First order methods in optimization (homework exercises)

Pourya Behmandpoor *

June 24, 2022

1 Part 1

For each of the following functions, find an expression for $\text{prox}_{\lambda f}(\boldsymbol{x})$ for any $\lambda > 0$ and $\boldsymbol{x}$.

1. $f(x) = 2x_{[1]} + x_{[2]}$. $\text{prox}_f((2,1,4,1,2,1)) =?$

   Solution:

   - $f(x) = \sigma_C$ where $C := \{w \in \mathbb{R}^n \mid \sum w_i = 3, 0 \le w_i \le 2,\ \forall i \in [n]\}$
   - $\text{prox}_{\lambda f}(x) = x - \lambda P_C(\frac{x}{\lambda})$
   - $P_C(x) = H_{a,b} \cap \text{Box}[\ell, u]$ where $\boldsymbol{a} = \boldsymbol{e}, b = 3, \boldsymbol{\ell} = \boldsymbol{0}, \boldsymbol{u} = \boldsymbol{2}$
   - the code is *ex1-1.py*

2. $f(x) = \begin{cases} \frac{1}{t}, & t > 0 \\ \infty, & \text{else} \end{cases}$

   Solution:

   - $\text{prox}_{\lambda f}(x) = \text{argmin}_{t>0}\{\frac{\lambda}{t} + \frac{1}{2}\|t - x\|^2\}$
   - then: $\text{prox}_{\lambda f}(x) = t_\star : t_\star^3 - x t_\star^2 - \lambda = 0$

3. $F(\boldsymbol{X}) = \begin{cases} \text{tr}(\boldsymbol{X}^{-1}), & \boldsymbol{X} \succ 0 \\ \infty, & \text{else} \end{cases}$

   Solution:

   - $F = f \circ \lambda : f(\boldsymbol{\lambda}) = \sum_{i=1}^n \frac{1}{\lambda_i}$
   - $$\begin{aligned} \text{prox}_{\alpha F}(\boldsymbol{X}) &= \boldsymbol{U}\,\text{diag}(\text{prox}_{\alpha f}(\boldsymbol{\lambda}(\boldsymbol{X})))\boldsymbol{U}^T \\ &= \boldsymbol{U}\,\text{diag}\left(\text{prox}_{\alpha f_1}(\boldsymbol{\lambda}_1(\boldsymbol{X})) \times \cdots \times \text{prox}_{\alpha f_n}(\boldsymbol{\lambda}_n(\boldsymbol{X}))\right)\boldsymbol{U}^T \end{aligned}$$
   - $f_i(x) = \frac{1}{x} : x > 0$. Hence: $\text{prox}_{\alpha f_i}(x) = t_\star$ from the previous exercise
   - $\text{prox}_F(\boldsymbol{X})$ with $\alpha = 1$ and $\boldsymbol{X} = \begin{pmatrix} 3 & 1 \\ 1 & 4 \end{pmatrix}$ is calculated using the code
   - the code is *ex1-2.py*

4. $f(\boldsymbol{x}) = (\|\boldsymbol{x}\|_2 - 1)^2$

   Solution:

   - $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2 - 2\|\boldsymbol{x}\|_2 + 1$
   - $\text{prox}_{\lambda g = -2\lambda\|\cdot\|_2}(\boldsymbol{x}) = \begin{cases} \left(1 + \frac{2\lambda}{\|\boldsymbol{x}\|}\right)\boldsymbol{x}, & \boldsymbol{x} \neq \boldsymbol{0}, \\ \{\boldsymbol{u} : \|\boldsymbol{u}\| = 2\lambda\}, & \boldsymbol{x} = \boldsymbol{0}. \end{cases}$
   - $\text{prox}_{\lambda f} = \text{prox}_{\lambda g + \lambda\|\cdot\|^2}(\boldsymbol{x}) = \text{prox}_{\frac{1}{2\lambda+1}g}(\frac{\boldsymbol{x}}{2\lambda+1}) = \begin{cases} \left(1 + \frac{2\lambda}{(2\lambda+1)\|\boldsymbol{x}\|}\right)\frac{\boldsymbol{x}}{2\lambda+1}, & \boldsymbol{x} \neq \boldsymbol{0}, \\ \{\boldsymbol{u} : \|\boldsymbol{u}\| = \frac{2\lambda}{2\lambda+1}\}, & \boldsymbol{x} = \boldsymbol{0}. \end{cases}$

