Comparative analysis of eigenvalue localization methods for stochastic matrices

Akshay Joshi
Department of Operations Research, North Carolina State University

Abstract

Eigevalue Localization(EL) refinements for stochastic matrices proposed by Chaoqian Li and Yaotang Li [5] are studied. The proof is reviewed in depth and gaps in the proof are filled for clarity. Proposed EL modifications [5] are compared with EL proposed by Cvetković [3] on different types of stochastic matrices to measure their effectiveness.

Keywords— stochastic matrix, non-negative matrices, eigenvalue localization

1 Introduction and motivation

In many problems, a sequence of events are independent of its preceding or succeeding events. In 1905, a Russian mathematician Andrei Markov developed a class of probabilistic models where events depend only on its immediate preceding event rather than other preceding events. A system of this type is called a Markov process or chain.

Following examples illustrate applications of markov chains in real life,

- The page rank of a webpage as used by Google is defined by a markov chain [2]
- Markov chains are used in macro-economics to model asset prices and market crashes [10]
- Markovian systems appear extensively in thermodynamics and statistical mechanics, whenever probabilities are used to represent unknown details of the system
- The LZMA lossless data compression algorithm combines markov chains with Lempel-Ziv compression to achieve very high compression ratios

Eigenvalue localization of stochastic matrices plays a key role in many application fields such as Computer Aided Geometric Design, Birth - Death Processes, and Markov chains [5]. Eigenvalue localization is used to bound the subdominant eigenvalue of a stochastic matrix with non-negative eigenvalues, which is crucial for bounding the convergence rate of stochastic processes.

2 Literature Review

Oscar Rojo et al [7] constructed a decreasing sequence of rectangles such that all the eigenvalues of a complex matrix are contained in each rectangle. Shen et al [9] modified the Gerchgorin circle set to localize the real eigenvalues (different from one) of any stochastic matrix. L.J. Cvetković et al [3] carried-out EL refinements for matrices with constant row or column sum. Chaoqian Li and Yaotang Li [5] provided a modification to EL for stochastic matrices which can be used to estimate moduli of the subdominant eigenvalue. This modification improves on EL refinements provided by Cvetković et al [3].

3 Eigenvalue localization of stochastic matrices

In markov chains, transitional probability is the conditional probability of the system in moving from one state to another. Stochastic matrices are used to describe these transitions in a markov chain. Essentially, they are entry-wise non-negative matrices in n dimensional real space whose row/columns add up to 1. Thus, each entry in a stochastic matrix represents a conditional probability of transition[8]. Let's begin by defining an upper bound on the spectrum of stochastic matrices.

Lemma 1. *Let A be a row stochastic matrix, i.e. $A \in R^{n\times n}$ Then $\lambda = 1$ is the dominant eigenvalue of A, i.e. $\mid \lambda \mid \leqslant 1$*

Proof. A is a row stochastic matrix, i.e., $\sum_{j=1}^{n} A_{ij} = 1$

$Ax = \lambda x$
$Let\ e = [1, 1, 1...1]^T$

$$Ae = \begin{bmatrix} \sum_{j=1}^{n} A_{1j} \\ \sum_{j=1}^{n} A_{2j} \\ . \\ . \\ \sum_{j=1}^{n} A_{nj} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ 1 \end{bmatrix}$$

$\therefore \lambda = 1$

Now to prove that $\lambda = 1$ is the dominant eigenvalue, consider the Geršchgorin circle theorem[4].

$G_r = \{\lambda \in C \mid (\lambda_i - a_{kk}) \leqslant \sum_{i\neq k} a_{ki}\}$

$Now\ \sum_{i\neq k} a_{ki} \geqslant 0$

$\therefore a_{kk} \in [0, 1]\ which\ is\ the\ center$

$and\ \sum_{i\neq k} a_{ki} = \sum_{j=1}^{n} a_{ij} - a_{kk} = 1 - a_{kk}\ which\ is\ the\ radius$

Therefore, each Geršchgorin disc will have 1 on its perimeter. Since, all eigenvalues lie on the union, $\mid \lambda \mid \leqslant 1$ □

L.J. Cvetković et al [3] presented a region for all eigenvalues of a stochastic matrix different from 1 by modifying the Geršchgorin discs of A.

