CS/ECE/ME532 Activity 9

Estimated Time: 25 minutes for P1, 30 minutes for P2

1. Consider the system of linear equations $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$ where $\boldsymbol{X} = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix}$, $\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ and $\boldsymbol{y} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$.

a) Sketch the set of all $\boldsymbol{w}$ that satisfy $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$ in the w_1-w_2 plane. Is the solution unique? What is the value of the squared error $\min_{\boldsymbol{w}} ||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$?

b) Use your sketch to find the $\boldsymbol{w}$ of minimum norm that satisfies the system of equations: $\min_{\boldsymbol{w}} ||\boldsymbol{w}||_2^2$ subject to $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$. Is this solution unique? What makes it unique? What is the value of the squared error $||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$ at this solution? What is the value of $||\boldsymbol{w}||^2$? *Hint:* The equation $||\boldsymbol{w}||_2^2 = c$ describes a circle in $\mathbb{R}^2$ with radius $\sqrt{c}$.

c) Algebraically find the $\hat{\boldsymbol{w}}$ that solves the Tikhonov-regularized (or ridge regression) problem $\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \{||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{w}||_2^2\}$ as a function of λ. *Hint:* Recall that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

d) Sketch the set solution to the Tikhonov-regularized problem in the w_1-w_2 plane as a function of λ for $0 < \lambda < \infty$. (Consider the solution for different values of λ in that range.) Find the squared error $||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$ and norm squared of the solution, $||\boldsymbol{w}||_2^2$ for $\lambda = 0$ and $\lambda = 5$. Compare the squared error and norm squared of the solution to those in part **b)**.

2. Let $\boldsymbol{X} = \begin{bmatrix} 1 & \gamma \\ 1 & -\gamma \\ 1 & -\gamma \\ 1 & \gamma \end{bmatrix}$.

a) Show that the columns of $\boldsymbol{X}$ are orthogonal to each other for any γ.

b) Express $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}$ where $\boldsymbol{U}$ is a 4-by-2 matrix with orthonormal columns and $\boldsymbol{\Sigma}$ is a 2-by-2 diagonal matrix (the non-diagonal entries are zero).

c) Express the solution to the least-squares problem $\min_{\boldsymbol{w}} ||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$ as a function of $\boldsymbol{U}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{y}$.

d) Let $\boldsymbol{y} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$. Find the weights $\boldsymbol{w}$ as a function of γ. What happens to $||\boldsymbol{w}||_2^2$ as $\gamma \to 0$?

e) The ratio of the largest to the smallest diagonal values in $\boldsymbol{\Sigma}$ is termed the condition number of $\boldsymbol{X}$. Find the condition number if $\gamma = 0.1$ and $\gamma = 10^{-8}$. Also find $||\boldsymbol{w}||_2^2$ for these two values of γ.

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

g) Now apply ridge regression, i.e., Tikhonov reqularization. Solve for $\boldsymbol{w}_o$ and $\boldsymbol{w}_\epsilon$ as a function of λ. Find $||\boldsymbol{w}_o||_2^2$ and $||\boldsymbol{w}_\epsilon||_2^2$ for $\lambda = 0.1$, $\epsilon = 0.01$ and $\gamma = 0.1$ and $\gamma = 10^{-8}$. Comment on the impact of regularization.

$$|| \underline{X}\, \underline{w} - \underline{y} ||_2^2 + \lambda || \underline{w} ||_2^2$$

532 Activity 9 - DEVIN BRESSER

1. Consider the system of linear equations $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$ where $\boldsymbol{X} = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix}$, $\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ and $\boldsymbol{y} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$.

   a) Sketch the set of all $\boldsymbol{w}$ that satisfy $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$ in the w_1-w_2 plane. Is the solution unique? What is the value of the squared error $\min_{\boldsymbol{w}} ||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$?

