Lecture 7: Lagrangian duality

Michael Patriksson

5 February 2004

The Relaxation Theorem

- Problem: find

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

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

- A *relaxation* to (1) is a problem of the following form: to find

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

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*

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

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: Frank–Wolfe algorithm (linearizing f yields lower bounds on f^*, see Chapter 12); exterior penalty methods yield lower bounds on f^* (see Chapter 13); Lagrangian relaxation yields lower bound on f^* (this chapter!).

Lagrangian relaxation

- Consider the optimization problem to find

$$f^* := \underset{\boldsymbol{x}}{\text{infimum}}\ f(\boldsymbol{x}), \tag{4a}$$
$$\text{subject to } \boldsymbol{x} \in X, \tag{4b}$$
$$g_i(\boldsymbol{x}) \leq 0, \qquad i = 1, \dots, m, \tag{4c}$$

where $f : \mathbb{R}^n \mapsto \mathbb{R}$ and $g_i : \mathbb{R}^n \mapsto \mathbb{R}$ $(i = 1, 2, \dots, m)$ 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*

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

- 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*

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

- 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 prove a Strong Duality Theorem later.

The Lagrangian dual problem associated with the Lagrangian relaxation

- Let

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

  be the *Lagrangian dual function*, defined by the infimum value of the Lagrange function over X.
- 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 $\boldsymbol{\mu} \geq \mathbf{0}^m$,

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

  equals $-\infty$.

- 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.*
- *Proof.* Let $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{\mu}, \bar{\boldsymbol{\mu}} \in \mathbb{R}^m$, and $\alpha \in [0,1]$. We have that

$$L(\boldsymbol{x}, \alpha\boldsymbol{\mu} + (1-\alpha)\bar{\boldsymbol{\mu}}) = \alpha L(\boldsymbol{x}, \boldsymbol{\mu}) + (1-\alpha) L(\boldsymbol{x}, \bar{\boldsymbol{\mu}}).$$

  Take the inf over $\boldsymbol{x} \in X$ on both sides; then,

$$\begin{aligned}
\inf_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \alpha\boldsymbol{\mu} + (1-\alpha)\bar{\boldsymbol{\mu}}) &= \inf_{\boldsymbol{x} \in X} \{\alpha L(\boldsymbol{x}, \boldsymbol{\mu}) + (1-\alpha) L(\boldsymbol{x}, \bar{\boldsymbol{\mu}})\} \\
&\geq \inf_{\boldsymbol{x} \in X} \alpha L(\boldsymbol{x}, \boldsymbol{\mu}) + \inf_{\boldsymbol{x} \in X} (1-\alpha) L(\boldsymbol{x}, \bar{\boldsymbol{\mu}}) \\
&= \alpha \inf_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \boldsymbol{\mu}) + (1-\alpha) \inf_{\boldsymbol{x} \in X} L(\boldsymbol{x}, \bar{\boldsymbol{\mu}}),
\end{aligned}$$

  since $\alpha \in [0,1]$, and "sum of inf" may be smaller than "inf of sum." Hence,

$$q(\alpha\boldsymbol{\mu} + (1-\alpha)\bar{\boldsymbol{\mu}}) \geq \alpha q(\boldsymbol{\mu}) + (1-\alpha) q(\bar{\boldsymbol{\mu}}).$$

This inequality has two implications: if $\boldsymbol{\mu}$ and $\bar{\boldsymbol{\mu}}$ belong to D_q, then so does $\alpha\boldsymbol{\mu} + (1-\alpha)\bar{\boldsymbol{\mu}}$, so D_q is convex, and further, 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,

