CS6210: Homework 1

Christopher Mertin

November 1, 2016

1. Often in practice, an approximation of the form

$$u(t) = \gamma_1 e^{\gamma_2 t}$$

is sought for a data fitting problem, where γ_1 and γ_2 are constants. Assume given data $(t_1, z_1), (t_2, z_2), \ldots, (t_m, z_m)$ where $z_i > 0$, $i = \{1, 2, \ldots, m\}$, and $m > 0$.

(a) Explain in one brief sentence why the techniques introduced in the present chapter cannot be directly applied to find this $u(t)$.

Solution:

Because the equation is not linear in the coefficients γ_1 and γ_2.

(b) Considering instead

$$v(t) = \ln(u(t)) = (\ln(\gamma_1)) + \gamma_2 t$$

it makes sense to define $b_i = \ln(z_i)$, $i = \{1, 2, \ldots, m\}$, and then find coefficients x_1 and x_2 such that $v(t) = x_1 + x_2 t$ is the best least squares fit for the data $(t_1, b_1), (t_2, b_2), \ldots, (t_m, b_m)$

Using this method, find $u(t)$ for the data

i	1	2	3
t_i	0.0	1.0	2.0
z_i	$e^{0.1}$	$e^{0.9}$	e^2

Solution:

We can take the exponential of both sides, giving

$$u(t) = e^{\gamma_1 + t\gamma_2} = e^{\gamma_1} e^{\gamma_2}$$

with

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} 1 \\ 0.9 \\ 2 \end{pmatrix}$$

In finding $\mathbf{x}$ in $A\mathbf{x} = \mathbf{b}$, we get

$$x = \begin{pmatrix} 0.05 \\ 0.95 \end{pmatrix}$$
$$u(t) = e^{0.05}e^{0.95t} = a_1 e^{ta_2}$$

2.

(a) Why can't one directly extend the LU-decomposition to a long and skinny matrix in order to solve least squares problems?

Solution:

It only works with square matrices.

(b) When writing $\mathbf{x} = A^{\dagger}\mathbf{b} = \cdots = R^{-1}Q^T\mathbf{b}$ we have somehow moved from the conditioning $[\kappa(A)]^2$ to the conditioning $\kappa(A)$ through mathematical equalities. Where is the improvement step hidden? Explain.

Solution:

For normal equations we have

$$\begin{aligned} A^TA\mathbf{x} &= A^T\mathbf{b} \\ \kappa\left(A^TA\right) &= \kappa\left(R^TQ^TQR\right) = \kappa\left(R^TR\right) \\ &= \kappa^2(R) = \kappa\left(A^2\right) \end{aligned}$$

For QR Factorization and sinc Q is orthogonal and can be omitted in κ

$$\kappa\left(R^{-1}Q^T\right) = \kappa\left(R^{-1}\right) = \kappa R = \kappa A$$

Since

$$\kappa R = ||R||\left|\left|R^{-1}\right|\right| = \left|\left|R^{-1}\right|\right| ||R|| = \kappa\left(R^{-1}\right)$$

3.

(a) Explain what may happen during the course of the Gram-Schmidt process if the matrix A is rank deficient.

Solution:

From page 158 of the text, it states that when A is rank deficient, $r_{i,i}$ will be zero for some i, and then in calculating q_i will lead to a division by zero.

(b) Show that classical Gram-Schmidt and the modified version are mathematically equivalent.

Solution:

Due to rounding errors, q_i's obtained during classical Gram-Schmidt are often *not orthogonal*. Therefore, we can rewrite it as

$$q_k = a_k - proj_{\mathbf{q_1}}\mathbf{a_k} - proj_{\mathbf{q_2}}\mathbf{a_k} - \cdots - proj_{\mathbf{q_{k-1}}}\mathbf{a_k}$$