   b) Use your sketch to find the $\boldsymbol{w}$ of minimum norm that satisfies the system of equations: $\min_{\boldsymbol{w}} ||\boldsymbol{w}||_2^2$ subject to $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$. Is this solution unique? What makes it unique? What is the value of the squared error $||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$ at this solution? What is the value of $||\boldsymbol{w}||^2$? *Hint:* The equation $||\boldsymbol{w}||_2^2 = c$ describes a circle in $\mathbb{R}^2$ with radius $\sqrt{c}$.

   c) Algebraically find the $\hat{\boldsymbol{w}}$ that solves the Tikhonov-regularized (or ridge regression) problem $\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \{||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{w}||_2^2\}$ as a function of λ. *Hint:* Recall that
   $$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

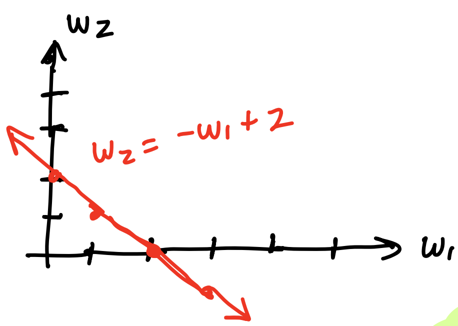
   d) Sketch the set solution to the Tikhonov-regularized problem in the w_1-w_2 plane as a function of λ for $0 < \lambda < \infty$. (Consider the solution for different values of λ in that range.) Find the squared error $||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$ and norm squared of the solution, $||\boldsymbol{w}||_2^2$ for $\lambda = 0$ and $\lambda = 5$. Compare the squared error and norm squared of the solution to those in part **b)**.

a.) $\underline{X}\,\underline{w} = \underline{y} = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix}\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = \begin{bmatrix} 2 \\ -4 \end{bmatrix}$

$\rightarrow \quad w_1 + w_2 = 2 \rightarrow w_2 = 2 - w_1$

$-2w_1 - 2w_2 = -4 \rightarrow 2w_2 = 4 - 2w_1 \rightarrow w_2 = 2 - w_1$

$\rightarrow$



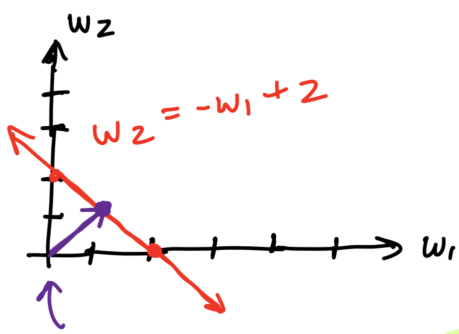
solution is nonunique, all $\begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$ so that $w_2 = 2 - w_1$ satisfy $\underline{X}\,\underline{w} = \underline{y}$

$\min ||\underline{X}\,\underline{w} = \underline{y}||_2^2 = 0$ for each solution

b) Use your sketch to find the $\boldsymbol{w}$ of minimum norm that satisfies the system of equations: $\min_{\boldsymbol{w}} ||\boldsymbol{w}||_2^2$ subject to $\boldsymbol{X}\boldsymbol{w} = \boldsymbol{y}$. Is this solution unique? What makes it unique? What is the value of the squared error $||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$ at this solution? What is the value of $||\boldsymbol{w}||^2$? *Hint:* The equation $||\boldsymbol{w}||_2^2 = c$ describes a circle in $\mathbb{R}^2$ with radius $\sqrt{c}$.

c) Algebraically find the $\hat{\boldsymbol{w}}$ that solves the Tikhonov-regularized (or ridge regression) problem $\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \{||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{w}||_2^2\}$ as a function of λ. *Hint:* Recall that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$



$$\min_{w} ||\underline{w}||_2^2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$\underline{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is a unique solution for $\min_{w} ||\underline{w}||_2^2$.

