CS 2243 Fall 2024 Scribes: <names>
11/13 Based on notes by: Yuyang Zhang

Lecture 20: Optimality of Approximate Message Passing for $\mathbb{Z}_2$ Synchronization

1 Recap $\mathbb{Z}_2$ Synchronization, AMP, And More

Let $X \sim \{\pm 1\}^n$ be the signal vector. Let $W \in \mathbb{R}^{n\times n}$ be the noise matrix with $W_{ij} \sim \mathcal{N}(0,1)$ for $i \neq j$ and $W_{ij} \sim \mathcal{N}(0,2)$ for $i = j$. We are given the following noisy version of the signal $XX^\top$

$$Y = \sqrt{\frac{\lambda}{n}} XX^\top + W,$$

where λ denotes the signal-to-noise ratio, and our goal is to learn the denoiser $\hat{X}$ minimizing the following loss

$$\mathrm{MSE}(\hat{X}) := \frac{1}{n^2}\mathbb{E}_{X,Y}\|\hat{X}(Y)\hat{X}(Y)^\top - XX^\top\|_F^2.$$

Inspired by the belief propagation (BP) algorithm, we introduced the following approximate message passing (AMP) algorithm

$$\begin{aligned} x_{t+1} &= \frac{1}{\sqrt{n}} Y f_t(x_t) - f_{t-1}(x_{t-1}) b_t, \quad b_t := \frac{1}{n}\sum_{j=1}^n f_t'(x_t^j), \\ \hat{x}_{t+1} &= f_{t+1}(x_{t+1}). \end{aligned}$$

Here $\frac{1}{\sqrt{n}} Y f_t(x_t)$ in the first equation can be seen as the zeroth-order (mean-field) approximation term of the message passing. Similarly, $-f_{t-1}(x_{t-1})b_t$ can be seen as the first-order approximation term, or as the Onsager correction term that corrects the error in the message passing. Following intuitions from the BP algorithm, we know that the output of our algorithm, i.e. $\hat{x}_{t+1}$, should approximate the marginal expectation of X following the posterior distribution. Namely, the following holds for the given signal Y

$$\hat{x}_{t+1}^i \approx \mathbb{E}_{X^i}[X^i | Y].$$

1.1 State Evolution Analysis

In the asymptotic setting where $n \to \infty$, the AMP algorithm can be analyzed using the "state evolution" method. Intuitively, when $n \to \infty$, the distribution of x_t can be well approximated by

$$x_t \sim \mathcal{N}(\mu_t X, \sigma_t^2 I), \tag{1}$$

which is independently sampled at every time step t. Here the state parameters μ_t and σ_t evolve according to the following dynamics

$$\mu_{t+1} := \sqrt{\lambda} \cdot \mathbb{E}[x f_t(\mu_t x + \sigma_t g)], \quad \sigma_{t+1} := \mathbb{E}[f_t(\mu_t x + \sigma_t g)^2]. \tag{2}$$

where $x \sim \{\pm 1\}$ and $g \sim \mathcal{N}(0,1)$. Therefore, the parameters μ_t's and σ_t's should indicate how close our estimation stays to the true value X.

The theorem listed below formalizes the above intuition, i.e. Equation 1, and we will directly use this intuition in the following proofs for simplicity.

Theorem 1. *[BM11a] If f_t's are Lipschitz, then for any "nice" test function $\psi : \mathbb{R}^2 \to \mathbb{R}$ and any t,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_i \psi(x_t^i, X^i) = \mathbb{E}[\psi(\mu_t x + \sigma_t g, x)]. \tag{3}$$

With this theorem, we can approximate the average error of all coordinates of our estimation (LHS of Equation 3) by some one-dimensional parameters that we can easily keep track of (RHS of Equation 3).

