Take Home Final - Analysis of High Dimensional Data

1a. $f(x) = \|x\|_1$ with domain of $f = \{x : \|x\|_\infty \leq 1\}$. Note that

$$\begin{aligned}
[\text{prox}_f(x)]_i &= \arg\min_u [(\|u\|_1 + \frac{1}{2}\|u - x\|_2^2)]_i \\
&= \begin{cases} x_i - 1 & \text{if } x_i > 1 \\ 0 & \text{if } |x_i| \leq 1 \\ x_i + 1 & \text{if } x_i < -1 \end{cases}
\end{aligned}$$

Now for $\text{prox}_f(x)$ such that $\{u : \|u\|_\infty \leq 1\}$ we have $\arg\min_u(\|u\|_1 + \frac{1}{2}\|u - x\|_2^2 + \|u\|_\infty)$. We can come up with an ADMM algorithm for this using the constraint that $z = u$. Thus we have

$$\arg\min \|u\|_1 + \frac{1}{2}\|u - x\|_2^2 + \|z\|_\infty \text{ such that } z = u$$

Then the augmented Lagrangain is

$$\mathcal{L}_\rho(u, z, \lambda) = \|u\|_1 + \frac{1}{2}\|u - x\|_2^2 + \|z\|_\infty - \mu^T(z - u) + \frac{\rho}{2}\|z - u\|_2^2$$

Hence we have that

$$\begin{aligned}
u^{k+1} &= \arg\min_u \mathcal{L}_\rho(u, z^k, \mu^k) \\
z^{k+1} &= \arg\min_z \mathcal{L}_\rho(u^{k+1}, z, \mu^k) \\
\mu^{k+1} &= \mu^k + \rho(z - u)
\end{aligned}$$

Now keeping everything other that u fixed we have

$$\begin{aligned}
\mathcal{L}_\rho(u, z, \mu) &= \|u\|_1 + \frac{1}{2}\|u - x\|_2^2 - \mu^T(z - u) + \frac{\rho}{2}\|z - u\|_2^2 \\
\implies \partial_u \mathcal{L}_\rho(u, z, \mu) &= s + (u - x) - \mu + \rho(u - z)
\end{aligned}$$

where $s_i = \text{sign}(u_i)$. Thus we can run a subgradient algorithm for u. Now keeping everthing other than z fixed we have

$$\begin{aligned}
\mathcal{L}_\rho(u, z, \mu) &= \|z\|_\infty - \mu^T(z-u) + \frac{\rho}{2}\|z-u\|_2^2 \\
\implies \min_z \mathcal{L}_\rho(u, z, \mu) &= \min_z \frac{1}{2}\|z\|_\infty - \mu^T z + \arg\min_z \frac{\rho}{2}\|z\|_2^2 - \rho u^T z \\
\implies \arg\min_z \mathcal{L}_\rho(u, z, \mu) &= \arg\max_{\mu \geq 0} -\frac{1}{2}\|\mu\|_1 + \arg\max_{\mu \geq 0} -\frac{1}{2\rho}\|\mu - 2\rho u\|_2^2 \\
&= \arg\min_{\mu \geq 0} \frac{1}{2}\|\mu\|_1 + \frac{1}{2\rho}\|\mu - 2\rho u\|_2^2
\end{aligned}$$

Thus the z update step is equivalent to solving a *lasso* problem.

1b. $f(x) = \max_{k=1,\ldots,n} x_k = \|x\|_\infty$. Thus

$$\text{prox}_f(x) = \arg\min_u (\|u\|_\infty + \frac{1}{2}\|u - x\|_2^2)$$

Note that

$$\begin{aligned}
&\min_u \|u\|_\infty + \frac{1}{2}\|u - x\|_2^2 \\
= &\min_u (\|u\|_\infty - x^T u) + \min_u \frac{1}{2}\|u\|_2^2
\end{aligned}$$

We can minimize the second part with respect to u by taking derivatives and getting that $u = 0$. Now $\arg\min_x \frac{1}{2}\|x\|_1 + \frac{1}{2}\|x - u\|_2^2 = S_{1/2}(u)$, which implies that

$$[S_{1/2}(u)]_i = \begin{cases} u_i - 1/2 & \text{if } u_i > 1/2 \\ 0 & \text{if } |u_i| \leq 1/2 \\ u_i + 1/2 & \text{if } u_i < -1/2 \end{cases}$$

2 We want to solve the following problem:

Primal: $\min(x_1 - x_2)$

subject to $x \in X = \{x : x_1 \geq 0, x_2 \geq 0\}, x_1 + 1 \leq 0, 1 - x_1 - x_2 \leq 0$

