DEVIN BRESSER

CS/ECE/ME532 Assignment 5

1. Here we continue the problem studied in Activity 11. Let a 4-by-2 matrix $\boldsymbol{X}$ have SVD $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ where $\boldsymbol{U} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$, $\boldsymbol{S} = \begin{bmatrix} 1 & 0 \\ 0 & \gamma \end{bmatrix}$, and $\boldsymbol{V} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

Let $\boldsymbol{y} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$.

a) The ratio of the largest to the smallest singular values is termed the condition number of $\boldsymbol{X}$. Find the condition number if $\gamma = 0.1$, and $\gamma = 10^{-8}$. Solve $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$ for $\boldsymbol{w}$ and find $||\boldsymbol{w}||_2^2$ for these two values of γ.

b) A system of linear equations with a large condition number is said to be "ill-conditioned". One consequence of an ill-conditioned system of equations is solutions with large norms as you found in the previous part of this problem. A second consequence is that the solution is very sensitive to small errors in $\boldsymbol{y}$ such as may result from measurement error or numerical error. Suppose $\boldsymbol{y} = \begin{bmatrix} 1+\epsilon \\ 0 \\ 0 \\ 1 \end{bmatrix}$. Write $\boldsymbol{w} = \boldsymbol{w}_o + \boldsymbol{w}_\epsilon$ where $\boldsymbol{w}_o$ is the solution for arbitrary γ when $\epsilon = 0$ and $\boldsymbol{w}_\epsilon$ is the perturbation in that solution due to some error $\epsilon \neq 0$. How does the norm of the perturbation due to $\epsilon \neq 0$, $||\boldsymbol{w}_\epsilon||_2^2$, depend on the condition number? Find $||\boldsymbol{w}_\epsilon||_2^2$ for $\epsilon = 0.01$ and $\gamma = 0.1$ and $\gamma = 10^{-8}$.

c) Now consider a "low-rank" inverse. Instead of writing

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \sum_{i=1}^{p} \frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^T$$

where p is the number of columns of $\boldsymbol{X}$ (assumed less than the number of rows), we approximate

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \approx \sum_{i=1}^{r} \frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^T$$

In this approximation we only invert the largest r singular values, and ignore all of them smaller than σ_r. Use $r = 1$ in the low-rank inverse to find $\boldsymbol{w} = \boldsymbol{w}_o + \boldsymbol{w}_\epsilon$

where $\boldsymbol{y} = \begin{bmatrix} 1+\epsilon \\ 0 \\ 0 \\ 1 \end{bmatrix}$ as in part b). Compare the results to part b).

DEVIN BRESSER

1. Here we continue the problem studied in Activity 11. Let a 4-by-2 matrix $\boldsymbol{X}$ have SVD $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T$ where $\boldsymbol{U} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$, $\boldsymbol{S} = \begin{bmatrix} 1 & 0 \\ 0 & \gamma \end{bmatrix}$, and $\boldsymbol{V} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$

Let $\boldsymbol{y} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$.

a) The ratio of the largest to the smallest singular values is termed the condition number of $\boldsymbol{X}$. Find the condition number if $\gamma = 0.1$, and $\gamma = 10^{-8}$. Solve $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$ for $\boldsymbol{w}$ and find $||\boldsymbol{w}||_2^2$ for these two values of γ.

$$\text{Condition \#} = \frac{\delta_{max}}{\delta_{min}}. \quad \text{If } \lambda = 0.1,\ C\# = 1/0.1 = 10$$

$$\text{If } \lambda = 10^{-8},\ C\# = 1/10^{-8} = 1 \cdot 10^{8}$$

$\underline{w}_{min} = \underline{V}\,\underline{\Sigma}^{-1}\,\underline{U}^T\underline{y}$ as shown in my solution for activity 11.

calculator

$$\underset{2\times1}{\underline{w}_{min}} = \underset{2\times2}{1/\sqrt{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}} \underset{2\times2}{\begin{bmatrix} 1 & 0 \\ 0 & 1/\lambda \end{bmatrix}} \underset{2\times4}{1/2\begin{bmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}} \underset{4\times1}{\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}} = \begin{bmatrix} 1/\sqrt{2} + 1/\sqrt{2}\lambda \\ 1/\sqrt{2} - 1/\sqrt{2}\lambda \end{bmatrix}$$