To simplify the parameter dynamics (Equation 2), we choose the following nonlinearity f_t's

$$f_t(y) := \mathbb{E}[x|\mu_t x + \sigma_t g = y],$$

which gives

$$\mu_{t+1} = \sqrt{\lambda} \cdot \mathbb{E}[x \cdot \mathbb{E}[x|\mu_t x + \sigma_t g]] = \sqrt{\lambda} \cdot \mathbb{E}[\mathbb{E}[x|\mu_t x + \sigma_t g]^2] = \sqrt{\lambda} \cdot \sigma_{t+1}^2. \tag{4}$$

Here the last equality follows directly from the definition of σ_t. Moreover, we define

$$\text{mmse}(\gamma) := \mathbb{E}[(x - \mathbb{E}[x|\sqrt{\gamma}x + g])^2] = 1 - \mathbb{E}[\mathbb{E}[x|\sqrt{\gamma}x + g]^2]. \tag{5}$$

When $\gamma = \mu_t^2/\sigma_t^2$, the random variable $\sqrt{\gamma}x + g = (\mu_t x + \sigma_t g)/\sigma_t$ can be seen as a scaled version of $\mu_t x + \sigma_t g$. Therefore, it naturally follows that

$$\sigma_{t+1}^2 = \mathbb{E}[\mathbb{E}[x|\mu_t x + \sigma_t g]^2] = 1 - \text{mmse}(\mu_t^2/\sigma_t^2). \tag{6}$$

Finally, we further simply Equations 4 and 6 by letting $\gamma_t = \mu_t^2/\sigma_t^2$. Substituting $\mu_t^2 = \gamma_t \sigma_t^2$ into the equations yields

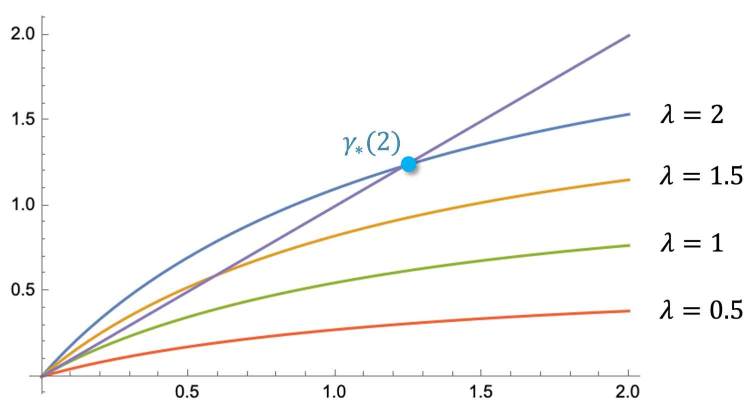
$$\gamma_t = \lambda \sigma_t^2, \quad \sigma_{t+1}^2 = 1 - \text{mmse}(\gamma_t),$$

which leads to the following dynamics

$$\gamma_{t+1} = \lambda(1 - \text{mmse}(\gamma_t)),$$

where λ is the signal-to-noise ratio we previously introduced.

To track the behavior of the γ_t's, a natural idea is to iterate over the above dynamics until we reach some fixed point. We denote the fixed point of γ for some given parameter λ as $\gamma^*(\lambda)$. The iteration results are plotted as follows Here the x-axis is γ and the y-axis is $f(\gamma) = \lambda(1 - \text{mmse}(\gamma))$. It is clear from the figure that for $\lambda > 1$, the



curve $y = f(\gamma)$ always intersects with line $y = \gamma$, leading to one non-zero fixed point. However, for $\lambda \le 1$, there is no non-zero fixed point. For such λ's, the problem becomes information-theoretically impossible.

2 Guarantees for AMP

Now that we have enough knowledge about the AMP algorithm and its convergence analysis in the asymptotic setting where $n \to \infty$, we now try to develop theoretical guarantees on the performance of its output. In this section, we will first compute the MSE achieved by AMP. We then show that this is also the MSE achieved by the Bayes-optimal estimator, i.e. the posterior mean, which in turn indicates that the AMP algorithm is optimal for the $\mathbb{Z}_2$ synchronization in the asymptotic setting.

2.1 MSE for AMP

In this subsection, we calculate the MSE achieved by AMP. Recall that t is the number of iterations of AMP, λ is the signal-to-noise ratio, and n is the dimension of the signal vector X. The MSE is then defined as

$$\begin{aligned} \mathrm{MSE}_{\mathrm{AMP}}(t;\lambda,n) &= \frac{1}{n^2}\mathbb{E}[\|XX^\top - \hat{x}^t(\hat{x}^t)^\top\|_F^2], \\ \mathrm{MSE}_{\mathrm{AMP}}(t;\lambda) &= \lim_{n\to\infty} \mathrm{MSE}_{\mathrm{AMP}}(t;\lambda,n). \end{aligned} \tag{7}$$

