Prior independent scheduler as Onsager correction term for AMP

THESIS

TO OBTAIN THE TITLE

POSTGRADUATE DIPLOMA

PRESENTS

KEN TANAKA HERNANDEZ

ktanaka@ictp.it

THE ABDUS SALAM
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS

QUANTITATIVE LIFE SCIENCE
Trieste, Italy

Supervisors:
Prof. Jean Barbier

17 August 2022

Acknowledgment

I want to thank the Abdus Salam International Centre for Theoretical Physics, for providing me with all this acquired knowledge not only by the Postgraduate Diploma Programme, but also through all the seminars, workshops and events, ranging for several years. Specially Prof. Jean Barbier, whose kind words have encouraged me to keep challenging myself and have led me become a better student. I also want to thank Prof. Matteo Marsili, and many more professors who have believed in and supported me through this entire process.

I am very thankful with all my friends, specially Adrielle T. Cusi, Rian F. Jalandoni, Patrizia Lazo, Dr. Francesco Camilli, Dr. Daria Tieplova, Valentina Bedoya, Roaa M. Y. Omer, Ali H. Umar, Dorcasse H. Yankam and Aavash Shakya with whom I have been encouraged and had wonderful moments of problem solving discussions and laughs.

Abstract

We consider the estimation problem of detecting a rank-one matrix (*spike*) in the presence of an additive perturbation (*noise*). We focus our attention to the spiked Wigner model and a novel structured noise.

Contents

List of Figures

Chapter 1

Introduction

1.1 Spiked Wigner model

One of the prototypical models in high-dimensional Bayesian inference problem is the Spiked Wigner model. The Spike Wigner model and the goal of this study consists of recovering from the data, a rank-one matrix (*spike*) with additive perturbation (*noise*) given as follows:

$$\mathbf{Y} = \sqrt{\frac{\lambda}{N}}\mathbf{x}^*(\mathbf{x}^*)^\intercal + \mathbf{W} \tag{1.1}$$

Each component of the spike $\mathbf{x}^*(\mathbf{x}^*)^\intercal$ of size N are sampled independently and are identically distributed (*i.i.d.*) from a probability distribution P_X. For our purposes the additive noise $\mathbf{W}$ to be considered is a matrix drawn from the Wigner ensemble of random matrices ($W_{ij} = W_{ji} \sim \mathcal{N}(\mu = 0, \sigma^2 = 1)\ \forall\ 1 \leq i \leq j \leq N$) and a novel *structured noise* [cite]. The *signal-to-noise* ratio (SNR) $\lambda \geq 0$ sets the strength of the signal with respect to the noise. Usually the structure of the signal is assumed, while the noise is often application-dependent [cite].

We shall consider four well known probability distributions for the spike: Bernoulli ($\mathrm{Ber}(\rho)$), Gaussian ($\mathcal{N}(\mu, \sigma^2)$), Rademacher ($\mathrm{Rad}(\rho)$), and discrete Uniform ($\mathcal{U}\{0, r-1\}$) and mainly focus on the Rademacher distribution.

$$X^*_{i\leq N} \sim P_X = \begin{cases} \mathrm{Ber}(\rho) := \rho\delta_1 + (1-\rho)\delta_0 \\ \mathcal{N}(\mu, \sigma^2) := \dfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\dfrac{1}{2}\left(\dfrac{x-\mu}{\sigma}\right)^2\right] \\ \mathrm{Rad}(\rho) := \dfrac{\rho}{2}(\delta_1 + \delta_{-1}) + (1-\rho)\delta_0 \\ \mathcal{U}\{0, r-1\} := \dfrac{1}{r}\displaystyle\sum_{i=0}^{r-1} \delta_i \end{cases}$$

1.1.1 Structured Noise

The purpose of the novel structured noise is to go beyond the independence assumption on the noise entries. The noise entries are drawn from a low-order polynomial orthogonal matrix ensemble [cite]. We study how the noise structure helps inferring the signal structure through a modified Approximate Message Passing algorithm.

The noise matrix $\mathbf{W}$ is drawn from a trace random matrix ensemble, defined by a certain potential $V : \mathbb{R} \to \mathbb{R}$. V is extended to diagonal matrices as follows: $V(\mathbf{A}) = \mathrm{diag}(V(a_1), \ldots, V(a_N))$. For real symmetric matrices $\mathbf{M} = \mathbf{U}\mathbf{A}\mathbf{U}^\intercal$, with $\mathbf{U}$ orthogonal, $V(\mathbf{M}) = \mathbf{U}V(\mathbf{A})\mathbf{U}^\intercal$. The density of the trace random matrix ensemble (with normalization constant C_V) is given as

$$dP_W(\mathbf{W}) = C_V \exp\left(-\frac{N}{2}\mathrm{Tr}V(\mathbf{W})\right)\prod_{i\leq j} dW_{ij} \tag{1.2}$$

Instances of such ensembles have a spectral decomposition $\mathbf{W} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\intercal$, with $\mathbf{V}$ uniformly distributed over $N \times N$ orthogonal matrices. The distribution of the eigenvalues in the diagonal matrix $\mathbf{\Lambda}$, which is independent of $\mathbf{V}$, can be explicitly written. We review the following subindexed matrix potentials:

$$V_4(x) = \frac{\mu}{2}x^2 + \frac{\gamma}{4}x^4 \tag{1.3}$$

$$V_6(x) = \frac{\mu}{2}x^2 + \frac{\gamma}{4}x^4 + \frac{\xi}{6}x^6 \tag{1.4}$$

which are called as quartic and sestic matrix potentials, respectively. μ, γ and ξ are non-negative real numbers [cite]. In order to have a coherent definition of SNR, we also fix $\int x^2 dP_X(x) = 1$; which amounts to rescale λ.

Quartic

When the noise $\mathbf{W}$ is drawn from the quartic ensemble has a known $N \to \infty$ asymptotic eigenvalue distribution [cite]:

$$\rho_4(x)dx = \frac{(\mu + 2a^2\gamma + \gamma x^2)\sqrt{4a^2 - x^2}}{2\pi}dx \tag{1.5}$$

where $a^2 = (\sqrt{\mu^2 + 12\gamma} - \mu)/(6\gamma)$. In order to have a coherent definition of SNR, we also fix $\int x^2 d\rho_4(x) = 1$, which implies

$$\gamma = \gamma(\mu) = \frac{8 - 9\mu + \sqrt{64 - 144\mu + 108\mu^2 - 27\mu^3}}{27} \tag{1.6}$$

When $\mu = 1$, (*i.e.* $\gamma = 0$) we recover the pure Wigner case. On the contrary, $\mu = 0$, (*i.e.* $\gamma = 16/27$) corresponds to a purely quartic case with unit variance (*i.e.* the "most structured") ensemble in this class. Therefore, μ allows us to interpolate between unstructured and structured noise ensembles.