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

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

- *Proof.* For all $\boldsymbol{\mu} \geq \mathbf{0}^m$ and $\boldsymbol{x} \in X$ with $\boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}^m$,

$$q(\boldsymbol{\mu}) = \operatorname*{infimum}_{\boldsymbol{z}\in X} L(\boldsymbol{z}, \boldsymbol{\mu}) \leq f(\boldsymbol{x}) + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}) \leq f(\boldsymbol{x}),$$

so

$$q^* = \operatorname*{supremum}_{\boldsymbol{\mu}\geq\mathbf{0}^m} q(\boldsymbol{\mu}) \leq \operatorname*{infimum}_{\boldsymbol{x}\in X:\boldsymbol{g}(\boldsymbol{x})\leq\mathbf{0}^m} f(\boldsymbol{x}) = f^*.$$

- 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 $q^* = f^*$, we say that there is *no duality gap*. If there exists a Lagrange multiplier vector, then by the weak duality theorem, this implies that there is no duality gap. The converse is not true in general: there may be cases where no Lagrange multiplier exists even when there is no duality gap; in that case though, the Lagrangian dual problem cannot have an optimal solution.

On the statement of the problem (4)

- There are several ways in which the problem can be defined.
- Constraints can be placed within the definition of the *ground set* X (kept intact), or within the explicit constraints defined by the functions g_i (Lagrangian relaxed).
- How to distinguish between the two, that is, how to decide whether a constraint should be kept or be Lagrangian relaxed, depends on several factors.
- Keeping more constraints within X may result in a smaller duality gap, and with fewer multipliers also result in a simpler Lagrangian dual problem.
- On the other hand, the Lagrangian subproblem defining the dual function then becomes more complex and difficult to solve.
- There are few rules to follow: experiment and experience!

Global optimality conditions

- The following system characterizes every optimal primal and dual solution. It is applicable only in the presence of Lagrange multipliers; in other words, the system (11) is consistent if and only if there exists a Lagrange multiplier but no duality gap.
- *The vector $(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ is a pair of optimal primal solution and Lagrange multiplier if and only if*

$$\begin{aligned}
\boldsymbol{\mu}^* &\geq \mathbf{0}^m, && \textit{(Dual feasibility)} && \text{(11a)}\\
\boldsymbol{x}^* &\in \arg\min_{x \in X} L(\boldsymbol{x}, \boldsymbol{\mu}^*), && \textit{(Lagrangian optimality)} && \text{(11b)}\\
\boldsymbol{x}^* \in X, \quad \boldsymbol{g}(\boldsymbol{x}^*) &\leq \mathbf{0}^m, && \textit{(Primal feasibility)} && \text{(11c)}\\
\mu_i^* g_i(\boldsymbol{x}^*) = 0, \quad i &= 1, \ldots, m. && \textit{(Complementary slackness)} && \text{(11d)}
\end{aligned}$$

- *Proof.* Suppose that (11) is satisfied. We apply the Relaxation Theorem as follows. Consider the identification in (10), with $\boldsymbol{\mu} = \boldsymbol{\mu}^* \geq \mathbf{0}^m$ and $\boldsymbol{x}_R = \boldsymbol{x}^*$. Note these equivalences:
  1. The relaxed solution, $\boldsymbol{x}_R$, is a Lagrangian optimal solution. (Lagrangian optimality is fulfilled.)
  2. That $\boldsymbol{x}_R \in S$ means that it is feasible in the primal problem. (Primal feasibility is fulfilled.)
  3. That $f^* = f_R^*$ means that $(\boldsymbol{\mu}^*)^{\mathrm{T}} \boldsymbol{g}(\boldsymbol{x}^*) = 0$. (Complementary slackness is fulfilled.)

  So, if (11) is satisfied, then the vector $(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ is a pair of optimal primal solution and Lagrange multiplier.

  Conversely, if $(\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ is a pair of optimal primal solution and Lagrange multiplier, then they are particularly primal and dual feasible, respectively. The last two equations in (11) then follow from the previous theorem on global optima. $\square$

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,*

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

  holds.

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 the problem (4), where $f : \mathbb{R}^n \mapsto \mathbb{R}$ and g_i ($i = 1, \dots, m$) are convex functions and $X \subseteq \mathbb{R}^n$ is a convex set.
- For this problem, we introduce the following Slater condition:

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

- *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.
- If f is linear then we can state the following: *If both the primal and dual problems have feasible solutions, then they both have optimal solutions, and their optimal values are the same.* We will prove this elegantly in the LP chapters!