Now let $\tilde{x} := \mathbb{E}[XX^\top|Y]$ be the Bayes-optimal estimator. Similarly, we define the MSE for this estimator as

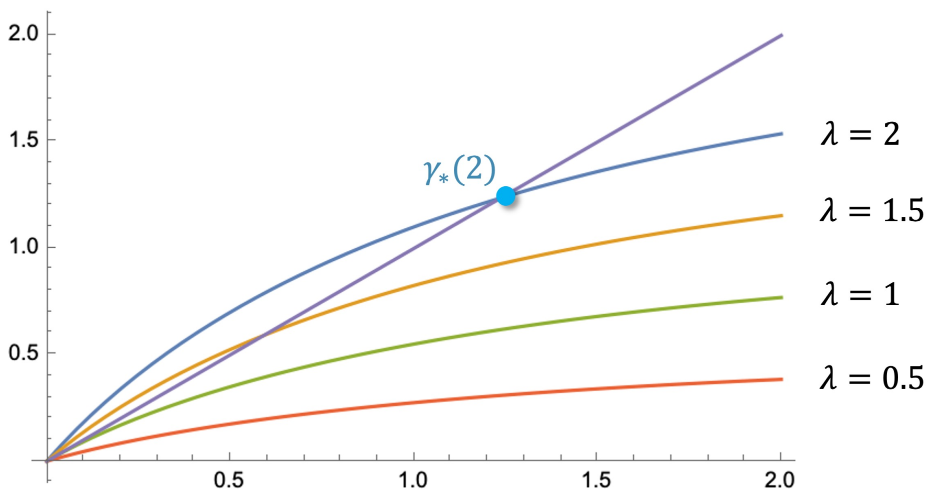
$$\mathrm{MMSE}(\lambda,n) = \frac{1}{n^2}\mathbb{E}[\|XX^\top - \tilde{x}\tilde{x}^\top\|_F^2], \quad \mathrm{MMSE}(\lambda) = \lim_{n\to\infty} \mathrm{MMSE}(\lambda,n). \tag{8}$$

We have the following nice guarantee on $\mathrm{MSE}_{\mathrm{AMP}}(t;\lambda)$, which we will prove shortly

Lemma 1.

$$MSE_{AMP}(t;\lambda) = 1 - \frac{\gamma_{t+1}^2}{\lambda^2}.$$

The following figure plots the MSE for the AMP algorithm for a problem with $n = 200$ and some large enough t It



is clear that this statistical physics-based theoretical result perfectly matches the algorithm behavior, which we haven't seen too much in other techniques.

Proof of Lemma 1. Direct expansion the Frobenius norm gives

$$\begin{aligned} \mathrm{MSE}_{\mathrm{AMP}}(t;\lambda) &= \lim_{n\to\infty} \frac{1}{n^2}\mathbb{E}_{X,Y}[\|XX^\top - \hat{x}_t\hat{x}_t^\top\|_F^2] \\ &= \lim_{n\to\infty} \frac{1}{n^2}\mathbb{E}_X[\|X\|_2^4] - \underbrace{2\mathbb{E}_{X,Y}[\frac{\langle \hat{x}_t, X\rangle^2}{n^2}]}_{I} + \underbrace{\frac{1}{n^2}\mathbb{E}_{X,Y}[\|\hat{x}_t\|_2^4]}_{II}. \end{aligned}$$

Since $X \in \{\pm 1\}^n$, we know that $\|X\|_2^2 = n$, which implies that the first term is 1.

Now we analyze the term I. Notice that

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\langle \hat{x}_t, X\rangle &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \hat{x}_t^i X^i = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f_t(x_t^i)X^i \\ &\overset{(i)}{\approx} \mathbb{E}_{x,g}[f_t(\mu_t x + \sigma_t g)x] \overset{(ii)}{=} \mu_{t+1}/\sqrt{\lambda} \\ &= \gamma_{t+1}/\lambda,\end{aligned}$$

where $x \sim \{\pm 1\}$ and $g \sim \mathcal{N}(0,1)$. (i) is because of the state evolution intuition we previously mentioned, i.e. Equation 1, and (ii) follows from the definition of μ_{t+1}, i.e. Equation 4. Therefore

$$I = 2\lim_{n\to\infty} \frac{1}{n^2}\mathbb{E}[\langle \hat{x}_t, X\rangle^2] = 2\gamma_{t+1}^2/\lambda^2.$$