With the modified version being written as

$$\begin{aligned}
q_k^{(1)} &= a_k - proj_{\mathbf{q_1}}\mathbf{a_k} \\
&\vdots \\
q_k^{(k-2)} &= q_k^{(k-3)} - proj_{\mathbf{q_{k-2}}}\left(\mathbf{q_k^{(k-3)}}\right) \\
q_k^{(k-1)} &= q_k^{(k-2)} - proj_{\mathbf{q_{k-1}}}\left(\mathbf{q_k^{(k-2)}}\right)
\end{aligned}$$

where, if we write $q_k^{(k-1)}$ recursively, we get back the classical version:

$$\begin{aligned}
q_k^{(k-1)} &= q_k^{(k-2)} - proj_{\mathbf{q_{k-1}}}\mathbf{q}_k^{(k-2)} \\
&= \left(q_k^{(k-3)} - proj_{\mathbf{q_{k-2}}}\mathbf{q}_k^{(k-3)}\right) - proj_{\mathbf{q_{k-1}}}\left(\mathbf{q_k^{(k-3)}} - proj_{\mathbf{q_{k-2}}}\mathbf{q_k^{(k-3)}}\right) \\
&= a_k - proj_{\mathbf{q_1}}\mathbf{a_k} - proj_{\mathbf{q_2}}\mathbf{a_k} - \cdots - proj_{\mathbf{q_{k-1}}}\mathbf{a_k} \quad \square
\end{aligned}$$

(c) Construct a 3×2 example, using a decimal floating point system with a 2-digit fraction, in which modified Gram-Schmidt proves to be more numerically stable than the classical version.

Solution:

A 3×2 matrix is not possible as both of these algorithms would produce the same result. This is due to the for loops, since both of these produce the same values for the first two columns in the matrix.

4. Let α be scalar, and consider the iterative scheme

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha\left(\mathbf{b} - A\mathbf{x}_k\right)$$

This is the gradient descent method with a fixed step size α.

(a) If $A = M - N$ is the splitting associated with this method, state what M and the iteration matrix T are.

Solution:

According to page 182 of the text, we have that $M^{-1} = \alpha\mathbb{1}$, so $M = \frac{1}{\alpha}\mathbb{1}$, giving the iteration matrix T as

$$\begin{aligned}
T &= \mathbb{1} - M^{-1}A \\
&= \mathbb{1} - \alpha A
\end{aligned}$$

(b) Suppose A is symmetrix positive definite and its eigenvalues are $\lambda_1 > \lambda_2 > \cdots > \lambda_n > 0$.

i. Derive a condition on α that guarentees convergence of the scheme to the solution $\mathbf{x}$ for any initial guess.
Solution:

$$x_{k+1} = x_k \left(\mathbb{1} - \alpha A\right) + \alpha \mathbf{b}$$

If x_k is an eigenvector of A, then

$$\begin{aligned} x_{k+1} &= \alpha\lambda_i x_k + \alpha\mathbf{b} = (1 - \alpha\lambda_i)x_k + \alpha\mathbf{b} \\ e_{k+1} &= e_k(1 - \alpha\lambda_i) \end{aligned}$$

For the series to be convergent, we need

$$\begin{aligned} \frac{e_{k+1}}{e_k} &< 1 \\ |1 - \alpha\lambda_i| &< 1 \end{aligned}$$

We have two inequalities: $1 - \alpha\lambda_i < 1$ and $1 - \alpha\lambda_i > -1$, giving the following

$$\begin{aligned} 1 - \alpha\lambda_i &< 1 \\ \Rightarrow \alpha &> 0 \\ \alpha &< \frac{2}{\lambda_i} \end{aligned}$$

If we put the biggest eigenvalue, λ_1, the inequality holds and thus does for all eigenvalues. In other words,

$$0 < \alpha < \frac{2}{\lambda_i} \quad i = \{1, 2, \ldots, n\}$$

ii. Express the condition on α in terms of the spectral condition number of A, $\kappa(A)$.
Solution:
We can simply substitute $\kappa(A) = \frac{\lambda_1}{\lambda_n}$ into the above inequality, which gives

$$0 < \alpha < \frac{2}{\lambda_n \kappa(A)}$$

iii. Show that the best value for the step size in terms of maximizing the speed of convergence is

$$\alpha = \frac{2}{\lambda_1 + \lambda_2}$$

Find the spectral radius of the iteration matrix in this case, and express it in terms of the condition number A.

