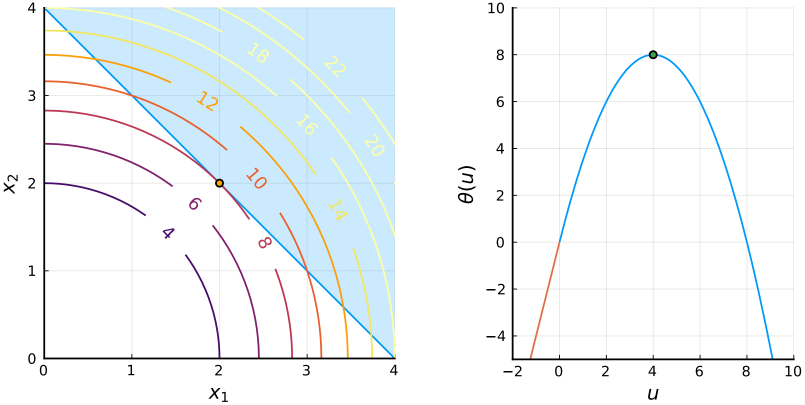
The Lagrangian dual function is given by:

$$\begin{aligned} \theta(u) &= \inf \left\{ x_1^2 + x_2^2 + u(-x_1 - x_2 + 4) : x_1, x_2 \geq 0 \right\} \\ &= \inf \left\{ x_1^2 - ux_1 : x_1 \geq 0 \right\} + \inf \left\{ x_2^2 - ux_2 : x_2 \geq 0 \right\} + 4u \\ &= \begin{cases} -1/2u^2 + 4u, & \text{if } u \geq 0 \\ -4u, \;\; 1 + 4u^2 & \text{if } u < 0. \end{cases} \end{aligned}$$

Solution of the Lagrangian dual problem $\sup_{u \geq 0} \{\theta(u)\}$ is $\overline{u} = 4$.

Duality gaps

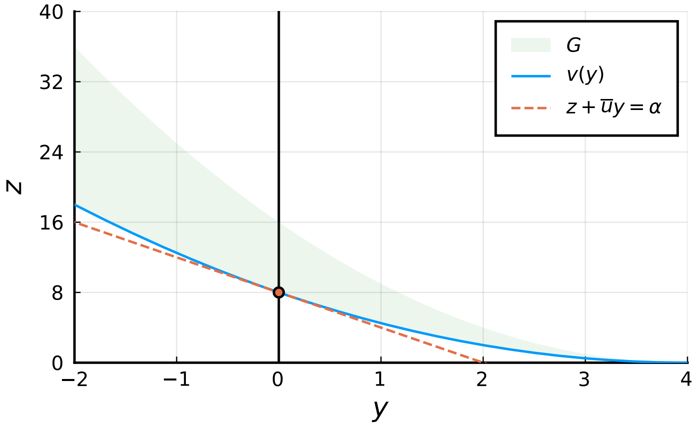
Example 1: notice that $f(\overline{x}_1, \overline{x}_2) = \theta(\overline{u}) = 8$.



Primal (left) and dual (right) problem representation

Duality gaps

If we draw the (g, f) map of X, we notice that $v(y) = (4 - y)^2/2$.
Note that $v(y) \geq v(0) - \overline{u}y$ holds for all $y \in \mathbb{R}$.



G mapping for Example 1, with $z = 8 - 4y$ and $(0, 8)$ matching at the primal and dual optima

Duality gaps

Example 2:

Let $X = \{(0,0),(0,4),(4,4),(4,0),(1,2),(2,1)\}$ and P be:

$$\begin{aligned} (P): \quad \text{min.} \quad & -2x_1 + x_2 \\ \text{subject to:} \quad & x_1 + x_2 = 3 \\ & x_1, x_2 \in X. \end{aligned}$$

- The optimal point is $(\overline{x}_1, \overline{x}_2) = (2,1)$.
- $f(x) = -2x_1 + x_2$, $h(x) = x_1 + x_2 - 3$.

