Convex Optimization for Computer Vision
Lecture: M. Möller and T. Möllenhoff
Exercises: E. Laude
Summer Semester 2016

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 2

Room: 02.09.023
Friday, 29.04.2016, 09:00-11:00
Submission deadline: Wednesday, 27.04.2016, 14:00, Room 02.09.023

Theory: The Subdifferential, optimality conditions and gradient descent (8+8 Points)

Exercise 1 (4 Points)**.** Let the convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be differentiable at $u \in \text{int}(\text{dom}(f))$. Show that

$$\partial f(u) = \{\nabla f(u)\}.$$

Hint: Use the definition of the subdifferential and the directional derivative. For f being differentiable at the interior of its domain, some direction $v \in \mathbb{R}^n$ and some point $x \in \text{int}(\text{dom}(f))$ the directional derivative $\partial_v f$ of f is given as

$$\partial_v f(x) := \lim_{\epsilon \to 0} \frac{f(x + \epsilon v) - f(x)}{\epsilon} = \lim_{\epsilon \to 0} \frac{f(x) - f(x - \epsilon v)}{\epsilon} = \langle \nabla f(x), v \rangle.$$

Solution. Recall that the subdifferential $\partial f(x)$ of some convex f at $x \in \text{dom}(f)$ is given as

$$\{p \in \mathbb{R}^n : f(y) \geq f(x) + \langle p, y - x \rangle, \, \forall \, y \in \text{dom}(f)\}.$$

Since $u \in \text{int}(\text{dom}(f))$, we find that for all $v \in \mathbb{R}^n$, $u + \epsilon v \in \text{dom}(f)$ for ϵ small enough since the interior of a set is open. By the definition of the subdifferential and setting $x := u$ and $y := u + \epsilon v$ or $y := u - \epsilon v$ we have that if $p \in \partial f(u)$ then

$$f(u + \epsilon v) \geq f(u) + \epsilon \langle p, v \rangle, \quad f(u - \epsilon v) \geq f(u) - \epsilon \langle p, v \rangle,$$

for all $v \in \mathbb{R}^n$ and ϵ small enough. This implies that

$$\lim_{\epsilon \to 0} \frac{f(u + \epsilon v) - f(u)}{\epsilon} \geq \langle p, v \rangle, \quad \lim_{\epsilon \to 0} \frac{f(u) - f(u - \epsilon v)}{\epsilon} \leq \langle p, v \rangle,$$

which means (using the hint)

$$\langle \nabla f(u), v \rangle \geq \langle p, v \rangle, \quad \langle \nabla f(u), v \rangle \leq \langle p, v \rangle$$

or

$$\langle \nabla f(u) - p, v \rangle \geq 0, \quad \langle \nabla f(u) - p, v \rangle \leq 0$$

for all $v \in \mathbb{R}^n$. For the particular choice of $v := \nabla f(u) - p$ we have that

$$\langle \nabla f(u) - p, \nabla f(u) - p \rangle = \|\nabla f(u) - p\|_2^2 = 0$$

which means $p = \nabla f(u)$. Clearly, $\partial f(u)$ is non-empty (and bounded) since $u \in \mathrm{int}(\mathrm{dom}(x))$ implies $u \in \mathrm{ri}(\mathrm{dom}(x))$ (see Thm. Subdifferentiability). Together this concludes the proof.

Definition (Karush-Kuhn-Tucker KKT conditions)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable. A point $x \in \mathbb{R}^n$ satisfies the KKT-conditions if there exists a Lagrange multiplier $\lambda \in \mathbb{R}^m$ s.t.

