Univ.-Prof. Dr. Radu Ioan Boţ

Exercises for the lecture course "Convex Optimization"

Winter term 2021/2022

Solutions are to be submitted by October 31, 2021:

1. Let $(H, \langle \cdot, \cdot \rangle)$ be a real Hilbert space and $D \subseteq H$ a nonempty set. For $i = 1, ..., k$, let $T_i : D \to H$ be α_i-averaged operators, $\alpha_i \in (0,1)$, and $\omega_i \geq 0$ such that $\sum_{i=1}^k \omega_i = 1$. Show that $\sum_{i=1}^k \omega_i T_i$ is α-averaged with $\alpha = \max_{i=1,...,k} \alpha_i$.

2. Let $(H, \langle \cdot, \cdot \rangle)$ be a real Hilbert space, $D \subseteq H$ a nonempty set, $T_1 : D \to H$ an α_1-averaged operator and $T_2 : D \to D$ an α_2-averaged operator, where $\alpha_1, \alpha_2 \in (0,1)$. Show that $T_1 T_2$ is α-averaged with
$$\alpha := \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2} \in (0,1).$$

3. Let $(H, \langle \cdot, \cdot \rangle)$ be a real Hilbert space, $D \subseteq H$ a nonempty set and $T_i : D \to D$ α_i-averaged operators, where $\alpha_i \in (0,1)$ for $i = 1, ..., m$. Show that $T_1 T_2 ... T_m$ is α-averaged with
$$\alpha := \frac{1}{1 + \frac{1}{\sum_{i=1}^m \frac{\alpha_i}{1-\alpha_i}}} \in (0,1).$$

4. Let $(H, \langle \cdot, \cdot \rangle)$ be a real Hilbert space, $f : H \to \overline{\mathbb{R}}$ a proper function and $\beta > 0$. Prove that the following statements are equivalent:
   (i) f is strongly convex with constant β;
   (ii) $f - \frac{\beta}{2}\| \cdot \|^2$ is convex.

5. Let $(H, \langle \cdot, \cdot \rangle)$ be a real Hilbert space, $U \subseteq H$ a nonempty, open and convex set, $f : U \to \mathbb{R}$ a Gâteaux differentiable function on U and consider the following statements:
   (i) f is strictly convex on U;
   (ii) $\langle \nabla f(x), y - x \rangle < f(y) - f(x)\ \forall x, y \in U, x \neq y$;
   (iii) $\langle \nabla f(y) - \nabla f(x), y - x \rangle > 0\ \forall x, y \in U, x \neq y$;
   (iv) if f is twice Gâteaux differentiable on U, then $\nabla^2 f(x)(d,d) > 0\ \forall x \in U\ \forall d \in H, d \neq 0$.

   Prove that $(iv) \Rightarrow (i) \Leftrightarrow (ii) \Leftrightarrow (iii)$ and show that the first implication is not an equivalence in general.

6. (a) For $\underline{\alpha}, \overline{\alpha}$ real numbers, such that $\underline{\alpha} < \overline{\alpha}$, calculate the proximal operator of
$$\sigma_{[\underline{\alpha},\overline{\alpha}]} : \mathbb{R} \to \mathbb{R},\ \sigma_{[\underline{\alpha},\overline{\alpha}]}(x) = \begin{cases} \underline{\alpha}x, & \text{if } x < 0 \\ 0, & \text{if } x = 0 \\ \overline{\alpha}x, & \text{if } x \geq 0. \end{cases}$$
   (b) Calculate the proximal operator of the function $\| \cdot \|_1 : \mathbb{R}^n \to \mathbb{R}$.

7. For $\alpha > 0$, calculate the proximal operator of the following real-valued functions defined on $\mathbb{R}$:
   (a) $x \mapsto \max\{|x| - \alpha, 0\}$;

(b)

$$x \mapsto \begin{cases} \frac{1}{2}x^2, & \text{if } |x| \leq \alpha \\ \alpha|x| - \frac{\alpha^2}{2}, & \text{otherwise.} \end{cases}$$

Solutions to be submitted by December 5, 2021:

8. Let H be a real Hilbert spaces.

   (a) For $f : H \to \overline{\mathbb{R}}$ a proper, convex and lower semicontinuous function and $\gamma > 0$, prove that

   $$\operatorname{prox}_{\gamma({}^1 f)}(x) = \frac{x + \gamma \operatorname{prox}_{(\gamma+1)f}(x)}{\gamma + 1},$$

   where

   $${}^1 f : H \to \mathbb{R}, \ {}^1 f(x) = \inf_{y \in H} \left\{ f(y) + \frac{1}{2}\|y - x\|^2 \right\},$$

   denotes the Moreau envelope of f of parameter 1.

   (b) For $C \subseteq H$ a nonempty, convex and closed set, calculate the proximal operator of the squared distance function $\frac{1}{2}d_C^2$.

9. Let H be a real Hilbert space.

   (a) For $f : \mathbb{R} \to \overline{\mathbb{R}}$ is a proper, convex and lower semicontinuous function and $u \in H$, determine the proximal operator of the function $x \mapsto f(\langle u, x\rangle)$.