Duality gaps

Example 2:

Let $X = \{(0,0),(0,4),(4,4),(4,0),(1,2),(2,1)\}$ and P be:

$$(P): \text{ min. } -2x_1 + x_2$$
$$\text{subject to: } x_1 + x_2 = 3$$
$$x_1, x_2 \in X.$$

- The optimal point is $(\overline{x}_1, \overline{x}_2) = (2,1)$.
- $f(x) = -2x_1 + x_2$, $h(x) = x_1 + x_2 - 3$.

The Lagrangian dual function is:

$$\theta(v) = \min\{(-2x_1 + x_2) + v(x_1 + x_2 - 3) : (x_1, x_2) \in X\}$$
$$= \begin{cases} -4 + 5v, \text{ if } v \leq -1 \\ -8 + v, 5 \text{ if } -1 \leq v \leq 2 \\ -3v, +p_1 \text{ if } v \geq 2. \end{cases}$$

Duality gaps

Example 2:

Let $X = \{(0,0),(0,4),(4,4),(4,0),(1,2),(2,1)\}$ and P be:

$$
\begin{aligned}
(P):\quad \text{min.}\quad & -2x_1 + x_2 \\
\text{subject to:}\quad & x_1 + x_2 = 3 \\
& x_1, x_2 \in X.
\end{aligned}
$$

- The optimal point is $(\overline{x}_1, \overline{x}_2) = (2,1)$.
- $f(x) = -2x_1 + x_2$, $h(x) = x_1 + x_2 - 3$.

The Lagrangian dual function is:

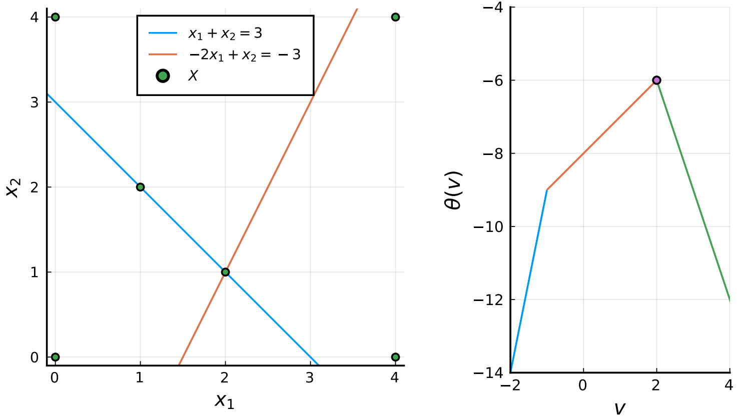
$$
\begin{aligned}
\theta(v) &= \min\{(-2x_1 + x_2) + v(x_1 + x_2 - 3) : (x_1, x_2) \in X\} \\
&= \begin{cases} -4 + 5v, & \text{if } v \leq -1 \\ -8 + v, 5 & \text{if } -1 \leq v \leq 2 \\ -3v, +p_1 & \text{if } v \geq 2. \end{cases}
\end{aligned}
$$

The solution of the Lagrangian dual problem $\sup\{\theta(v)\}$ is $\overline{v} = 2$.

Duality gaps

Example 2:

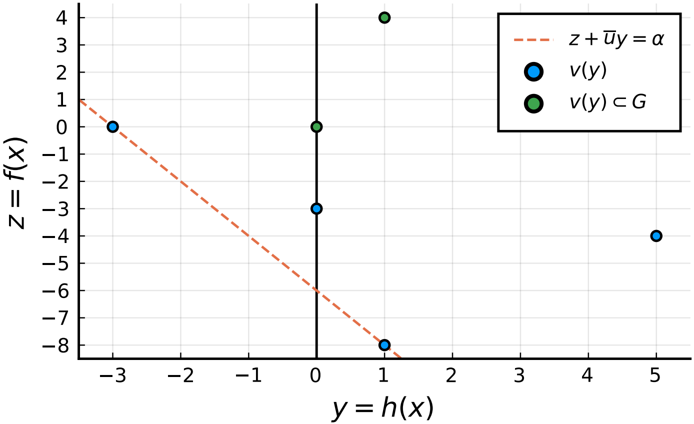
Notice that now $-3 = f(\overline{x}) > \theta(\overline{v}) = -6$.