Similarly, for term II, we first notice that

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n}\|\hat{x}_t\|_2^2 &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} (\hat{x}_t^i)^2 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f_t(x_t^i)^2 \\ &\overset{(i)}{\approx} \mathbb{E}_{x,g}[f_t(\mu_t x + \sigma_t g)^2] \overset{(ii)}{=} \sigma_{t+1}^2 \\ &= \gamma_{t+1}/\lambda.\end{aligned}$$

Again, (i) is because of the state evolution intuition in Equation 1, and (ii) follows from the definition of σ_{t+1}, i.e. Equation 6. Therefore,

$$II = \lim_{n\to\infty} \frac{1}{n^2}\mathbb{E}_{X,Y}[\|\hat{x}_t\|_2^4] = \gamma_{t+1}^2/\lambda^2. \tag{9}$$

Substituting the two terms back into Equation 2.1 gives

$$\mathrm{MSE}_{\mathrm{AMP}}(t;\lambda) = 1 - 2\gamma_{t+1}^2/\lambda^2 + \gamma_{t+1}^2/\lambda^2 = 1 - \gamma_{t+1}^2/\lambda^2. \tag{10}$$

□

2.2 AMP Is Optimal for $\mathbb{Z}_2$ Synchronization

Now we show the MSE achieved by AMP (Lemma 1) is optimal by proving the following lemma

Lemma 2.

$$MMSE(\lambda) = \lim_{t\to\infty} MSE_{AMP}(t;\lambda).$$

The key to the proof is the "I-MMSE relation" given by the following lemma

Lemma 3. *[GSV05]*

$$\frac{1}{n}\cdot\frac{\partial}{\partial\lambda} I(XX^\top; Y) = \frac{1}{4} MMSE(\lambda, n).$$

Here $I(X;Y)$ is the mutual information between X and Y. Intuitively, it indicates:

As λ increases, the information present in Y about the signal $XX^\top$ increases at a rate proportional to MMSE(λ).

Proof of Lemma 2. First, notice that

$$\lim_{n\to\infty} \frac{1}{n} I(XX^\top; Y)|_{\lambda=0} = 0, \quad \lim_{n\to\infty} \frac{1}{n} I(XX^\top; Y)|_{\lambda\to\infty} = \log 2.$$

This is because when $\lambda \to \infty$, knowing Y directly tells the value of $XX^\top$. Therefore there is full information of $XX^\top$ in Y. When $\lambda = 0$, Y has nothing to do with the signal $XX^\top$. So there is no mutual information. Now applying Lemma 3 gives

$$\begin{aligned}
&\frac{1}{4}\lim_{n\to\infty}\int_0^\infty \mathrm{MMSE}(\lambda, n)\mathrm{d}\lambda \\
=\ & \lim_{n\to\infty}\frac{1}{n}I(XX^\top; Y)|_{\lambda\to\infty} - \lim_{n\to\infty}\frac{1}{n}I(XX^\top; Y)|_{\lambda=0} \\
=\ & \log 2.
\end{aligned}$$

On the other hand, since the MMSE is the optimal estimator, we trivially lower bound the MSE of AMP as follows

$$\frac{1}{4}\lim_{n\to\infty}\lim_{t\to\infty}\int_0^\infty \mathrm{MSE}_{\mathrm{AMP}}(t;\lambda,n)\mathrm{d}\lambda \ge \frac{1}{4}\lim_{n\to\infty}\int_0^\infty \mathrm{MMSE}(\lambda,n)\mathrm{d}\lambda \ge \log 2. \tag{11}$$

For the rest of the proof, we will prove that the LHS of Equation 11 can be upper bounded by $\log 2$. Therefore, all inequalities are tight, which gives

$$\frac{1}{4}\lim_{n\to\infty}\lim_{t\to\infty}\int_0^\infty \mathrm{MSE}_{\mathrm{AMP}}(t;\lambda,n)\mathrm{d}\lambda = \frac{1}{4}\lim_{n\to\infty}\int_0^\infty \mathrm{MMSE}(\lambda,n)\mathrm{d}\lambda.$$

Substituting in the definitions of $\mathrm{MSE}(\lambda)$ (Equation 7) and $\mathrm{MMSE}_{\mathrm{AMP}}(t;\lambda)$ (Equation 8) completes the proof.

