The Comparison of HIO and GEC in Solving Phase Retrieval

Ziqi Zhang

August 2018

1 Introduction

The phase retrieval problem aims to estimate the phase of the signal's complex measurement given the modulus of the measurement and prior information of the signal (prior distribution, real, positive, etc.). In 1982, Fienup introduced the Hybrid Input Output(HIO) algorithm[1], which is still widely used today and is believed to be one of the best algorithm to solve this problem. On the other hand, using Bayesian inference can also offer an optimal estimation of the signal in the MMSE perspective. However, the calculation of the posterior distribution is intractable due to the large scale of the signal, which makes the estimation hard to be implemented. Generalized Expectation Consistent(GEC) method[2][3] proposes a new way to approximate the posterior distribution of the signal in a tractable way. The newly proposed GEC algorithm is considered to be a new approach to solve the phase retrieval problem. However, there is very little article which mentions the differences of performance between the widely used HIO algorithm and newly proposed GEC algorithm.

The purpose of this note is to discuss the performance of those two algorithms when they are used to reconstruct a Complex Gaussian signal. The comparison includes the measurement ratio, the MSE of the reconstructed signal, the convergence speed and robustness to the different initialization.

2 System Model

Consider a generalized linear model

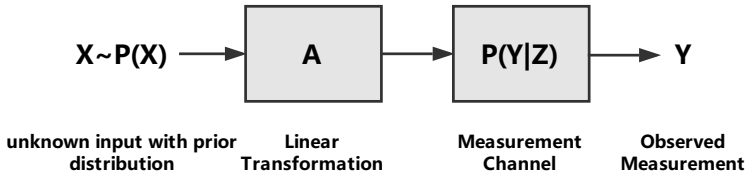

Figure 1: The generalized linear model

$\boldsymbol{x}$ is the unknown input with a prior distribution that we have already known. In this note, we consider $\boldsymbol{x}$ to be i.i.d with each entry following the Normalized Complex Gaussian distribution, i.e.$\boldsymbol{x} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$. $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ is the sensing matrix. $\boldsymbol{z}$ is the hidden output of the signal transform:

$$\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x} \tag{1}$$

$\boldsymbol{z}$ is measured through a noiseless channel and the output $\boldsymbol{y}$ is the modulus of $\boldsymbol{z}$:

$$\boldsymbol{y} = |\boldsymbol{z}| \tag{2}$$

The problem is to reconstruct the N dimension signal vector $\boldsymbol{x}$ through the M dimension observed measurement $\boldsymbol{y}$ when both M and N approaches infinity, with the measurement ratio,

$$\delta = \lim_{M,N\to+\infty} \frac{M}{N} \tag{3}$$

3 Hybrid Input Output Algorithm

This method considered two different kinds of constraints[4]: the support constraint $\boldsymbol{Sup}$ and the modulus constraint $\boldsymbol{Mod}$. Using an iterative method to find the signal $\boldsymbol{x}$ which satisfies:

$$\boldsymbol{x} = \boldsymbol{Sup} \cap \boldsymbol{Mod} \tag{4}$$

Denote that $\boldsymbol{z^t}$ is the reconstructed $\boldsymbol{z}$ in the $\boldsymbol{t}$ iteration. The support constraint can be expressed as:

$$\boldsymbol{z^t} \in \boldsymbol{range(M)} \tag{5}$$

The modulus constraint can be expressed as:

$$|\boldsymbol{z^t}| = \boldsymbol{y} \tag{6}$$

We can derive two projection onto $\boldsymbol{Sup}$ and $\boldsymbol{Mod}$:

$$\boldsymbol{P_S} = \boldsymbol{A}(\boldsymbol{A^H A})^{-1}\boldsymbol{A^H x} \tag{7}$$

$$\boldsymbol{P_M} = \boldsymbol{y}.*\boldsymbol{arg(z)} \tag{8}$$

$$\boldsymbol{arg(z)} = \frac{\boldsymbol{z}}{|\boldsymbol{z}|} \tag{9}$$

The iteration algorithm can be expressed as:

$$\boldsymbol{z^{t+1}} = \boldsymbol{z^t} + \boldsymbol{\beta P_M}(\boldsymbol{2P_S}(\boldsymbol{z^t}) - \boldsymbol{z^t}) - \boldsymbol{P_S}(\boldsymbol{z^t}) \tag{10}$$

where $\boldsymbol{\beta}$ is a damping parameter and it usually equals to 0.5. (Note that the final output of z should be $\boldsymbol{z^\infty} = \boldsymbol{P_M}(\boldsymbol{2P_S}(\boldsymbol{z^\infty}) - \boldsymbol{z^\infty})$.)

4 Generalized Expectation Consistent Algorithm

The GEC algorithm include three blocks and transfer extrinsic information between the blocks using the turbo method. Module A calculate the posterior mean and variance of $\boldsymbol{z}$ based on the measurement $\boldsymbol{y}$ and prior information $\boldsymbol{r_{1z}}$,$\boldsymbol{v_{1z}}$. Module C constraints the $\boldsymbol{z}$ and $\boldsymbol{x}$ in the linear space $\boldsymbol{z} = \boldsymbol{Ax}$. Module B calculate the posterior mean and variance of $\boldsymbol{x}$ using the prior distribution of $\boldsymbol{x}$.

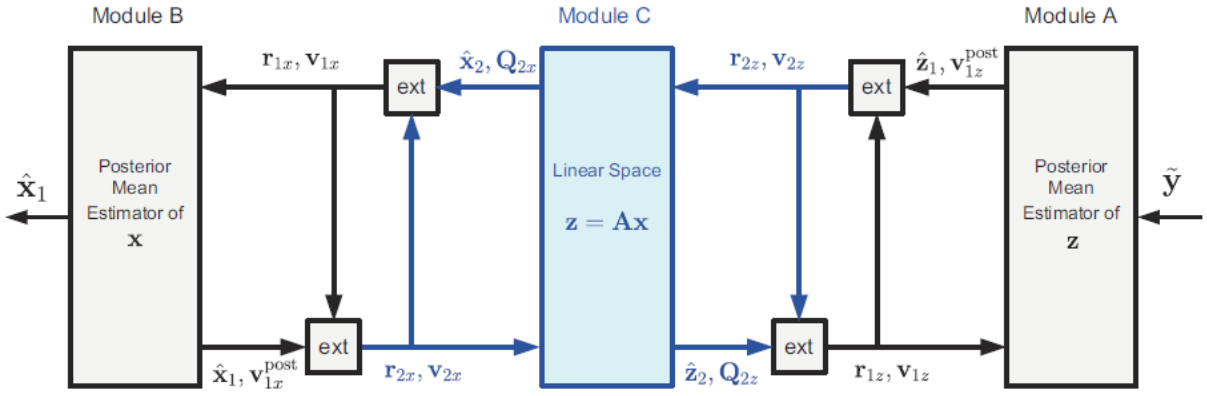

Figure 2: The block diagram of GEC algorithm[3]

Since it is the Gaussian signal that we are discussing. $\boldsymbol{r_{2z}} = 0, \boldsymbol{v_{2z}} = 1$ in each iteration and the Module C can be reduced. Thus, the reduced module can be derived as below.

