Matrix Factorization and Applications

Reminder

We have studied in class the following problem: given a vector $\mathbf{x}^*$, where each component x_i^* is taken from $P_X(x)$, one is given a noisy symmetric $N \times N$ matrix $\mathbf{Y}$ such that

$$Y_{ij} = \sqrt{\frac{\lambda}{N}} x_i^* x_j^* + \omega_{ij} \tag{1}$$

where λ is the signal to noise ratio, and where $\omega_{ij} = \omega_{ji}$ is a noise taken from a Gaussian distribution $\mathcal{N}(0,1)$. We have seen that the free entropy reads $\Phi(\lambda) = \min_m \Phi_{\rm RS}(m;\lambda)$ where $m \geq 0$ and

$$\Phi_{\rm RS}(m;\lambda) = -\frac{\lambda m^2}{4} + \int P_X(x^*)\,\mathrm{d}x^* \int \frac{\mathrm{e}^{-z^2/2}}{\sqrt{2\pi}}\,\mathrm{d}z \, \log\left\{\int P_X(x)\,\mathrm{d}x \, \exp\left(-\frac{\lambda m}{2}x^2 + \lambda m x^* x + \sqrt{\lambda m} z x\right)\right\}$$

and the maximizer $m^*(\lambda)$ is the magnetization $m = \mathbb{E}_{\mathbf{x}^*,\boldsymbol{\omega}}\left[\left\langle \frac{1}{N}\sum_i x_i^* x_i \right\rangle_{\mathbf{x}^*,\boldsymbol{\omega}}\right]$. We remind that the Minimum Mean Square Error is given by

$$\mathsf{MMSE}(\lambda) = \mathbb{E}_{x^*}\left[(x^*)^2\right] - m^*(\lambda)$$

A. Application to a binary estimation problem

(a) Assume now that each $x_i^* = \pm 1$ with equal probabilities, and show that one can then drastically simplify the expression of the free entropy such that only one integral (over z) remains.

(b) Compute (numerically) the MMSE as function of λ and show that there is a phase transition at $\lambda = 1$.

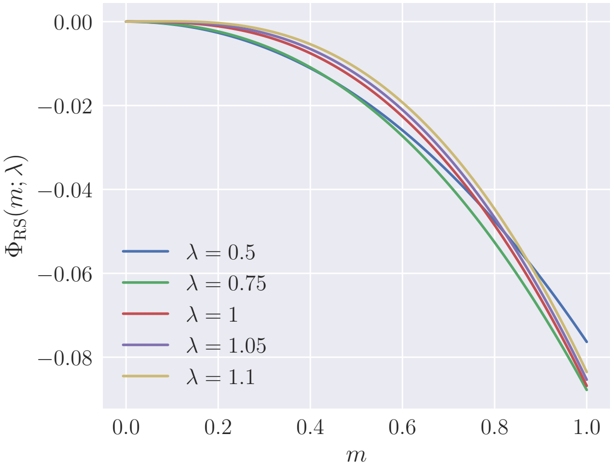
Solution.

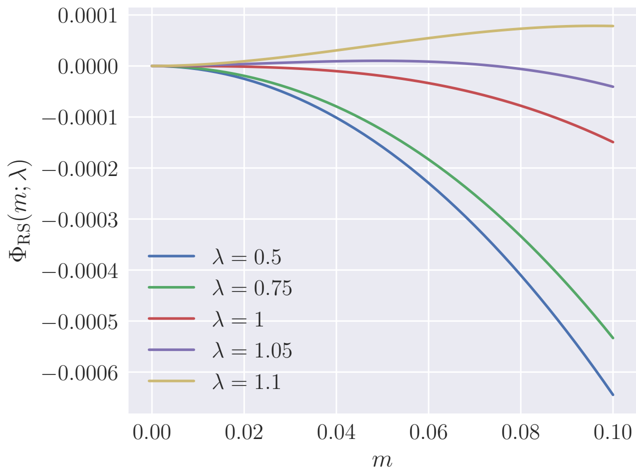
(a) Since now we have $P_X(x) = \frac{1}{2}\left[\delta(x-1) + \delta(x+1)\right]$, we can simplify as

$$\begin{aligned}
&\int P_X(x)\,\mathrm{d}x \, \exp\left(-\frac{\lambda m}{2}x^2 + \lambda m x^* x + \sqrt{\lambda m} z x\right) \\
&\qquad = \frac{1}{2}\left[\exp\left(-\frac{\lambda m}{2} + \lambda m x^* + \sqrt{\lambda m} z\right) + \exp\left(-\frac{\lambda m}{2} - \lambda m x^* - \sqrt{\lambda m} z\right)\right] \\
&\qquad = \mathrm{e}^{-\frac{\lambda m}{2}} \cosh\left(\lambda m x^* + \sqrt{\lambda m} z\right)
\end{aligned}$$

$$\begin{aligned}
\Phi_{\rm RS}(m;\lambda) &= -\frac{\lambda m^2}{4} + \int P_X(x^*)\,\mathrm{d}x^* \int \frac{\mathrm{e}^{-z^2/2}}{\sqrt{2\pi}}\,\mathrm{d}z \, \log\left[\mathrm{e}^{-\frac{\lambda m}{2}} \cosh\left(\lambda m x^* + \sqrt{\lambda m} z\right)\right] \\
&= -\frac{\lambda m^2}{4} + \int P_X(x^*)\,\mathrm{d}x^* \int \frac{\mathrm{e}^{-z^2/2}}{\sqrt{2\pi}}\,\mathrm{d}z \, \left[-\frac{\lambda m}{2} + \log\left\{\cosh\left(\lambda m x^* + \sqrt{\lambda m} z\right)\right\}\right] \\
&= -\frac{\lambda m^2}{4} + \frac{1}{2}\int \frac{\mathrm{e}^{-z^2/2}}{\sqrt{2\pi}}\,\mathrm{d}z \, \left[-\frac{\lambda m}{2} + \log\left\{\cosh\left(\lambda m + \sqrt{\lambda m} z\right)\right\}\right] \\
&\quad + \frac{1}{2}\int \frac{\mathrm{e}^{-z^2/2}}{\sqrt{2\pi}}\,\mathrm{d}z \, \left[-\frac{\lambda m}{2} + \log\left\{\cosh\left(-\lambda m + \sqrt{\lambda m} z\right)\right\}\right] \\
&= -\frac{\lambda m^2}{4} - \frac{\lambda m}{2} + \frac{1}{2}\mathbb{E}_z\left[\log\left\{\frac{1}{2}\left(\cosh(2\lambda m) + \cosh(2z\sqrt{\lambda m})\right)\right\}\right]
\end{aligned}$$

(b) The $\Phi_{\text{RS}}(m;\lambda)$ function is plotted over range $m \in [0,1]$ and $m \in [0, 0.1]$ with $\lambda \in \{0.5, 0.75, 1, 1.05, 1.1\}$.