Primal (left) and dual (right) problem representation

Duality gaps

Also, note that $v(y) \geq v(0) - \overline{u}y$ for all $y \in \mathbb{R}$ does not hold.



G mapping for Example 2, with $z = -6 - 2y$ and $(0, -3)$, not matching at the optima

Strong duality

Under convexity assumptions and constraint qualification, strong duality holds.

Strong duality

Under convexity assumptions and constraint qualification, strong duality holds.

Theorem 6

Let $X \subseteq \mathbb{R}^n$ *be a nonempty convex set. Moreover, let* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $g : \mathbb{R}^n \to \mathbb{R}^m$ *be convex functions, and let* $h : \mathbb{R}^n \to \mathbb{R}^l$ *be an affine function:* $h(x) = Ax - b$*. Suppose that Slater's constraint qualification holds true. Then*

$$\inf \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\} = \sup \{\theta(u, v) : u \geq 0\}.$$

Furthermore, if $\inf \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$ *is finite and achieved at* $\overline{x}$*, then* $\sup \{\theta(u, v) : u \geq 0\}$ *is achieved at* $(\overline{u}, \overline{v})$ *with* $\overline{u} \geq 0$ *and* $\overline{u}^\top g(\overline{x}) = 0$*.*

Strong duality

Proof outline.

1. Let $\gamma = \inf\{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$. Suppose that $-\infty < \gamma < \infty$, hence finite (left-hand side trivial, right-hand side by assumption).

Strong duality

Proof outline.

1. Let $\gamma = \inf \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$. Suppose that $-\infty < \gamma < \infty$, hence finite (left-hand side trivial, right-hand side by assumption).
2. Formulate the inconsistent system:

$$f(x) - \gamma < 0, \quad g(x) \leq 0, \quad h(x) = 0, \quad x \in X.$$

Strong duality

Proof outline.

1. Let $\gamma = \inf \{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$. Suppose that $-\infty < \gamma < \infty$, hence finite (left-hand side trivial, right-hand side by assumption).
2. Formulate the inconsistent system:

$$f(x) - \gamma < 0, \quad g(x) \leq 0, \quad h(x) = 0, \quad x \in X.$$

3. Use the separation theorem (or a form of Farkas' theorem) to show that (u_0, u, v) with $u_0 > 0$ and $u \geq 0$ exists such that, after scaling $u_0 = 1$: $f(x) + \overline{u}^\top g(x) + \overline{v}^\top h(x) \geq \gamma, \; x \in X$.

Strong duality

Proof outline.

1. Let $\gamma = \inf\{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$. Suppose that $-\infty < \gamma < \infty$, hence finite (left-hand side trivial, right-hand side by assumption).
2. Formulate the inconsistent system:
$$f(x) - \gamma < 0, \quad g(x) \leq 0, \quad h(x) = 0, \quad x \in X.$$
3. Use the separation theorem (or a form of Farkas' theorem) to show that (u_0, u, v) with $u_0 > 0$ and $u \geq 0$ exists such that, after scaling $u_0 = 1$: $f(x) + \overline{u}^\top g(x) + \overline{v}^\top h(x) \geq \gamma,\ x \in X$.
4. From weak duality (Theorem 3), we have $\theta(\overline{u}, \overline{v}) = \gamma$.

Strong duality

Proof outline.