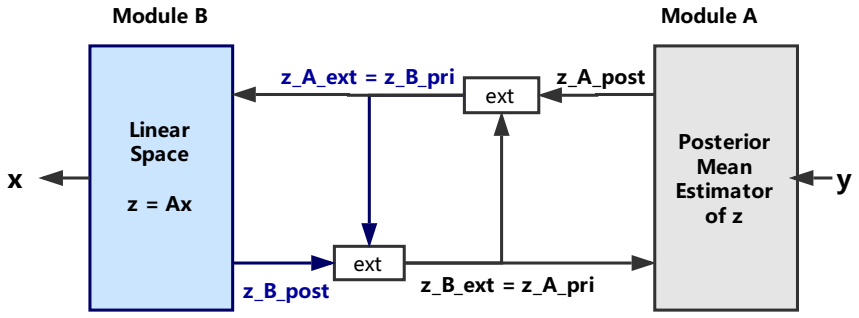

Figure 3: The reduced GEC module

Algorithm: General Expectation Consistent Algorithm

Initialization: for any $v_A^{pri} < \delta^{-1}$, $\boldsymbol{z}_A^{pri} \sim \mathbb{CN}(0, (\delta^{-1} - v_A^{pri}))$.
Note that the original signal $\boldsymbol{x} \sim \mathbb{CN}(0, 1)$ and as a result the corresponding $\boldsymbol{z} \sim \mathbb{CN}(0, \delta^{-1})$ under normalized $\boldsymbol{A}$
For *iteration* $= 1 : T_{max}$
1) Block A: calculate the posteriori mean and variance

$$\boldsymbol{z}_A^{post} = E[\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{z}_A^{pri}, v_A^{pri}] = yR_0(2\boldsymbol{y}|\boldsymbol{z}_A^{pri}|/v_A^{pri}) * arg(\boldsymbol{z}_A^{pri}), \tag{11a}$$

$$\boldsymbol{v}_A^{post} = Var[\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{z}_A^{pri}, v_A^{pri}] = \boldsymbol{y}^2 - |\boldsymbol{z}_A^{post}|^2, \tag{11b}$$

2) Calculate the extrinsic information of Block A

$$v_B^{pri} = v_A^{ext} = ((v_A^{post})^{-1} - (v_A^{pri})^{-1})^{-1}, \tag{12a}$$

$$\boldsymbol{z}_B^{pri} = \boldsymbol{z}_A^{ext} = v_A^{ext}(\boldsymbol{z}_A^{post}/v_A^{post} - \boldsymbol{z}_A^{pri}/v_A^{pri}), \tag{12b}$$

3) Block B: calculate the posteriori mean and variance

$$\boldsymbol{z}_B^{post} = ((\boldsymbol{v}_B^{pri})^{-1} + 1)^{-1}\boldsymbol{A}\boldsymbol{A}^{\boldsymbol{H}} z_B^{pri}, \tag{13a}$$

$$\boldsymbol{v}_B^{post} = (1 + (\boldsymbol{v}_B^{pri})^{-1})^{-1}\delta^{-1}, \tag{13b}$$

4) Calculate the extrinsic information of Block B

$$\boldsymbol{v}_A^{pri} = \boldsymbol{v}_B^{ext} = ((\boldsymbol{v}_B^{post})^{-1} - (\boldsymbol{v}_B^{pri})^{-1})^{-1}, \tag{14a}$$

$$\boldsymbol{z}_A^{pri} = \boldsymbol{z}_B^{ext} = \boldsymbol{v}_B^{ext}(\boldsymbol{z}_B^{post}/\boldsymbol{v}_B^{post} - \boldsymbol{z}_B^{pri}/\boldsymbol{v}_B^{pri}), \tag{14b}$$

Output: $\hat{\boldsymbol{x}} = (v_B^{pri} + 1)^{-1}\boldsymbol{A}^{\boldsymbol{H}}\boldsymbol{z}_B^{pri}$

5 Comparison

5.1 The Measurement Ratio

The two algorithms are initialized with the same random Gaussian signal, the iteration time is set to be 3000 times. The relationship between the measurement ratio δ and the success rate of the reconstruction is shown as below:

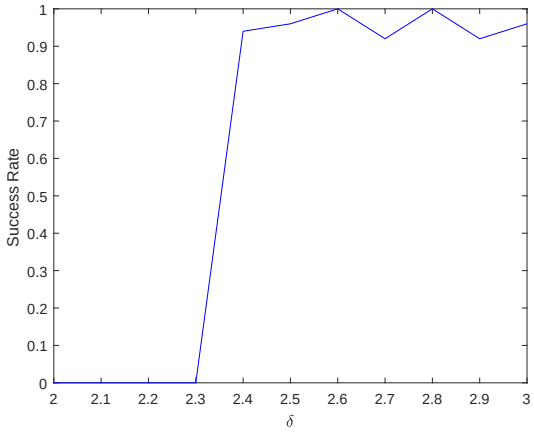

Figure 4: The change of success rate in GEC algorithm with the change of δ

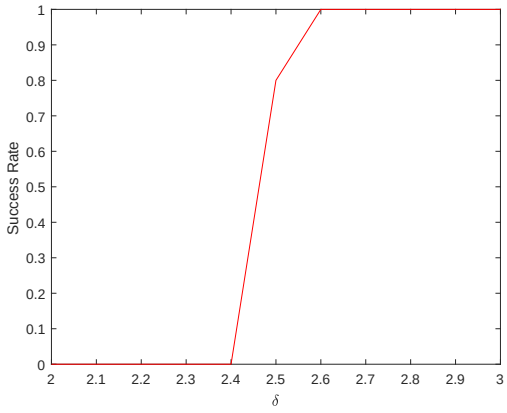

Figure 5: The change of success rate in HIO algorithm with the change of δ

From the plot, it is apparent that the necessary measurement ratio needed for both algorithm is almost the same and GEC is actually a little bit better than the HIO algorithm.

5.2 The Convergence Speed and MSE

Under the same circumstance, i.e. random initialization, the same measurement ratio $\delta = 3$. The relationship between MSE of reconstructed signal and iteration time is shown as below:

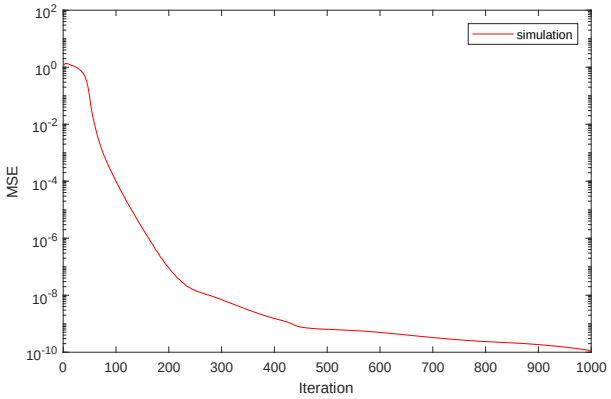

Figure 6: The relationship between iteration and MSE of HIO algorithm

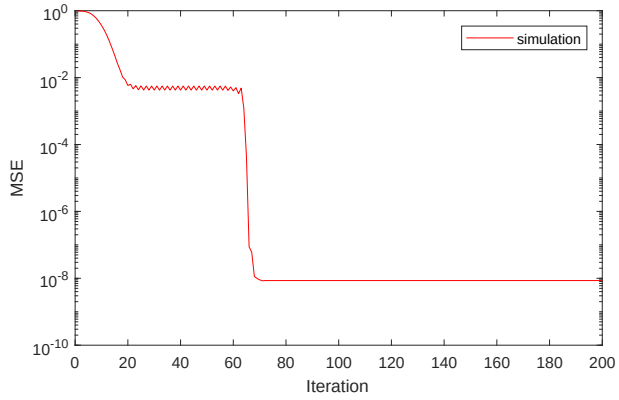

Figure 7: The relationship between iteration and MSE of GEC algorithm

While the final MSE of the both algorithm is almost identical, it is apparent that the convergence speed of the GEC algorithm is much faster than the HIO algorithm.