Theorem 2. *Consider A as in the previous case. Let s_i be the minimal element among the off-diagonal entries of the i^{th} column of A, that is, $s_i = \min_{j\neq i} a_{ji}$. Taking $\gamma(A) = \max_{i\in N}(a_{ii} - s_i)$,*

Then for any $\lambda \in \sigma(A) \setminus \{1\}$,

$$\mid \lambda - \gamma(A) \mid < r(A) = 1 - trace(A) + (n-1)\gamma(A) \tag{1}$$

Complete proof can be found in [3], theorem 3.4.

The localization circle provided by Cvetković is ineffective in certain cases. Consider the following class of matrices (say SM_0) where all the diagonals and atleast one value in each column are set to zero. This is possible in certain markov chains where transition probabilities for some states are set to zero due to practical constraints. For this class of matrices,

trace(A) $= 0$ and $\gamma(A) = 0$.

Using equation (1), $\mid \lambda \mid < 1$, which is trivial. Thus, Cvetković circle fails to improve localization for SM_0 class of matrices.

4 Modification of eigenvalue localization for stochastic matrices

Let A be a row stochastic matrix $A \in R^{n \times n}$. As proved in lemma 1, 1 is the dominant eigenvalue of A. Row stochastic matrices belong to a broader class of constant row sum matrices such that $Ae = \lambda_k e$ (A is stochastic when λ_k is 1). This section tries to localize eigenvalues different from λ_k.

Lemma 3. *Let $A = [a_{ij}] \in R^{n \times n}$ such that $Ae = \lambda_k e$ and let $\mu \in \sigma(A) \setminus \{\lambda_k\},$. Then for any real number $d_i, i \in N$, we have $-\mu \in \sigma(B)$, where*

$$B = diag(d_1, d_2, ..., d_n)ee^T - A^T \tag{2}$$

Proof. We know that $-\mu \in \sigma(A)$. Let $x \neq 0$ be the corresponding eigenvector of μ.

$A^T x = \mu x$

$Now, consider\ \mu(x^T e) = (\mu x)^T e = (A^T x)^T e = x^T A e = x^T \lambda e = \lambda x^T e$

$But,\ we\ know\ that\ \mu \neq \lambda. This\ implies\ that\ x^T e = e x^T = 0.$

$B.x = diag(d_1, d_2, ..., d_n)ee^T x - A^T x = -A^T x = -\mu x$

$Hence\ proved, -\mu \in \sigma(B).$ □

Cvetković et al [3] proved that a prudent choice of d_i would refine the Geršchgorin circle of A. Chaoqian Li and Yaotang Li [5] proposed that for a non-negative matrix A, the best choice of d_i will be

$$d_i = a_{ii} + \max_{i \in N}\{S_i - a_{ii}\},\ where\ S_i = \max_{j \neq i} a_{ji} \tag{3}$$

This choice of d_i will improve the eigenvalue localization for all eigenvalues other than 1 for a stochastic matrix. In order to define a localization circle for all such eigenvalues, three lemmas are required which will be proved in the following section. In lemma 4, two terms $R_i(B)$ and $S_i(B)$ are defined to establish the relationship between off - diagonal terms of A and B. lemma 5 provides a localization circle for matrices with constraint $S_i \geqslant a_{ii}$. Finally, lemma 6 extends this result to all stochastic matrices.

Lemma 4. *Let $A = [a_{ij}] \in R^{n\times n}$ be non-negative. Let $B = [b_{ij}]$ be the matrix given in (1) with $d_i = S_i$ for each i. Then, as proved in lemma 3, $-\mu \in \sigma(B)$ and,*

$R_i(B) = (n-1)S_i - C_i(A), \quad C_j(B) = \sum_{k\neq j}(S_k - R_j(A)$

where $R_i(B) = \sum_{j\neq i} \mid b_{ij} \mid$ *and* $C_j(B) = R_i(B^T)$.

Proof. By construction, $b_{ij} = S_i - a_{ji} \geqslant 0$ for all i,j. Hence,

$R_i(B) = \sum_{j\neq i}(S_i - a_{ji}) = (n-1)S_i - \sum_{j\neq i} a_{ji} = (n-1)S_i - C_i(A)$

Also,

$C_j(B) = \sum_{k\neq j}(S_k - R_j(A) = \sum_{k\neq j}(S_k - a_{jk}) = \sum_{k\neq j} S_k - R_j(A)$ □

Lemma 5. *Let $A = [a_{ij}] \in R^{n\times n}$ be non-negative and $S_i \geqslant a_{ii}$. Taking $\tilde{\gamma}(A) = \max_{i\in N}\{S_i - a_{ii}\}$, then*

$$\mid \mu + \tilde{\gamma}(A) \mid \leqslant trace(A) + (n-1)\tilde{\gamma}(A) - 1 \tag{4}$$

Proof. Let $d_i = a_{ii} + \tilde{\gamma}(A)$.