1. Let $\gamma = \inf\{f(x) : g(x) \leq 0, h(x) = 0, x \in X\}$. Suppose that $-\infty < \gamma < \infty$, hence finite (left-hand side trivial, right-hand side by assumption).
2. Formulate the inconsistent system:

$$f(x) - \gamma < 0, \quad g(x) \leq 0, \quad h(x) = 0, \quad x \in X.$$

3. Use the separation theorem (or a form of Farkas' theorem) to show that (u_0, u, v) with $u_0 > 0$ and $u \geq 0$ exists such that, after scaling $u_0 = 1$: $f(x) + \overline{u}^\top g(x) + \overline{v}^\top h(x) \geq \gamma,\ x \in X$.
4. From weak duality (Theorem 3), we have $\theta(\overline{u}, \overline{v}) = \gamma$.
5. Finally, an optimal $\overline{x}$ solving the primal problem implies that $g(\overline{x}) \leq 0$, $h(\overline{x}) = 0$, $\overline{x} \in X$, and $f(x) = \gamma$. From 3, we have $\overline{u}^\top g(\overline{x}) \geq 0$. As $g(\overline{x}) \leq 0$ and $\overline{u} \geq 0$, $\overline{u}^\top g(\overline{x}) \geq 0 = 0$. □

Strong duality and convergence

Weak duality provides a stopping criterion of solution methods that can generate both primal and dual feasible solutions (primal-dual pairs).

For feasible x and (u, v), one can bound how suboptimal $f(x)$ is, as

$$f(x) - f(\overline{x}) \leq f(x) - \theta(u, v)$$

This establishes that x is ϵ-optimal, with $\epsilon = f(x) - \theta(u, v)$.

Strong duality and convergence

Weak duality provides a stopping criterion of solution methods that can generate both primal and dual feasible solutions (primal-dual pairs).

For feasible x and (u, v), one can bound how suboptimal $f(x)$ is, as

$$f(x) - f(\overline{x}) \leq f(x) - \theta(u, v)$$

This establishes that x is ϵ-optimal, with $\epsilon = f(x) - \theta(u, v)$.

- That is, (u, v) is a certificate of (sub-)optimality of x, as (u, v) proves that x is ϵ-optimal.
- In case strong duality holds, under the conditions of Theorem 6, one can expect ϵ converge to zero.

Strong duality and KKT conditions

Assume that strong duality holds. Observe the following:

1. Theorem 6 shows that, the complementarity conditions $\overline{u}^\top g(\overline{x}) \geq 0 = 0$ hold for an optimal primal-dual pair $(\overline{x}, (\overline{u}, \overline{v}))$.

Strong duality and KKT conditions

Assume that strong duality holds. Observe the following:

1. Theorem 6 shows that, the complementarity conditions $\overline{u}^\top g(\overline{x}) \geq 0 = 0$ hold for an optimal primal-dual pair $(\overline{x}, (\overline{u}, \overline{v}))$.
2. By definition, $\overline{x}$ and $(\overline{u}, \overline{v})$ are primal and dual feasible, respectively.

Strong duality and KKT conditions

Assume that strong duality holds. Observe the following:

1. Theorem 6 shows that, the complementarity conditions $\overline{u}^\top g(\overline{x}) \geq 0 = 0$ hold for an optimal primal-dual pair $(\overline{x}, (\overline{u}, \overline{v}))$.
2. By definition, $\overline{x}$ and $(\overline{u}, \overline{v})$ are primal and dual feasible, respectively.
3. Finally, notice that, if $\overline{x}$ is a minimiser for $\phi(x, \overline{u}, \overline{v}) = f(x) + \overline{u}^\top g(x) + \overline{v}^\top h(x)$, then we must have

$$\nabla f(\overline{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\overline{x}) + \sum_{i=1}^{l} v_i \nabla h_i(\overline{x}) = 0$$

Strong duality and KKT conditions

Assume that strong duality holds. Observe the following:

1. Theorem 6 shows that, the complementarity conditions $\overline{u}^\top g(\overline{x}) \geq 0 = 0$ hold for an optimal primal-dual pair $(\overline{x}, (\overline{u}, \overline{v}))$.
2. By definition, $\overline{x}$ and $(\overline{u}, \overline{v})$ are primal and dual feasible, respectively.
3. Finally, notice that, if $\overline{x}$ is a minimiser for $\phi(x, \overline{u}, \overline{v}) = f(x) + \overline{u}^\top g(x) + \overline{v}^\top h(x)$, then we must have

$$\nabla f(\overline{x}) + \sum_{i=1}^{m} u_i \nabla g_i(\overline{x}) + \sum_{i=1}^{l} v_i \nabla h_i(\overline{x}) = 0$$

Combining (1)–(3), we notice that $(\overline{x}, (\overline{u}, \overline{v}))$ satisfies the KKT conditions (Lecture 8), which are necessary and sufficient optimality conditions under the assumptions of Theorem 6.

