Lectures 3–4: Lagrangian duality and algorithms for the Lagrangian dual problem

Ann-Brith Strömberg

2008-09-09

The Relaxation Theorem

- Problem: find

$$f^* = \underset{\mathbf{x}}{\text{infimum}} \; f(\mathbf{x}), \tag{1a}$$
$$\text{subject to } \; \mathbf{x} \in S, \tag{1b}$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ is a given function and $S \subseteq \mathbb{R}^n$

- A *relaxation* to (1a)–(1b) has the following form: find

$$f_R^* = \underset{\mathbf{x}}{\text{infimum}} \quad f_R(\mathbf{x}), \tag{2a}$$
$$\text{subject to} \quad \mathbf{x} \in S_R, \tag{2b}$$

where $f_R : \mathbb{R}^n \mapsto \mathbb{R}$ is a function with $f_R \leq f$ on S and $S_R \supseteq S$.

Relaxation example—Maximization!!

- The binary knapsack problem

$$\begin{array}{rlcl} z^* = \displaystyle\operatorname*{maximize}_{\mathbf{x}\in\{0,1\}^4} & 7x_1 + 4x_2 + 5x_3 + 2x_4 & & \\ \text{subject to} & 3x_1 + 3x_2 + 4x_3 + 2x_4 & \leq & 5 \end{array}$$

  has the optimal solution $\mathbf{x}^* = (1,0,0,1)$, $z^* = 9$

- Its continuous relaxation

$$\begin{array}{rlcl} z^*_{\mathrm{LP}} = \displaystyle\operatorname*{maximize}_{\mathbf{x}\in[0,1]^4} & 7x_1 + 4x_2 + 5x_3 + 2x_4 & & \\ \text{subject to} & 3x_1 + 3x_2 + 4x_3 + 2x_4 & \leq & 5 \end{array}$$

  has the optimal solution $\mathbf{x}^*_R = (1, \frac{2}{3}, 0, 0)$, $z^*_R = 9\frac{2}{3} > z^*$

- $\mathbf{x}^*_R$ is *not feasible* in the binary problem

The relaxation theorem

1. [relaxation] $f_R^* \leq f^*$
2. [infeasibility] *If* (2) *is infeasible, then so is* (1)
3. [optimal relaxation]
   If the problem (2) *has an optimal solution,* $\mathbf{x}_R^*$, *for which it holds that*
   $$\mathbf{x}_R^* \in S \qquad \text{and} \qquad f_R(\mathbf{x}_R^*) = f(\mathbf{x}_R^*),$$
   then $\mathbf{x}_R^*$ *is an optimal solution to* (1) *as well.*

- *Proof portion.* For *3.*, note that

$$f(\mathbf{x}_R^*) = f_R(\mathbf{x}_R^*) \leq f_R(\mathbf{x}) \leq f(\mathbf{x}), \qquad \mathbf{x} \in S$$

Lagrangian relaxation

- Consider the optimization problem:

$$f^* = \operatorname*{infimum}_{\mathbf{x}} f(\mathbf{x}), \tag{3a}$$
$$\text{subject to } \mathbf{x} \in X, \tag{3b}$$
$$g_i(\mathbf{x}) \leq 0, \qquad i = 1, \ldots, m, \tag{3c}$$

  where $f : \mathbb{R}^n \mapsto \mathbb{R}$ and $g_i : \mathbb{R}^n \mapsto \mathbb{R}$ $(i = 1, 2, \ldots, m)$ are given functions, and $X \subseteq \mathbb{R}^n$

- Here we assume that

$$-\infty < f^* < \infty, \tag{4}$$

  that is, that f is bounded from below and that the problem has at least one feasible solution

- For a vector $\boldsymbol{\mu} \in \mathbb{R}^m$, we define the *Lagrange function*

$$L(\mathbf{x}, \boldsymbol{\mu}) = f(\mathbf{x}) + \sum_{i=1}^{m} \mu_i g_i(\mathbf{x}) = f(\mathbf{x}) + \boldsymbol{\mu}^{\mathrm{T}} \mathbf{g}(\mathbf{x})$$

- We call the vector $\boldsymbol{\mu}^* \in \mathbb{R}^m$ a *Lagrange multiplier* if it is non-negative and if $f^* = \inf_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\mu}^*)$ holds.

Lagrange multipliers and global optima

- *Let* $\boldsymbol{\mu}^*$ *be a Lagrange multiplier. Then,* $\mathbf{x}^*$ *is an optimal solution to*
  $f^* = \inf\{f(\mathbf{x})\,|\mathbf{x} \in X, g_i(\mathbf{x}) \leq 0, i = 1, \ldots, m\}$,
  if and only if it is feasible and

$$\mathbf{x}^* \in \arg\min_{\mathbf{x}\in X} L(\mathbf{x}, \boldsymbol{\mu}^*), \text{ and } \mu_i^* g_i(\mathbf{x}^*) = 0, \; i = 1, \ldots, m$$

- Notice the resemblance to the KKT conditions!
  If $X = \mathbb{R}^n$ and all functions are in C^1 then "$\mathbf{x}^* \in \arg\min_{\mathbf{x}\in X} L(\mathbf{x}, \boldsymbol{\mu}^*)$" is the same as the force equilibrium condition, the first row of the KKT conditions. The second item, "$\mu_i^* g_i(\mathbf{x}^*) = 0$ for all i" is the complementarity conditions

The Lagrangian dual problem associated with the Lagrangian relaxation

- The *Lagrangian dual function* is

$$q(\boldsymbol{\mu}) = \operatorname*{infimum}_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\mu})$$

- The *Lagrangian dual problem* is to

$$q^* = \operatorname*{maximize}_{\boldsymbol{\mu} \geq \mathbf{0}^m} q(\boldsymbol{\mu}) \tag{5}$$

- For some $\boldsymbol{\mu}$, $q(\boldsymbol{\mu}) = -\infty$ is possible. If this is true for all $\boldsymbol{\mu} \geq \mathbf{0}^m$ then

$$q^* = \operatorname*{supremum}_{\boldsymbol{\mu} \geq \mathbf{0}^m} q(\boldsymbol{\mu}) = -\infty$$

- The *effective domain* of q is
  $D_q = \{ \boldsymbol{\mu} \in \mathbb{R}^m \mid q(\boldsymbol{\mu}) > -\infty \}$

  [Theorem] D_q is convex, and q is concave on D_q $\square$

- That the Lagrangian dual problem always is convex is very good news!
- We indeed maximize a concave function