From Lemma 1, we know that

$$\begin{aligned}
&\frac{1}{4}\lim_{n\to\infty}\lim_{t\to\infty}\int_0^\infty \mathrm{MSE}_{\mathrm{AMP}}(t;\lambda,n)\mathrm{d}\lambda \\
=&\ \frac{1}{4}\lim_{t\to\infty}\int_0^\infty (1-\frac{\gamma_t^2}{\lambda^2})\mathrm{d}\lambda \\
=&\ \frac{1}{4}\int_0^\infty (1-\frac{\gamma_*^2(\lambda)}{\lambda^2})\mathrm{d}\lambda.
\end{aligned} \tag{12}$$

To do the integral calculation, we let $\psi(\gamma,\lambda) := \frac{\lambda}{4} + \frac{\gamma^2}{4\lambda} - \frac{\gamma}{2} + I(\gamma)$, which we will prove to be the anti-derivative of $1 - \frac{\gamma_*^2(\lambda)}{\lambda^2}$ at $\gamma = \gamma_*(\lambda)$. Here $I(\gamma) := I(X; \sqrt{\gamma}X + g)$ for $g \sim \mathcal{N}(0,1)$. By definition, we have

$$\begin{aligned}
\frac{\partial}{\partial\lambda}\psi(\gamma,\lambda) &= \frac{1}{4}(1-\frac{\gamma^2}{\lambda^2}),\\
\frac{\partial}{\partial\gamma}\psi(\gamma,\lambda) &\overset{(i)}{=} \frac{\gamma}{2\lambda} - \frac{1}{2} + \frac{1}{2}\mathrm{mmse}(\gamma) = \frac{1}{2\lambda}\Big(\gamma - \lambda(1-\mathrm{mmse}(\gamma))\Big).
\end{aligned}$$

Here in (i), we have used the fact that $I'(\gamma) = \frac{1}{2}\mathrm{mmse}(\gamma)$, where $\mathrm{mmse}(\gamma)$ is defined in Equation 5. Notice that for γ_*, we have $\gamma_* = \lambda(1-\mathrm{mmse}(\gamma_*))$. Then combining the two terms together gives

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\lambda}\psi(\gamma_*(\lambda),\lambda) &= \frac{1}{4}(1-\frac{\gamma_*^2}{\lambda^2}) + \frac{1}{2\lambda}\Big(\gamma_* - \lambda(1-\mathrm{mmse}(\gamma_*))\Big)\\
&= \frac{1}{4}(1-\frac{\gamma_*^2}{\lambda^2}).
\end{aligned}$$

The clever choise of ψ enables us to cancel the terms w.r.t. $\frac{\partial}{\partial\gamma}\psi(\gamma,\lambda)$ at $\gamma = \gamma_*$ and conclude that $\psi(\lambda_*(\lambda),\lambda)$ is the anti-derivative we are looking for.

Finally, substituting back into Equation 12 gives

$$\frac{1}{4}\int_0^\infty (1-\frac{\gamma_*^2(\lambda)}{\lambda^2})\mathrm{d}\lambda = \psi(\lambda_*(\lambda),\lambda)\Big|_0^\infty. \tag{13}$$

When $\lambda \to 0$, we have $\mathrm{mmse}(\gamma_*(\lambda)) \geq 0$ and $I(\gamma_*(\lambda)) \to 0$. It naturally follows that

$$\lim_{\lambda\to 0} \gamma_*(\lambda) = \lim_{\lambda\to 0} \lambda(1 - \mathrm{mmse}(\gamma_*(\lambda))) \leq \lim_{\lambda\to 0} \lambda = 0.$$

Therefore,

$$\lim_{\lambda\to 0} \psi(\gamma_*(\lambda), \lambda) = \lim_{\lambda\to 0} \frac{\lambda}{4} + \frac{\gamma^2}{4\lambda} - \frac{\gamma}{2} + I(\gamma) = 0. \tag{14}$$

When $\lambda \to \infty$, we have $\mathrm{mmse}(\gamma_*(\lambda)) \to 0$ and $I(\gamma_*(\lambda)) \to \log 2$. It is clear that,

$$\lim_{\lambda\to \infty} \gamma_*(\lambda) = \lim_{\lambda\to \infty} \lambda(1 - \mathrm{mmse}(\gamma_*(\lambda))) = \lim_{\lambda\to \infty} \lambda.$$

