Lecture 3
Lagrangian duality, part II: Algorithms for the Lagrangian dual problem

Ann-Brith Strömberg

2018–01–19

Subgradients of convex functions

Definition of a subgradient

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

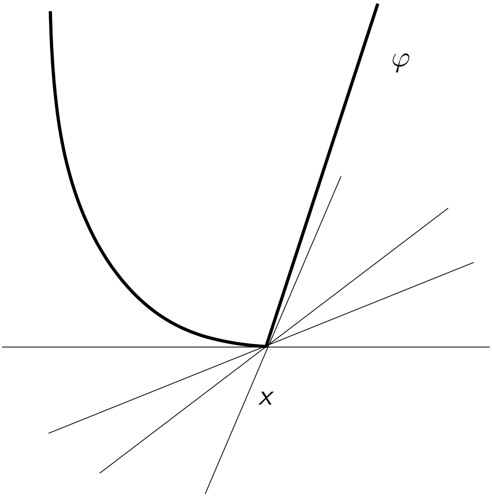
$$\varphi(\mathbf{y}) \geq \varphi(\mathbf{x}) + \mathbf{p}^\top(\mathbf{y} - \mathbf{x}), \qquad \mathbf{y} \in \mathbb{R}^n \tag{1}$$

- The set of such vectors $\mathbf{p}$ defines the *subdifferential* of φ at $\mathbf{x}$, and is denoted $\partial\varphi(\mathbf{x})$
- $\partial\varphi(\mathbf{x})$ is the collection of "slopes" of the function φ at $\mathbf{x}$

Properties of the subdifferential

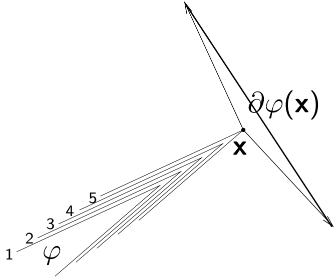
For every $\mathbf{x} \in \mathbb{R}^n$, $\partial\varphi(\mathbf{x})$ is a non-empty, convex, and compact set

Subgradients of convex functions—illustration



Figur: Four possible slopes of the convex function φ at x

Subdifferential of a convex function—illustration



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

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

Differentiability of the Lagrangian dual function

- Consider the problem

$$f^* := \operatorname*{infimum}_{\mathbf{x}} f(\mathbf{x}), \tag{2a}$$
$$\text{subject to } \mathbf{x} \in X, \tag{2b}$$
$$g_i(\mathbf{x}) \leq 0, \qquad i = 1, \dots, m, \tag{2c}$$

and assume that

$$f \text{ and } g_i, i = 1, \dots, m, \text{ are continuous;} \tag{3a}$$
$$X \text{ is nonempty and compact.} \tag{3b}$$

The set of solutions to the *Lagrangian subproblem*

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

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

Subgradients and gradients of q

- *Suppose that* (3) *holds (f, g_i continuous; $X \neq \emptyset$, compact) in the problem* (2):
  $f^* = \inf_{\mathbf{x}} \left\{ f(\mathbf{x}) \;\middle|\; \mathbf{x} \in X;\; g_i(\mathbf{x}) \leq 0,\; i = 1, \ldots, m \right\}$
- 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

Theorem: subgradient of the dual function

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. It follows that

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

Subgradients and gradients of q, cont'd

- Recall the *subgradient inequality* (1) for a *convex* function φ: $\mathbf{p}$ is a subgradient of φ at $\mathbf{x}$ if

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

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

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

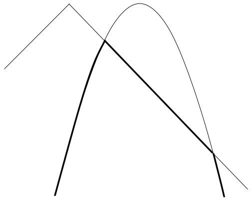
- The function $q(\boldsymbol{\mu}) + (\bar{\boldsymbol{\mu}} - \boldsymbol{\mu})^\top \mathbf{g}(\mathbf{x})$ is linear w.r.t. $\bar{\boldsymbol{\mu}}$ and *overestimates* $q(\boldsymbol{\mu})$ over $\mathbb{R}^m$

Example — a concave function

Let $\quad h_1(x) = 4 - |x| \quad$ and $\quad h_2(x) = 4 - (x-2)^2$

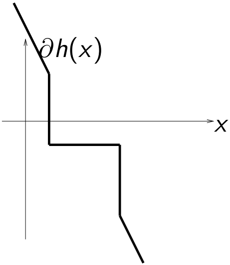
Define the function $h : \mathbb{R} \mapsto \mathbb{R}$ as $\quad h(x) := \min\left\{h_1(x), h_2(x)\right\}$

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



Example, cont'd — supdifferential of a concave 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 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, $\partial q(\boldsymbol{\mu}) = \operatorname{conv} \left\{ \mathbf{g}(\mathbf{x}) \mid \mathbf{x} \in X(\boldsymbol{\mu}) \right\}$

- Let $\boldsymbol{\mu} \in \mathbb{R}^m$. The dual function q is differentiable at $\boldsymbol{\mu}$ *if and only if* $\left\{ \mathbf{g}(\mathbf{x}) \mid \mathbf{x} \in X(\boldsymbol{\mu}) \right\}$ 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 $\forall i$ (e.g., f quadratic, X polyhedral, g_i linear)

How do we write the subdifferential of h?