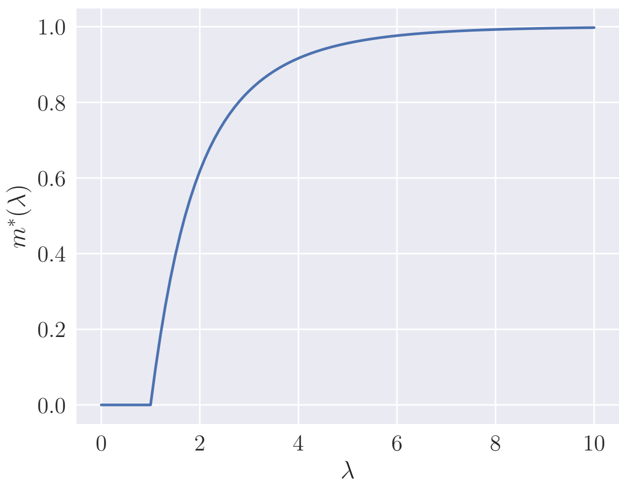
To compute m^* as a function of λ, notice that it is the maximizer of $\Phi_{\text{RS}}(m;\lambda)$, so we have the consistency equation by solving

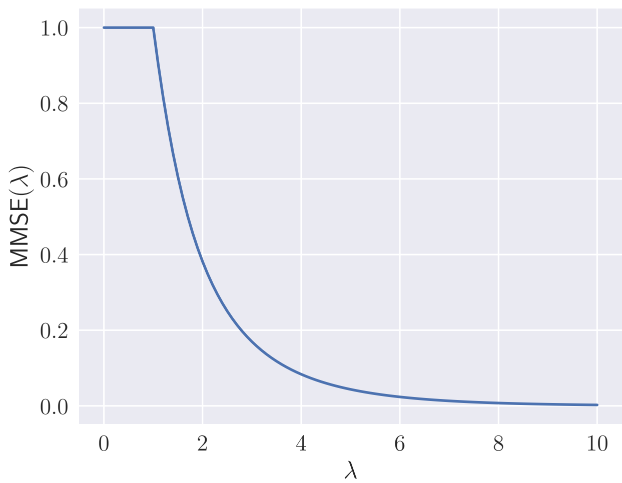
$$0 = \frac{\partial}{\partial m}\Phi_{\text{RS}}(m;\lambda) = -\frac{\lambda m}{2} - \frac{\lambda}{2} + \frac{1}{2}\mathbb{E}_z\left[\frac{2\lambda\sinh\left(2\lambda m\right) + \frac{\lambda z}{\sqrt{\lambda m}}\sinh\left(2z\sqrt{\lambda m}\right)}{\cosh(2\lambda m) + \cosh(2z\sqrt{\lambda m})}\right]$$

$$\Rightarrow \quad m = \mathbb{E}_z\left[\frac{2\sinh\left(2\lambda m\right) + \frac{z}{\sqrt{\lambda m}}\sinh\left(2z\sqrt{\lambda m}\right)}{\cosh(2\lambda m) + \cosh(2z\sqrt{\lambda m})}\right] - 1$$

Finally, for the MMSE, it is straightforward to see

$$\mathbb{E}_{x^*}\left[(x^*)^2\right] = 1 \quad \Rightarrow \quad \mathsf{MMSE}(\lambda) = \mathbb{E}_{x^*}\left[(x^*)^2\right] - m^*(\lambda) = 1 - m^*(\lambda)$$





Again, it is easy to see that $m^*(\lambda)$ leaves 0 continuously when $\lambda \geq 1$, indicating a second-order phase transition.

$\square$

B. Comparison to a spectral algorithm

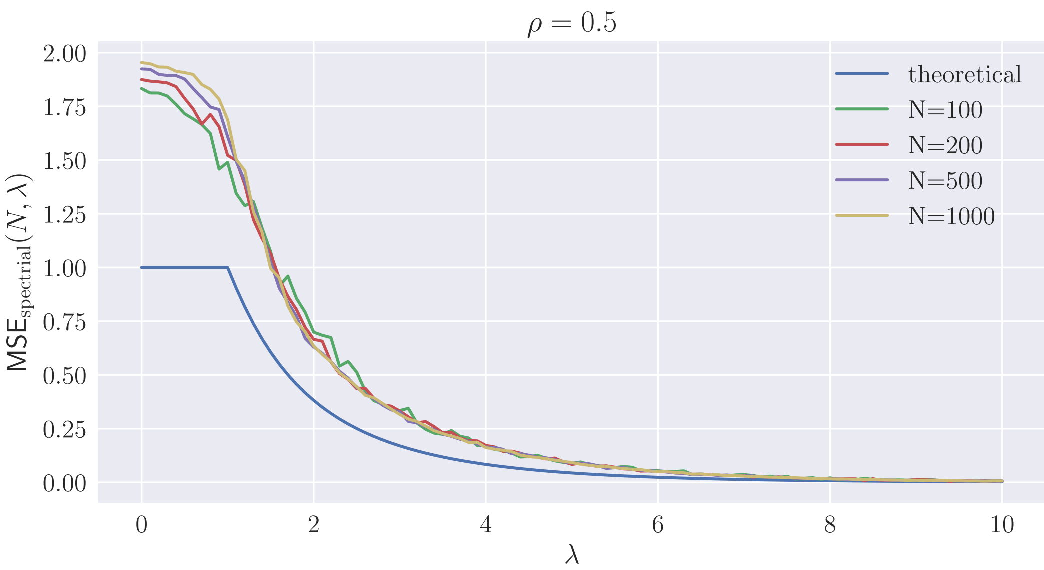
Let us compare the theoretical performance with the one of a particular algorithm: spectral clustering. In spectral clustering, we estimate the unknown $\mathbf{x}^*$ as follow: (i) we compute the leading (largest) eigenvalue of $\mathbf{Y}$ and (ii) we estimate $\hat{x}_i = \mathsf{sign}(v_i)$, where $\mathbf{v}$ is the eigenvector corresponding the largest eigenvalue.

For several values of N compute the mean square error of the spectral method as a function of λ (note that since we cannot know if the solution is $\hat{\mathbf{x}}$ or $-\hat{\mathbf{x}}$ we should try both):

$$\mathsf{MSE}_{\text{spectral}}(N, \lambda) = \min\left(\frac{1}{N}\sum_i (x_i^* - \hat{x}_i)^2, \frac{1}{N}\sum_i (x_i^* + \hat{x}_i)^2\right)$$

and compare $\mathsf{MSE}_{\text{spectral}}(N, \lambda)$ with the theoretical predictions at $N \to \infty$. How would you rate the performance of the spectral algorithm?

