Sampling from Spin Glass Gibbs Measures via Anisotropic Gaussian Processes

Victor Kawasaki-Borruat[*], Yuansi Chen[†]

Abstract

This paper offers an alternative solution to the problem of sampling from the Gibbs measure of spin glasses in the high-temperature regime with no external field using diffusion processes. In particular, we propose a sampling scheme which is driven by an anisotropic Gaussian process. With careful choice of the covariance matrix, we are able to show that while we are sampling from a Gibbs measure of inverse temperature β_*, the algorithm tracks an *effective inverse temperature* $\beta_{\text{eff}} \leq \beta_*$. We show equivalence of this sampling method to previous ones for the Sherrington-Kirkpatrick model, as well as also discuss potential breaching of conjectured barriers to current diffusion-based sampling of p-spin glasses, thanks to this discrepancy between the sampled and effective inverse temperatures.

[*]Department of Electrical Engineering and Information Technology, ETH Zurich
[†]Seminar for Statistics, Department of Mathematics, ETH Zurich

Contents

1 Introduction

The Sherrington-Kirkpatrick (SK) model [SK75] is a historically important toy example of a spin glass , as it was introduced as the first solvable model of its kind. It is described by a collection of $n \in \mathbb{N}$ 'spins' pointing either 'up' or 'down' (± 1) on a $n \times n$ lattice. For a given configuration of the system $\mathbf{x} \in \Sigma_n := \{-1, +1\}^n$, the Hamiltonian of the SK model is given by

$$H_n(\mathbf{x}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J_{i,j} \mathbf{x}_i \mathbf{x}_j, \tag{1.1}$$

where $J_{i,j}$ are independent and identically distributed (i.i.d.) standard Gaussian random variables. The Hamiltonian is thus a random function, and enjoys a 'distributional' symmetry around zero. In the statistical physics literature, a quantity of particular interest is the *Gibbs measure* of a system, which corresponds to the equilibrium distribution of the system in configuration space for a given average energy [1] $\langle H_n(\mathbf{x}) \rangle$. By a classical result, the Gibbs measure $\mu(\mathbf{x}) \propto \exp(-\beta H_n(\mathbf{x}))$, where $\beta > 0$ is the inverse temperature of a system. For all means and purposes, we will simply consider β as a parameter of the problem of interest. The SK Gibbs measure is thus given by a probability distribution over Σ_n as follows:

$$\mu_{\mathbf{A}}(\mathbf{x}) := \frac{1}{Z(\mathbf{A}, \beta)} \exp\left(\frac{\beta}{2} \langle \mathbf{x}, \mathbf{A}\mathbf{x} \rangle\right), \tag{1.2}$$

where $\beta \geq 0$ is the inverse temperature of the model, $\mathbf{A} \sim \mathrm{GOE}(n)$ is drawn from the *Gaussian orthogonal ensemble*. That is, $\mathbf{A} = (\sqrt{2n})^{-1}(\mathbf{G} + \mathbf{G}^\intercal)$, where $\mathbf{G} \in \mathbb{R}^{n \times n}$ is a matrix with i.i.d. entries drawn from a standard normal distribution. We thus have $A_{i,j} = A_{j,i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/n)$ for $i \neq j$ and $A_{i,i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 2/n)$. $Z(\mathbf{A}, \beta)$ is a normalization factor, typically referred to as *partition function* by physicists.

1.1 Background: Sampling and Spin Glasses

This paper will consider the problem of efficiently sampling from the above Gibbs measure. Typically, the problem of sampling such physical systems has been treated with Markov Chain Monte Carlo (MCMC) methods. In this case, Glauber dynamics ([Gla63]) – also known as Gibbs sampling – is a very well-studied and popular choice of sampler. This method simulates the evolution of the system up to its equilibrium state by sequentially randomly choosing a spin, and flipping it with a probability proportional to the energy decrease said flip would induce. However, while physics literature points to fast convergence to equilibrium for certain observables for any $\beta < 1$, little rigorous proofs thereof exist. Bauerschmidt and Bodineau's work in [BB19] implies a sampling algorithm for $\beta < 1/4$, and Eldan, Koehler, Zeitouni [EKZ22] demonstrate that in the same region $\beta < 1/4$ for Ising models with quadratic Hamiltonians the mixing of Glauber dynamics happens in $\mathcal{O}(n^2)$ spin flips. This was proven by satisfying a Poincaré inequality, directly allowing control of the mixing time. The latter result was improved by [Ana+21] to $\mathcal{O}(n \log(n))$ by a modified log-Sobolev inequality. Such difficulty to bound mixing times arises from specific physical conditions. Namely, the Almeida-Thouless (AT) line [AT78; AM22; Tal11; MPV87] with no external field

$$1 = \beta^2 \int \frac{\cosh(\beta z)}{\cosh^2(\beta z) - 1} \phi(z) \mathrm{d}z, \quad \phi := \text{field distribution}, \tag{1.3}$$

[1] We denote the average energy here by $\langle \cdot \rangle$ for consistency with physics notation. However, throughout the rest of the paper $\langle \cdot, \cdot \rangle$ will denote standard scalar products in $\mathbb{R}^n$.

is the conjectured line in the temperature-magnetic field plane separating the high-temperature regime and the so-called glassy phase. In the SK model, ϕ is the standard Gaussian distribution, and this line is thus given by $\beta = 1$. Above this line (i.e. $\beta < 1$), the SK model exhibits replica symmetry [Pan12], allowing a simpler analysis of the system (hence rapid mixing). Moreover, in the high temperature regime, the SK model's energy landscape (described by the TAP equations [TAP77]) is smoother, and local minima are less harsh [Pan12; MPV87]. This allows energy fluctuations to escape local minima, resulting in faster mixing to equilibrium [AH87], supporting the aforementioned claim.

1.2 Background: Diffusion-Based Sampling

Our approach will be different. Rather than sample with traditional MCMC methods, we follow the idea presented in [AMS24], [NSZ22] by sampling from the Gibbs measure using diffusion processes (originally motivated by stochastic localization). This problem has successfully been treated and solved for $\beta < \frac{1}{2}$ and improved up to $\beta < 1$ in [AMS24] and [Cel22] respectively, by sampling with diffusion processes and running Approximate Message Passing (AMP) and Natural Gradient Descent (NGD) as subroutines.

Diffusion processes have been quite prevalent in the Machine Learning (particularly Deep Learning) literature, labelled as denoising diffusion probabilistic models (DDPMs) [HJA20], [Soh+15], [Son+21], [Cam+22]. These generative models typically work by slowly injecting noise to samples, eventually destroying all structure in the samples. Then, when prompted to generate a sample, the model reverses the process, initialized with random noise. The result is an approximate sample of the desired distribution (e.g. a particular image, a person's voice). The mathematical foundations are explained in Section 2. The reversal is possible by knowing the score of the prior distribution [Che+23], which is typically learned with a Neural Network (NN) and using massive amounts of data modelling the typically unknwon prior $p_{\text{data}} \approx p_{\text{prior}}$. Song and Ermon [SE20] proposed a Noise Conditional Score Network to estimate the score, and used Langevin MCMC to produce samples. In [Son+21], the framework of DDPMs was extended to SDEs. This method still requires NNs to solve the forward score (denoising score matching), or various other methods, such as slice score matching [Son+19]. In this paper, we will leverage the fact that our prior is known (albeit random) for a given parameter $\mathbf{A}$, and the structure of the forward diffusion process to circumvent the need for NNs. In this paper, we instead we use AMP to estimate the score, due to Proposition 2.1. This offers a different perspective on sampling from known yet complex distributions without relying on learning, and also ties into important statistical estimation problems [Mon23; LKZ17]. Since AMP will be the main algorithm of interest in this paper, we offer a short historical introduction to the topic. As well as the current state of research.

1.3 Background: Approximate Message Passing

Approximate Message Passing was first inspired by the work of Bolthausen [Bol12], in which an iterative solution to the TAP equations for the SK model is presented. It highlighted interesting dependencies between the mean-spins m_i and the random couplings g_{ij}, and explained the Onsager term as an Itô-type correction. The key concept introduced was denoted 'conditioning technique', and allowed for statistical simplifications in the asymptotic regime. This further inspired Bayati and Montanari to develop a generalization of message-passing algorithms for compressed sensing [BM11], an extension to the proposed algorithms in [DMM09]. Message-passing algorithms are popular in the subject of Coding Theory, and provided an interesting tool called density evolution [RU08], which held only on locally tree-like sparse graphs. These new algorithms were shown to successfully approximate message passing algorithms in the large system limit, and exhibited an analogous one-dimensional *state evolution*, but on dense graphs. Since then, AMP has received attention as a general-purpose statistical estimation algorithm, mostly regarded as a 'generalized power method'. Other than in compressed sensing [DJM13], AMP has found success in various topics. Of which include communications (multi-user detection) [Jeo+15], [Kab03] and low-rank matrix estimation [MV19], [LKZ17], [MT13], [DAM15]. The latter having shown interesting ties to our problem of spin glasses [Les+17], [AMS24]. Ties between the TAP free energy $\mathcal{F}_{\text{TAP}}$ of the SK model and the statistical estimation problem of $\mathbb{Z}_2$-synchronization have been the subject of recent research [DAM15; FMM20; CFM23]. The TAP free energy functional's global minimizer $\mathbf{m}_*$ is conjectured to be the Bayes-optimal estimator of a signal $\mathbf{x}_0 \in \mathbb{R}^n$ given a *Wigner spike model* observation:

$$\mathbf{Y} = \mathbf{x}_0\mathbf{x}_0^{\intercal} + \mathbf{W}, \quad \mathbf{W} \sim \text{GOE}(n). \tag{1.4}$$

It is a known result that AMP a local minimum of $\mathcal{F}_{\text{TAP}}$ in a finite number of iterations, but it is still an open problem to determine whether it actually converges to it. In [CFM23] the statement is conjectured to be true, supported by numerical evidence. Moreover, the AMP mapping is shown to be stable at the global minimum. Recently, Celentano also proved in [Cel22] that the TAP free energy in the above spike estimation problem is locally convex around the region to which AMP converges.

1.4 Contents and Contributions

We propose a variation on the sampling algorithm, which exhibits equivalent precision and complexity in SK models, and shows potential improvements in efficiently sampling from less trivial spin glass Gibbs measures. Namely, we point towards a breaching of the barriers to sampling with diffusion processes presented in [LKZ17], [Ghi+23] for p-spin Ising models.

The core of our technique acknowledges the equivalence of Gaussian diffusion processes and stochastic localization of [Eld19]; first presented by Alaoui et al. in [AMS24] and generalized in [Mon23] to further types of processes. We leverage the use of *anisotropic Gaussian noise* and control its covariance matrix to efficiently sample from a certain inverse temperature β_*, while running the algorithm on an *effective* inverse temperature $\beta_{\text{eff}} < \beta_*$. The current barriers are due to AMP being algorithmically difficult in the metastable regions of the replica symmetric (RS) entropy. The discrepancy between β_* and the effective inverse temperature could allow the algorithm trajectory to avoid such troublesome regions (see Figure 2 for a visual explanation). Most importantly, we present a new 'hybrid' version of AMP, which will be the main topic of study in this paper, and is crucial in the context of sampling using anisotropic Gaussian noise.