- $0 = \nabla f(x) + \sum_{i=1}^m \lambda_i \nabla g_i(x)$
- $\lambda_i \geq 0$, $g_i(x) \leq 0$, $\lambda_i g_i(x) = 0$ for $1 \leq i \leq m$

Definition (Guignard Constraint Qualification GCQ)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable and convex. Let

$$X := \{x \in \mathbb{R}^n : g_i(x) \leq 0,\, 1 \leq i \leq m\}$$

denote the feasible set and $x \in X$. Then the condition

$$\begin{aligned} N_X(x) &:= \{v \in \mathbb{R}^n : \langle v, y - x \rangle \leq 0,\, \forall\, y \in X\} \\ &= \left\{ \sum_{i \in \mathcal{A}(x)} \lambda_i \nabla g_i(x) : \lambda_i \geq 0,\, i \in \mathcal{A}(x) \right\}, \end{aligned}$$

is called GCQ. $N_X(x)$ is called the normal cone of the set X at the point $x \in X$ and $\mathcal{A}(x)$ is the set of active constraints at the point x:

$$\mathcal{A}(x) := \{i : 1 \leq i \leq m,\, g_i(x) = 0\}.$$

Definition (Slater's condition)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable and convex. Let $X := \{x \in \mathbb{R}^n : g_i(x) \leq 0,\, 1 \leq i \leq m\}$ denote the feasible set. The condition

$$\exists\, x \in X \text{ s.t. } g_i(x) < 0,\, \forall\, 1 \leq i \leq m$$

is called Slater's condition.

Proposition. Slater's condition is a constraint qualification CQ, i.e. it implies GCQ.

Exercise 2 (8 Points)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable and convex and let $X := \{x \in \mathbb{R}^n : g_i(x) \leq 0,\, 1 \leq i \leq m\}$ denote the feasible set. Let Slater's condition be satisfied. Show that X is convex and then prove the equivalence of the following statements:

- x solves

$$\min_{x\in\mathbb{R}^n} f(x) + \iota_X(x), \qquad \iota_X(x) := \begin{cases} 0 & \text{if } x \in X \\ \infty & \text{otherwise} \end{cases}. \tag{1}$$

- $-\nabla f(x)$ is an element of the normal cone $N_X(x)$ of X at x.
- x satisfies the KKT-conditions.

Hint: Use the proposition stated above. Explain why Slater's condition allows you to apply the sum rule for the subdifferential.

Solution. Starting with the definition of convexity it is straight forward to show that $X = \text{dom}(\iota_X)$ is a convex set. Since further f is a convex function, x solves problem (1) iff the optimality condition

$$0 \overset{!}{\in} \partial(f(x) + \iota_X(x))$$

holds. Slater's condition means that $\text{ri}(\text{dom}(\iota_X)) = \text{ri}(X) \neq \emptyset$ and therefore

$$\text{ri}(\text{dom}(f)) \cap \text{ri}(\text{dom}(\iota_X)) \neq \emptyset.$$

We can therefore apply the sum rule of the subdifferential s.t.

$$0 \in \partial f(x) + \partial \iota_X(x)$$

and using the result of Ex. 1 we can equivalently rewrite the optimality condition as

$$-\nabla f(x) \in \partial \iota_X(x).$$

We proceed showing that $\partial\iota_X(x)$ is the normal cone $N_X(x)$ of X at x. Since $x \in X$ (otherwise $-\nabla f(x) \in \partial\iota_X(x) = \emptyset$) we have

$$\begin{aligned} \partial\iota_X(x) &= \{d \in \mathbb{R}^n : \iota_X(y) - \iota_X(x) \geq \langle d, y - x\rangle, \forall\, y \in \mathbb{R}^n\} \\ &= \{d \in \mathbb{R}^n : \iota_X(y) \geq \langle d, y - x\rangle, \forall\, y \in \mathbb{R}^n\} \end{aligned}$$

Since for $y \notin X$ the inequality $\iota_X(y) \geq \langle d, y - x\rangle$ is trivially satisfied we can rewrite $\partial\iota_X(x)$ as

$$\partial\iota_X(x) = \{d \in \mathbb{R}^n : 0 \geq \langle d, y - x\rangle, \forall\, y \in X\} = N_X(x).$$