Solution. For the simulation, I choose $N \in \{100, 200, 500, 1000\}$, over $\lambda \in [0, 10]$, each data point is averaged over 100 trials.



It can be seen that the MSE of spectral algorithm is about twice larger than the theoretical curve. This is because the MMSE estimation uses posterior mean, which is a 'soft' decision, to give a explicit clustering partition one needs to make hard decision over those posterior mean. Spectral algorithm, on the other hand, gives hard decision directly, where the event $\hat{x}_i = 0$ occurs with probability zero.

Actually, to give a fair comparison, one needs to implement AMP algorithm to compute 'soft' $\hat{\mathbf{x}}$, make hard decision based on it and then compute the MSE of AMP algorithm. □

C. A sparse variation

Repeat the previous analysis when $x_i^* = \sqrt{(1-\rho)/\rho}$ with probability $\rho = 0.075$ and $-\sqrt{\rho/(1-\rho)}$ with probability $1 - \rho = 0.925$. Adapt the clustering algorithm in this case, and compare its performance with the theoretical curve. How would you rate the performance of the spectral algorithm this time?

Solution. Using similar tricks as the first problem, we obtain

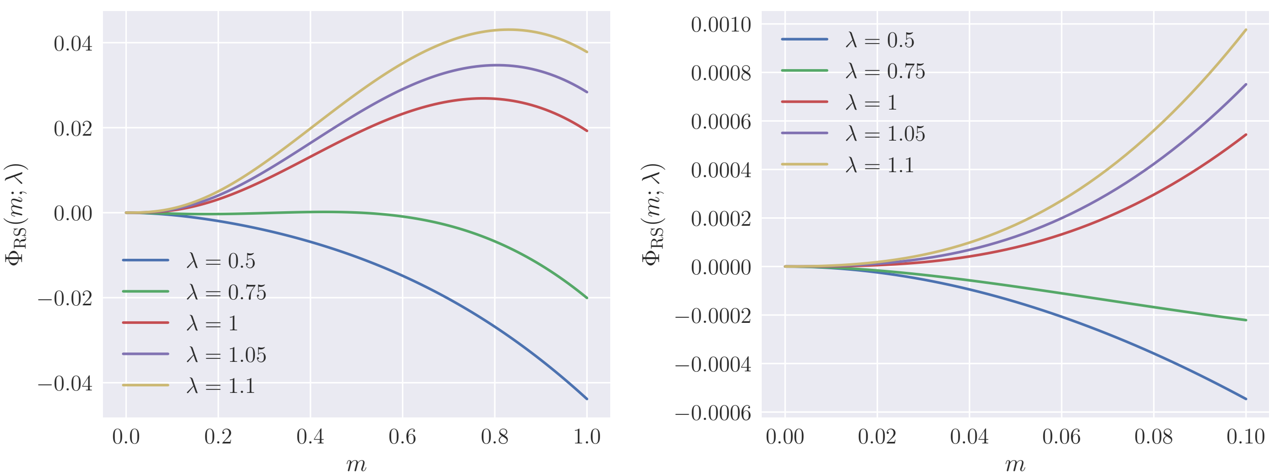
$$
\begin{aligned}
\int P_X(x)\,\mathrm{d}x\ \ \exp\left(-\frac{\lambda m}{2}x^2+\lambda m x^* x+\sqrt{\lambda m}zx\right)
&= \rho\exp\left(-\frac{\lambda m}{2}\frac{1-\rho}{\rho}+\lambda m\sqrt{\frac{1-\rho}{\rho}}x^*+\sqrt{\lambda m}\sqrt{\frac{1-\rho}{\rho}}z\right)\\
&\quad+(1-\rho)\exp\left(-\frac{\lambda m}{2}\frac{\rho}{1-\rho}-\lambda m\sqrt{\frac{\rho}{1-\rho}}x^*-\sqrt{\lambda m}\sqrt{\frac{\rho}{1-\rho}}z\right)
\end{aligned}
$$

Unfortunately, there is no easy way to simplify this expression to a more compact form.

Plugging in the above results yields

$$
\begin{aligned}
\Phi_{\mathrm{RS}}(m;\lambda) = &-\frac{\lambda m^2}{4}\\
&+\rho\,\mathbb{E}_z\left[\log\left\{\rho\exp\left(\frac{\lambda m}{2}\frac{1-\rho}{\rho}+\sqrt{\lambda m}\sqrt{\frac{1-\rho}{\rho}}z\right)\right.\right.\\
&\qquad\left.\left.+(1-\rho)\exp\left(-\frac{\lambda m}{2}\left(\frac{\rho}{1-\rho}+2\right)-\sqrt{\lambda m}\sqrt{\frac{\rho}{1-\rho}}z\right)\right\}\right]\\
&+(1-\rho)\,\mathbb{E}_z\left[\log\left\{\rho\exp\left(-\frac{\lambda m}{2}\left(\frac{1-\rho}{\rho}+2\right)+\sqrt{\lambda m}\sqrt{\frac{1-\rho}{\rho}}z\right)\right.\right.\\
&\qquad\left.\left.+(1-\rho)\exp\left(\frac{\lambda m}{2}\frac{\rho}{1-\rho}-\sqrt{\lambda m}\sqrt{\frac{\rho}{1-\rho}}z\right)\right\}\right]
\end{aligned}
$$

The $\Phi_{\mathrm{RS}}(m;\lambda)$ function is plotted over range $m\in[0,1]$ and $m\in[0,0.1]$ with $\lambda\in\{0.5,0.75,1,1.05,1.1\}$.



Similarly, to compute $m^*(\lambda)$, we need to find the self-consistency equation

$$
0=\frac{\partial}{\partial m}\Phi_{\mathrm{RS}}(m;\lambda)=-\frac{\lambda m}{2}+\mathbb{E}_z\left[\rho T_1+(1-\rho)T_2\right]
$$

$$
T_1=\frac{\left(\frac{\lambda(1-\rho)}{2}+\frac{z}{2}\sqrt{\frac{\lambda\rho(1-\rho)}{m}}\right)\mathrm{e}^{\frac{\lambda m}{2}\frac{1-\rho}{\rho}+\sqrt{\lambda m}\sqrt{\frac{1-\rho}{\rho}}z}+\left(-\frac{\lambda(2-\rho)}{2}-\frac{z}{2}\sqrt{\frac{\lambda\rho(1-\rho)}{m}}\right)\mathrm{e}^{-\frac{\lambda m}{2}\left(\frac{\rho}{1-\rho}+2\right)-\sqrt{\lambda m}\sqrt{\frac{\rho}{1-\rho}}z}}{\rho\mathrm{e}^{\frac{\lambda m}{2}\frac{1-\rho}{\rho}+\sqrt{\lambda m}\sqrt{\frac{1-\rho}{\rho}}z}+(1-\rho)\mathrm{e}^{-\frac{\lambda m}{2}\left(\frac{\rho}{1-\rho}+2\right)-\sqrt{\lambda m}\sqrt{\frac{\rho}{1-\rho}}z}}
$$