For $j \neq i$,
$b_{ij} = d_i - a_{ji} = a_{ii} + \tilde{\gamma}(A) - a_{ji} \geqslant a_{ii} + \tilde{\gamma}(A) - S_i = \tilde{\gamma}(A) - (S_i - a_{ii}) \geqslant 0,$
and
$b_{ii} = d_i - a_{ii} = a_{ii} + \tilde{\gamma}(A) - a_{ii} = \tilde{\gamma}(A)$

Using the above result,

$C_i(B) = \sum_{k\neq i}(d_k - R_i(A)) = \sum_{k\neq i}(d_k - (1 - a_{ii})) = \sum_{k\in N}(d_k - 1 + (a_{ii} - d_i))$

Note that $d_i = a_i i + \tilde{\gamma}(A)$, *then*,

$C_i(B) = \sum_{k\in N}(d_k - 1 - \tilde{\gamma}(A) = \sum_{k\in N}(a_{kk} + \tilde{\gamma}(A)) - 1 - \tilde{\gamma}(A)$
$C_i(B) = trace(A) + (n-1) - 1 - \tilde{\gamma}(A) - 1$

We know that $b_{ii} = \tilde{\gamma}(A)$ and $\sum_{j\neq i} \mid b_{ij} \mid = R_i(B)$.

By Geršchgorin theorem [4], all the eigenvalues of B are contained in the region $\Gamma(B)$ where,

$$\Gamma(B) = \{z \in C: \ \mid z - \tilde{\gamma}(A) \mid \leqslant trace(A) + \tilde{\gamma}(A) - 1\} \tag{5}$$

Using this and lemma 3, the result follows. □

As shown above, lemma 5 provides a localization circle for all eigenvalues different from 1 for a stochastic matrix A with the restriction $S_i \geqslant a_{ii}, i \in N$. In fact, it can proved that the same result holds true for all stochastic matrices.

Lemma 6. *Let $A \in R^{n\times n}$ be a stochastic matrix and let $\Delta = \{i \in N \colon a_{ii} > S_i\}$. If $\Delta \neq \emptyset$, then,*

$$A_\delta = \frac{1}{1+(n-1)\delta}(A + \delta(ee^T - I)) \tag{6}$$

where $\delta = -\min_{i\in N}(S_i - a_{ii})$ is a stochastic matrix. Also, for any $\mu \in \sigma(A)\setminus\{1\}$, $\frac{\mu-\delta}{1+(n-1)\delta} \in \sigma(A\delta)$.

Proof. First, let's prove that A_δ is a stochastic matrix,

Let $A_\delta = [\tilde{a}_{ij}]$

$[\tilde{a}_{ij}] = \frac{a_{ij}}{1+(n-1)\delta}, \quad for\ j \neq i$
$[\tilde{a}_{ij}] = \frac{a_{ij}+\delta}{1+(n-1)\delta}, \quad for\ j = i$

Since $\Delta \neq \emptyset$, we have that $\delta > 0$, and all $[\tilde{a}_{ij}] \geqslant 0$. Hence,$A_\delta$ is non-negative.

To prove that A_δ is stochastic,

$$A_\delta e = \frac{1}{1+(n-1)\delta}(A + \delta(ee^T - I))e = \frac{1}{1+(n-1)\delta}(1.e + (n-1)\delta e)) = \frac{(1+(n-1)\delta)e}{1+(n-1)\delta} = e \tag{7}$$

As $A_\delta e = e$, $A_\delta e$ is a non-negative stochastic matrix.