It remains to show the equivalence of the last two bullet points. Since the above proposition states that Slater's condition is a CQ one can express the normal cone $N_X(x)$ as

$$N_X(x) = \left\{ \sum_{i\in\mathcal{A}(x)} \lambda_i \nabla g_i(x) : \lambda_i \geq 0,\, i \in \mathcal{A}(x) \right\}.$$

Let $-\nabla f(x) \in N_X(x)$. That means there exist $\lambda_i \geq 0$, with $i \in \mathcal{A}(x)$ s.t.

$$-\nabla f(x) = \sum_{i\in\mathcal{A}(x)} \lambda_i \nabla g_i(x).$$

Defining the missing entries in $\lambda \in \mathbb{R}^m$ s.t. $\lambda_j = 0$ for all $j \notin \mathcal{A}(x)$ and bringing $\nabla f(x)$ to the other side yields:

$$0 = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x).$$

Further we have $\lambda_i \geq 0$ and $x \in X$ i.e. $g_i(x) \leq 0$. For all $i \in \mathcal{A}(x)$ we have $g_i(x) = 0$ and therefore $\lambda_i g_i(x) = 0$ and for all $j \notin \mathcal{A}(x)$ we have $\lambda_j = 0$ and therefore $\lambda_j g_j(x) = 0$. Let conversely x satisfy the KKT conditions. That means there exists $\lambda \in \mathbb{R}^m$ with $\lambda_i \geq 0$ s.t.

$$-\nabla f(x) = \sum_{i=1}^{m} \lambda_i \nabla g_i(x)$$

and for all $i \notin \mathcal{A}(x)$ we have $g_i(x) < 0$ and therefore (since the product $g_i(x)\lambda_i = 0$) we find $\lambda_i = 0$ so that we can rewrite

$$\sum_{i=1}^{m} \lambda_i \nabla g_i(x) = \sum_{i \in \mathcal{A}(x)} \lambda_i \nabla g_i(x).$$

This concludes the proof.

Exercise 3 (4 points)**.** Let the function $E : \mathbb{R}^n \to \mathbb{R}$ be given as

$$E(u) := t(u) + h(u).$$

where the function $h : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$h(u) := g(Du), \qquad g(v) = \sum_{i=1}^{2n} \varphi(v_i), \qquad \varphi(x) = \sqrt{x^2 + \epsilon^2},$$

with $D \in \mathbb{R}^{2n \times n}$ being a finite difference gradient operator and $t : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$t(u) := \frac{\lambda}{2}\|u - f\|^2.$$

1. Show that the function E is L-smooth with $L = \lambda + \frac{\|D\|^2}{\epsilon}$.
2. Show that the function E is m-strongly convex, with $m = \lambda$.

Solution. To compute the (smallest) Lipschitz constant of ∇E we separately compute the (smallest) Lipschitz constants of both $\nabla t(u)$ and $\nabla h(u)$: We first show that h is $\frac{\|D\|^2}{\epsilon}$-smooth and begin computing the gradient of the function h using the chain rule and the quotient rule for φ:

$$\nabla h(u) = D^\top \nabla g(Du), \quad \nabla g(v) = (\varphi'(v_i))_{i=1}^{2n}, \quad \varphi'(x) = \frac{x}{\sqrt{x^2 + \epsilon^2}}.$$

Starting with the left-hand side of the definition we have:

$$\begin{aligned}\|\nabla h(u) - \nabla h(v)\| &= \left\|D^\top \nabla g(Du) - D^\top \nabla g(Dv)\right\| \\ &\leq \|D\| \cdot \|\nabla g(Du) - \nabla g(Dv)\| \\ &= \|D\| \cdot \sqrt{\sum_{i=1}^{2n} \left(\varphi'((Du)_i) - \varphi'((Dv)_i)\right)^2}.\end{aligned}$$