Solution:

We have that $T = \mathbb{1} - \alpha A$, so we can compute $\kappa(T)$ as

$$\kappa(T) = \frac{|1 - \lambda_1 \alpha|}{|1 - \lambda_n \alpha|}$$

Where the best value for the step size would happen with $\kappa(T) = 1$, so

$$\begin{aligned} 1 &= \frac{|1 - \lambda_1 \alpha|}{|1 - \lambda_n \alpha|} \\ (1 - \lambda_1 \alpha)^2 &= (1 - \lambda_n \alpha)^2 \\ \alpha^2 \left(\lambda_1^2 - \lambda_n^2\right) - 2\alpha \left(\lambda_1 - \lambda_n\right) &= 0 \\ \alpha^2 \left(\lambda_1 + \lambda_n\right) &= 2\alpha \end{aligned}$$

Since $\alpha > 0$, the above equation results in

$$\alpha = \frac{2}{\lambda_1 + \lambda_n}$$

iv. Determine whether the following statement is true or false. Justify your answer

"If A is strictly diagonally dominant and $\alpha = 1$, then the iterative scheme converges to the solution for any initial guess $\mathbf{x}_0$."

Solution:

This can be done with an example

$$A = \begin{pmatrix} 3 & 0 \\ 0 & 4 \end{pmatrix}$$

Matrix A is strictly diagonally dominant and also since it is diagonal its eigenvalues are $\{3, 4\}$, so it is positive semi-definite as well. With its largest eigenvalue being 4, for this to be convergent α needs to be restricted by $0 < \alpha < 1/4$. The question states that $\alpha = 1$, so therefore it is not convergent.

5. Consider the linear system $A\mathbf{x} = \mathbf{b}$, where A is a symmetric matrix. Suppose that $M - N$ is a splitting of A, where M is symmetric positive definite and N is symmetric. Show that if $\lambda_{\min}(M) > \rho(N)$, then the iterative scheme $M\mathbf{x}_{k+1} = N\mathbf{x}_k + \mathbf{b}$ converges to $\mathbf{x}$ for any initial guess $\mathbf{x}_0$.

Solution:

The question states that $\lambda_{\min} > \rho(N)$, so we can use the fact that

$$\lambda_{\min}(M) = \frac{1}{\lambda_{\max}\left(M^{-1}\right)}$$

By substituting this equation into the previous, we get

$$\begin{aligned}
\lambda_{\min}(M) &> \rho(N) \\
&\Rightarrow 1 > \lambda_{\max}\left(M^{-1}\right)\rho(N) \\
&= \rho\left(M^{-1}\right)\rho(N) \\
&\Rightarrow 1 > \rho\left(M^{-1}\right)\rho(N) \qquad (1)
\end{aligned}$$

Now, to show that it is convergent with any starting point

$$\begin{aligned}
Mx_{k+1} &= Nx_k + \mathbf{b} \\
&\Rightarrow x_{k+1} = M^{-1}Nx_k + M^{-1}\mathbf{b} \\
e_{k+1} &= \left(M^{-1}N\right)e_k \\
&\Rightarrow ||e_{k+1}|| = \left|\left|M^{-1}Ne_k\right|\right| \\
&\leq \left|\left|M^{-1}\right|\right| ||N||\,||e_k|| \\
\frac{||e_{k+1}||}{||e_k||} &\leq \left|\left|M^{-1}\right|\right| ||N||
\end{aligned}$$

Equation 1 completes the proof, giving

$$\begin{aligned}
\left|\left|M^{-1}\right|\right| ||N|| &< 1 \\
&\Rightarrow \frac{||e_{k+1}||}{||e_k||} < 1
\end{aligned}$$

6.

(a) Write a program for solving the linear least squares problems that arise throughout the iterations of the GMRES method, using Givens rotations, where the matrix is a non-square $(k+1) \times k$ upper Hessenberg matrix. Specifically, solve

$$\min_{\mathbf{z}} ||\rho e_1 - H_{k+1,k}\mathbf{z}||$$

Provide a detailed explanation of how this is done in your program, and state what Q and R in the associated QR-Factorization are.

Solution:

The upper Hessenberg matrix is formed via iterative QR Factorizations and is of the form $H_k = Q_k R_k$, and can be formed from Givens Rotation Matrix. Following which, we can get the solution to the minimization problem by solving the system

$$R_k\mathbf{x}_k = Q_k^T\mathbf{b}$$

which can be iterated over for to get the values. In our instance, $\rho = \mathbf{b} - A\mathbf{x}_k$ and $e_1 = (1, 0, 0, \ldots, 0) \in \mathbb{R}^n$. This gives the minimization problem that we have to solve as being

$$R_k \mathbf{z} = Q_k^T \rho e_1$$

Where the code for this can be found in `prob6a.py`

(b) Given $H_{k+1,k}$, suppose we perform a step of the Arnoldi process. As a result, we now have a new upper Hessenberg matrix $H_{k+2,k+1}$. Describe the relationship between the old and the new upper Hessenberg matrices and explain how this relationship can be used to solve the new least squares problem in an economical fashion.