$\lambda = 0.1$:

$$\begin{bmatrix} 1/\sqrt{2} + 1/\sqrt{2}\cdot 0.1 \\ 1/\sqrt{2} - 1/\sqrt{2}\cdot 0.1 \end{bmatrix} = \begin{bmatrix} 7.78 \\ -6.36 \end{bmatrix}, \quad ||w_{min}||_2^2 = 101.$$

$\lambda = 10^{-8}$:

$$\begin{bmatrix} 1/\sqrt{2} + 1/\sqrt{2}\cdot 10^{-8} \\ 1/\sqrt{2} - 1/\sqrt{2}\cdot 10^{-8} \end{bmatrix} = \begin{bmatrix} 7.071\cdot 10^{7} \\ -7.071\cdot 10^{7} \end{bmatrix}, \quad ||w_{min}||_2^2 = 1 \cdot 10^{16}$$

b) A system of linear equations with a large condition number is said to be "ill-conditioned". One consequence of an ill-conditioned system of equations is solutions with large norms as you found in the previous part of this problem. A second consequence is that the solution is very sensitive to small errors in $\boldsymbol{y}$ such as may result from measurement error or numerical error. Suppose $\boldsymbol{y} = \begin{bmatrix} 1+\epsilon \\ 0 \\ 0 \\ 1 \end{bmatrix}$. Write $\boldsymbol{w} = \boldsymbol{w}_o + \boldsymbol{w}_\epsilon$ where $\boldsymbol{w}_o$ is the solution for arbitrary γ when $\epsilon = 0$ and $\boldsymbol{w}_\epsilon$ is the perturbation in that solution due to some error $\epsilon \neq 0$. How does the norm of the perturbation due to $\epsilon \neq 0$, $||\boldsymbol{w}_\epsilon||_2^2$, depend on the condition number? Find $||\boldsymbol{w}_\epsilon||_2^2$ for $\epsilon = 0.01$ and $\gamma = 0.1$ and $\gamma = 10^{-8}$.

b.)

$$\tilde{y} = \begin{bmatrix} 1+\epsilon \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad \underline{w}_{min} = \begin{bmatrix} \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}\lambda} \\ \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}\lambda} \end{bmatrix}$$

$$\underset{2\times2}{\underline{V}} \cdot \underset{2\times2}{\underline{\Sigma}^{-1}} \cdot \underset{2\times4}{\underline{U}^T} = \frac{1}{2\sqrt{2}} \begin{bmatrix} 1+\frac{1}{\lambda} & 1-\frac{1}{\lambda} & 1-\frac{1}{\lambda} & 1+\frac{1}{\lambda} \\ 1-\frac{1}{\lambda} & 1+\frac{1}{\lambda} & 1+\frac{1}{\lambda} & 1-\frac{1}{\lambda} \end{bmatrix}$$

$$\Rightarrow \underline{V}\,\underline{\Sigma}^{-1}\,\underline{U}^T \tilde{y} = \frac{1}{2\sqrt{2}} \underset{2\times4}{\begin{bmatrix} 1+\frac{1}{\lambda} & 1-\frac{1}{\lambda} & 1-\frac{1}{\lambda} & 1+\frac{1}{\lambda} \\ 1-\frac{1}{\lambda} & 1+\frac{1}{\lambda} & 1+\frac{1}{\lambda} & 1-\frac{1}{\lambda} \end{bmatrix}} \underset{4\times1}{\begin{bmatrix} 1+\epsilon \\ 0 \\ 0 \\ 1 \end{bmatrix}}$$

$$= \frac{1}{2\sqrt{2}} \underset{2\times1}{\begin{bmatrix} (1+\frac{1}{\lambda})(1+\epsilon) + (1+\frac{1}{\lambda}) \\ (1-\frac{1}{\lambda})(1+\epsilon) + (1-\frac{1}{\lambda}) \end{bmatrix}} \leftarrow \underline{w}$$

$$\underline{w} = \frac{1}{2\sqrt{2}} \left(\underbrace{\begin{bmatrix} 1+\frac{1}{\lambda} \\ 1-\frac{1}{\lambda} \end{bmatrix}}_{\uparrow\ \underline{w}_o} + \underbrace{\begin{bmatrix} (1+\frac{1}{\lambda})(1+\epsilon) \\ (1-\frac{1}{\lambda})(1+\epsilon) \end{bmatrix}}_{\uparrow\ \underline{w}_\epsilon} \right)$$

b.) continued:

$$\underline{w} = \frac{1}{2\sqrt{2}}\left(\underbrace{\begin{bmatrix} 1+1/\lambda \\ 1-1/\lambda \end{bmatrix}}_{\underline{w}_0} + \underbrace{\begin{bmatrix} (1+1/\lambda)(1+\epsilon) \\ (1-1/\lambda)(1+\epsilon) \end{bmatrix}}_{\underline{w}_\epsilon} \right)$$

$\epsilon = 0.01$, $\lambda = 0.1$:

$$\|\underline{w}_\epsilon\|_2^2 = \left\| \frac{1}{2\sqrt{2}} \begin{bmatrix} (1+1/0.1)(1+0.01) \\ (1-1/0.1)(1+0.01) \end{bmatrix} \right\|_2^2$$

$\rightarrow$ calculator $=$ 25.76.

