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 4

Room 01.09.014
Friday, 13.5.2016, 09:00-11:00
Submission deadline: Wednesday, 11.5.2016, 14:00, Room: 02.09.023

Theory: Projected gradient descent (8+8 Points)

Exercise 1 (4 Points)**.** Let $A \in \mathbb{R}^{n \times n}$ be orthonormal, meaning that $A^\top A = AA^\top = I$. Let the convex set C be given as

$$C := \{u \in \mathbb{R}^n : \|Au\|_\infty \leq 1\}.$$

Compute a formula for the projection onto C given as

$$\Pi_C(v) := \operatorname{argmin}_{u \in \mathbb{R}^n} \frac{1}{2}\|u - v\|_2^2, \quad \text{s.t. } u \in C.$$

Hint: Show that the ℓ_2-norm of a vector is invariant under a multiplication with an orthonormal matrix A, meaning that $\|u\|_2 = \|Au\|_2$.

Solution. We begin proving the hint:

$$\|Ax\|_2^2 = \langle Ax, Ax \rangle = \langle A^\top Ax, x \rangle = \langle x, x \rangle = \|x\|_2^2.$$

The idea is to rewrite the projection onto the set C in terms of the projection $\Pi_{\tilde{C}}$ onto the unit ball of the ℓ_∞-norm $\tilde{C} := \{x \in \mathbb{R}^n : \|x\|_\infty \leq 1\}$. With the substitution

$$w := Au \iff u = A^\top w$$

and using the hint we obtain:

$$\begin{aligned}
\Pi_C(v) &= \operatorname{argmin}_{\|Au\|_\infty \leq 1} \frac{1}{2}\|v - u\|^2 \\
&= A^\top \operatorname{argmin}_{\|w\|_\infty \leq 1} \frac{1}{2}\|v - A^\top w\|^2 \\
&= A^\top \operatorname{argmin}_{\|w\|_\infty \leq 1} \frac{1}{2}\|A(v - A^\top w)\|^2 \\
&= A^\top \operatorname{argmin}_{\|w\|_\infty \leq 1} \frac{1}{2}\|Av - AA^\top w\|^2 \\
&= A^\top \operatorname{argmin}_{\|w\|_\infty \leq 1} \frac{1}{2}\|Av - w\|^2 \\
&= A^\top \Pi_{\tilde{C}}(Av).
\end{aligned}$$

Exercise 2 (8 Points)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be coercive, differentiable and let ∇f be locally Lipschitz continuous i.e. for each $x \in \mathbb{R}^n$ there exists $\epsilon > 0$ and a constant $L_\epsilon > 0$, such that ∇f is L_ϵ-Lipschitz on $B_\epsilon(x)$.

1. Give an example of a function $f : \mathbb{R}^n \to \mathbb{R}$ that meets the above assumptions, but for which ∇f is not Lipschitz continuous. (With a proof).

2. Let f be an arbitrary function meeting the above assumptions. Show that for any $\alpha \geq \min_x f(x)$ there exists a constant L_α such that f is L_α-smooth on the sublevel set $S_\alpha := \{x : f(x) \leq \alpha\}$.
   Hint: You can use the topological definition of compactness: Any subset $X \subset \mathbb{R}^n$ is called compact (closed and bounded in the Euclidean case that we consider) if each of its open covers has a finite subcover (see `https://en.wikipedia.org/wiki/Compact_space`). One possible open cover of X is the union of all ϵ-balls with any ϵ: $\bigcup_{x \in X} B_\epsilon(x)$.

3. Show that $x^+ := x - \tau \nabla f(x) \in S_\alpha$ if $\tau < \frac{1}{L_\alpha}$ and $x \in S_\alpha$.

4. Conclude that for each initialization x_0, there exists a τ_{x_0} such that gradient descent with the constant step size τ_{x_0} converges.

Solution.

1. A possible choice is the function $f(x) := x^4$. The first and second order derivatives of f are the given as

$$f'(x) = 4x^3, \quad f''(x) = 12x^2.$$

   Then for any $x \in \mathbb{R}$ and any $\epsilon > 0$ we have

$$\sup_{y \in B_\epsilon(x)} 12x^2 = 12(|x| + \epsilon)^2 =: L_\epsilon.$$

   Thus $f''(y) \leq L_\epsilon$ for $y \in B_\epsilon(x)$ which means that f is L_ϵ-smooth on $B_\epsilon(x)$.

