Homework assignment 4

Exercises 1 and 2 are Hand-in exercises.

1. Compute the proximal operator to the following functions f. For each example do the following:

(i) Indicate if the proximal operator is separable.

(ii) Provide a rough estimate (if possible) of the computational cost of evaluating the proximal operator (or point out the most costly operation).

(iii) Consider computing the prox of $f \circ L$ where L is an arbitrary linear operator. Estimate if the computational cost of computing this prox is significantly increased compared to computing the prox of f.

(iv) Decide if the function satisfies the properties needed to guarantee convergence using a forward-step (gradient-step) of its gradient. That is, decide if f differentiable with ∇f cocoercive (or equivalently, decide if f smooth).

(v) If a forward-step in (iv) is OK, compare the computational cost of the forward-step and the prox-step (with and without composition).

a. $f(x) = \|x\|_1$.

b. $f(x) = \|x\|_0$ (counts the number of nonzero elements (nonconvex)).

c. $f(x) = \frac{1}{2}x^T Hx$ where H is full symmetric and positive semi-definite.

d. $f(x) = \|x\|$.

e. $f(x) = \frac{1}{2}\|x\|^2$.

f. $f(x) = \|x\|_1 + \frac{1}{2}\|x\|^2$.

g. $f(x) = \iota_C(x)$ where $C \neq \emptyset$ is a general closed and convex set.

h. $f(x) = \iota_V(x)$ where $V = \{x \mid Lx = b\} \neq \emptyset$.

i. $f(x) = \iota_B(x)$ where $B = \{x \mid l \leq x \leq u\} \neq \emptyset$.

2. Moreau's identity and generalizations.

a. Assume that A is a maximal monotone operator and that $\gamma \in (0,\infty)$. Prove the generalized Moreau identity:

$$J_{\gamma A} + \gamma J_{\gamma^{-1}A^{-1}} \circ (\gamma^{-1}\mathrm{Id}) = \mathrm{Id}.$$

b. Let f be proper closed and convex and L be a linear operator and let $\gamma \in (0,\infty)$. Further assume that $f \circ L$ is proper. Show that if

$$s^* \in \underset{s}{\mathrm{Argmin}}\{g^*(s) + \tfrac{\gamma}{2}\|L^*s - \gamma^{-1}z\|^2\}$$

exists, then

$$\mathrm{prox}_{\gamma(f\circ L)}(z) = z - \gamma L^* s^*$$

3. Assume that $\beta \in (0,1)$ and that T is $\frac{1}{\beta}$-cocoercive. Show that $T+(1-\beta)\mathrm{Id}$ is $\frac{\beta}{2}$-averaged.

4. Suppose that f is proper closed and convex.

a. Assume that f is σ-strongly convex with $\sigma \in (0,\infty)$. Then the reflected resolvent R_f is α-negatively averaged. Provide a value for α.

b. Assume that f is β-smooth with $\beta \in (0,\infty)$. Then the reflected resolvent R_f is α-averaged. Provide a value for α.

5. Consider the problem

$$\text{minimize } \tfrac{1}{2}\|Ax-b\|^2 + \lambda\|Dx\|_1 \tag{P}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $D \in \mathbb{R}^{p\times n}$, $x \in \mathbb{R}^n$, and $\lambda \in (0,\infty)$. Assume that A and D always have full rank. We will consider two different dual problems for (P). The primal problem can be written on the following general form

$$\text{minimize } f(x) + g(Lx).$$

Let $f(x) = \frac{1}{2}\|Ax-b\|^2$ and $g_1(x) = \lambda\|Dx\|_1$ to get the first dual problem

$$\text{minimize } f^*(-\mu) + g_1^*(\mu). \tag{D1}$$

Let $f(x) = \frac{1}{2}\|Ax-b\|^2$, $g_2(y) = \lambda\|y\|_1$, and $L = D$ to get the second dual problem

$$\text{minimize } f^*(-L^*\mu) + g_2^*(\mu). \tag{D2}$$

Consider the following cases:

(i) $m \geq n$ and $D = I$
(ii) $m \geq n$ and $p < n$ and D general structure
(iii) $m \geq n$ and $p \geq n$ and D general structure
(iv) $m < n$ and $D = I$
(v) $m < n$ and $p < n$ and D general structure
(vi) $m < n$ and $p \geq n$ and D general structure

a. For each of these cases, motivate if the problem can be solved by forward-backward splitting applied to (P), (D1), and/or (D2). Also provide bounds on γ in each case and algorithm for which the algorithm is guaranteed to converge.

b. For each of these cases, motivate which forward-backward method that gets cheapest iterations. That is, if forward-backward splitting applied to (P), (D1), or (D2) (if applicable) gets the cheapest iteration cost.

c. For each problem and (feasible) forward-backward splitting method (that is FB applied to (P), (D1), (D2)), motivate if the convergence is linear or sublinear.

d. Implement the algorithms using random A and D (where applicable).