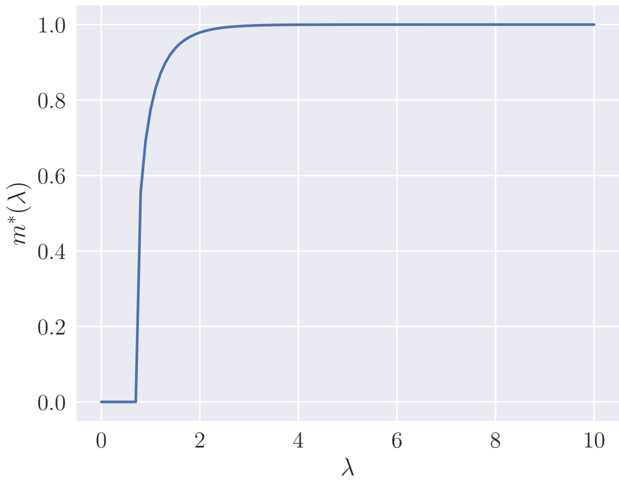
$$
T_2=\frac{\left(-\frac{\lambda(1+\rho)}{2}+\frac{z}{2}\sqrt{\frac{\lambda\rho(1-\rho)}{m}}\right)\mathrm{e}^{-\frac{\lambda m}{2}\left(\frac{1-\rho}{\rho}+2\right)+\sqrt{\lambda m}\sqrt{\frac{1-\rho}{\rho}}z}+\left(\frac{\lambda\rho}{2}-\frac{z}{2}\sqrt{\frac{\lambda\rho(1-\rho)}{m}}\right)\mathrm{e}^{\frac{\lambda m}{2}\frac{\rho}{1-\rho}-\sqrt{\lambda m}\sqrt{\frac{\rho}{1-\rho}}z}}{\rho\mathrm{e}^{-\frac{\lambda m}{2}\left(\frac{1-\rho}{\rho}+2\right)+\sqrt{\lambda m}\sqrt{\frac{1-\rho}{\rho}}z}+(1-\rho)\mathrm{e}^{\frac{\lambda m}{2}\frac{\rho}{1-\rho}-\sqrt{\lambda m}\sqrt{\frac{\rho}{1-\rho}}z}}
$$

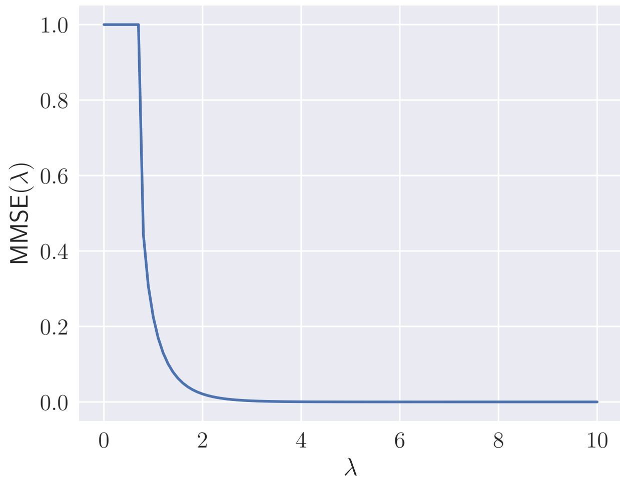
So the self-consistency equation is

$$m = \frac{2}{\lambda}\mathbb{E}_z\left[\rho T_1 + (1-\rho)T_2\right]$$

Solving for $m^*(\lambda)$ numerically, finally we get to the MMSE, it is straightforward to see

$$\mathbb{E}_{x^*}\left[(x^*)^2\right] = \rho\frac{1-\rho}{\rho} + (1-\rho)\frac{\rho}{1-\rho} = 1 \quad \Rightarrow \quad \mathsf{MMSE}(\lambda) = \mathbb{E}_{x^*}\left[(x^*)^2\right] - m^*(\lambda) = 1 - m^*(\lambda)$$





The phase transition occurs at $\lambda \approx 0.746$.

To make the spectral algorithm works for the unbalanced case, we need to slightly change the algorithm as follows:

(1) Find the eigenvector $\mathbf{v}$ associated to the largest eigenvalue in the sense of absolute value

(2) Find the set of indices of first $N\rho$ smallest elements in $\mathbf{v}$, call it $\mathcal{I}^+$, similarly do this for $-\mathbf{v}$ and get $\mathcal{I}^-$.

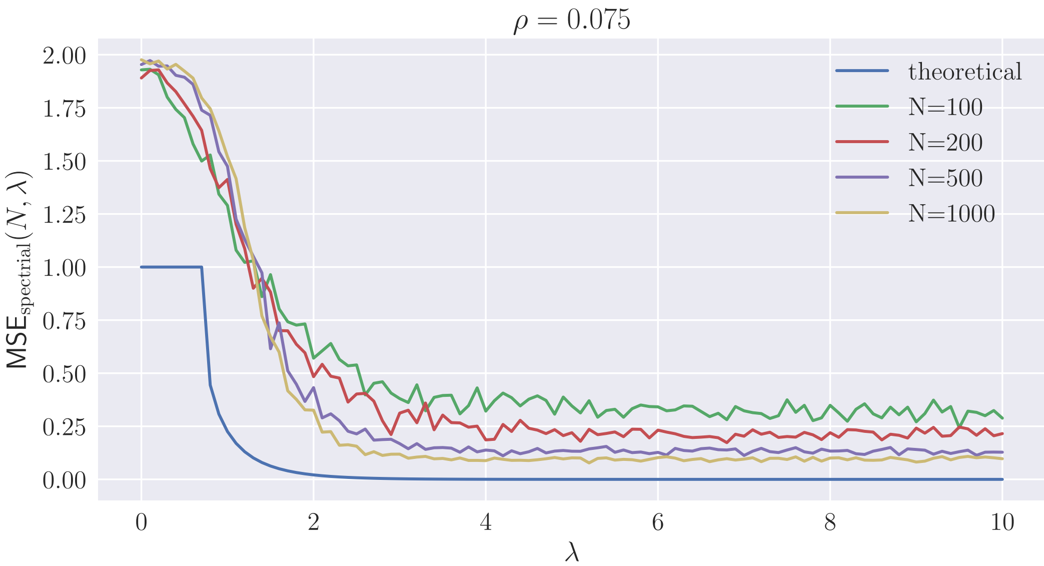
(3) Compute $\hat{\mathbf{x}}^+$ and $\hat{\mathbf{x}}^-$ by

$$\hat{x}_i^+ = \begin{cases} \sqrt{\frac{1-\rho}{\rho}}, & \text{if } i \in \mathcal{I}^+ \\ -\sqrt{\frac{\rho}{1-\rho}}, & \text{if } i \notin \mathcal{I}^+ \end{cases}, \qquad \hat{x}_i^- = \begin{cases} \sqrt{\frac{1-\rho}{\rho}}, & \text{if } i \in \mathcal{I}^- \\ -\sqrt{\frac{\rho}{1-\rho}}, & \text{if } i \notin \mathcal{I}^- \end{cases},$$