$$\mathbf{J}_6(\mathbf{Y}) = \xi\lambda\mathbf{Y}^5 - \xi\lambda^2\mathbf{Y}^4 + \xi\lambda^2\mathbf{Y}^2$$

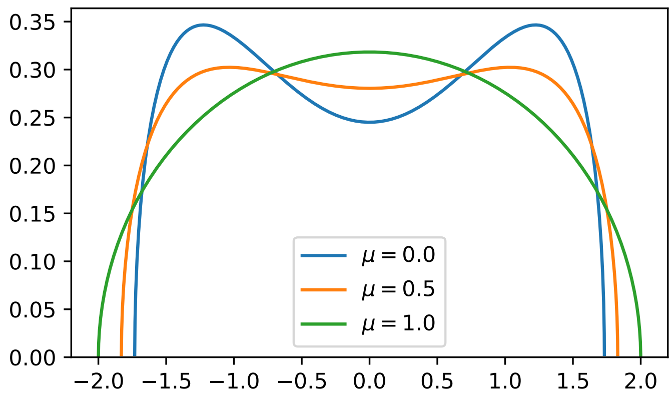


Figure 1.1: Asymptotic spectral density (1.5) of the random noise ensemble defined by the potential (1.3) from less structured (with independent entries) at ($\mu = 1$), corresponding to the standard semi-circle law, to the more structured ($\mu = 0$).

Sestic

When the noise $\mathbf{W}$ is drawn from the sestic ensemble, as we are interested in the most structured and accessible case, we set the coefficients of the lower order monomials to 0 ($i.e.$ $\mu = \gamma = 0$).

$$V_6(x) = \frac{\xi}{6}x^6 \tag{1.7}$$

The asymptotic eigenvalue distribution is given as

$$\begin{aligned} \rho_6(x)dx =& \frac{(\mu + 2a^2\gamma + 6a^4\xi + (\gamma + 2a^2\xi)x^2 + \xi x^4)\sqrt{4a^2 - x^2}}{2\pi}dx \\ =& \frac{\xi(6a^4 + 2a^2x^2 + x^4)\sqrt{4a^2 - x^2}}{2\pi}dx \end{aligned} \tag{1.8}$$

In order to have a coherent definition of SNR, we also fix $\int x^2 d\rho_6(x) = 1$, which implies

$$\xi = \frac{27}{80}, \qquad a = \sqrt{\frac{2}{3}}$$

$$\mathbf{J}_4(\mathbf{Y}) = \mu\lambda\mathbf{Y} - \gamma\lambda^2\mathbf{Y}^2 + \gamma\lambda\mathbf{Y}^3$$

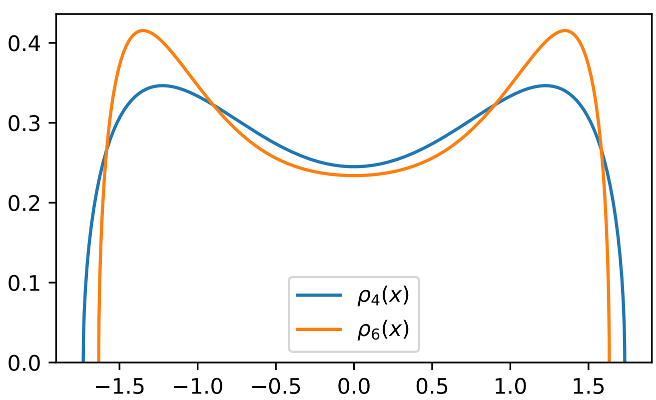


Figure 1.2: Asymptotic spectral density comparison between (1.5) and (1.8) at the most structured ($\mu = 0$).

1.2 Approximate Message Passing algorithm

Approximate Message Passing (AMP) is a very well known low complexity iterative algorithm that produces iterates $\mathbf{p}^t, \mathbf{q}^t, \mathbf{P}^t, \mathbf{Q}^t \in \mathbb{R}^N$ and recovers the spike $\mathbf{x}^*(\mathbf{x}^*)^\intercal$ from the given input data $\mathbf{Y}$ after a finite number of iterations $t \in \mathbb{N}_+$. The iterates $\mathbf{p}^t$ and $\mathbf{q}^t$ are called *Marginal* Mean and Variance, respectively. The marginals are computed by the function $\eta(\cdot,\cdot)$, also known as the *Denoiser* and its partial derivative. The denoiser is a scalar function component-wise defined. The denoiser takes as input the iterates $\mathbf{P}^t$, $\mathbf{Q}^t \in \mathbb{R}^N$. The iterates $\mathbf{P}^t$ and $\mathbf{Q}^t$ are known as Mean and Variance *Cavity Fields*, respectively. The marginal mean iterate $\mathbf{p}^t$ is computed by the denoiser as a function of the cavity field iterates $\mathbf{P}^t$ and $\mathbf{Q}^t$, whereas the marginal variance $\mathbf{q}^t$ is computed by the partial derivative of the denoiser with respect to the mean cavity field, and evaluated at the cavity field iterates. The Bayes-optimal iterative procedure for a generic prior is given as follows:

$$\mathbf{p}^{t+1} := \eta(\mathbf{P}^t, \mathbf{Q}^t), \quad \mathbf{q}^{t+1} := \partial_{\mathbf{P}^t}\eta(\mathbf{P}^t, \mathbf{Q}^t)$$

The marginal mean iterate $\mathbf{p}^t$ converge as $N \to \infty$ to Gaussian random variables (R.V.) with a prescribed mean and variance. [cite] The marginal mean iterate recovers the spike after a few iterations. In other words, the marginal mean $\mathbf{p}^t \otimes \mathbf{p}^t$ converges to the spike $\mathbf{x}^*(\mathbf{x}^*)^\intercal$. Therefore, the AMP algorithm can be run for a finite number of iterations until a certain threshold δ of the *distance* is met. The distance is computed by the average of the difference between two consecutive marginal mean iterates: $\mathbb{E}[\mathbf{p}^t - \mathbf{p}^{t-1}] > \delta$.

1.2.1 Cavity Fields

Mean

The cavity fields offer a unique structure in which the far right term has a particular name called *Onsager* correction. The Onsager correction term $\lambda \mathbf{Y}^2\mathbf{q}^t/N$ is dependent to the marginal variance at iteration t and it is multiplied to the marginal mean at iteration $t-1$ as presented in the following equation:

$$\mathbf{P}^{t+1} := \sqrt{\frac{\lambda}{N}}\mathbf{Y}\mathbf{p}^t - \mathbf{p}^{t-1} \circ \left[\frac{\lambda}{N}\mathbf{Y}^2\mathbf{q}^t\right] \tag{1.9}$$

where λ is the SNR, $\mathbf{Y}$ is the input data of size $N \times N$, $\mathbf{p}$'s are the marginal means computed at the iterations t and $t-1$ respectively, $\mathbf{q}^t$ is the marginal variance computed at the iteration t and the operation $\circ$ is the Hadamard (entrywise) product as well as the power raising exponent: $\mathbf{Y}^2 \equiv \mathbf{Y} \circ \mathbf{Y}$. The initialization marginals for the AMP algorithm are $\mathbf{p}^0 \equiv \mathbf{0}$, whereas the components of $\mathbf{p}^1$ and $\mathbf{q}^1$ are sampled from the Normal distribution $\mathcal{N}(0, \varepsilon)$ with $\varepsilon \ll 1$.