5.3 The Performance of Algorithm with the Change of Initialization

The performance of HIO and GEC algorithm can vary a lot under different initialization. To test the robustness of the algorithm with the change of initialization, the initialization is designed as below.

$$\hat{\boldsymbol{z}} = \frac{\frac{1}{\delta}}{(\frac{1}{\delta} + v_0)} * (\boldsymbol{z} + \sqrt{v_0}\mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})) \tag{15}$$

The performance of the algorithm is measured using the minimal δ that makes the algorithm converge. The change of v_0 denotes the "distance" between the initialization and original signal. The result is shown as below.

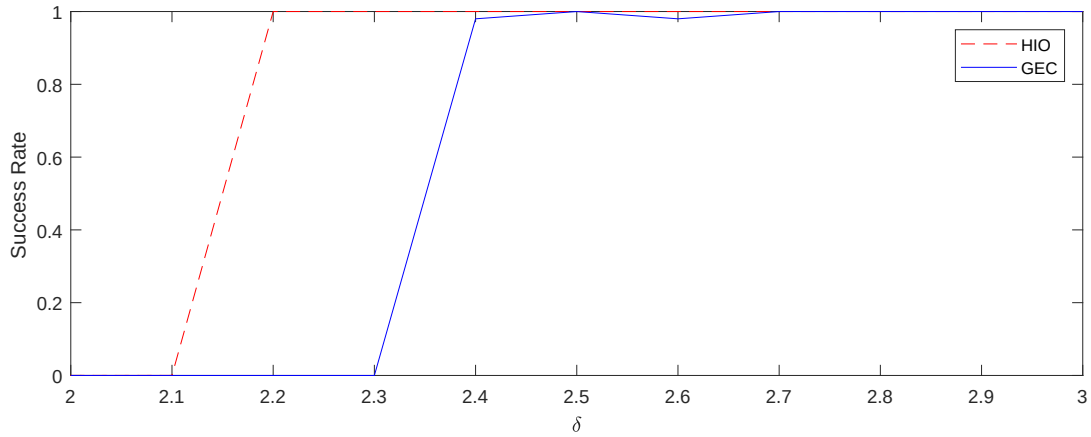

Figure 8: The Comparison of performance between two algorithms under $v_0 = 1$

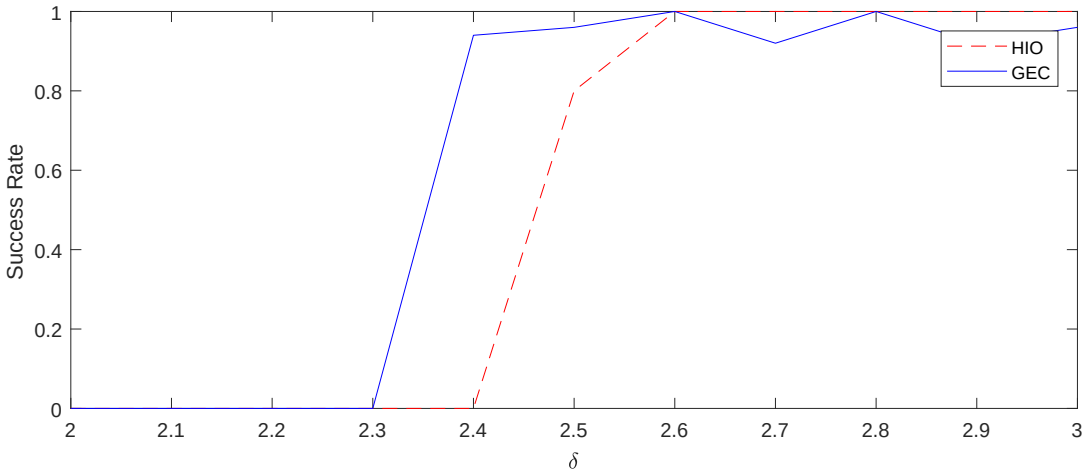

Figure 9: The Comparison of performance between two algorithms under $v_0 = 100$

Due to the space limit, the performance of both algorithm when v_0 is between 1 and 100 will not be shown in this note.

The performance of GEC algorithm does not vary too much with the change of v_0, while the performance of HIO algorithm change abruptly with the change of v_0. Under low v_0, the HIO algorithm perform better, but as v_0 increases, the performance of HIO algorithm can deteriorate rapidly.

6 State Evolution

For GEC algorithm, there actually exist a mathematical approach to predict the performance of the algorithm under different circumstance. The change of MSE with the iteration time, the fix point of convergence, the relationship between the measurement ratio and possibility of convergence can all be estimated through the state evolution[5].

6.1 Using State Evolution to Predict the Convergence of the Algorithm

The two blocks in Fig. 3 are treated separately and the relationship between the input and output of each models is tested, i.e. the variances $v_B^{ext} = v_A^{pri}, v_A^{ext} = v_B^{pri}$ are tested.

The input-output curve of those two modules are plotted using different method. Since there is no closed-form expression for module A, the monte carlo method is used: $\boldsymbol{z}$ is produced randomly and $\boldsymbol{z} \sim \mathbb{CN}(\boldsymbol{0}, \frac{1}{\delta}\boldsymbol{I})$, then $\boldsymbol{y}$ is produced following $\boldsymbol{y} = |\boldsymbol{z}|$. The expectation of the output of module A is calculated from a large quantity of random $\boldsymbol{z}$ under different input.

On the other hand, there exist closed-form expression between the input and output of module B, which is

$$v_B^{ext} = \frac{v_B^{pri}}{\delta v_B^{pri} + \delta - 1} \tag{16}$$

The proof of this closed form equation is listed in the appendixA. Then relationship between the input and output is shown as below.

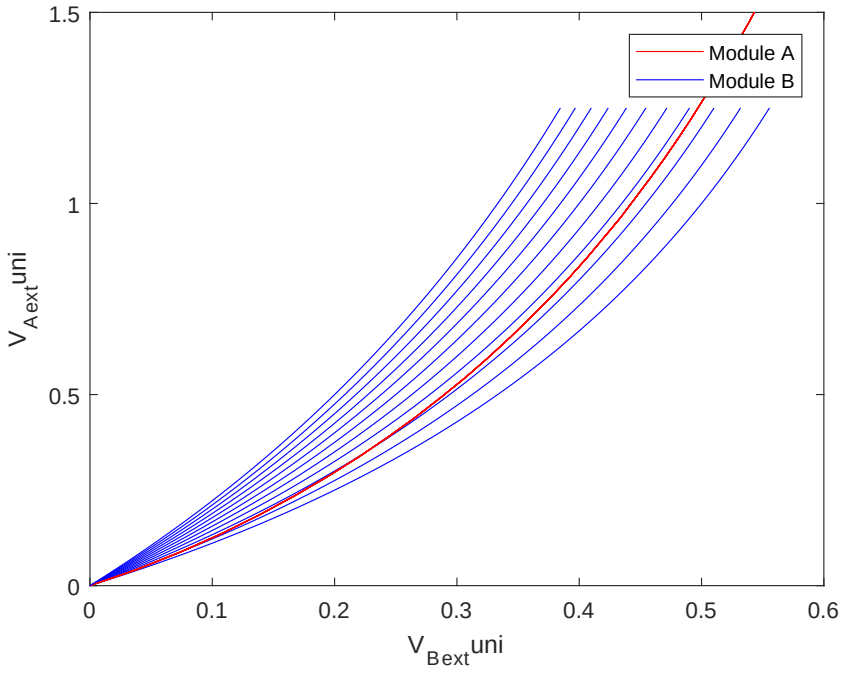

Figure 10: The state evolution curve

The blue curve denote module B while the red curve denote module A. The red curve is a unitary curve and does not change with δ. The blue curves increase with the increase of δ and the ten blue curves separately denote the performance of module B when

$$\delta = 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0 \tag{17}$$

the plot can be separated into three phases:

- Phase1: When the blue curve lies below the red curve, the iteration will diverge regardless of the initialization. From the plot we can conclude that it is impossible for GEC algorithm to reconstruct the signal when $\delta < 2.0$ no matter how close the initialization is to the true signal.
- Phase2: As the blue curve increase, it will intersect with the red curve when v_B^{ext} is close to 0, i.e. the initialization is close to the original signal enough, the algorithm will converge. $\delta > 2.0$ Corresponds to this phase.