As explained above, $\beta_{\max} = 1$ is a bound that cannot be improved upon for the SK model, and has already been reached. In the p-spin case, however, methods so far allow us to sample using diffusion processes up to a certain β_1 ([AMS23]), which is strictly lower than the 'dynamical phase transition critical temperature' β_{dyn}, shown by [Ghi+23] to be the actual theoretical limit to diffusion-based sampling. This will be the subject of discussion of Section 3.2. Current literature [AMS23] also suggests algorithmic stability up to $\beta_c \geq \beta_{\text{dyn}} > \beta_1$, allowing still much room for improvement as alluded to above. We first present the general theory of diffusion processes, and how they may be employed for sampling purposes, also thanks to their ties with stochastic localization. Next, we present the case of sampling using anisotropic processes, which is also a known result ([Mon23]), and offer a specific construction for the covariance matrix, allowing us to efficiently sample from the SK Gibbs measure. We then engage in a discussion around the implications of our result, and provide examples from existing literature demonstrating the intuition behind our hypothesis. We then replicate Section 4 of [AMS24], which proves our main Theorem revolving around the convergence of the new AMP algorithm. General background information on AMP is provided in Appendix A.

1.5 Notation

Most of this paper will revolve around random variables and vectors. We typically denote one-dimensional random variables with capital letters e.g. X, Y, while random vectors will be lowercase bold letters e.g. $\mathbf{x}, \mathbf{y}$. Expectations $\mathbb{E}$ are taken with the respect to the most explicit distributions or joint distributions. $X \sim \mathrm{Rad}$ denotes a symmetric Bernoulli (or Rademacher) random variable, distributed as

$$P(X = 1) = P(X = -1) = \frac{1}{2}. \tag{1.5}$$

Matrices of all sorts will be denoted with bold uppercase letter e.g. $\mathbf{A}, \mathbf{H}$. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix, and $\mathbf{1}_n$ the n-dimensional all-ones vector. A single exception will be $\mathbf{B}(t)$, used to denote multidimensional Brownian motion. We will use $o_n(1)$ to denote a quantity tending to zero in probability. $\mathcal{B}(\mathbf{x}, \delta)$ denotes the n-dimensional δ-ball around a point $\mathbf{x} \in \mathbb{R}^n$.

2 Sampling with Diffusion Processes

The idea of sampling with diffusion processes, first introduced by Sohl-Dickstein in [Soh+15], was motivated by time-reversal, and mostly popularized in the machine learning literature in [HJA20]. [Son+21] then adapted the framework to be modelled with SDEs, and [Che+23] solidified the theory, effectively reducing the problem of sampling from a distribution to that of estimating its score (gradient of the log-likelihood). This section will act as a primer on sampling with diffusion processes, as well as establish the connection to stochastic localization as in [Mon23]. Then, we will recall the method presented in [AMS24] to sample from the SK Gibbs measure, and present our new method.

2.1 Forward & Reverse Process

The forward process will be the process turning samples of our target distribution to noise. In particular, we will consider a general one-dimensional diffusion process:

$$\mathrm{d}X_t = f_t(X_t)\mathrm{d}t + \sigma_t \mathrm{d}W_t, \tag{2.1}$$

where f_t is a drift term, σ_t is the diffusion coefficient, and W_t is a standard Brownian motion. We can then reverse (2.1) over $t \in [0, T]$ by considering the process

$$X_t^{\leftarrow} := X_{T-t}, \quad t \in [0, T]. \tag{2.2}$$

It follows from [And82] that the reverse process is also a diffusion process, given by

$$\mathrm{d}X_t^{\leftarrow} = f_t^{\leftarrow}(X_t^{\leftarrow})\mathrm{d}t + \sigma_{T-t}\mathrm{d}W_t, \tag{2.3}$$

where the reverse drift term satisfies (we omit arguments for ease of notation):

$$f_t + f_{T-t}^{\leftarrow} = \sigma_t \sigma_t^{\intercal} \nabla \ln(q_t), \quad X_t \sim q_t. \tag{2.4}$$

In particular, we get that the reverse SDE is completely described by the law of the forward process.

The Ornstein-Uhlenbeck (OU) process is a well-known process which turns any starting distribution to Gaussian noise exponentially fast. For any initial distribution $\mu \in \mathcal{P}(\mathbb{R})$, it is given by

$$\mathrm{d}X_t := -X_t \mathrm{d}t + \sqrt{2}\mathrm{d}B_t, \quad X_0 \sim \mu \tag{2.5}$$

and its reverse is thus

$$\begin{aligned} \mathrm{d}X_t^{\leftarrow} &:= \left[X_t^{\leftarrow} + 2\nabla \ln\left(q_{T-t}(X_t^{\leftarrow})\right)\right]\mathrm{d}t + \mathrm{d}B_t \\ &= \left[X_t^{\leftarrow} + 2\nabla \ln\left(q_{T-t}(X_{T-t})\right)\right]\mathrm{d}t + \mathrm{d}B_t. \end{aligned} \tag{2.6}$$

Proposition 2.1 (Tweedie's Formula [Efr11])**.** *For $Y = X + Z$, where $Z \sim \mathcal{N}(0, \gamma^2)$, we have that*

$$\mathbb{E}[X|Y = y] = y + \gamma^2 \nabla_y \ln(p_Y(y)) \tag{2.7}$$

Proof. For $X \sim \mu$ and $Y = X + Z$, $Z \sim \mathcal{N}(0, \gamma^2)$, it holds that

$$p_Y(y) = \int \mu(x) \frac{1}{\sqrt{2\pi}\gamma} \exp\left(\frac{(y-x)^2}{2\gamma^2}\right) \mathrm{d}x. \tag{2.8}$$

Thus,

$$
\begin{aligned}
\frac{\mathbb{E}[X|Y=y]-y}{\gamma^2} &= \mathbb{E}\left[\frac{X-Y}{\gamma^2}|Y=y\right] \\
&= \int \frac{x-y}{\gamma^2} p_{X|Y=y}(x)\mathrm{d}x \\
&= \int \left(\frac{x-y}{\gamma^2}\right) \frac{p_{X,Y}(x,y)}{p_Y(y)}\mathrm{d}x \\
&= \int \left(\frac{x-y}{\gamma^2}\right) \frac{\mu(x)p_\gamma(y-x)}{p_Y(y)}\mathrm{d}x
\end{aligned} \tag{2.9}
$$

with $p_\gamma(\xi) = \phi(\frac{\xi}{\gamma})$ the normal probability distribution function. $\square$

Now notice that if we integrate the OU process (2.5) using Itô's formula,

$$
X_t = e^{-t}X_0 + \sqrt{1-e^{-2t}}Z, \quad X_0 \sim \mu, \tag{2.10}
$$

we now can apply Tweedie's formula,

$$
\nabla \ln(q_t(X_t)) = \frac{\mathbb{E}\left[e^{-t}x|e^{-t}x + \sqrt{1-e^{-2t}}Z = y\right] - y}{1-e^{-2t}} \tag{2.11}
$$

giving us the full reverse SDE:

$$
\mathrm{d}X_t^{\leftarrow} = -\frac{1+e^{-2(T-t)}}{1-e^{-2(T-t)}}X_t^{\leftarrow}\mathrm{d}t + \frac{2}{e^{-2(T-t)}}\mathbb{E}\left[e^{-(T-t)X_0}|X_t^{\leftarrow} = X_t\right]\mathrm{d}t + \sqrt{2}\mathrm{d}B_t. \tag{2.12}
$$

Applying a change of variables (as in [Mon23]) gives us the following (cleaner) reverse process:

$$
\mathrm{d}X_t^{\leftarrow} = \left(-\frac{1+t}{t(1+t)}X_t^{\leftarrow} + \frac{m(\sqrt{t(1+t)}X_t^{\leftarrow};t)}{\sqrt{t(1+t)}}\right)\mathrm{d}t + \frac{1}{\sqrt{t(1+t)}}\mathrm{d}B_t, \tag{2.13}
$$

where

$$
m(y;t) := \mathbb{E}[x|tx + \sqrt{t}W = y], \quad (x,W) \sim \mu \otimes \mathcal{N}(0,1) \tag{2.14}
$$

is convenient notation for the posterior expectation of x given a Gaussian observation.

2.2 Ties to Stochastic Localization

Now consider a different process, namely

$$
y(t) = tx + W_t, \tag{2.15}
$$

where W_t is a standard Brownian motion. This is a simple stochastic localization process as described in [Eld19]. We can easily write the posterior law of x given observation $y(t)$ using Bayes' Theorem:

$$
\mu_t(x) := \frac{P_t(y(t)|x)\mu(x)}{p(y)} \tag{2.16}
$$

which by characterization of Brownian motion corresponds to a random tilt of the initial probability measure μ:

$$\mu_t(x) = \frac{1}{Z'}\mu(x)\exp\left(-\frac{1}{2t}\|y(t)-x\|_2^2\right)$$
$$= \frac{1}{Z}\mu(x)\exp\left(-\langle y(t),x\rangle - \frac{t}{2}\|x\|_2^2\right) \tag{2.17}$$

Thus, using our construction 2.15 we can conclude from the above calculation that

$$m(y(t);t) = \int x\mu_t(x)\mathrm{d}x, \tag{2.18}$$

i.e. that $m(y(t);t)$ is the mean of the tilted measure.

Proposition 2.2 (Lemma 2.1 [Eld13], Proposition 1.1 [Mon23]). *Now suppose that μ has finite second moment. Then, $(y(t))_{t\geq 0}$ is the unique solution to the following SDE:*

$$\mathrm{d}Y_t = m(Y_t;t)\mathrm{d}t + \mathrm{d}B_t, \quad Y_0 = 0, \tag{2.19}$$

where B_t is standard Brownian motion.

Notice that by the following change of variable:

$$Y_t = \sqrt{t(1+t)}X_t^{\leftarrow}, \tag{2.20}$$

the process described in (2.2) is the same as in (2.13).

It follows from properties of stochastic localization that $m(y(t);t) \to x \sim \mu$ as $t \to \infty$. To produce an approximate sample of the distribution μ, we can thus either output $y(T)/T$ or $m(y(T);T)$ for large enough T, and round the coordinates to be on a corner of the hypercube $\{-1,+1\}^n$.