Properties of Lagrangian dual function

Strong duality motivates the use of Lagrangian duals to solve (often harder) primal optimisation problems.

This is partially related to the two main properties of Lagrangian functions: concavity and subdifferentiability.

Properties of Lagrangian dual function

Strong duality motivates the use of Lagrangian duals to solve (often harder) primal optimisation problems.

This is partially related to the two main properties of Lagrangian functions: concavity and subdifferentiability.

Theorem 7 (Concavity of Lagrangian dual functions)

Let $X \subseteq \mathbb{R}^n$ *be a nonempty compact set, and let* $f : \mathbb{R}^n \to \mathbb{R}$ *and* $\beta : \mathbb{R}^n \to \mathbb{R}^{m+l}$, *with* $w^\top \beta(x) = \binom{u}{v}^\top \binom{g(x)}{h(x)}$ *be continuous. Then* $\theta(w) = \inf \left\{ f(x) + w^\top \beta(x) : x \in X \right\}$ *is concave in* $\mathbb{R}^{m+l}$.

Properties of Lagrangian dual function

Strong duality motivates the use of Lagrangian duals to solve (often harder) primal optimisation problems.

This is partially related to the two main properties of Lagrangian functions: concavity and subdifferentiability.

Theorem 7 (Concavity of Lagrangian dual functions)

Let $X \subseteq \mathbb{R}^n$ be a nonempty compact set, and let $f : \mathbb{R}^n \to \mathbb{R}$ and $\beta : \mathbb{R}^n \to \mathbb{R}^{m+l}$, with $w^\top \beta(x) = \binom{u}{v}^\top \binom{g(x)}{h(x)}$ be continuous. Then $\theta(w) = \inf \left\{ f(x) + w^\top \beta(x) : x \in X \right\}$ is concave in $\mathbb{R}^{m+l}$.

Remarks:

1. Global optimality conditions hold for $\theta(w)$.
2. The dual function $\theta(w)$ is not explicitly available.
3. Concavity implies that $\theta(w)$ has subgradients everywhere.

Properties of Lagrangian dual function

Subgradients of Lagrangian dual functions are easily available. Let $X(w) = \left\{x \in X : x = \arg\min \left\{f(x) + w^\top \beta(x)\right\}\right\}$.

Theorem 8

Let $X \subseteq \mathbb{R}^n$ be a nonempty compact set, and let $f : \mathbb{R}^n \to \mathbb{R}$ and $\beta : \mathbb{R}^n \to \mathbb{R}^{m+l}$, with $w^\top \beta(x) = \binom{u}{v}^\top \binom{g(x)}{h(x)}$ be continuous.
If $\overline{x} \in X(\overline{w})$, then $\beta(\overline{x})$ is a subgradient of $\theta(w)$ at $\overline{w}$.

Properties of Lagrangian dual function

Subgradients of Lagrangian dual functions are easily available. Let $X(w) = \left\{x \in X : x = \arg\min\left\{f(x) + w^\top \beta(x)\right\}\right\}$.

Theorem 8

Let $X \subseteq \mathbb{R}^n$ be a nonempty compact set, and let $f : \mathbb{R}^n \to \mathbb{R}$ and $\beta : \mathbb{R}^n \to \mathbb{R}^{m+l}$, with $w^\top \beta(x) = \binom{u}{v}^\top \binom{g(x)}{h(x)}$ be continuous. If $\overline{x} \in X(\overline{w})$, then $\beta(\overline{x})$ is a subgradient of $\theta(w)$ at $\overline{w}$.