- But we need still to show how a Lagrangian dual optimal solution can be used to generate a primal optimal solution

Weak Duality Theorem

Let $\mathbf{x}$ *and* $\boldsymbol{\mu}$ *be* feasible *in*

$f^* = \inf\{f(\mathbf{x}) \,|\mathbf{x} \in X, g_i(\mathbf{x}) \leq 0, i = 1, \ldots, m\}$

and $\quad q^* = \max\{\, q(\boldsymbol{\mu}) \,|\boldsymbol{\mu} \geq \mathbf{0}^m\},\quad$ *respectively.*

Then,

$$q(\boldsymbol{\mu}) \leq f(\mathbf{x})$$

In particular,

$$q^* \leq f^*$$

If $q(\boldsymbol{\mu}) = f(\mathbf{x})$, *then the pair* $(\mathbf{x}, \boldsymbol{\mu})$ *is optimal in the respective problem and* $q^* = q(\boldsymbol{\mu}) = f(\mathbf{x}) = f^*$ $\square$

- Weak duality is also a consequence of the Relaxation Theorem: For any $\boldsymbol{\mu} \geq \mathbf{0}^m$, let

$$\begin{aligned} S &= X \cap \{\, \mathbf{x} \in \mathbb{R}^n \mid \mathbf{g}(\mathbf{x}) \leq \mathbf{0}^m \,\}, \\ S_R &= X, \\ f_R &= L(\boldsymbol{\mu}, \cdot) \end{aligned}$$

  Apply the Relaxation Theorem

- If $q^* = f^*$, there is *no duality gap*.

- If there exists a Lagrange multiplier vector, then by the weak duality theorem, there is no duality gap.

Global optimality conditions

- *The vector* $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ *is a pair of an optimal primal solution and a Lagrange multiplier if and only if*

$$\boldsymbol{\mu}^* \geq \mathbf{0}^m, \qquad \text{(Dual feasibility)} \tag{6a}$$

$$\mathbf{x}^* \in \arg\min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\mu}^*), \qquad \text{(Lagrangian optimality)} \tag{6b}$$

$$\mathbf{x}^* \in X, \quad \mathbf{g}(\mathbf{x}^*) \leq \mathbf{0}^m, \qquad \text{(Primal feasibility)} \tag{6c}$$

$$\mu_i^* g_i(\mathbf{x}^*) = 0, \quad i = 1, \ldots, m \qquad \text{(Complementary slackness)} \tag{6d}$$

- If $\exists (\mathbf{x}^*, \boldsymbol{\mu}^*)$ that fulfil (6), then there is a zero duality gap and Lagrange multipliers exist

Saddle points

- *The vector* $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ *is a pair of an optimal primal solution and a Lagrange multiplier if and only if* $\mathbf{x}^* \in X$, $\boldsymbol{\mu}^* \geq \mathbf{0}^m$, *and* $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ *is a* saddle point *of the Lagrangian function on* $X \times \mathbb{R}^m_+$, *that is,*

$$L(\mathbf{x}^*, \boldsymbol{\mu}) \leq L(\mathbf{x}^*, \boldsymbol{\mu}^*) \leq L(\mathbf{x}, \boldsymbol{\mu}^*), \quad (\mathbf{x}, \boldsymbol{\mu}) \in X \times \mathbb{R}^m_+,$$

  holds

- If $\exists(\mathbf{x}^*, \boldsymbol{\mu}^*)$, equivalent to the global optimality conditions, the existence of Lagrange multipliers, and a zero duality gap

Strong duality for convex programs, introduction

- Convexity of the dual problem comes with very few assumptions on the original, primal problem
- The characterization of the primal–dual set of optimal solutions is also quite easily established
- To establish *strong duality*—sufficient conditions under which there is no duality gap—takes much more
- In particular—as with the KKT conditions—we need regularity conditions (constraint qualifications) and separation theorems

Strong duality Theorem

- Consider the problem (3), that is,

$$f^* = \inf\{f(\mathbf{x})\,|\mathbf{x} \in X, g_i(\mathbf{x}) \leq 0, i = 1, \ldots, m\},$$

  where $f : \mathbb{R}^n \mapsto \mathbb{R}$ and g_i $(i = 1, \ldots, m)$ are *convex* and $X \subseteq \mathbb{R}^n$ is a *convex* set

- Introduce the following constraint qualification (CQ):

$$\exists \mathbf{x} \in X \text{ with } \mathbf{g}(\mathbf{x}) < \mathbf{0}^m \tag{7}$$

Strong duality Theorem

Suppose that $-\infty < f^* < \infty$, *and that the CQ* (7) *holds for the (convex) problem* (3)

(a) *There is no duality gap and there exists at least one Lagrange multiplier* $\boldsymbol{\mu}^*$. *Moreover, the set of Lagrange multipliers is bounded and convex*

(b) *If infimum in* (3) *is attained at some* $\mathbf{x}^*$, *then the pair* $(\mathbf{x}^*, \boldsymbol{\mu}^*)$ *satisfies the global optimality conditions* (6)

(c) *If the functions* f *and* g_i *are in* C^1 *and* X *is open (for example,* $X = \mathbb{R}^n$*) then* (6) *equal the KKT conditions*

If all constraints are linear we can remove the CQ (7).

Example I: An explicit, differentiable dual problem

- Consider the problem to

$$\begin{aligned} \underset{\mathbf{x}}{\operatorname{minimize}} \quad & f(\mathbf{x}) := x_1^2 + x_2^2, \\ \text{subject to} \quad & x_1 + x_2 \geq 4, \\ & x_j \geq 0, \qquad j = 1, 2 \end{aligned}$$

- Let $g(\mathbf{x}) = -x_1 - x_2 + 4$ and $X = \{\, (x_1, x_2) \mid x_j \geq 0,\ j = 1, 2 \,\} = \mathbb{R}_+^2$

- The Lagrangian dual function is

$$\begin{aligned} q(\mu) &= \min_{\mathbf{x}\in X} L(\mathbf{x}, \mu) := f(\mathbf{x}) + \mu(-x_1 - x_2 + 4) \\ &= 4\mu + \min_{\mathbf{x}\geq\mathbf{0}} \{x_1^2 + x_2^2 - \mu x_1 - \mu x_2\} \\ &= 4\mu + \min_{x_1\geq 0} \{x_1^2 - \mu x_1\} + \min_{x_2\geq 0} \{x_2^2 - \mu x_2\}, \;\; \mu \geq 0 \end{aligned}$$

- For a fixed $\mu \geq 0$, the minimum is attained at $x_1(\mu) = \frac{\mu}{2}, x_2(\mu) = \frac{\mu}{2}$
- Substituting this expression into $q(\mu)$, we obtain that $q(\mu) = f(\mathbf{x}(\mu)) + \mu(-x_1(\mu) - x_2(\mu) + 4) = 4\mu - \frac{\mu^2}{2}$
- Note that q is *strictly concave*, and it is differentiable everywhere (due to the fact that f, g are differentiable and $\mathbf{x}(\mu)$ is unique)

- Recall the dual problem

$$q^* = \max_{\mu \geq 0} q(\mu) = \max_{\mu \geq 0} \left(4\mu - \frac{\mu^2}{2} \right)$$

- We have that $q'(\mu) = 4 - \mu = 0 \iff \mu = 4$.
  As $4 \geq 0$, this is the optimum in the dual problem!