Variance

The variance cavity field also contains the Onsager correction term as shown in the following:

$$\begin{aligned}\mathbf{Q}^{t+1} :=& \frac{\lambda}{N}\left((\mathbb{I}_N - \mathbf{Y}^2)\mathbf{q}^t + \|\mathbf{p}^t\|^2\mathbf{1}\right) \\ =& \frac{\lambda}{N}\left(\mathbf{q}^t + \|\mathbf{p}^t\|^2\mathbf{1}\right) - \left[\frac{\lambda}{N}\mathbf{Y}^2\mathbf{q}^t\right]\end{aligned} \tag{1.10}$$

where $\mathbf{1}$ represents the vector of size N with each component valued with 1, $\|\cdot\|^2$ represents the L_2 norm squared and for this particular case, it can be expressed as a combination of Hadamard and dot $(\cdot)$ products as $\|\mathbf{v}\|^2 = (\mathbf{v} \circ \mathbf{v}) \cdot \mathbf{1} = \mathbf{v}^2 \cdot \mathbf{1}$.

Simplifications

It is known that the Onsager correction term can be simplified into scalar as the mean value of the Onsager correction term: $\mathbb{E}[\lambda \mathbf{Y}^2 \mathbf{q}^t / N]\mathbf{1}$. This simplification is still dependent on the input data $\mathbf{Y}$. By considering a C amount of different input data (*samples*), we are able to average out every mean Onsager correction term and *learn* statistically the values of the mean Onsager correction term per iteration $\mathbb{O}(t)\mathbf{1}$.

$$\mathbb{O}(t) = \frac{\lambda}{N} \left\langle \mathbb{E}\left[\mathbf{Y}^2 \mathbf{q}^t\right] \right\rangle_C \tag{1.11}$$

where the $\langle \cdot \rangle_C$ represents the average value with respect to each sample C. There is no prior reasoning to establish that any *similar* function to $\mathbb{O}(t)$ will be able to recover the spike in a finite number of iterations. We recall that the marginal variance $\mathbf{q}^t$ present in the Onsager correction term is dependent on the partial derivative of the denoiser. While the denoiser is strictly dependent on the prior distribution of the signal $\mathbf{x}^*$. Therefore, the ability of the AMP algorithm to recover the spike in a finite number of iterations regardless to the partial derivative of the denoiser will mainly depend on two plausible statements: AMP algorithm is *robust* enough to denoise the input data or the *Gaussianity* of the input data allows the AMP algorithm to denoise the input data. We shall evaluate the use of an arbitrary function $f(t)$ called *scheduler*, which follows the behavior of $\mathbb{O}(t)$ as the Onsager correction term per iteration. By making use of the scheduler $f(t)$, we are able to completely disregard the computation of the partial derivative which in many cases it would be hard to obtain or is non-existent. We shall track quantitatively a *comparison* of the obtained iterates $\mathbf{p}^t$ to the Information-theoretically optimal result of the AMP algorithm.

1.2.2 Marginals and Denoiser

The Bayes-optimal setting is encountered when defining the Minimal Mean Square Error (MMSE) denoiser $\eta : \mathbb{R}^N \to \mathbb{R}^N$ as the posterior mean of a Gaussian channel:

$$\begin{aligned} \eta(b, v) :=& \mathbb{E}\left[X | Xv + W\sqrt{v} = b\right] \\ =& \frac{\int dP_X(x) x \exp(xb - x^2 v/2)}{\int dP_X(x) \exp(xb - x^2 v/2)} \end{aligned} \tag{1.12}$$

We recall that the marginals mean and variance strictly depend on the denoiser and its partial derivative evaluated with the cavity fields. As the denoiser represents the posterior mean of a Gaussian channel, the denoiser is clearly dependent on the prior of the input data, in which for this case study we look into $\mathrm{Ber}(\rho)$, $\mathcal{N}(\mu, \sigma^2)$ and $\mathrm{Rad}(\rho)$. We present a list of the multiple denoisers with its respective partial derivative for each prior distribution of the signal.

When $X_i^* \sim \mathrm{Ber}(\rho) := (1-\rho)\delta_0 + \rho\delta_1,\ \rho \in (0, 1]$:

$$\begin{aligned} \eta(b, v) =& \frac{\rho \exp(b - v/2)}{\rho \exp(b - v/2) + (1-\rho)} \\ =& \left(1 + \frac{1-\rho}{\rho} \exp(v/2 - b)\right)^{-1} \end{aligned} \tag{1.13}$$

$$\begin{aligned} \eta\prime(b, v) =& -\frac{\exp(v/2 - b)(1-\rho)/\rho}{\left(1 + \frac{1-\rho}{\rho} \exp(v/2 - b)\right)^2} \\ =& \left(\frac{\rho - 1}{\rho} \exp(v/2 - b)\right) \eta^2(b, v) \end{aligned} \tag{1.14}$$

When $X_i^* \sim \mathcal{N}(\mu, \sigma) := \exp\left(-\left[(x-\mu)/\sigma\right]^2/2\right)/\sqrt{2\pi\sigma^2}$:

$$\eta(b,v) = \frac{b\sigma^2+\mu}{v\sigma^2+1} \tag{1.15}$$

$$\begin{aligned}\eta\prime(b,v) &= \frac{\sigma^2}{v\sigma^2+1}\\ &= \left(\frac{\sigma^2}{b\sigma^2+\mu}\right)\eta(b,v)\\ &= \frac{(v\sigma^2+1)\sigma^2}{(b\sigma^2+\mu)^2}\eta^2(b,v)\end{aligned} \tag{1.16}$$