Proof.

Since f and β are continuous and X is compact, $X(\overline{w}) \neq \emptyset$ for any $w \in \mathbb{R}^{m+l}$. Now, let $\overline{w} \in \mathbb{R}^{m+l}$ and $\overline{x} \in X(\overline{w})$. Then

$$\begin{aligned}
\theta(w) &= \inf\left\{f(x) + w^\top \beta(x) : x \in X\right\} \\
&\leq f(\overline{x}) + w^\top \beta(\overline{x}) \\
&= f(\overline{x}) + (w - \overline{w})^\top \beta(\overline{x}) + \overline{w}^\top \beta(\overline{x}) \\
&= \theta(\overline{w}) + (w - \overline{w})^\top \beta(\overline{x}).
\end{aligned}$$

□

Solving Lagrangian duals

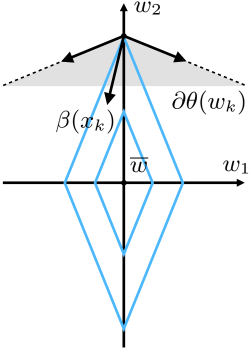
Solving Lagrangian dual is challenging since $\theta(w)$ is typically nonsmooth, requiring an adaptation of the gradient method.

Solving Lagrangian duals

Solving Lagrangian dual is challenging since $\theta(w)$ is typically nonsmooth, requiring an adaptation of the gradient method.

The subgradient method uses subgradients $\beta(\overline{x})$ to solve max. $\{\theta(\overline{x}) : w \in W\}$, where $W = \{w = (u, v) : u \geq 0\}$.

- Subgradients are not necessarily ascent directions for nonsmooth concave functions.
- However, for sufficiently small steps size, the distance $|w - \overline{w}|$ to a maximiser $\overline{w}$ decreases.



Level curves of a nonsmooth $\theta(w_k)$ and a subgradient $\beta(x_k)$ at (w_k).

The subgradient method

Algorithm Subgradient method

```
initialise. tolerance ϵ > 0, initial point w_0, iteration count k = 0.
while ||β(x_k)||_2 > ϵ do
    x_k ← arg min_x {θ(w_k) = inf {f(x) + w_k^⊤ β(x)}}
    LB_k = max {LB_k, θ(w_k)}
    update λ_k
    w_{k+1} = w_k + λ_k β(x_k).
    k ← k + 1.
end while
return LB_k = θ(w_k).
```

The subgradient method

Algorithm Subgradient method

1: **initialise.** tolerance $\epsilon > 0$, initial point w_0, iteration count $k = 0$.
2: **while** $||\beta(x_k)||_2 > \epsilon$ **do**
3: $\quad x_k \leftarrow \arg\min_x \left\{\theta(w_k) = \inf\left\{f(x) + w_k^\top \beta(x)\right\}\right\}$
4: $\quad LB_k = \max\left\{LB_k, \theta(w_k)\right\}$
5: $\quad$ update λ_k
6: $\quad w_{k+1} = w_k + \lambda_k \beta(x_k)$.
7: $\quad k \leftarrow k + 1$.
8: **end while**
9: **return** $LB_k = \theta(w_k)$.

Remarks:

1. The stop condition $||\beta(x_k)||_2 > \epsilon$ emulates $0 \in \partial\theta(w_k)$
2. Theoretical convergence is guaranteed if Step 5 generates a sequence $\{\lambda_k\}$ such that $\sum_{k=0}^{\infty} \lambda_k \to \infty$ and $\lim_{k\to\infty} \lambda_k = 0$.
3. To guarantee that $w \in W$, one can use the projection $w_k^i = \max\left\{0, w_k^i\right\}, i = 1, \dots, m$.

The subgradient method

A common rule for step size updates is the Polyak rule:

$$\lambda_{k+1} = \frac{\alpha_k(LB_k - \theta(w_k))}{||\beta(x_k)||^2}$$

with $\alpha_k \in (0, 2)$ and LB_k being a lower bound on $\theta(\overline{w})$.