What is the dual of an LP problem?

- Consider the linear program to

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}}\ & \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}, & (14\text{a})\\ \text{subject to } & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, & (14\text{b})\\ & \boldsymbol{x} \geq \boldsymbol{0}^n, & (14\text{c}) \end{aligned}$$

where $\boldsymbol{A}$ is an $m \times n$ matrix and $\boldsymbol{c}$ (respectively, $\boldsymbol{b}$) is an n-vector (respectively, m-vector). If we let $X := \mathbb{R}^n_+$, then the Lagrangian dual problem is to

$$\begin{aligned} \underset{\boldsymbol{\lambda} \in \mathbb{R}^m}{\text{maximize}}\ & \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda}, & (15\text{a})\\ \text{subject to } & \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} \leq \boldsymbol{c}. & (15\text{b}) \end{aligned}$$

- The reason why we can write it in this form is that

$$q(\boldsymbol{\lambda}) := \underset{\boldsymbol{x} \geq \boldsymbol{0}^n}{\text{infimum}} \left\{ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} + \boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}) \right\} = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda} + \underset{\boldsymbol{x} \geq \boldsymbol{0}^n}{\text{infimum}}\, (\boldsymbol{c} - \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda})^{\mathrm{T}}\boldsymbol{x},$$

so that

$$q(\boldsymbol{\lambda}) = \begin{cases} \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda}, & \text{if } \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} \leq \boldsymbol{c}, \\ -\infty, & \text{otherwise.} \end{cases}$$

- More: $\boldsymbol{\lambda}$ unrestricted, because we can write $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ as $\boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b}$; $-\boldsymbol{A}\boldsymbol{x} \leq -\boldsymbol{b}$. Introduce $\boldsymbol{\mu}^+$ and $\boldsymbol{\mu}^-$. Lagrange function: $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} + (\boldsymbol{\mu}^+ - \boldsymbol{\mu}^-)^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x})$. Substitute $\boldsymbol{\lambda} = \boldsymbol{\mu}^+ - \boldsymbol{\mu}^-$.
- If both the primal and dual problems have feasible solutions, then they both have optimal solutions, satisfying strong duality ($\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}^* = \boldsymbol{b}^{\mathrm{T}}\boldsymbol{\lambda}^*$).
- More about dual problems in the LP chapters!

Examples, I: An explicit, differentiable dual problem

- Consider the problem to

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

- We consider the first constraint to be the complicated one, and hence define $g(\boldsymbol{x}) := -x_1 - x_2 + 4$ and let $X := \{ (x_1, x_2) \mid x_j \geq 0,\ j = 1, 2 \}$.

- The Lagrangian dual function is

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

- 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 $\nabla q(\mu) = 4 - \mu = 0 \iff \mu = 4$. 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

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

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

- Consider Lagrangian relaxing the first constraint, obtaining

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

$$q(\mu) = -3\mu + \operatorname*{minimum}_{0\le x_1\le 2}\{(-1+2\mu)x_1\} + \operatorname*{minimum}_{0\le x_2\le 1}\{(-1+4\mu)x_2\}$$
$$= \begin{cases} -3+5\mu, & 0\le\mu\le 1/4,\\ -2+\ \mu, & 1/4\le\mu\le 1/2,\\ \quad -3\mu, & 1/2\le\mu. \end{cases}$$

- 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 $\{\binom{2\alpha}{0} \mid 0\le\alpha\le 1\}$. Among the subproblem solutions, we next have to find one that is primal feasible as well as complementary.

- Primal feasibility means that $2\cdot 2\alpha + 2\cdot 0 \le 3 \iff \alpha \le 3/4$.
- Further, complementarity means that $\mu^*\cdot(2x_1^* + 4x_2^* - 3) = 0 \iff \alpha = 3/4$, 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 $f^* = q^*$.

