MTH786U/P, Semester C, 2020/21

Assignment 5 Solutions **M. Poplavskyi**

This week marks a mixture of theoretical and practical coursework.

Smoothed modulus function

As we have seen in the lecture, the modulus/absolute value function is an important element of the LASSO problem. However, due to its non-differentiability, it is a main reason of new optimisation methods being introduced in the module. In this exercise we consider one of them, namely changing the modulus with its smooth version

1. Prove that for any $x \in \mathbb{R}$ one has $|x| = \max_{p\in[-1,1]} xp$.

2. Let us now take any $\tau > 0$ and introduce the smoothed modulus function $|x|_\tau$ via

$$|x|_\tau = \max_{p\in[-1,1]} xp - \frac{\tau}{2}p^2.$$

   Show that the smoothed modulus function has the closed-form solution

$$|x|_\tau = \begin{cases} |x| - \frac{\tau}{2} & |x| > \tau \\ \frac{1}{2\tau}|x|^2 & |x| \leq \tau \end{cases}.$$

3. Sketch the plot of $|x|_\tau$.

4. For any vector $\mathbf{w} \in \mathbb{R}^n$ we define its Huber loss as

$$H_\tau(\mathbf{w}) = \sum_{j=1}^{n} |w_j|_\tau,$$

   and its soft-tresholding function as

$$\mathrm{soft}_\tau(\mathbf{w}) = \mathbf{w} - \tau \cdot \nabla H_\tau(\mathbf{w}).$$

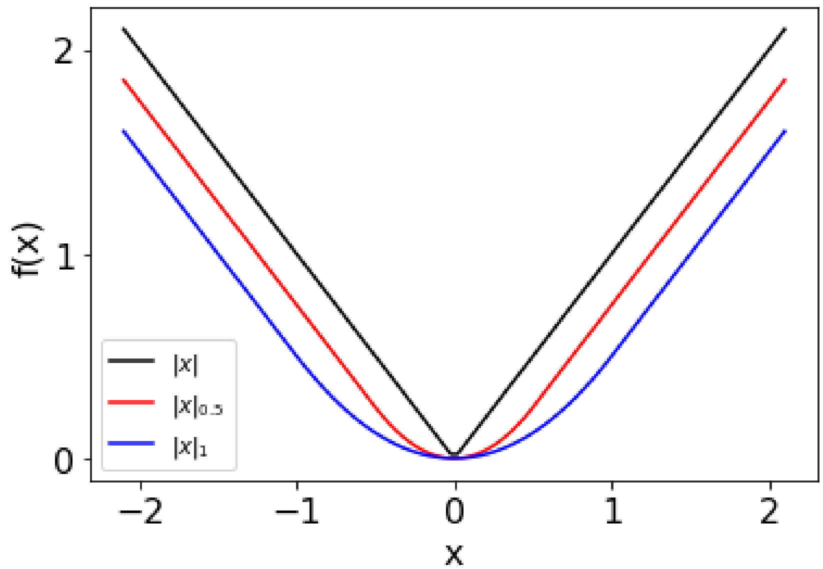
   Evaluate $\mathrm{soft}_\tau(\mathbf{w})$ explicitly.

Solutions:

1. We prove the statement by the case study, i.e. we consider two cases $x \geq 0$ and $x < 0$ and evaluate $\max_{p\in[-1,1]} xp$ in both cases. If $x \geq 0$, then the function $f_x(p) = xp$ is non-decreasing, and thus $\max_{p\in[-1,1]} = f_x(1) = x$. On the other hand, if $x < 0$ then the function $f_x(p) = xp$ is decreasing, and thus $\max_{p\in[-1,1]} = f_x(-1) = -x$. This shows that $|x| = \max_{p\in[-1,1]} xp$ for all $x \in \mathbb{R}$.