We will show that φ is $\frac{1}{\epsilon}$-smooth, so that

$$\begin{aligned}\|D\| \cdot \sqrt{\sum_{i=1}^{2n} (\varphi'((Du)_i) - \varphi'((Dv)_i))^2} &\leq \|D\| \cdot \sqrt{\sum_{i=1}^{2n} \left(\frac{1}{\epsilon}((Du)_i - (Dv)_i)\right)^2} \\ &= \frac{\|D\|}{\epsilon} \cdot \sqrt{\sum_{i=1}^{2n} ((Du)_i - (Dv)_i)^2} \\ &= \frac{\|D\|}{\epsilon} \cdot \|Du - Dv\| \\ &\leq \frac{\|D\|^2}{\epsilon} \cdot \|u - v\|\end{aligned}$$

This means that h is $\frac{\|D\|^2}{\epsilon}$-smooth. It remains to show that φ is $\frac{1}{\epsilon}$-smooth. We do that by giving an upper bound on the absolute value of the second order derivative φ'' of φ: Using the quotient rule we obtian:

$$|\varphi''(x)| = \varphi''(x) = \frac{1 \cdot \sqrt{x^2 + \epsilon^2} - x \cdot \frac{1}{2} \frac{1}{\sqrt{x^2+\epsilon^2}} \cdot 2x}{x^2 + \epsilon^2} = \frac{\frac{x^2+\epsilon^2-x^2}{\sqrt{x^2+\epsilon^2}}}{x^2 + \epsilon^2} = \frac{\epsilon^2}{(x^2 + \epsilon^2)^{\frac{3}{2}}}$$

Clearly the maximum of φ'' is attained for $x = 0$ s.t.

$$\varphi''(x) \leq \frac{1}{\epsilon}.$$

The data term $t(u)$ is λ-smooth since the Hessian of

$$\frac{\lambda}{2}\|u\|^2 - \frac{\lambda}{2}\|u - f\|^2$$

is $\mathbf{0}$ which clearly is negative semidefinite. Overall we obtain using the triangle inequality:

$$\begin{aligned}\|\nabla E(u) - \nabla E(v)\| &= \|\nabla(t + h)(u) - \nabla(t + h)(v)\| \\ &= \|\nabla t(u) + \nabla h(u) - \nabla t(v) - \nabla h(v)\| \\ &\leq \|\nabla t(u) - \nabla t(v)\| + \|\nabla h(u) - \nabla h(v)\| \\ &\leq \lambda\|u - v\| + \frac{\|D\|^2}{\epsilon}\|u - v\| = \left(\lambda + \frac{\|D\|^2}{\epsilon}\right)\|u - v\|.\end{aligned}$$

This concludes the proof of the first part of this exercise.

For the second part we first show that the data term $t(u) = \frac{\lambda}{2}\|u - f\|^2$ is λ-strongly convex since the Hessian of

$$\frac{\lambda}{2}\|u - f\|^2 - \frac{\lambda}{2}\|u\|^2$$

is $\mathbf{0}$ which clearly is positive semidefinite. Since $h(u)$ is also convex (this follows from a straight forward computation starting with the definition of a convex function) and, according to the lecture, the sum of two convex functions is convex we have that

$$\frac{\lambda}{2}\|u - f\|^2 - \frac{\lambda}{2}\|u\|^2 + h(u)$$

is also convex and therefore the energy $E(u)$ is λ-strongly convex.

Programming: Image denoising (12 Points)

Exercise 4 (12 Points)**.** Denoise the noisy input image f, given in the file `noisy_input.png` by minimizing the energy from Ex. 3:

$$E(u) = \frac{\lambda}{2}\left\|u - f\right\|^2 + \sum_{i=1}^{2n}\sqrt{(Du)_i^2 + \epsilon^2}$$

with gradient descent. To guarantee convergence choose your step size τ so that

$$0 < \tau \leq \frac{2}{m + L}.$$

Use MATLABs `normest` to estimate the norm $\|D\|$ of your finite difference gradient operator D. Here, n is the number of pixels times the number of color channels.