2.3 Using Isotropic Gaussian Processes

Following the principle above, it is straightforward to generalize to the vector case. We can consider for any $t > 0$ the vector process $\mathbf{y}_t \in \mathbb{R}^n$ given by

$$\mathbf{y}_t = t\mathbf{x} + \mathbf{B}(t), \quad \forall t > 0, \tag{2.21}$$

which is the unique solution to the SDE

$$\mathrm{d}\mathbf{y}_t = \mathbf{m}(\mathbf{y};t)\mathrm{d}t + \mathrm{d}\mathbf{B}(t), \tag{2.22}$$

where $\mathbf{B}(t)$ is a standard n-dimensional Brownian motion and $\mathbf{m}(\mathbf{y};t)$ is the conditional expectation of $\mathbf{x}$ given observation $\mathbf{y}$:

$$\mathbf{m}(\mathbf{y};t) = \mathbb{E}[\mathbf{x}|t\mathbf{x} + \sqrt{t}\mathbf{g} = \mathbf{y}], \quad \mathbf{g} \sim \mathcal{N}(0, \mathbf{I}_n). \tag{2.23}$$

And our tilted measure described in (2.17) is now given by a random tilt of (1.2):

$$\mu_{\mathbf{y},t}(\mathrm{d}\mathbf{x}) := \frac{1}{Z(\mathbf{y})}\exp\left(\langle \mathbf{y},\mathbf{x}\rangle - \frac{t}{2}\|\mathbf{x}\|_2^2\right)\mu_{\mathbf{A}}(\mathrm{d}\mathbf{x}), \tag{2.24}$$

whose mean vector coincides with (2.23). A sampling procedure using this method would require two things:

1. Efficient computation of the posterior mean $\mathbf{m}(\mathbf{y}_t; t)$, which is performed using Approximate Message Passing (AMP) (see [BM11], [DAM15], [Fen+21], implemented in [AMS24])

2. Proper discretization of the SDE.

Indeed, in [AMS24], the use of AMP followed by NGD ensures a precise posterior mean computation (see Theorem 2.3 [CFM23]), and the resulting estimator is proven to be sufficiently regular in $\mathbf{y}$. Moreover, the SDE is discretized using Euler discretization. We now present our method.

2.4 Using Anisotropic Gaussian Processes

Now we will consider a different scheme, proposed in Section 4.2 of [Mon23]. We let our observation process be defined as

$$\mathbf{y}_t = \int_0^t \mathbf{Q}(s)\mathbf{x}\mathrm{d}s + \int_0^t \mathbf{Q}(s)\mathrm{d}\mathbf{B}(s), \tag{2.25}$$

which is the unique solution to the SDE

$$\mathrm{d}\mathbf{y}_t = \mathbf{Q}(t)\mathbf{m}(\mathbf{y}_t; \mathbf{\Omega}(t)) + \mathbf{Q}(t)^{\frac{1}{2}}\mathrm{d}\mathbf{B}(t), \tag{2.26}$$

where $\mathbf{Q}(s)$ is PSD, $\mathbf{\Omega}(t) := \int_0^t \mathbf{Q}(s)\mathrm{d}s$ and

$$\mathbf{m}(\mathbf{y}; \mathbf{\Omega}) := \mathbb{E}[\mathbf{x}|\mathbf{\Omega}\boldsymbol{x} + \mathbf{\Omega}^{\frac{1}{2}}\mathbf{g} = \mathbf{y}]. \tag{2.27}$$

Now following the same logic as above, (2.27) is the mean of the tilted measure

$$\mu_{\mathbf{y},t}(\mathrm{d}\mathbf{x}) = \frac{1}{Z(\mathbf{A}, \mathbf{y})} \exp\left(\langle \mathbf{y_t}, \mathbf{x}\rangle - \frac{t}{2}\|\mathbf{\Omega}(t)\mathbf{x}\|^2_{\Omega(t)^{-1}}\right) \mu_{\mathbf{A}}(\mathrm{d}\mathbf{x}), \tag{2.28}$$

where the dependence of the normalizing factor on β is omitted.

To produce a sample from μ, we apply the same logic as above, which requires us again to overcome two issues:

1. Efficient computation of the posterior mean (2.27), which will be done with a new version of AMP, presented in subsequent sections. This is a main contribution of the present paper.

2. Proper discretization of the SDE (2.26). This is done analogously to [AMS24].

The interest of this method is that, as will be shown later, the covariance matrix can be chosen to our advantage and can help overcome requirements on β.

3 Main Results and Implications on p-spin Models

In this section we will present our findings. Namely, we will present the sampling scheme using anisotropic Gaussian noise, and establish its equivalence to that of [AMS24] for the SK model, as well as address the limitation it can overcome in the case of sampling from p-spin glass Gibbs measures. Our sampling scheme will require a modification of the AMP algorithm, which will be provided, but fully derived proven later in Section 4.

3.1 Sampling Algorithm for $\beta < 1$

Here we will present our sampling algorithm, which will first require computing an estimator of the mean of the tilted measure

$$\mu_{\mathbf{y},t}(\mathrm{d}\mathbf{x}) = \frac{1}{Z(\mathbf{A},\mathbf{y})} \exp\left(\langle \mathbf{y_t}, \mathbf{x} \rangle - \frac{t}{2} \|\mathbf{\Omega}(t)\mathbf{x}\|^2_{\Omega(t)^{-1}} \right) \mu_{\mathbf{A}}(\mathrm{d}\mathbf{x}), \quad \mathbf{x} \in \{-1,+1\}^n. \tag{3.1}$$

We begin the discussion with a proposition on a useful form of $\mathbf{\Omega}(t)$.

Proposition 3.1. *Let $\mathbf{Q}(s) = \lambda \mathbf{I}_n - \beta' \mathbf{A}$ for some $\beta' > 0$ and $\lambda > \beta' \cdot \lambda_{\max}(\mathbf{A})$, then the tilted measure in (2.28) becomes*

$$\mu_{\mathbf{y},t}(\mathrm{d}\mathbf{x}) = \frac{1}{Z(\mathbf{A},\mathbf{y})} \exp\left(\frac{\beta_{\text{eff}}(t)}{2} \langle \mathbf{x}, \mathbf{A}\mathbf{x} \rangle + \langle \mathbf{y}_t, \mathbf{x} \rangle - \frac{t}{2} \|\lambda \mathbf{x}\|_2^2 \right) \tag{3.2}$$

for $\beta_{\text{eff}}(t) := \beta - \beta' t$.

Proof. By the fact that $\lambda_{\max}(\mathbf{A}) \to 2$ as $n \to \infty$ (see standard properties of GOE matrices [Ver18]), we conclude that $\mathbf{Q}(s)$ is well-defined and is positive semidefinite. The claim follows. □

Remark 3.1. *The main point of this anisotropic process is that while we sample from a certain SK measure with inverse temperature β_**

$$\mu_{\mathbf{A},\beta_*}(\mathbf{x}) = \frac{1}{Z(\mathbf{A},\beta_*)} \exp\left(\frac{\beta_*}{2} \langle \mathbf{x}, \mathbf{A}\mathbf{x} \rangle \right), \tag{3.3}$$

the actual tilted measure we are giving to the algorithm is

$$\mu_{\mathbf{y}_t,t}(\mathbf{x}) = \frac{1}{Z(\mathbf{A},\mathbf{y_t})} \mu_{\mathbf{A},\beta_{\text{eff}}(t)}(\mathbf{x}) e^{\langle \mathbf{y}_t, \mathbf{x} \rangle + \frac{t}{2} \|\lambda \mathbf{x}\|_2^2} \tag{3.4}$$

with inverse temperature $\beta_{\text{eff}} \leq \beta_*$ by the above Proposition. We thus allow AMP to appropriately estimate the posterior mean, even in β-ranges previously unreachable. Note that the λ-term in the exponent is absorbed in the normalizer, as we only consider point $\mathbf{x}$ on the corners of the unit hypercube, i.e. $\|\mathbf{x}\|_2^2 = n$. The potentially far-reaching consequences of this will be discussed in Section 3.2.

3.1.1 Approximate Posterior Mean Computation

We will denote our approximate computation of the posterior mean (2.27) by $\widehat{\mathbf{m}}(\mathbf{A},\mathbf{y})$, while the actual posterior mean will be denoted $\mathbf{m}(\mathbf{A},\mathbf{y})$. Similarly to [AMS24], the computation in Algorithm 1 is performed in two steps:

1. An AMP stage, which will compute an estimate of the posterior mean. Notice that it is different from a usual AMP iteration, as it actually computes two AMP iterations in parallel. This is our main novelty, as we use AMP simultaneouly to transform the anisotropic Gaussian side information $\mathbf{y}$ into a 'simpler' side channel, becoming asymptotically 'compatible' with regular AMP iterates. This will be explicited in Section 4.

2. A NGD stage, which will refine our estimation, by leveraging the relationship between Bayes optimality of the posterior mean and the global minimizer of the TAP free energy [TAP77; Tal11]:

Algorithm 1: Mean of Tilted Gibbs Measure

Input: $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\mathbf{y} \in \mathbb{R}^n$, parameters $\beta, \eta > 0$, $q \in (0,1)$, iteration numbers $K_{\text{AMP}}, K_{\text{NGD}}$

$\widehat{\mathbf{m}}^{-1} = \mathbf{z}^0 = 0$,

for $k = 0, ..., K_{\text{AMP}} - 1$ **do**

  $\widehat{\mathbf{m}}^{-1} = \mathbf{z}^0 = 0$,

  $\mathbf{s}^{k+1} = \tanh\left(\mathbf{H}^* \mathbf{r}^k - s^k\right)$,

  $\mathbf{r}^{k+1} = \mathbf{y} - \mathbf{H}\mathbf{s}^k + \frac{1}{n}\sum_{i=1}^n \left(1 - \tanh^2\left(s_i^k\right)\right)$,

  $\widehat{\mathbf{m}}^k = \tanh(\mathbf{z}^k)$,

  $\mathbf{z}^{k+1} = \beta \mathbf{A}\widehat{\mathbf{m}}^k + \mathbf{r}^k - \frac{\beta^2}{n}\sum_{i=1}^n \left(1 - \tanh^2\left(z_i^k\right)\right)$

end

$\mathbf{u}^0 = \mathbf{z}^{K_{\text{AMP}}}$,

for $k = 0, ..., K_{\text{NGD}} - 1$ **do**

  $\mathbf{u}^{k+1} = \mathbf{u}^k - \eta \cdot \mathcal{F}_{\text{TAP}}(\widehat{\mathbf{m}}^{+,k}; \mathbf{y}, q)$,

  $\widehat{\mathbf{m}}^{+,k} = \tanh(\mathbf{u}^{k+1})$

end

return $\widehat{\mathbf{m}}^{+,K_{\text{NGD}}}$

$$\mathcal{F}_{\text{TAP}}(\mathbf{m}; \mathbf{y}, q) := -\frac{\beta}{2}\langle \mathbf{m}, \mathbf{A}\mathbf{m}\rangle - \langle \mathbf{y}, \mathbf{m}\rangle - \sum_{i=1}^n h(m_i) - \frac{n\beta^2(1-q)(1+q-2Q(\mathbf{m}))}{4}, \tag{3.5}$$

$$Q(\mathbf{m}) := \frac{\|\mathbf{m}\|_2^2}{n}, \qquad h(m) = -\frac{1+m}{2}\log\left(\frac{1+m}{2}\right) - \frac{1-m}{2}\log\left(\frac{1-m}{2}\right). \tag{3.6}$$

The second stage is exactly the same as in [AMS24], as no modifications are required here. The goal of this stage is to refine the output of the previous AMP stage, as it is still not exact. We may run NGD due to convexity of the $\mathcal{F}_{\text{TAP}}$ function around AMP iterates for all $\beta < 1$ [Cel22]. NGD finds the local minimum, which is known to be close to the optimal Bayes estimator of the planted model (see Section 4.1 and [CFM23; FMM20]).