It is the single point along the solution line that is aimed at the origin orthogonal to the line (intersects $\min(\ell_2 \text{ norm})$ circle)

The value of $||\underline{X}\,\underline{w} - \underline{y}||_2^2 = 0$ at this solution, because it is an exact solution.

$$||\underline{w}||_2^2 = \sqrt{(1^2)+(1^2)}^{\,2} = 2$$

c) Algebraically find the $\hat{\boldsymbol{w}}$ that solves the Tikhonov-regularized (or ridge regression) problem $\hat{\boldsymbol{w}} = \arg\min_{\boldsymbol{w}} \{||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{w}||_2^2\}$ as a function of λ. *Hint:* Recall that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

d) Sketch the set solution to the Tikhonov-regularized problem in the w_1-w_2 plane as a function of λ for $0 < \lambda < \infty$. (Consider the solution for different values of λ in that range.) Find the squared error $||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$ and norm squared of the solution, $||\boldsymbol{w}||_2^2$ for $\lambda = 0$ and $\lambda = 5$. Compare the squared error and norm squared of the solution to those in part **b**).

c) $$\underline{\hat{w}} = \arg\min_{\underline{w}} \left\{ \|\underline{X}\,\underline{w} - \underline{y}\|_2^2 + \lambda \|\underline{w}\|_2^2 \right\}$$

(check)

$$\rightarrow \text{fact: } \|\underline{a}\|_2^2 + \|\underline{b}\|_2^2 = \underline{a}^T\underline{a} + \underline{b}^T\underline{b} = \left\| \begin{bmatrix} \underline{a} \\ \underline{b} \end{bmatrix} \right\|_2^2$$

$$\rightarrow \left\| \begin{bmatrix} \underline{X}\,\underline{w} - \underline{y} \\ \lambda^{1/2}\,\underline{w} \end{bmatrix} \right\|_2^2 \rightarrow \min_{\underline{w}} \Bigg\| \underbrace{\begin{bmatrix} \underline{X} \\ \lambda^{1/2}\underline{\underline{I}} \end{bmatrix}}_{\underline{\tilde{X}}} \underline{w} - \underbrace{\begin{bmatrix} \underline{y} \\ 0 \end{bmatrix}}_{\underline{\tilde{y}}} \Bigg\|_2^2$$

$$\rightarrow \min_{\underline{w}} \|\underline{\tilde{X}}\,\underline{w} - \underline{\tilde{y}}\| = \left(\underline{\tilde{X}}^T\underline{\tilde{X}}\right)^{-1}\underline{\tilde{X}}^T\underline{\tilde{y}} =$$

$$= \left(\begin{bmatrix} \underline{X}^T & \lambda^{1/2}\underline{\underline{I}} \end{bmatrix} \begin{bmatrix} \underline{X} \\ \lambda^{1/2}\underline{\underline{I}} \end{bmatrix} \right)^{-1} \begin{bmatrix} \underline{X}^T & \lambda^{1/2}\underline{\underline{I}} \end{bmatrix} \begin{bmatrix} \underline{y} \\ 0 \end{bmatrix}$$

$$\underline{w}_{min} = \left(\underline{X}^T\underline{X} + \lambda\underline{\underline{I}}\right)^{-1}\underline{X}^T\underline{y}$$

$$\underline{w}_{min} = \left(\begin{bmatrix} 5 & 5 \\ 5 & 5 \end{bmatrix} + \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & -2 \\ 1 & -2 \end{bmatrix} \begin{bmatrix} 2 \\ -4 \end{bmatrix}$$

$$= \begin{pmatrix} 5+\lambda & 5 \\ 5 & 5+\lambda \end{pmatrix}^{-1} \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

$$= \frac{1}{(5+\lambda)^2 - 25} \begin{bmatrix} 5+\lambda & -5 \\ -5 & 5+\lambda \end{bmatrix} \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