- Phase3: When $\delta \geq 2.4$, the algorithm converges regardless of the initialization and the simulation Fig. 4. Shows the same result.

From the calculation of MATLAB, it can be analysed that $\delta = 2$ is the threshold that separate phase 1 and phase 2. When $\delta = 2$, the blue curve(Block B) lies completely below the red curve(Block A) in the interval $v_A^{ext} \in (0, 1)$ and two curve coincide when v_A^{ext} approach 0 and 1.

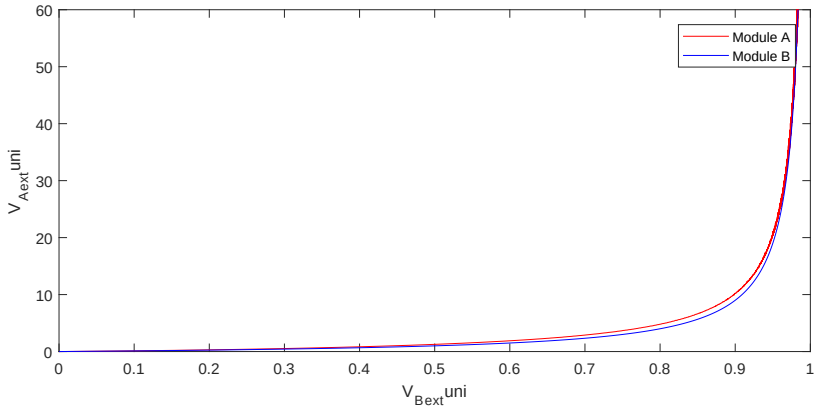

Figure 11: The relationship of two curves when $\delta = 2$

This result is of great important as it demonstrate the threshold that can distinguish

- When

$$\lim_{t \to +\infty} \mathbb{E}\left[\frac{\langle \hat{\boldsymbol{x}}^t, \boldsymbol{x}\rangle}{\|\hat{\boldsymbol{x}}^t\|_2 \|\boldsymbol{x}\|_2}\right] = 0 \tag{18}$$

- When there start to exist a convergence point locally when the initialized v_A^{pri} approaches zero.

This result is considered to be somehow provable and it is discussed in detail as below.

6.1.1 The Local Convergence of the Algorithm when v_A^{pri} Approaches 0

Theorem 1 *There exist the threshold $\delta = 2$, such that*

$$\left.\frac{\partial v_A^{ext}}{\partial v_A^{pri}}\right|_{v_A^{pri} \to 0} = \left.\frac{\partial v_B^{pri}}{\partial v_B^{ext}}\right|_{v_B^{ext} \to 0} \tag{19}$$

(Note that $v_A^{pri} = v_B^{ext}; v_B^{pri} = v_A^{ext}$.)

The proof of the *Theorem*1 is listed as below:

Proof 1 *Firstly we normalized the two blocks regarding to δ(for simplicity, we denote $\widetilde{v}$ as $\widetilde{v}_A^{pri}$)*

$$\widetilde{\boldsymbol{z}} = \sqrt{\delta}\boldsymbol{z}, \widetilde{\boldsymbol{z}} \sim \mathbb{CN}(\boldsymbol{0}, \boldsymbol{I}) \tag{20a}$$
$$\widetilde{\boldsymbol{y}} = |\widetilde{\boldsymbol{z}}| \tag{20b}$$
$$\widetilde{v} = \delta v \tag{20c}$$
$$\widetilde{\boldsymbol{z}}_A^{pri} = \sqrt{\delta}\boldsymbol{z}_A^{pri} = (1 - \widetilde{v})\widetilde{\boldsymbol{z}} + \sqrt{\widetilde{v}(1 - \widetilde{v})}\boldsymbol{w}, \boldsymbol{w} \sim \mathbb{CN}(\boldsymbol{0}, \boldsymbol{I}) \tag{20d}$$
$$\widetilde{v}_A^{ext} = \delta v_A^{ext} \tag{20e}$$
$$\widetilde{v}_B^{ext} = \delta v_B^{ext} \tag{20f}$$
$$\tag{20g}$$

Normalization separates δ from block A and put the effect of δ all onto block B. We can derive the normalized closed form equation of Module B.

$$\widetilde{v} = \frac{\widetilde{v}_B^{pri}}{\widetilde{v}_B^{pri} + \delta - 1} \tag{21}$$

Then to prove 19 is to prove:

$$\left.\frac{\partial \widetilde{v}_A^{ext}}{\partial \widetilde{v}}\right|_{\widetilde{v}\to 0} = \left.\frac{\partial \widetilde{v}_B^{pri}}{\partial \widetilde{v}}\right|_{\widetilde{v}\to 0} \tag{22}$$

Then we can derive the partial derivative of $\widetilde{v}_B^{pri}$ according to $\widetilde{v}$ when $\widetilde{v}$ approaches 0, which is

$$\left.\frac{\partial \widetilde{v}_B^{pri}}{\partial \widetilde{v}}\right|_{\widetilde{v}\to 0} = \left.\frac{\delta - 1}{1 - \widetilde{v}}\right|_{\widetilde{v}\to 0} = \delta - 1 \tag{23}$$

Then we just need to prove that when $\delta = 2$:

$$\left.\frac{\partial \widetilde{v}_A^{ext}}{\partial \widetilde{v}}\right|_{\widetilde{v}\to 0} = 1 \tag{24}$$

To prove the equation above, firstly we calculate the partial derivative of $\mathbb{E}[\widetilde{v}_A^{post}]$ regarding to $\widetilde{v}$ when $\widetilde{v}$ approaches 0.(Note that the state evolution of block A uses a monte carlo method, as a result all the variable in 44 except $\widetilde{v}$ should be the expectation of joint distribution of $\boldsymbol{z}$ and $\boldsymbol{z}_A^{pri}$). Using the equation listed in [6]:

$$\mathbb{E}[\widetilde{v}_A^{post}] = \mathbb{E}[|\widetilde{\boldsymbol{y}}|^2] - \mathbb{E}[|\widetilde{\boldsymbol{z}}_A^{post}|^2] \tag{25}$$

$$|\widetilde{\boldsymbol{z}}_A^{post}|^2 = \widetilde{\boldsymbol{y}}^2 R_0^2\left(\frac{2\widetilde{\boldsymbol{y}}|\widetilde{\boldsymbol{z}}_A^{pri}|}{\widetilde{v}}\right) \tag{26}$$

R_0 denotes the ratio of the first order modified bessel function and the zero order modified bessel function, i.e.

$$R_0(\rho) = \frac{I_1(\rho)}{I_0(\rho)} \tag{27}$$

Since $\widetilde{\boldsymbol{y}}$ is irrelevant to $\widetilde{v}$, we derive that

$$\lim_{\widetilde{v}\to 0}\frac{\partial \mathbb{E}[\widetilde{v}_A^{post}]}{\partial \widetilde{v}} = -\lim_{\widetilde{v}\to 0}\frac{\partial}{\partial \widetilde{v}}\left(\mathbb{E}\left[\widetilde{\boldsymbol{y}}^2 R_0^2\left(\frac{2\widetilde{\boldsymbol{y}}|\widetilde{\boldsymbol{z}}_{\boldsymbol{A}}^{\boldsymbol{pri}}|}{\widetilde{v}}\right)\right]\right) \tag{28}$$