Theorem

If $h(\mathbf{x}) := \min_{i=1,\dots,m}\{h_i(\mathbf{x})\}$, 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 $\mathcal{I}(\mathbf{x}) \subseteq \{1,\dots,m\}$ by the *active* segments at $\mathbf{x}$:

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

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

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

Optimality conditions for the dual problem

- For a differentiable, concave function h it holds that

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

Theorem

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

$$\mathbf{x}^* \in \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmax}}\ h(\mathbf{x}) \quad\Longleftrightarrow\quad \mathbf{0}^n \in \partial h(\mathbf{x}^*)$$

Proof

$\Leftarrow$ Suppose that $\mathbf{0}^n \in \partial h(\mathbf{x}^*) \Longrightarrow h(\mathbf{x}) \leq h(\mathbf{x}^*) + (\mathbf{0}^n)^\top(\mathbf{x}-\mathbf{x}^*)$ for all $\mathbf{x} \in \mathbb{R}^n$, that is, $h(\mathbf{x}) \leq h(\mathbf{x}^*)$ for all $\mathbf{x} \in \mathbb{R}^n$

$\Rightarrow$ Suppose that $\mathbf{x}^* \in \operatorname{argmax}_{\mathbf{x}\in\mathbb{R}^n} h(\mathbf{x}) \Longrightarrow$ $h(\mathbf{x}) \leq h(\mathbf{x}^*) = h(\mathbf{x}^*) + (\mathbf{0}^n)^\top(\mathbf{x}-\mathbf{x}^*)$ for all $\mathbf{x} \in \mathbb{R}^n$, that is, $\mathbf{0}^n \in \partial h(\mathbf{x}^*)$

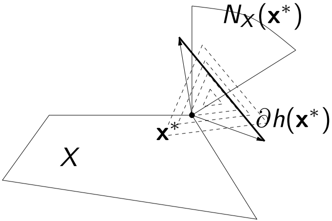
Optimality conditions for the dual problem, cont'd

- The example: $0 \in \partial h(1) \implies x^* = 1$

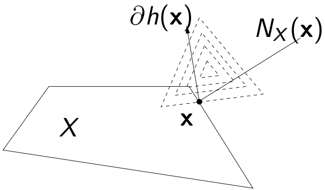
Generalization of the KKT conditions:

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

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 $\mathbf{x}^*$



Figur: An optimal solution $\mathbf{x}^*$



A non-optimal solution $\mathbf{x}$

Optimality conditions for the dual problem, cont'd

- The dual problem has 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:

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

- 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

$$\begin{aligned}
\boldsymbol{\mu}^{k+1} &:= \mathrm{proj}_{\mathbb{R}^m_+}\left[\boldsymbol{\mu}^k + \alpha_k \mathbf{g}^k\right] \\
&= \left[\boldsymbol{\mu}^k + \alpha_k \mathbf{g}^k\right]_+ \\
&= \left(\max\left\{0;\ \left(\boldsymbol{\mu}^k\right)_i + \alpha_k \left(\mathbf{g}^k\right)_i\right\}\right)_{i=1}^m,
\end{aligned} \tag{4}$$

where k is the iteration counter and $\mathbf{g}^k \in \partial q(\boldsymbol{\mu}^k)$ is an arbitrary subgradient of q at $\boldsymbol{\mu}^k$

A subgradient method for the dual problem, cont'd

- We often write $\mathbf{g}^k = \mathbf{g}(\mathbf{x}^k)$, where $\mathbf{x}^k \in \operatorname{argmin}_{\mathbf{x} \in X} L(\mathbf{x}, \boldsymbol{\mu}^k)$
- Main difference to C^1 case: an arbitrary subgradient $\mathbf{g}^k$ *may be a non-ascent direction!*

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

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

$$\left\| \boldsymbol{\mu}^{k+1} - \boldsymbol{\mu}^* \right\| < \left\| \boldsymbol{\mu}^k - \boldsymbol{\mu}^* \right\| \tag{5}$$

  holds for every step length α_k *in the interval*

$$\alpha_k \in \left(0, \; \frac{2[q^* - q(\boldsymbol{\mu}^k)]}{\|\mathbf{g}^k\|^2} \right) \tag{6}$$