Therefore,

$$\lim_{\lambda\to \infty} \psi(\gamma_*(\lambda), \lambda) = \lim_{\lambda\to \infty} \frac{\lambda}{4} + \frac{\gamma^2}{4\lambda} - \frac{\gamma}{2} + I(\gamma) = \log 2. \tag{15}$$

Substituting Equation 14 and 15 back into Equation 13 gives

$$\frac{1}{4} \lim_{n\to\infty} \lim_{t\to\infty} \int_0^\infty \mathrm{MSE}_{\mathrm{AMP}}(t; \lambda, n) \mathrm{d}\lambda = \frac{1}{4} \int_0^\infty (1 - \frac{\gamma_*^2(\lambda)}{\lambda^2}) \mathrm{d}\lambda = \log 2.$$

Finally, combining the above result with Equation 11 finishes the proof. □

3 Free Energy Perspective for AMP

In this section, we will establish an optimization-based interpretation of AMP, which relates the algorithm to the optimization of certain free energy.

To begin with, we first analyze the following simpler version of AMP with only a mean-field approximation term

$$x_{t+1} = \frac{1}{\sqrt{n}} Y \tanh(\sqrt{\lambda} x_t). \tag{16}$$

It turns out that the fixed point x_* of the above dynamics satisfies $\nabla G_{\mathrm{MF}}(x_*) = 0$ for the following Gibbs free energy $G_{\mathrm{MF}}(x)$

$$G_{\mathrm{MF}}(x) = -H(\nu) - \frac{\sqrt{\lambda}}{2\sqrt{n}} \mathbb{E}_{z\sim\nu}[z^\top Y z],$$

where ν is the product distribution with marginal expectations given by x. Namely, this algorithm (Equation 16) is minimizing $G_{\mathrm{MF}}(x)$ over product distributions.

Similarly, the fixed point x_* of the AMP algorithm

$$x_{t+1} = \frac{1}{\sqrt{n}} Y \tanh(\sqrt{\lambda} x_t) - \tanh(\sqrt{\lambda} x_{t-1}) \cdot b_t \tag{17}$$

satisfies $\nabla G_{\mathrm{TAP}}(x_*) = 0$ for the following TAP free energy $G_{\mathrm{TAP}}(x)$

$$G_{\mathrm{TAP}}(x) = -H(\nu) - \frac{\sqrt{\lambda}}{2\sqrt{n}} \mathbb{E}_{z\sim\nu}[z^\top Y z] - \frac{n\lambda}{4}(1 - Q(\mathbb{E}_{z\sim\nu}[z]))^2$$

where ν is the product distribution with marginal expectations given by x and $Q(v) = \frac{1}{n}\|V\|^2$. Namely, AMP is minimizing $G_{\mathrm{TAP}}(x)$ over product distributions. Here the last term corresponds to the Onsager correction term $-\tanh(\sqrt{\lambda} x_{t-1}) \cdot b_t$ in AMP (Equation 17).

4 Other Applications And Algorithms

4.1 Beyond $\mathbb{Z}_2$ Synchronization

When the distribution of X is not uniform over $\{\pm 1\}^n$, AMP is not necessarily Bayes-optimal. The following figure [Mio18] plots the MSE for AMP, where every coordinate of X follows distribution $\mathbb{P}(X^i = 1) = 0.05, \mathbb{P}(X^i = -1) = 0.95$. Unlike the Bayes-optimal estimator MMSE, AMP starts to get non-trivial results only after $\lambda \geq 1$.

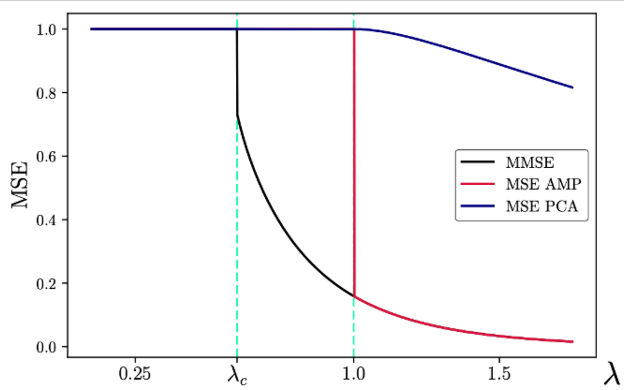


Figure 3.1: Mean Squared Errors for the Spiked Wigner model with prior P_0 given by (3.2.5) with $p = 0.05$.