*KU Leuven, Department of Electrical Engineering ESAT-STADIUS – Kasteelpark Arenberg 10, bus 2446, B-3001 Leuven, Belgium (pourya.behmandpoor@esat.kuleuven.be).

2 Part 2

2.1 exercise 0

1. $g(x) = \lambda_2 \|x\|_1$ then $\text{prox}_g(x) = \mathcal{T}_{\frac{\lambda_2}{L_f}}(x)$

2. $g(x) = \frac{\lambda_1}{2}\|x\|_2^2 + \lambda_2\|x\|_1$ then $\text{prox}_g(x) = \mathcal{T}_{\frac{\lambda_2}{L_f+\lambda_1}}(\frac{x}{\frac{\lambda_1}{L_f}+1})$

3. the code is *ex-p2-e0.py*

4. V-FISTA 1, i.e. on $g(x) = \lambda_2\|x\|_1$, performs the best

5. $x^{sol}_{PGM}(1:4) = (-0.4397, 0.0197, 1.4228, -0.8781)$
   $x^{sol}_{FISTA}(1:4) = (-0.4320, 0.0288, 1.4337, -0.9066)$
   $x^{sol}_{V-FISTA1}(1:4) = (-0.43210748, 0.02959795, 1.43437278, -0.90583604)$
   $x^{sol}_{V-FISTA2}(1:4) = (-0.43210796, 0.02959784, 1.43437041, -0.90583332)$

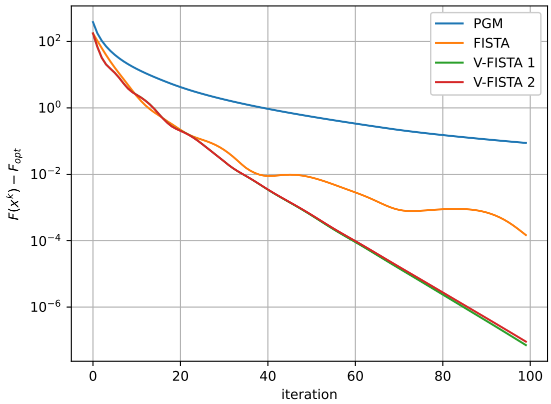

Figure 1: Part 2, exercise 0, V-FISTA 1 is with $g = \lambda_2\|\cdot\|_1$, and V-FISTA 2 is with $g = \frac{\lambda_1}{2}\|\cdot\|_2^2 + \lambda_2\|\cdot\|_1$

2.2 exercise 1

1. $H(\boldsymbol{x}) := \boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + 2\boldsymbol{b}^T\boldsymbol{x} + c$
   $\nabla H(\boldsymbol{x}) = 2\boldsymbol{b} + 2\boldsymbol{Q}\boldsymbol{x} = 0$, then $\bar{\boldsymbol{x}} = -\boldsymbol{Q}^{-1}\boldsymbol{b}$
   $H(\bar{\boldsymbol{x}}) = \boldsymbol{b}^T\boldsymbol{Q}^{-T}\boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{b} - 2\boldsymbol{b}^T\boldsymbol{Q}^{-1}\boldsymbol{b} + c \geq 0$
   $\Rightarrow c \geq \boldsymbol{b}^T\boldsymbol{Q}^{-1}\boldsymbol{b}$

2. as Q is positive definite, one can decompose it to have
   $f(\boldsymbol{x}) = \sqrt{\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + 2\boldsymbol{b}^T\boldsymbol{x} + c} = \sqrt{\|Px - v\|_2^2 + c'^2} = \| \begin{pmatrix} Px - v \\ c' \end{pmatrix} \|_2$
   where if $Q = U\Lambda U^T$, $P := \Lambda^{\frac{1}{2}}U^T$, $v := -\Lambda^{\frac{-1}{2}}U^Tb$, and $c'^2 := c - b^TQ^{-1}b$. As ℓ_p norm with $p \geq 1$ is convex, and composition of a convex function with an affine mapping is also convex, this norm is convex. Also sum of two convex function is convex, hence, the problem is convex.

