Lecture 3: Lagrangian duality and algorithms for the Lagrangian dual problem

Michael Patriksson

19 March 2004

The Relaxation Theorem

- Problem: find

$$f^* := \operatorname*{infimum}_{\boldsymbol{x}} \; f(\boldsymbol{x}), \tag{1a}$$
$$\text{subject to } \boldsymbol{x} \in S, \tag{1b}$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ given function, $S \subseteq \mathbb{R}^n$.

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

$$f_R^* := \operatorname*{infimum}_{\boldsymbol{x}} \; f_R(\boldsymbol{x}), \tag{2a}$$
$$\text{subject to } \boldsymbol{x} \in S_R, \tag{2b}$$

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 Theorem: (a) [relaxation] $f_R^* \leq f^*$.
  (b) [infeasibility] *If* (2) *is infeasible, then so is* (1).
  (c) [optimal relaxation] *If the problem* (2) *has an optimal solution,* $\boldsymbol{x}_R^*$, *for which it holds that*

$$\boldsymbol{x}_R^* \in S \qquad \text{and} \qquad f_R(\boldsymbol{x}_R^*) = f(\boldsymbol{x}_R^*), \tag{3}$$

  then $\boldsymbol{x}_R^*$ *is an optimal solution to* (1) *as well.*

- *Proof portion.* For (c), note that

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

- Applications: exterior penalty methods yield lower bounds on f^*; Lagrangian relaxation yields lower bound on f^*.

Lagrangian relaxation

- Consider the optimization problem to find

$$f^* := \operatorname*{infimum}_{\boldsymbol{x}} f(\boldsymbol{x}), \tag{4a}$$

$$\text{subject to } \boldsymbol{x} \in X, \tag{4b}$$

$$g_i(\boldsymbol{x}) \leq 0, \qquad i = 1, \ldots, m, \tag{4c}$$

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 $X \subseteq \mathbb{R}^n$.

- For this problem, we assume that

$$-\infty < f^* < \infty, \tag{5}$$

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*

$$L(\boldsymbol{x}, \boldsymbol{\mu}) := f(\boldsymbol{x}) + \sum_{i=1}^{m} \mu_i g_i(\boldsymbol{x}) = f(\boldsymbol{x}) + \boldsymbol{\mu}^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{x}). \quad (6)$$

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

Lagrange multipliers and global optima

- *Let* $\boldsymbol{\mu}^*$ *be a Lagrange multiplier. Then,* $\boldsymbol{x}^*$ *is an optimal solution to* (4) *if and only if* $\boldsymbol{x}^*$ *is feasible in* (4) *and*

$$\boldsymbol{x}^* \in \arg\min_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\mu}^*), \text{ and } \mu_i^* g_i(\boldsymbol{x}^*) = 0, \; i = 1, \ldots, m. \tag{7}$$

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

- Seems to imply that there is a hidden convexity assumption here. Yes, there is. We show a Strong Duality Theorem later.

The Lagrangian dual problem associated with the Lagrangian relaxation

$$q(\boldsymbol{\mu}) := \operatorname*{infimum}_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\mu}) \tag{8}$$

is the *Lagrangian dual function*. The *Lagrangian dual problem* is to

$$\operatorname*{maximize}_{\boldsymbol{\mu}} \; q(\boldsymbol{\mu}), \tag{9a}$$

$$\text{subject to } \boldsymbol{\mu} \geq \mathbf{0}^m. \tag{9b}$$

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

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

- The *effective domain* of q is $D_q := \{ \boldsymbol{\mu} \in \mathbb{R}^m \mid q(\boldsymbol{\mu}) > -\infty \}$.
- *The effective domain D_q of q is convex, and q is concave on D_q.* □
- That the Lagrangian dual problem always is convex (we indeed maximize a concave function!) is very good news!
- 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* $\boldsymbol{x}$ *and* $\boldsymbol{\mu}$ *be feasible in* (4) *and* (9), *respectively. Then,*

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

In particular,

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

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

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

$$S := X \cap \{\, \boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}^m \,\}, \tag{10a}$$

$$S_R := X, \tag{10b}$$

$$f_R := L(\boldsymbol{\mu}, \cdot). \tag{10c}$$

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. There may be cases where no Lagrange multiplier exists even when there is no duality gap; in that case, the Lagrangian dual problem cannot have an optimal solution.