2. Let $f_\tau(p) = xp - \frac{\tau}{2}p^2$. Then its global maximiser is a solution of $f'_\tau(p) = 0 \Leftrightarrow x - \tau p \Leftrightarrow p = \frac{x}{\tau}$. If $|x| \leq \tau$, then $\frac{x}{\tau} \in [-1, 1]$ and $\max\limits_{p\in[-1,1]} f_\tau(p) = f_\tau\left(\frac{x}{\tau}\right) = \frac{x^2}{2\tau}$. If $|x| > \tau$, then the maximum of $f_\tau(x)$ on $[-1, 1]$ is attained at one of the endpoints, and thus
$$\max_{p\in[-1,1]} f_\tau(p) = \max_{p=\pm1} f_\tau(p) = \max_{p=\pm1} [xp] - \frac{\tau}{2} = |x| - \frac{\tau}{2}.$$

3. The plot of $|x|_\tau$ for $\tau = 0.5$ and $\tau = 1$ are as follows



4. The gradient of the Huber loss function is given by
$$\nabla H_\tau(w) = \left(\frac{\mathrm{d}}{\mathrm{d}w_1}|w_1|_\tau, \frac{\mathrm{d}}{\mathrm{d}w_2}|w_2|_\tau, \ldots, \frac{\mathrm{d}}{\mathrm{d}w_n}|w_n|_\tau\right).$$

   Thus the soft-tresholding function of vector $\mathbf{w}$ can be written as an element-wise application of the soft-tresholding:
$$\mathrm{soft}_\tau(\mathbf{w}) = (\mathrm{soft}_\tau(w_1), \ldots, \mathrm{soft}_\tau(w_n)).$$

   For a one dimensional argument w, we can use the explicit form of the smoothed absolute value function $|\cdot|_\tau$ to find:
$$\frac{\mathrm{d}}{\mathrm{d}w}|w|_\tau = \begin{cases} 1, & w > \tau, \\ \frac{1}{\tau}w, & |w| \leq \tau, \\ -1, & w < -\tau, \end{cases}$$

   that yields
$$\mathrm{soft}_\tau(w) = \begin{cases} w - \tau, & w > \tau, \\ 0, & |w| \leq \tau, \\ w + \tau, & w < -\tau. \end{cases}$$

Proximal maps

Another approach to deal with the LASSO we will meet in the module is a so-called proximal gradient descent. It is an iterative method of solving the problem

$$\hat{w} = \arg\min_{w} \left\{ L\left(w\right) + R\left(w\right) \right\},$$

where $E\left(w\right)$ is a differentiable, convex function, while R is just a convex function, continuous function that may be non-differentiable (such as modulus function for example). The main tool of the method are the proximal map prox_R defined as

$$\text{prox}_{\tau R}\left(z\right) = \left(I + \tau \partial R\right)^{-1}\left(z\right) = \arg\min_{x \in \mathbb{R}^n} \left[\frac{1}{2}\left\|x - z\right\|^2 + \tau R\left(x\right)\right].$$

This maps the variable z to the minimiser of the above function.

1. Let $n = 1$, that means $R : \mathbb{R} \to \mathbb{R}$ is a one-dimensional function of one-dimensional argument and

   $$\text{prox}_{\tau R}\left(z\right) = \arg\min_{x \in \mathbb{R}} \left[\frac{1}{2}\left(x - z\right)^2 + \tau R\left(x\right)\right].$$

   Find the proximal map for

   - $R\left(x\right) := x^2$
   - $R\left(x\right) := \alpha\left|x\right|$
     Hint: if z is your input argument, make the assumption $x = \lambda z$ for your solution of the proximal map, for a scalar $\lambda \in \mathbb{R}$.
   - $R\left(x\right) := \begin{cases} 0 & x \in [-1, 1] \\ \infty & x \notin [-1, 1] \end{cases}$
   - $R\left(x\right) := \begin{cases} 0 & x \in \mathcal{C} \\ \infty & x \notin \mathcal{C} \end{cases}$, for some convex $\mathcal{C} \subset \mathbb{R}$.

2. Compute the proximal map for $R(x) := \frac{1}{2}\|Dx\|^2$ for some matrix $D \in \mathbb{R}^{m \times n}$.

3. Write the proximal map for $R(x) := aS(x - y) + b$, for $y \in \mathbb{R}^n$, constants $a, b \in \mathbb{R}$ with $a > 0$ and a convex function S in terms of the proximal map of S.

