Assignment 5

1. Here we continue the problem studied in Activity 11. Let a 4-by-2 matrix $\boldsymbol{X}$ have SVD $\boldsymbol{X} = \boldsymbol{USV}^T$ 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 $\boldsymbol{Xw} = \boldsymbol{y}$ for $\boldsymbol{w}$ and find $||\boldsymbol{w}||_2^2$ for these two values of γ.

$\gamma = 0.1$: $S = \begin{bmatrix} 1 & 0 \\ 0 & 0.1 \end{bmatrix}$

The diagonal values in the S matrix are the singular values.

Condition Number $= \frac{1}{0.1} = 10$

Solve $Xw = y \rightarrow w = VS^{-1}U^Ty$

$S^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix}$

$w = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix}\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

$= \frac{1}{2\sqrt{2}}\begin{bmatrix} 1 & 10 \\ 1 & -10 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

$= \frac{1}{2\sqrt{2}}\begin{bmatrix} 11 & -9 & -9 & 11 \\ -9 & 11 & 11 & -9 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

$w = \frac{1}{2\sqrt{2}}\begin{bmatrix} 22 \\ -18 \end{bmatrix}$

$||w||_2^2 = \left(\frac{1}{2\sqrt{2}}(22)\right)^2 + \left(\frac{1}{2\sqrt{2}}(-18)\right)^2$

$= \frac{11^2}{2} + \frac{(-9)^2}{2}$

$= \frac{121}{2} + \frac{81}{2} = \frac{202}{2} = 101$

$\gamma = 10^{-8}$: $S = \begin{bmatrix} 1 & 0 \\ 0 & 10^{-8} \end{bmatrix}$

Condition number $= \frac{1}{10^{-8}} = 10^{8}$

Solve for $Xw = y \rightarrow w = V S^{-1} U^T y$

$S^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 10^{8} \end{bmatrix}$

$$w = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 10^{8} \end{bmatrix} \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$= \frac{1}{2\sqrt{2}} \begin{bmatrix} 1 & 10^{8} \\ 1 & -10^{8} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$= \frac{1}{2\sqrt{2}} \begin{bmatrix} 1+10^{8} & 1-10^{8} & 1-10^{8} & 1+10^{8} \\ 1-10^{8} & 1+10^{8} & 1+10^{8} & 1-10^{8} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$w = \frac{1}{2\sqrt{2}} \begin{bmatrix} 2 + 2(10^{8}) \\ 2 - 2(10^{8}) \end{bmatrix}$$

$$\begin{aligned} \|w\|_2^2 &= \left(\tfrac{1}{2\sqrt{2}}(2+2(10^{8}))\right)^2 + \left(\tfrac{1}{2\sqrt{2}}(2-2(10^{8}))\right)^2 \\ &= \tfrac{1}{8}(2+2(10^{8}))^2 + \tfrac{1}{8}(2-2(10^{8}))^2 \\ &= \tfrac{1}{8}(4 + 8(10^{8}) + 4(10^{16})) \\ &\quad + \tfrac{1}{8}(4 - 8(10^{8}) + 4(10^{16})) \\ &= \tfrac{1}{8}(8 + 8(10^{16})) \\ &= 1 + 10^{16} \end{aligned}$$

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 $\gamma = 10^{-8}$.

$$w_o = V S^{-1} U^T \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{2\sqrt{2}} \begin{bmatrix} 2 + \frac{2}{\gamma} \\ 2 - \frac{2}{\gamma} \end{bmatrix}$$

$$w_\epsilon = V S^{-1} U^T \begin{bmatrix} \epsilon \\ 0 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{2\sqrt{2}} \begin{bmatrix} \epsilon + \frac{\epsilon}{\gamma} \\ \epsilon - \frac{\epsilon}{\gamma} \end{bmatrix}$$

$$w = \frac{1}{2\sqrt{2}} \begin{bmatrix} 2 + \epsilon + \frac{2}{\gamma} + \frac{\epsilon}{\gamma} \\ 2 + \epsilon - \frac{2}{\gamma} - \frac{\epsilon}{\gamma} \end{bmatrix}$$

The norm of the perturbation due to error will be much greater if the condition number is large and the system of linear equations is "ill-conditioned".

The solution weights norm is blown up if the condition number is large, so the corresponding w_ε norm will also be large.