Solution:

The system produced in the minimization problem is

$$AQ_n = Q_{n+1}\widetilde{H}_n$$

and the produced upper Hessenberg matrix is of dimensions $(m+1) \times m$ while the original matrix AQ_n is $m \times n$. Therefore, it produces a smaller matrix with the new iteration which can be used to save on computation time.

(c) The least squares problems throughout the iterations can be solved using a QR decomposition approach. Show that the upper triangular factor cannot be singular unless $\mathbf{x}_k = \mathbf{x}$, the exact solution.

Solution:

$$\begin{aligned} A^T A\mathbf{x} &= A^T\mathbf{b} \\ R^T Q^T Q R\mathbf{x} &= R^T Q^T \mathbf{b} \\ R^T R\mathbf{x} &= R^T Q^T \mathbf{b} \quad \left(Q^T Q = \mathbb{1}\right) \\ R\mathbf{x} &= Q^T\mathbf{b} \\ \mathbf{x} &= R^{-1} Q^T \mathbf{b} \end{aligned}$$

where R is non-singular, or else it wouldn't be invertable.

7. A column-stochastic matrix P is a matrix whose entries are nonnegative and whose column sums are all equal to 1. In practice, such matrices are often large and sparse.

Let E be a matrix of the same size as P, say, $n \times n$, all of whose entries are equal to $1/n$, and let α be a scalar, $0 < \alpha < 1$.

(a) Show that $A(\alpha) = \alpha P + (1-\alpha)E$ is also a column-stochastic matrix.

Solution:

We can write any arbitrary element of each matrix in the above given equation, giving

$$A_{i,j} = \alpha P_{i,j} + (1-\alpha)E_{i,j}$$

Now in taking a summation of an arbitrary column from above

$$\begin{aligned}\sum_{i=1}^{n} A_{i,j} &= \alpha \sum_{i=1}^{n} P_{i,j} + (1-\alpha)\sum_{i=1}^{n} E_{i,j} \\ &= \alpha + (1-\alpha) = 1\end{aligned}$$

Which is due to the fact that E and P are column-stochastic matrices.

(b) What are the largest eigenvalue and corresponding eigenvector of $A(\alpha)$?

Solution:

The text states that the largest eigenvalue of a *row-stochastic matrix* is 1. $A(\alpha)$ is a *column-stochastic matrix*, so its transpose is a *row-stochastic matrix* with a largest eigenvalue of 1. Therefore, the largest eigenvalue of $A(\alpha)$ is also 1 since the eigenvalues of a matrix and its transpose are the same as it's based on the linear dependence of the rows/columns.

(c) Show that the second largeset eigenvalue of $A(\alpha)$ is bounded (in absolute value) by α.

Solution:

[Extracted from `http://nlp.stanford.edu/pubs/secondeigenvalue.pdf`]

We know that the second eigenvector $\mathbf{x}_2$ of A must be an eigenvector $\mathbf{y}_i$ of P^T, and the corresponding eigenvalue is $\gamma_i = \lambda_2/\alpha$.

So, since $\lambda_2 = \alpha\gamma_i$, and since we know P is stochastic, we have $|\gamma_i| \leq 1$, and therefore $|\lambda_2| \leq \alpha$.

(d) Suppose the dominant eigenvector of $A(\alpha)$ is to be computed using the power method. This vector, if normalized so that its ℓ_1-norm is equal to 1, is called the stationary distribution vector.

i. Show how matrix-vector products with $P(\alpha)$ can be performed in an efficient manner in terms of storage. (Assume n is very large, and recall that E is dense).

Solution:

If $P(\alpha)$ has elemental values of $1/n$, a matrix vector multiplication can be computed efficiently in storage by simply storing the value of α/n in a single float, and the resulting scalar vector multiplicatio from α/n and a vector would produce the same results as if you did the enitre matrix vector multiplication since all elements in the matrix have the same value.