We will not go into detail of the second stage, and will refer the interested reader to Section 2.1 of [AMS24] for a direct explanation, and to [FMM20], [CFM23], [Cel22] for more ties between Approximate Message Passing and the TAP free energy.

3.1.2 Sampling Algorithm via Anisotropic Gaussian Processes

We are now equipped to present the full sampling algorithm in Algorithm 2. Our new sampling algorithm makes use of two constants $q_k(\beta, t), \tau_k$ defined by the recursions:

$$q_{k+1} = \mathbb{E}\left[\tanh\left(\beta^2 q_k + t + \sqrt{\beta^2 q_k + t}W\right)^2\right], \quad q_0 = 0 \tag{3.7}$$

$$\tau_{k+1}^2 = 1 + \mathbb{E}\left[\left(\tanh(\tau_k^{-2} + \tau_k^{-1}W) - 1\right)^2\right], \quad \tau_0 = 1, \tag{3.8}$$

where $W \sim \mathcal{N}(0,1)$. Both may be efficiently computed with a one-dimensional integral. Mostly, we are interested in their asymptotic limit:

$$q_* := \lim_{n\to\infty} q_k, \tag{3.9}$$

$$\tau_* := \lim_{n\to\infty} \tau_k, \tag{3.10}$$

both of which are approached exponentially fast (see Lemma 4.1).

The sampling algorithm tracks the SDE driven by anisotropic noise (2.26) for $t \in [0, L_0]$, and then switches to the isotropic process (2.22) for $t \in (L_0, L]$. The SDE (2.26) is tracked by means of Euler discretization

$$\widehat{\mathbf{y}}_{\ell+1} = \widehat{\mathbf{y}}_\ell + \mathbf{\Omega}\delta\widehat{\mathbf{m}}(\mathbf{A}, \widehat{\mathbf{y}}_\ell) + \sqrt{\mathbf{\Omega}\delta}\mathbf{w}_{\ell+1}. \tag{3.11}$$

Finally, we output $\widehat{\mathbf{m}}(\mathbf{A}, \widehat{\mathbf{y}}_L)$ and round its coordinates to produce an approximate sample of $\mathbf{x}_* \sim \mu_{\mathbf{A},\beta_*}$.

Algorithm 2: Sampling from the SK Gibbs Measure via Anisotropic Diffusion Processes

Input: $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, $\mathbf{y} \in \mathbb{R}^n$, parameters $\beta, \beta', \eta, \delta, L_0, L > 0$, $q \in (0, 1)$, iteration numbers $K_{\mathrm{AMP}}, K_{\mathrm{NGD}}$

$\widehat{\mathbf{y}}(0) = 0$,
$\mathbf{\Omega} = 3\beta'\mathbf{I}_n - \beta'\mathbf{A}$
for $\ell = 0, ..., L_0 - 1$ **do**
  Draw $\mathbf{w}_{\ell+1} \sim \mathcal{N}(0, 1)$ independent of anything else so far
  Set $\beta_{\mathrm{eff}} := \beta - \ell\delta\beta'$
  Set $q = q_k(\beta, t = \ell\delta \cdot \tau_k^{-2}(\beta_{\mathrm{eff}}, t))$
  Set $\widehat{\mathbf{m}}(\mathbf{A}, \widehat{\mathbf{y}}_\ell)$ the output of Algorithm 1 with parameters $(\beta_{\mathrm{eff}}, \eta, q, K_{\mathrm{AMP}}, K_{\mathrm{NGD}})$
  Update $\widehat{\mathbf{y}}_{\ell+1} = \widehat{\mathbf{y}}_\ell + \mathbf{\Omega}\delta\widehat{\mathbf{m}}(\mathbf{A}, \widehat{\mathbf{y}}_\ell) + \sqrt{\mathbf{\Omega}\delta}\mathbf{w}_{\ell+1}$
end
Set $\beta_0 := \beta - L_0\delta\beta'$
for $\ell = L_0, ..., L - 1$ **do**
  Draw $\mathbf{w}_{\ell+1} \sim \mathcal{N}(0, \mathbf{I}_n)$ independent of anything else so far
  Set $q = q_k(\beta_0, t = \ell\delta)$
  Set $\widehat{\mathbf{m}}(\mathbf{A}, \widehat{\mathbf{y}}_\ell)$ the output of Algorithm 1 of [AMS24] with parameters $(\beta_0, \eta, q, K_{\mathrm{AMP}}, K_{\mathrm{NGD}})$
  Update $\widehat{\mathbf{y}}_{\ell+1} = \widehat{\mathbf{y}}_\ell + \delta\widehat{\mathbf{m}}(\mathbf{A}, \widehat{\mathbf{y}}_\ell) + \sqrt{\delta}\mathbf{w}_{\ell+1}$
end
Set $\widehat{\mathbf{m}}(\mathbf{A}, \widehat{\mathbf{y}}_L)$ the output of Algorithm 1 of [AMS24] with parameters $(\beta_0, \eta, q, K_{\mathrm{AMP}}, K_{\mathrm{NGD}})$
Draw $\{x_i^{\mathrm{alg}}\}_{i\le n}$ conditionally independent with $\mathbb{E}[x_i^{\mathrm{alg}} | \mathbf{y}, \{\mathbf{w}_\ell\} = \widehat{m}_i(\mathbf{A}, \widehat{\mathbf{y}}_L)]$
return $\mathbf{x}^{\mathrm{alg}}$

3.1.3 Equivalence to Isotropic Sampling for SK

Here we state that our anisotropic noise sampling algorithm efficiently samples from the SK Gibbs measure, just as the regular one.

Theorem 3.1 (Equivalence of Methods). *For any $\varepsilon > 0$ and $\beta_* < 1$, there exists $\eta, K_{\mathrm{AMP}}, K_{\mathrm{NGD}}, L, L_0, \delta$ independent of n such that for all $\beta < \beta_*$, the sampling algorithm 2 takes as input $\mathbf{A} \sim \mathrm{GOE}(n)$ and parameters $(\eta, K_{\mathrm{AMP}}, K_{\mathrm{NGD}}, L, L_0, \delta)$ and outputs a random point $\mathbf{x}_*$ with law $\tilde{\mu}_{\mathbf{A},\beta}$ such that with probability $1 - o_n(1)$ over $\mathbf{A} \sim \mathrm{GOE}(n)$,*

$$W_{2,n}(\tilde{\mu}_{\mathbf{A},\beta}, \mu_{\mathbf{A},\beta}) \le \varepsilon. \tag{3.12}$$

The complexity of this algorithm is $\mathcal{O}(n^2)$.

Remark 3.2. *This theorem is exactly the same statement as Theorem 2.1 of [AMS24], but states it for our Algorithm 2. In a sense, no new result on sampling is achieved for the SK model. This discussion will be treated in the next section.*

3.2 Enhanced Performance of Anisotropic Gaussian Sampling in p-spin Models

In this section we will discuss how our sampling algorithm offers solutions to 'barriers to diffusion models', proposed in [Ghi+23]. We will first recall how the inverse temperature parameter β affects the ability of the AMP to converge to the global maximum of the free energy, as well as other bottlenecks that arise from this. Finally, we explain how our method allows us to circumvent these issues.

3.2.1 Limitations of Sampling p-spin Gibbs Measures with Standard Diffusions

Spin glasses[2] have long been the object of interest in the study of phase transitions. A very interesting line of work is to observe when the spin glass will undergo a 'dynamical phase transition' for certain inverse temperatures β_{dyn}. Moreover, this β_{dyn} has been shown to depend on yet another parameter β_c, which breaks the replica symmetry of the system (see [Pan12] for an in-depth explanation of the topic). It is also the case that $\beta_{\text{dyn}} \leq \beta_c$. As explained in [AMS23], we have that

$$\beta_{\text{dyn}} \asymp \Theta\left(\sqrt{\frac{\log(p)}{p}}\right), \tag{3.13}$$

which will be the main interest of this section. Several restrictions on β were pointed out in [AMS23], of which one still persists and currently prohibits sampling for all $\beta < \beta_{\text{dyn}}$. The problematic inverse temperature $\beta_1 \asymp \frac{1}{\sqrt{p}} < \beta_{\text{dyn}}$ is the condition on the uniqueness of a fixed point q_* in Eq. (3.7) (with a slight adaptation for p-spin models), prohibiting the sampling algorithm of [AMS23] to work in $[\beta_1, \beta_{\text{dyn}})$. Interestingly, [Ghi+23] has also raised this issue, demonstrating that this threshold is a triple point of the RS entropy of the planted model (see Section 4.1), with additional Gaussian information. This splits the phase diagram into regions of a single local maximum, and metastable regions (phase coexistence with two maxima), shown in Figure 1 as the green and red/orange regions respectively.

3.2.2 AMP Trajectory & Phase Diagram

The 'trajectory' of AMP can be expressed as a straight line, as it computes the mean of the tilted measure (2.24) μ_t for a *fixed* value of β and increasing t, thus confirming that β_1 is the threshold on sampling from p-spin models with isotropic diffusion processes.

Remark 3.3 (The case of the SK model)**.** *In the SK model, we have that $\beta_1 = \beta_{\text{dyn}} = 1$, which alleviates the problem described above. This is also the reason why our method does not offer a wider sampling region in this particular case.*

[2] This discussion will solely be in the context of p-spin Ising models, but an extension to spherical models is also possible and rather straightforward. For more phase diagrams of different models, we refer to [Ghi+23].

[4] Available at `https://github.com/IdePHICS/DiffSamp`.

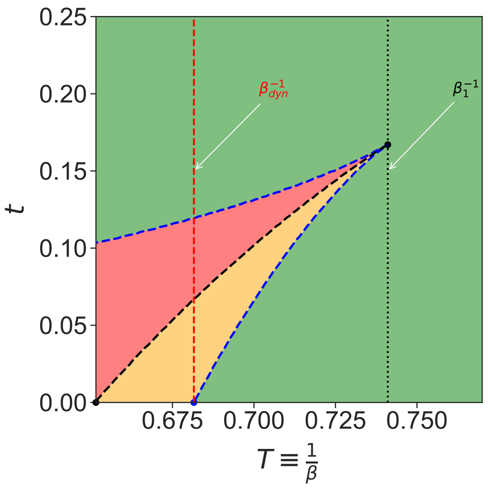

Figure 1: Phase diagram of the p-spin Ising model ($p = 3$) free entropy with temperature on the x-axis and SNR t on the y-axis made using code from [Ghi+23].[4] While the green region exhibits a single maximum, the orange and red have two, and the black dotted curve marks a swapping of local and global maxima (called IT threshold). Moreover, running AMP in the red region is strongly believed to be algorithmically hard [ZK16]. We can also see $\beta_1 \approx \frac{1}{0.741}$ and $\beta_{\text{dyn}} \approx \frac{1}{682}$.