Global optimality conditions

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

$$\boldsymbol{\mu}^* \geq \mathbf{0}^m, \qquad \text{(Dual feasibility)} \tag{11a}$$

$$\boldsymbol{x}^* \in \arg\min_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\mu}^*), \qquad \text{(Lagrangian optimality)} \tag{11b}$$

$$\boldsymbol{x}^* \in X, \quad \boldsymbol{g}(\boldsymbol{x}^*) \leq \mathbf{0}^m, \qquad \text{(Primal feasibility)} \tag{11c}$$

$$\mu_i^* g_i(\boldsymbol{x}^*) = 0, \quad i = 1, \dots, m. \qquad \text{(Complementary slackness)} \tag{11d}$$

- If $\exists(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$, equivalent to zero duality gap and existence of Lagrange multipliers.

Saddle points

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

$$L(\boldsymbol{x}^*, \boldsymbol{\mu}) \leq L(\boldsymbol{x}^*, \boldsymbol{\mu}^*) \leq L(\boldsymbol{x}, \boldsymbol{\mu}^*), \qquad (\boldsymbol{x}, \boldsymbol{\mu}) \in X \times \mathbb{R}^m_+, \tag{12}$$

  holds.

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

Strong duality for convex programs, introduction

- Results so far have been rather non-technical to achieve: convexity of the dual problem comes with very few assumptions on the original, primal problem, and the characterization of the primal–dual set of optimal solutions is simple and also quite easily established.
- In order to establish *strong duality*, that is, to establish sufficient conditions under which there is no duality gap, however takes much more.
- In particular, as is the case with the KKT conditions we need regularity conditions (that is, constraint qualifications), and we also need to utilize separation theorems.

Strong duality Theorem

- Consider problem (4), 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 Slater condition:

$$\exists \boldsymbol{x} \in X \text{ with } \boldsymbol{g}(\boldsymbol{x}) < \mathbf{0}^m. \tag{13}$$

- *Suppose that* (5) *and Slater's CQ* (13) *hold for the (convex) problem* (4).
- (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 the infimum in* (4) *is attained at some* $\boldsymbol{x}^*$, *then the pair* $(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ *satisfies the global optimality conditions* (11).
- (c) *If the functions* f *and* g_i *are in* C^1 *then the condition* (11b) *can be written as a variational inequality. If further* X *is open (for example,* $X = \mathbb{R}^n$*) then the conditions* (11) *are the same as the KKT conditions.*
- Similar statements for the case of also having linear equality constraints.
- If all constraints are linear we can remove the Slater condition.

Examples, I: An explicit, differentiable dual problem

- Consider the problem to

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

- Let $g(\boldsymbol{x}) := -x_1 - x_2 + 4$ and $X := \{ (x_1, x_2) \mid x_j \geq 0,\ j = 1, 2 \}$.

- The Lagrangian dual function is

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

- 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 $q(\mu) = f(\boldsymbol{x}(\mu)) - \mu(x_1(\mu) + x_2(\mu) - 4) = 4\mu - \frac{\mu^2}{2}$.
- Note that q is strictly concave, and it is differentiable everywhere (due to the fact that f, g are differentiable and $\boldsymbol{x}(\mu)$ is unique).