(4) Choose the one with smaller error, i.e.

$$\mathsf{MSE}_{\text{spectral}}(N,\lambda) = \min\left(\frac{1}{N}\sum_i \left(x_i^* - \hat{x}_i^+\right)^2, \frac{1}{N}\sum_i \left(x_i^* - \hat{x}_i^-\right)^2\right)$$

At the end, let's compare the performance between spectral algorithm and the theoretical curve:



In contrast, the theoretical MMSE curve is much lower, which means spectral algorithm performs far away from the theoretical optimal bound. □

D. Community Detection variation

This time picture the values $x_i^* = 1$ as being in one group, and $x_i^* = -1$ as being in another group. Instead on observing Y_{ij} as in eq. (1) we have that

$$P\left(Y_{ij} = 1\right) = p_{\text{out}} + \frac{\mu}{\sqrt{N}} x_i^* x_j^*$$
$$P\left(Y_{ij} = 0\right) = 1 - p_{\text{out}} - \frac{\mu}{\sqrt{N}} x_i^* x_j^*$$

(a) Using the universality principle we have seen in class, show that this is a particular case of the Stochastic Block Model as defined in the very first lecture.

(b) Use the universality theorem that maps a generic output channel into the Gaussian one using Fisher information. Give the expression between λ, and p_{out} and μ in this mapping.

Solution.

(a) We have nodes in two groups with $x_i^* \in \{\pm 1\}$ with $x_i \sim P_X(x) = (1-\rho)\,\delta(x+1) + \rho\,\delta(x-1)$. Also, we can view Y_{ij} as the indicator of whether there is an edge connecting node i and node j Therefore, the problem is converted into

$$\begin{cases} P\left((ij) \in E \,\middle|\, x_i^*, x_j^*\right) = P(Y_{ij} = 1) = p_{x_i^*, x_j^*} \\ P\left((ij) \notin E \,\middle|\, x_i^*, x_j^*\right) = P(Y_{ij} = 0) = 1 - p_{x_i^*, x_j^*} \end{cases} \quad \Rightarrow \quad p_{x_i^*, x_j^*} = p_{\text{out}} + \frac{\mu}{\sqrt{N}} x_i^* x_j^*$$

Now, we have

- number of groups: 2
- the faction of each community: $p_{-1} = 1 - \rho$, $p_{+1} = \rho$
- a symmetric 2×2 matrix of edge probabilities

$$\begin{bmatrix} p_{\text{out}} + \frac{\mu}{\sqrt{N}} & p_{\text{out}} - \frac{\mu}{\sqrt{N}} \\ p_{\text{out}} - \frac{\mu}{\sqrt{N}} & p_{\text{out}} + \frac{\mu}{\sqrt{N}} \end{bmatrix}$$

which is sufficient to define an SBM.

(b) Mapping this problem into a symmetric vector-spin glass model:

$$P(\mathbf{x} \mid \mathbf{Y}) = \frac{1}{Z(\mathbf{Y})} \prod_i P_X(x_i) \prod_{i \le j} \mathrm{e}^{g\left(Y_{ij}, \frac{x_i x_j}{\sqrt{N}}\right)}$$

where

$$g(y, z) = \log\left((p_{\text{out}} + \mu z)^y (1 - p_{\text{out}} - \mu z)^{1-y}\right) = \log(1 - p_{\text{out}} - \mu z) + y \log\left(\frac{p_{\text{out}} + \mu z}{1 - p_{\text{out}} - \mu z}\right)$$

Define

$$S(Y_{ij}) \equiv \left.\frac{\partial g(Y_{ij}, z)}{\partial z}\right|_{z=0} = \frac{(Y_{ij} - p_{\text{out}})\mu}{p_{\text{out}}(1 - p_{\text{out}})}$$

Hence, evaluating at $Y_{ij} \in \{0, 1\}$ gives

$$S(Y_{ij} = 1) = \frac{\mu}{p_{\text{out}}}, \quad S(Y_{ij} = 0) = -\frac{\mu}{1 - p_{\text{out}}}$$

The inverse Fisher information is defined as

$$\mathbb{E}_{Y|z=0}\left[\left(\frac{\partial}{\partial z} \log(P(Y \mid z))\Big|_{z=0}\right)^2\right] = \mathbb{E}_{Y|z=0}\left[\left(\frac{\partial g(Y, z)}{\partial z}\Big|_{z=0}\right)^2\right] = \mathbb{E}_{Y|z=0}\left[S^2(Y)\right]$$
$$= (1 - p_{\text{out}})S^2(0) + p_{\text{out}} S^2(1) = \frac{\mu^2}{p_{\text{out}}(1 - p_{\text{out}})}$$

On the other hand, consider the channel in the class, we have

$$Y_{ij} = \sqrt{\lambda}\frac{x_i^* x_j^*}{\sqrt{N}} + \xi_{ij}, \qquad P(y \mid z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(y - \sqrt{\lambda} z\right)^2\right)$$

In this case we have $Y \mid z = 0 \sim \mathcal{N}(0, 1)$ and the inverse Fisher information is

$$\mathbb{E}_{Y|z=0}\left[\left(\frac{\partial}{\partial z} \log(P(Y \mid z))\Big|_{z=0}\right)^2\right] = \int \frac{\mathrm{e}^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \mathrm{d}y \left[\sqrt{\lambda}\left(y - \sqrt{\lambda} z\right)\Big|_{z=0}\right]^2$$
$$= \lambda \int y^2 \frac{\mathrm{e}^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \mathrm{d}y = \lambda$$

Finally, the relationship between λ, p_{out} and μ is

$$\lambda = \frac{\mu^2}{p_{\text{out}}(1 - p_{\text{out}})}$$

□

E. State Evolution

We have seen in class an algorithm called Approximate Message Passing. The interesting things about it is that we can predict exactly its dynamics by analyzing the so-called "state evolution" equation that states that

$$m^{(t+1)} = \mathbb{E}_{x^*, z}\left[\left(f\left(\lambda m^{(t)}, \lambda m^{(t)} x^* + \sqrt{\lambda m^{(t)}} z\right)\right)^2\right]$$

where

$$f(A,B) = \frac{\int \mathrm{d}x\; x P_X^0(x) \mathrm{e}^{-\frac{A}{2}x^2 + Bx}}{\int \mathrm{d}x\; P_X^0(x) \mathrm{e}^{-\frac{A}{2}x^2 + Bx}}$$

Show that the state evolution equation is a fixed point of the replica free entropy.