$\Rightarrow$ $\mu^* = 4$ and $\mathbf{x}^* = (x_1(\mu^*), x_2(\mu^*))^{\mathrm{T}} = (2, 2)^{\mathrm{T}}$

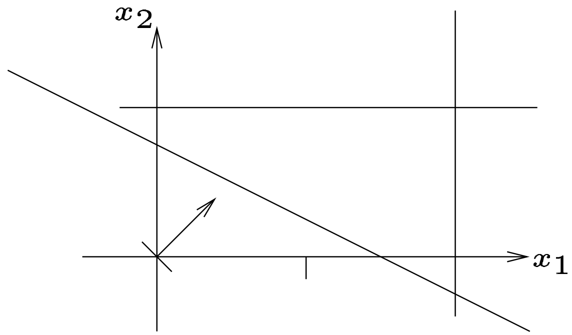
- Also: $f(\mathbf{x}^*) = q(\mu^*) = 8$
- In this example, the dual function is *differentiable*.
  The optimum $\mathbf{x}^*$ is also unique and automatically given by $\mathbf{x}^* = \mathbf{x}(\mu^*)$

Example II: Implicit non-differentiable dual problem

- Consider the linear programming problem to

$$
\begin{aligned}
\underset{\mathbf{x}}{\text{minimize}} \;\; & f(\mathbf{x}) := -x_1 - x_2, \\
\text{subject to} \;\; & 2x_1 + 4x_2 \leq 3, \\
& 0 \leq x_1 \leq 2, \\
& 0 \leq x_2 \leq 1
\end{aligned}
$$

- The optimal solution is $\mathbf{x}^* = (3/2, 0)^{\mathrm{T}}, f(\mathbf{x}^*) = -3/2$

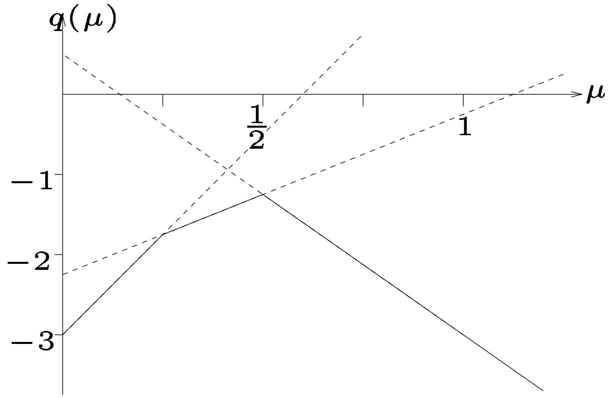


Lagrangian relax the first constraint:

$$L(\mathbf{x}, \mu) = -x_1 - x_2 + \mu(2x_1 + 4x_2 - 3);$$

$$q(\mu) = -3\mu + \min_{0 \le x_1 \le 2} \{(-1 + 2\mu)x_1\} + \min_{0 \le x_2 \le 1} \{(-1 + 4\mu)x_2\}$$

$$= \begin{cases} -3 + 5\mu, & 0 \le \mu \le 1/4, \quad \Leftrightarrow \quad x_1(\mu) = 2, x_2(\mu) = 1 \\ -2 + \mu, & 1/4 \le \mu \le 1/2, \quad \Leftrightarrow \quad x_1(\mu) = 2, x_2(\mu) = 0 \\ -3\mu, & 1/2 \le \mu \quad \Leftrightarrow \quad x_1(\mu) = x_2(\mu) = 0 \end{cases}$$



$$\mu^* = \frac{1}{2},\ q(\mu^*) = -\frac{3}{2}$$

- For linear (convex) programs strong duality holds, but how obtain $\mathbf{x}^*$ from μ^*?
- q is non-differentiable at $\mu^* \Rightarrow$ Utilize the characterization in (6)
- First, the subproblem solution set at μ^* is $X(\mu^*) = \{ \binom{2\alpha}{0} \mid 0 \le \alpha \le 1 \}$.
- Among the subproblem solutions, we next have to find one that is primal feasible as well as complementary
- Primal feasibility means that $2 \cdot 2\alpha + 4 \cdot 0 \le 3 \Longleftrightarrow \alpha \le 3/4$
- Further, complementarity means that $\mu^* \cdot (2x_1^* + 4x_2^* - 3) = 0 \Longleftrightarrow \alpha = 3/4$, since $\mu^* \neq 0$.
- Conclusion: the only primal vector $\mathbf{x}$ that satisfies the system (6) together with the dual solution $\mu^* = 1/2$ is $\mathbf{x}^* = (3/2, 0)^{\mathrm{T}}$
- Observe finally that $f^* = q^*$

A theoretical argument for $\mu^* = 1/2$

- Due to the global optimality conditions, the optimal solution must in this convex case be among the subproblem solutions
- Since x_1^* is not in one of the "corners" of X $(0 < x_1^* < 2)$, the value of μ^* must be such that the cost term for x_1 in $L(\mathbf{x}, \mu^*)$ is zero! That is, $-1 + 2\mu^* = 0 \Rightarrow \mu^* = 1/2$!
- A non-coordinability phenomenon—a non-unique subproblem solution means that the optimal solution is not obtained automatically
- In non-convex cases (e.g., integrality constraints) the optimal solution may not be among the points in $X(\boldsymbol{\mu}^*)$ (the set of subproblem solutions at μ^*)
- What do we do then??