If, they infact, do not have the same values, then you can store only the nonzero elements and their indices, and simply do a matrix vector multpilcation that way, where you would ignore all the zeros as they wouldn't produce any meaningful results. Something such as CSR matrix format.

ii. Show that if the power method is applied, then if the initial guess $\mathbf{v}_0$ satisfies $||\mathbf{v}_0||_1 = 1$, then all subsequent iterates $\mathbf{v}_k$ also have a unit ℓ_1-norm, and hence there is no need to normalize throughout the iteration.
Solution:
Since each row in A is scaled to sum upto n, if each row in $\mathbf{v}_0$ is scaled to sum up to 1, so will the result of $A \cdot \mathbf{v}$. This is due to the fact that since each column will scale proportionally to those values in $\mathbf{v}$ such that they will also scale to 1. With this scaling, none of the norms will have to be taken since they'll be 1. In other words

$$\sum_{i=1}^{n} \left(A^{(k)}\mathbf{x}\right)_i = \sum_{i=1}^{n} (\mathbf{x})_i$$

8. Consider the least squares problem

$$\min_x ||\mathbf{b} - A\mathbf{x}||_2$$

where we know that A is ill-conditioned. Consider the *regularization* approach that replaces the normal equations by the modified, better-conditioned system

$$\left(A^T A + \gamma I\right) \mathbf{x}_\gamma = A^T \mathbf{b},$$

where $\gamma > 0$ is a parameter.

(a) Show that $\kappa_2^2(A) \geq \kappa_2\left(A^T A + \gamma I\right)$
Solution:
Since $A^T A$ is symmetrix positive-definite, its eigenvalues are such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. We have

$$\begin{aligned} A^T A\mathbf{x} &= \lambda\mathbf{x} \\ &\Rightarrow A^T A\mathbf{x} + \gamma\mathbf{x} = \lambda\mathbf{x} + \gamma\mathbf{x} \\ \mathbf{x}\left(A^T A + \gamma\right) &= \mathbf{x}\left(\lambda + \gamma\right) \\ \kappa_2\left(A^T A + \gamma\mathbb{1}\right) &= \frac{\lambda_1 + \gamma}{\lambda_n + \gamma} \end{aligned}$$

Since $\kappa^2(A) = \kappa\left(A^T A\right)$

$$\kappa^2(A) = \frac{\lambda_1}{\lambda_n}$$

To prove $\kappa_2^2 \geq \kappa_2\left(A^TA + \gamma\mathbb{1}\right)$, we show that $\kappa_2^2(A) - \kappa_2\left(A^TA + \gamma\mathbb{1}\right) \geq 0$

$$\begin{aligned}\kappa_2^2(A) - \kappa_2\left(A^TA + \gamma\mathbb{1}\right) &= \frac{\lambda_1}{\lambda_n} - \frac{\lambda_1 + \gamma}{\lambda_n + \gamma} \\ &= \frac{\lambda_1\lambda_n + \lambda_1\gamma - \lambda_1\lambda_n - \lambda_n\gamma}{(\lambda_n)^2 + \lambda_n\gamma} \\ &= \frac{(\lambda_1 - \lambda_n)\,\gamma}{(\lambda_n)^2 + \lambda_n\gamma}\end{aligned}$$

which is greater than or equal to zero since $\gamma > 0$ and all eigenvalues are greater than or equal to zero.

(b) Reformulate the equations for $\mathbf{x}_\gamma$ as a linear least squares problem.
Solution:

$$\begin{aligned}\underbrace{\left(A^TA + \gamma I\right)}_{M}\mathbf{x}_\gamma &= A^T\mathbf{b} \\ \underbrace{AM}_{W}\mathbf{x}_\gamma &= \underbrace{\left(AA^T\right)\mathbf{b}}_{\mathbf{d}} \\ W\mathbf{x}_\gamma &= \mathbf{d} \\ &\Rightarrow \min_{\mathbf{x}_\gamma} ||\mathbf{d} - W\mathbf{x}_\gamma||_2\end{aligned}$$

(c) Show that $||\mathbf{x}_\gamma||_2 \leq ||\mathbf{x}||_2$.
Solution:
From the two equations, we have

$$\begin{aligned}A^TA\mathbf{x}_\gamma + \gamma\mathbb{1}\mathbf{x}_\gamma &= A^TA\mathbf{b} \\ A^TA\mathbf{x} &= A^T\mathbf{b}\end{aligned}$$

