

Lectures 4: Algorithms for the Lagrangian dual problem

Ann-Brith Strömberg

2009-09-11

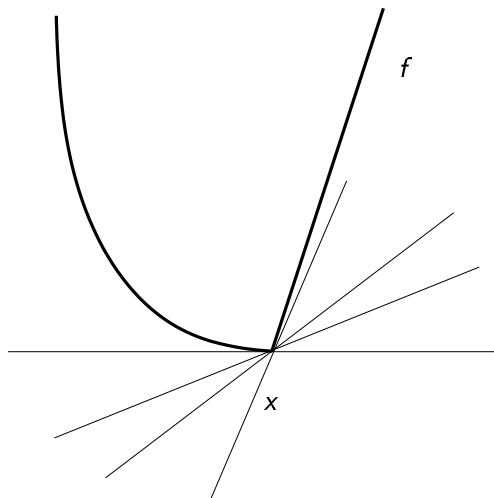
Subgradients of convex functions

- ▶ 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

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \mathbf{p}^T(\mathbf{y} - \mathbf{x}), \quad \mathbf{y} \in \mathbb{R}^n \quad (1)$$

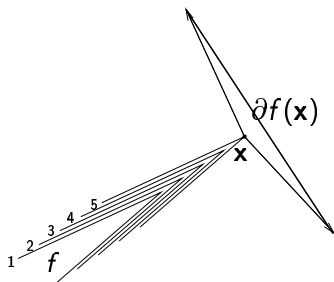
- ▶ The set of such vectors \mathbf{p} defines the *subdifferential* of f at \mathbf{x} , and is denoted $\partial f(\mathbf{x})$
- ▶ $\partial f(\mathbf{x})$ is the collection of “slopes” of the function f at \mathbf{x}
- ▶ For every $\mathbf{x} \in \mathbb{R}^n$, $\partial f(\mathbf{x})$ is a non-empty,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

Subgradients of convex functions, II



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

Subgradients of convex functions, III



Figur: 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

$$f^* = \underset{\mathbf{x}}{\text{infimum}} f(\mathbf{x}), \quad (2a)$$

$$\text{subject to } \mathbf{x} \in X, \quad (2b)$$

$$g_i(\mathbf{x}) \leq 0, \quad i = 1, \dots, m, \quad (2c)$$

and assume that

$$f, g_i (\forall i) \text{ continuous; } X \text{ nonempty and compact} \quad (3)$$

- ▶ 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 $\boldsymbol{\mu} \in \mathbb{R}^m$

Subgradients and gradients of q

- ▶ Suppose that (3) holds ($f, g_i, \forall i$ continuous; X nonempty and compact) in the problem (2):
 $f^* = \inf_{\mathbf{x}} \{f(\mathbf{x}) \mid \mathbf{x} \in X, g_i(\mathbf{x}) \leq 0, i = 1, \dots, m\}$
- ▶ 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

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

Subgradients and gradients of q , cont'd

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

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

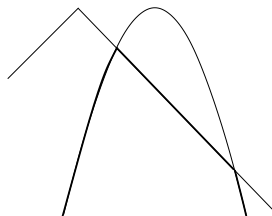
- ▶ The function $f(\mathbf{x}) + \mathbf{p}^T(\mathbf{y} - \mathbf{x})$ is linear w.r.t. \mathbf{y} and *underestimates* $f(\mathbf{y})$ over \mathbb{R}^n
- ▶ Here, we have a *concave* function q and the opposite inequality: $\mathbf{g}(\mathbf{x})$ is a subgradient (actually, supgradient) of q at $\boldsymbol{\mu}$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

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

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

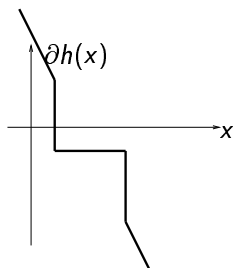
Example

- ▶ 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, $h(x) = \begin{cases} 4 - x, & 1 \leq x \leq 4, \\ 4 - (x - 2)^2, & x \leq 1, x \geq 4 \end{cases}$



Example, cont'd

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

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

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

- ▶ Holds in particular if the Lagrangian subproblem has a unique solution \Leftrightarrow The solution set $X(\mu)$ 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} :

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

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

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

Optimality conditions for the dual problem

- ▶ For a differentiable, concave function h it holds that

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

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

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

- ▶ *Proof.*

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

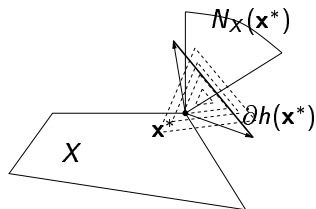
Suppose that $\mathbf{x}^* \in \arg \max_{\mathbf{x} \in \mathbb{R}^n} h(\mathbf{x}) \implies$
 $h(\mathbf{x}) \leq h(\mathbf{x}^*) = h(\mathbf{x}^*) + (\mathbf{0}^n)^T(\mathbf{x} - \mathbf{x}^*)$ for all $\mathbf{x} \in \mathbb{R}^n$, that is,
 $\mathbf{0}^n \in \partial h(\mathbf{x}^*)$ □