Using the expansion in [7]:

$$\lim_{\rho\to+\infty} R_0(\rho) = 1 - \frac{1}{2\rho} + \mathcal{O}\left(\rho^{-1}\right) \tag{29}$$

We derive that

$$-\lim_{\widetilde{v}\to 0}\frac{\partial}{\partial \widetilde{v}}\left(\mathbb{E}\left[\widetilde{\boldsymbol{y}}^2 R_0^2\left(\frac{2\widetilde{\boldsymbol{y}}|\widetilde{\boldsymbol{z}}_A^{pri}|}{\widetilde{v}}\right)\right]\right) = -\lim_{\widetilde{v}\to 0}\frac{\partial}{\partial \widetilde{v}}\mathbb{E}\left[\left(\widetilde{\boldsymbol{y}} - \frac{\widetilde{v}}{4|\widetilde{\boldsymbol{z}}_A^{pri}|}\right)^2\right] \tag{30}$$

Expand $\widetilde{\boldsymbol{z}}_A^{pri}$ using 20d and we derive that:

$$\mathbb{E}\left[\left(\widetilde{\boldsymbol{y}}-\frac{\widetilde{v}}{4|\widetilde{\boldsymbol{z}}_A^{pri}|}\right)^2\right]=\mathbb{E}\left[\widetilde{\boldsymbol{y}}^2+\frac{\widetilde{v}^2}{16|\widetilde{\boldsymbol{z}}_A^{pri}|^2}-\frac{\widetilde{\boldsymbol{y}}\widetilde{v}}{2|\widetilde{\boldsymbol{z}}_A^{pri}|}\right] \tag{31a}$$

$$=\mathbb{E}[|\widetilde{\boldsymbol{z}}|^2]+\frac{\widetilde{v}^2}{16\mathbb{E}[|\widetilde{\boldsymbol{z}}_A^{pri}|^2]}-\frac{\widetilde{v}}{2}\mathbb{E}\left[\left|\frac{\widetilde{\boldsymbol{z}}}{\widetilde{\boldsymbol{z}}_A^{pri}}\right|\right] \tag{31b}$$

$$=\mathbb{E}[|\widetilde{\boldsymbol{z}}|^2]+\frac{\widetilde{v}^2}{16[(1-\widetilde{v})\widetilde{v}\mathbb{E}[\boldsymbol{w}\boldsymbol{w}^*]+(1-\widetilde{v})^2\mathbb{E}[\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^*]+0]} \tag{31c}$$

$$-\frac{\widetilde{v}}{2}\left(\frac{\mathbb{E}[\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^*]}{(1-\widetilde{v})^2\mathbb{E}[\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^*]+\widetilde{v}(1-\widetilde{v})\mathbb{E}[\boldsymbol{w}\boldsymbol{w}^*]+0}\right)^{\frac{1}{2}} \tag{31d}$$

$$=1+\frac{\widetilde{v}^2}{16[(1-\widetilde{v})\widetilde{v}+(1-\widetilde{v})^2]}-\frac{\widetilde{v}}{2}\left(\frac{1}{(1-\widetilde{v})^2+\widetilde{v}(1-\widetilde{v})}\right)^{\frac{1}{2}} \tag{31e}$$

$$=1+\frac{\widetilde{v}^2}{16(1-\widetilde{v})}-\frac{\widetilde{v}}{2}\left(\frac{1}{1-\widetilde{v}}\right)^{\frac{1}{2}} \tag{31f}$$

Then we can derive that

$$-\lim_{\widetilde{v}\to 0}\frac{\partial}{\partial\widetilde{v}}\mathbb{E}\left[\left(\widetilde{\boldsymbol{y}}-\frac{\widetilde{v}}{2|\widetilde{\boldsymbol{z}}_A^{pri}|}\right)^2\right]=-(0+0-\frac{1}{2})=\frac{1}{2} \tag{32}$$

As a result,

$$\lim_{\widetilde{v}\to 0}\frac{\partial\mathbb{E}[\widetilde{v}_A^{post}]}{\partial\widetilde{v}}=\frac{1}{2} \tag{33}$$

For Module A, it is easy to calculate from 12a that:

$$\frac{\partial\mathbb{E}[\widetilde{v}_A^{ext}]}{\partial\widetilde{v}}=\frac{1}{\left(\frac{\widetilde{v}}{\mathbb{E}[\widetilde{v}_A^{post}]}-1\right)^2}\left(\frac{\widetilde{v}^2}{\mathbb{E}[(\widetilde{v}_A^{post})]^2}\frac{\partial\mathbb{E}[\widetilde{v}_A^{post}]}{\partial\widetilde{v}}-1\right) \tag{34}$$

From [7] we know that

$$\lim_{\rho\to+\infty}R_0(\rho)=1 \tag{35}$$

Then it is easily to derive that:

$$\lim_{\widetilde{v}\to 0}\mathbb{E}[\widetilde{v}_A^{post}]=\lim_{\widetilde{v}\to 0}\left(\mathbb{E}[|\widetilde{\boldsymbol{y}}|^2]-\mathbb{E}[|\widetilde{\boldsymbol{z}}_A^{post}|^2]\right) \tag{36a}$$

$$=\lim_{\widetilde{v}\to 0}\left(\mathbb{E}[|\widetilde{\boldsymbol{y}}|^2]-\mathbb{E}\left[\widetilde{\boldsymbol{y}}^2R_0^2\left(\frac{2\widetilde{\boldsymbol{y}}|\widetilde{\boldsymbol{z}}_A^{pri}|}{\widetilde{v}}\right)\right]\right) \tag{36b}$$

$$=0 \tag{36c}$$

Since we have already known 33 and 36c, then we can expand $\mathbb{E}[\widetilde{v}_A^{post}]$ as a function of $\widetilde{v}$ using first-order Taylor series expansion, i.e.

$$\lim_{\widetilde{v}\to 0}\mathbb{E}[\widetilde{v}_A^{post}]=0+\frac{1}{2}\widetilde{v}+\mathcal{O}(\widetilde{v}) \tag{37}$$

Put 37 into 34,we derive that

$$\lim_{\widetilde{v}\to 0}\frac{\partial\mathbb{E}[\widetilde{v}_A^{ext}]}{\partial\widetilde{v}}=\frac{1}{(2-1)^2}\left(2^2*\frac{1}{2}-1\right)=1 \tag{38}$$

When $\delta=2$,

$$\left.\frac{\partial\mathbb{E}[\widetilde{v}_A^{ext}]}{\partial\widetilde{v}}\right|_{\widetilde{v}\to 0}=\left.\frac{\partial\widetilde{v}_B^{pri}}{\partial\widetilde{v}}\right|_{\widetilde{v}\to 0}=1 \tag{39}$$

The theorem is proved.

Theorem 1 tells us that when $\delta = 2$, the derivative of curve A coincide with that of curve B as $\widetilde{v}$ approaches 0. As δ increases, the derivative of block B, i.e. $\delta - 1$ also increases while the derivative of curve A remain the same. As a result, the curve B will no longer beneath the curve A in the small local area near zero and there will be an intersection point $\widetilde{v}_1$(unstable fix point) of two curves near zero and this will lead to a local convergence of the algorithm if the initialization lies within $(0, \widetilde{v}_1)$. On the other hand, if $\delta < 2.0$, then the derivative of curve B will be smaller than that of curve A and no local convergence area will be formed.