- We then have that $q'(\mu) = 4 - \mu = 0 \iff \mu = 4$. As $\mu = 4 \geq 0$, it is the optimum in the dual problem! $\mu^* = 4; \boldsymbol{x}^* = (x_1(\mu^*), x_2(\mu^*))^{\mathrm{T}} = (2, 2)^{\mathrm{T}}$.
- Also: $f(\boldsymbol{x}^*) = q(\mu^*) = 8$.
- This is an example where the dual function is differentiable. In this particular case, the optimum $\boldsymbol{x}^*$ is also unique, and is automatically given by $\boldsymbol{x}^* = \boldsymbol{x}(\mu)$.

Examples, II: An implicit, non-differentiable dual problem

- Consider the linear programming problem to

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

- The optimal solution is $\boldsymbol{x}^* = (3/2, 0)^{\mathrm{T}}$, $f(\boldsymbol{x}^*) = -3/2$.
- Consider Lagrangian relaxing the first constraint,

obtaining

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

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

$$= \begin{cases} -3 + 5\mu, & 0 \le \mu \le 1/4, \\ -2 + \mu, & 1/4 \le \mu \le 1/2, \\ -3\mu, & 1/2 \le \mu. \end{cases}$$

- We have that $\mu^* = 1/2$, and hence $q(\mu^*) = -3/2$. For linear programs, we have strong duality, but how do we obtain the optimal primal solution from μ^*? q is non-differentiable at μ^*. We utilize the characterization given in (11).

- First, at μ^*, it is clear that $X(\mu^*)$ is the set $\{ \binom{2\alpha}{0} \mid 0 \leq \alpha \leq 1 \}$. Among the subproblem solutions, we next have to find one that is primal feasible as well as complementary.
- Primal feasibility means that $2 \cdot 2\alpha + 2 \cdot 0 \leq 3 \Longleftrightarrow \alpha \leq 3/4$.
- Further, complementarity means that $\mu^* \cdot (2x_1^* + 4x_2^* - 3) = 0 \Longleftrightarrow \alpha = 3/4$, since $\mu^* \neq 0$. We conclude that the only primal vector $\boldsymbol{x}$ that satisfies the system (11) together with the dual optimal solution $\mu^* = 1/2$ is $\boldsymbol{x}^* = (3/2, 0)^{\mathrm{T}}$.
- Observe finally that $f^* = q^*$.

- Why must $\mu^* = 1/2$? According 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” (it is between 0 and 2), the value of μ^* has to be such that the cost term for x_1 in $L(\boldsymbol{x}, \mu^*)$ is identically zero! That is, $-1 + \mu^* \cdot 2 = 0$ implies that $\mu^* = 1/2$!

- A non-coordinability phenomenon (non-unique subproblem solution means that the optimal solution is not obtained automatically).
- In non-convex cases, the optimal solution may not be among the points in $X(\boldsymbol{\mu}^*)$. What do we do then??

Subgradients of convex functions

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

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \boldsymbol{p}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}), \qquad \boldsymbol{y} \in \mathbb{R}^n. \tag{14}$$