Subgradients of convex functions

- Let $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a convex function.
  A vector $\mathbf{p} \in \mathbb{R}^n$ is a *subgradient* of f at $\mathbf{x} \in \mathbb{R}^n$ if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{p}^{\mathrm{T}}(\mathbf{y} - \mathbf{x}), \qquad \mathbf{y} \in \mathbb{R}^n \tag{8}$$

- The set of such vectors $\mathbf{p}$ defines the *subdifferential* of f at $\mathbf{x}$, and is denoted $\partial f(\mathbf{x})$
- $\partial f(\mathbf{x})$ is the collection of "slopes" of the function f at $\mathbf{x}$
- *For every* $\mathbf{x} \in \mathbb{R}^n$, $\partial f(\mathbf{x})$ *is a non-empty, convex, and compact set*

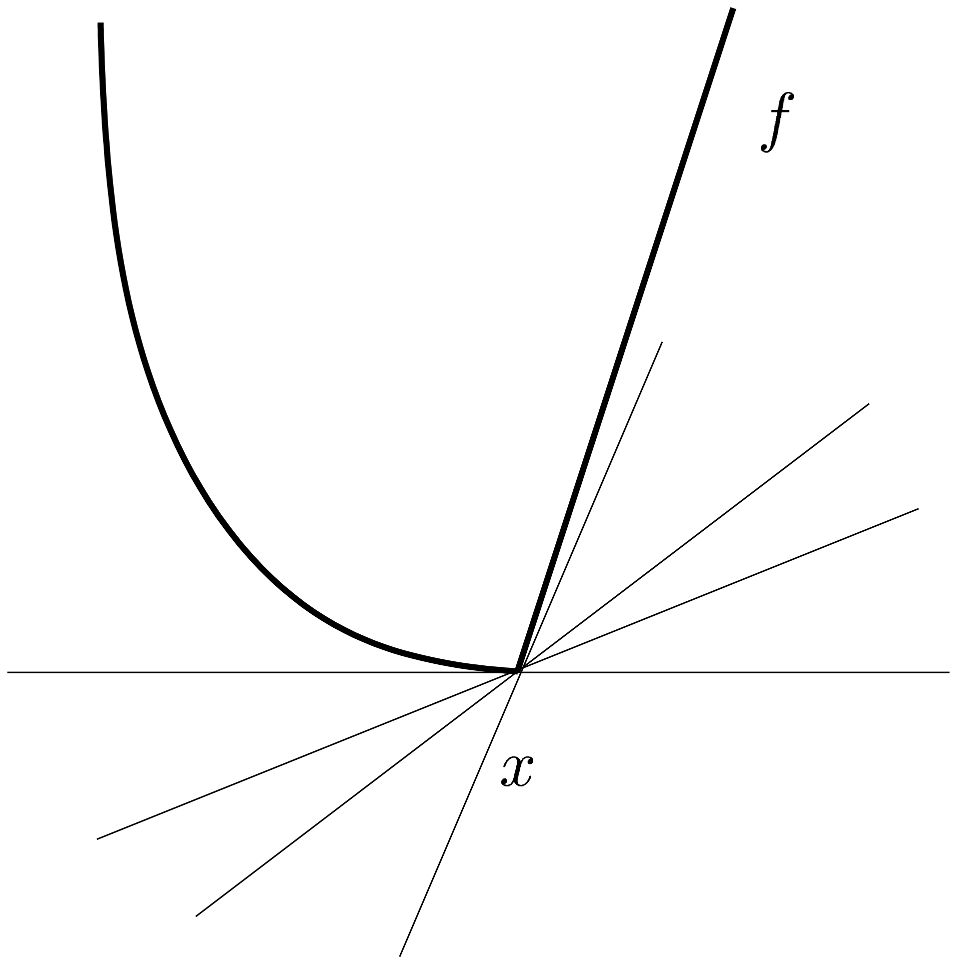


Figure 1: Four possible slopes of the convex function f at x

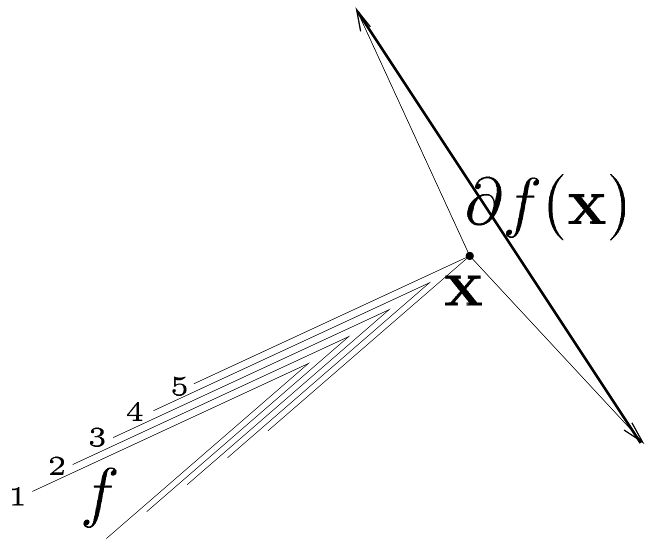


Figure 2: The subdifferential of a convex function f at $\mathbf{x}$.
f is indicated by level curves.

- *The convex function f is differentiable at $\mathbf{x}$ if there exists exactly one subgradient of f at $\mathbf{x}$ which then equals the gradient of f at $\mathbf{x}$, $\nabla f(\mathbf{x})$*

Differentiability of the Lagrangian dual function

- Consider the problem (3):
  $f^* = \inf\{ f(\mathbf{x}) \,|\, \mathbf{x} \in X, g_i(\mathbf{x}) \leq 0, i = 1, \ldots, m \}$,
  and assume

$$f, g_i \, (\forall i) \text{ continuous; } X \text{ nonempty and compact} \qquad (9)$$