3. $f(\boldsymbol{x}) = \sqrt{\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + 2\boldsymbol{b}^T\boldsymbol{x} + c}, \quad g(\boldsymbol{x}) = 0.2\|\boldsymbol{D}\boldsymbol{x}\|_1$
   $\nabla f(\boldsymbol{x}) = \frac{2\boldsymbol{b}+2\boldsymbol{Q}\boldsymbol{x}}{2\sqrt{H(\boldsymbol{x})}}$
   $\text{prox}_{\frac{0.2}{L_f}\|\cdot\|_1}(\boldsymbol{x}) = \mathcal{T}_{\frac{0.2}{L_f}}(\boldsymbol{x})$

$$\begin{aligned}\text{prox}_{\frac{g}{Lf}}(\boldsymbol{x}) &= \text{prox}_{\frac{0.2}{L_f}\|\cdot\|_1\circ\boldsymbol{D}}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{D}^T\,\text{prox}_{\frac{0.2}{L_f}\|\cdot\|_1}(\boldsymbol{D}\boldsymbol{x}) - \boldsymbol{D}\boldsymbol{x} \\ &= \text{prox}_{\frac{0.2}{L_f}\|\cdot\|_1\circ\boldsymbol{D}}(\boldsymbol{x}) = \boldsymbol{x} + \boldsymbol{D}^T\mathcal{T}_{\frac{0.2}{L_f}}(\boldsymbol{D}\boldsymbol{x}) - \boldsymbol{D}\boldsymbol{x}\end{aligned}$$

4. to estimate L_f we upper bound the Hessian of the function f. We have:
   $\nabla^2 f(x) = \frac{1}{\sqrt{\|Px-v\|^2+c'^2}}P^T\left[I - \frac{(Px+b)(Px+b)^T}{\|Px-v\|^2+c'^2}\right]P$

the eigenvalues of the bracket belong to $\{0,1\}$, as the matrix inside is a product of one vector by itself. We upper bound the Hessian as: $\nabla^2 f(x) \leq \frac{\|P\|^2}{\sqrt{\|Px-v\|^2+c'^2}} \leq \frac{\|P\|^2}{c'}$, Hence,
$L_f = \frac{\|\Lambda\|}{\sqrt{c-b^T Q^{-1} b}} = \frac{\|Q\|}{\sqrt{c-b^T Q^{-1} b}}$