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

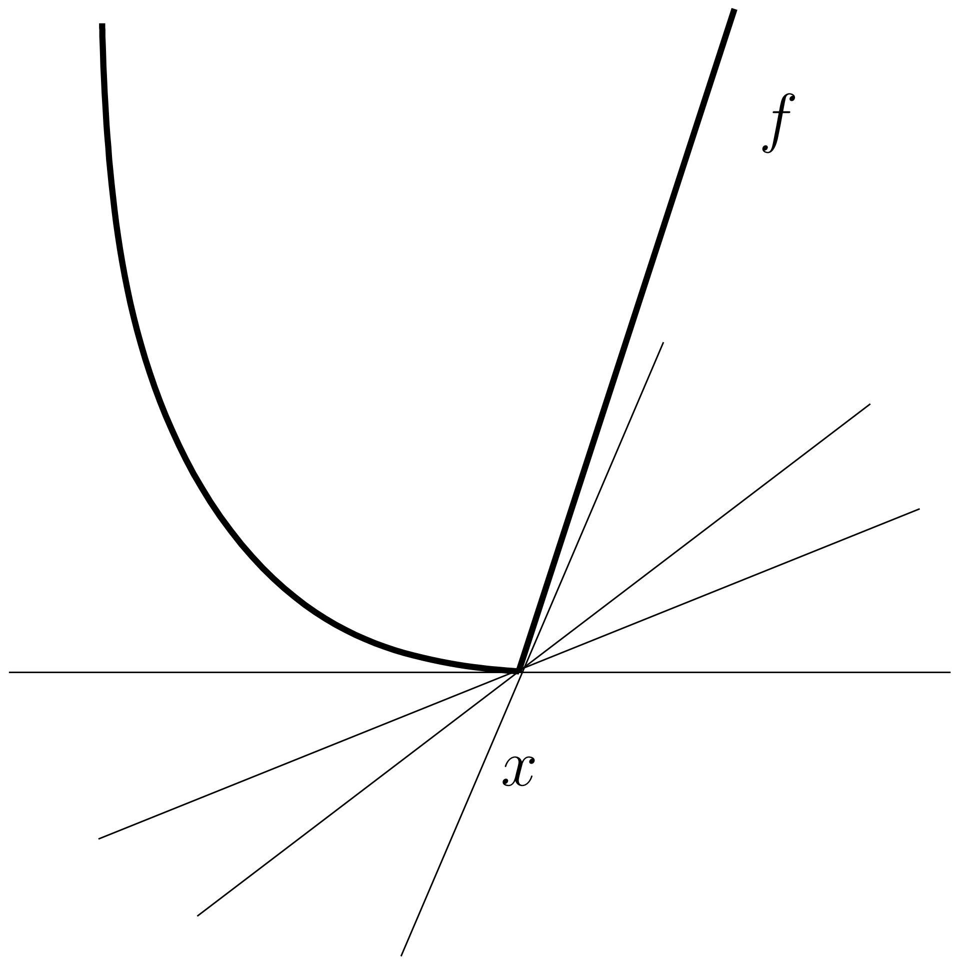


Figure 1: Four possible slopes of the convex function f at x.