4. Write the proximal map for a function $R : \mathbb{R}^n \to \mathbb{R}^n$ defined as $R(x) := S(x) + \alpha\|x\|^2 + \langle x, y\rangle$, for fixed $y \in \mathbb{R}^n$ and $\alpha > 0$, in terms of the proximal map of S.

Solutions:

1. The proximal maps for the following one dimensional functions can be calculated as follows:

   - for $R\left(x\right) = x^2$ we have

     $$\text{prox}_{\tau R}\left(z\right) = \arg\min_{x \in \mathbb{R}} \left[\underbrace{\frac{1}{2}\left(x - z\right)^2 + \tau x^2}_{E(x)}\right].$$

The minimiser of the energy function $E(x)$ can be found as a root of the derivative of E:

$$E'(x) = 0 \Leftrightarrow x - z + 2\tau x = 0 \Rightarrow x = \frac{1}{2\tau + 1} z.$$

Thus, $\text{prox}_{\tau R}(z) = \frac{1}{2\tau+1} z$.

- for $R(x) = \alpha |x|$ we have

$$\text{prox}_{\tau R}(z) = \arg\min_{x \in \mathbb{R}} \left[\underbrace{\frac{1}{2}(x-z)^2 + \tau\alpha |x|}_{E(x)} \right].$$

To minimise the energy function $E(x)$ we first change the variables via introducing $x = \lambda z$.

Remark: if $z = 0$, then obviously $x = 0$ is the minimiser of $E(x)$ and thus we can still write $x = \lambda z$ for any λ.

Changing the variables into $E(x)$ one gets

$$E\left(\lambda = \frac{x}{z}\right) = \frac{1}{2} z^2 (\lambda - 1)^2 + \tau\alpha |\lambda| |z|.$$

Let us now consider two cases $\lambda \geq 0$ and $\lambda < 0$. For $\lambda \geq 0$ the energy function can be rewritten as

$$E\left(\lambda = \frac{x}{z}\right) = \frac{1}{2} z^2 \lambda^2 + \left(\tau\alpha |z| - z^2\right)\lambda + \frac{z^2}{2}.$$

This is minimised at $\lambda = -\frac{\tau\alpha|z| - z^2}{2 \cdot \frac{1}{2} z^2} = 1 - \frac{\tau\alpha}{|z|}$. If this value satisfies $1 - \frac{\tau\alpha}{|z|} \geq 0$, then this is a proper minimiser on the set $\lambda \geq 0$, otherwise the minimiser is $\lambda = 0$. Analogously, for $\lambda < 0$ the energy function can be rewritten as

$$E\left(\lambda = \frac{x}{z}\right) = \frac{1}{2} z^2 \lambda^2 - \left(\tau\alpha |z| + z^2\right)\lambda + \frac{z^2}{2},$$

which is minimised at $\lambda = \frac{\tau\alpha|z| + z^2}{2 \cdot \frac{1}{2} z^2} = 1 + \frac{\tau\alpha}{|z|} > 0$. Thus on the set $\lambda < 0$ the energy function is minimised at $\lambda = 0$. Combining all the above we obtain

$$\min_{\lambda} E(\lambda) = \begin{cases} E\left(1 - \frac{\tau\alpha}{|z|}\right), & |z| \geq \tau\alpha, \\ E(0), & |z| \leq \tau\alpha. \end{cases}$$

Plugging back $x = \lambda z$ one obtains

$$\text{prox}_{\tau R}(z) = \begin{cases} z - \text{sign}(z)\,\tau\alpha, & |z| \geq \tau\alpha, \\ 0, & |z| \leq \tau\alpha. \end{cases} = \begin{cases} z - \tau\alpha, & z \geq \tau\alpha, \\ 0, & |z| \leq \tau\alpha, \\ z + \tau\alpha, & z \leq -\tau\alpha. \end{cases}$$