   (b) For H a finite-dimensional space, $e_1, ..., e_n$ an orthonormal basis of H and $f : H \to \overline{\mathbb{R}}$, $f(x) = \sum_{i=1}^n f_i(\langle x, e_i\rangle)$, where $f_i : \mathbb{R} \to \overline{\mathbb{R}}, i = 1, ..., n$, are proper, convex and lower semicontinuous functions, determine the proximal operator of f.

10. Let H be a real Hilbert space, $f : H \to \overline{\mathbb{R}}$ a proper, convex and lower semicontinuous function and $\gamma > 0$.

    (a) Prove that $\operatorname{prox}_{({}^\gamma f)}(x) = x + \frac{1}{\gamma+1}\left(\operatorname{prox}_{(\gamma+1)f}(x) - x\right) \ \forall x \in H$.

    (b) For $g : H \to \mathbb{R}, g(x) = \frac{1}{2\gamma}\|x\|^2 - ({}^\gamma f)(x)$, prove that $g(x) = (f + \frac{1}{2\gamma}\|\cdot\|^2)^* \left(\frac{1}{\gamma}x\right)$ for all $x \in H$, and deduce from here that g is convex.

    (c) Show that $\operatorname{prox}_g(x) = x - \frac{1}{\gamma}\operatorname{prox}_{\frac{\gamma^2}{\gamma+1}f}\left(\frac{\gamma}{\gamma+1}x\right) \ \forall x \in H$.

11. Implement the proximal point algorithm. Apply the algorithm to minimize the convex function

    $$f : \mathbb{R}^n \to \mathbb{R}, \ f(x) = \frac{\alpha}{2}\|x\|^2 + \|x\|_1,$$

    (i) by considering different values for the dimension n (for instance, $n = 1, 10, 100, 1000$) and for the starting point x^0;

    (ii) by considering different values for the parameter α (for instance, $\alpha = 0, 1, 10, 100, 1000, 10000$);

    (iii) by using as stopping criterion $\|x^k - x^*\| \leq 10^{-6}$, where x^* denotes the unique minimizer of f;

    (iv) by using the following choices for the stepsizes: $\gamma_k = 1, \forall k \geq 0$; $\gamma_k = \frac{1}{k+1}, \forall k \geq 0$; $\gamma_k = k+1, \forall k \geq 0$; $\gamma_k = (k+1)^2, \forall k \geq 0$; $\gamma_k = e^k, \forall k \geq 0$.

    Display the fixed point residual ($\|x^{k+1} - x^k\|, k = 0, 1, 2, ...$), the distance to the optimal solution ($\|x^k - x^*\|, k = 0, 1, 2, ...$), and the objective function values ($f(x^k) - f(x^*), k = 0, 1, 2, ...$) as functions of the number of iterations k (in separate plots).

12. Implement the proximal-gradient algorithm. Apply the algorithm to solve the minimization problem

$$\min_{x \in \mathbb{R}^n} f(x) + g(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}, f(x) = \frac{\alpha}{2}\|x\|^2 + \|x\|_1$, with $\alpha \geq 0$, and $g : \mathbb{R}^n \to \mathbb{R}, g(x) = \frac{1}{2}\|Ax - b\|^2$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$,

(i) by considering different values for the dimensions m and n;

(ii) by independently generating the entries of A and of b using a standard normal distribution;

(iii) by considering different values for the parameter α (for instance, $\alpha = 0, 1, 10, 100, 1000, 10000$);

(iv) by using different values for the and for the starting point x^0 and the stepsize $\gamma \in \left(0, \frac{2}{L_{\nabla g}}\right)$, taking also into consideration the restriction on the step size required in order to obtain covergence rates.

For every instance of the optimization problem, a given starting point and a given step size in $\left(0, \frac{1}{L_{\nabla g}}\right)$, first let the algorithm run for 10000 iterations, and set $x^* := x^{10000}$. Further, for various starting points and various stepsizes, stop the algorithm after 300 iterations and display $(\|x^k - x^*\|, k = 0, 1, 2, ..., 300)$ and $((f+g)(x^k) - (f+g)(x^*), k = 0, 1, 2, ..., 300)$ as functions of the number of iterations k (in separate plots).

Solutions to be submitted by January 23, 2022:

13. Implement the subgradient algorithm. Apply the algorithm to solve the minimization problem

$$\min_{(x_1, x_2) \in \mathbb{R}^2} g(x_1, x_2) = |x_1 + 2x_2| + |3x_1 + 4x_2|,$$

(i) by using $x^0 = (1, 2)$ as starting point;

(ii) by considering different choices for the sequence of step sizes $(t_k)_{k \geq 0}$, including the choice $t_k = \frac{g(x^k)}{\|\xi^k\|}$, where $\xi^k \in \partial g(x_1^k, x_2^k)$, for all $k \geq 0$.