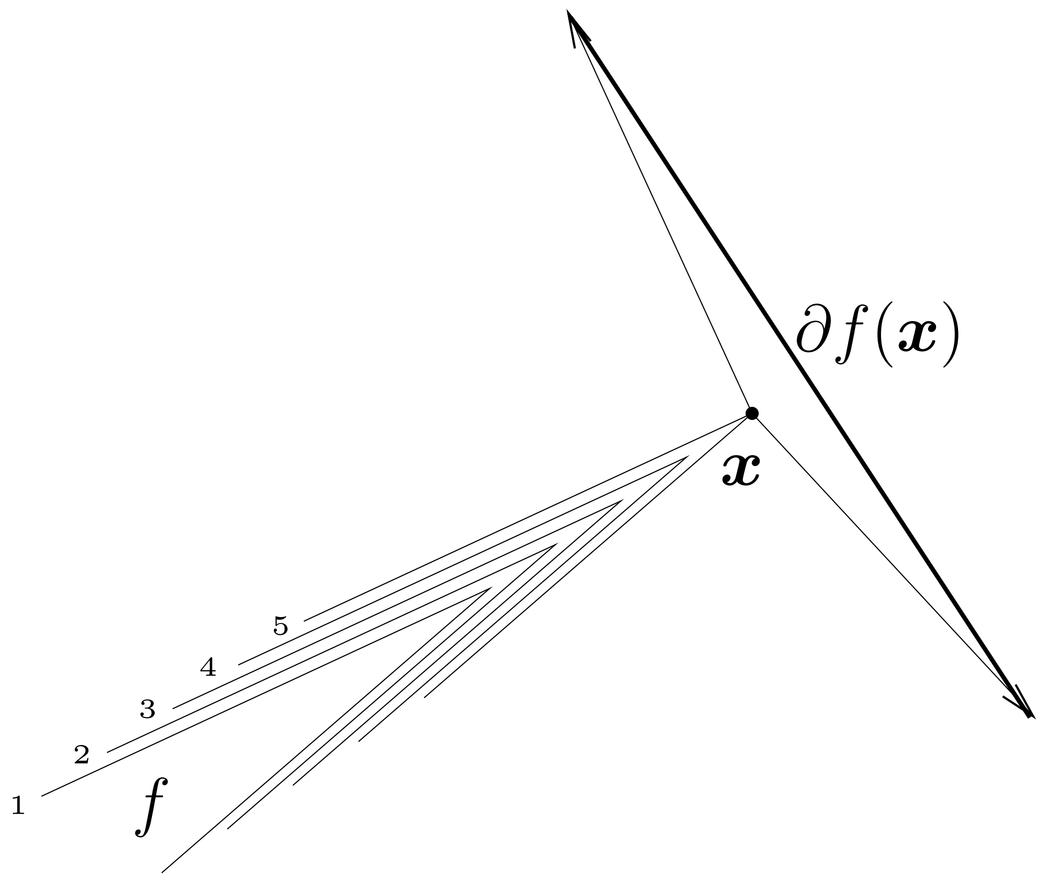


Figure 2: The subdifferential of a convex function f at $\boldsymbol{x}$.

- *The convex function f is differentiable at $\boldsymbol{x}$ exactly when there exists one and only one subgradient of f at $\boldsymbol{x}$, which then is the gradient of f at $\boldsymbol{x}$, $\nabla f(\boldsymbol{x})$.*

Differentiability of the Lagrangian dual function: Introduction

- Consider the problem (4), under the assumption that

$$f, g_i \, (i = 1, \ldots, m) \text{ continuous; } X \text{ nonempty and compact.} \tag{15}$$

- Then, the set of solutions to the Lagrangian subproblem,

$$X(\boldsymbol{\mu}) := \arg \operatorname*{minimum}_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\mu}), \qquad \boldsymbol{\mu} \in \mathbb{R}^m, \tag{16}$$

  is non-empty and compact for every $\boldsymbol{\mu}$.
- We develop the *sub*-differentiability properties of the function q.

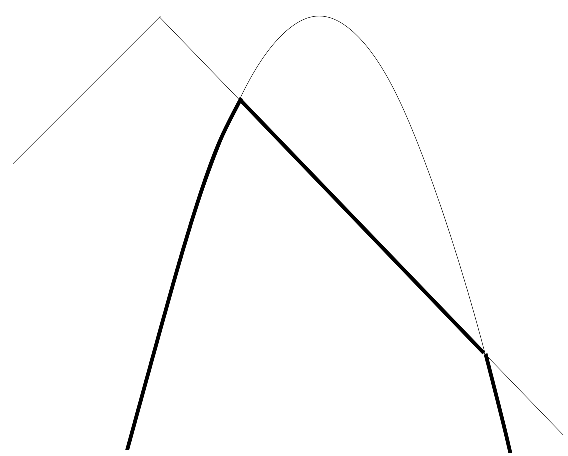
Subgradients and gradients of q

- *Suppose that, in the problem* (4), (15) *holds.*
- *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 $\boldsymbol{x} \in X(\boldsymbol{\mu})$, then $\boldsymbol{g}(\boldsymbol{x})$ is a subgradient to q at $\boldsymbol{\mu}$, that is, $\boldsymbol{g}(\boldsymbol{x}) \in \partial q(\boldsymbol{\mu})$.*
- *Proof.* Let $\bar{\boldsymbol{\mu}} \in \mathbb{R}^m$ be arbitrary. We have that

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

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, $h(x) = \begin{cases} 4 - x, & 1 \le x \le 4, \\ 4 - (x-2)^2, & x \le 1,\ x \ge 4. \end{cases}$



