CSC4008 Assignment 2

CHEN Ang (118010009)

2.1

Recall from the lectures that if $p \geq 1$, $||\cdot||_p$ is a norm. Suppose given vectors $\boldsymbol{x}, \boldsymbol{y}$, one has for all $\lambda \in [0,1]$ that

$$\begin{aligned} ||\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}||_p &\leq ||\lambda \boldsymbol{x}||_p + ||(1-\lambda)\boldsymbol{y}||_p \quad &\text{(Triangle Inequality)} \\ &= \lambda||\boldsymbol{x}||_p + (1-\lambda)||\boldsymbol{y}||_p \quad &\text{(Scalability)} \end{aligned}$$

If $0 < p < 1$, consider $\boldsymbol{x} = [0;1]$, $\boldsymbol{y} = [1;0]$ for $\mathbb{R}^2$, and $\boldsymbol{x} = [0;1;\boldsymbol{0}]$ and $\boldsymbol{y} = [1;0;\boldsymbol{0}]$ for higher dimensional spaces where $\boldsymbol{0}$ denotes zero vectors of suitable lengths. Then with $\lambda = 1/2$,

$$||\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}||_p = 2^{1/p-1} > 1 = \lambda||\boldsymbol{x}||_p + (1-\lambda)||\boldsymbol{y}||_p$$

2.6

Denote by s and k the spark and krank of $\boldsymbol{A}$, respectively. By definition of krank, k is the maximum number of columns one can arbitrary choose from $\boldsymbol{A}$ such that these columns are always linearly independent. Thus if one gets to choose one more column, namely $k+1$ columns, the requirement of krank would fail as there exists a set of $k+1$ columns from $\boldsymbol{A}$ that are linear dependent. In other words, there exists $\boldsymbol{x} \neq \boldsymbol{0}$ with $||\boldsymbol{x}||_0 = k+1$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$. Hence we have $s \leq k+1$. Further, because any k-subset of columns of $\boldsymbol{A}$ is always linearly independent, there exists no $\boldsymbol{y} \neq \boldsymbol{0}$ with $||\boldsymbol{y}||_0 = k$ such that $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0}$. Hence $s > k$, and we conclude that $s = k+1$.

2.10

Consider the SVD of matrix $\boldsymbol{A}$ with full a row rank:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top = \boldsymbol{U}[\boldsymbol{\Sigma}_1 \ \ \boldsymbol{O}] \begin{bmatrix} \boldsymbol{V}_1^\top \\ \boldsymbol{V}_2^\top \end{bmatrix} = \boldsymbol{U}\boldsymbol{\Sigma}_1\boldsymbol{V}_1^\top$$

The solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ with the minimum 2-norm is given by

$$\boldsymbol{x}_m = \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{y} = \boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1}\boldsymbol{U}^\top\boldsymbol{y}$$

From last assignment, we know

$$\nabla f(\boldsymbol{x}) = 2(\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x} - \boldsymbol{A}^\top\boldsymbol{y})$$

Writing $\beta := 2\alpha$, GD generates the sequence

$$
\begin{aligned}
\boldsymbol{x}_0 &= \mathbf{0} \\
\boldsymbol{x}_k &= \boldsymbol{B}\boldsymbol{x}_{k-1} + \boldsymbol{c} \\
&= \boldsymbol{B}^2\boldsymbol{x}_{k-2} + \boldsymbol{B}\boldsymbol{c} + \boldsymbol{c} \\
&\vdots \\
&= \boldsymbol{B}^k\boldsymbol{x}_0 + \sum_{i=0}^{k-1}\boldsymbol{B}^i\boldsymbol{c} \\
&= \sum_{i=0}^{k-1}\boldsymbol{B}^i \cdot \boldsymbol{c}
\end{aligned}
$$

where $\boldsymbol{B} := \boldsymbol{I} - \beta\boldsymbol{A}^\top\boldsymbol{A}$ and $\boldsymbol{c} := \beta\boldsymbol{A}^\top\boldsymbol{y}$. Notice that we cannot directly take $k \to \infty$ as