- The set of solutions to the Lagrangian subproblem

$$X(\boldsymbol{\mu}) = \arg\min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\mu})$$

  is non-empty and compact for every $\boldsymbol{\mu} \in \mathbb{R}^m$

- Next: *sub*-differentiability properties of the function q

Subgradients and gradients of q

- *Suppose that* (9) *holds in the problem* (3)
- *The dual function q is finite, continuous and concave on $\mathbb{R}^m$. If its supremum over $\mathbb{R}^m_+$ is attained, then the optimal solution set therefore is closed and convex*
- *Let $\boldsymbol{\mu} \in \mathbb{R}^m$. If $\mathbf{x} \in X(\boldsymbol{\mu})$, then $\mathbf{g}(\mathbf{x})$ is a subgradient to q at $\boldsymbol{\mu}$, that is, $\mathbf{g}(\mathbf{x}) \in \partial q(\boldsymbol{\mu})$*
- *Proof.* Let $\bar{\boldsymbol{\mu}} \in \mathbb{R}^m$ be arbitrary. We have that

$$\begin{aligned} q(\bar{\boldsymbol{\mu}}) &= \operatorname*{infimum}_{\mathbf{y} \in X} L(\mathbf{y}, \bar{\boldsymbol{\mu}}) \leq f(\mathbf{x}) + \bar{\boldsymbol{\mu}}^{\mathrm{T}} \mathbf{g}(\mathbf{x}) \\ &= f(\mathbf{x}) + (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{g}(\mathbf{x}) + \boldsymbol{\mu}^{\mathrm{T}} \mathbf{g}(\mathbf{x}) \\ &= q(\boldsymbol{\mu}) + (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu})^{\mathrm{T}} \mathbf{g}(\mathbf{x}) \end{aligned}$$

- Recall the *subgradient inequality* (8) for a *convex* function f: $\mathbf{p}$ is a subgradient of f at $\mathbf{x}$ if

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{p}^{\mathrm{T}}(\mathbf{y} - \mathbf{x}), \qquad \mathbf{y} \in \mathbb{R}^n$$

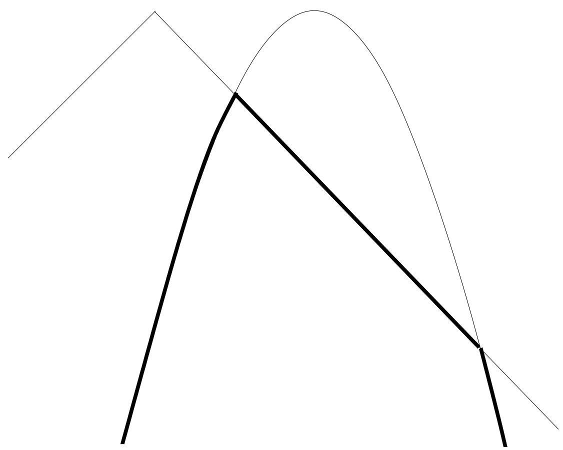
- The function $f(\mathbf{x}) + \mathbf{p}^{\mathrm{T}}(\mathbf{y} - \mathbf{x})$ is linear w.r.t. $\mathbf{y}$ and *underestimates* $f(\mathbf{y})$ over $\mathbb{R}^n$
- Here, we have a *concave* function q and the opposite inequality: $\mathbf{g}(\mathbf{x})$ is a subgradient (actually, supgradient) of q at $\boldsymbol{\mu}$ if $\mathbf{x} \in X(\boldsymbol{\mu})$ and

$$q(\bar{\boldsymbol{\mu}}) \leq q(\boldsymbol{\mu}) + (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu})^{\mathrm{T}}\mathbf{g}(\mathbf{x}), \qquad \bar{\boldsymbol{\mu}} \in \mathbb{R}^m$$

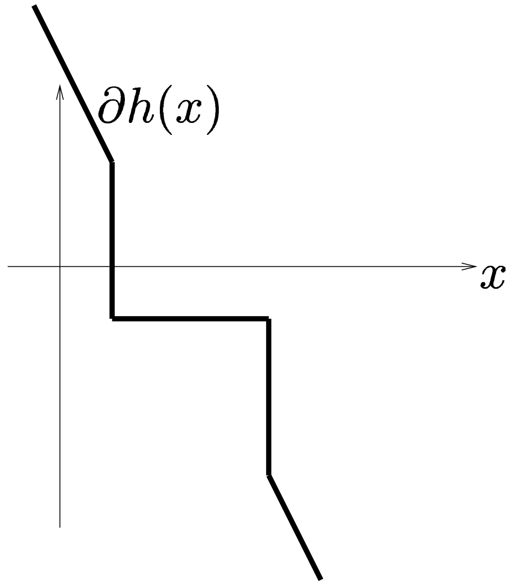
- The function $q(\boldsymbol{\mu}) + (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu})^{\mathrm{T}}\mathbf{g}(\mathbf{x})$ is linear w.r.t. $\bar{\boldsymbol{\mu}}$ and *overestimates* $q(\boldsymbol{\mu})$ over $\mathbb{R}^m$

Example

- Let $h(x) = \min\{h_1(x), h_2(x)\}$, where $h_1(x) = 4 - |x|$ and $h_2(x) = 4 - (x-2)^2$
- Then, $h(x) = \begin{cases} 4 - x, & 1 \leq x \leq 4, \\ 4 - (x-2)^2, & x \leq 1,\ x \geq 4 \end{cases}$



- h is non-differentiable at $x = 1$ and $x = 4$, since its graph has non-unique supporting hyperplanes there



$$\partial h(x) = \begin{cases} \{-1\}, & 1 < x < 4 \\ \{4 - 2x\}, & x < 1,\ x > 4 \\ [-1, 2]\,, & x = 1 \\ [-4, -1]\,, & x = 4 \end{cases}$$

- The subdifferential is here either a singleton (at differentiable points) or an interval (at non-differentiable points)

The Lagrangian dual problem

- *Let* $\boldsymbol{\mu} \in \mathbb{R}^m$. *Then,* $\partial q(\boldsymbol{\mu}) = \operatorname{conv}\{\, \mathbf{g}(\mathbf{x}) \mid \mathbf{x} \in X(\boldsymbol{\mu})\,\}$
- *Let* $\boldsymbol{\mu} \in \mathbb{R}^m$. *The dual function* q *is differentiable at* $\boldsymbol{\mu}$ if and only if $\{\, \mathbf{g}(\mathbf{x}) \mid \mathbf{x} \in X(\boldsymbol{\mu})\,\}$ *is a singleton set. Then,*

$$\nabla q(\boldsymbol{\mu}) = \mathbf{g}(\mathbf{x}),$$

for every $\mathbf{x} \in X(\boldsymbol{\mu})$

- *Holds in particular if the Lagrangian subproblem has a unique solution* $\Leftrightarrow$ *The solution set* $X(\boldsymbol{\mu})$ *is a singleton True, e.g., when* X *is convex,* f *strictly convex on* X, *and* g_i *convex on* X $\forall i$

How do we write the subdifferential of h?