For $\mu \in \sigma(A) \setminus \{1\}$, let x be the corresponding eigenvector of A. By lemma 3, we have that $e^T x = xe^t = 0$. Then,

$$A_\delta x = \frac{1}{1+(n-1)\delta}(A + \delta(ee^T - I))x = \frac{(\mu-\delta)}{1+(n-1)\delta}x \tag{8}$$

Hence, it follows that $\frac{(\mu-\delta)}{1+(n-1)\delta} \in \sigma(A_\delta)$. □

Thus, all the lemmas required for the defining a localization circle [5] have been proved. Now, parameters of the localization circle need to be defined. This is the proposed modification over Cvetković circle[5].

Theorem 7. *Let $A \in R^{n\times n}$ be a stochastic matrix. Taking $\tilde{\gamma}(A) = \max_{i\in N}\{S_i - a_{ii}\}$, if $\mu \in \sigma(A) \setminus \{1\}$, then,*

$$\mid \mu + \tilde{\gamma}(A) \mid \leqslant trace(A) + (n-1)\tilde{\gamma}(A) - 1$$

Proof. Without loss of generality, we can assume $\Delta \neq \emptyset$. This follows from if $\Delta = \emptyset$, then by lemma 5, inquality (8) will hold.
Let,

$$A_\delta = \frac{1}{1+(n-1)\delta}(A + \delta(ee^T - I)) \tag{9}$$

where $\delta = -\min_{i\in N}(S_i - a_{ii})$. From lemma 6, A_δ is stochastic. For all j$\in N$,

$$S_j(A_\delta) = \frac{S_j + \delta}{1+(n-1)\delta} = \frac{S_j - \min_{i\in N}(S_i - a_{ii})}{1+(n-1)\delta} \geqslant \frac{S_j - (S_j - a_{ii})}{1+(n-1)\delta} = \frac{a_{jj}}{1+(n-1)\delta} = \tilde{a}_{jj} \tag{10}$$

Hence, $S_j(A_\delta) \geqslant \tilde{a}_{jj}$. By lemma 6, we know that if $\mu \in \sigma(A) \setminus \{1\}$, then $\frac{(\mu-\delta)}{1+(n-1)\delta} \in \sigma(A_\delta)$.

Now, applying lemma 5 to A_δ and using the above result, we get,

$$\mid \frac{(\mu-\delta)}{1+(n-1)\delta} + \tilde{\gamma}(A) \mid \leqslant trace(A) + (n-1)\tilde{\gamma}(A) - 1 \tag{11}$$

Now, using equation (10), $\tilde{\gamma}(A_\delta)$ can be written as

$$\tilde{\gamma}(A_\delta) = \max_{i\in N}(S_i(A_\delta) - \tilde{a}_{ii}) = \max_{i\in N}(\frac{S_i+\delta}{1+(n-1)\delta} - \frac{a_{ii}}{1+(n-1)\delta}) = \frac{\delta + \tilde{\gamma}(A)}{1+(n-1)\delta} \tag{12}$$

Combining equations (11) and (12), we have,

$$\mid \frac{(\mu-\delta)}{1+(n-1)\delta} + \frac{\delta+\tilde{\gamma}(A)}{1+(n-1)\delta} \mid \leqslant \frac{trace(A)}{1+(n-1)\delta - 1} + (n-1)\frac{\delta+\tilde{\gamma}(A)}{1+(n-1)\delta - 1} \tag{13}$$

Simplifying, we get

$$\mid \mu + \tilde{\gamma}(A) \mid \leqslant trace(A) + (n-1)\tilde{\gamma}(A) - 1 \tag{14}$$

The proof is complete.

□

5 Comparative Analysis

The modification proposed by Chaoqian Li and Yaotang Li [5] is claimed to be an improvement over EL method proposed by Cvetković [3]. Cvetković 's bounds dont hold in cases where $s_i = 0$ and $a_{ii} = 0$, as shown in previous sections. However, there are many cases in which both the EL methods give meaningful results. Selection of a method in these cases depends on the type of matrices. To evaluate their comparative performance, both the methods were implemented in R. The baseline model for reference is set by Geršchgorin discs. In each run of the experiment, all three methods are plotted on the same graph with the eigenvalues. The performance is then evaluated by comparing the radius of circle plotted by both the Cvetković method and the propsed modification (sometimes referred to as Chaoqian circle).