Now the Lagrangian is

$$\begin{aligned}
\mathcal{L}(x,\lambda) &= (x_1 - x_2) + \lambda_1(-x_1) + \lambda_2(-x_2) \\
&\quad + \lambda_3(x_1+1) + \lambda_4(1 - x_1 - x_2) \\
&= x_1(1 - \lambda_1 + \lambda_3 - \lambda_4) + x_2(-1 - \lambda_2 - \lambda_4) + \lambda_3 + \lambda_4 \\
\implies g(\lambda_1,\lambda_2,\lambda_3,\lambda_4) &= \inf_x \mathcal{L}(x,\lambda) \\
&= \begin{cases} -\infty & \text{if } 1 - \lambda_1 + \lambda_3 - \lambda_4 \neq 0 \text{ or } -1 - \lambda_2 - \lambda_4 \neq 0 \\ \lambda_3 + \lambda_4 & \text{else} \end{cases}
\end{aligned}$$

Thus the dual problem is

$$\begin{aligned}
\textbf{Dual:} \max_{\lambda_i \geq 0} &\ \lambda_3 + \lambda_4 \\
&\text{subject to } 1 - \lambda_1 + \lambda_3 - \lambda_4 = 0 \\
&\text{and } -1 - \lambda_2 - \lambda_4 = 0
\end{aligned}$$

The constraints have infinitely many solutions. Hence this is clearly maximized when $\lambda_i = \infty, i = 3,4$ which implies that the dual is unbounded.

3a. Let $A_j \in \mathbb{R}^{m\times n}, y_j \in \mathbb{R}$. We want to solve the following problem:

$$\begin{aligned}
\textbf{Primal:} \min_{X\in\mathbb{R}^{m\times n}} &\sum_{j=1}^{J} (y_j - tr(X'A_j))^2 \\
&\text{subject to } \|X\|_* \leq t
\end{aligned}$$

or equivalently,

$$\min_{X\in\mathbb{R}^{m\times n}} \sum_{j=1}^{J} (y_j - tr(X'A_j))^2 + \lambda\|X\|_*$$

For our simulations we used, $m = 10, n = 5$ and $J = 5$. In addition, A_j are matrices with elements randomly generated from $\mathcal{U}(-5,5)$. To generate the true X we generated a matrix with elements from $\mathcal{U}(-5,5)$ took the SVD as $U\Sigma V'$ and set $X = UV'$. We then calculated $y_j = tr(X'A_j) + \epsilon_j$ where $\epsilon_j \sim \mathcal{N}(0,1)$. Note that $\frac{d}{dX}\sum_{j=1}^{J}(y_j - tr(X'A_j))^2 = -2\sum_{j=1}^{J}(y_j - tr(X'A_j))A_j$.

Thus a proximal gradient method would be

$$X^{k+1} = X^k + t(2\sum_{j=1}^{J}(y_j - tr((X^k)'A_j))A_j) \tag{1}$$

$$X^{k+1} = S_{t\lambda}(X^{k+1}) \tag{2}$$

where t is the stepsize and $S_{t\lambda}(X^{k+1}) = US_{t\lambda}(\Sigma)V'$. Here,

$$\begin{aligned} X &= U\Sigma V' \\ &= Udiag(\sigma_1, ..., \sigma_p)V' \end{aligned}$$

is the SVD of X and $S_{t\lambda}(\Sigma) = diag(S_{t\lambda}(\sigma_1), ..., S_{t\lambda}(\sigma_p))$ is the soft thresholding operator applied elementwise to the diagonal of Σ. Using a cross-validation procedure we find that $\lambda \approx 550$ minimizes the objective function as shown in Figure 1. Doing 10 replications using this value of λ converges in about $500 - 700$ iterations and gives us an average Frobenius norm loss of

$$F = \sum_{r=1}^{10} \frac{\|\hat{X}_r - X_r\|_F}{\|X_r\|_F} = 0.9858716.$$

4. We want to solve the following problem for $i = 1, ..., p$:

$$\begin{aligned} &\min \|\beta^1\|_1 + \|\beta^2\|_1 \\ &\text{subject to } \|S\beta^1 - e_i\|_\infty \leq t_1, \|S\beta^2 - e_i\|_\infty \leq t_2 \\ &\text{and } \|\beta^2 - \beta^1\|_\infty \leq t_3 \end{aligned}$$

Note that this is equivalent to

$$\begin{aligned} &\min \|\beta^1\|_1 + \|\beta^2\|_1 \\ &\text{subject to } -t_1 1 \leq S\beta^1 - e_i \leq t_1 1 \\ &\text{and } -t_2 1 \leq S\beta^2 - e_i \leq t_2 1 \\ &\text{and } -t_3 1 \leq \beta^2 - \beta^1 \leq t_3 1 \end{aligned}$$