6.1.2 The Circumstance when v_A^{pri}Approaches 1

Theorem 2 *There exist the threshold $\delta = 2$, such that*

$$\left.\frac{\partial v_A^{ext}}{\partial v_A^{pri}}\right|_{v_A^{pri}\to 1} = \left.\frac{\partial v_B^{pri}}{\partial v_B^{ext}}\right|_{v_B^{ext}\to 1} \tag{40}$$

The proof of the *Theorem*2 is listed as below:

Proof 2 *We normalize both block A and block B as before. To prove 40 is to prove*

$$\left.\frac{\partial \widetilde{v}_A^{ext}}{\partial \widetilde{v}}\right|_{\widetilde{v}\to 1} = \left.\frac{\partial \widetilde{v}_B^{pri}}{\partial \widetilde{v}}\right|_{\widetilde{v}\to 1} \tag{41}$$

Then we can derive the partial derivative of $\widetilde{v}_B^{pri}$ according to $\widetilde{v}$, using 21, which is

$$\frac{\partial \widetilde{v}_B^{pri}}{\partial \widetilde{v}} = \frac{\delta - 1}{(1-\widetilde{v})^2}\left(= \frac{1}{(1-\widetilde{v})^2}, \delta = 2\right) \tag{42}$$

Then we just need to prove that under $\delta = 2$:

$$\left.\frac{\partial \widetilde{v}_A^{ext}}{\partial \widetilde{v}}\right|_{\widetilde{v}\to 1} = \frac{1}{(1-\widetilde{v})^2} \tag{43}$$

For Module A, it is easy to calculate from 12a that:

$$\frac{\partial \mathbb{E}[\widetilde{v}_A^{ext}]}{\partial \widetilde{v}} = \frac{1}{\left(\frac{\widetilde{v}}{\mathbb{E}[\widetilde{v}_A^{post}]} - 1\right)^2}\left(\frac{\widetilde{v}^2}{\mathbb{E}[(\widetilde{v}_A^{post})]^2}\frac{\partial \mathbb{E}[\widetilde{v}_A^{post}]}{\partial \widetilde{v}} - 1\right) \tag{44}$$

(Under monte carlo method, all the variable in block A except for the input $\widetilde{v}$ should be the expectation of joint distribution of $\boldsymbol{z}$ and $\boldsymbol{z}_A^{pri}$)

From [7], we can derive that:

$$\lim_{\rho\to 0} R_0(\rho) = 0 \tag{45}$$

In addition, from 20d, we know that $\widetilde{\boldsymbol{z}}_A^{pri}$ is a function of $\widetilde{v}$ and

$$\lim_{\widetilde{v}\to 1} \widetilde{\boldsymbol{z}}_A^{pri} = 0 \tag{46}$$

As a result,

$$\lim_{\widetilde{v}\to 1} \mathbb{E}[\widetilde{v}_A^{post}] = \lim_{\widetilde{v}\to 1}\left(\mathbb{E}[|\widetilde{\boldsymbol{y}}|^2] - \mathbb{E}[|\widetilde{\boldsymbol{z}}_A^{post}|^2]\right) \tag{47a}$$

$$= \mathbb{E}[|\widetilde{\boldsymbol{y}}|^2] - \lim_{\widetilde{v}\to 1}\mathbb{E}\left[\widetilde{\boldsymbol{y}}^2 R_0^2\left(\frac{2\widetilde{\boldsymbol{y}}|\widetilde{\boldsymbol{z}}_A^{pri}|}{\widetilde{v}}\right)\right] \tag{47b}$$

$$= \mathbb{E}[|\widetilde{\boldsymbol{z}}|^2] - 0 \tag{47c}$$

$$= 1 \tag{47d}$$

Then proving 43 can be transformed into proving

$$\frac{\partial \mathbb{E}[\widetilde{v}_A^{post}]}{\widetilde{v}} = 2 \tag{48}$$

Using 25 and 26, and since $\widetilde{\boldsymbol{y}}$ *is irrelevant to* $\widetilde{v}$*, we can derive that:*

$$\frac{\partial \mathbb{E}[\widetilde{v}_A^{post}]}{\partial \widetilde{v}} = -\frac{\partial \mathbb{E}[|\widetilde{\boldsymbol{z}}_A^{post}|^2]}{\partial \widetilde{v}} = -\frac{\partial \mathbb{E}\left[\widetilde{\boldsymbol{y}}^2 R_0^2\left(\frac{2\widetilde{\boldsymbol{y}}|\widetilde{\boldsymbol{z}}_A^{pri}|}{\widetilde{v}}\right)\right]}{\partial \widetilde{v}} \tag{49}$$

From 46, it is apparent that:

$$\lim_{\widetilde{v}\to 1} \frac{2\widetilde{\boldsymbol{y}}|\widetilde{\boldsymbol{z}}_A^{pri}|}{\widetilde{v}} = 0 \tag{50}$$

Under the first-order Taylor series expansion[7], the ratio of modified bessel function can be simplified as

$$\lim_{\widetilde{v}\to 1} R_0\left(\frac{2\widetilde{\boldsymbol{y}}|\widetilde{\boldsymbol{z}}_A^{pri}|}{\widetilde{v}}\right) = \frac{\widetilde{\boldsymbol{y}}|\widetilde{\boldsymbol{z}}_A^{pri}|}{\widetilde{v}} + \mathcal{O}\left(\frac{\boldsymbol{2}\widetilde{\boldsymbol{y}}|\widetilde{\boldsymbol{z}}_A^{pri}|}{\widetilde{v}}\right) \tag{51}$$

The problem is simplified into solving the expectation of a polynomial regarding to $\widetilde{\boldsymbol{y}}$*,* $\widetilde{\boldsymbol{z}}_A^{pri}$ *and* $\widetilde{v}$*, i.e. proving*

$$-\lim_{\widetilde{v}\to 1} \frac{\partial \mathbb{E}\left[\widetilde{\boldsymbol{y}}^4\left(\frac{|\widetilde{\boldsymbol{z}}_A^{pri}|}{\widetilde{v}}\right)^2\right]}{\partial \widetilde{v}} = -2 \tag{52}$$

and expand $\widetilde{\boldsymbol{z}}_A^{pri}$ *using 20d:*

$$|\widetilde{\boldsymbol{z}}_A^{pri}|^2 = (1-\widetilde{v})\widetilde{v}|\boldsymbol{w}|^2 + (1-\widetilde{v})^2|\widetilde{\boldsymbol{z}}|^2 + (1-\widetilde{v})\sqrt{(1-\widetilde{v})\widetilde{v}}(\widetilde{\boldsymbol{z}}^*\boldsymbol{w} + \boldsymbol{w}^*\widetilde{\boldsymbol{z}}) \tag{53}$$

Since $\widetilde{\boldsymbol{z}}$ *and* $\boldsymbol{w}$ *is irrelevant and both follow* $\mathbb{CN}(\boldsymbol{0}, \boldsymbol{I})$*, then we can derive that:*

$$-\mathbb{E}\left[\widetilde{\boldsymbol{y}}^4\left(\frac{|\widetilde{\boldsymbol{z}}_A^{pri}|}{\widetilde{v}}\right)^2\right] = -\mathbb{E}\left[|\widetilde{\boldsymbol{z}}|^6\frac{(1-\widetilde{v})^2}{\widetilde{v}^2}\right] - \mathbb{E}\left[\frac{(1-\widetilde{v})\widetilde{v}}{\widetilde{v}^2}|\widetilde{\boldsymbol{z}}|^4|w|^2\right] + 0 \tag{54a}$$

$$= -\left(\frac{1-\widetilde{v}}{\widetilde{v}}\right)^2\mathbb{E}[|\widetilde{\boldsymbol{z}}|^6] - \frac{\widetilde{v}(1-\widetilde{v})}{\widetilde{v}^2}\mathbb{E}[|\widetilde{\boldsymbol{z}}|^4] * 1 \tag{54b}$$

