

Exercise sheet 5

Exercise 1 [Convex subdifferential, until 08.11]

Let $g: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ be a convex function. The convex subdifferential $\partial g(x)$ of g in $x \in \mathbb{R}^n$ is defined by

$$\partial g(x) = \{\gamma \in \mathbb{R}^n \mid \langle \gamma, y - x \rangle + g(x) \leq g(y) \quad \forall y \in \mathbb{R}^n\}.$$

a) Prove that $\partial g(x)$ with $g(x) = \|x\|_1$ is given by

$$(\partial g(x))_i = \begin{cases} \{-1\} & x_i < 0, \\ [-1, 1] & x_i = 0, \\ \{1\} & x_i > 0 \end{cases}$$

for $i = 1, \dots, n$. Moreover calculate $\partial g(x)$ for the indicator function of the convex set $C \subset \mathbb{R}^n$ defined by

$$g(x) = \begin{cases} 0 & x \in C, \\ \infty & \text{else.} \end{cases}$$

- b) Prove: $\hat{x} \in \mathbb{R}^n$ solves $\min_{x \in \mathbb{R}^n} g(x)$ iff $0 \in \partial g(\hat{x})$
- c) Let $f: \mathbb{R}^n \rightarrow \mathbb{R}$ be continuously differentiable and convex. Show $\partial f(x) = \{\nabla f(x)\}$.
- d) Prove that $\partial(f(x) + g(x)) = \{\nabla f(x)\} + \partial g(x)$ holds true and conclude that $\hat{x} \in \mathbb{R}^n$ solves $\min_{x \in \mathbb{R}^n} f(x) + g(x)$ iff $-\nabla f(\hat{x}) \in \partial g(\hat{x})$. Hint: Use the definition of directional differentiability and the convexity of g to prove $\partial(f(x) + g(x)) \subset \{\nabla f(x)\} + \partial g(x)$.

Exercise 2 [Proximal mapping, until 08.11]

The proximal mapping $P_L g: \mathbb{R}^n \rightarrow \mathbb{R}^n$ of g is defined by

$$P_L g(y) = \operatorname{argmin}_{x \in \mathbb{R}^n} \left(g(x) + \frac{L}{2} \|x - y\|_2^2 \right).$$

Note that it can be proven that g is continuous.

- a) Assume that $g \neq \infty$ and that g is bounded from below. Note that it can be proven that g is continuous, since it is convex on \mathbb{R}^n and real valued (∞ is excluded). Why is $P_L g$ a mapping?

b) Show that $P_L g$ for $g(x) = \alpha \|x\|_1$ is given by

$$(P_L g(y))_i = \max(0, y_i - \alpha/L) + \min(0, y_i + \alpha/L)$$

for $i = 1, \dots, n$.

c) Let C additionally be closed. What is $P_L g$ for the indicator function of C .

Let additionally $f \in C^{1,1}$ with the Lipschitz constant of the gradient L_f . Then we consider the following proximal gradient method for $\min_{x \in \mathbb{R}^n} f(x) + g(x) = F(x)$:

```

for  $k = 0, 1, \dots$  do
    | Set  $x^{k+1} = P_{L_f g} \left( x^k - \frac{1}{L_f} \nabla f(x^k) \right)$ ;
end

```

Algorithm 1: Proximal gradient method

for $x^0 \in \mathbb{R}^n$.

Exercise 3 [ISTA, until 15.11]

In the following we analyze the convergence properties of ISTA in the non-smooth case.

a) Prove: For $L \geq L_f$ and for any $x, y \in \mathbb{R}^n$ there holds

$$\begin{aligned}
 F(x) - F \left(P_L g \left(y - \frac{1}{L} \nabla f(y) \right) \right) &\geq \frac{L}{2} \left\| P_L g \left(y - \frac{1}{L} \nabla f(y) \right) - y \right\|_2^2 \\
 &\quad + L \langle y - x, P_L g \left(y - \frac{1}{L} \nabla f(y) \right) - y \rangle
 \end{aligned}$$

Hint: Use Lemma 2.26 from the lecture and the optimality condition for the optimization problem in the definition of the proximal mapping.

b) Prove the following statement by using a) and modifying the proof of Theorem 2.29 from the lecture: Let $\{x^k\}$ be the sequence generated by Algorithm (1). Then for any $k \geq 1$

$$F(x^k) - F(\bar{x}) \leq \frac{L_f \|x^0 - \bar{x}\|^2}{2k} \quad \forall \bar{x} \in X^*,$$

where X^* is defined in the lecture.

Next we consider the FISTA algorithm which is given by

```

Set  $y^0 = x^0$  and  $t^1 = 1$ 
for  $k = 1, 2, \dots, \text{maxiter}$  do
    | Set  $x^k = P_{L_f g} \left( y^k - \frac{1}{L_f} \nabla f(y^k) \right)$ ;
    |    $t^{k+1} = \frac{1 + \sqrt{1 + 4(t^k)^2}}{2}$ ;
    |    $y^{k+1} = x^k + \left( \frac{t^k - 1}{t^{k+1}} \right) (x^k - x^{k-1})$ ;
end

```

Algorithm 2: FISTA

for $x^0 \in \mathbb{R}^n$.

Exercise 4 [Programming exercise: ISTA and FISTA]

In this programming exercise we consider the setting of Exercise sheet 2, exercise 5. In particular we consider the problem

$$\min_{x \in \mathbb{R}^n} \|Cu - y_n\|_2^2 + \alpha \|u\|_1 = f(u) + g(u) \quad (1)$$

- a) Implement Algorithm (1) and (2) for (1) in Matlab/Octave.
- b) Plot the iterates x^{maxiter} for Algorithm (1) and (2). Moreover display the convergence behavior of the functional values for both algorithms in a semilog-plot. Describe your results. Hint: Replace \bar{x} by x^{maxiter} . Include a plot of c_1/k and c_2/k^2 with the correct constants c_i .
- c) Which method would you use and why?

Hand in by email (philip.trautmann@uni-graz.at) until 15.11.2019, 23:59 o'clock.