2. S_α is non-empty since $\alpha \geq \min_x f(x)$. Moreover, $S_\alpha = f^{-1}((-\infty, \alpha])$ is closed since it is the preimage of the closed set $(-\infty, \alpha]$ under the continuous function f. A characterization of continuity of a function $f : X \to Y$ is that preimages of open sets are open under f. Using this characterization one can show that the same holds for closed sets: If $D \subseteq Y$ is closed we have that $Y \setminus D$ is open and since f continuous

$$f^{-1}(Y \setminus D) = X \setminus f^{-1}(D)$$

   is open. And therefore $X \setminus (X \setminus f^{-1}(D)) = f^{-1}(D)$ is closed. Since f is coercive S_α is bounded. Otherwise there would exist a sequence $\{x_n\}_{n \in \mathbb{N}} \subset S_\alpha$ with $\|x_n\| \to \infty$ for $n \to \infty$ and then, since f coercive and continuous, $f(x_n) \to \infty$ for $n \to \infty$. So overall S_α is compact and therefore the closure of the convex hull $\mathrm{cl}(\mathrm{conv}(S_\alpha))$ of S_α is compact too. Now, for any $x, y \in S_\alpha$ consider the closed line $M := [x, y] := \mathrm{conv}(\{x, y\}) \subset \mathrm{cl}(\mathrm{conv}(S_\alpha))$. Now

$M \subset \bigcup_{x \in M} B_{\epsilon_x}(x)$ (where ϵ_x is chosen s.t. ∇f locally Lipschitz on $B_{\epsilon_x}(x)$ with constant L_{ϵ_x}) has a finite subcover

$$M \subset \bigcup_{i=1}^{N} B_{\epsilon_{x_i}}(x_i).$$

We can now construct a finite sequence $\{y_i\}_{i=1}^{N+1} \subset M$ with points along the line with

$$\begin{aligned} x = y_1 \in B_{\epsilon_{x_1}}(x_1), \quad y = y_{N+1} \in B_{\epsilon_{x_N}}(x_N) \\ y_i \in B_{\epsilon_{x_{i-1}}}(x_{i-1}) \cap B_{\epsilon_{x_i}}(x_i), \quad 2 \leq i \leq N. \end{aligned}$$

Then we have for $L_\alpha := \max_{1 \leq i \leq N} L_{\epsilon_{x_i}}$

$$\begin{aligned} \|\nabla f(x) - \nabla f(y)\|_2 &= \left\| \sum_{i=1}^{N} \nabla f(y_i) - \nabla f(y_{i+1}) \right\|_2 \\ &\leq \sum_{i=1}^{N} \underbrace{\|\nabla f(y_i) - \nabla f(y_{i+1})\|_2}_{y_i, y_{i+1} \in B_{\epsilon_{x_i}}(x_i)} \\ &\leq \sum_{i=1}^{N} L_{\epsilon_{x_i}} \|y_i - y_{i+1}\|_2 \\ &\leq \sum_{i=1}^{N} L_\alpha \|y_i - y_{i+1}\|_2 \\ &= L_\alpha \sum_{i=1}^{N} \|y_i - y_{i+1}\|_2 \\ &= L_\alpha \|x - y\|_2 \end{aligned}$$

Thus ∇f is L_α-Lipschitz on S_α.

3. We define

$$g(t) := -f(x) + f(x - t\nabla f(x))$$

Then, since f is differentiable we obtain

$$g'(t) = -\langle \nabla f(x), \nabla f(x - t\nabla f(x)) \rangle$$

Suppose $\nabla f(x) \neq 0$, otherwise the assertion would be trivially satisfied. Then $g'(0) = -\|\nabla f(x)\|^2 < 0$ which means (due to the continuity of f) there exists $t > 0$, such that $g'(\tau) < 0$ and (due to Taylor's theorem) $g(\tau) < 0$ for all $\tau \in (0, t]$. Suppose there exists a $\tau' \in (t, 1/L_\alpha]$ so that $g(\tau') > 0$. From the intermediate value theorem it follows that there exists a point $\tau'' \in (\tau, \tau')$ with $g(\tau'') = 0$ and from the mean value theorem of f it follows that there exists a

$\xi \in (\tau, \tau'')$ with $g'(\xi) > 0$ and $g(\xi) < 0$. As $g(\xi) = f(x) + f(x - \xi\nabla f(x)) < 0$ we have that $x - \xi\nabla f(x) \in S_\alpha$. Then we get (since ∇f is L_α-Lipschitz)

$$\begin{aligned}
g'(\xi) &= -\langle \nabla f(x), \nabla f(x) - (\nabla f(x) - \nabla f(x - \xi\nabla f(x)) \rangle \\
&= -\|\nabla f(x)\|_2^2 + \langle \nabla f(x), \nabla f(x) - \nabla f(x - \xi\nabla f(x)) \rangle \\
&\leq -\|\nabla f(x)\|_2^2 + \|\nabla f(x)\|_2 \|\nabla f(x) - \nabla f(x - \xi\nabla f(x))\|_2 \\
&\leq -\|\nabla f(x)\|_2^2 + \|\nabla f(x)\|_2 L_\alpha \|x - x + \xi\nabla f(x)\|_2 \\
&= -\|\nabla f(x)\|_2^2 + \|\nabla f(x)\|_2 L_\alpha \xi \|\nabla f(x)\|_2 \\
&= (L_\alpha \xi - 1)\|\nabla f(x)\|_2^2 < 0
\end{aligned}$$