The subgradient method

A common rule for step size updates is the Polyak rule:

$$\lambda_{k+1} = \frac{\alpha_k (LB_k - \theta(w_k))}{||\beta(x_k)||^2}$$

with $\alpha_k \in (0, 2)$ and LB_k being a lower bound on $\theta(\overline{w})$. This rule is inspired by the following result.

Proposition 8.1 (Improving step size)

If w_k is not optimal, then, for all optimal dual solutions $\overline{w}$, we have

$$||w_{k+1} - \overline{w}|| < ||w_k - \overline{w}||$$

for all step sizes λ_k such that

$$0 < \lambda_k < \frac{2(\theta(\overline{w}) - \theta(w_k))}{||\beta(x_k)||^2}.$$

The subgradient method

A common rule for step size updates is the Polyak rule:

$$\lambda_{k+1} = \frac{\alpha_k (LB_k - \theta(w_k))}{||\beta(x_k)||^2}$$

with $\alpha_k \in (0, 2)$ and LB_k being a lower bound on $\theta(\overline{w})$. This rule is inspired by the following result.

Proposition 8.1 (Improving step size)

If w_k is not optimal, then, for all optimal dual solutions $\overline{w}$, we have

$$||w_{k+1} - \overline{w}|| < ||w_k - \overline{w}||$$

for all step sizes λ_k such that

$$0 < \lambda_k < \frac{2(\theta(\overline{w}) - \theta(w_k))}{||\beta(x_k)||^2}.$$

Remark: $\theta(\overline{w})$ is replaced by an approximation LB_k and α_k acts as a correction term for the difference $\theta(\overline{w}) - \theta(w_k)$.

The subgradient method

Proof.

We have that $||w_{k+1} - \overline{w}||^2 = ||w_k + \lambda_k \beta(x_k) - \overline{w}||^2 =$

$$||w_k - \overline{w}||^2 - 2\lambda_k(\overline{w} - w_k)^\top \beta(x_k) + (\lambda_k)^2 ||\beta(x_k)||^2.$$

By the subgradient inequality: $\theta(\overline{w}) - \theta(w_k) \leq (\overline{w} - w_k)^\top \beta(x_k)$. Thus

$$||w_{k+1} - \overline{w}||^2 \leq ||w_k - \overline{w}||^2 - 2\lambda_k(\theta(\overline{w}) - \theta(w_k))^\top \beta(x_k) + (\lambda_k)^2 ||\beta(x_k)||^2.$$

The subgradient method

Proof.

We have that $||w_{k+1} - \overline{w}||^2 = ||w_k + \lambda_k \beta(x_k) - \overline{w}||^2 =$

$$||w_k - \overline{w}||^2 - 2\lambda_k(\overline{w} - w_k)^\top \beta(x_k) + (\lambda_k)^2 ||\beta(x_k)||^2.$$

By the subgradient inequality: $\theta(\overline{w}) - \theta(w_k) \leq (\overline{w} - w_k)^\top \beta(x_k)$. Thus

$$||w_{k+1} - \overline{w}||^2 \leq ||w_k - \overline{w}||^2 - 2\lambda_k(\theta(\overline{w}) - \theta(w_k))^\top \beta(x_k) + (\lambda_k)^2 ||\beta(x_k)||^2.$$

Parametrising the last two terms by $\gamma_k = \frac{\lambda_k ||\xi||^2}{\theta(\overline{w}) - \theta(w_k)}$ leads to

$$||w_{k+1} - \overline{w}||^2 \leq ||w_k - \overline{w}||^2 - \frac{\gamma_k(2 - \gamma_k)(\theta(\overline{w}) - \theta(w_k))^2}{||\xi_k||^2}.$$

Notice that if $0 < \lambda_k < \frac{2(\theta(\overline{w}) - \theta(w_k))}{||\beta(x_k)||^2}$ then $0 < \gamma_k < 2$ and, thus, $||w_{k+1} - \overline{w}|| < ||w_k - \overline{w}||$.


THANK YOU