Where we can set these two equal to each other, with $A^TA = M$, giving

$$\begin{aligned}M\mathbf{x}_\gamma + \gamma\mathbb{1}\mathbf{x}_\gamma &= M\mathbf{x} \\ \mathbf{x}_\gamma + M^{-1}\gamma\mathbb{1}\mathbf{x}_\gamma &= \mathbf{x} \\ \mathbf{x} - \mathbf{x}_\gamma &= M^{-1}\gamma\mathbb{1}\mathbf{x}_\gamma \\ \mathbf{x} - \mathbf{x}_\gamma &\geq 0\end{aligned}$$

Where the last term comes about since A is postive definite and the right hand side of the equation would be greater than zero. Since this is true, we can say $||\mathbf{x}_\gamma||_2 \leq ||\mathbf{x}||_2$.

(d) Find a bound for the relative error $\frac{||\mathbf{x}-\mathbf{x}_\gamma||_2}{||\mathbf{x}||_2}$ in terms of either the largest or the smallest singular value of the matrix A.

State a sufficient condition on the value of γ that would guarentee that the relative error is bounded below a given value ϵ.

Solution:

$$\begin{aligned}
\mathbf{x}_\alpha &= \mathbf{x} - \left(A^T A\right)^{-1} \gamma \mathbf{x}_\gamma \\
\mathbf{x} - \mathbf{x}_\alpha &= \left(A^T A\right)^{-1} \gamma \mathbf{x}_\gamma \\
||\mathbf{x} - \mathbf{x}_\alpha|| &= \left|\left|\left(A^T A\right)^{-1} \gamma \mathbf{x}_\gamma\right|\right| \\
&\leq \gamma \left|\left|\left(A^T A\right)^{-1}\right|\right| ||\mathbf{x}_\gamma||
\end{aligned}$$

But, since $||\mathbf{x}_\gamma||_2 \leq ||\mathbf{x}||_2$

$$||\mathbf{x} - \mathbf{x}_\alpha|| \leq \gamma \left|\left|\left(A^T A\right)^{-1}\right|\right| ||\mathbf{x}||$$

Finally

$$\frac{||\mathbf{x} - \mathbf{x}_\alpha||}{||\mathbf{x}||} \leq \gamma \left|\left|\left(A^T A\right)^{-1}\right|\right|$$

We need the relative error to be less than ϵ, so

$$\begin{aligned}
\gamma \left|\left|\left(A^T A\right)^{-1}\right|\right| &\leq \epsilon \\
\Rightarrow \gamma &\leq \frac{\epsilon}{\left|\left|(A^T A)^{-1}\right|\right|}
\end{aligned}$$

giving

$$\frac{||\mathbf{x} - \mathbf{x}_\alpha||}{||\mathbf{x}||} \leq \gamma \left|\left|\left(A^T A\right)^{-1}\right|\right| \leq \epsilon$$

(e) Write a short program to solve the 5×4 problem in Example 8.8 regularized as above, using MATLAB's backslash command. Try $\gamma = 10^{-j}$ for $j = \{0, 3, 6, 12\}$. For each γ, calculate the ℓ_2-norms of the residual, $||B\mathbf{x}_\gamma - \mathbf{b}||$, and the solution, $||\mathbf{x}_\gamma||$. Compare to the results for $\gamma = 0$ and to those using SVD as reported in Example 8.8. What are your conclusions?

Solution:

(f) For large ill-conditioned least squares problems, what is the potential advantage of the regularization with γ presented here over minimum norm truncated SVD?

Solution:

In calculating the SVD, the higher the condition number of the matrix, the better. With using a minimum norm truncated SVD on a matrix with a low condition

number, the ratio between the max singular value and the minimum singular value is low. In doing a minimum norm, there is no stability on the number of eigenvalues that are being removed. For example, if you're looking for a norm truncation of 0.9, and all the values are below 1, you will have to remove a good majority of the eigenvalues to get the norm that low.

With this method, it is independent on the removal of the eigenvalues and essentially "feeds" any error back into the system with $\gamma\mathbb{1}$, and removes error. Also, there's no possibility of removing *all* of the eigenvalues in the system.