$\epsilon = 0.01$, $\lambda = 10^{-8}$:

$$\|\underline{w}_\epsilon\|_2^2 = \left\| \frac{1}{2\sqrt{2}} \begin{bmatrix} (1+1/10^{-8})(1+0.01) \\ (1-1/10^{-8})(1+0.01) \end{bmatrix} \right\|_2^2$$

$\rightarrow$ calculator $=$ $2.55 \cdot 10^{15}$

As the condition # increases, $\|\underline{w}_\epsilon\|_2^2$ increases.

c) Now consider a "low-rank" inverse. Instead of writing

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T = \sum_{i=1}^{p} \frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^T$$

where p is the number of columns of $\boldsymbol{X}$ (assumed less than the number of rows), we approximate

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T \approx \sum_{i=1}^{r} \frac{1}{\sigma_i}\boldsymbol{v}_i\boldsymbol{u}_i^T$$

In this approximation we only invert the largest r singular values, and ignore all of them smaller than σ_r. Use $r = 1$ in the low-rank inverse to find $\boldsymbol{w} = \boldsymbol{w}_o + \boldsymbol{w}_\epsilon$ where $\boldsymbol{y} = \begin{bmatrix} 1+\epsilon \\ 0 \\ 0 \\ 1 \end{bmatrix}$ as in part b). Compare the results to part b).

In this case: $$(\underline{X}^T\underline{X})^{-1}\underline{X}^T \approx \sum_{i=1}^{1} \frac{1}{\delta_i}\underline{v}_i\underline{u}_i^T = \frac{1}{1}\cdot \frac{1}{\sqrt{2}}\underbrace{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}_{2\times 1}\frac{1}{2}\underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}}_{1\times 4} = \frac{1}{2\sqrt{2}}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

Now, $\underline{w}_{min} = (\underline{X}^T\underline{X})^{-1}\underline{X}^T y$

$$\frac{1}{2\sqrt{2}}\underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}}_{2\times 4}\underbrace{\begin{bmatrix} 1+\epsilon \\ 0 \\ 0 \\ 1 \end{bmatrix}}_{4\times 1} = \frac{1}{2\sqrt{2}}\begin{bmatrix} 1+\epsilon+1 \\ 1+\epsilon+1 \end{bmatrix} = \begin{bmatrix} 2+\epsilon \\ 2+\epsilon \end{bmatrix}$$

So, $$\underline{w} = \frac{1}{2\sqrt{2}}\begin{bmatrix} 2+\epsilon \\ 2+\epsilon \end{bmatrix} = \frac{1}{2\sqrt{2}}\left(\underbrace{\begin{bmatrix} 2 \\ 2 \end{bmatrix}}_{\underline{w}_o} + \underbrace{\begin{bmatrix} \epsilon \\ \epsilon \end{bmatrix}}_{\underline{w}_\epsilon} \right)$$

In this case, because we do not consider the singular value associated with λ, the expressions for $\underline{w}$, $\underline{w}_o$, and $\underline{w}_\epsilon$ are greatly simplified.

$$\underline{w} = \frac{1}{2\sqrt{2}}\left(\underbrace{\begin{bmatrix} 1+1/\lambda \\ 1-1/\lambda \end{bmatrix}}_{\underline{w}_o} + \underbrace{\begin{bmatrix} (1+1/\lambda)(1+\epsilon) \\ (1-1/\lambda)(1+\epsilon) \end{bmatrix}}_{\underline{w}_\epsilon} \right) \rightarrow \frac{1}{2\sqrt{2}}\left(\underbrace{\begin{bmatrix} 2 \\ 2 \end{bmatrix}}_{\underline{w}_o} + \underbrace{\begin{bmatrix} \epsilon \\ \epsilon \end{bmatrix}}_{\underline{w}_\epsilon} \right)$$

And, we are no longer troubled by very large $\underline{w}$ when λ is very small.