$$
\boldsymbol{I} - \boldsymbol{B} = \beta\boldsymbol{A}^\top\boldsymbol{A}
$$

is singular. However, using SVD

$$
\begin{aligned}
\boldsymbol{x}_k &= \beta\sum_{i=0}^{k-1}(\boldsymbol{I} - \beta\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\boldsymbol{V}^\top)^i \cdot \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{y} \\
&= \beta\sum_{i=0}^{k-1}(\boldsymbol{V}\boldsymbol{V}^\top - \beta\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}\boldsymbol{V}^\top)^i \cdot \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{y} \\
&= \beta\sum_{i=0}^{k-1}(\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^\top)^i \cdot \boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{y} \\
&= \beta\sum_{i=0}^{k-1}\boldsymbol{V}\boldsymbol{D}^i\boldsymbol{V}^\top\boldsymbol{V}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{y} \\
&= \beta\sum_{i=0}^{k-1}\boldsymbol{V}\boldsymbol{D}^i\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{y}
\end{aligned}
$$

where $\boldsymbol{D} := \boldsymbol{I} - \beta\boldsymbol{\Sigma}^\top\boldsymbol{\Sigma}$. Multiplying $\boldsymbol{V}^\top$ on both sides, we obtain

$$
\begin{aligned}
\tilde{\boldsymbol{x}}_k := \boldsymbol{V}^\top\boldsymbol{x}_k &= \beta\sum_{i=0}^{k-1}\boldsymbol{D}^i \cdot \boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{y} \\
&= \beta\begin{bmatrix} \sum_{i=0}^{k-1}(\boldsymbol{I} - \beta\boldsymbol{\Sigma}_1^2)^i & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} \end{bmatrix}\boldsymbol{\Sigma}^\top\boldsymbol{U}^\top\boldsymbol{y}
\end{aligned}
$$

If the step size is chosen such that for all singular value σ of $\boldsymbol{A}$,

$$
|1 - \beta\sigma^2| < 1 \iff \beta \in (0, 2\sigma^{-2}) \iff \alpha \in (0, \sigma^{-2})
$$

then as $k \to \infty$,

$$
\sum_{i=0}^{k-1}(\boldsymbol{I} - \beta\boldsymbol{\Sigma}_1^2)^i \to \beta^{-1}\boldsymbol{\Sigma}_1^{-2}
$$

and so

$$
\tilde{\boldsymbol{x}}_k \to \begin{bmatrix} \boldsymbol{\Sigma}_1^{-1} \\ \boldsymbol{O} \end{bmatrix}\boldsymbol{U}^\top\boldsymbol{y}
$$

Multiplying by $\boldsymbol{V}$ back,

$$
\begin{aligned}
\boldsymbol{x}_k \to \boldsymbol{x}_\star &= \boldsymbol{V}\begin{bmatrix} \boldsymbol{\Sigma}_1^{-1} \\ \boldsymbol{O} \end{bmatrix}\boldsymbol{U}^\top\boldsymbol{y} \\
&= \boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1}\boldsymbol{U}^\top\boldsymbol{y} \\
&= \boldsymbol{x}_m
\end{aligned}
$$

and we're done.

2.13

1

"$\implies$":

Suppose $\boldsymbol{z} = \boldsymbol{B}\boldsymbol{x} + \boldsymbol{e}$ has a solution $(\boldsymbol{x}, \boldsymbol{e})$ with $||\boldsymbol{e}||_0 = k$. Then by multiplying $\boldsymbol{A}$ through in $(2.6.10)$,

$$\boldsymbol{A}\boldsymbol{z} = \boldsymbol{A}\boldsymbol{B}\boldsymbol{x} + \boldsymbol{A}\boldsymbol{e} = \boldsymbol{A}\boldsymbol{e}$$

"$\impliedby$":

Suppose $\boldsymbol{A}\boldsymbol{e} = \boldsymbol{A}\boldsymbol{z}$ has a solution $\boldsymbol{e}$ with $||\boldsymbol{e}||_0 = k$. Then

$$\boldsymbol{A}(\boldsymbol{e} - \boldsymbol{z}) = \boldsymbol{0} \iff \boldsymbol{e} - \boldsymbol{z} \in \ker(\boldsymbol{A})$$

It's been given that the rows of $\boldsymbol{A}$ span the left null space of $\boldsymbol{B}$, i.e., $\mathrm{im}(\boldsymbol{A}^\top) = \ker(\boldsymbol{B}^\top)$. We also know by relationship of fundamental subspaces that $\ker(\boldsymbol{A}) = \mathrm{im}(\boldsymbol{A}^\top)^\perp$ and $\mathrm{im}(\boldsymbol{B}) = \ker(\boldsymbol{B}^\top)^\perp$. Hence $\ker(\boldsymbol{A}) = \mathrm{im}(\boldsymbol{B})$, and so

$$\boldsymbol{e} - \boldsymbol{z} \in \mathrm{im}(\boldsymbol{B}) \iff \boldsymbol{e} - \boldsymbol{z} = \boldsymbol{B}\boldsymbol{x} \iff \boldsymbol{z} = \boldsymbol{B}(-\boldsymbol{x}) + \boldsymbol{e}$$

for some $\boldsymbol{x} \in \mathbb{R}^r$.