5.1 Performance vs Matrix Sizes

In many practical applications, the transition matrix grows over time as the system complexity increases. Its interesting to evaluate the performance of EL methods against increasing size of the transition matrices. Matrices were generated in R using randomly generated values from uniform distribution. Row elements are non-negative and add up to 1. Thus, row stochastic matrices of size any n > 2 can be generated. It can be shown that as size of transition matrix increases, modification proposed by Chaoqian Li and Yaotang Li remains effective while bounds proposed by Cvetković circle fail. The size of Cvetković circle is more than the baseline radii set by Geršchgorin discs.

This is a direct consequence of radius of Cvetković circle depending on the size of the matrix.

Radius $= 1 - trace(A) + (n-1)\gamma(A)$

As the matrix size increases, the radius of Cvetković circle becomes more than one, thus rendering it trivial. In the proposed modification, $\tilde{\gamma}(A)$ decreases as the matrix size increases, thereby controlling the radius. It also provides sharper bounds than baseline Geršchgorin discs. Refer to fig. 1 for visual representation of the results. Colored discs are baseline Gershgorin circles while the black and grey discs denote the Cvetkoić and Chaoqian circle respectively.

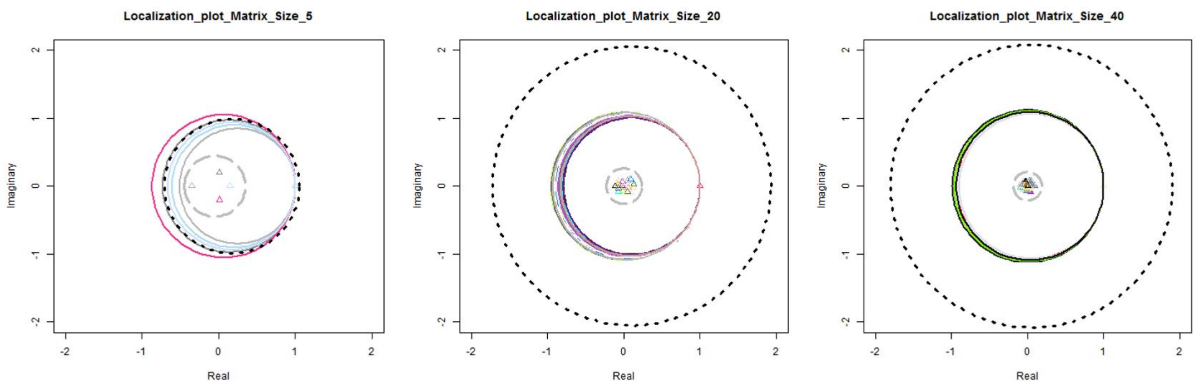

Figure 1: Effectiveness of the proposed modification(grey disc) improves with matrix size

5.2 Performance vs Eigen Gradient

Eigen gradient is defined as the change in magnitude of eigenvalues for a matrix. Gradient of eigenvalues depends on the nature of transition matrix, which in turn, is a function of the system. Certain applications of Markov chains have a steeper eigen gradient. For example, systems with transition probabilities similar for all the states tend to have a much steeper eigen gradient. In order to evaluate effectiveness of the EL methods, 5×5 doubly stochastic matrices were generated for a defined set of real positive eigenvalues as prescribed by L.F. Martignon[6]. EL methods were implemented on generated matrices with both steep and gentle eigen gradient. For matrices with a steep eigen gradient, Cvetković circle performs better than the proposed modification. The radius is smaller than baseline radii and the bound is consistently sharper than the proposed modification(fig. 2).

On the other hand, with a gentle eigen gradient, both the methods fail to localize the subdominant eigenvalues (fig. 3). This is a result of the higher magnitude of all the eigenvalues which drive up the trace of the matrix. As a result, the radii of the localization circles are much higher than the gershgorin discs. Thus, in case of gentle eigen gradient, the baseline geršchgorin discs should be used for eigenvalue localization.

5.3 Performance vs Equal subdominant eigenvalues

In stochastic matrices, we call μ a subdominant eigenvalue if $1 > |\mu| \geqslant |\eta|$ for all eigenvalues η different than 1 and μ. In case of matrices with equal subdominant eigenvalues, performance of