Theorem: *If* $h(\mathbf{x}) = \min_{i=1,\ldots,m} h_i(\mathbf{x})$, *where each function* h_i *is concave and differentiable on* $\mathbb{R}^n$, *then* h *is a concave function on* $\mathbb{R}^n$

- Define the set $\mathcal{I}(\mathbf{x}) \subseteq \{1,\ldots,m\}$ by the active segments at $\mathbf{x}$:

$$\begin{cases} i \in \mathcal{I}(\mathbf{x}) & \text{if } h(\mathbf{x}) = h_i(\mathbf{x}), \\ i \notin \mathcal{I}(\mathbf{x}) & \text{if } h(\mathbf{x}) < h_i(\mathbf{x}), \end{cases} \qquad i \in \{1,\ldots,m\}$$

- Then, the subdifferential $\partial h(\mathbf{x})$ is the *convex hull* of the gradients $\{\nabla h_i(\mathbf{x}) \mid i \in \mathcal{I}(\mathbf{x})\}$:

$$\partial h(\mathbf{x}) = \left\{ \xi = \sum_{i \in \mathcal{I}(\mathbf{x})} \lambda_i \nabla h_i(\mathbf{x}) \,\middle|\, \sum_{i \in \mathcal{I}(\mathbf{x})} \lambda_i = 1;\; \lambda_i \geq 0, i \in \mathcal{I}(\mathbf{x}) \right\}$$

Optimality conditions for the dual problem

- For a differentiable, concave function h it holds that

$$\mathbf{x}^* \in \arg\max_{\mathbf{x}\in\mathbb{R}^n} h(\mathbf{x}) \quad \Longleftrightarrow \quad \nabla h(\mathbf{x}^*) = \mathbf{0}^n$$

- Theorem: *Assume that h is concave on $\mathbb{R}^n$. Then,*

$$\mathbf{x}^* \in \arg\max_{\mathbf{x}\in\mathbb{R}^n} h(\mathbf{x}) \quad \Longleftrightarrow \quad \mathbf{0}^n \in \partial h(\mathbf{x}^*)$$

- *Proof.* Suppose that $\mathbf{0}^n \in \partial h(\mathbf{x}^*) \Longrightarrow$
$h(\mathbf{x}) \leq h(\mathbf{x}^*) + (\mathbf{0}^n)^{\mathrm{T}}(\mathbf{x} - \mathbf{x}^*)$ for all $\mathbf{x} \in \mathbb{R}^n$, that is,
$h(\mathbf{x}) \leq h(\mathbf{x}^*)$ for all $\mathbf{x} \in \mathbb{R}^n$
Suppose that $\mathbf{x}^* \in \arg\max_{\mathbf{x}\in\mathbb{R}^n} h(x) \Longrightarrow$
$h(\mathbf{x}) \leq h(\mathbf{x}^*) = h(\mathbf{x}^*) + (\mathbf{0}^n)^{\mathrm{T}}(\mathbf{x} - \mathbf{x}^*)$ for all $\mathbf{x} \in \mathbb{R}^n$,
that is, $\mathbf{0}^n \in \partial h(\mathbf{x}^*)$ $\square$

- The example: $0 \in \partial h(1) \Longrightarrow x^* = 1$
- For optimization with constraints the KKT conditions are generalized:

$$\mathbf{x}^* \in \arg\max_{\mathbf{x}\in X} h(\mathbf{x}) \quad \Longleftrightarrow \quad \partial h(\mathbf{x}^*) \cap N_X(\mathbf{x}^*) \neq \emptyset,$$

where $N_X(\mathbf{x}^*)$ is the normal cone to X at $\mathbf{x}^*$, that is, the conical hull of the active constraints' normals at $\mathbf{x}^*$

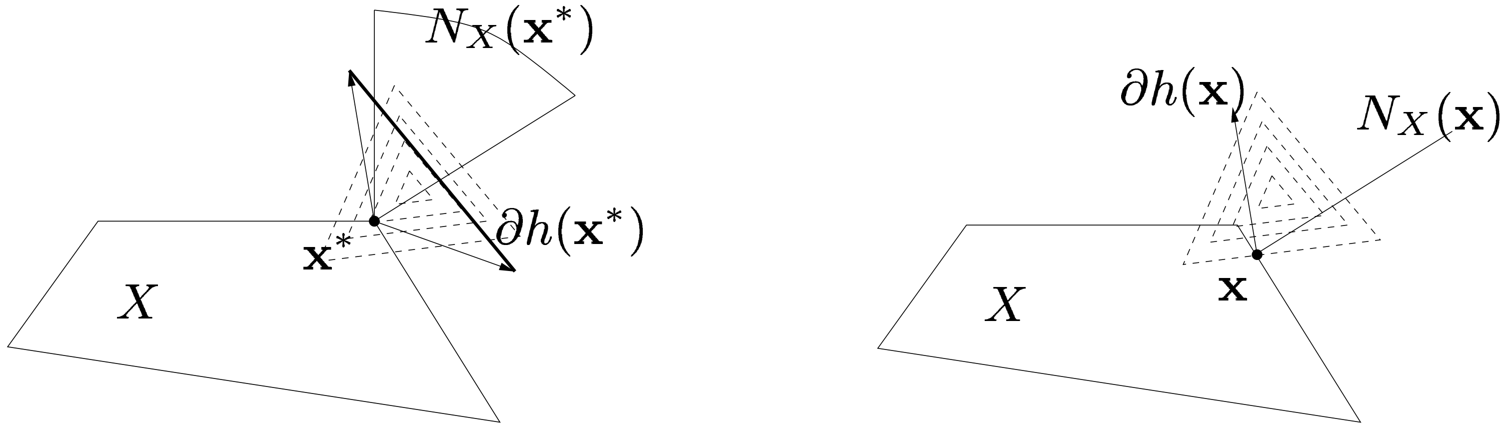


Figure 3: An optimal solution $\mathbf{x}^*$ A non-optimal solution $\mathbf{x}$

- In the case of the dual problem we have only sign conditions $\boldsymbol{\mu} \geq \mathbf{0}^m$

- Consider the dual problem

$$q^* = \underset{\boldsymbol{\mu} \geq \mathbf{0}^m}{\text{maximize}}\ q(\boldsymbol{\mu})$$

- $\boldsymbol{\mu}^* \geq \mathbf{0}^m$ is then optimal if and only if there exists a subgradient $\mathbf{g} \in \partial q(\boldsymbol{\mu}^*)$ for which the following holds:

$$\mathbf{g} \leq \mathbf{0}^m; \quad \mu_i^* g_i = 0,\ i = 1, \ldots, m$$

- Compare with a one-dimensional max-problem (h concave):

$$x^* \geq 0 \text{ is optimal} \quad \Leftrightarrow \quad h'(x^*) \leq 0; \quad x^* \cdot h'(x^*) = 0$$