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

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

  for every $\boldsymbol{x} \in X(\boldsymbol{\mu})$.
- *Holds in particular if the Lagrangian subproblem has a unique solution [*$X(\boldsymbol{\mu})$ *is a singleton set]. In particular, satisfied if* X *is convex,* f *is strictly convex on* X*, and* g_i $(i = 1, \ldots, m)$ *are convex;* q *then even in* C^1. □

- How do we write the subdifferential of h?
- Theorem: *If* $h(\boldsymbol{x}) = \min_{i=1,\ldots,m} h_i(\boldsymbol{x})$, *where each function* h_i *is concave and differentiable on* $\mathbb{R}^n$, *then* h *is a concave function on* $\mathbb{R}^n$.
- Let $\mathcal{I}(\bar{\boldsymbol{x}}) \subseteq \{1, \ldots, m\}$ be defined by $h(\bar{\boldsymbol{x}}) = h_i(\bar{\boldsymbol{x}})$ for $i \in \mathcal{I}(\bar{\boldsymbol{x}})$ and $h(\bar{\boldsymbol{x}}) < h_i(\bar{\boldsymbol{x}})$ for $i \notin \mathcal{I}(\bar{\boldsymbol{x}})$ (the active segments at $\bar{\boldsymbol{x}}$).
- Then, the subdifferential $\partial h(\bar{\boldsymbol{x}})$ is the convex hull of $\{\nabla h_i(\bar{\boldsymbol{x}}) \mid i \in \mathcal{I}(\bar{\boldsymbol{x}})\}$, that is,

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

Optimality conditions for the dual problem

- For a differentiable, concave function h it holds that

$$\boldsymbol{x}^* \in \arg\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{maximum}}\ h(\boldsymbol{x}) \quad \iff \quad \nabla h(\boldsymbol{x}^*) = \mathbf{0}^n.$$

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

$$\boldsymbol{x}^* \in \arg\underset{\boldsymbol{x}\in\mathbb{R}^n}{\text{maximum}}\ h(\boldsymbol{x}) \quad \iff \quad \mathbf{0}^n \in \partial h(\boldsymbol{x}^*).$$

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

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

$$\boldsymbol{x}^* \in \arg\operatorname*{maximum}_{\boldsymbol{x} \in X} \; h(\boldsymbol{x}) \quad \Longleftrightarrow \quad \partial h(\boldsymbol{x}^*) \cap N_X(\boldsymbol{x}^*) \neq \emptyset,$$

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

- In the case of the dual problem we have only sign conditions.
- Consider the dual problem (9), and let $\boldsymbol{\mu}^* \geq \mathbf{0}^m$. It is then optimal in (9) if and only if there exists a subgradient $\boldsymbol{g} \in \partial q(\boldsymbol{\mu}^*)$ for which the following holds:

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

- Compare with a one-dimensional max-problem (h concave): x^* is optimal if and only if

$$h'(x^*) \leq 0; \quad x^* \cdot h'(x^*) = 0; \quad x^* \geq 0.$$

A subgradient method for the dual problem

- Subgradient methods extend gradient projection methods from the 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

$$\begin{aligned} \boldsymbol{\mu}_{k+1} &= \mathrm{Proj}_{\mathbb{R}^m_+}[\boldsymbol{\mu}_k + \alpha_k \boldsymbol{g}_k] && \text{(17a)} \\ &= [\boldsymbol{\mu}_k + \alpha_k \boldsymbol{g}_k]_+ && \text{(17b)} \\ &= (\text{maximum}\,\{0, (\boldsymbol{\mu}_k)_i + \alpha_k (\boldsymbol{g}_k)_i\})_{i=1}^m, && \text{(17c)} \end{aligned}$$

where $\boldsymbol{g}_k \in \partial q(\boldsymbol{\mu}_k)$ is arbitrarily chosen.