When $X_i^* \sim \mathrm{Rad}(\rho) := (1-\rho)\delta_0 + \frac{\rho}{2}(\delta_1+\delta_{-1}),\ \rho \in (0,1]$:

$$\begin{aligned}\eta(b,v) &= \frac{\rho(\exp(b-v/2)-\exp(-b-v/2))/2}{\rho(\exp(b-v/2)+\exp(-b-v/2))/2+(1-\rho)}\\ &= \frac{\sinh(b)}{\cosh(b)+\exp(v/2)(1-\rho)/\rho}\\ &= \frac{\tanh(b)}{1+2(1-\rho)/\rho(\exp(b-v/2)+\exp(-b-v/2))}\end{aligned} \tag{1.17}$$

$$\begin{aligned}\eta\prime(b,v) &= \frac{(\cosh(b)+\exp(v/2)(1-\rho)/\rho)\cosh(b)-\sinh(b)\sinh(b)}{(\cosh(b)+\exp(v/2)(1-\rho)/\rho)^2}\\ &= \frac{1+\exp(v/2)\cosh(b)(1-\rho)/\rho}{(\cosh(b)+\exp(v/2)(1-\rho)/\rho)^2}\\ &= \left(\frac{1}{\sinh^2(b)}+\frac{1-\rho}{\rho}\frac{\exp(v/2)}{\sinh(b)\tanh(b)}\right)\eta^2(b,v)\end{aligned} \tag{1.18}$$

When $X_i^* \sim \mathrm{Dice}(\rho) := \frac{1}{\rho}\sum_{i=0}^{\rho-1}\delta_i,\ \rho \in \mathcal{N}_+$:

$$\begin{aligned}\eta(b,v) &= \frac{\sum_{i=0}^{\rho-1} i\exp(ib-i^2v/2)/\rho}{\rho(\exp(b-v/2)+\exp(-b-v/2))/2+(1-\rho)}\\ &= \frac{\rho\exp(-v/2)(\exp(b)-\exp(-b))/2}{\rho\exp(-v/2)\left[(\exp(b)+\exp(-b))/2+\exp(v/2)(1-\rho)/\rho\right]}\\ &= \frac{(\exp(b)-\exp(-b))/2}{(\exp(b)+\exp(-b))/2+\exp(v/2)(1-\rho)/\rho}\\ &= \frac{\sinh(b)}{\cosh(b)+\exp(v/2)(1-\rho)/\rho}\\ &= \frac{\tanh(b)}{1+\exp(v/2)(1-\rho)/\rho\cosh(b)}\\ &= \frac{\tanh(b)}{1+2(1-\rho)/\rho(\exp(b-v/2)+\exp(-b-v/2))}\end{aligned} \tag{1.19}$$

$$\begin{aligned}
\eta\prime(b,v) =& \frac{(\cosh(b)+\exp(v/2)(1-\rho)/\rho)\cosh(b)-\sinh(b)\sinh(b)}{(\cosh(b)+\exp(v/2)(1-\rho)/\rho)^2} \\
=& \frac{\cosh^2(b)-\sinh^2(b)+\cosh(b)\exp(v/2)(1-\rho)/\rho}{(\cosh(b)+\exp(v/2)(1-\rho)/\rho)^2} \\
=& \frac{1+\exp(v/2)\cosh(b)(1-\rho)/\rho}{(\cosh(b)+\exp(v/2)(1-\rho)/\rho)^2} \\
=& \left(1+\frac{1-\rho}{\rho}\exp(v/2)\cosh(b)\right)\left(\frac{\eta(b,v)}{\sinh(b)}\right)^2 \\
=& \left(\frac{1}{\sinh^2(b)}+\frac{1-\rho}{\rho}\frac{\exp(v/2)}{\sinh(b)\tanh(b)}\right)\eta^2(b,v)
\end{aligned} \tag{1.20}$$

In both $\mathrm{Rad}(\rho)$ and $\mathrm{Ber}(\rho)$ distributions the marginal variance can be decomposed into a multiplication of the denoiser squared and an auxiliary function $f(t)$. As our aim is the usage of an arbitrary function regardless of the partial derivative of the denoiser, we shall follow the behavior of this auxiliary function and evaluate the AMP algorithm with a different but similar in behavior scheduler.

Simplifications

In the case when the prior is $\mathrm{Rad}(\rho=1)\equiv(\delta_{-1}+\delta_1)/2$, the denoiser and its partial derivative can be simplified as follows:

$$\begin{aligned}
\eta(b,v) =& \tanh(b) \\
\eta(b) =& \\
\eta\prime(b,v) =& (\cosh(b))^{-2} \\
\eta\prime(b) =&
\end{aligned} \tag{1.21}$$

This simplification dismisses the use of the variance cavity field $\mathbf{Q}^t$ in the denoiser and its partial derivative. Hence, we can simplify the Onsager correction term by simplifying the marginal variance as a multiplication of a scheduler times the square of the prior. By updating all the computation needed in the mean cavity field we have:

$$\begin{aligned}
\mathbf{P}^{t+1} =& \sqrt{\frac{\lambda}{N}}\mathbf{Y}\eta(\mathbf{P}^t)-\eta(\mathbf{P}^{t-1})\circ\left[\frac{\lambda}{N}\mathbf{Y}^2\eta\prime(\mathbf{P}^t)\right] \\
=& \sqrt{\frac{\lambda}{N}}\mathbf{Y}\eta(\mathbf{P}^t)-\eta(\mathbf{P}^{t-1})\circ\mathcal{O}^t\left[\frac{\lambda}{N}\mathbf{Y}^2\eta^2(\mathbf{P}^t)\right]
\end{aligned} \tag{1.22}$$

1.2.3 BAMP: Bayes-optimal AMP

In order to achieve Bayes-optimal performance, one should consider the BAMP iteration which is of the form

$$\mathbf{P}^t=\mathbf{J}(\mathbf{Y})\mathbf{p}^t-\sum_{i=1}^{t}c_{t,i}\mathbf{p}^i, \quad \mathbf{p}^{t+1}=\eta(\mathbf{P}^t), \quad t\geq 1 \tag{1.23}$$

1.3 Measurements

1.3.1 State Evolution (SE)

The key property of AMP is that it admits *asymptotically exact* characterization in the high-dimensional limit where $N\to\infty$. These prescribed mean and variance parameters evolve according to deterministic recursions, jointly termed "state evolution" Let $q^0 = \epsilon \ll 1$. For $t \in \mathbb{N}_0$:

$$q^{t+1} = \mathbb{E}\left[\eta(\lambda q^t X^* + \sqrt{\lambda q^t} Z, \lambda q^t) X^*\right]$$

$$q^{t+1} = \begin{cases} \rho^2 \mathbb{E}\left[\left(\rho + (1-\rho)\exp\left(-\frac{\lambda q^2}{2} - \sqrt{\lambda q^t} Z\right)\right)^{-1}\right] & : \mathrm{Ber}(\rho) \\ \rho^2 \mathbb{E}\left[\frac{\tanh(\lambda q^t + \sqrt{\lambda q^t} Z)}{\rho + \frac{\exp(\lambda q^t/2)}{2\cosh(\lambda q^t + \sqrt{\lambda q^t} Z)(1-\rho)}}\right] & : \mathrm{Rad}(\rho) \end{cases}$$