A subgradient method for the dual problem, cont'd

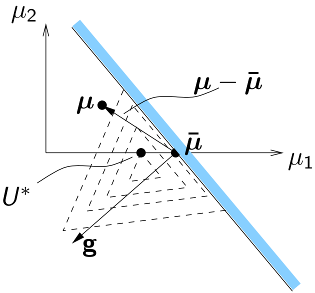
- Why? Let $\mathbf{g} \in \partial q(\bar{\boldsymbol{\mu}})$, and let U^* be the set of optimal solutions to $\max_{\boldsymbol{\mu} \geq \mathbf{0}^m} q(\boldsymbol{\mu})$. Then,

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

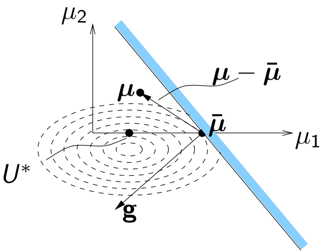
  In other words, $\mathbf{g}$ defines a half-space that contains the set of optimal solutions.

- The good news: If the step length α_k is small enough we get closer to the set of optimal solutions! [i.e., (6) $\to$ (5)]

Each (sub)gradient defines a halfspace that contains the optimal set



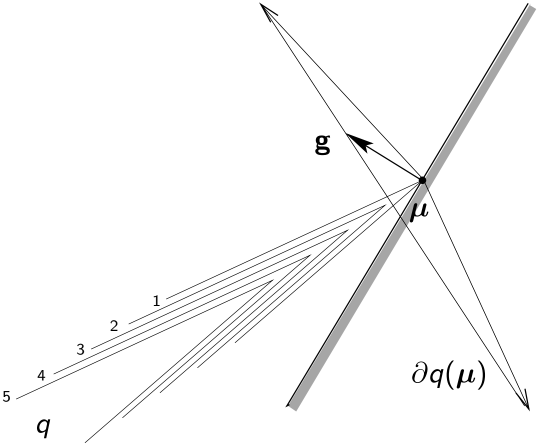
Figur: q non-differentiable



q differentiable

$$\mathbf{g} \in \partial q(\bar{\boldsymbol{\mu}}) \quad \Rightarrow \quad U^* \subseteq \left\{ \boldsymbol{\mu} \in \mathbb{R}^m \;\middle|\; \mathbf{g}^\top (\boldsymbol{\mu} - \bar{\boldsymbol{\mu}}) \geq 0 \right\}$$

Each (sub)gradient defines a halfspace that contains the optimal set



Figur: The half-space defined by a subgradient $\mathbf{g} \in q(\boldsymbol{\mu})$. Note that this subgradient is *not an ascent direction*

Polyak's step length rule

- Choose the step length α_k such that

$$\sigma \leq \alpha_k \leq \frac{2\left[q^* - q(\boldsymbol{\mu}^k)\right]}{\left\|\mathbf{g}^k\right\|^2} - \sigma, \qquad k = 1, 2, \ldots \tag{7}$$

- $\sigma > 0 \Rightarrow$ step lengths α_k don't converge to 0 or to a too large value

- Bad news: Utilizes knowledge of the optimal value q^*!

- But: q^* can be replaced by and approximation $\bar{q}_k \geq q^*$

The divergent series step length rule

- Choose the step length α_k such that

$$\alpha_k > 0, \; k = 1, 2, \ldots ; \qquad \lim_{k \to \infty} \alpha_k = 0; \quad \sum_{s=1}^{\infty} \alpha_s = +\infty \tag{8}$$

- Additional condition, added to ensure convergence to a *point*:

$$\sum_{s=1}^{\infty} \alpha_s^2 < +\infty \tag{9}$$

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*

$$f^* = \inf \left\{ f(\mathbf{x}) \mid \mathbf{x} \in X, \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \right\} \tag{10}$$

(a) *Let $\{\boldsymbol{\mu}^k\}$ be generated by the method on p. 15, under the Polyak step length rule (7), where $\sigma > 0$ is small.*
Then, $\{\boldsymbol{\mu}^k\} \to \boldsymbol{\mu}^* \in U^*$

(b) *Let $\{\boldsymbol{\mu}^k\}$ be generated by the method on p. 15, under the divergent series step length rule (8).*
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. 15, under the divergent series step length rule (8) and (9).*
Then, $\{\boldsymbol{\mu}^k\} \to \boldsymbol{\mu}^* \in U^*$

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 subproblem be $\mathbf{x}^k = \mathbf{x}(\boldsymbol{\mu}^k)$
4. Calculate a subgradient $\mathbf{g}(\mathbf{x}^k) \in \partial q(\boldsymbol{\mu}^k)$
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 (see previous page)
6. [1]Set any negative components of this vector to 0 $\Rightarrow \boldsymbol{\mu}^{k+1}$
7. Let $k := k + 1$ and repeat from 2.

[1]Euclidean projection onto $\mathbb{R}^m_+$

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
- Recovery of primal solutions by utilizing weighted averages of subproblem solutions