This contradicts the assumption and therefore

$$g(t) \leq 0, \text{ for all } t \in \left[0, \frac{1}{L_\alpha}\right].$$

4. Set $\alpha := f(x_0)$ and

$$\tau_{x_0} \in \left(0, \frac{1}{L_\alpha}\right].$$

Then, for $x_k \in S_\alpha$ we have

$$x_{k+1} := x_k - \tau_{x_0}\nabla f(x_k) \in S_\alpha,$$

and using the result from the previous exercise $\{f(x_k)\}_{k\in\mathbb{N}}$ is monotone decreasing and bounded since S_α compact and f and continuous. Together this yields that $\{f(x_k)\}_{k\in\mathbb{N}}$ converges.

Exercise 3 (4 Points)**.** Let C_i, $1 \leq i \leq n$ be a family of closed convex sets such that

$$\bigcap_{1\leq i\leq n} C_i \neq \emptyset.$$

Show that the problem of finding an element u^* in the intersection

$$u^* \in \bigcap_{1\leq i\leq n} C_i$$

can be formulated as the following optimization problem:

$$u^* \in \arg\min_{u\in\bigcap_{i\in\mathcal{I}} C_i} \sum_{\substack{j\notin\mathcal{I} \\ 1\leq j\leq n}} d^2(u, C_j),$$

where $\mathcal{I} \subseteq \{1, 2, \ldots, n\}$ can be arbitrary (including the empty set) and $d(z, X)$ is the distance of a point z to the closed convex set X defined as

$$d(z, X) := \min_{x\in X} \|x - z\|_2.$$

Solution. Since all C_i are closed and convex, $d(u, C_i)$ is well defined. Since $d^2(u, C_j) \geq 0$ and $d^2(u, C_j) = 0 \iff u \in C_j$,

$$u \in \bigcap_{\substack{j \notin \mathcal{I} \\ 1 \leq j \leq n}} C_j \iff \sum_{\substack{j \notin \mathcal{I} \\ 1 \leq j \leq n}} d^2(u, C_j) = 0.$$

This yields that

$$0 = \sum_{\substack{j \notin \mathcal{I} \\ 1 \leq j \leq n}} d^2(u^*, C_j) + \iota_{\bigcap_{i \in \mathcal{I}} C_i}(u^*)$$

iff $u^* \in \bigcap_{1 \leq i \leq n} C_i$. Since $\bigcap_{1 \leq i \leq n} C_i$ non-empty

$$\operatorname{argmin}_{u \in \bigcap_{i \in \mathcal{I}} C_i} \sum_{\substack{j \notin \mathcal{I} \\ 1 \leq j \leq n}} d^2(u, C_j) \subset \bigcap_{1 \leq i \leq n} C_i.$$

Programming: SUDOKU (12 Points)

Exercise 4 (12 Points)**.** Solve the SUDOKUs given in the files `exampleSudoku1.mat` and `exampleSudoku2.mat` with projected gradient descent. For that you need to find a point

$$u^* \in \bigcap_{1 \leq i \leq n+m+1} C_i$$

where the convex sets in the intersection are given as

$$C_i := \{u \in \mathbb{R}^{729} : \langle a_i, u \rangle = 1\}, \quad 1 \leq i \leq n,$$

$$C_i := \{u \in \mathbb{R}^{729} : u_j = 1,\ j \in \mathcal{B}\}, \quad n+1 \leq i \leq n+m,$$

and $\mathcal{B}$ is the set of indexes corresponding to the known numbers and

$$C_{n+m+1} := \{u \in \mathbb{R}^{729} : u_j \in [0,1],\ \forall\, 1 \leq j \leq 729\}.$$

For a more precise definition of the constraint sets see lecture.
Solve the programming assignment in the spirit of exercise 3 using the following two partitions of the indexes $\{1, 2, \ldots, n+m+1\}$:

1. $\mathcal{I}_1 := \{n+m+1\}$ and
2. $\mathcal{I}_2 := \{n+1, n+2, \ldots, n+m+1\}$,

and plot the resulting energy decays.
Hint: Show that for the linear constraint sets C_i, $1 \leq i \leq n$ the distance $d(z, C_i)$ of a point z to the set C_i is equal to $|\langle a_i, z \rangle - 1|$.