Since $\mathbb{E}[|\boldsymbol{z}|^6]$ *and* $\mathbb{E}[|\boldsymbol{z}|^4]$ *are the higher order moments of z under Rayleigh distribution, i.e.* $|\boldsymbol{z}| \sim Rayl(\frac{1}{\sqrt{2}})$*, the result is*

$$-\lim_{\widetilde{v}\to 1}\mathbb{E}\left[\widetilde{\boldsymbol{y}}^4\left(\frac{|\widetilde{\boldsymbol{z}}_A^{pri}|}{\widetilde{v}}\right)^2\right] = -\lim_{\widetilde{v}\to 1}(1-\widetilde{v})^2 * 6 + \widetilde{v}(1-\widetilde{v}) * 2 \tag{55}$$

Then it is simple to calculate the partial derivative of 55 regarding to $\widetilde{v}$

$$\lim_{\widetilde{v}\to 1}\frac{\partial \mathbb{E}[\widetilde{v}_A^{post}]}{\widetilde{v}} = \lim_{\widetilde{v}\to 1}\frac{\partial[-(1-\widetilde{v})^2 * 6 - \widetilde{v}(1-\widetilde{v}) * 2]}{\partial \widetilde{v}} = 2 \tag{56}$$

The theorem is proved.

The derivatives of both $\widetilde{v}_A^{ext}$ and $\widetilde{v}_B^{pri}$ under two modules are now proved to be the same when $\widetilde{v}$ approaches 1, then it can be easily proved that both $\widetilde{v}_A^{ext}$ and $\widetilde{v}_B^{pri}$ approach to the same number when $\widetilde{v}$ approaches 1. As δ increases, the derivative of curve B will increase and the derivative of curve A will remain the same. So there will exist an intersection point $\widetilde{v}_2$(stable point) in the vicinity of 1.

Thus, We can make an assumption from both theorem 1 and 2:

Assumption 1 *when* $\delta > 2$,

$$\lim_{t\to+\infty} \mathbb{E}\left[\frac{\langle \hat{\boldsymbol{x}}^t, \boldsymbol{x}\rangle}{\|\hat{\boldsymbol{x}}^t\|_2\|\boldsymbol{x}\|_2}\right] > 0 \tag{57}$$

On the other hand, when $\delta < 2$, *wherever the initialization points,*

$$\lim_{t\to+\infty} \mathbb{E}\left[\frac{\langle \hat{\boldsymbol{x}}^t, \boldsymbol{x}\rangle}{\|\hat{\boldsymbol{x}}^t\|_2\|\boldsymbol{x}\|_2}\right] = 0 \tag{58}$$

(t is the iteration of the algorithm)

Note that to completely prove the assumption above, another theorem also need to be proved:

Theorem 3 *When* $\delta \leq 2$, *it is always true that* $\widetilde{v}_B^{pri} \leq \widetilde{v}_A^{ext}$ *under the same* $\widetilde{v} \in (0,1)$.*(remains unproved)*

However, it can be easily noticed from statistical analysis of MATLAB11 that the curve of block B lies always below the curve of block A when $\delta = 2$. We can say that there is no intersection point lies within $(0,1)$ when $delta \leq 2$, as the two curves only coincide when $\widetilde{v}$ approaches 1 and 0 when $delta = 2$.
The numerical result from MATLAB coincide with the mathematical prediction from Assumption 1.

- When $\delta < 2$, curve B lies completely below curve A and the iteration of the algorithm will let $\widetilde{v}$ approaches 1, which leads to 58:

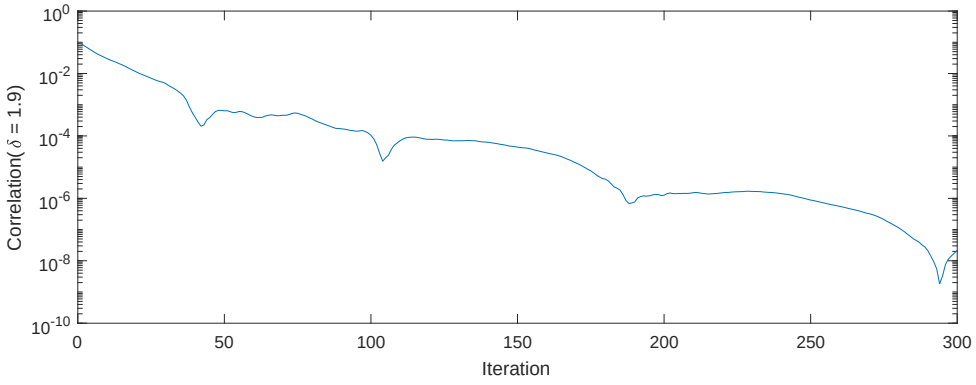

Figure 12: The correlation between the reconstructed signal and the original signal when $\delta = 1.9$

- When $\delta > 2$, there is an stable fix point in the vicinity of 1 and the iteration of the algorithm will let $\widetilde{v}$ approaches to the stable fix point, which leads to 57:

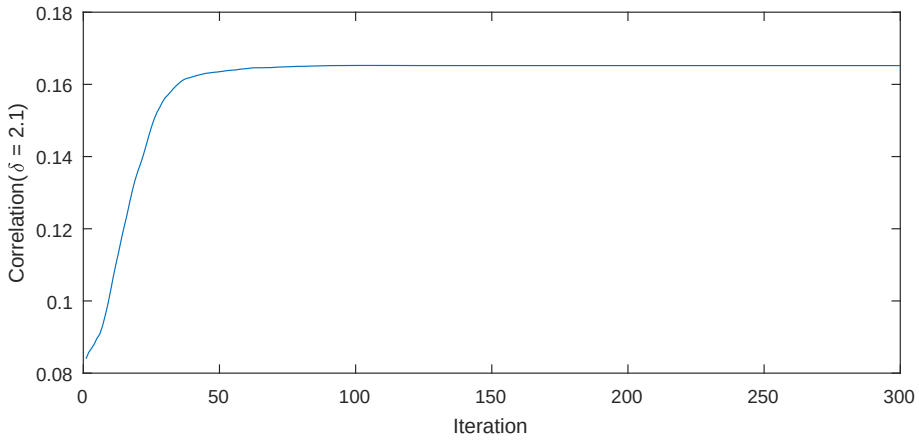

Figure 13: The correlation between the reconstructed signal and the original signal when $\delta = 2.1$

6.2 Using State Evolution to Predict the Change of MSE of the Algorithm

The changes of MSE with the iteration can also be analysed through the state evolution:

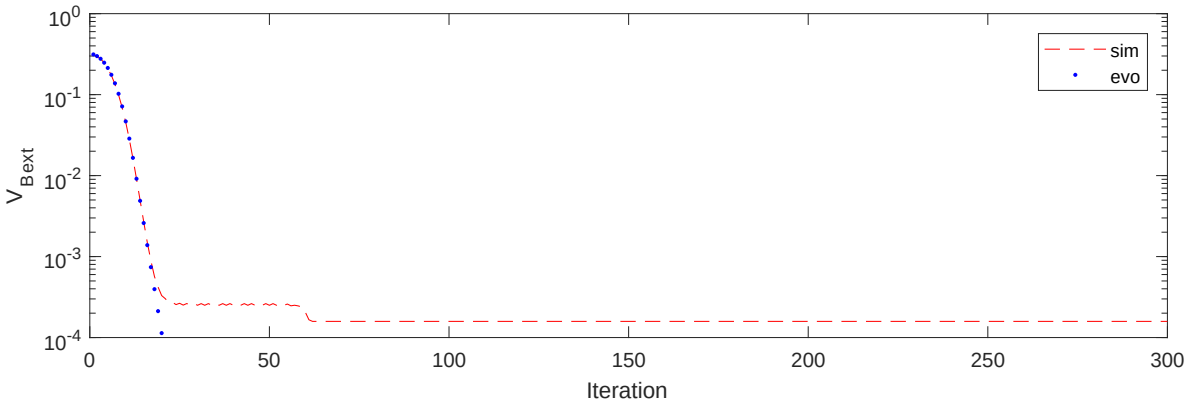

Figure 14: The change of MSE with iteration(semilogy)

From Fig. 14. we notice that the state evolution can accurately estimate the simulation until the MSE is below 10^{-5}. And the divergence of those two curves is considered to most probably result from the rounding of numbers in the MATLAB during the simulation process. The result is well enough for the real application.

A The Proof of Closed Form Equation of Block B

From 13a 13b 14a 14b, we derive the closed form equation of $\boldsymbol{z}_B^{ext}$ regarding to $\boldsymbol{z}_B^{pri}$ and v_B^{pri}:

$$\boldsymbol{z}_B^{ext} = (\delta v_B^{pri} + \delta - 1)^{-1}(\delta \boldsymbol{A}\boldsymbol{A}^{\boldsymbol{H}} \boldsymbol{z}_B^{pri} - \boldsymbol{z}_B^{pri}) \tag{59}$$

For ease of notation, we simplify the equation above as:

$$\boldsymbol{z}_B^{ext} = c\boldsymbol{B}\boldsymbol{z}_B^{pri}, \tag{60a}$$

$$c = (\delta v_B^{pri} + \delta - 1)^{-1}, \tag{60b}$$

$$B = (\delta \boldsymbol{A}\boldsymbol{A}^{\boldsymbol{H}} - \boldsymbol{I}). \tag{60c}$$

From [3], we know that

$$\boldsymbol{z}_B^{pri} = \boldsymbol{A}\boldsymbol{x} + \sigma\boldsymbol{w}, \boldsymbol{w} \sim \mathbb{CN}(\boldsymbol{0}, \boldsymbol{I}), \sigma = v_B^{pri} \tag{61}$$

As a result,

$$\boldsymbol{z}_B^{ext} = c\boldsymbol{B}\boldsymbol{A}\boldsymbol{x} + c\sigma\boldsymbol{B}\boldsymbol{w} \tag{62}$$

Using

$$\boldsymbol{B}\boldsymbol{A} = (\delta \boldsymbol{A}\boldsymbol{A}^{\boldsymbol{H}} - \boldsymbol{I})\boldsymbol{A} = (\delta - 1)A \tag{63}$$

In addition, it is also obvious that $c\sigma\boldsymbol{B}\boldsymbol{w}$ is still a Complex Guassian signal and we can note it as:

$$c\sigma\boldsymbol{B}\boldsymbol{w} = \lambda\widetilde{\boldsymbol{w}}, \widetilde{\boldsymbol{w}} \sim \mathbb{CN}(\boldsymbol{0}, \boldsymbol{I}) \tag{64}$$

And we can calculate the λ^2:

$$\lambda^2 = Var[c\sigma\boldsymbol{B}\boldsymbol{w}] \tag{65a}$$

$$= E[\|\sigma\boldsymbol{B}\boldsymbol{w}\|^2] - 0 \tag{65b}$$

$$= c^2\sigma^2 \lim_{M\to+\infty} \frac{tr(\boldsymbol{B}\boldsymbol{w}\boldsymbol{w}^{\boldsymbol{H}}\boldsymbol{B}^{\boldsymbol{H}})}{M} \tag{65c}$$

$$= c^2\sigma^2 \lim_{M\to+\infty} \frac{tr((\delta \boldsymbol{A}\boldsymbol{A}^{\boldsymbol{H}} - \boldsymbol{I})(\delta \boldsymbol{A}\boldsymbol{A}^{\boldsymbol{H}} - \boldsymbol{I})^H)}{M} \tag{65d}$$

$$= c^2\sigma^2 \lim_{M\to+\infty} \frac{(\delta^2 - 2\delta)tr(\boldsymbol{A}^{\boldsymbol{H}}\boldsymbol{A}) + M}{M} \tag{65e}$$

$$= c^2\sigma^2(\delta - 1) \tag{65f}$$

We can simplify the notation in the equation of $\boldsymbol{z}_B^{ext}$:

$$\boldsymbol{z}_B^{ext} = c(\delta - 1)\boldsymbol{A}\boldsymbol{x} + \lambda\widetilde{\boldsymbol{w}} = c(\delta - 1)\boldsymbol{z} + \lambda\widetilde{\boldsymbol{w}} \tag{66}$$

Using the definition of the variance and 60b 65f, we derive that:

$$\boldsymbol{v}_B^{ext} = \frac{\|\boldsymbol{z}_B^{ext} - \boldsymbol{z}\|^2}{M} \tag{67a}$$

$$= \frac{\|(c\delta - c - 1)\boldsymbol{z} + \lambda\widetilde{\boldsymbol{w}}\|^2}{M} \tag{67b}$$

$$= \frac{(c\delta - c - 1)^2(\boldsymbol{x}^{\boldsymbol{H}}\boldsymbol{A}^{\boldsymbol{H}}\boldsymbol{A}\boldsymbol{x}) + \lambda^2\widetilde{\boldsymbol{w}}^{\boldsymbol{H}}\widetilde{\boldsymbol{w}}}{M} \tag{67c}$$

$$= (c\delta - c - 1)^2 * \delta^{-1} + \lambda^2 * 1 \tag{67d}$$

$$= \frac{v_B^{pri}}{\delta v_B^{pri} + \delta - 1} \tag{67e}$$

The closed form equation is proved.

References

[1] J. R. Fienup, "Phase retrieval algorithms: a comparison," *Appl. Opt.*, vol. 21, no. 15, pp. 2758–2769, Aug 1982. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-21-15-2758

[2] A. Fletcher, M. Sahraee-Ardakan, S. Rangan, and P. Schniter, "Expectation consistent approximate inference: Generalizations and convergence," in *2016 IEEE International Symposium on Information Theory (ISIT)*, July 2016, pp. 190–194.

[3] H. He, C. Wen, and S. Jin, "Generalized expectation consistent signal recovery for nonlinear measurements," *CoRR*, vol. abs/1701.04301, 2017. [Online]. Available: http://arxiv.org/abs/1701.04301

[4] H. H Bauschke, P. Combettes, and D. Luke, "Phase retrieval, gerchberg-saxton algorithm, and fienup variants: A view from convex optimization," vol. 19, pp. 1334–45, 08 2002.

[5] J. Ma, X. Yuan, and L. Ping, "Turbo compressed sensing with partial dft sensing matrix," *IEEE Signal Processing Letters*, vol. 22, no. 2, pp. 158–161, Feb 2015.

[6] P. Schniter and S. Rangan, "Compressive phase retrieval via generalized approximate message passing," *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 1043–1055, Feb 2015.

[7] C. Robert, "Modified bessel functions and their applications in probability and statistics," *Statistics and Probability Letters*, vol. 9, no. 2, pp. 155 – 161, 1990. [Online]. Available: http://www.sciencedirect.com/science/article/pii/016771529290011S