Class Title Goes Here
Homework Number

Your Name Goes Here

Date Goes Here

1. Problem 1

1-(c) A matrix $\boldsymbol{W}$ is in the subgradient if it satisfies:

$$\operatorname{trace}\{(\boldsymbol{Y}-\boldsymbol{X})^\top \boldsymbol{M}\} \leq \|\boldsymbol{Y}\|_{2,1} - \|\boldsymbol{X}\|_{2,1}.$$

This can be rewritten as:

$$\sum_{j=1}^{n}(\boldsymbol{Y}_{.,j}-\boldsymbol{X}_{.,j})^\top \boldsymbol{M}_{.,j} \leq \sum_{j=1}^{n}\|\boldsymbol{Y}_{.,j}\|_2 - \sum_{i=1}^{n}\|\boldsymbol{X}_{.,j}\|_2. \tag{1.1}$$

The previous equation tells us that if $\boldsymbol{M}_{.,j}$ for $j=1,\ldots,n$ are in the subgradients of $\|\boldsymbol{X}_{.,j}\|_2$ for $j=1,\ldots,n$ respectively then $\boldsymbol{M}$ is in the subgradient of $\|\boldsymbol{X}\|_{2,1}$. However, this is only a *sufficient* condition. To prove it is also *necessary*, observe that if $\boldsymbol{M}$ is in the subgradient of $\|\boldsymbol{X}\|_{2,1}$ then it satisfies eqn. (1.1). Next, since the equation is supposed to hold for all $\boldsymbol{Y}$, we can take $\boldsymbol{Y}_{.,j}=\boldsymbol{0}$ everywhere except for $j=k$. Then, eqn. (1.1) becomes:

$$(\boldsymbol{Y}_{.,k}-\boldsymbol{X}_{.,k})^\top \boldsymbol{M}_{.,k} \leq \|\boldsymbol{Y}_{.,k}\|_2 - \|\boldsymbol{X}_{.,k}\|_2$$

so that $\boldsymbol{M}_{.,k}$ is in the subgradient of $\|\boldsymbol{X}_{.,k}\|_2$. Since this must hold for $k=1,\ldots,n$, using problem 1(b) we conclude:

$$(\partial\|\boldsymbol{X}\|_{2,1})_{.,j} = \begin{cases} \frac{\boldsymbol{X}_{.,j}}{\|\boldsymbol{X}_{.,j}\|_2} & \text{if } \boldsymbol{X}_{.,j} \neq \boldsymbol{0} \\ \{\boldsymbol{M}_{.,j} : \|\boldsymbol{M}_{.,j}\|_2 \leq 1\} & \text{if } \boldsymbol{X}_{.,j} = \boldsymbol{0} \end{cases},$$

Equivalently:

$$(\partial\|\boldsymbol{X}\|_{2,1})_{i,j} = \begin{cases} \frac{\boldsymbol{X}_{i,j}}{\|\boldsymbol{X}_{.,j}\|_2} & \text{if } \boldsymbol{X}_{.,j} \neq \boldsymbol{0} \\ \{\boldsymbol{M}_{i,j} : \|\boldsymbol{M}_{.,j}\|_2 \leq 1\} & \text{if } \boldsymbol{X}_{.,j} = \boldsymbol{0} \end{cases}.$$

as desired. Note: the homework uses $\boldsymbol{W}$ instead of $\boldsymbol{M}$.