3.2.3 AMP Trajectory with Anisotropic Gaussian Diffusions

Recalling Remark 3.1, we notice that using a covariance matrix $\mathbf{\Omega} = \lambda \mathbf{I}_n - \beta' \mathbf{A}$ for a certain β', the tilted measure now becomes

$$\begin{aligned}\mu_{\mathbf{y},t}(\mathrm{d}\mathbf{x}) &= \frac{1}{Z(\mathbf{A})} \mu_{\mathbf{A},\beta_*} e^{\langle \mathbf{y}_t, \mathbf{x} \rangle - \frac{t}{2} \langle \mathbf{x}, \mathbf{\Omega} \mathbf{x} \rangle} \\ &= \frac{1}{Z(\mathbf{A},\mathbf{y})} \exp\left(\frac{\beta_{\text{eff}}(t)}{2} \langle \mathbf{x}, \mathbf{A}\mathbf{x} \rangle + \langle \mathbf{y}_t, \mathbf{x} \rangle - \frac{t}{2} \|\lambda \mathbf{x}\|_2^2 \right) \end{aligned} \tag{3.14}$$

for $\beta_{\text{eff}}(t) := \beta - \beta' t$. We thus immediately see that with respect to Figure 1, our sampling algorithm begins at an initial inverse temperature β_* and the tilt then represents a different SK measure, with effective inverse temperature $\beta_{\text{eff}}(t)$. While we are technically still sampling from $\mu_{\mathbf{A},\beta_*}$, AMP will be computing means of $\mu_{\mathbf{A},\beta_{\text{eff}}(t)}$ with side information $\mathbf{y}_t$ as defined in (2.25). We note however that while $\mathbf{y}_t$ is a function of t, the CDMA-AMP procedure ($\mathbf{r}_k$ and $\mathbf{s}_k$ iterates of Algorithm 1, explained in Section 4.2.1) asymptotically converts it to an i.i.d. Rademacher vector with i.i.d. noise $\tau_*^{-2} t$

$$\mathbf{y}_t \xrightarrow{\text{CDMA-AMP}} \mathbf{x}_0 + \tau_*^{-2} t \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, \mathbf{I}_n). \tag{3.15}$$

This defines the *effective SNR* $\rho(t) := \tau_*^{-2} t$. Thus, the *effective parameters* of AMP are $(\beta_{\text{eff}}(t), \rho(t))$. We must now adapt Figure 1 to have $\rho(t)$ on the y-axis. The resulting trajectory of our AMP algorithm is a non-linear concave trajectory, which we can run until $\beta_{\text{eff}}(t) < \beta_1$, before switching back to regular isotropic AMP.

Remark 3.4. *Figure 2 is not from an actual implementation of the anisotropic AMP on the 3-spin Ising model. It is a plot of the trajectory for the SK model, on the same phase diagram as Figure 1. The depiction is made to highlight our hypothesis, and is not a rigorous result.*

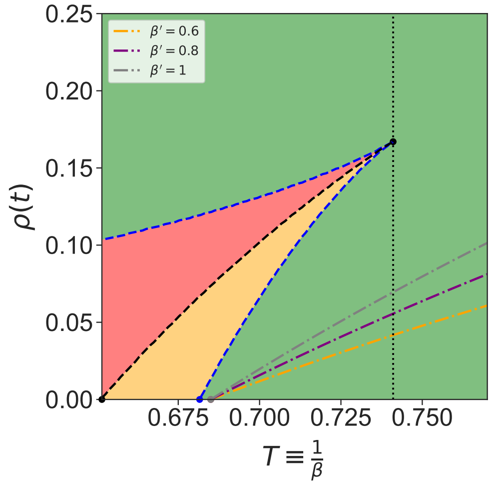

Figure 2: The trajectory of lines 3-8 of our new anisotropic Gaussian process-driven AMP algorithm (purple dashed line) for $\beta_* = 0.685$, $\beta' \in \{0.6, 0.8, 1\}$ and $\tau_* \approx 1.26038$. We can also see that for chosen values of β', β_{eff} remains in the computationally feasible region, allowing us to successfully sample from lower temperatures. Furthermore, we also see that for e.g. $\beta' = 0.8$, a choice of $L_0 \approx 0.06$ of Section 3.1.2 is a sensible point to switch to isotropic sampling.

4 Analysis of the Approximate Message Passing Algorithm

This section will focus on formally deriving Algorithm 1, and revisit AMP results from [AMS24] adapated to our case. We will also follow the approach of working with a *planted model*, under which all properties transfer to our SK model by contiguity (Section 6 [Vaa98]).

4.1 The Planted Model

Using similar arguments to [AMS24], we can perform our analysis of the algorithm on a simpler planted model. We describe two joint distributions of $(\mathbb{A}, \mathbf{y})$:

$$\mathbb{Q} : \begin{cases} \mathbf{A} \sim \mu_{\text{GOE}}, \\ \mathbf{y}(t) = \int_0^t \mathbf{Q}(s)\mathbf{m}(\mathbf{A}, \mathbf{y}(s))\mathrm{d}s + \int_0^t \mathbf{Q}(s)^{\frac{1}{2}}\mathrm{d}\mathbf{B}(s) \end{cases} \tag{4.1}$$

$$\mathbb{P} : \begin{cases} \mathbf{x}_0 \sim \bar{\nu}, \quad \bar{\nu} = \text{Unif}(\{+1, -1\}^n) \\ \mathbf{A} \sim \mathbf{x}_0\mathbf{x}_0^\intercal + \mathbf{W}, \quad \mathbf{W} \sim \mu_{\text{GOE}}, \\ \mathbf{y}(t) = t\mathbf{\Omega}\mathbf{x}_0 + \sqrt{\mathbf{\Omega}}\mathbf{B}(t) \end{cases} \tag{4.2}$$

where $\mathbb{Q}$ is the random model and $\mathbb{P}$ is the planted model.

Proposition 4.1. *For $\mathbb{P}$ and $\mathbb{Q}$ as given above, then they are mutually contiguous for $\beta < 1$.*

Proof. Notice that both $\mathbf{y}(t)$ processes are equivalent, as one is a restatement of the other. The claim thus follows by Proposition 4.2 [AMS24]. □

4.2 Hybrid Approximate Message Passing

Here we will analyse the AMP iteration of Algorithm 1. We begin the discussion with a few remarks.

Remark 4.1. *Let X be a Rademacher random variable, and $\mathbf{x}$ be a random vector with i.i.d. Rademacher entries. Then for any scalar $t \in \mathbb{R}$, the entries of $t\mathbf{x}$ are uniformly distributed in $\{-t, +t\}$. Thus,*

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{t\mathbf{x}_i} \longrightarrow_{n\to\infty} \mathcal{L}(Y), \tag{4.3}$$

where $Y \sim tX$.

Remark 4.2. *In the AMP procedures presented in [AMS23], [AMS24], [MW23], [DAM15] which make use of an extra observation process $\mathbf{y}$, this observation process was always distributed as*

$$Y_i \overset{\text{i.i.d.}}{\sim} t \cdot X_0 + \sqrt{t}Z \tag{4.4}$$

for $X_0 \sim \mathrm{Rad}$ and $Z \sim \mathcal{N}(0,1)$. Now consider a 'traditional' AMP iteration for $\mathbb{Z}_2$-synchronization [Fen+21; DAM15]:

$$\begin{aligned} \widehat{\mathbf{m}}^{-1} &= \mathbf{z}^0 = 0, \\ \widehat{\mathbf{m}}^{k} &= \eta_k(\mathbf{z}^k), \\ \mathbf{z}^{k+1} &= \beta \mathbf{A}\widehat{\mathbf{m}}^k - \frac{\beta^2}{n}\sum_{i=1}^{n} \eta_k^{'}(z_i^k)\widehat{\mathbf{m}}^{k-1} \end{aligned} \tag{4.5}$$

The crux of AMP, as described in [BM11], is that the intermediate iterates $\mathbf{z}^k$ are asymptotically distributed as $\gamma_k X_0 + \sqrt{\gamma_k}W$ for $W \sim \mathcal{N}(0,1)$. Thus, replacing the last line with an extra $\mathbf{y}$ will give us that each coordinate of the intermediate iterates are distributed as

$$\Bigg(\underbrace{\beta \mathbf{A}\widehat{\mathbf{m}}^k + \mathbf{y} - \frac{\beta^2}{n}\sum_{i=1}^{n} \eta_k^{'}(z_i^k)\widehat{\mathbf{m}}^{k-1}}_{:=\mathbf{z}^{k+1}}\Bigg)_i \overset{\text{i.i.d.}}{\sim} (\gamma_{k+1}+t)X_0 + \sqrt{\gamma_{k+1}+t}Z, \tag{4.6}$$

for all $i \in \{1, ..., n\}$ by independence of the driving Brownian motions.

Remark 4.3. *For the anisotropic process $\mathbf{y}$ given in (2.25), it does not follow in general that the elements of $\mathbf{y}$ are i.i.d. scaled Rademacher random variables with Gaussian noise. This is obvious from the matrix multiplication.*

The main novelty of our analysis lies in procedurally decorrelating the elements of $\mathbf{y}$ using a parallel instance of AMP for Code-Division Multiple Access (CDMA) (previously treated in [Mon11], [Jeo+15], [MT06]), and feeding the intermediate iterates to our $\mathbb{Z}_2$-synchronization with side information AMP iteration.

4.2.1 AMP for CDMA

For a given vector $\mathbf{x}_0 \in \mathbb{R}^n$ and sparsity matrix $\mathbf{H} \in \mathbb{R}^{n\times n}$, we can define the received vector as

$$\mathbf{y} = \mathbf{H}\mathbf{x_0} + \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n). \tag{4.7}$$

AMP has been shown to be the optimal MMSE receiver in the large system limit ([BM11], [Jeo+15]) under sub-gaussianity conditions on the sparsity matrix $\mathbf{H}$. We can write out the iterations for $\mathbf{x_0}$

a vector of iid Rademacher RVs:

$$
\begin{aligned}
\mathbf{s}^{k+1} &= \tanh\left(\mathbf{H}^{*}\mathbf{r}^{k} - s^{k}\right), \\
\mathbf{r}^{k+1} &= \mathbf{y} - \mathbf{H}\mathbf{s}^{k} + \frac{1}{n}\sum_{i=1}^{n}\left(1-\tanh^{2}\left(s_{i}^{k}\right)\right)\mathbf{s}^{k-1}
\end{aligned}
\tag{4.8}
$$

By the decoupling principle ([Mon11]), the correlation between the signals decay in the large system limit, and our state evolution is given by

$$
\tau_{k+1}^{2} = \sigma^{2} + \mathbb{E}\left[(\tanh(\tau_{k}^{-2} + \tau_{k}^{-1}Z) - 1)^{2}\right]. \tag{4.9}
$$

Again, the decoupling gives us that the intermediate iterates $(r_i^k)_{i=1}^n$ are asymptotically distributed as $\tau_k^{-2}X_0 + \tau_k^{-1}Z$. The existence and uniqueness of a fixed point of this iteration is proven in [Mon11] for a sub-gaussian signature matrix $\mathbf{H}$.