$$\cancel{25} + 10\lambda + \lambda^2 \cancel{-25}$$

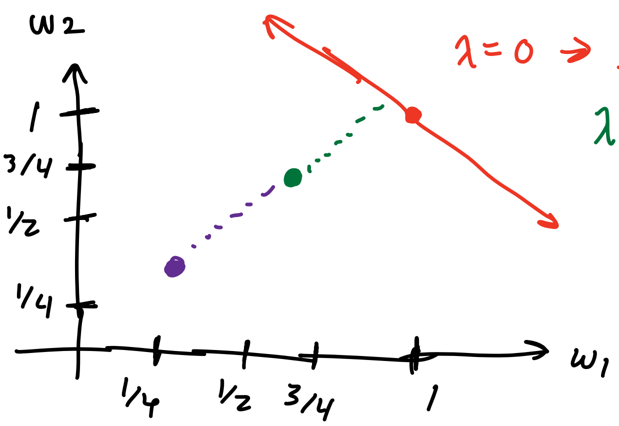
$$= \frac{1}{10\lambda + \lambda^2} \begin{bmatrix} 5+\lambda & -5 \\ -5 & 5+\lambda \end{bmatrix} \begin{bmatrix} 10 \\ 10 \end{bmatrix}$$

$$\frac{10}{10+\lambda} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \frac{1}{10\lambda + \lambda^2} \begin{bmatrix} 10\lambda \\ 10\lambda \end{bmatrix}$$

$$= \frac{10\lambda}{10\lambda + \lambda^2} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{\lambda(10)}{\lambda(10+\lambda)} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{10}{10+\lambda} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

d) Sketch the set solution to the Tikhonov-regularized problem in the w_1-w_2 plane as a function of λ for $0 < \lambda < \infty$. (Consider the solution for different values of λ in that range.) Find the squared error $||\boldsymbol{Xw} - \boldsymbol{y}||_2^2$ and norm squared of the solution, $||\boldsymbol{w}||_2^2$ for $\lambda = 0$ and $\lambda = 5$. Compare the squared error and norm squared of the solution to those in part **b**).



$\lambda = 0 \rightarrow \underline{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$\lambda = 5 \rightarrow \underline{w} = \begin{bmatrix} 2/3 \\ 2/3 \end{bmatrix}$

$\lambda = 20 \rightarrow \underline{w} = \begin{bmatrix} 1/3 \\ 1/3 \end{bmatrix}$

$\underline{\lambda = 0}$:

$||\underline{X}\,\underline{w} - \underline{y}||_2^2 = 0$

$||\underline{w}||_2^2 = 2$

$\underline{\lambda = 5}$:

$||\underline{X}\,\underline{w} - \underline{y}||_2^2 = 2.222$

$||\underline{w}||_2^2 = 0.888$

2. Let $\boldsymbol{X} = \begin{bmatrix} 1 & \gamma \\ 1 & -\gamma \\ 1 & -\gamma \\ 1 & \gamma \end{bmatrix}$.

a) Show that the columns of $\boldsymbol{X}$ are orthogonal to each other for any γ.

b) Express $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}$ where $\boldsymbol{U}$ is a 4-by-2 matrix with orthonormal columns and $\boldsymbol{\Sigma}$ is a 2-by-2 diagonal matrix (the non-diagonal entries are zero).

a.) $\begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \beta \begin{bmatrix} \gamma \\ -\gamma \\ -\gamma \\ \gamma \end{bmatrix} = \beta\gamma - \beta\gamma - \beta\gamma + \beta\gamma = 0$ for all β

So $\underline{x}_1, \underline{x}_2$ are orthogonal

b.) $\underline{X} = \underline{U}\,\underline{\Sigma}$

$$\underbrace{\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \\ 1/2 & -1/2 \\ 1/2 & 1/2 \end{bmatrix}}_{\underline{U}} \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & 2\gamma \end{bmatrix}}_{\underline{\Sigma}} = \underbrace{\begin{bmatrix} 1 & \gamma \\ 1 & -\gamma \\ 1 & -\gamma \\ 1 & -\gamma \end{bmatrix}}_{\underline{X}}$$

c) Express the solution to the least-squares problem $\min_{\boldsymbol{w}} ||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$ as a function of $\boldsymbol{U}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{y}$.