$\gamma = 0.1$: $w_0 = \frac{1}{2\sqrt{2}}\begin{bmatrix} 22 \\ -18 \end{bmatrix}$

$$w_\varepsilon = \frac{1}{2\sqrt{2}}\begin{bmatrix} 11 & -9 & -9 & 11 \\ -9 & 11 & 11 & -9 \end{bmatrix}\begin{bmatrix} \varepsilon \\ 0 \\ 0 \\ 0 \end{bmatrix} = \frac{1}{2\sqrt{2}}\begin{bmatrix} 11\varepsilon \\ -9\varepsilon \end{bmatrix}$$

$$w = w_0 + w_\varepsilon = \frac{1}{2\sqrt{2}}\begin{bmatrix} 22 \\ -18 \end{bmatrix} + \frac{1}{2\sqrt{2}}\begin{bmatrix} 11\varepsilon \\ -9\varepsilon \end{bmatrix}$$
$$= \frac{1}{2\sqrt{2}}\begin{bmatrix} 22 + 11\varepsilon \\ -18 - 9\varepsilon \end{bmatrix}$$

$$\|w_\varepsilon\|_2^2 = \left(\frac{1}{2\sqrt{2}}(11\varepsilon)\right)^2 + \left(\frac{1}{2\sqrt{2}}(-9\varepsilon)\right)^2$$
$$= \frac{121}{8}\varepsilon^2 + \frac{81}{8}\varepsilon^2 = \frac{202}{8}\varepsilon^2$$

when $\varepsilon = 0.01$: $w_\varepsilon = \frac{1}{2\sqrt{2}}\begin{bmatrix} 0.11 \\ -0.09 \end{bmatrix}$ $\|w_\varepsilon\|_2^2 = \frac{202}{8}(0.01)^2$
$= 0.002525$

$\gamma = 10^8$: $w_0 = \frac{1}{2\sqrt{2}}\begin{bmatrix} 2 + 2(10^8) \\ 2 - 2(10^8) \end{bmatrix}$

$$w_\varepsilon = \frac{1}{2\sqrt{2}}\begin{bmatrix} 1+10^8 & 1-10^8 & 1-10^8 & 1+10^8 \\ 1-10^8 & 1+10^8 & 1+10^8 & 1-10^8 \end{bmatrix}\begin{bmatrix} \varepsilon \\ 0 \\ 0 \\ 0 \end{bmatrix}$$
$$= \frac{1}{2\sqrt{2}}\begin{bmatrix} \varepsilon + 10^8\varepsilon \\ \varepsilon - 10^8\varepsilon \end{bmatrix}$$

$$w = w_0 + w_\varepsilon = \frac{1}{2\sqrt{2}}\begin{bmatrix} 2 + 2(10^8) \\ 2 - 2(10^8) \end{bmatrix} + \frac{1}{2\sqrt{2}}\begin{bmatrix} \varepsilon + 10^8\varepsilon \\ \varepsilon - 10^8\varepsilon \end{bmatrix}$$
$$= \frac{1}{2\sqrt{2}}\begin{bmatrix} 2 + \varepsilon + 2(10^8) + \varepsilon(10^8) \\ 2 + \varepsilon - 2(10^8) - \varepsilon(10^8) \end{bmatrix}$$

$$\|w_\varepsilon\|_2^2 = \left(\frac{\varepsilon}{2\sqrt{2}}(1+10^8)\right)^2 + \left(\frac{\varepsilon}{2\sqrt{2}}(1-10^8)\right)^2$$
$$= \left(\frac{\varepsilon^2}{8}(1 + 2(10^8) + 10^{16})\right) + \left(\frac{\varepsilon^2}{8}(1 - 2(10^8) + 10^{16})\right)$$
$$= \frac{\varepsilon^2}{8}(2 + 2(10^{16})) = \frac{\varepsilon^2}{4}(1 + 10^{16})$$

when $\varepsilon = 0.01$: $w_\varepsilon = \frac{1}{2\sqrt{2}}\begin{bmatrix} 0.01 + 10^6 \\ 0.01 - 10^6 \end{bmatrix}$

$$\|w_\varepsilon\|_2^2 = \frac{(0.01)^2}{4}(1 + 10^{16})$$
$$= \frac{1}{4}(0.0001 + 10^{12})$$

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).

$$w_0 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 \end{bmatrix}\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad w_\varepsilon = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 \end{bmatrix}\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} \varepsilon \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$= \frac{1}{2\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad = \frac{1}{2\sqrt{2}}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} \varepsilon \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$= \frac{1}{2\sqrt{2}}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad = \frac{1}{2\sqrt{2}}\begin{bmatrix} \varepsilon \\ \varepsilon \end{bmatrix}$$

$$= \frac{1}{2\sqrt{2}}\begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$w = \frac{1}{2\sqrt{2}}\begin{bmatrix} 2+\varepsilon \\ 2+\varepsilon \end{bmatrix}$$

part b: $w = \frac{1}{2\sqrt{2}}\begin{bmatrix} 2+\varepsilon+\frac{2}{\gamma}+\frac{\varepsilon}{\gamma} \\ 2+\varepsilon-\frac{2}{\gamma}-\frac{\varepsilon}{\gamma} \end{bmatrix}$

The low rank w got rid of the γ terms and instead only uses a constant and the error term. Therefore, if a very small γ causes the matrix to become ill-conditioned, using a low rank approximation will make the matrix less impacted by the error term.