This homework is due at **2pm, November 23**.

Problem 1

Let $(x_1, y_1), \dots, (x_n, y_n) \in \mathbb{R}^p \times \mathbb{R}$ be the data. Consider the ℓ_1-penalized ℓ_1-regression problem:

$$\hat{\beta} \in \operatorname*{arg\,min}_{\beta} \left\{ f(\beta) + g(\beta) \,\middle|\, \beta \in \mathbb{R}^p \right\},$$

where

$$f(\beta) := \sum_{i=1}^{n} |y_i - \langle x_i, \beta \rangle|, \quad g(\beta) := \lambda \|\beta\|_1,$$

for some penalization parameter $\lambda > 0$.

1. (10 points) The Huber loss is given by

$$H_\mu(z) := \begin{cases} \frac{z^2}{2\mu}, & |z| \le \mu, \\ |z| - \frac{\mu}{2}, & \text{otherwise}, \end{cases},$$

   for every $\mu > 0$ and $z \in \mathbb{R}$. **Show that the Huber loss is the Moreau envelope of the absolute value function, i.e.,**

$$H_\mu(z) := \min_{w} \left\{ |w| + \frac{1}{2\mu}(w - z)^2 \,\middle|\, w \in \mathbb{R} \right\}.$$

2. (10 points) Define

$$f_\mu(\beta) := \sum_{i=1}^{n} H_\mu(\langle x_i, \beta \rangle - y_i).$$

   Show that f_μ is L_μ-smooth with $L_\mu := (1/\mu)\|X\|_{2\to 2}^2$ with respect to the 2-norm, where $X \in \mathbb{R}^{n \times p}$ denotes the matrix whose i-th row is given by x_i^{T}.

   HINT: You may consider first showing that $h_\mu(\beta) := \sum_{j=1}^{p} H_\mu(\beta^{(j)})$ is the Moreau envelope of the ℓ_1-norm function, and it is $(1/\mu)$-smooth with respect to the 2-norm.

3. (10 points) **Show that**

$$f(\beta) - \frac{n\mu}{2} \le f_\mu(\beta) \le f(\beta), \quad \forall \beta \in \mathbb{R}^p.$$

4. (20 points) **Provide an algorithm that, given $\varepsilon > 0$, finds some $\tilde{\beta}$ such that**

$$(f + g)(\tilde{\beta}) - (f + g)(\hat{\beta}) \le \varepsilon,$$

   after calling the first-order oracle associated with f_μ for $O(1/\varepsilon)$ times.

Problem 2

The *alternating direction method of multipliers (ADMM)* is an optimization method that solves the problem

$$(x_1^\star, x_2^\star) \in \operatorname*{arg\,min}_{(x_1, x_2)} \left\{ \varphi(x_1) + \psi(x_2) \,\middle|\, x_1 \in \mathbb{R}^{p_1}, x_2 \in \mathbb{R}^{p_2}, A_1 x_1 + A_2 x_2 = b \right\},$$

Problem 2 continued on the next page...

for given $A_1 \in \mathbb{R}^{p\times p_1}$, $A_2 \in \mathbb{R}^{p\times p_2}$ and $b \in \mathbb{R}^p$, where φ and ψ are proper closed convex functions. Let $\kappa > 0$, $\lambda_0 \in \mathbb{R}^p$ and $x_{2,0} \in \mathbb{R}^{p_2}$. The ADMM iterates as, for every $t = 0,1,\dots$,

$$
\begin{aligned}
x_{1,t+1} &\leftarrow \operatorname*{arg\,min}_{x_1} \left\{ \varphi(x_1) + \langle \lambda_t, A_1 x_1 \rangle + \frac{\kappa}{2} \| A_1 x_1 + A_2 x_{2,t} - b \|_2^2 \,\middle|\, x_1 \in \mathbb{R}^{p_1} \right\}, \\
x_{2,t+1} &\leftarrow \operatorname*{arg\,min}_{x_2} \left\{ \psi(x_2) + \langle \lambda_t, A_2 x_2 \rangle + \frac{\kappa}{2} \| A_1 x_{1,t+1} + A_2 x_2 - b \|_2^2 \,\middle|\, x_2 \in \mathbb{R}^{p_2} \right\}, \\
\lambda_{t+1} &\leftarrow \lambda_t + \kappa \left(A_1 x_{1,t+1} + A_2 x_{2,t+1} - b \right).
\end{aligned}
$$

The ADMM is guaranteed to converge for any $\kappa > 0$; the iteration complexity of the ADMM is $O(1/\varepsilon)$ in general [2, 1], and can be $O(\log(1/\varepsilon))$ if either f or g is strongly convex [3].

