REGULARIZATION IN LINEAR SYSTEM

By

Justin Sun

MATH 166 NUMERICAL ANALYSIS
FINAL PROJECT

Submitted to

Howard Levinson

Dec. 9th, 2022

I. Introduction

In this project, I will be exploring the basics of the Regularization of linear systems, more specifically, the Tikhonov Regularization. The reason why this topic attracted me is because of its utilization in noise reduction. As an electrical engineering major student, I encountered many issues that cause some disturbing output responses while the signal is transmitting through the circuit, for example, EMI (short for electromagnetic interference). There are many noise sources in an environment that are not appealing to the functioning of circuits, even a tiny disturbance would shift the result by a large degree. Thus, appropriate numerical methods such as regularization are highly desirable. The Tikhonov Regularization contains many topics covered in the course Numerical Analysis including norms, ill-conditioning, condition numbers, and methods to numerically solve a system. The main portion of this project is discretized into three sections: 1) An overview of the Tikhonov Regulation theory and some of its applications, 2) Two related problems and solutions that evaluate the theory, and 3) Reflection and conclusion on this project.

II. Theory - Tikhonov Regularization

1. Let's first consider a linear system as follows[1]:

$$\frac{1}{60}\begin{bmatrix} 60 & 30 & 20 \\ 30 & 20 & 15 \\ 20 & 15 & 12 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} b \\ c \\ d \end{bmatrix} \tag{1}$$

The matrix is a symmetric and integer 3x3 matrix with solutions:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 110 \\ 65 \\ 47 \end{bmatrix} \rightarrow \begin{bmatrix} x \\ y \\ z \end{bmatrix} \rightarrow \begin{bmatrix} 60 \\ 60 \\ 60 \end{bmatrix} \tag{2}$$

However, if we change $[a\ b\ c]^T$ to be $[11\ 65\ 47]^T$, the solution becomes:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 111 \\ 65 \\ 47 \end{bmatrix} \rightarrow \begin{bmatrix} x \\ y \\ z \end{bmatrix} \rightarrow \begin{bmatrix} 69 \\ 24 \\ 90 \end{bmatrix} \tag{3}$$

Which is very different from the previous solutions.

Such a situation means that:

1) The solution is very sensitive to noise in the data.

2) We are not sure whether a = 110 is the true data or not a = 111.

Let's consider another example,

$$\begin{bmatrix} 1 & 10000 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \Downarrow \quad \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{4}$$

If we vary the data slightly:

$$\begin{bmatrix} 1 & 10000 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0.001 \end{bmatrix}$$
$$\Downarrow$$
$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -9 \\ 0.001 \end{bmatrix} \quad (5)$$

WOW, even an error of 0.001 would largely change the solution!

2. So, why is this happening? What causes this problem?

Recall from MATH 166: if a system is very sensitive to a slight change or noise, then we call such a system ill-conditioned. And, to check if a system is ill-conditioned, we need to look at the condition number **K**(A), defined as

$$K(A) = \mathrm{cond}(A) = \|A\| \cdot \|A^{-1}\| \quad (6)$$

Defining 2-norm for real EVs

$$\|A\|_2 = \max_i \left\{ \sqrt{|\lambda_i|} : \lambda_i \text{ is eigenvalue of } \underbrace{A^T A}_{\substack{\text{symmetric} \\ \hookrightarrow \text{ all eigenvalue} \\ \text{are real!}}} \right\} \quad (6)$$

Using ||A|| with 2-norm, we can find the condition number for (4) is **K**(A) = 10000, a pretty large number, which means that the system is ill-conditioned and will be largely affected by noise.

Relevant recall from MATH 166:

Thm let A be invertable $A\vec{x} = \vec{b}$

let $\vec{e} = \vec{x} - \tilde{x}$ and $\vec{r} = \vec{b} - A\tilde{x}$

Note: $K(A) \geq 1$

then,

$$\frac{\|\vec{e}\|}{\|\vec{x}\|} \leq K(A) \frac{\|\vec{r}\|}{\|\vec{b}\|} \tag{7}$$

"relative" error

"relative" residual ← measurable.