- We often use $\boldsymbol{g}_k = \boldsymbol{g}(\boldsymbol{x}_k)$, where $\boldsymbol{x}_k \in \arg\operatorname{minimum}_{\boldsymbol{x} \in X} \; L(\boldsymbol{x}, \boldsymbol{\mu}_k)$.
- Main difference to C^1 case: an arbitrary subgradient $\boldsymbol{g}_k$ may not be an ascent direction!
- Cannot do line searches; must use predetermined step lengths α_k.
- *Suppose that* $\boldsymbol{\mu} \in \mathbb{R}^m_+$ *is not optimal in* (9). *Then, for every optimal solution* $\boldsymbol{\mu}^* \in U^*$ *in* (9),

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

holds for every step length α_k *in the interval*

$$\alpha_k \in (0, 2[q^* - q(\boldsymbol{\mu}_k)]/\|\boldsymbol{g}_k\|^2). \qquad (18)$$

- Why? Let $\boldsymbol{g} \in \partial q(\bar{\boldsymbol{\mu}})$, and let U^* be the set of optimal solutions to (9). Then,

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

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

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

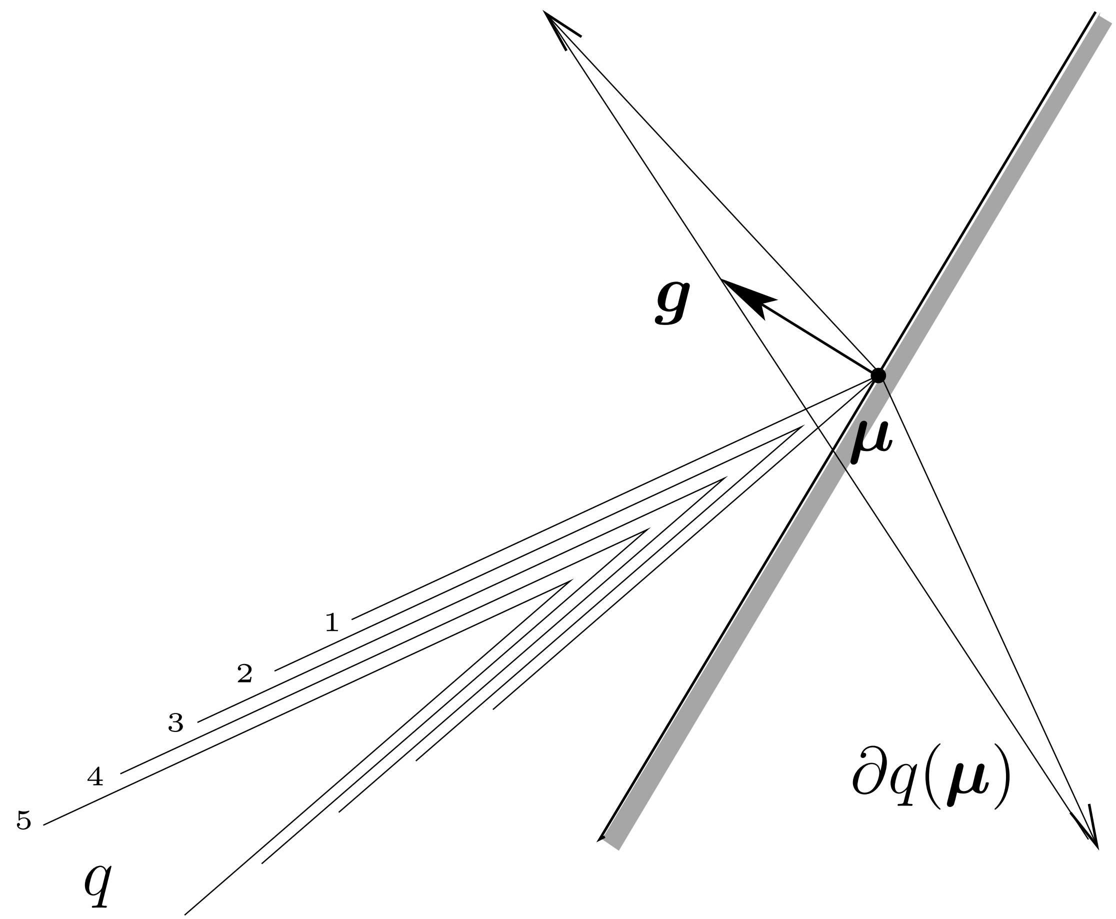


Figure 3: The half-space defined by the subgradient $\boldsymbol{g}$ of q at $\boldsymbol{\mu}$. Note that the subgradient is not an ascent direction.

- *Polyak step length rule*:

$$\sigma \leq \alpha_k \leq 2[q^* - q(\boldsymbol{\mu}_k)]/\|\boldsymbol{g}_k\|^2 - \sigma, \quad k = 1, 2, \ldots . \tag{19}$$

- $\sigma > 0$ makes sure that we do not allow the step lengths to converge to zero or a too large value.
- Bad news: Utilizes knowledge of the optimal value q^*! (Can be replaced by an upper bound on q^* which is updated.)
- The *divergent series* step length rule:

$$\alpha_k > 0, \ k = 1, 2, \ldots ; \quad \lim_{k \to \infty} \alpha_k = 0; \quad \sum_{s=1}^{\infty} \alpha_s = +\infty. \tag{20}$$

- Additional condition often added:

$$\sum_{s=1}^{\infty} \alpha_s^2 < +\infty. \tag{21}$$

- Suppose that the problem (4) is feasible, and that (15) and (13) hold.