2

Note that **1** also holds for L1 norm. (Simply replace $||\boldsymbol{e}||_0$ by $||\boldsymbol{e}||_1$ in the proof).

Let $\hat{\boldsymbol{x}}$ be a solution to $\min_{\boldsymbol{x}} ||\boldsymbol{B}\boldsymbol{x} - \boldsymbol{z}||_1$. Then $\hat{\boldsymbol{e}} = \boldsymbol{z} - \boldsymbol{B}\hat{\boldsymbol{x}}$ satisfies

$$\boldsymbol{A}\hat{\boldsymbol{e}} = \boldsymbol{A}\boldsymbol{z} - \boldsymbol{A}\boldsymbol{B}\hat{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{z}$$

Moreover, $\hat{\boldsymbol{e}}$ is optimal to $(2.6.13)$. To see that, assume there exists some $\boldsymbol{e} \neq \hat{\boldsymbol{e}}$ such that $\boldsymbol{A}\boldsymbol{e} = \boldsymbol{A}\boldsymbol{z}$ and $||\boldsymbol{e}||_1 < ||\hat{\boldsymbol{e}}||_1$. By "$\impliedby$", there must also exists $\boldsymbol{x}'$ such that $\boldsymbol{z} = \boldsymbol{B}\boldsymbol{x}' + \boldsymbol{e}$. But then

$$\min_{\boldsymbol{x}} ||\boldsymbol{B}\boldsymbol{x} - \boldsymbol{z}||_1 \leq ||\boldsymbol{B}\boldsymbol{x}' - \boldsymbol{z}||_1 = ||-\boldsymbol{e}||_1 = ||\boldsymbol{e}||_1 < ||\hat{\boldsymbol{e}}||_1 = ||\boldsymbol{B}\hat{\boldsymbol{x}} - \boldsymbol{z}||_1$$

contradicting the optimality of $\hat{\boldsymbol{x}}$.

Now suppose $\hat{\boldsymbol{e}}$ is a solution to $(2.6.13)$. Due to the constraint it is also a solution to equation $(2.6.11)$. By "$\impliedby$", there exists $\boldsymbol{x}'$ such that $\boldsymbol{z} = \boldsymbol{B}\boldsymbol{x}' + \hat{\boldsymbol{e}}$. We claim that $\boldsymbol{x}'$ is a solution to $(2.6.12)$ with $\hat{\boldsymbol{e}} = \boldsymbol{z} - \boldsymbol{B}\boldsymbol{x}'$.

By way of contradiction assume there exists $\boldsymbol{x}_* \neq \boldsymbol{x}'$ such that $k := ||\boldsymbol{B}\boldsymbol{x}_* - \boldsymbol{z}||_1 < ||\boldsymbol{B}\boldsymbol{x}' - \boldsymbol{z}||_1$. Clearly, $(\boldsymbol{x}_*, \boldsymbol{z} - \boldsymbol{B}\boldsymbol{x}_*)$ is a solution to $(2.6.10)$ with $||\boldsymbol{z} - \boldsymbol{B}\boldsymbol{x}_*||_1 = k$. Thus by "$\implies$", $\boldsymbol{z} - \boldsymbol{B}\boldsymbol{x}_*$ is also a solution to $(2.6.11)$ with $||\boldsymbol{z} - \boldsymbol{B}\boldsymbol{x}_*||_1 = k$. But then

$$\min_{\boldsymbol{e}} ||\boldsymbol{e}||_1 \leq ||\boldsymbol{z} - \boldsymbol{B}\boldsymbol{x}_*||_1 = k < ||\boldsymbol{B}\boldsymbol{x}' - \boldsymbol{z}||_1 = ||-\hat{\boldsymbol{e}}||_1 = ||\hat{\boldsymbol{e}}||_1$$

contradicting the optimality of $\hat{\boldsymbol{e}}$.