1.3.2 Mean Square Error (MSE)

$$\mathrm{MSE}^t_{\mathrm{AMP}} \equiv \frac{1}{N^2}\left\|\mathbf{x}^* \otimes \mathbf{x}^* - \mathbf{p}^t \otimes \mathbf{p}^t\right\|_F^2 = \frac{1}{N^2}\left(\|\mathbf{x}^*\|_2^4 + \|\mathbf{b}^t\|_2^4 - 2(\mathbf{x}^* \cdot \mathbf{b}^t)^2\right)$$

From State Evolution (SE) and the independence of the parameters x_i^* we can predict that almost surely:

$$\begin{aligned} \lim_{N \to +\infty} \mathrm{MSE}^t_{\mathrm{AMP}} =& \rho^2 + (q^t)^2 - 2(q^t_*)^2 \\ =& \rho^2 - (q^t)^2 \end{aligned}$$

Whenever the fixed point q^∞ of state evolution reached from the non-informative initialization matches the global minimizer $q_0(\lambda, \rho)$ of the replica symmetric potential, then AMP is optimal in the sense that its estimate $\mathbf{b}^\infty \otimes \mathbf{b}^\infty$ of the spike almost surely leads to the spike-MMSE in the limit $n \to +\infty$:

$$\begin{aligned} &\text{If } q^0 = \varepsilon \;\; \& \;\; q^\infty = q_0(\lambda, \rho) \\ &\text{Then } \lim_{t \to +\infty} \lim_{N \to +\infty} \mathrm{MSE}^t_{\mathrm{AMP}} = \lim_{N \to +\infty} \mathrm{MMSE}(\mathbf{X}^* \otimes \mathbf{X}^* | \mathbf{Y}) \end{aligned}$$

Methodology

All algorithms are run for $N = 5000$ and the results are averaged over $C = 30$ independent trials. The following algorithms describe the

Algorithm 1: Spike Wigner Model (Quartic)

Input : λ, ρ, μ

$\mathbf{x}^* \leftarrow \mathbf{X}^*_{i\leq N} \sim P_X(\rho)$

$\gamma \leftarrow \frac{8-9\mu+\sqrt{64-144\mu+108\mu^2-27\mu^3}}{27}$

$a^2 \leftarrow \frac{\sqrt{\mu^2+12\gamma}-\mu}{6\gamma}$

$\mathbf{V} \leftarrow$ Orthogonal matrix

$\mathbf{\Lambda} \leftarrow \mathrm{diag}\left(W_i \sim \rho_4(x) = \frac{(\mu+2a^2\gamma+\gamma x^2)\sqrt{4a^2-x^2}}{2\pi}\right)$

$\mathbf{Y} \leftarrow \frac{\lambda}{N}\mathbf{x}^*\mathbf{x}^{*\intercal} + \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\intercal$

Output: $\mathbf{Y}$

Algorithm 2: Spike Wigner Model (Sestic)

Input : λ, ρ

$\mathbf{x}^* \leftarrow \mathbf{X}^*_{i\leq N} \sim P_X(\rho)$

$\xi \leftarrow 27/80$

$a^2 \leftarrow 2/3$

$\mathbf{V} \leftarrow$ Orthogonal matrix

$\mathbf{\Lambda} \leftarrow$
$\mathrm{diag}\left(W_i \sim \rho_6(x) = \frac{\xi(6a^4+2a^2x^2+x^4)\sqrt{4a^2-x^2}}{2\pi}\right)$

$\mathbf{Y} \leftarrow \frac{\lambda}{N}\mathbf{x}^*\mathbf{x}^{*\intercal} + \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\intercal$

Output: $\mathbf{Y}$

Algorithm 3: Spike Wigner Model (Normal)

Input : λ, ρ

$\mathbf{x}^* \leftarrow \mathbf{X}^*_{i\leq N} \sim P_X(\rho)$

$\mathbf{W} \leftarrow W_{i\leq j\leq N} \sim \mathcal{N}(0,1)$

$\mathbf{Y} \leftarrow \sqrt{\frac{\lambda}{N}}\mathbf{x}^*\mathbf{x}^{*\intercal} + \mathbf{W}$

Output: $\mathbf{Y}$

Algorithm 4: Approximate Message Passing

Input : $\lambda, \rho, \mathbf{Y}, \delta$

$\mathbf{p}_0 \leftarrow \mathbf{X}_{i\leq N} \sim \mathcal{N}(0,\varepsilon)$

$\mathbf{p}_1 \leftarrow \mathbf{X}_{i\leq N} \sim \mathcal{N}(0,\varepsilon)$

$\mathbf{q} \leftarrow \mathbf{X}_{i\leq N} \sim \mathcal{N}(0,\varepsilon)$

while $(\mathbf{p}^1 - \mathbf{p}^0)\cdot \mathbf{1}/N > \delta$ **do**

$\quad\mathbf{O} \leftarrow \frac{\lambda}{N}\mathbf{Y}^2\mathbf{q}$

$\quad\mathbf{P} \leftarrow \sqrt{\frac{\lambda}{N}}\mathbf{Y}\mathbf{p}_1 - \mathbf{p}_0 \circ \mathbf{O}$

$\quad\mathbf{Q} \leftarrow \frac{\lambda}{N}\left(\mathbf{q} + \|\mathbf{p}_1\|_2^2\mathbf{1}\right) - \mathbf{O}$

$\quad\mathbf{p}_0 \leftarrow \mathbf{p}_1$

$\quad\mathbf{p}_1 \leftarrow \eta(\mathbf{P},\mathbf{Q})$

$\quad\mathbf{q} \leftarrow \left\{\eta^2(\mathbf{P},\mathbf{Q})f(t),\ \eta\prime(\mathbf{P},\mathbf{Q})\right\}$

end

Algorithm 5: AMP (Structured noise)

Input : $\lambda, \rho, \mathbf{Y}, \delta$

$\mathbf{p}_0 \leftarrow \mathbf{X}_{i\leq N} \sim \mathcal{N}(0,\varepsilon)$

$\mathbf{p}_1 \leftarrow \mathbf{X}_{i\leq N} \sim \mathcal{N}(0,\varepsilon)$

$\mathbf{q}_1 \leftarrow \mathbf{X}_{i\leq N} \sim \mathcal{N}(0,\varepsilon)$

while $(\mathbf{p}^1 - \mathbf{p}^0)\cdot \mathbf{1}/N > \delta$ **do**

$\quad\mathbf{P} \leftarrow \mathbf{J}(\mathbf{Y})\mathbf{p}_t - \sum_{i=1}^t f_t(i)\mathbf{p}_i$