Suppose that now we would like to solve the optimization problem

$$
x^\star \in \operatorname*{arg\,min}_{x} \left\{ f(x) + g(x) \,\middle|\, x \in \mathbb{R}^p \right\}, \tag{1}
$$

where

$$
f(x) := \sum_{i=1}^{n} f_i(x),
$$

for proper closed convex functions $f_1, \dots, f_n$, and g is a proper closed convex function.

1. (20 points) **Show that the optimization problem** (1) **can be solved via the following method.** Let $\kappa > 0$, $\lambda_{1,0}, \lambda_{2,0}, \dots, \lambda_{n,0}, z_0 \in \mathbb{R}^p$. For every $t = 0,1,\dots$,

$$
\begin{aligned}
x_{i,t+1} &\leftarrow \operatorname*{arg\,min}_{x_i} \left\{ f_i(x_i) + \langle \lambda_{i,t}, x_i \rangle + \frac{\kappa}{2} \| x_i - z_t \|_2^2 \,\middle|\, x_i \in \mathbb{R}^p \right\}, \quad \forall i = 1,2,\dots,n, \\
z_{t+1} &\leftarrow \operatorname*{arg\,min}_{z} \left\{ g(z) - \sum_{i=1}^{n} \left(\langle \lambda_{i,t}, z \rangle - \frac{\kappa}{2} \| x_{i,t+1} - z \|_2^2 \right) \,\middle|\, z \in \mathbb{R}^p \right\}, \\
\lambda_{i,t+1} &\leftarrow \lambda_{i,t} + \kappa \left(x_{i,t+1} - z_{t+1} \right), \quad \forall i = 1,2,\dots,n.
\end{aligned} \tag{2}
$$

2. (10 points) Indeed, computing z_{t+1} in (2) corresponds to computing the proximal mapping associated with g. Define, for every t,

$$
\bar{x}_t := \frac{1}{n} \sum_{i=1}^{n} x_{i,t}, \quad \bar{\lambda}_t := \frac{1}{n} \sum_{i=1}^{n} \lambda_{i,t}.
$$

Show that (2) **can be equivalently written as**

$$
z_{t+1} \leftarrow \operatorname{prox}_{g/(n\kappa)} \left(\bar{x}_{t+1} + \frac{1}{\kappa} \bar{\lambda}_t \right).
$$

3. (10 points) Suppose that $n = 4m$ for some positive integer m. **Show how we can solve** (1) **using four processors, all of which are connected to a central unit.**

Recall the lasso, which corresponds to the case where

$$
f(x) := \| y - Ax \|_2^2, \quad g(x) := \tau \| x \|_1,
$$

for given $y \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times p}$, and penalization parameter $\tau > 0$. Suppose that $n = 4m$ for some positive integer m. Decompose y and A into four equal-sized blocks as

$$
y := \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}, \quad A := \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \end{bmatrix}.
$$

Problem 2 continued on the next page…

4. (10 points) **Show that we can compute the lasso via the following algorithm.** Let $\kappa > 0$, $\lambda_{1,0}, \lambda_{2,0}, \lambda_{3,0}, \lambda_{4,0}, z_0 \in \mathbb{R}^p$. For every $t = 0, 1, \ldots$,

$$\begin{aligned}
\bar{\lambda}_t &\leftarrow \frac{1}{4} \sum_{i=1}^{4} \lambda_{i,t} \\
x_{i,t+1} &\leftarrow \left(A_i^{\mathrm{T}} A_i + \frac{\kappa}{2} I \right)^{-1} \left[A_i^{\mathrm{T}} y_i + \frac{\kappa}{2} \left(z_t - \frac{1}{\kappa} \lambda_{i,t} \right) \right], \quad \forall\, i = 1, 2, 3, 4, \\
\bar{x}_{t+1} &\leftarrow \frac{1}{4} \sum_{i=1}^{4} x_{i,t+1} \\
z_{t+1} &\leftarrow \mathrm{soft}_{\tau/(n\kappa)} \left(\bar{x}_{t+1} + \frac{1}{\kappa} \bar{\lambda}_t \right), \\
\lambda_{i,t+1} &\leftarrow \lambda_{i,t} + \kappa \left(x_{i,t+1} - z_{t+1} \right), \quad \forall\, i = 1, 2, \ldots, 4,
\end{aligned}$$

where $I \in \mathbb{R}^{p \times p}$ denotes the identity matrix, and $\mathrm{soft}(\cdot)$ is the soft-thresholding operator defined in the lecture slides.

References

[1] HE, B., AND YUAN, X. On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method. *SIAM J. Numer. Anal. 50*, 2 (2012), 700–709.

[2] HE, B., AND YUAN, X. On non-ergodic convergence rate of Douglas-Rachford alternating direction method of multipliers. *Numer. Math. 130* (2015), 567–577.

[3] NISHIHARA, R., LESSARD, L., RECHT, B., PACKARD, A., AND JORDAN, M. I. A general analysis of the convergence of ADMM. In *Proc. 32nd Int. Conf. Machine Learning* (2015).