- $\tilde{x}$ is $\vec{x}$ with a small change in data.

In practice, if **K**(A) > 10,000, then the system is considered ill-conditioned.

3. A good solution: Tikhonov Regularization

Consider a linear system A**x** = **b**, the Tikhonov Regularization minimizes the cost of the problem with the general form[2]

$$\min_{\mathbf{x}\in\mathbb{R}^n}\{\|A\mathbf{x}-\mathbf{b}\|^2+\lambda^2\|L\mathbf{x}\|^2\}. \tag{8}$$

This replacement is known as Tikhonov regularization. The parameter $\boldsymbol{\lambda} > 0$ is the regularization parameter that balances the first and second terms. In Tikhonov Regularization, we use 2-norm for calculation. Note that when $\boldsymbol{\lambda} = 0$, the system will not be changed. In the second term, the regularization matrix **L** is a p$\times$n matrix where p is an arbitrary positive integer.

Tikhonov Regularization standard form sets **L** = **I** (identity matrix)

$$||Ax - b||^2 + \lambda^2 ||x||^2 \qquad (9)$$

But we are interested in the general form of Tikhonov Regularization because a suitable choice of regularization matrix **L** ≠ **I** can be a much better approximation of the exact solution **x** than the solution with **L** = **I**. For example, one suitable choice of **L** (n×n) can be

$$L = \begin{bmatrix} 1 & 0 & \cdots & \\ -1 & 1 & \cdots & \\ & \ddots & \ddots & \\ 0 & & \ddots & \ddots \end{bmatrix} \qquad (10)$$

By modifying the generalized Tikhonov, we get

$$(A^T A + \lambda^2 L^T L)\vec{x} = A^T \vec{b} \qquad (11)$$

which has the unique solution

$$\vec{x} = (A^T A + \lambda^2 L^T L)^{-1} A^T \vec{b} \qquad (12)$$

Well done!

4. Example of Tikhonov Regularization

To illustrate how it works, let's go back to our example in Part I.

$$\frac{1}{60}\begin{bmatrix} 60 & 30 & 20 \\ 30 & 20 & 15 \\ 20 & 15 & 12 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 110 \\ 65 \\ 47 \end{bmatrix} \tag{13}$$

The system yields an exact solution since the A is invertible:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 60 \\ 60 \\ 60 \end{bmatrix} \tag{14}$$

Now, we again perturb the data

$$\frac{1}{60}\begin{bmatrix} 60 & 30 & 20 \\ 30 & 20 & 15 \\ 20 & 15 & 12 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 111 \\ 65 \\ 47 \end{bmatrix} \tag{15}$$

The solution now is

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 69 \\ 24 \\ 90 \end{bmatrix} \tag{16}$$

But we want to find a smooth solution to this noise sensitivity system. So, let's use generalized Tikhonov with a suitable choice of **L** and matrices A and **b** as follow

$$A = \frac{1}{60}\begin{bmatrix} 60 & 30 & 20 \\ 30 & 20 & 15 \\ 20 & 15 & 12 \end{bmatrix} \quad L = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \quad b = \begin{bmatrix} 111 \\ 65 \\ 47 \end{bmatrix} \tag{17}$$

Then we solve the following system:

$$(A^T A + \lambda^2 L^T L)\vec{x} = A^T \vec{b} \tag{18}$$

With **λ** = 0, **λ** = 0.01, and **λ** = 0.1, we get the corresponding solutions:

$$\begin{aligned}
\lambda = 0 \quad & \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 69. \\ 24. \\ 90. \end{bmatrix} \rightarrow A\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 110. \\ 65. \\ 47. \end{bmatrix} \\
\lambda = 0.01 \quad & \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 62.44 \\ 60.64 \\ 55.51 \end{bmatrix} \rightarrow A\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 109.84 \\ 65.82 \\ 47.87 \end{bmatrix} \\
\lambda = 0.1 \quad & \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 55.43 \\ 65.00 \\ 65.73 \end{bmatrix} \rightarrow A\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 110.96 \\ 65.11 \\ 46.93 \end{bmatrix}
\end{aligned} \tag{19}$$

Thus we have found a smooth solution that satisfies the system almost exactly!