However, in such cases, AMP is still conjectured to be optimal among all polynomial-time algorithms. In other words, people conjecture that the failure of AMP in some regimes actually indicates computational hardness.

4.2 Beyond Low-Rank Matrix Estimation

Listed below are many other rigorous applications of AMP

- Compressed sensing: [DMM09], [BM11b].
- Generalized linear models: [Ran11], [SR14].
- Mixed linear regression: [TV23].
- Planted clique: [DM15].
- Group synchronization: [PWBM16].
- Nonnegative PCA, sparse PCA, etc.: [MR16], [DM14].
- Random polynomial optimization: [Sub23], [Mon19], [EAMS21].

4.3 Other BP-Inspired Algorithms

There is another kind of BP-inspired algorithm based on non-backtracking operators [KMM+13]. Recall that the AMP algorithm approximates belief propagation in the regime where the interaction matrix is dense but every message is relatively small. In another regime where we only have a sparse interaction matrix but every entry has nontrivial strength, we can still try to approximate belief propagation with the non-backtracking operator.

Consider a given graph G and the following nonbacktracking matrix B, whose entries are indexed by edges and satisfy:

$$B_{(i,j),(k,l)} = 1[j = k \cap i \neq l]. \tag{18}$$

Intuitively, this entry is only non-zero if we can walk along i to j and along j to l without backtracking to i. Such matrices can be used for a well-studied problem called "community detection" [Moo17, Abb17, MNS14, YP23].

4.4 Connecting Back to Distribution Learning

In the next unit, we will see how to use the Bayes-optimal denoising algorithms, of which the AMP algorithm is an instance, to get distribution learning guarantees via diffusion generative modeling.

References

[Abb17] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

[BM11a] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.

[BM11b] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.

[DM14] Yash Deshpande and Andrea Montanari. Sparse pca via covariance thresholding. *Advances in Neural Information Processing Systems*, 27, 2014.

[DM15] Yash Deshpande and Andrea Montanari. Finding hidden cliques of size n/e n/e in nearly linear time. *Foundations of Computational Mathematics*, 15:1069–1128, 2015.

[DMM09] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[EAMS21] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *The Annals of Probability*, 49(6):2922–2960, 2021.

[GSV05] Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in gaussian channels. *IEEE transactions on information theory*, 51(4):1261–1282, 2005.

[KMM+13] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.

[Mio18] Léo Miolane. Phase transitions in spiked matrix estimation: information-theoretic analysis. *arXiv preprint arXiv:1806.04343*, 2018.

[MNS14] Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. In *Conference on Learning Theory*, pages 356–370. PMLR, 2014.

[Mon19] Andrea Montanari. Optimization of the sherrington-kirkpatrick hamiltonian. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1417–1433. IEEE, 2019.

[Moo17] Cristopher Moore. The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *arXiv preprint arXiv:1702.00467*, 2017.

[MR16] Andrea Montanari and Emile Richard. Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Transactions on Information Theory*, 62(3):1458–1484, 2016.

[PWBM16] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Optimality and sub-optimality of pca for spiked random matrices and synchronization. *arXiv preprint arXiv:1609.05573*, 2016.

[Ran11] Sundeep Rangan. Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2168–2172. IEEE, 2011.

[SR14] Philip Schniter and Sundeep Rangan. Compressive phase retrieval via generalized approximate message passing. *IEEE Transactions on Signal Processing*, 63(4):1043–1055, 2014.

[Sub23] Eliran Subag. The free energy of spherical pure p-spin models: computation from the tap approach. *Probability Theory and Related Fields*, pages 1–20, 2023.

[TV23] Nelvin Tan and Ramji Venkataramanan. Mixed linear regression via approximate message passing. In *International Conference on Artificial Intelligence and Statistics*, pages 2116–2131. PMLR, 2023.

[YP23] Qian Yu and Yury Polyanskiy. Ising model on locally tree-like graphs: Uniqueness of solutions to cavity equations. *IEEE Transactions on Information Theory*, 2023.