Exercise sheet 3

You can return your solutions to questions 1 and 6 to get them marked. If so, please upload them on Moodle before Monday 14/3 at 12noon.

1. (*Bregman subgradient method*) Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a smooth and strictly convex function, and let D_ϕ be its Bregman divergence. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a potentially nonsmooth convex function, and consider the following Bregman subgradient method:

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathbb{R}^n} \{t_k \langle g_k, x - x_k \rangle + D_\phi(x|x_k)\}$$

where $g_k \in \partial f(x_k)$.
(a) Show that for $\phi(x) = \|x\|_2^2/2$ we recover the usual subgradient method.
(b) Let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^n$. We assume that ϕ is 1-strongly convex with respect to $\|\cdot\|$. Show that the iterates of the Bregman subgradient method satisfy:

$$D_\phi(x^*|x_{k+1}) \le D_\phi(x^*|x_k) + \frac{1}{2}\|t_k g_k\|_*^2 + t_k(f(x^*) - f(x_k))$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$. Deduce:

$$f_{\text{best},k} - f^* \le \frac{D_\phi(x^*\|x_0)}{\sum_{i=0}^k t_i} + \frac{\sum_{i=0}^k t_i^2 \|g_i\|_*^2}{\sum_{i=0}^k t_i}.$$

where $f_{\text{best},k} = \min\{f(x_0), \dots, f(x_k)\}$.

2. (*Conjugate functions*) Let f be convex and lower semicontinuous. Define the conjugate of f by

$$f^*(\xi) = \sup_{x \in \mathbf{dom}(f)} \{\langle \xi, x \rangle - f(x)\}.$$

Show that

- Biduality: $f^{**} = f$
- $f(x) + f^*(\xi) = \langle \xi, x \rangle \iff x \in \partial f^*(\xi) \iff \xi \in \partial f(x)$
- Moreau's identity: $\mathbf{prox}_{f^*}(x) = x - \mathbf{prox}_f(x)$
- If f is m-strongly convex with respect to $\|\cdot\|$, then $\mathbf{dom}(f^*) = \mathbb{R}^n$ and f^* is smooth with

$$\nabla(f^*)(\xi) = \operatorname*{argmax}_{x \in \mathbf{dom}(f)} \{\langle \xi, x \rangle - f(x)\}.$$

Moreover, $\nabla(f^*)$ is $(1/m)$-Lipschitz with respect to $\|\cdot\|$, i.e., $\|\nabla f^*(\xi_1) - \nabla f^*(\xi_2)\| \le \|\xi_1 - \xi_2\|_*$, where $\|\cdot\|_*$ is the dual norm.

3. (*Mirror descent*) Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex, and let $\phi : \mathbb{R}^n \to \mathbb{R}$ be smooth strongly convex. Consider the following iterative method to minimize $f(x)$:

$$x_{k+1} = \nabla\phi^*(\nabla\phi(x_k) - t_k g_k) \tag{1}$$

where $g_k \in \partial f(x_k)$ and where ϕ^* denotes the conjugate function of ϕ. Show that (1) is equivalent to the Bregman subgradient method considered in the first question.

4. (*Smoothing via conjugate functions*)

(a) Assume f is a convex function given as $f(x) = h^*(Ax + b)$ where h is convex lower-semicontinuous, defined on a *compact domain* D, i.e.,

$$f(x) = \max_{y \in D} \left\{ y^T(Ax + b) - h(y) \right\}.$$

Let d be convex function defined on D which is 1-strongly convex with respect to the Euclidean norm, and consider for $\mu > 0$ the function

$$f_\mu(x) = (h + \mu d)^*(Ax + b).$$

Show that f_μ is smooth, with smoothness parameter (with respect to Euclidean norm) $L = \|A\|^2/\mu$ where $\|A\|$ is the operator norm of A. Further, show that

$$f - \mu R \leq f_\mu \leq f$$

where $R = \max_{d \in D} d(x)$.

(b) Examples: (i) let $f(x) = \|Ax + b\|_1$ which we can write as $f(x) = h^*(Ax + b)$ where h is the indicator function of the unit ℓ_∞ ball. Compute $f_\mu(x)$ explicitly for $d(y) = \|y\|_2^2/2$, and for $d(y) = \sum_i 1 - \sqrt{1 - y_i^2}$ (check that both functions are 1-strongly convex).

5. (*Newton's method*) Let $f : \mathbb{R}^n \to \mathbb{R}$ be m-strongly convex and L-smooth (i.e., $mI \preceq \nabla^2 f(x) \preceq LI$ for all $x \in \mathbb{R}^n$). Consider Newton's method with constant step size $t_k = m/L$

$$x^+ = x - \frac{m}{L} \nabla^2 f(x)^{-1} \nabla f(x).$$

Show that $f(x^+) - f(x) \leq -c\|\nabla f(x)\|_2^2$ for some constant $c > 0$ that depends only on m and L that you should specify.

6. (*2D Total variation denoising*) An image is represented by a matrix $b = (b_{ij})$ of size $N \times N$, where each entry b_{ij} represents the (i, j) pixel intensity. To denoise a noisy image b, we consider the following optimization-based approach (known as the Rudin-Osher-Fatemi model):

$$\min_{x \in \mathbb{R}^{N \times N}} \sum_{ij} (x_{i,j} - b_{i,j})^2 + \lambda \sum_{1 \leq i,j \leq N-1} \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2}.$$

The solution of this optimization problem is the candidate denoised image. Discuss algorithms you can use to solve this problem. Extra: Implement the methods you propose with $b = b_0 + \epsilon$ where b_0 is a clean image, and ϵ is some randomly generated Gaussian noise.