Note that since A is invertible, there is an exact solution to this system, but it is not necessary to have an exact solution because, in the real world, most systems are either overdamped or underdamped.

III. IntroWrite a Homework Assignment

Problem 1 (Multi-parts) (Calculator allowed):

Given a linear system, with an exact solution.

$$A = \begin{bmatrix} 2 & 10000 \\ 0 & 1 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} \qquad \vec{x} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

a. Find the 2-norm

b. Find the condition number **K**(A) using 2-norm

c. Now let's try to vary the data slightly with a new **b** = [2; 0.001], what is the solution of the system now? What do you observe? What can you conclude? (Hint: Showing if the system is sensitive to noise, or ill-conditioned.)

d. Applying the standard form Tikhonov Regularization with **λ** = 0 and **λ** = 0.1 when **b** = [2; 0.001], what are the corresponding solutions? What do you observe? What can you conclude? (Calculator allowed)

Problem 2 (Programming/MATLAB):

Given a noise-sensitive linear system

$$\frac{1}{100}\begin{bmatrix} 3 & 200 & 61 \\ 1 & 11 & 2 \\ 1 & 33 & 10 \end{bmatrix} \vec{x} = \begin{bmatrix} 264 \\ 14 \\ 44 \end{bmatrix}$$

With an exact solution in this case **x** = [100; 100; 100] since A is invertible.

Use a suitable regularization matrix

$$L = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}$$

a. Using Tikhonov Regularization with multiple **λ** from [0,1] with log increment ie. [0…0.001…1], what do you observe when there is no noise interference?

b. Now let's add a small noise to the system data with **b** = [264; 14; 45]. Redo part a with multiple **λ** from [0,1]. Compared with part a, what does the solution seem to converge to? Or can you find a "smooth" solution to the system?

IV. Solutions

Problem 1 Solution:

a.

$$A = \begin{bmatrix} 2 & 10000 \\ 0 & 1 \end{bmatrix} \qquad A^T = \begin{bmatrix} 2 & 0 \\ 10000 & 1 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 2 & 0 \\ 10000 & 1 \end{bmatrix}\begin{bmatrix} 2 & 10000 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 20000 \\ 20000 & 100000001 \end{bmatrix}$$

$$\det(\lambda I - A^TA) = 0 \qquad \lambda I - A^TA = \begin{bmatrix} \lambda - 4 & -20000 \\ -20000 & \lambda - 100000001 \end{bmatrix}$$

$$\Downarrow$$

$$(\lambda - 4)(\lambda - 100000001) - 4 \times 10^8 = 0$$

solve for $\lambda_1 = 0$ or $\underline{\lambda_2 = 1 \times 10^8}$ ← pick this

$$\boxed{\|A\|_2 = \sqrt{1 \times 10^8} = 1 \times 10^4 \text{ or } 10000}$$

b. $K(A) = \|A\|_2 \cdot \|A^{-1}\|_2$

$$A^{-1} = \frac{1}{2}\begin{bmatrix} 1 & -10000 \\ 0 & 2 \end{bmatrix} \qquad (A^{-1})^T = \begin{bmatrix} \frac{1}{2} & 0 \\ -5000 & 1 \end{bmatrix}$$

$$(A^{-1})^T A^{-1} = 1\times10^7 \cdot \begin{bmatrix} 0 & -3\times10^{-4} \\ -3\times10^{-4} & 2.5 \end{bmatrix}$$

$$\det(\lambda I - (A^{-1})^T A^{-1}) = 0 \qquad \lambda I - (A^{-1})^T A^{-1} = \begin{bmatrix} \lambda & -3000 \\ -3000 & \lambda - 2.5\times10^7 \end{bmatrix}$$

$$\Downarrow$$

$$\lambda(\lambda - 2.5\times10^7) - 9\times10^6 = 0$$

solve $\lambda_1 = -0.36$ or $\underline{\lambda_2 = 25\times10^6}$

$$\|A^{-1}\|_2 = \sqrt{2.5\times10^7} = 5000 \qquad \boxed{\begin{aligned} K(A) &= 10000 \times 5000 \\ &= 5\times10^7 \end{aligned}}$$