$\quad\mathbf{p}_{t-1} \leftarrow \mathbf{p}_t$

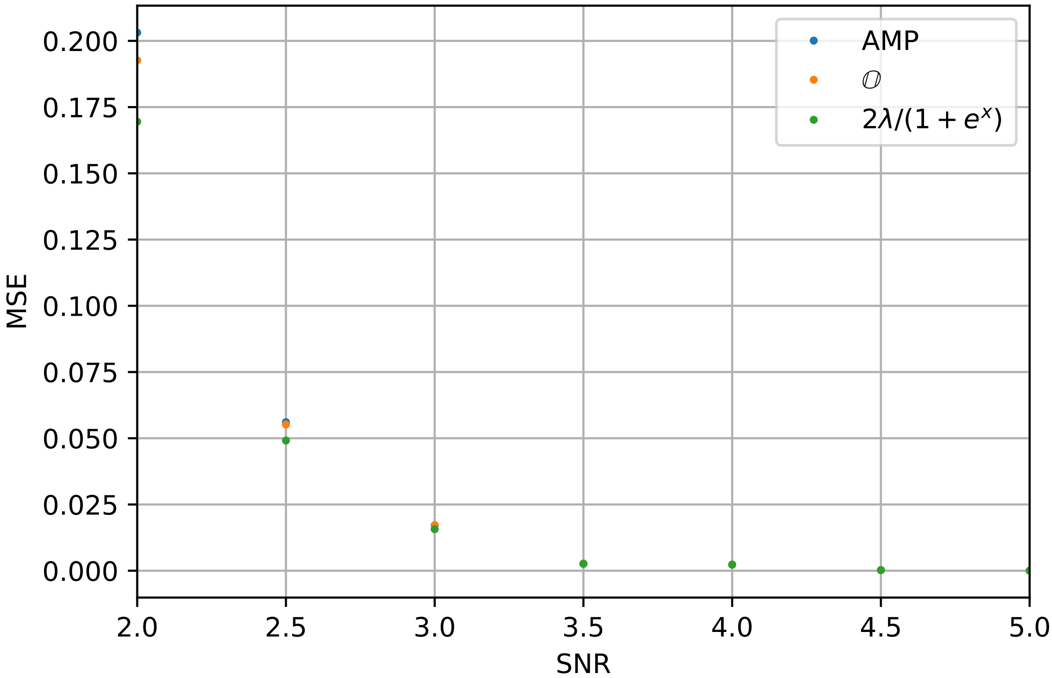
$\quad\mathbf{p}_t \leftarrow \eta(\mathbf{P})$

$\quad\mathbf{q} \leftarrow \left\{\eta^2(\mathbf{P})f(t),\ \eta\prime(\mathbf{P})\right\}$

end

Results

The


AMP
$\mathbb{O}$
$2\lambda/(1+e^x)$
MSE
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
2.0
2.5
3.0
3.5
4.0
4.5
5.0
SNR

Chapter 4

Conclusions

We succesfully reproduced the well known results of the Spiked Wigner model. By applying the prior independent scheduler in the Spiked Wigner model resulted succesful in reaching the Bayes-optimality. The application of the prior independent scheduler resulted in longer iterations than the well known procedure of AMP. From these results we are able to conclude that the AMP algorithm, at least for the Spiked Wigner model is robust enough to converge to the spike after several iterations. Due to the time and computational limitations of this quick research, the author invites the reader for a thorough reproduction of these results as well as for the Structured noise procedure.

The application of a prior independent scheduler following the well known procedure of the AMP in the Spiked Wigner model with Structured noise, resulted in a suboptimal convergence. This algorithm was not capable of reaching the optimal convergence of the well known AMP, nor the Bayes-optimal procedure for the Structured noise. It is possible that the algorithm was incorrectly implemented and if corrected the algorithm does converge. Once again, due to the time and computational limitations of this quick research, the author invites the reader for a thorough reproduction of these results reader.

If the prior independent scheduler does converge to the well known AMP, this would open an interesting question regarding the usage of the denoiser as we know it. This would also result in newer algorithms which are able to recover the spike and a family of schedulers with different efficacies.

When $X_i^* \sim \mathrm{Beta}(\alpha,\beta) := \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \alpha,\beta \in \mathbb{N}_+$:

$$
\begin{aligned}
\eta(b,v) =& \frac{\int_0^1 dx x^{\alpha}(1-x)^{\beta-1}e^{xb-x^2v/2}}{\int_0^1 dx x^{\alpha-1}(1-x)^{\beta-1}e^{xb-x^2v/2}} \\
\eta\prime(b,v) =& \frac{\left(\int_0^1 dx x^{\alpha-1}(1-x)^{\beta-1}e^{xb-x^2v/2}\right)\left(x^{\alpha}(1-x)^{\beta-1}e^{xb-x^2v/2}\right) - \left(\int_0^1 dx x^{\alpha}(1-x)^{\beta-1}e^{xb-x^2v/2}\right)\left(x^{\alpha-1}(1-x)^{\beta-\cdots}\right.}{\left(\int_0^1 dx x^{\alpha-1}(1-x)^{\beta-1}e^{xb-x^2v/2}\right)^2} \\
=& e^{xb-x^2v/2}x^{\alpha-1}(1-x)^{\beta-1}\frac{\left(\int_0^1 dx x^{\alpha-1}(1-x)^{\beta-1}e^{xb-x^2v/2}\right)x - \left(\int_0^1 dx x^{\alpha}(1-x)^{\beta-1}e^{xb-x^2v/2}\right)}{\left(\int_0^1 dx x^{\alpha-1}(1-x)^{\beta-1}e^{xb-x^2v/2}\right)^2} \\
=& \left(e^{xb-x^2v/2}x^{\alpha-1}(1-x)^{\beta-1}\frac{\left(\int_0^1 dx x^{\alpha-1}(1-x)^{\beta-1}e^{xb-x^2v/2}\right)x - \left(\int_0^1 dx x^{\alpha}(1-x)^{\beta-1}e^{xb-x^2v/2}\right)}{\left(\int_0^1 dx x^{\alpha}(1-x)^{\beta-1}e^{xb-x^2v/2}\right)^2}\right)\eta^2(b,v) \\
=& \left(e^{xb-x^2v/2}x^{\alpha-1}(1-x)^{\beta-1}\frac{\left(\int_0^1 dx x^{\alpha-1}(1-x)^{\beta-1}e^{xb-x^2v/2}\right)x - x\left(\int_0^1 dx x^{\alpha-1}(1-x)^{\beta-1}e^{xb-x^2v/2}\right) + \int_0^1 dx \cdots}{\left(\int_0^1 dx x^{\alpha}(1-x)^{\beta-1}e^{xb-x^2v/2}\right)^2}\right. \\
=& \left(e^{xb-x^2v/2}x^{\alpha-1}(1-x)^{\beta-1}\frac{\int_0^1 dx \int_0^1 dx x^{\alpha-1}(1-x)^{\beta-1}e^{xb-x^2v/2}}{\left(\int_0^1 dx x^{\alpha}(1-x)^{\beta-1}e^{xb-x^2v/2}\right)^2}\right)\eta^2(b,v)
\end{aligned} \tag{.1}
$$