- (a) *Let* $\{\boldsymbol{\mu}_k\}$ *be generated by the method* (17), *under the Polyak step length rule* (19), *where* σ *is a small positive number. Then,* $\{\boldsymbol{\mu}_k\}$ *converges to an optimal solution to* (9).
- (b) *Let* $\{\boldsymbol{\mu}_k\}$ *be generated by the method* (17), *under the divergent step length rule* (20). *Then,* $\{q(\boldsymbol{\mu}_k)\} \rightarrow q^*$, *and* $\{\text{dist}_{U^*}(\boldsymbol{\mu}_k)\} \rightarrow 0$.
- (c) *Let* $\{\boldsymbol{\mu}_k\}$ *be generated by the method* (17), *under the divergent step length rule* (20), (21). *Then,* $\{\boldsymbol{\mu}_k\}$ *converges to an optimal solution to* (9). $\square$

Application to the Lagrangian dual problem

- Given $\boldsymbol{\mu}_k \geq \mathbf{0}^m$.
- Solve the Lagrangian subproblem to minimize $L(\boldsymbol{x}, \boldsymbol{\mu}_k)$ over $\boldsymbol{x} \in X$.
- Let an optimal solution to this problem be $\boldsymbol{x}_k$.
- Calculate $\boldsymbol{g}(\boldsymbol{x}_k) \in \partial q(\boldsymbol{\mu}_k)$.
- Take a (positive) step in the direction of $\boldsymbol{g}(\boldsymbol{x}_k)$ from $\boldsymbol{\mu}_k$, according to a step length rule.
- Set any negative components of this vector to zero.
- We have obtained $\boldsymbol{\mu}_{k+1}$.

Additional algorithms

- We can choose the subgradient more carefully, such that we will obtain ascent directions. This amounts to gathering several subgradients at nearby points and solving quadratic programming problems to find the best convex combination of them. (*Bundle methods.*)
- Pre-multiply the subgradient obtained by some positive definite matrix. We get methods similar to Newton methods. (*Space dilation methods.*)
- Pre-project the subgradient vector (onto the tangent cone of $\mathbb{R}^m_+$) so that the direction taken 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).