c. $b = \begin{bmatrix} 2 \\ 0.001 \end{bmatrix}$

$$\begin{bmatrix} 2 & 10000 \\ 0 & 1 \end{bmatrix} \vec{x} = \begin{bmatrix} 2 \\ 0.001 \end{bmatrix}$$

$$\text{solve } \vec{x} = \begin{bmatrix} 2 & 10000 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 0.001 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{2} & -5000 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0.001 \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} -4 \\ 0.001 \end{bmatrix}$$

With a slight change in the data, the solution now is very different compared to the exact solution. And as we saw $\kappa(A) = 5 \times 10^7$ which is very large. Thus the system is ill-conditioned and sensitive to noise.

d).

$$L = I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\vec{x} = \left(\begin{bmatrix} 2 & 0 \\ 10000 & 1 \end{bmatrix} \begin{bmatrix} 2 & 10000 \\ 0 & 1 \end{bmatrix} + 0^2 \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 2 & 0 \\ 10000 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0001 \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} -4 \\ 0.001 \end{bmatrix}$$

$$\vec{x} = \left(\begin{bmatrix} 2 & 0 \\ 10000 & 1 \end{bmatrix} \begin{bmatrix} 2 & 10000 \\ 0 & 1 \end{bmatrix} + (0.1)^2 \cdot \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} 2 & 0 \\ 10000 & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0001 \end{bmatrix}$$

$$\vec{x} = \begin{bmatrix} -16 \times 10^{-6} \\ 2 \times 10^{-4} \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Compare to the exact solution $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, we can see that the solution have improved with Tikhonov Regularization. In general, Tikhonov method will gives a better approximation when the system is under noise.

Problem 2 Solution (MATLAB):

V. Homework Narrative Reflection

Personally, I believe the two problems I wrote cover the essential parts of the Regularization of linear systems, especially focusing on the most widely used method-Tikhonov Regularization. In the first question, I brought up the concepts of norms and conditional number to explain the meaning of ill-conditioned/ noise-sensitive systems. Then, I included the usage of Tikhonov Regularization in its standard form to evaluate how this method would help to reduce the noise effect. The selection of the system could be improved to have a better result explaining the Tikhonov Regularization. In the second problem, which is highly based on programming skills, I offered students an opportunity to try constructing the Tikhonov Regularization in its general form by themselves with the choice of $\boldsymbol{\lambda}$. This problem also reinforces the concept by comparing the effect of Tikhonov Regularization with and without noise.

Those two problems teach both the theory and its potential applications. Students will find it easier to understand the topic with the practice of these problems. In the real world, systems can be expressed in matrix form, for example, control circuits. And since Tikhonov Regularization aims to reduce the noise effect, it is very useful in designing and analyzing electrical signals and feedback. In addition, the level of these two problems is fair for the undergraduate level with the assistance of the calculator (if needed) and programming software such as MATLAB.

VI. Conclusion

First of all, this project is very amazing and interesting. Being able to explore a new topic by myself is surely helpful in terms of learning, applying, and, more importantly, teaching the concepts. One thing that surprised me during the project was the broadness of methods to solve one kind of problem. For instance, in the regularization of linear systems, the Tikhonov Regularization is the most widely used method, but there are many other methods such as Truncated SVD. And, different methods perform well in different cases, just like the root-finding problem. I have been always interested in the logic or algorithm behind every machine. Finding better methods or algorithms will forever drive scientists and engineers to improve the world we live in.

Bibliography

[1] Yagle, A. E. (2005). Regularized matrix computations. *matrix*, *500*, 10. https://web.eecs.umich.edu/~aey/recent/regular.pdf

[2] Park, Y., Reichel, L., Rodriguez, G., & Yu, X. (2018). Parameter determination for Tikhonov regularization problems in general form. Journal of Computational and Applied Mathematics, 343, 12-25. https://par.nsf.gov/servlets/purl/10062075