4.2.2 Combining CDMA with $\mathbb{Z}_2$-synchronization

Now using the fact that in the large system limit, state evolution for CDMA implies that

$$
\mathbf{y}_{i} \overset{\text{i.i.d.}}{\sim} X_{0} + \tau_{k}Z, \tag{4.10}
$$

we may substitute $\mathbf{y}$ with $\mathbf{r}^k$. The new parallel AMP algorithm thus reads:

$$
\begin{aligned}
\widehat{\mathbf{m}}^{-1} &= \mathbf{z}^{0} = 0, \\
\mathbf{s}^{k+1} &= \tanh\left(\mathbf{H}^{*}\mathbf{r}^{k} - s^{k}\right), \\
\mathbf{r}^{k+1} &= \mathbf{y} - \mathbf{H}\mathbf{s}^{k} + \frac{1}{n}\sum_{i=1}^{n}\left(1-\tanh^{2}\left(r_{i}^{k}\right)\right)\mathbf{s}^{k-1}, \\
\widehat{\mathbf{m}}^{k} &= \tanh(\mathbf{z}^{k}), \\
\mathbf{z}^{k+1} &= \beta\mathbf{A}\widehat{\mathbf{m}}^{k} + \mathbf{r}^{k} - \frac{\beta^{2}}{n}\sum_{i=1}^{n}\left(1-\tanh^{2}\left(z_{i}^{k}\right)\right)\widehat{\mathbf{m}}^{k-1}
\end{aligned}
\tag{4.11}
$$

for $\mathbf{H}t := \Omega(t)$. We are now ready to analyse Algorithm 1.

4.2.3 State Evolution

Using the general state evolution framework of [BM11][5], we can now write our hybrid state evolution

$$
\begin{aligned}
\gamma_{k}(\beta, t, \tau_{k}) &= \beta^{2}\cdot\mathbb{E}\left[\tanh(\gamma_{k}(\beta,t) + \tau_{k}^{-2}t + G_{k})\right], \\
\Sigma_{k+1,\ell+1} &= \beta^{2}\cdot\mathbb{E}\left[\tanh(\gamma_{k}(\beta,t) + \tau_{k}^{-2}t + G_{k})\tanh(\gamma_{\ell}(\beta,t) + \tau_{\ell}^{-2}t + G_{\ell})\right], \\
\tau_{k+1}^{2} &= 1 + \mathbb{E}\left[(\tanh(\tau_{k}^{-2} + \tau_{k}^{-1}Z) - 1)^{2}\right]
\end{aligned}
\tag{4.12}
$$

where $(G_1, ..., G_k)^\intercal := \mathbf{g}_k$ are jointly Gaussian with zero mean and covariance matrix $\boldsymbol{\Sigma}_{\leq \boldsymbol{k}} + t\boldsymbol{\tau}_{\leq \boldsymbol{k}}^{-1}\boldsymbol{\tau}_{\leq \boldsymbol{k}}^{-1\intercal}$, where the $^{-1}$ is applied element-wise and $\boldsymbol{\Sigma}_{\leq \boldsymbol{k}} := (\Sigma_{i,j})_{i,i\leq k}$. Moreover, the general behaviour of the iterates can be described as follows:

[5] A more thorough explanation for AMP & state evolution in low-rank matrix estimation is given in Appendix A

Proposition 4.2 (Theorem 1 [BM11]). *For $(\mathbf{A}, \mathbf{y}) \sim \mathbb{P}$ and any $k \in \mathbb{Z}_{\geq 0}$, the empirical distribution of the coordinate of the AMP iterates converges almost surely in $W_2(\mathbb{R}^{k+2})$ as follows:*

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{z_i^1,\ldots,z_i^k,x_i,r_i^k} \xrightarrow{W_2} \mathcal{L}(\boldsymbol{\gamma}_{\leq \boldsymbol{k}} X + \mathbf{g} + R_k \mathbf{1}, X, R_k), \tag{4.13}$$

$$\boldsymbol{\gamma}_{\leq k} \; = \; (\gamma_1, \gamma_2, ..., \gamma_k)^{\intercal}, \quad \mathbf{g} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\leq k}), \tag{4.14}$$

where $X \sim \mathrm{Rad}, R_k = \tau_k^{-2} X + \tau_k^{-1} W$ with $W \sim \mathcal{N}(0,1)$ and $X, \boldsymbol{g}, W$ are mutually independent.

4.3 Convergence Analysis of Hybrid AMP

Similary to Eq. (4.20) of [AMS24], we get that

$$\begin{aligned}
\gamma_{k+1} &= \beta^2 \cdot \mathbb{E}\left[\tanh(\underbrace{\gamma_k(\beta,t) + \tau_k^{-2} t}_{:=\alpha_k(\beta,t)} + \sqrt{\gamma_k(\beta,t) + \tau_k^{-2} t Z})\right] \\
&= \beta^2 \mathbb{E}[\tanh(\alpha_k(\beta,t) + \sqrt{\alpha_k(\beta,t)} Z)^2] \\
&= \beta^2 \cdot (1 - \mathrm{mmse}(\alpha_k(\beta,t)))\,.
\end{aligned} \tag{4.15}$$

Lemma 4.1 (Fixed Points). *Consider the first and third state evolution equations (4.12). Then it holds that for all $t \geq 0$ the following fixed points exist:*

$$\tau_*^2 := 1 + \mathbb{E}\left[\left(\tanh(\tau_*^{-2} + \tau_*^{-1} Z) - 1\right)^2\right], \tag{4.16}$$

$$\gamma_* \; = \; \beta^2 \cdot \left(1 - \mathrm{mmse}(\gamma_* + \frac{t}{\tau_*^2})\right). \tag{4.17}$$

Proof. The first equation is a direct consequence of Section 2.3 [BM11], [MT06] since it is independent of γ_k. We note however that τ_* does indeed depend on the dimensions of the sparsity (or design) matrix $\mathbf{H}$ as well as the variance of the initial added noise. Due to the symmetries of our case, it is simplified. The second follows from point (b) of Lemma 4.5 [AMS24]; a consequence of properties of the mmse$(\cdot)$ function, as the statement holds for any $t \geq 0$, we can let $\tilde{t} = \frac{t}{\tau_k^2}$ and conclude. $\square$

Remark 4.4. *Note that our state evolution γ_k is **not** the same as $\gamma_k(\beta, t)$ of [AMS24], although it serves the same purpose. We kept the same notation for ease of understanding, as the principles do not change.*

Remark 4.5. *Another alternative to (4.11) would be to simply run AMP to denoise $\mathbf{y}(t)$, and replace $\mathbf{r}^k$ with the resulting output of CDMA-AMP. This does not affect the fixed points of either iteration. Convergence to the fixed points are exponential in both cases [BM11].*

Lemma 4.2. *Let $\mathrm{MSE}_{\mathrm{AMP}}(k; \beta, t) = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\|\mathbf{x} - \widehat{\mathbf{m}}^k(\mathbf{A}, \mathbf{y})\|_2^2]$, for $\widehat{\mathbf{m}}^k$ the k-th iterate of our hybrid AMP algorithm. Then, we have*

$$\mathrm{MSE}_{\mathrm{AMP}}(k; \beta, t) = 1 - \frac{\gamma_{k+1}(\beta, t)}{\beta^2} \tag{4.18}$$

Proof. By state evolution, we get that

$$\begin{aligned}
\mathrm{MSE}_{\mathrm{AMP}}(k;\beta,t) &= \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\|\mathbf{x}-\widehat{\mathbf{m}}^k(\mathbf{A},\mathbf{y})\|_2^2] \\
&= \mathbb{E}[(\tanh(\alpha_k X+\sqrt{\alpha_k}W)-X)^2] \\
&= 1-2\mathbb{E}[\tanh(\alpha_k X+\sqrt{\alpha_k}W)]+\mathbb{E}[\tanh(\alpha_k X+\sqrt{\alpha_k}W)^2] \\
&= 1-2\frac{\gamma_{k+1}}{\beta^2}+\frac{\gamma_{k+1}}{\beta^2} \\
&= 1-\frac{\gamma_{k+1}}{\beta^2} \qquad (4.19)
\end{aligned}$$

□

Proposition 4.3 (Proposition 4.7 [AMS24])**.** *Fix $\beta > 0$ and $t \geq 0$, we have*

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\|\mathbf{x}-\mathbf{m}(\mathbf{A},\mathbf{y}(t)\|_2^2\right] = \frac{\gamma_*}{\beta^2} \tag{4.20}$$

Proof. See proof of Proposition 4.7 [AMS24], plugging in our γ in place of theirs. □

We can finally show that our hybrid AMP successfully approximates the posterior mean $\mathbf{m}(\mathbf{A},\mathbf{y}(t))$ in the same sense as with the isotropic observation process.

Proposition 4.4. *Fix $\beta < 1$, $T > 0$, and let $t \in [0,T]$. Then for $\widehat{\mathbf{m}}^k(\mathbf{A},\mathbf{y}(t))$ the output of hybrid AMP after k iterations, and $\mathbf{z}^k$ given by Equation (4.11), we have*

$$\lim_{k\to\infty}\sup_{t\in(0,T)}\operatorname*{p-lim}_{n\to\infty}\frac{\|\mathbf{m}(\mathbf{A},\mathbf{y}(t))-\widehat{\mathbf{m}}^k(\mathbf{A},\mathbf{y}(t))\|_2}{\|\mathbf{m}(\mathbf{A},\mathbf{y}(t))\|_2} = 0, \tag{4.21}$$

$$\lim_{k\to\infty}\sup_{t\in(0,T)}\operatorname*{p-lim}_{n\to\infty}\frac{\|\mathbf{z}^k-\mathbf{z}^{k+1}\|_2}{\|\mathbf{z}\|_2} = 0 \tag{4.22}$$

Proof. We begin by using the bias-variance decomposition of $\mathrm{MSE}_{\mathrm{AMP}}(k;\beta,t)$,

$$\mathrm{MSE}_{\mathrm{AMP}}(k;\beta,t) = \operatorname*{p-lim}_{n\to\infty}\left\{\frac{1}{n}\left[\mathbb{E}\left[\|\mathbf{m}(\mathbf{A},\mathbf{y}(t))-\widehat{\mathbf{m}}^k(\mathbf{A},\mathbf{y}(t))\|_2^2\right]+\mathbb{E}\left[\|\mathbf{m}(\mathbf{A},\mathbf{y}(t))-\mathbf{x}\|_2^2\right]\right]\right\} \tag{4.23}$$

Applying Lemma 4.2 to the first term, and Proposition 4.3, we get

$$\mathrm{MSE}_{\mathrm{AMP}}(k;\beta,t) = \frac{\gamma_k}{\beta^2}, \tag{4.24}$$

and the claim follows from Lemma 4.5 of [AMS24]. □

Lemma 4.3 (Regularity in t)**.** *For fixed $\beta < 1$, $0 \leq t_1 < t_2 \leq T$, we have*

$$\begin{aligned}
\operatorname*{p-lim}_{n\to\infty}\sup_{t\in[t_1,t_2]}\frac{1}{n}\mathbb{E}\left[\|\mathbf{m}(\mathbf{A},\mathbf{y}(t))-\mathbf{m}(\mathbf{A},\mathbf{y}(t_1))\|\right] &= \operatorname*{p-lim}_{n\to\infty}\frac{1}{n}\mathbb{E}\left[\|\mathbf{m}(\mathbf{A},\mathbf{y}(t_2))-\mathbf{m}(\mathbf{A},\mathbf{y}(t_1))\|\right] \\
&= \frac{\gamma_*(\beta,t_2)-\gamma_*(\beta,t_1)}{\beta^2} \qquad (4.25)
\end{aligned}$$