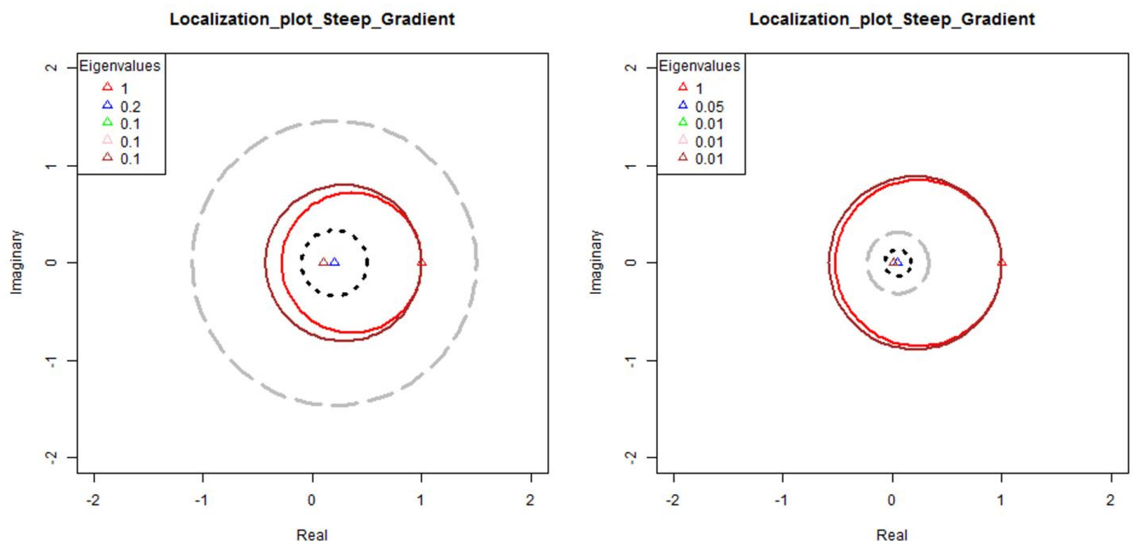

Figure 2: Cvetković circle(black disc) performs better than the proposed modification(grey disc) for matrices with steep eigen-gradient

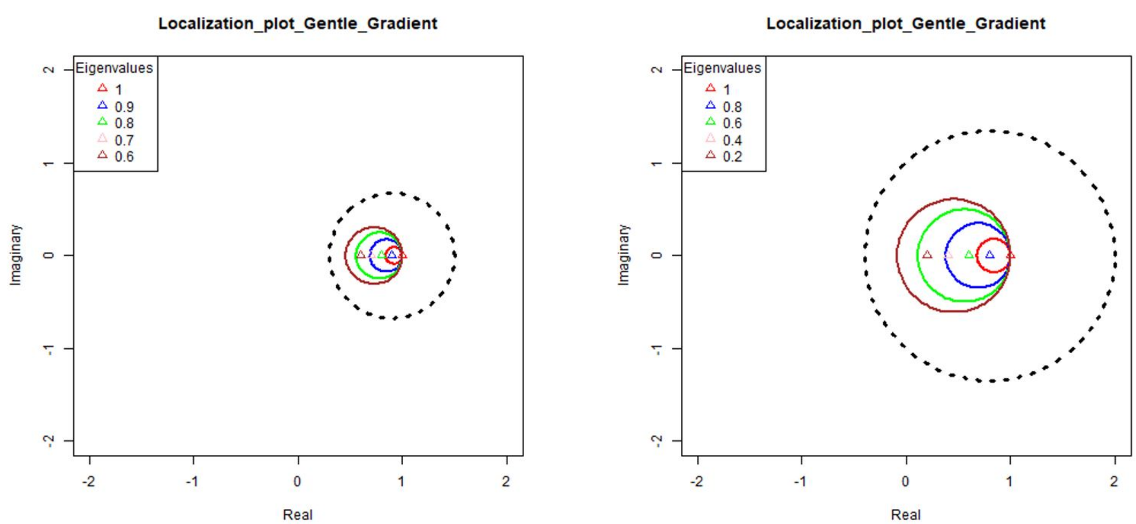

Figure 3: Both the Cvetković circle(black disc) and the proposed modification(grey disc) fail to improve performance for matrices with gentle eigen-gradient

the proposed modification depends on the magnitude of the subdominant eigenvalues. Improved performance is achieved with Chaoquins circle when the magnitude is less than 0.1. However, for larger subdominant eigenvalues, again, both the methods fail to improve localization (fig. 4). This is, again, because of higher trace of the matrix that increases the radii of the localization circles.