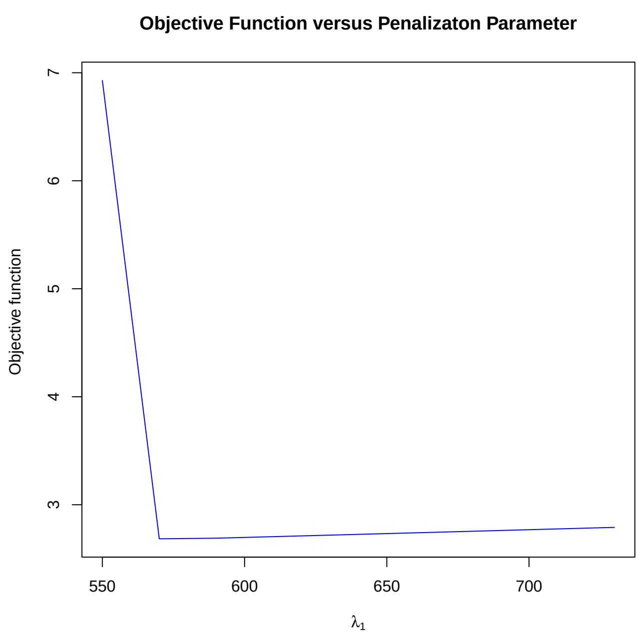

Figure 1: Value of objective function for predicted X_λ versus λ for trace norm constrained optimization problem

Now reparametrize by letting $u^1 = |\beta^1|_1$ and $u^2 = |\beta^2|_1$. Then the problem is

$$
\begin{aligned}
&\min \sum_{j=1}^{p} u_j^1 + \sum_{j=1}^{p} u_j^2 \\
&\text{subject to } -\beta^1 \leq u^1 \leq \beta^1, -\beta^2 \leq u^2 \leq \beta^2 \\
&\text{ and } -t_1 1 \leq S\beta^1 - e_i \leq t_1 1 \\
&\text{and } -t_2 1 \leq S\beta^2 - e_i \leq t_2 1 \\
&\text{ and } -t_3 1 \leq \beta^2 - \beta^1 \leq t_3 1
\end{aligned}
$$

or equivalently,

$$
\begin{aligned}
&\min \sum_{j=1}^{p} u_j^1 + \sum_{j=1}^{p} u_j^2 \\
&\text{subject to } 0 \leq 2u^1 + (\beta^1 - u^1) \leq 2\beta^1, 0 \leq 2u^2 + (\beta^2 - u^2) \leq 2\beta^2 \\
&\text{ and } S(\beta^1 + u^1) - Su^1 \leq e_i + t_1 1 \\
&\text{ and } -(S(\beta^1 + u^1) - Su^1) \leq t_1 1 - e_i \\
&\text{ and } S(\beta^2 + u^2) - Su^2 \leq e_i + t_2 1 \\
&\text{ and } -(S(\beta^2 + u^2) - Su^2) \leq t_2 1 - e_i \\
&\text{ and } (\beta^1 - u^1) - (\beta^2 - u^2) + u^1 - u^2 \leq -t_3 1 \\
&\text{ and } (\beta^2 - u^2) - (\beta^1 - u^1) - u^1 + u^2 \leq t_3 1
\end{aligned}
$$

Hence we can write the above problem as a linear programming problem:

$$
\begin{aligned}
&\min c^T x \\
&\text{subject to } Ax \leq b \\
&\text{and } x \geq 0
\end{aligned}
$$

where

$$
x = \begin{pmatrix} u^1 \\ u^2 \\ \beta^1 + u^1 \\ \beta^2 + u^2 \end{pmatrix},
$$

$$
c^T = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix},
$$

$$A = \begin{pmatrix} -2I & 0 & I & 0 \\ 0 & 0 & -I & 0 \\ 0 & -2I & 0 & I \\ 0 & 0 & 0 & -I \\ -S & 0 & S & 0 \\ S & 0 & -S & 0 \\ 0 & -S & 0 & S \\ 0 & S & 0 & -S \\ I & -I & -I & I \\ -I & I & I & -I \end{pmatrix},$$

and

$$b^T = \begin{pmatrix} 0 & 0 & 0 & 0 & t_1 1 + e_i & t_1 1 - e_i & t_2 1 + e_i & t_2 1 - e_i & t_3 1 & -t_3 1 \end{pmatrix}.$$

Then set $\hat{B}^1 = (\hat{B}^1 + \hat{u}^1) - \hat{u}^1$ and $\hat{B}^2 = (\hat{B}^2 + \hat{u}^2) - \hat{u}^2$ and let the solutions to the i-th problem be the i-th columns of $\hat{\Omega}^1$ and $\hat{\Omega}^2$. To use *lpSolve* in R, we need all the variables to be ≥ 0. Thus replace the $Ax \leq b$ with $b - Ax \geq 0$.