Stop the algorithm after at most 100 iterations and display $(g(x_1^k, x_2^k), k = 0, 1, 2, ..., 100)$ and $(g_{\text{best}}^k, k = 0, 1, 2, ..., 100)$ as functions of the number of iterations k in one plot, and $((x_1^k, x_2^k), k = 0, 1, 2, ..., 100)$ as functions of the number of iterations over the contour lines of the function g.

14. Implement the fast proximal-gradient algorithm. Apply the algorithm to solve the minimization problem

$$\min_{x \in \mathbb{R}^n} f(x) + g(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}, f(x) = \|x\|_1$ and $g : \mathbb{R}^n \to \mathbb{R}, g(x) = \frac{1}{2}\|Ax - b\|^2$, where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$,

(i) by considering different values for the dimensions m and n;

(ii) by independently generating the entries of A and of b using a standard normal distribution;

(iii) by using the Nesterov rule and the Chambolle-Dossal rule for the momentum parameters;

(iv) by using different values for the starting point x^0.

For every instance of the optimization problem, a given starting point and step size $\gamma = \frac{1}{L_{\nabla g}}$, first let the algorithm run for 10000 iterations, and set $x^* := x^{10000}$. Further, for various starting points, run the fast proximal-gradient algorithm and the proximal-gradient algorithm with step size $\gamma = \frac{1}{L_{\nabla g}}$, both for 200 iterations. Display for both algorithms $(\|x^k - x^*\|, k = 0, 1, 2, ..., 200)$ and $((f+g)(x^k) - (f+g)(x^*), k = 0, 1, 2, ..., 200)$ as functions of the number of iterations k.

15. Solve the minimization problem

$$\min_{x \in \mathbb{R}^n} g(x),$$

where

$$g : \mathbb{R}^n \to \mathbb{R}, \quad g(x) = \begin{cases} \frac{1}{c}\|x\| - \frac{1}{2c^2}, & \text{if } \|x\| \geq \frac{1}{c}, \\ \frac{1}{2}\|x\|^2, & \text{otherwise,} \end{cases}$$

with the gradient algorithm and the fast gradient algorithm,

(i) by considering different values for the dimension $n \in \{1, 10, 50, 500, 5000\}$ and the parameter $c > 0$;

(ii) by using the Nesterov rule and the Chambolle-Dossal rule for the momentum parameters;

(iii) by using different values for the starting point x^0.

Run the fast proximal-gradient algorithm and the proximal-gradient algorithm with a given starting point and step size $\gamma = \frac{1}{L_{\nabla g}}$ for 200 iterations. Display for both algorithms $(\|x^k\|, k = 0, 1, 2, ..., 200)$ and $(g(x^k), k = 0, 1, 2, ..., 200)$ as functions of the number of iterations k.

16. (MAP versus DR) Implement the method of alternating projections (MAP) and the Douglas-Rachford (DR) algorithm for determining an element in the intersection of two sets. Apply the algorithms to find an element in the intersection of the sets

$$S = \mathbb{R}^2_+ \text{ and } T = \{(u, v) \in \mathbb{R}^2 : u + 5v = 6\},$$

(i) by using $d_T(x_k) \leq 10^{-4}$ as stopping criterion for the method of alternating projections.;

(ii) by using $d_S(P_T(x_k)) \leq 10^{-4}$ as stopping criterion for the Douglas-Rachford algorithm;

(iii) by choosing $x_0 \in \{(u_0, v_0) \in \mathbb{Z} \times \mathbb{Z} : u_0 \in [0, 100], v_0 \in [-100, 0]\}$ as starting points for both algorithms.

For each of the algorithms and all starting points display in a colored array the number of iterations needed to satisfy the stopping criteria.

17. Implement the Chambolle-Pock algorithm. Apply the algorithm to solve the minimization problem

$$\min_{x \in \mathbb{R}^n} f(x) + g(Ax) + h(x),$$

where $f : \mathbb{R}^n \to \mathbb{R}, f(x) = \|x\|_1$, $h : \mathbb{R}^n \to \mathbb{R}, h(x) = \frac{\alpha}{2}\|x\|^2$ with $\alpha \geq 0$, and $A \in \mathbb{R}^{m \times n}$,

(i) for $g : \mathbb{R}^m \to \mathbb{R}, g(y) = \|y\|$, and $g : \mathbb{R}^m \to \mathbb{R}, g(y) = \|y\|_1$;

(i) by considering different values for the dimensions m and n;

(ii) by independently generating the entries of A using a standard normal distribution;

(iii) by considering different values for the parameter α (for instance, $\alpha = 0, 1, 10, 100, 1000, 10000$);

(iv) by using different values for the starting point (x^0, y^0) and the stepsizes $c > 0$ and $\tau > 0$ chosen such that $1 > c\tau\|A\|^2$.

For various starting points and various stepsizes, run the algorithm for 300 iterations. Display $(\|x^k\|, k = 0, 1, 2, ..., 300)$, $(\|y^k\|, k = 0, 1, 2, ..., 300)$ and $((f + g \circ A + h)(\overline{x}^k), k = 1, 2, ..., 300)$, where $\overline{x}^k := \frac{1}{k}\sum_{i=1}^{k} x^k$, as functions of the number of iterations k (in separate plots).