$$\min_{w} ||\underline{X}\,\underline{w} - \underline{y}||_2^2 = \min_{w} ||\underline{U}\,\underline{\Sigma}\,\underline{w} - \underline{y}||_2^2$$

$$= (\underline{X}^T\underline{X})^{-1}\underline{X}^T\underline{y} = ((\underline{U}\,\underline{\Sigma})^T\underline{X})^{-1}(\underline{U}\,\underline{\Sigma})^T\underline{y}$$

$$= (\underline{\Sigma}^T\underline{U}^T\underline{U}\,\underline{\Sigma})^{-1}\underline{\Sigma}^T\underline{U}^T\underline{y}$$

$$= (\underline{\Sigma}^T\underline{\Sigma})^{-1}\underline{\Sigma}^T\underline{U}^T\underline{y}$$

c) Express the solution to the least-squares problem $\min_{\boldsymbol{w}} ||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$ as a function of $\boldsymbol{U}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{y}$.

d) Let $\boldsymbol{y} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$. Find the weights $\boldsymbol{w}$ as a function of γ. What happens to $||\boldsymbol{w}||_2^2$ as $\gamma \to 0$?

d.) $\underline{w}_{min} = (\underline{\Sigma}^T \underline{\Sigma})^{-1} \underline{\Sigma}\, \underline{U}^T \underline{y}$ $\quad \rightarrow (\Sigma^T \Sigma)^{-1} = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/\gamma^2 \end{bmatrix}$

$\rightarrow \underbrace{(\Sigma^T \Sigma)^{-1} \Sigma\, U^T}_{2 \times 4} = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25/\gamma & -0.25/\gamma & -0.25/\gamma & 0.25/\gamma \end{bmatrix}$

$\rightarrow \quad '' \cdot y = \begin{bmatrix} 0.5 \\ 0.5/\gamma \end{bmatrix}$

$\rightarrow \quad \underline{w}\, min = \begin{bmatrix} 0.5 \\ 0.5/\gamma \end{bmatrix}$

$||\underline{w}\, min||_2^2 = \sqrt{(0.5)^2 + (0.5/\gamma)^2}^{\,2} = 0.25 + 0.25/\gamma^2$

$\boxed{\text{As } \gamma \rightarrow 0, \ ||w_{min}||_2^2 \rightarrow \infty}$

e) The ratio of the largest to the smallest diagonal values in $\boldsymbol{\Sigma}$ is termed the condition number of $\boldsymbol{X}$. Find the condition number if $\gamma = 0.1$ and $\gamma = 10^{-8}$. Also find $||\boldsymbol{w}||_2^2$ for these two values of γ.

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

g) Now apply ridge regression, i.e., Tikhonov reqularization. Solve for $\boldsymbol{w}_o$ and $\boldsymbol{w}_\epsilon$ as a function of λ. Find $||\boldsymbol{w}_o||_2^2$ and $||\boldsymbol{w}_\epsilon||_2^2$ for $\lambda = 0.1$, $\epsilon = 0.01$ and $\gamma = 0.1$ and $\gamma = 10^{-8}$. Comment on the impact of regularization.

e.) $\underline{\Sigma} = \begin{bmatrix} 2 & 0 \\ 0 & 2\gamma \end{bmatrix}$

$$\Sigma_{\gamma=0.1} = \begin{bmatrix} 2 & 0 \\ 0 & 0.2 \end{bmatrix} \rightarrow \text{condition \#} \ \ 2/0.2 = 10$$

$$||\underline{w}_{min}||_2^2 \ _{\gamma=0.1} = 0.25 + 0.25/0.1^2 = 25.25$$

$$\Sigma_{\gamma=10^{-8}} = \begin{bmatrix} 2 & 0 \\ 0 & 2\cdot 10^{-8} \end{bmatrix} \rightarrow \text{condition \#} \ \ 2/2\cdot 10^{-8} = 1\cdot 10^{8}$$

$$||\underline{w}_{min}||_2^2 \ _{\gamma=0.1} = 0.25 + 0.25/(10^{-8})^2 = 2.5\cdot 10^{15}$$

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

g) Now apply ridge regression, i.e., Tikhonov reqularization. Solve for $\boldsymbol{w}_o$ and $\boldsymbol{w}_\epsilon$ as a function of λ. Find $||\boldsymbol{w}_o||_2^2$ and $||\boldsymbol{w}_\epsilon||_2^2$ for $\lambda = 0.1$, $\epsilon = 0.01$ and $\gamma = 0.1$ and $\gamma = 10^{-8}$. Comment on the impact of regularization.