*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

$$f(\boldsymbol{y}) \ge f(\boldsymbol{x}) + \boldsymbol{p}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}), \qquad \boldsymbol{y} \in \mathbb{R}^n. \tag{16}$$

- 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.*
- *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

$$f, g_i\,(i=1,\dots,m) \text{ are continuous; } X \text{ is nonempty and compact.} \tag{17}$$

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

$$X(\boldsymbol{\mu}) := \arg\operatorname*{minimum}_{\boldsymbol{x}\in X} L(\boldsymbol{x},\boldsymbol{\mu}), \qquad \boldsymbol{\mu}\in\mathbb{R}^m, \tag{18}$$

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), (17) 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

$$q(\bar{\boldsymbol{\mu}}) = \underset{\boldsymbol{y}\in X}{\text{infimum}}\ L(\boldsymbol{y}, \bar{\boldsymbol{\mu}}) \leq f(\boldsymbol{x}) + \bar{\boldsymbol{\mu}}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x})$$
$$= f(\boldsymbol{x}) + (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}) = q(\boldsymbol{\mu}) + (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{g}(\boldsymbol{x}),$$

which implies that $\boldsymbol{g}(\boldsymbol{x}) \in \partial q(\boldsymbol{\mu})$. □

- *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, that is, if the value of the vector of constraint functions is invariant over the set of solutions $X(\boldsymbol{\mu})$ to the Lagrangian subproblem. Then, we have that*

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

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

- *This result holds in particular if the Lagrangian subproblem has a unique solution, that is, $X(\boldsymbol{\mu})$ is a singleton set. In particular, this property is satisfied if further X is a convex set, f is strictly convex on X, and g_i $(i = 1, \dots, m)$ are convex, in which case q is even in C^1.* □

*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

$$\boldsymbol{\mu}_{k+1} = \mathrm{Proj}_{\mathbb{R}^m_+}[\boldsymbol{\mu}_k + \alpha_k \boldsymbol{g}_k] \tag{19a}$$
$$= [\boldsymbol{\mu}_k + \alpha_k \boldsymbol{g}_k]_+ \tag{19b}$$
$$= (\text{maximum}\,\{0, (\boldsymbol{\mu}_k)_i + \alpha_k (\boldsymbol{g}_k)_i\})_{i=1}^m, \tag{19c}$$

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 $\boldsymbol{x}_k \in \arg\text{minimum}_{\boldsymbol{x}\in X}\ L(\boldsymbol{x}, \boldsymbol{\mu}_k)$.

- 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

$$\alpha_k \in (0, 2[q^* - q(\boldsymbol{\mu}_k)]/\|\boldsymbol{g}_k\|^2). \tag{20}$$

- Good news: If the step length is small enough we get closer to every optimal solution!

- *Polyak step length rule*:

$$\sigma \leq \alpha_k \leq 2[q^* - q(\boldsymbol{\mu}_k)]/\|\boldsymbol{g}_k\|^2 - \sigma, \qquad k = 1, 2, \dots . \tag{21}$$

- $\sigma > 0$ makes sure that we do not allow the step lengths to converge to zero or a too large value.
- Bad news: Utilizes the knowledge of the optimal value q^*!
- Exists also other step length rules:

$$\alpha_k > 0, \quad k = 1, 2, \dots, \qquad \lim_{k \to \infty} \alpha_k = 0; \qquad \sum_{s=1}^{\infty} \alpha_s = +\infty. \tag{22}$$

- Called the *divergent series* step length rule. Additional condition often added:

$$\sum_{s=1}^{\infty} \alpha_s^2 < +\infty. \tag{23}$$

- Suppose that the problem (4) is feasible, and that (17) and (13) hold.
- (a) *Let* $\{\boldsymbol{\mu}_k\}$ *be generated by the method* (19), *under the Polyak step length rule* (21), *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* (19), *under the divergent step length rule* (22). *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* (19), *under the divergent step length rule* (22), (23). *Then,* $\{\boldsymbol{\mu}_k\}$ *converges to an optimal solution to* (9). $\square$