Proof. Similarly to Proposition 4.3, we may reuse the original proof of [AMS24] (Propositiom 4.9), due to the simple rescaling $\tilde{t} = t\tau_*^{-2}$, and apply our results to $\tilde{t}_1, \tilde{t}_2$. □

Finally, we will conclude on an existing result extending Lemma 4.4. Recalling point 1) of Lemma 4.10 of [AMS24], we get that for any fixed $\beta < \frac{1}{2}$ and $T > 0$, there exists $\varepsilon_0 = \varepsilon(\beta, T)$ such that for all $\epsilon \in (0, \varepsilon_0)$ there exists $K_{\text{AMP}} =_{\text{AMP}} (\beta, T, \epsilon)$ such that the following holds. Denoting $\widehat{\mathbf{m}}^{K_{\text{AMP}}}$ the K_{AMP}-th iterate of their AMP procedure, we get that with probability $1 - o_n(1)$ under the planted model, for all $t \in (0, T]$ and all $\widehat{\mathbf{y}} \in \mathcal{B}(\mathbf{y}(t), c\sqrt{\varepsilon t n}/4)$, setting $q_* := q_*(\beta, t)$, $\widehat{\mathbf{m}}^{K_{\text{AMP}}}$ is in a $\sqrt{\varepsilon t n}/2$-neighbourhood of a local minimum of $\mathcal{F}_{\text{TAP}}(\mathbf{m}; \widehat{y}, q_*)$. Again, this result is easily extendable to our case, as the hybrid AMP iterates (4.11) reduce to the same problem, with scaled SNR $t \mapsto \tau_*^{-2} t$. Also, we note that due to Celentano's results in [Cel22], we may safely extend the results to $\beta > 1$. We will write out this result for convenience.

Lemma 4.4. *Fix $\beta < 1$ and $T > 0$, there exists $\varepsilon_0 = \varepsilon(\beta, T)$ such that for all $\epsilon \in (0, \varepsilon_0)$ there exists $K_{\text{AMP}} =_{\text{AMP}} (\beta, T, \epsilon)$ such that the following holds. Denoting $\widehat{\mathbf{m}}^{K_{\text{AMP}}}$ the K_{AMP}-th iterate of (4.11), we get that with probability $1 - o_n(1)$ under the planted model, for all $t \in (0, T]$ and all $\widehat{\mathbf{y}} \in \mathcal{B}(\mathbf{y}(t), c\sqrt{\varepsilon t n}/4)$, setting $t_* = \tau_*^{-2} t$ and $q_* := q_*(\beta, t_*)$, the function*

$$\mathbf{m} \mapsto \mathcal{F}_{\text{TAP}}(\mathbf{m}; \widehat{\mathbf{y}}, q_*) \tag{4.26}$$

restricted to $\mathcal{B}\left(\widehat{\mathbf{m}}^{K_{\text{AMP}}}, \sqrt{\varepsilon n t_*}/2\right) \cap (-1, 1)^n$ has a unique stationary point

$$\mathbf{m}_*(\mathbf{A}, \widehat{\mathbf{y}}) \in \mathcal{B}\left(\widehat{\mathbf{m}}^{K_{\text{AMP}}}, \sqrt{\varepsilon n t_*}/2\right) \cap (-1, 1)^n \tag{4.27}$$

which is also a local minimum.

The proof of this result is omitted, as it can be found in Appendix A of [AMS24] by considering $t \mapsto t_*$, and that the uniqueness of the local minimum is now due to [Cel22].

Remark 4.6. *Notice that our state evolution from (4.12) $\gamma_k(\beta, t, \tau_k) = \gamma_k(\beta, \frac{t}{\tau_k^2})$ of [AMS24]. It essentially matches, up to a scaled time factor. In practice, this would require running AMP for more iterations. From this observation, we may carry over all results on Natural Gradient Descent (Section 4.3 [AMS24]).*

4.4 Continuity and Proof of Theorem 3.1

Here we will present adaptations of arguments in Section 4.4 [AMS24] to prove Theorem 3.1. We will do so by showing a 'crude' Lipschitz continuity of the maps $\mathbf{y} \mapsto \mathbf{s}(\mathbf{A}, \mathbf{y})$ and $\mathbf{y} \mapsto \widehat{\mathbf{m}}(\mathbf{A}, \mathbf{y})$.

Lemma 4.5 (Lipschitz continuity for CDMA-AMP)**.** *Let $\mathbf{H} \in \mathbb{R}^{n \times n}$ be a square signature matrix in the CDMA setup of Section (4.2.1). Moreover, we will consider that its operator bound satisfies $\|\mathbf{H}\|_{\text{op}} \leq C < \infty$.*[6] *Let* CDMA-AMP$(\mathbf{A}, \mathbf{y}; k)$ *denote the k-th iteration of the CDMA AMP procedure. Then,*

$$\|\text{CDMA-AMP}(\mathbf{H}, \mathbf{y}; k) - \text{CDMA-AMP}(\mathbf{H}, \widehat{\mathbf{y}}; k)| \leq j(C+1)^j \|\mathbf{y} - \widehat{\mathbf{y}}\|. \tag{4.28}$$

Proof. We set $0 \leq j \leq k$, we denote

$$\begin{aligned}
\mathbf{s}^j &= \text{CDMA-AMP}(\mathbf{H}, \mathbf{y}; j), \\
\mathbf{r}^j &= \tanh^{-1}(\widehat{\mathbf{m}}^j), \\
d_j &= \frac{1}{n} \sum_{i=1}^{n} \left(1 - \tanh^2(s_i^k)\right),
\end{aligned} \tag{4.29}$$

[6] In the specific case of $\mathbf{\Omega} = 2\beta' \mathbf{I}_n + \beta' \mathbf{A}$, we could choose $C = 5\beta'$

and

$$
\begin{aligned}
\widehat{\mathbf{s}}^j &= \text{CDMA-AMP}(\mathbf{H}, \widehat{\mathbf{y}}; j), \\
\widehat{\mathbf{r}}^j &= \tanh^{-1}(\widehat{\mathbf{m}}^j), \\
\widehat{d}_j &= \frac{1}{n}\sum_{i=1}^{n}\left(1 - \tanh^2(\widehat{s}_i^k)\right).
\end{aligned} \tag{4.30}
$$

Now we write

$$
\begin{aligned}
\|\mathbf{r}^{j+1} - \widehat{\mathbf{r}}^{j+1}\| &\le \|\mathbf{H}(\mathbf{s}^j - \widehat{\mathbf{s}}^j)\| + \|\mathbf{y} - \widehat{\mathbf{y}}\| + \|d_j(\mathbf{s}^{j-1} - \widehat{\mathbf{s}}^{j-1})\| + \|(d_j + \widehat{d}_j)\widehat{\mathbf{s}}^{j-1}\| \\
&\le C\|\mathbf{r}^j - \widehat{\mathbf{r}}^j\| + \|\mathbf{y} - \widehat{\mathbf{y}}\| + d_j\|\mathbf{r}^{j-1} - \widehat{\mathbf{r}}^{j-1}\| + |d_j - \widehat{d}_j|\sqrt{n}.
\end{aligned} \tag{4.31}
$$

Let $D_{j+1} := \max_{i \le j} \|\mathbf{r}^j - \widehat{\mathbf{r}}^j\|$, and we get

$$
D_{j+1} \le (C+1)D_j + \|\mathbf{y} - \widehat{\mathbf{y}}\|, \tag{4.32}
$$

by the assumption that $C > 1$, which by induction yields

$$
D_j \le j(C+1)^j\|\mathbf{y} - \widehat{\mathbf{y}}\| \tag{4.33}
$$

$\square$

Lemma 4.6. *If $\|\mathbf{A}\|_{\text{op}} \le 3$, then for any $\mathbf{y}, \widehat{\mathbf{y}} \in \mathbb{R}^n$, we have*

$$
\|\tanh^{-1}(\text{AMP}(\mathbf{A}, \mathbf{y}; k)) - \tanh^{-1}(\text{AMP}(\mathbf{A}, \widehat{\mathbf{y}}; k))\| \le j\tilde{C}^j\|\mathbf{y} - \widehat{\mathbf{y}}\|, \tag{4.34}
$$

where $\text{AMP}(\mathbf{A}, \mathbf{y}; k)$ denotes the k-th iteration of the iteration (4.11) with inputs $(\mathbf{A}, \mathbf{y})$, and some constant $C' > 1$.

Proof. We set $0 \le j \le k$ and write out

$$
\begin{aligned}
\mathbf{m}^j &= \text{AMP}(\mathbf{A}, \mathbf{y}; j), \\
\mathbf{z}^j &= \tanh^{-1}(\mathbf{m}^j), \\
b_j &= \frac{\beta^2}{n}\sum_{i=1}^{n}\left(1 - \tanh^2(z_i^k)\right),
\end{aligned} \tag{4.35}
$$

and

$$
\begin{aligned}
\widehat{\mathbf{m}}^j &= \text{AMP}(\mathbf{A}, \widehat{\mathbf{y}}; j), \\
\widehat{\mathbf{z}}^j &= \tanh^{-1}\widehat{\mathbf{m}}^j, \\
\widehat{b}_j &= \frac{\beta^2}{n}\sum_{i=1}^{n}\left(1 - \tanh^2(\widehat{z}_i^k)\right).
\end{aligned} \tag{4.36}
$$

Also, we will use all definitions mentioned in the proof of Lemma 4.5. We can then write out

$$
\begin{aligned}
\|\mathbf{z}^{j+1} - \widehat{\mathbf{z}}^{j+1}\| &\le \|\beta\mathbf{A}(\mathbf{m}^j - \widehat{\mathbf{m}}^j)\| + \|\mathbf{r}^j - \widehat{\mathbf{r}}^j\| \\
&\qquad + b_j\mathbf{m}^{j-1} - b_j\widehat{\mathbf{m}}^{j-1}\| + \|b_j\widehat{\mathbf{m}}^{j-1} - \widehat{b}_j\widehat{\mathbf{m}}^{j-1}\| \\
&\le 3\beta\|\mathbf{z}^j - \widehat{\mathbf{z}}^j\| + \|\mathbf{r}^j - \widehat{\mathbf{r}}^j\| + d_j\|\mathbf{m}^j - \widehat{\mathbf{m}}^j\| + |d_j - \widehat{d}_j|\sqrt{n}
\end{aligned} \tag{4.37}
$$

We now define $E_j := \max_{i\le j} \|\mathbf{z}^{i+1} - \widehat{\mathbf{z}}^{i+1}\|$, and by the fact that $|1-\tanh(x)| \le 1$ for all $x \in \mathbb{R}, \beta \le 1$ and $\tanh(\cdot)$ is 1-Lipschitz, we get

$$\begin{aligned} E_{j+1} &\le (3\beta^2 + 3\beta)E_j + \|\mathbf{r}^j - \widehat{\mathbf{r}}^j\| \\ &\le 6E_j + \|\mathbf{r}^j - \widehat{\mathbf{r}}^j\|. \end{aligned} \tag{4.38}$$