$$
\begin{aligned}
&\int_0^1 dx x^{\alpha}(1-x)^{\beta-1}\exp(xb-x^2v/2) = \\
&\qquad x\exp(xb-x^2v/2)\int_0^1 dx x^{\alpha-1}(1-x)^{\beta-1} - \int_0^1 dx x^{\alpha-1}(1-x)^{\beta-1}\exp(xb-x^2v/2)(1+x(b\cdots \\
&\qquad = \\
&\quad x\exp(xb-x^2v/2)\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} - \int_0^1 dx x^{\alpha-1}(1-x)^{\beta-1}\exp(xb-x^2v/2) - \int_0^1 dx x^{\alpha}(1-x)^{\beta-1}\exp(xb-x^2v/\cdots \\
&(b+1)\int_0^1 dx x^{\alpha}(1-x)^{\beta-1}\exp(xb-x^2v/2) = \\
&\quad x\exp(xb-x^2v/2)\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} - \int_0^1 dx x^{\alpha-1}(1-x)^{\beta-1}\exp(xb-x^2v/2) + v\int_0^1 dx x^{\alpha+1}(1-x)^{\beta-1}\exp(xb - \cdots
\end{aligned}
$$

$$
\begin{aligned}
\eta\prime(b,v) =& \frac{1+\exp(v/2)\cosh(b)(1-\rho)/\rho}{(\cosh(b)+\exp(v/2)(1-\rho)/\rho)^2} \\
=& \left(\frac{1}{\sinh^2(b)} + \frac{1-\rho}{\rho}\frac{\exp(v/2)}{\sinh(b)\tanh(b)}\right)\eta^2(b,v)
\end{aligned} \tag{.2}
$$

$$
\begin{aligned}
\mathbf{p}^{t+1} =& \eta\left(\sqrt{\frac{\lambda}{N}}\mathbf{Y}\mathbf{p}^t - \left[\sqrt{\frac{\lambda}{N}}\mathbf{Y}f(t)\circ\mathbf{p}^t\right]^2 \circ \mathbf{p}^{t-1}, \frac{\lambda}{N}\left(\mathbb{I}_N - \mathbf{Y}^2\right)(f(t)\circ\mathbf{p}^t)^2 + \frac{\lambda}{N}\|\mathbf{p}^t\|^2\mathbf{1}\right) \\
=& \frac{\tanh\left(\sqrt{\frac{\lambda}{N}}\mathbf{Y}\mathbf{p}^t - \left[\sqrt{\frac{\lambda}{N}}\mathbf{Y}f(t)\circ\mathbf{p}^t\right]^2\circ\mathbf{p}^{t-1}\right)}{1+2(1-\rho)/(\rho(\exp(b-v/2)+\exp(-b-v/2)))} \\
=& \frac{\tanh\left(\sqrt{\frac{\lambda}{N}}\mathbf{Y}\left[\mathbf{1}-\sqrt{\frac{\lambda}{N}}f^2(t)\circ\mathbf{p}^t\circ\mathbf{p}^{t-1}\right]\circ\mathbf{p}^t\right)}{1+\frac{1-\rho}{\rho}\frac{\exp\left[\frac{\lambda\left((\mathbb{I}_N-\mathbf{Y}^2)(f(t)\circ\mathbf{p}^t)^2+\|\mathbf{p}^t\|^2\mathbf{1}\right)}{2N}\right]}{\cosh\left(\sqrt{\frac{\lambda}{N}}\mathbf{Y}\left[\mathbf{1}-\sqrt{\frac{\lambda}{N}}f^2(t)\circ\mathbf{p}^t\circ\mathbf{p}^{t-1}\right]\circ\mathbf{p}^t\right)}}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{p}^{t+1} =& \eta(\mathbf{P}^t, \mathbf{Q}^t), \quad \mathbf{q}^{t+1} = f^2(t+1) \circ \eta^2(\mathbf{P}^t, \mathbf{Q}^t), \quad (\mathbf{v})^2 \equiv \mathbf{v} \circ \mathbf{v}, \quad \|\mathbf{v}\|^2 = \mathbf{v}^2 \cdot \mathbf{1} \\
\mathbf{p}^{t+1} =& \eta(\mathbf{P}^t, \mathbf{Q}^t), \quad \mathbf{q}^{t+1} = \left(f(t+1) \circ \mathbf{p}^{t+1}\right)^2 \\
\mathbf{p}^{t+1} =& \eta\left(\sqrt{\frac{\lambda}{N}}\mathbf{Y}\mathbf{p}^t - \left[\frac{\lambda}{N}\mathbf{Y}^2\mathbf{q}^t\right] \circ \mathbf{p}^{t-1}, \frac{\lambda}{N}\left(\mathbf{q}^t + \|\mathbf{p}^t\|^2\mathbf{1}\right) - \left[\frac{\lambda}{N}\mathbf{Y}^2\mathbf{q}^t\right]\right) \\
=& \eta\left(\sqrt{\frac{\lambda}{N}}\mathbf{Y}\mathbf{p}^t - \left[\sqrt{\frac{\lambda}{N}}\mathbf{Y}f(t) \circ \mathbf{p}^t\right]^2 \circ \mathbf{p}^{t-1}, \left(\sqrt{\frac{\lambda}{N}}f(t) \circ \mathbf{p}^t\right)^2 + \frac{\lambda}{N}\|\mathbf{p}^t\|^2\mathbf{1} - \left[\sqrt{\frac{\lambda}{N}}\mathbf{Y}f(t) \circ \mathbf{p}^t\right]^2\right) \\
=& \eta\left(\sqrt{\frac{\lambda}{N}}\mathbf{Y}\mathbf{p}^t - \left[\sqrt{\frac{\lambda}{N}}\mathbf{Y}f(t) \circ \mathbf{p}^t\right]^2 \circ \mathbf{p}^{t-1}, \frac{\lambda}{N}\left(\mathbb{I}_N - \mathbf{Y}^2\right)\left(f(t) \circ \mathbf{p}^t\right)^2 + \frac{\lambda}{N}\|\mathbf{p}^t\|^2\mathbf{1}\right) \\
=& \left(1 + \frac{1-\rho}{\rho}\exp\left[\frac{\|\mathbf{p}^t\|^2\mathbf{1} - 2\sqrt{N/\lambda}\mathbf{Y}\mathbf{p}^t - (\mathbf{Y}^2 - 2\mathbf{Y}^2\mathbf{p}^{t-1}\mathbf{1}^\intercal \circ \mathbb{I}_N - \mathbb{I}_N)(f(t) \circ \mathbf{p}^t)^2}{2N/\lambda}\right]\right) \\
=& \left(1 + \frac{1-\rho}{\rho}\exp\left[\frac{\mathbf{G}^\intercal\mathbf{G}\mathbf{1} - \mathbf{Y}^\intercal\mathbf{Y}\mathbf{1}/\lambda - (\mathbf{Y}^2 - 2\mathbf{Y}^2\mathbf{p}^{t-1}\mathbf{1}^\intercal \circ \mathbb{I}_N - \mathbb{I}_N(f(t) \circ \mathbf{p}^t)^2}{2N/\lambda}\right]\right) \\
=& \left(1 + \frac{1-\rho}{\rho}\exp\left[-\frac{(a-(b\pm c)^2)^2}{2N/\lambda}\right]\right)
\end{aligned}
$$