- because $R(x)$ take infinite value outside of the interval $[-1, 1]$ we can write the corresponding proximal map as

$$\text{prox}_{\tau}(z) = \arg \min_{x \ in[-1,1]} \frac{1}{2}(x - z)^2.$$

Obviously, if $z\ in\,[-1,1]$, then the minimiser is $x = z$. If $z \notin [-1,1]$, then then x should be the closest point to z of the interval $[-1,1]$. This can be either 1 or -1 depending on which side of the interval is z. All together this implies

$$\operatorname{prox}_\tau(z) = \begin{cases} 1, & z > 1, \\ z, & |z| \leq 1, \\ -1, & z < -1. \end{cases}$$

- Any convex subset of $\mathbb{R}$ is an interval. Let us denote this interval as $[a, b]$. Then analogously to the above

$$\operatorname{prox}_\tau(z) = \begin{cases} b, & z > b, \\ z, & |z| \leq 1, \\ a, & z < a. \end{cases}$$

2. The proximal mapping for $f(x) := \frac{1}{2}\|Dx\|^2$ reads

$$(I + \partial f)^{-1}(z) = \arg\min_x \left\{ E(x) := \frac{1}{2}\|x - z\|^2 + \frac{1}{2}\|Dx\|^2 \right\} .$$

As the function E is differentiable and convex, we can simply compute the gradient of E and set it to zero in order to determine the argument that minimises the expression. The gradient is

$$\nabla E(x) = x - z + D^T Dx \, .$$

Setting this equation to zero and solving for x yields $x = (I + D^T D)^{-1} z$. Hence, the proximal mapping reads as

$$(I + \partial f)^{-1}(z) = (I + D^T D)^{-1} z \, .$$

3. The proximal mapping reads

$$\begin{aligned} (I + \partial f)^{-1}(z) &= \arg\min_x \left\{ \frac{1}{2}\|x - z\|^2 + aS(x - y) + b \right\} \\ &= \arg\min_x \left\{ \frac{1}{2}\|x - z\|^2 + aS(x - y) \right\} , \end{aligned}$$

where we got rid of the constant as it does not affect the solution of the minimisation problem. For the substitution $x = w + y$ we can conclude

$$\begin{aligned} (I + \partial f)^{-1}(z) &= y + \arg\min_w \left\{ \frac{1}{2}\|w - (z - y)\|^2 + aS(w) \right\} \\ &= y + (I + \partial(aS))^{-1}(z - y) \end{aligned}$$

4. The proximal map reads

$$
\begin{aligned}
(I + \partial f)^{-1}(z) &= \arg\min_x \left\{ \frac{1}{2}\|x - z\|^2 + S(x) + \alpha\|x\|^2 + \langle x, y\rangle \right\} \\
&= \arg\min_x \left\{ \frac{1}{2}\|x\|^2 - \langle x, z\rangle + S(x) + \alpha\|x\|^2 + \langle x, y\rangle \right\} \\
&= \arg\min_x \left\{ \frac{1 + 2\alpha}{2}\|x\|^2 - \langle x, z - y\rangle + S(x) \right\} \\
&= \arg\min_x \left\{ \frac{1 + 2\alpha}{2} \left(\|x\|^2 - 2 \left\langle x, \frac{z - y}{1 + 2\alpha} \right\rangle \right) + S(x) \right\} \\
&= \arg\min_x \left\{ \frac{1 + 2\alpha}{2} \left(\|x\|^2 - 2 \left\langle x, \frac{z - y}{1 + 2\alpha} \right\rangle \right) + S(x) \right\} \\
&= \arg\min_x \left\{ \frac{1 + 2\alpha}{2} \left\| x - \frac{z - y}{1 + 2\alpha} \right\|^2 + S(x) \right\} \\
&= \arg\min_x \left\{ \frac{1}{2} \left\| x - \frac{z - y}{1 + 2\alpha} \right\|^2 + \frac{1}{1 + 2\alpha} S(x) \right\} \\
&= \left(I + \frac{1}{1 + 2\alpha} \partial g \right)^{-1} \left(\frac{z - y}{1 + 2\alpha} \right) .
\end{aligned}
$$