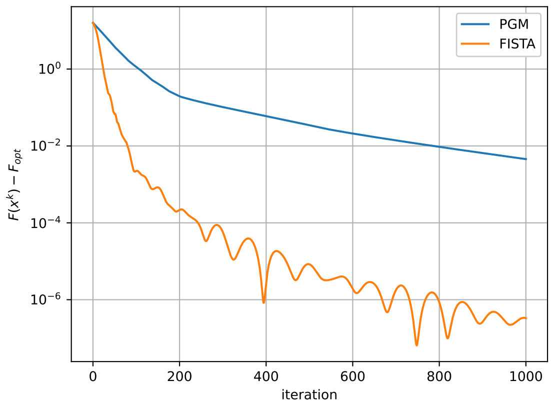

Figure 2: Part 2, exercise 1

5. the stepsize used in PGM is $\gamma = \frac{1}{L_f} = 0.02027373$

6. iterates:
   PGM:
   $iter\ 1:\ F(x^k) - F_{opt}: 15.969037688894332$
   $iter\ 101:\ F(x^k) - F_{opt}: 1.138840371899061$
   $iter\ 201:\ F(x^k) - F_{opt}: 0.19536585866318035$
   $iter\ 301:\ F(x^k) - F_{opt}: 0.10256671457047162$
   $iter\ 401:\ F(x^k) - F_{opt}: 0.05981160382616579$
   $iter\ 501:\ F(x^k) - F_{opt}: 0.034541962530351356$
   $iter\ 601:\ F(x^k) - F_{opt}: 0.021032280748006116$
   $iter\ 701:\ F(x^k) - F_{opt}: 0.013937040643956067$
   $iter\ 801:\ F(x^k) - F_{opt}: 0.009469387714531763$
   $iter\ 901:\ F(x^k) - F_{opt}: 0.0065280945179999605$
   $iter\ 1001:\ F(x^k) - F_{opt}: 0.004544681858035915$

   FISTA:
   $iter\ 1:\ F(x^k) - F_{opt}: 15.616787710263878$
   $iter\ 101:\ F(x^k) - F_{opt}: 0.002229198785176578$
   $iter\ 201:\ F(x^k) - F_{opt}: 0.00021635498722005764$
   $iter\ 301:\ F(x^k) - F_{opt}: 5.744173655841678e-05$
   $iter\ 401:\ F(x^k) - F_{opt}: 2.7570780005703455e-06$
   $iter\ 501:\ F(x^k) - F_{opt}: 8.335118572233569e-06$
   $iter\ 601:\ F(x^k) - F_{opt}: 1.9008889147187347e-06$
   $iter\ 701:\ F(x^k) - F_{opt}: 1.7973270658444562e-06$
   $iter\ 801:\ F(x^k) - F_{opt}: 8.541816285401183e-07$
   $iter\ 901:\ F(x^k) - F_{opt}: 2.646209011913925e-07$
   $iter\ 1001:\ F(x^k) - F_{opt}: 3.302678983629903e-07$

7. solutions:

   PGM: $(-0.17302056, -0.13170504, 0.2853231, -0.26288016, -0.53923318,$
   $-0.41890723, 0.34400762, -0.51913402, 0.38304034, -0.8028364,$
   $-0.156584, -0.27559503, 0.06735174, -0.17116589, -0.07071933,$
   $-0.07055706, -0.0558151, -0.13077329, -0.10362827, 0.16621555,$
   $-0.74600957, -0.41574597, 0.62302217, -0.33120591, -0.24841553,$
   $-0.63743909, -0.34855206, 0.2553829, -0.13449554, -0.0940163)$

FISTA: $(-0.23010618, -0.13942272, 0.31427133, -0.27633859, -0.59820196,$
$-0.49575369, 0.34118456, -0.5456482, 0.38702184, -0.91169326,$
$-0.17665491, -0.32830946, 0.02846801, -0.10373275, -0.14537967,$
$-0.07446206, -0.03326546, -0.1831482, -0.08021404, 0.26755059,$
$-0.84049738, -0.48603082, 0.65971966, -0.25551334, -0.2822796,$
$-0.75444863, -0.35508949, 0.28620091, -0.20633458, -0.18046632)$

2.3 exercise 3

1. $f(x) = \frac{1}{2}\|x\|_2^2$, and it is strongly convex with $\sigma = 1$
2. $g(Ax) := C\sum_{i=1}^{n} \max\{0, 1 - y_i \boldsymbol{w}^T \boldsymbol{x}_i\}$, take $C = 1$
   $A = \text{diag}(\boldsymbol{y})\boldsymbol{X}$, where $\boldsymbol{y} = (y_1, \ldots, y_n)$ and $\boldsymbol{X}$ has rows equal to $\boldsymbol{x}_i$
   $L_f = \|A\|^2/\sigma$
   $g(z) = h(1 - z)$ where we define:
   $h(z) := \sum_{i=1}^{n}\{0, z_i\} = \sigma_{\text{Box}[0,1]}(z)$
   $\text{prox}_{L_f h}(x) = z - L_f P_{\text{Box}[0,1]}(\frac{x}{L_f})$
   $\text{prox}_{L_f g}(x) = 1 - \text{prox}_{L_f h}(1 - x) = x + L_f P_{\text{Box}[0,1]}(\frac{1-x}{L_f})$
3. DPG iterates:

   $x^k = A^T y^k$
   $y^{k+1} = y^k - \frac{1}{L_f}Ax^k + \frac{Ax^k - L_f y^k}{L_f} + P_{\text{Box}[0,1]}(\frac{1 - Ax^k + L_f y^k}{L_f})$
4. solutions:

   by DPG: $(2.04031854, -0.91262279)$
   by FDPG: $(2.13382443, -1.01718168)$

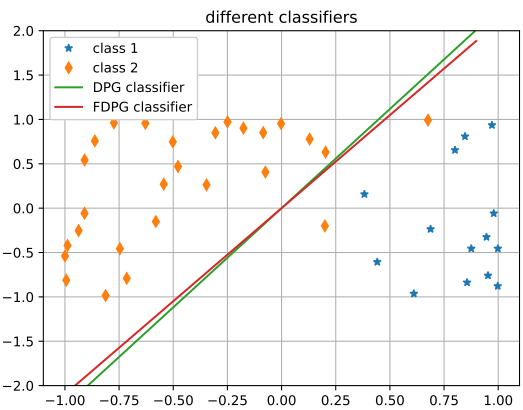

Figure 3: Part 2, exercise 3

3 Part 3

3.1 exercise 2 - part 1

1. ADLPMM: $f(x) + g(z)$ where $x + Bz = 0$ and $g = 0$
   with $B = (-A, -A, -D1, -D2)$
   A is a matrix with rows equal to a_i

$$D_1 = \begin{bmatrix} 1 & -1 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdot & 1 & -1 & 0 \end{bmatrix} \text{ and } D_2 = \begin{bmatrix} 0 & 1 & -1 & \cdots & 0 \\ 0 & 0 & 1 & -1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdot & 1 & -1 \end{bmatrix}$$

$f(z,w,v,u) = -\sum_{i=1}^m \log(a_i^T z - b_i) + \delta_{C:Aw\geq b}(w) + \sum_{i=1}^{n-2}\sqrt{v_i^2+u_i^2}$

$$\begin{aligned}\operatorname{prox}_{\lambda f}(z,w,v,u) = &\left(\frac{z_j - b_j + \sqrt{(z_j-b_j)^2+4\lambda}}{2} + b_j\right)_{j=1}^n \\ &\times (\max\{w_j, b_j\})_{j=1}^n \\ &\times \left((1-\frac{\lambda}{\max\{\sqrt{v_j^2+u_j^2},\lambda\}})v_j\right)_{j=1}^n \\ &\times \left((1-\frac{\lambda}{\max\{\sqrt{v_j^2+u_j^2},\lambda\}})u_j\right)_{j=1}^n\end{aligned}$$

$\operatorname{prox}_{\lambda g}(z) = z$
In the algorithm, $\alpha = \rho$ and $\beta = \rho\lambda_{max}(B^TB)$

2. CP: $f(Qx)+g(x)$ where $g=0$ and f is defined as above
with $Q=(A, A, D1, D2)$
In the algorithm, using Moreau Decomposition, $\operatorname{prox}_{\sigma f^*}(x) = x - \sigma\operatorname{prox}_{\frac{1}{\sigma}f}(\frac{x}{\sigma})$