$$\underline{y} = \begin{bmatrix} 1+\epsilon \\ 0 \\ 0 \\ 1 \end{bmatrix} \xrightarrow{\text{calculator}} \underline{w} = \begin{bmatrix} 0.5 + 0.25\epsilon \\ \\ \dfrac{0.5+0.25\epsilon}{\gamma} \end{bmatrix} = \underbrace{\begin{bmatrix} 0.5 \\ \\ 0.5/\gamma \end{bmatrix}}_{\underline{w}_o} + \underbrace{\begin{bmatrix} 0.25\epsilon \\ \\ 0.25\epsilon/\gamma \end{bmatrix}}_{\underline{w}_\epsilon}$$

$$||\underline{w}_\epsilon||_2^2 = \sqrt{(0.25\epsilon)^2 + (0.25\epsilon/\gamma)^2}^{\,2} = 0.125\epsilon^2 + 0.125\epsilon^2/\gamma^2$$

As γ increases $>>2$, condition # decreases & $||\underline{w}_\epsilon||_2^2 \to 0.125\epsilon^2$

As γ decreases $<<2$, condition # increases. & $||\underline{w}_\epsilon||_2^2 \to \infty$

$$||w_\epsilon||_2^2{}_{[\epsilon=0.01,\ \gamma=0.1]} = 0.125(0.01)^2 + 0.125(0.01)^2/(0.1)^2$$

$$= 0.001263$$

$$||w_\epsilon||_2^2{}_{[\epsilon=0.01,\ \gamma=10^{-8}]} = \cdots = 1.25 \cdot 10^{11}$$

g) Now apply ridge regression, i.e., Tikhonov reqularization. Solve for $\boldsymbol{w}_o$ and $\boldsymbol{w}_\epsilon$ as a function of λ. Find $||\boldsymbol{w}_o||_2^2$ and $||\boldsymbol{w}_\epsilon||_2^2$ for $\lambda = 0.1$, $\epsilon = 0.01$ and $\gamma = 0.1$ and $\gamma = 10^{-8}$. Comment on the impact of regularization.

$$\underline{w}_{min} = \begin{bmatrix} 0.5 + 0.25\epsilon \\ \frac{0.5+0.25\epsilon}{\gamma} \end{bmatrix} \quad \text{when} \quad \underline{y} = \begin{bmatrix} 1+\epsilon \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\min_{\underline{w}} \quad \| \underline{X}\,\underline{w} - \underline{y} \|_2^2 + \lambda \| \underline{w} \|_2^2$$

$$\rightarrow \underline{w}_{min} = (\underline{X}^T \underline{X} + \lambda \underline{I})^{-1} \underline{X}^T \underline{y}$$ known result.

$$\rightarrow \left.\begin{matrix} \lambda = 0.1 \\ \epsilon = 0.01 \\ \gamma = 0.1 \end{matrix}\right\} \xrightarrow{\text{calculator}} \quad \underline{w}_o = \begin{bmatrix} 0.4902 \\ 1.436 \end{bmatrix}$$

$$\underline{w}_\epsilon = \begin{bmatrix} 0.487 \\ 1.428 \end{bmatrix}, \quad \|w_\epsilon\|_2^2 = 2.27$$

$$\left.\begin{matrix} \lambda = 0.1 \\ \epsilon = 0.01 \\ \gamma = 10^{-8} \end{matrix}\right\} \xrightarrow{\text{calculator}} \quad \underline{w}_{min} = \begin{bmatrix} 0.4902 \\ 2.01 \cdot 10^{-7} \end{bmatrix}$$

$$\underline{w}_e = \begin{bmatrix} 0.487 \\ 2 \cdot 10^{-7} \end{bmatrix}, \quad \|\underline{w}_e\|_2^2 = 0.237$$

Adding the λ term reduces $\underline{w}_e$, the perturbation due to errors ϵ.