Convex Optimization for Machine Learning and Computer Vision

Lecture: Dr. Virginia Estellers
Exercises: Emanuel Laude
Winter Semester 2017/18

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 3

Room: 02.09.023

Friday, 17.11.2017, 09:15-11:00

Submission deadline: Monday, 13.11.2017, 10:15, Room 02.09.023

Theory: Lipschitz continuity, fixed point iterations and gradient descent (12+4 Points)

Exercise 1 (4 Points)**.** We call a function $E : \mathbb{R}^n \to \mathbb{R}$ *absolutely one-homogeneous* if

$$E(\alpha u) = |\alpha| E(u)$$

holds for all $u \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Prove that

$$\partial E(u) = \{p \in \mathbb{R}^n \mid \langle p, u \rangle = E(u), \quad E(v) \geq \langle p, v \rangle \ \forall v \in \mathbb{R}^n\}.$$

Solution. Let $u \in \mathbb{R}^n$ and let $p \in \partial E(u)$. That means that $E(v) - E(u) \geq \langle p, v - u \rangle$ for all $v \in \mathbb{R}^n$. Let $v = 0$. Then $E(v) = 0$. Therefore

$$E(u) \leq \langle p, u \rangle. \tag{1}$$

Conversely

$$E(u) = E(2u) - E(u) \geq \langle p, u \rangle.$$

This means $E(u) = \langle p, u \rangle$. Then also $E(v) - \langle p, u \rangle \geq \langle p, v - u \rangle$ and therefore $E(v) \geq \langle p, v - u \rangle$.

Exercise 2 (4 Points)**.** Find examples for the following functions and explain why your example is correct:

- A continuously differentiable convex function that is not L-smooth.
- A Lipschitz continuous function that is not a contraction.
- A function that is not differentiable, but Lipschitz continuous.
- A convex L-smooth function E and a step size τ for which G defined by $G(u) = u - \tau \nabla E$ is not a non-expansive function.

Solution.

- $|x|^{3/2}$ or x^4.
- $5x$

- $|x|$.
- $E(u) = 1/2u^2$, $\tau = 3$, $G(u) = 2u$

Exercise 3 (4 Points)**.** Show that for any $a, b \in \mathbb{R}^n$, $\theta \in \mathbb{R}$ it holds that

$$\|(1-\theta)a + \theta b\|^2 = (1-\theta)\|a\|^2 + \theta\|b\|^2 - \theta(1-\theta)\|a-b\|^2$$

Solution.

$$\begin{aligned}
\|(1-\theta)a + \theta b\|^2 &= (1-\theta)^2\|a\|^2 + \langle(1-\theta)a, \theta b\rangle + \theta^2\|b\|^2 \\
&= (1-2\theta+\theta^2)\|a\|^2 + (\theta-\theta^2)\langle a,b\rangle + \theta^2\|b\|^2 \\
&= \theta^2\|a\|^2 + (1-\theta)\|a\|^2 - \theta\|a\|^2 - \theta^2\langle a,b\rangle + \theta\langle a,b\rangle + \theta^2\|b\|^2 \\
&= \theta^2\|a-b\|^2 + (1-\theta)\|a\|^2 - \theta\|a\|^2 + \theta\langle a,b\rangle - \theta\|b\|^2 + \theta\|b\|^2 \\
&= (\theta^2-\theta)\|a-b\|^2 + (1-\theta)\|a\|^2 + \theta\|b\|^2 \\
&= (1-\theta)\|a\|^2 + \theta\|b\|^2 - (1-\theta)\theta\|a-b\|^2
\end{aligned}$$

Exercise 4 (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

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

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.

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

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:

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

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 5 (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.