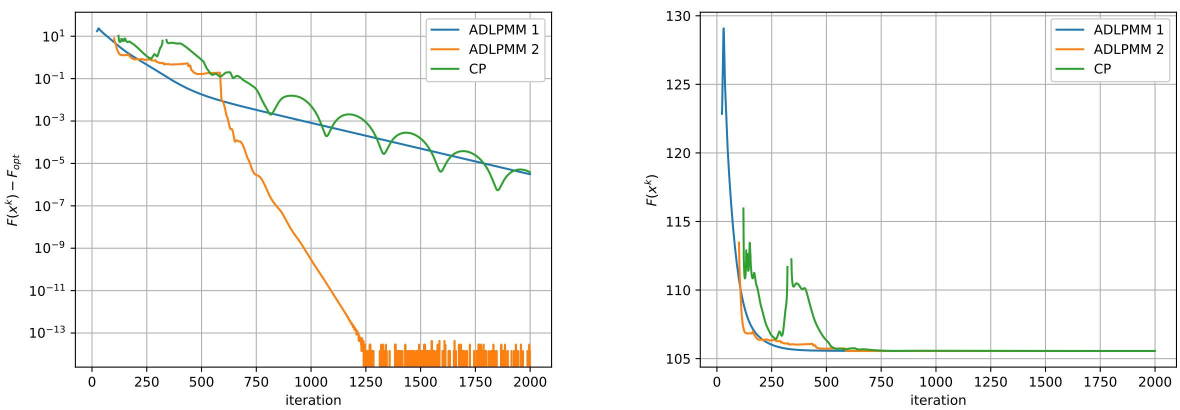

Figure 4: Part 3, exercise 2, section (c), ADLPMM-1 is with $\rho = 1$, ADLPMM-2 is with $\rho = \frac{1}{\|A\|_2}$, and CP is with $\tau = \sigma = \frac{1}{\|A\|_2}$

3. Solutions:
ADLPMM 1: $(-5.2156251, 1.41843662, 1.40372643)$
ADLPMM 2: $(-5.2148905, 1.41837664, 1.40349131)$
CP: $(-5.21427958, 1.41827063, 1.40324038)$

3.2 exercise 2 - part 2

1. ACP: $f(Qx)+g(x)$ where $g(x) = \frac{1}{2}\|x\|_2^2$, Q and f is defined as above
$\operatorname{prox}_{\lambda g}(x) = \frac{x}{1+\lambda}$
the strong convexity parameter $\gamma = 1$

2. FDPG: $L_f = \frac{\|Q\|_2^2}{\gamma}$ where $\gamma = 1$ is the strong convexity parameter
The update: $u^k = \operatorname{argmax}_u\{\langle u, Q^Tw^k\rangle - g(u)\} = Q^Tw^k$

3. Solutions:
ACP: $(-4.26504195, 1.15973671, 1.98584402)$
FDPG: $(-4.264957, 1.1596312, 1.98584261)$

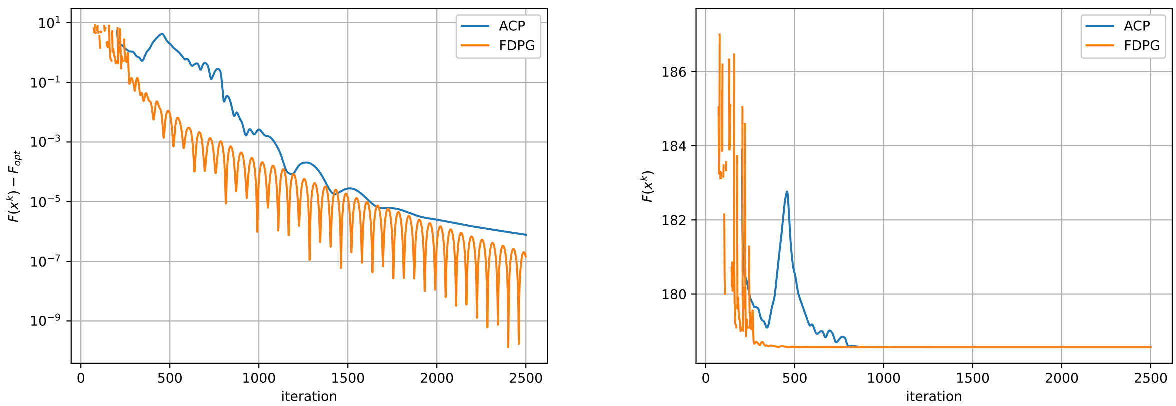

Figure 5: Part 3, exercise 2, section (e)