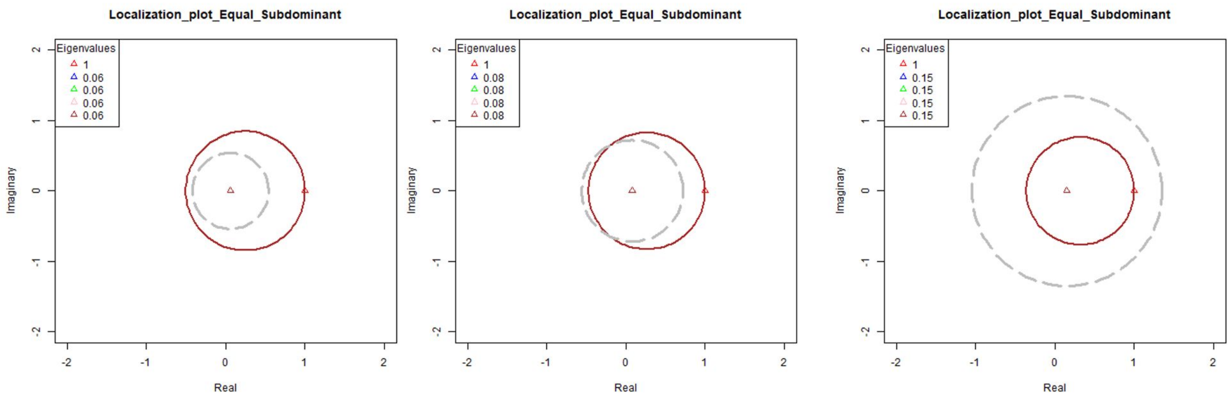

Figure 4: Performance of the proposed modification(grey disc) for matrices with equal sub - dominant eigenvalues depends on magnitude of μ

6 Summary

In conclusion, the propsed modification by Choaqian Li et al [3] is effective for large stochastic matrices and and matrices with small subdominant eigenvalues. Thus, this method may be useful for large and complex markov chains with less or no absorbing states. EL method proposed by Cvetković [5] is still effective for matrices with steep eigen-gradient. Therefore, Cvetković method can be used for simpler markov chains. Both methods fail to improve localization over baseline Gershgorin discs for matrices with gentle eigen-gradient. This,in turn, suggests that an eigenvalue localization method with less dependency on trace of the stochastic matrix will be effective.

7 Notes

The author is grateful to the subject teacher for his useful and constructive feedback. This project was submitted towards completion of MA 723 Theory and Applications of Matrices (Spring 2017). Primary reference paper for this project was *A modification of eigenvalue localization for stochastic matrices* [5] by Chaoqian Li and Yaotang Li. Implementation R code for both the EL Methods and other comparative analysis can be found in the author's GitHub repository[1].

References

[1] Implementation r codes. URL: https://github.com/akshayjoshi942/eigenvalue-localization.

[2] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Seventh International World-Wide Web Conference (WWW 1998)*, 1998.

[3] LJ. Cvetković, V. Kostić, and J. M. Pena. Eigenvalue localization refinements for matrices related to positivity. *SIAM Journal on Matrix Analysis and Applications*, 32(3):771–784, 2011.

[4] S. Geršgorin (S. Gerschgorin). Uber die abgrenzung der eigenwerte einer matrix. *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na*, pages 749–754, 1931.

[5] Chaoqian Li and Yaotang Li. A modification of eigenvalue localization for stochastic matrices. *Linear Algebra and its Applications*, 460:231 – 241, 2014.

[6] L.F. Martignon. Doubly stochastic matrices with prescribed positive spectrum. *Linear Algebra and its Applications*, 61:11 – 13, 1984.

[7] Oscar Rojo, Ricardo L. Soto, and Héctor Rojo. A decreasing sequence of eigenvalue localization regions. *Linear Algebra and its Applications*, 196:71 – 84, 1994.

[8] J.K. Sharma. *Operations Research: Theory and Applications*. Macmillan Publishers India Ltd., 2013.

[9] Shu-Qian Shen, Juan Yu, and Ting-Zhu Huang. Some classes of nonsingular matrices with applications to localize the real eigenvalues of real matrices. *Linear Algebra and its Applications*, 447:74 – 87, 2014.

[10] Nikolay Timofeev and Galina Timofeeva. *Estimation of Loan Portfolio Risk on the Basis of Markov Chain Model*, pages 207–216. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.