A subgradient method for the dual problem

- Subgradient methods extend gradient projection methods from C^1 to general convex (or, concave) functions, generating a sequence of dual vectors in $\mathbb{R}^m_+$ using a single subgradient in each iteration
- The simplest type of iteration has the form

$$\begin{aligned} \boldsymbol{\mu}^{k+1} &= \mathrm{Proj}_{\mathbb{R}^m_+}[\boldsymbol{\mu}^k + \alpha_k \mathbf{g}^k] \\ &= [\boldsymbol{\mu}^k + \alpha_k \mathbf{g}^k]_+ \\ &= (\mathrm{maximum}\,\{0, (\boldsymbol{\mu}^k)_i + \alpha_k (\mathbf{g}^k)_i\})_{i=1}^m, \end{aligned} \tag{10}$$

where k is the iteration counter and $\mathbf{g}^k \in \partial q(\boldsymbol{\mu}^k)$ is arbitrarily chosen

- We often write $\mathbf{g}^k = \mathbf{g}(\mathbf{x}^k)$, where $\mathbf{x}^k \in \arg\min_{\mathbf{x} \in X} \, L(\mathbf{x}, \boldsymbol{\mu}^k)$
- Main difference to C^1 case: an arbitrary subgradient $\mathbf{g}^k$ *may not be an ascent direction!*

$\Rightarrow$ Cannot make line searches; must use predetermined step lengths α_k

- *Suppose that* $\boldsymbol{\mu} \in \mathbb{R}^m_+$ *is not optimal in* $\max_{\boldsymbol{\mu} \geq \mathbf{0}^m} q(\boldsymbol{\mu})$ *Then, for every optimal solution* $\boldsymbol{\mu}^* \in U^*$

$$\|\boldsymbol{\mu}^{k+1} - \boldsymbol{\mu}^*\| < \|\boldsymbol{\mu}^k - \boldsymbol{\mu}^*\|$$

holds for every step length α_k *in the interval*

$$\alpha_k \in (0, 2[q^* - q(\boldsymbol{\mu}^k)]/\|\mathbf{g}^k\|^2)$$

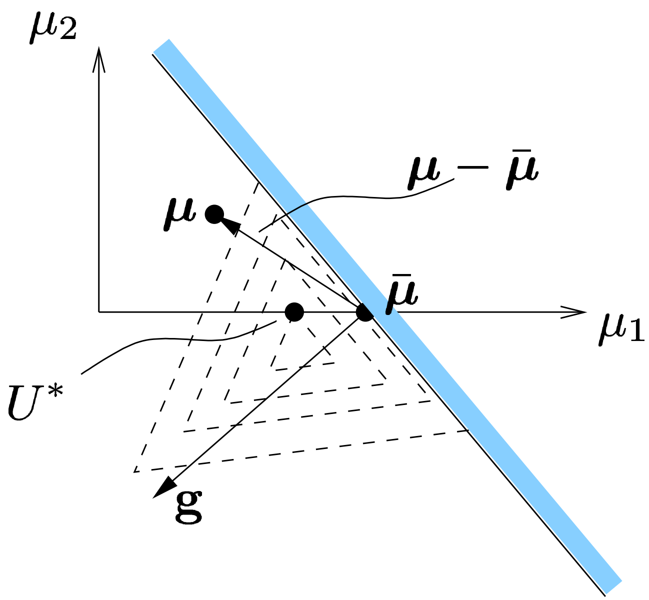
- Why? Let $\mathbf{g} \in \partial q(\bar{\boldsymbol{\mu}})$, and let U^* be the set of optimal solutions to $\max_{\boldsymbol{\mu} \geq \mathbf{0}^m} q(\boldsymbol{\mu})$ Then,

$$U^* \subseteq \{\, \boldsymbol{\mu} \in \mathbb{R}^m \mid \mathbf{g}^{\mathrm{T}}(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}) \geq 0 \,\}$$

  In other words, $\mathbf{g}$ defines a half-space that contains the set of optimal solutions.

- Good news: If the step length α_k is small enough we get closer to the set of optimal solutions!

A (sub)gradient defines a halfspace containing the optimal set



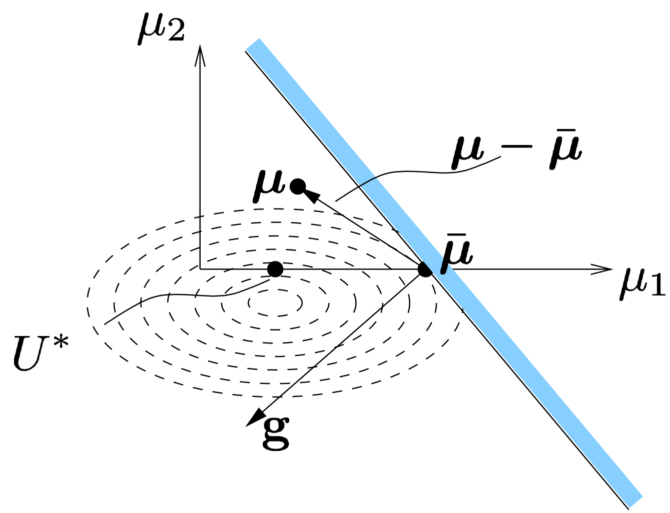


Figure 4: q non-differentiable $\qquad$ q differentiable

$$\mathbf{g} \in \partial q(\bar{\boldsymbol{\mu}}) \quad \Rightarrow \quad U^* \subseteq \{ \boldsymbol{\mu} \in \mathbb{R}^m \mid \mathbf{g}^{\mathrm{T}}(\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}) \geq 0 \}$$