Given this, which one of the problems discussed so far should be solved by AMP (in the sense that AMP reached the Bayes-Optimal MSE)? Would AMP outperforms spectral methods in question B and C ? If yes, in which sense.

Solution.

(a) We start with the replica free entropy at Bayes-optimal case where $P_X \equiv P_X^0$:

$$\Phi_{\mathrm{RS}}(m;\lambda) = -\frac{\lambda m^2}{4} + \mathbb{E}_{x^*,z}\left[\log\left(\int P_X^0(x)\,\mathrm{d}x\; \mathrm{e}^{-\frac{\lambda m}{2}x^2 + \left(\lambda m x^* + \sqrt{\lambda m} z\right)x}\right)\right]$$

Therefore, taking partial derivative w.r.t. m and set it to zero gives

$$\begin{aligned}
0 &= \frac{\partial}{\partial m}\Phi_{\mathrm{RS}}(m;\lambda) \\
&= -\frac{\lambda m}{2} + \mathbb{E}_{x^*,z}\left[\frac{\frac{\partial}{\partial m}\int P_X^0(x)\,\mathrm{d}x\; \mathrm{e}^{-\frac{\lambda m}{2}x^2 + \left(\lambda m x^* + \sqrt{\lambda m} z\right)x}}{\int P_X^0(x)\,\mathrm{d}x\; \mathrm{e}^{-\frac{\lambda m}{2}x^2 + \left(\lambda m x^* + \sqrt{\lambda m} z\right)x}}\right] \\
&= -\frac{\lambda m}{2} + \mathbb{E}_{x^*,z}\left[\frac{\int P_X^0(x)\,\mathrm{d}x\; \mathrm{e}^{-\frac{\lambda m}{2}x^2 + \left(\lambda m x^* + \sqrt{\lambda m} z\right)x}\left[-\frac{\lambda}{2}x^2 + \lambda x^* x + \sqrt{\frac{\lambda}{m}}\frac{z}{2}x\right]}{\int P_X^0(x)\,\mathrm{d}x\; \mathrm{e}^{-\frac{\lambda m}{2}x^2 + \left(\lambda m x^* + \sqrt{\lambda m} z\right)x}}\right] \\
&= -\frac{\lambda m}{2} + \mathbb{E}_{x^*,z}\left[\left(\lambda x^* + \sqrt{\frac{\lambda}{m}}z\right) f\left(\lambda m, \lambda m x^* + \sqrt{\lambda m} z\right) - \frac{\lambda}{2}\frac{\int \mathrm{d}x\; x^2 P_X^0(x)\mathrm{e}^{-\frac{A}{2}x^2+Bx}}{\int \mathrm{d}x\; P_X^0(x)\mathrm{e}^{-\frac{A}{2}x^2+Bx}}\right] \\
&\overset{(a)}{=} -\frac{\lambda m}{2} + \mathbb{E}_{x^*,z}\Bigg[\lambda x^* f\left(\lambda m, \lambda m x^* + \sqrt{\lambda m} z\right) + \frac{1}{2}\sqrt{\frac{\lambda}{m}}\frac{\partial}{\partial z} f\left(\lambda m, \lambda m x^* + \sqrt{\lambda m} z\right) \\
&\qquad - \frac{\lambda}{2}\frac{\int \mathrm{d}x\; x^2 P_X^0(x)\mathrm{e}^{-\frac{A}{2}x^2+Bx}}{\int \mathrm{d}x\; P_X^0(x)\mathrm{e}^{-\frac{A}{2}x^2+Bx}}\Bigg] \\
&\overset{(b)}{=} -\frac{\lambda m}{2} + \mathbb{E}_{x^*,z}\left[\lambda x^* f\left(\lambda m, \lambda m x^* + \sqrt{\lambda m} z\right) - \frac{\lambda}{2}\left(f\left(\lambda m, \lambda m x^* + \sqrt{\lambda m} z\right)\right)^2\right] \\
&\overset{(c)}{=} -\frac{\lambda m}{2} + \frac{\lambda}{2}\mathbb{E}_{x^*,z}\left[\left(f\left(\lambda m, \lambda m x^* + \sqrt{\lambda m} z\right)\right)^2\right]
\end{aligned}$$

where

(a) Uses the Stein's lemma that $z \sim \mathcal{N}(0,1)$, then

$$\mathbb{E}_z\left[z f(z)\right] = \mathbb{E}_z\left[\frac{\partial}{\partial z} f(z)\right]$$

(b) Uses the fact that

$$\frac{\partial}{\partial B} f(A,B) = \frac{\int \mathrm{d}x\; x P_X^0(x)\mathrm{e}^{-\frac{A}{2}x^2+Bx}}{\int \mathrm{d}x\; P_X^0(x)\mathrm{e}^{-\frac{A}{2}x^2+Bx}} - [f(A,B)]^2$$

(c) Uses the Nishimori identity that

$$\mathbb{E}_{x^*,z}\left[x^*\langle x\rangle_{x^*,z}\right]=\mathbb{E}_{x^*,z}\left[\langle x^*x\rangle_{x^*,z}\right]=\mathbb{E}_{x^*,z}\left[\left\langle x^{(1)}x^{(2)}\right\rangle_{x^*,z}\right]=\mathbb{E}_{x^*,z}\left[\left\langle x^{(1)}\right\rangle_{x^*,z}\left\langle x^{(2)}\right\rangle_{x^*,z}\right]$$

where $x^{(1)}, x^{(2)}$ are two i.i.d. replicas distributed as $P\left(x\,\middle|\,\sqrt{\lambda}x^*+z\right)$ and the posterior mean is denoted as

$$\langle x\rangle_{x^*,z}=\frac{\int \mathrm{d}x\; xP_X^0(x)\mathbb{e}^{-\frac{\lambda m}{2}x^2+\left(\lambda m x^*+\sqrt{\lambda m}z\right)x}}{\int \mathrm{d}x\; P_X^0(x)\mathbb{e}^{-\frac{\lambda m}{2}x^2+\left(\lambda m x^*+\sqrt{\lambda m}z\right)x}}=f\left(\lambda m, \lambda m x^*+\sqrt{\lambda m}z\right)$$

Hence, finally we have the stationary condition

$$m=\mathbb{E}_{x^*,z}\left[\left(f\left(\lambda m, \lambda m x^*+\sqrt{\lambda m}z\right)\right)^2\right]=\mathbb{E}_{x^*,z}\left[x^*f\left(\lambda m, \lambda m x^*+\sqrt{\lambda m}z\right)\right],$$

which is exactly the state evolution equation.