$$
\mathbf{a} \circ \mathbf{b} = \mathbf{a}\mathbf{1}^\intercal\mathbf{b}, \quad a\mathbf{1}\mathbf{1}^\intercal\mathbf{1} = Na\mathbf{1}
$$

$$
\begin{aligned}
(\mathbf{Y}/\sqrt{\lambda} - \mathbf{p}^t\mathbf{1}^\intercal/\sqrt{N})^\intercal(\mathbf{Y}/\sqrt{\lambda} - \mathbf{p}^t\mathbf{1}^\intercal/\sqrt{N})\mathbf{1} =& (\mathbf{Y}^\intercal\mathbf{Y}/\lambda - \mathbf{Y}^\intercal\mathbf{p}^t\mathbf{1}^\intercal/\sqrt{N\lambda} - (\mathbf{p}^t\mathbf{1}^\intercal)^\intercal\mathbf{Y}/\sqrt{N\lambda} + (\mathbf{p}^t\mathbf{1}^\intercal)^\intercal\mathbf{p}^t\mathbf{1}^\intercal/N)\mathbf{1} \\
=& (\mathbf{Y}^\intercal\mathbf{Y}/\lambda - \mathbf{Y}^\intercal\mathbf{p}^t\mathbf{1}^\intercal/\sqrt{N\lambda} - \mathbf{Y}^\intercal\mathbf{p}^t\mathbf{1}^\intercal/\sqrt{N\lambda} + \mathbf{1}(\mathbf{p}^t)^\intercal\mathbf{p}^t\mathbf{1}^\intercal/N)\mathbf{1} \\
=& (\mathbf{Y}^\intercal\mathbf{Y}/\lambda - \mathbf{Y}^\intercal\mathbf{p}^t\mathbf{1}^\intercal/\sqrt{N\lambda} - \mathbf{Y}^\intercal\mathbf{p}^t\mathbf{1}^\intercal/\sqrt{N\lambda} + \|\mathbf{p}\|^2\mathbf{1}\mathbf{1}^\intercal/N)\mathbf{1} \\
=& (\mathbf{Y}^\intercal\mathbf{Y}/\lambda - 2/\sqrt{N\lambda}\mathbf{Y}^\intercal\mathbf{p}^t\mathbf{1}^\intercal + \|\mathbf{p}\|^2\mathbf{1}\mathbf{1}^\intercal/N)\mathbf{1} \\
=& \mathbf{Y}^\intercal\mathbf{Y}\mathbf{1}/\lambda - 2\sqrt{N/\lambda}\mathbf{Y}\mathbf{p}^t + \|\mathbf{p}\|^2\mathbf{1} \\
\mathbf{G}^\intercal\mathbf{G}\mathbf{1} =&
\end{aligned}
$$

The following equations summarizes the different cavity fields to be used:

$$
\begin{aligned}
\mathbf{P}^t &= \sqrt{\frac{\lambda}{N}}\mathbf{Y}\mathbf{p}^t - \left[\frac{\lambda}{N}\mathbf{Y}^2\mathbf{q}^t\right] \circ \mathbf{p}^{t-1}, \quad \mathbf{Q}^t = \frac{\lambda}{N}\left(\mathbf{q}^t + \|\mathbf{p}^t\|^2\mathbf{1}\right) - \left[\frac{\lambda}{N}\mathbf{Y}^2\mathbf{q}^t\right] \\
\mathbf{P}^t_{\mathbb{E}} &:= \sqrt{\frac{\lambda}{N}}\mathbf{Y}\mathbf{p}^t - \mathbb{E}\left[\frac{\lambda}{N}\mathbf{Y}^2\mathbf{q}^t\right]\mathbf{p}^{t-1}, \quad \mathbf{Q}^t_{\mathbb{E}} := \frac{\lambda}{N}\left(\mathbf{q}^t + \|\mathbf{p}^t\|^2\mathbf{1}\right) - \mathbb{E}\left[\frac{\lambda}{N}\mathbf{Y}^2\mathbf{q}^t\right]\mathbf{1} \\
\mathbf{P}^t_{\mathbb{O}} &:= \sqrt{\frac{\lambda}{N}}\mathbf{Y}\mathbf{p}^t - \mathbb{O}(t)\mathbf{p}^{t-1}, \quad \mathbf{Q}^t_{\mathbb{O}} := \frac{\lambda}{N}\left(\mathbf{q}^t + \|\mathbf{p}^t\|^2\mathbf{1}\right) - \mathbb{O}(t)\mathbf{1} \\
\mathbf{P}^t_{f} &:= \sqrt{\frac{\lambda}{N}}\mathbf{Y}\mathbf{p}^t - f(t)\mathbf{p}^{t-1}, \quad \mathbf{Q}^t_{f} = \frac{\lambda}{N}\left(\mathbf{q}^t + \|\mathbf{p}^t\|^2\mathbf{1}\right) - f(t)\mathbf{1}
\end{aligned}
$$

We recall that the denoiser is a scalar function component-wise defined [cite]:

$$
\mathbb{R}^N \times \mathbb{R}^N \ni \mathbf{v} \mapsto \eta_t(\mathbf{v}) = (\eta_t(v_1), \ldots, \eta_t(v_N))^T
$$

Uniform

$$
\frac{e^{-(\nu+4\beta\rho)/(8\rho^2)}\left(2(1-e^{\beta/\rho})/\sqrt{2\pi\nu} + e^{(\nu+2\beta\rho)^2/(8\nu\rho^2)}\frac{\beta}{\nu}\left[erf\left(\frac{\nu-2\beta\rho}{2\sqrt{2\nu}\rho}\right) + erf\left(\frac{\nu+2\beta\rho}{2\sqrt{2\nu}\rho}\right)\right]\right)}{e^{\beta^2/(2\nu)}\left[erf\left(\frac{\nu-2\beta\rho}{2\sqrt{2\nu}\rho}\right) - erf\left(\frac{-\beta-\nu/(2\rho)}{\sqrt{2\nu}}\right)\right]}
$$

Normal