1-(d) If we wish to minimize the objective function we first observe that the function is convex. In fact, it is *strictly convex* because the term $\|\boldsymbol{X} - \boldsymbol{A}\|_F^2$ is strictly convex in $\boldsymbol{A}$ and, as we have shown in 1(b), $\|\boldsymbol{A}\|_{2,1}$ is convex in $\boldsymbol{A}$. Furthermore, the sum of a strictly convex function and a convex function is again a strictly convex function. Therefore, our objective function has a unique solution. Next, observe that the subgradient of our objective function is:

$$-(\boldsymbol{X} - \boldsymbol{A}) + \tau\partial\|\boldsymbol{A}\|_{2,1}.$$

By part 1(c) we can further write the subgradient as:

$$-(\boldsymbol{X} - \boldsymbol{A}) + \tau\boldsymbol{M}$$

where:

$$\boldsymbol{M}_{.,j} = \begin{cases} \frac{\boldsymbol{A}_{i,j}}{\|\boldsymbol{A}_{.,j}\|_2} & \text{if } \boldsymbol{A}_{.,j} \neq \boldsymbol{0} \\ \{\boldsymbol{M}_{i,j} : \|\boldsymbol{M}_{.,j}\|_2 \leq 1\} & \text{if } \boldsymbol{A}_{.,j} = \boldsymbol{0} \end{cases}. \tag{1.2}$$

Recall that a matrix $\boldsymbol{A}^*$ is a minimizer of our objective function if the subgradient of the objective function at $\boldsymbol{A}^*$ contains the matrix $\boldsymbol{0}$. This can be written as:

$$\tau\boldsymbol{M} = (\boldsymbol{X} - \boldsymbol{A}^*) \tag{1.3}$$

for some matrix $\boldsymbol{M}$ defined by eqn. (1.2). Now, since we have already shown the objective function has a unique minimizer, all we must do is plug in the suggested solution for $\boldsymbol{A}^*$ and see if it satisfies eqn. (1.3). Plugging in:

$$\begin{aligned}(\boldsymbol{X} - \boldsymbol{A}^*)_{.,j} &= [\boldsymbol{X} - \boldsymbol{X}\mathcal{S}_\tau(\operatorname{diag}(\boldsymbol{x}))\operatorname{diag}(\boldsymbol{x})^{-1}]_{.,j} \\ &= \begin{cases} \tau\frac{\boldsymbol{X}_{.,j}}{\|\boldsymbol{X}_{.,j}\|_2} & \text{if } \|\boldsymbol{X}_{.,j}\|_2 > \tau \\ \boldsymbol{X}_{.,j} & \text{if } \|\boldsymbol{X}_{.,j}\|_2 \leq \tau \end{cases}.\end{aligned}$$

if we assume $\tau > 0$. Now, observe that:

$$\|\boldsymbol{X}_{.,j}\|_2 \leq \tau \Rightarrow \boldsymbol{A}^*_{.,j} = \boldsymbol{0} \Rightarrow \|\tau\boldsymbol{M}_{.,j}\|_2 \leq \tau$$

subject to $\|\boldsymbol{M}_{.,j}\|_2 \leq 1$. Consequently, we can take $\tau\boldsymbol{M}_{.,j} = \boldsymbol{X}_{.,j}$ whenever $\|\boldsymbol{X}_{.,j}\|_2 \leq \tau$. In the other case, we have:

$$\|\boldsymbol{X}_{.,j}\|_2 > \tau \Rightarrow \boldsymbol{A}^*_{.,j} \neq \boldsymbol{0} \Rightarrow \tau\boldsymbol{M}_{.,j} = \tau\frac{\boldsymbol{X}_{.,j}}{\|\boldsymbol{X}_{.,j}\|_2},$$

this shows that if we take:

$$\tau\boldsymbol{M}_{.,j} = \begin{cases} \tau\frac{\boldsymbol{X}_{.,j}}{\|\boldsymbol{X}_{.,j}\|_2} & \text{if } \|\boldsymbol{X}_{.,j}\|_2 > \tau \\ \boldsymbol{X}_{.,j} & \text{if } \|\boldsymbol{X}_{.,j}\|_2 \leq \tau \end{cases},$$

then eqn. (1.3) is satisfied. Therefore, $\boldsymbol{A}^* = \boldsymbol{X}\mathcal{S}_\tau(\operatorname{diag}(\boldsymbol{x}))\operatorname{diag}(\boldsymbol{x})^{-1}$ is the unique solution to our optimization problem.