(b) For Problem B and C, let's see what is their state evolution fixed point. First we have the prior

$$P_X^0(x)=\rho\delta\left(x-\sqrt{\frac{1-\rho}{\rho}}\right)+(1-\rho)\delta\left(x+\sqrt{\frac{\rho}{1-\rho}}\right)$$

$$\begin{aligned}f(A,B)&=\frac{\sqrt{\rho(1-\rho)}\left(\mathbb{e}^{-\frac{A}{2}\frac{1-\rho}{\rho}+B\sqrt{\frac{1-\rho}{\rho}}}-\mathbb{e}^{-\frac{A}{2}\frac{\rho}{1-\rho}-B\sqrt{\frac{\rho}{1-\rho}}}\right)}{\rho\mathbb{e}^{-\frac{A}{2}\frac{1-\rho}{\rho}+B\sqrt{\frac{1-\rho}{\rho}}}+(1-\rho)\mathbb{e}^{-\frac{A}{2}\frac{\rho}{1-\rho}-B\sqrt{\frac{\rho}{1-\rho}}}}\\&=\sqrt{\rho(1-\rho)}\frac{2\sinh\left(h(A,B)\right)}{\mathbb{e}^{-h(A,B)}+2\rho\sinh\left(h(A,B)\right)}\end{aligned}$$

where

$$h(A,B)=-\frac{A}{4}\frac{1-2\rho}{\rho(1-\rho)}+\frac{B}{2}\left(\sqrt{\frac{1-\rho}{\rho}}+\sqrt{\frac{\rho}{1-\rho}}\right)=-\frac{A}{4}\frac{1-2\rho}{\rho(1-\rho)}+\frac{B}{2\sqrt{\rho(1-\rho)}}$$

Therefore, for short we write $A\triangleq\lambda m$ and $B_\pm=\pm\lambda m\sqrt{\left(\frac{1-\rho}{\rho}\right)^{\pm1}}+\sqrt{\lambda m}z$

$$h(A,B_+)=h\left(\lambda m, \lambda m\sqrt{\frac{1-\rho}{\rho}}+\sqrt{\lambda m}z\right)=\frac{1}{4}\left[\frac{\lambda m}{\rho(1-\rho)}+2z\sqrt{\frac{\lambda m}{\rho(1-\rho)}}\right]$$

$$h(A,B_-)=h\left(\lambda m, -\lambda m\sqrt{\frac{\rho}{1-\rho}}+\sqrt{\lambda m}z\right)=\frac{1}{4}\left[-\frac{\lambda m}{\rho(1-\rho)}+2z\sqrt{\frac{\lambda m}{\rho(1-\rho)}}\right]$$

Notice that $z\sim\mathcal{N}(0,1)$ is symmetric around 0, thus we have $\mathbb{E}_z[g(z)]=\mathbb{E}_z[g(-z)]$ for any function g

$$
\begin{aligned}
\mathbb{E}_z\left[f(A,B_-)\right] &= \mathbb{E}_z\left[\sqrt{\rho(1-\rho)}\frac{2\sinh\left(h(A,B_-)\right)}{\mathrm{e}^{-h(A,B_-)}+2\rho\ \sinh\left(h(A,B_-)\right)}\right] \\
&= \mathbb{E}_z\left[\frac{\sqrt{\rho(1-\rho)}\cdot 2\sinh\left(\frac{1}{4}\left[-\frac{\lambda m}{\rho(1-\rho)}+2z\sqrt{\frac{\lambda m}{\rho(1-\rho)}}\right]\right)}{\mathrm{e}^{-\frac{1}{4}\left[-\frac{\lambda m}{\rho(1-\rho)}+2z\sqrt{\frac{\lambda m}{\rho(1-\rho)}}\right]}+2\rho\ \sinh\left(\frac{1}{4}\left[-\frac{\lambda m}{\rho(1-\rho)}+2z\sqrt{\frac{\lambda m}{\rho(1-\rho)}}\right]\right)}\right] \\
&= \mathbb{E}_z\left[\frac{\sqrt{\rho(1-\rho)}\cdot 2\sinh\left(\frac{1}{4}\left[-\frac{\lambda m}{\rho(1-\rho)}-2z\sqrt{\frac{\lambda m}{\rho(1-\rho)}}\right]\right)}{\mathrm{e}^{-\frac{1}{4}\left[-\frac{\lambda m}{\rho(1-\rho)}-2z\sqrt{\frac{\lambda m}{\rho(1-\rho)}}\right]}+2\rho\ \sinh\left(\frac{1}{4}\left[-\frac{\lambda m}{\rho(1-\rho)}-2z\sqrt{\frac{\lambda m}{\rho(1-\rho)}}\right]\right)}\right] \\
&= \mathbb{E}_z\left[\sqrt{\rho(1-\rho)}\frac{2\sinh\left(-h(A,B_+)\right)}{\mathrm{e}^{h(A,B_+)}+2\rho\ \sinh\left(-h(A,B_+)\right)}\right] \\
&= -\mathbb{E}_z\left[\sqrt{\rho(1-\rho)}\frac{2\sinh\left(h(A,B_+)\right)}{\mathrm{e}^{h(A,B_+)}-2\rho\ \sinh\left(h(A,B_+)\right)}\right]
\end{aligned}
$$