Optimality conditions for the dual problem, cont'd

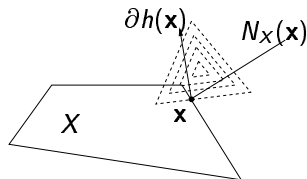
- ▶ The example: $0 \in \partial h(1) \implies x^* = 1$
- ▶ For optimization with constraints the KKT conditions are generalized:

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

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 $\mu \geq \mathbf{0}^m$
- ▶ Consider the dual problem

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

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

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

- ▶ 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

$$\begin{aligned}\boldsymbol{\mu}^{k+1} &= \text{Proj}_{\mathbb{R}_+^m}[\boldsymbol{\mu}^k + \alpha_k \mathbf{g}^k] \\ &= [\boldsymbol{\mu}^k + \alpha_k \mathbf{g}^k]_+ \\ &= (\text{maximum}\{0, (\boldsymbol{\mu}^k)_i + \alpha_k (\mathbf{g}^k)_i\})_{i=1}^m,\end{aligned}\tag{4}$$

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

A subgradient method for the dual problem, cont'd

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

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

A subgradient method for the dual problem, cont'd

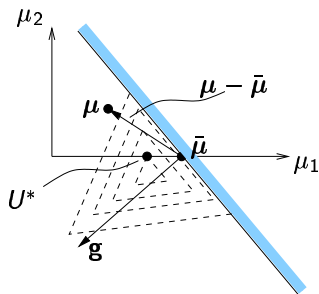
- ▶ Why? Let $\mathbf{g} \in \partial q(\bar{\boldsymbol{\mu}})$, and let U^* be the set of optimal solutions to $\max_{\boldsymbol{\mu} \geq \mathbf{0}^m} q(\boldsymbol{\mu})$. Then,

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

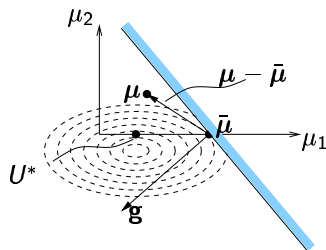
In other words, \mathbf{g} defines a half-space that contains the set of optimal solutions.

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

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



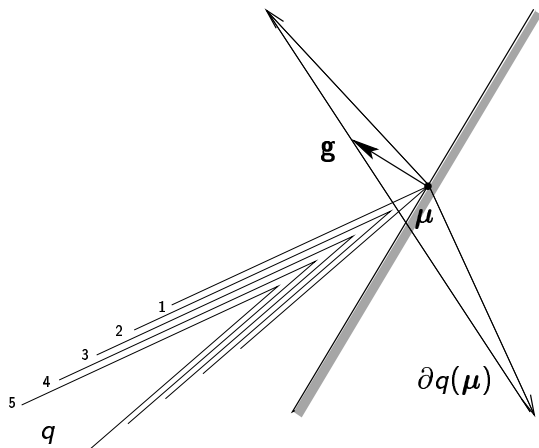
Figur: q non-differentiable



q differentiable

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

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



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

Polyak's step length rule

- ▶ Choose the step length α_k such that

$$\sigma \leq \alpha_k \leq 2[q^* - q(\boldsymbol{\mu}^k)]/\|\mathbf{g}^k\|^2 - \sigma, \quad k = 1, 2, \dots \quad (5)$$

- ▶ $\sigma > 0 \Rightarrow$ 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 an approximation $\bar{q}_k \geq q^*$

The divergent series step length rule

- ▶ Choose the step length α_k such that

$$\alpha_k > 0, k = 1, 2, \dots; \quad \lim_{k \rightarrow \infty} \alpha_k = 0; \quad \sum_{s=1}^{\infty} \alpha_s = +\infty \quad (6)$$

- ▶ Additional condition often added:

$$\sum_{s=1}^{\infty} \alpha_s^2 < +\infty \quad (7)$$

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

$$f^* = \inf\{f(\mathbf{x}) \mid \mathbf{x} \in X, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\} \quad (8)$$

- (a) Let $\{\boldsymbol{\mu}^k\}$ be generated by the method on p. 15, under the Polyak step length rule (5), where $\sigma > 0$ is small.
Then, $\{\boldsymbol{\mu}^k\} \rightarrow \boldsymbol{\mu}^* \in U^*$
- (b) Let $\{\boldsymbol{\mu}^k\}$ be generated by the method on p. 15, under the divergent series step length rule (6).
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 on p. 15, under the divergent series step length rule (6), (7).
Then, $\{\boldsymbol{\mu}^k\} \rightarrow \boldsymbol{\mu}^* \in U^*$

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 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 μ^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*)

- ▶ 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)