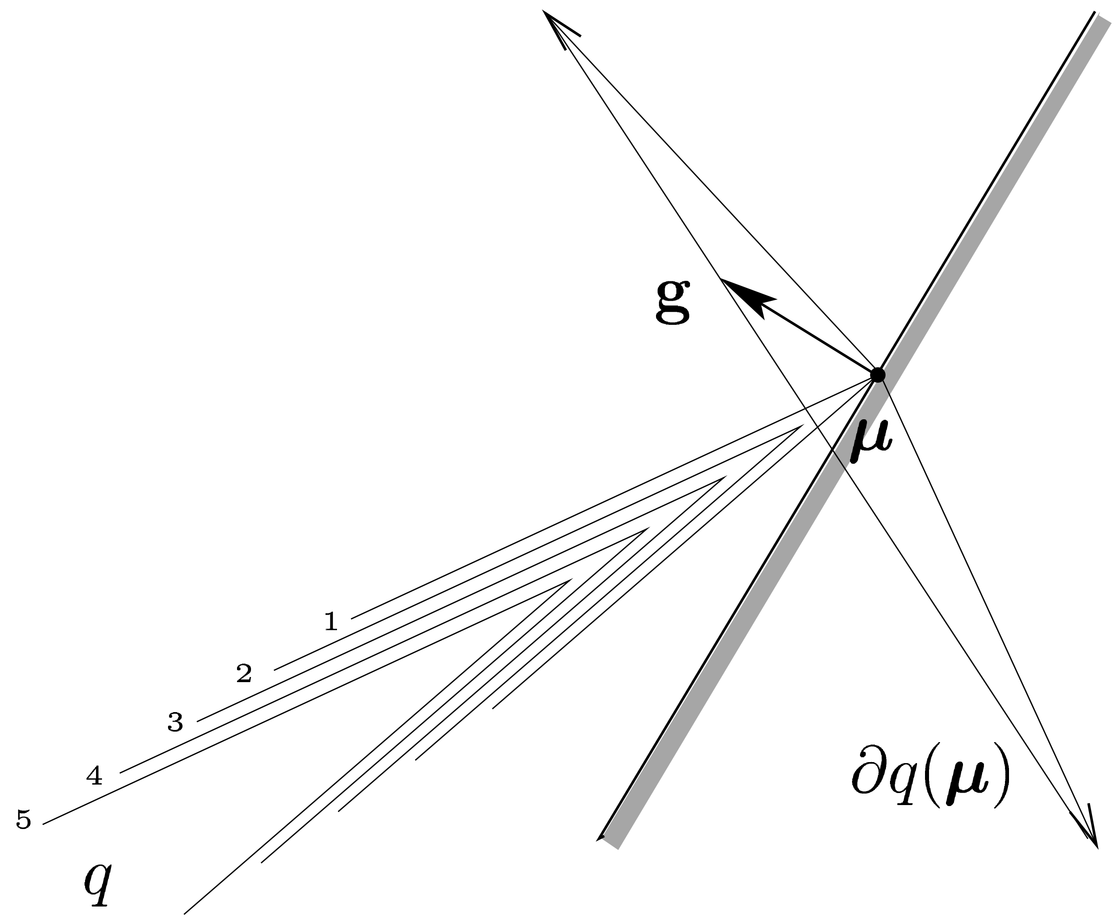


Figure 5: The half-space defined by a subgradient $\mathbf{g} \in q(\boldsymbol{\mu})$.

Note that this subgradient is *not an ascent direction*

Polyak step length rule:

$$\sigma \leq \alpha_k \leq 2[q^* - q(\boldsymbol{\mu}^k)]/\|\mathbf{g}^k\|^2 - \sigma, \quad k = 1, 2, \ldots \tag{11}$$

- $\sigma > 0 \Rightarrow$ step lengths α_k don't converge to 0 or a too large value
- Bad news: Utilizes knowledge of the optimal value q^*! But: q^* can be replaced by $\bar{q}_k \geq q^*$

The divergent series step length rule:

$$\alpha_k > 0, \; k = 1, 2, \ldots ; \quad \lim_{k \to \infty} \alpha_k = 0; \quad \sum_{s=1}^{\infty} \alpha_s = +\infty \tag{12}$$

- Additional condition often added:

$$\sum_{s=1}^{\infty} \alpha_s^2 < +\infty \tag{13}$$

Convergence results

- Suppose that f and $\mathbf{g}$ are continuous, X is compact, $\exists \mathbf{x} \in X : \mathbf{g}(\mathbf{x}) < \mathbf{0}$, and consider the problem

$$f^* = \inf\{ f(\mathbf{x}) \,|\, \mathbf{x} \in X, \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \} \tag{14}$$

(a) *Let* $\{\boldsymbol{\mu}^k\}$ *be generated by the method on p. 36, under the Polyak step length rule* (11), *where* $\sigma > 0$ *is small.*
Then, $\{\boldsymbol{\mu}^k\}$ *converges to an optimal solution to* (14)

(b) *Let* $\{\boldsymbol{\mu}^k\}$ *be generated by the method on p. 36, under the divergent series step length rule* (12).
Then, $\{q(\boldsymbol{\mu}^k)\} \to q^*$, *and* $\{\mathrm{dist}_{U^*}(\boldsymbol{\mu}^k)\} \to 0$

(c) *Let* $\{\boldsymbol{\mu}^k\}$ *be generated by the method on p. 36, under the divergent series step length rule* (12), (13).
Then, $\{\boldsymbol{\mu}^k\}$ *converges to an optimal solution to* (14)

Application to the Lagrangian dual problem

1. Given $\boldsymbol{\mu}^k \geq \mathbf{0}^m$
2. Solve the Lagrangian subproblem: $\min_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\mu}^k)$
3. Let an optimal solution to this problem be $\mathbf{x}^k = \mathbf{x}(\boldsymbol{\mu}^k)$
4. Calculate $\mathbf{g}(\mathbf{x}^k) \in \partial q(\boldsymbol{\mu}^k)$
5. Take a step $\alpha_k > 0$ in the direction of $\mathbf{g}(\mathbf{x}^k)$ from $\boldsymbol{\mu}^k$, according to a step length rule
6. Set any negative components of this vector to $0 \Rightarrow \boldsymbol{\mu}^{k+1}$
7. Let $k := k + 1$ and repeat from 2.

Additional algorithms

- We can choose the subgradient more carefully, to obtain *ascent* directions.
- Gather several subgradients at nearby points $\boldsymbol{\mu}^k$ and solve quadratic programming problems to find the best convex combination of them (*Bundle methods*)
- Pre-multiply the subgradient by some positive definite matrix $\Rightarrow$ methods similar to Newton methods (*Space dilation methods*)
- Pre-project the subgradient vector (onto the tangent cone of $\mathbb{R}^m_+$) $\Rightarrow$ step direction is a *feasible direction* (*Subgradient-projection methods*)

More to come

- Discrete optimization: The size of the duality gap, and the relation to the continuous relaxation.
- Convexification
- Primal feasibility heuristics
- Global optimality conditions for discrete optimization (and general problems)