Hence, the state evolution equation can be rewritten as

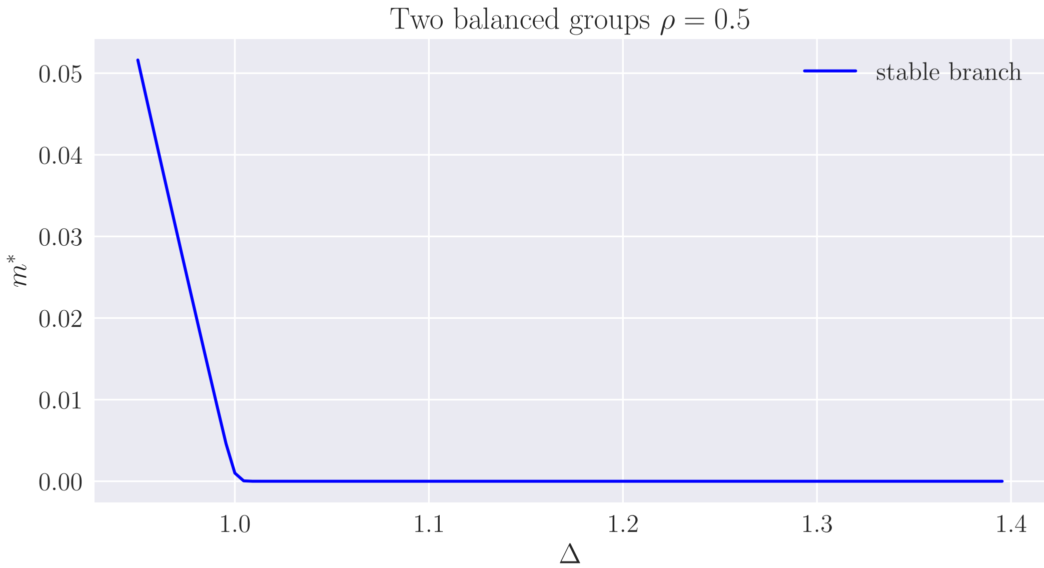
$$
\begin{aligned}
&\mathbb{E}_{x^*,z}\left[x^* f\left(\lambda m, \lambda m x^* + \sqrt{\lambda m} z\right)\right] \\
&= \rho\sqrt{\frac{1-\rho}{\rho}}\mathbb{E}_z\left[f(A,B_+)\right] - (1-\rho)\sqrt{\frac{\rho}{1-\rho}}\mathbb{E}_z\left[f(A,B_-)\right] \\
&= 2\rho(1-\rho)\mathbb{E}_z\left[\frac{\sinh\left(h(A,B_+)\right)}{\mathrm{e}^{-h(A,B_+)}+2\rho\sinh\left(h(A,B_+)\right)}+\frac{\sinh\left(h(A,B_+)\right)}{\mathrm{e}^{h(A,B_+)}-2\rho\sinh\left(h(A,B_+)\right)}\right] \\
&= 2\rho(1-\rho)\mathbb{E}_z\left[\frac{2\sinh\left(h(A,B_+)\right)\cosh\left(h(A,B_+)\right)}{1+4\rho(1-\rho)\sinh^2\left(h(A,B_+)\right)}\right] \\
&= 2\rho(1-\rho)\mathbb{E}_z\left[\frac{\sinh\left(2h(A,B_+)\right)}{1+2\rho(1-\rho)\left[\cosh\left(2h(A,B_+)\right)-1\right]}\right] \\
&= \mathbb{E}_z\left[\frac{2\rho(1-\rho)\sinh\left(\frac{\lambda m}{2\rho(1-\rho)}+z\sqrt{\frac{\lambda m}{\rho(1-\rho)}}\right)}{1+2\rho(1-\rho)\left[\cosh\left(\frac{\lambda m}{2\rho(1-\rho)}+z\sqrt{\frac{\lambda m}{\rho(1-\rho)}}\right)-1\right]}\right]
\end{aligned}
$$

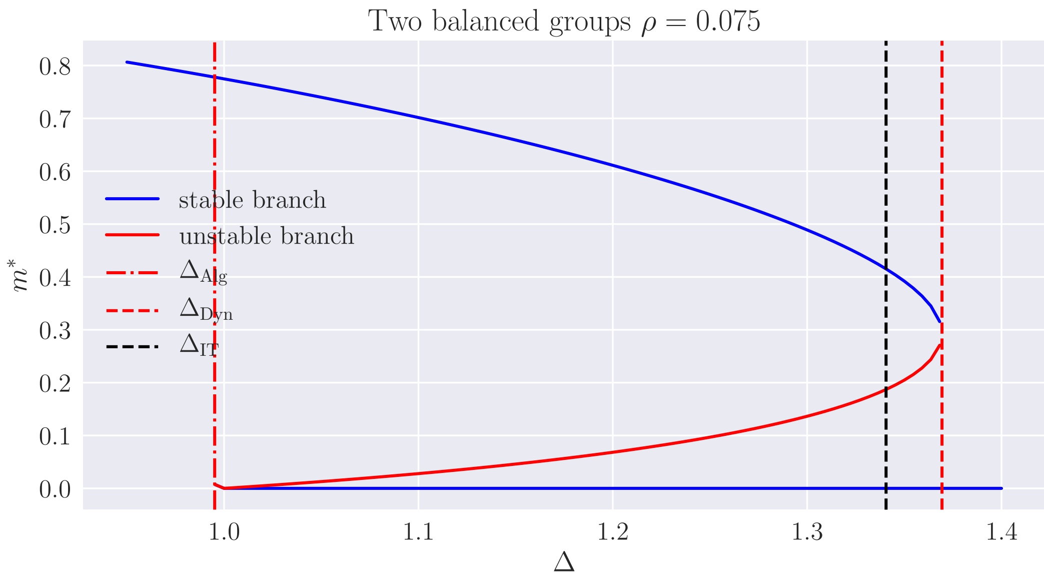
i.e. we can use the following state evolution iteration

$$
m^{(t+1)} = f^{\mathrm{SE}}\left(\lambda m^{(t)}\right), \quad f^{\mathrm{SE}}(x) = \int \frac{\mathrm{e}^{-\frac{z^2}{2}}}{\sqrt{2\pi}}\mathrm{d}z\,\frac{2\rho(1-\rho)\sinh\left(\frac{x}{2\rho(1-\rho)}+z\sqrt{\frac{x}{\rho(1-\rho)}}\right)}{1+2\rho(1-\rho)\left[\cosh\left(\frac{x}{2\rho(1-\rho)}+z\sqrt{\frac{x}{\rho(1-\rho)}}\right)-1\right]}
$$

AMP algorithm achieves the information theoretical MSE the problem is not in the hard phase. Hence, we can plot m^* versus $\Delta = 1/\lambda$ to see if there exists a hard phase, i.e. if the interval $[\Delta_{\mathrm{Alg}}, \Delta_{\mathrm{IT}}]$ exists.

It can be seen that when $\rho = 0.5$, there is no hard phase so that AMP can achieve the Bayes-optimal MSE; when $\rho = 0.075$, the hard phase is $\Delta \in [1, 1.3406]$, AMP converges to a strictly larger error than the Bayes-optimal MSE.


Two balanced groups $\rho = 0.5$
stable branch
m^*
Δ
0.00
0.01
0.02
0.03
0.04
0.05
1.0
1.1
1.2
1.3
1.4


Two balanced groups $\rho = 0.075$
stable branch
unstable branch
Δ_{Alg}
Δ_{Dyn}
Δ_{IT}
m^*
Δ
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
1.0
1.1
1.2
1.3
1.4

$\square$