Now, let $R_j := \max_{i\le j}\{E_j, D_j\}$ with D_j as defined above, we notice that

$$\begin{aligned} R_{j+1} &\le 6E_j + (C+1)D_j + \|\mathbf{y} - \widehat{\mathbf{y}}\| \\ &\le 6(C+1)R_j + \|\mathbf{y} - \widehat{\mathbf{y}}\|, \end{aligned} \tag{4.39}$$

where C is the upper-bounding constant used in the previous lemma. Induction yields

$$R_{j+1} \le j\,(6(C+1))^j\,\|\mathbf{y} - \widehat{\mathbf{y}}\|. \tag{4.40}$$

Setting $j = k$ and $\tilde{C} = 6(C+1)$ completes the proof. $\square$

We now fix (β, T) and choose $K_{\mathrm{AMP}} = K_{\mathrm{AMP}}(\beta, T, \varepsilon)$, $\rho_0 = \rho_0(\beta, T, \varepsilon, K_{\mathrm{AMP}})$, $\rho \in (0, \rho_0)$ an K_{NGD} such that Lemma 4.10 [AMS24] holds. If we consider the coupled discretized process $(\widehat{\mathbf{y}}_\ell)_{\ell\ge 0}$ of (3.11) to $(\mathbf{y}(t))_{t\ge 0}$ via the driving noise (setting $\mathbf{w}_{\ell+1} := \frac{1}{\sqrt{\delta}}\int_{\ell\delta}^{(\ell+1)\delta} \mathrm{d}\mathbf{B}(t)$), we can derive bounds on the random approximation errors

$$A_\ell := \frac{1}{\sqrt{n}}\|\widehat{\mathbf{y}}_\ell - \mathbf{y}(\ell\delta)\|, \tag{4.41}$$

$$B_\ell := \frac{1}{\sqrt{n}}\|\widehat{\mathbf{m}}(\mathbf{A}, \widehat{\mathbf{y}}_\ell) - \mathbf{m}(\mathbf{A}, \mathbf{y}(\ell\delta))\|, \tag{4.42}$$

by appealing to Lemma 4.14 [AMS24] (which holds by previous lemmas). We get

$$A_\ell \le \Lambda e^{\Lambda\ell\delta}\ell\delta(\rho\sqrt{\ell\delta} + \sqrt{\delta}) + \xi(n), \tag{4.43}$$

$$B_\ell \le \Lambda e^{\Lambda\ell\delta}\ell\delta(\rho\sqrt{\ell\delta} + \sqrt{\delta}) + \Lambda\rho\sqrt{\ell\delta} + \xi(n), \tag{4.44}$$

where $\Lambda \equiv \Lambda(\beta) < \infty$ is a constant, and $\xi(n)$ is a non-negative deterministic sequence such that $\lim_{n\to\infty} \xi(n) = 0$. We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1. We first present a proof outline.

1. Recall that our hybrid AMP lands in a convex $\sqrt{\epsilon n t_*}/2$-neighbourhood of $\mathcal{F}_{\mathrm{TAP}}$ after a sufficiently large (but finite) number of iterations K_{AMP} (Lemma 4.4).

2. We then run NGD as in [AMS24], initialized at this K_{AMP}-th iterate. From Lemma 4.10 [AMS24], we know that it successfully approximately the posterior mean given parameters $\mathbf{A}, \widehat{\mathbf{y}}_\ell$.

3. We now consider the discretized process $\widehat{\mathbf{y}}_\ell$, and show that $\widehat{\mathbf{m}}(\mathbf{A}, \widehat{\mathbf{y}}_\ell)$ is close to $\mathbf{m}(\mathbf{A}, \mathbf{y}(\ell\delta))$. We set $\ell = L = \frac{T}{\delta}$ and $\rho = \sqrt{\delta}$ in (4.44). We can now write

$$\begin{aligned} \mathbb{E}\left[W_{2,n}(\mu_{\mathbf{A}}, \mathcal{L}(\widehat{\mathbf{m}}(\mathbf{A}, \widehat{\mathbf{y}}_L)))\right] &\le \mathbb{E}\left[W_{2,n}(\mu_{\mathbf{A}}, \mathcal{L}(\mathbf{m}(\mathbf{A}, \mathbf{y}(T))))\right] \\ &\quad + \mathbb{E}\left[W_{2,n}(\mathcal{L}(\mathbf{m}(\mathbf{A}, \mathbf{y}(T))), \mathcal{L}(\widehat{\mathbf{m}}(\mathbf{A}, \widehat{\mathbf{y}}_L)))\right] \\ &\le T^{-\frac{1}{2}} + \Lambda(\beta, T)\sqrt{\delta} + o_n(1). \end{aligned} \tag{4.45}$$

For sufficiently large T, small enough δ and in the large limit system (large n), we get

$$\mathbb{E}\left[W_{2,n}(\mu_{\mathbf{A}}, \mathcal{L}(\widehat{\mathbf{m}}_{\text{NGD}}(\mathbf{A}, \widehat{\mathbf{y}}_L)))\right] \leq \frac{\varepsilon^2}{4} \tag{4.46}$$

for any $\varepsilon > 0$. The first term of (4.45) arises from Lemma 3.3 [AMS24], and the second from (4.44).

4. We now consider the produced samples, which are roundings of the produced estimators.

   For convenience, we will recall Lemma 4.15 of [AMS24] below.

   Lemma 4.7. *Suppose probability distributions μ_1, μ_2 on $[-1,1]^n$ are given. Sample $\mathbf{m}_1 \sim \mu_1$ and $\mathbf{m}_2 \sim \mu_2$ and let $\mathbf{x}_1, \mathbf{x}_2 \in \{-1,+1\}^n$ be standard randomized roundings of $\mathbf{m}_1$ resp. $\mathbf{m}_2$. Then,*

$$W_{2,n}(\mathcal{L}(\mathbf{x}_1), \mathcal{L}(\mathbf{x}_2)) \leq 2\sqrt{W_{2,n}(\mu_1, \mu_2)} \tag{4.47}$$

   We can apply this lemma to (4.46) to obtain

$$\mathbb{E}\left[W_{2,n}(\mu_{\mathbf{A}}, \mathcal{L}(\widehat{\mathbf{m}}_{\text{NGD}}(\mathbf{A}, \widehat{\mathbf{y}}_L)))\right] \leq \epsilon. \tag{4.48}$$

   Markov's inequality thus proves Theorem 3.1.

□

5 Conclusions and Further Outcomes

We studied the problem of sampling from the Sherrington-Kirkpatrick Gibbs measure using anisotropic Gaussian diffusion processes. The resulting sampling algorithm extended the usage of Approximate Message Passing to compute the tilted mean (2.23) by simultaneously denoising the anisotropic observation process. By the decoupling principle, our observation reduces to i.i.d. Gaussian observations of a Rademacher random variable, resulting in a similar state evolution of the posterior mean AMP. This algorithm is used as a subroutine to sample from an inverse temperature β_* which differs from the effective inverse temperature β_{eff} used in AMP, allowing us to create a non-linear algorithm trajectory in phase space. An extension of this algorithm to the p-spin model seems to be feasible, taking inspiration from [AMS23], allowing us to sample beyond the threshold of $\beta < \beta_1$, possibly pushing to $\beta < \beta_{\text{dyn}}$. Another interesting problem would be to generalize the local TAP convexity presented in [Cel22], and extend it to the p-spin model. A final related but more general problem is that of the necessity of running a Natural Gradient Descent phase after AMP. It is conjectured in [CFM23] that AMP converges to the global minimum of the TAP free energy, supported by numerical experiments, but not proven. While this would not reduce the complexity of the sampling algorithm by any polynomial factor, it is nevertheless of interest for understanding.

Acknowledgements

This project was completed in the context of a Semester Project for the Master's Degree in the Department of Electrical Engineering and Information Technology at ETH Zürich. I would like to thoroughly thank Professor Yuansi Chen, for providing kind and insightful supervision of the project, as well as Professor Hans-Andrea Loeliger for allowing me to carry out this project in the Seminar for Statistics, Department of Mathematics.

References

[Gla63] Roy J. Glauber. "Time Dependent Statistics of the Ising Model". In: *Journal of Mathematical Physics* 4.2 (1963), pp. 294–307.

[SK75] David Sherrington and Scott Kirkpatrick. "Solvable Model of a Spin-Glass". In: *Physical Review Letters* 35 (Dec. 1975), pp. 1792+. DOI: 10.1103/PhysRevLett.35.1792.

[TAP77] Thouless, Anderson, and Palmer. "Solution of 'Solvable model of a spin glass'". In: *The Philosophical Magazine: A Journal of Theoretical Experimental and Applied Physics* 35.3 (1977), pp. 593–601. DOI: 10.1080/14786437708235992. eprint: https://doi.org/10.1080/14786437708235992. URL: https://doi.org/10.1080/14786437708235992.

[AT78] J R L de Almeida and D J Thouless. "Stability of the Sherrington–Kirkpatrick solution of a spin glass model". In: *Spin Glass Theory and Beyond*. 1978, pp. 129–136. DOI: 10.1142/9789812799371_0012. eprint: https://www.worldscientific.com/doi/pdf/10.1142/9789812799371_0012. URL: https://www.worldscientific.com/doi/abs/10.1142/9789812799371_0012.

[And82] Brian D.O. Anderson. "Reverse-time diffusion equation models". In: *Stochastic Processes and their Applications* 12.3 (1982), pp. 313–326. ISSN: 0304-4149. DOI: https://doi.org/10.1016/0304-4149(82)90051-5. URL: https://www.sciencedirect.com/science/article/pii/0304414982900515.

[AH87] M. Aizenman and R. Holley. "Rapid Convergence to Equilibrium of Stochastic Ising Models in the Dobrushin Shlosman Regime". In: *Percolation Theory and Ergodic Theory of Infinite Particle Systems*. Ed. by Harry Kesten. New York, NY: Springer New York, 1987, pp. 1–11. ISBN: 978-1-4613-8734-3. DOI: 10.1007/978-1-4613-8734-3_1. URL: https://doi.org/10.1007/978-1-4613-8734-3_1.

[MPV87] M. Mezard, G. Parisi, and M.A. Virasoro. *Spin Glass Theory And Beyond: An Introduction To The Replica Method And Its Applications*. World Scientific Lecture Notes In Physics. World Scientific Publishing Company, 1987. ISBN: 9789813103917. URL: https://books.google.ch/books?id=DwY8DQAAQBAJ.

[Vaa98] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[Kab03] Yoshiyuki Kabashima. "A CDMA multiuser detection algorithm on the basis of belief propagation". In: *Journal of Physics A: Mathematical and General* 36 (Oct. 2003), p. 11111. DOI: 10.1088/0305-4470/36/43/030.

[MT06] Andrea Montanari and David Tse. *Analysis of Belief Propagation for Non-Linear Problems: The Example of CDMA (or: How to Prove Tanaka's Formula)*. 2006. arXiv: cs/0602028 [cs.IT].

[RU08] Tom Richardson and Rüdiger Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008.

[DMM09] David L. Donoho, Arian Maleki, and Andrea Montanari. "Message-passing algorithms for compressed sensing". In: *Proceedings of the National Academy of Sciences* 106.45 (Nov. 2009), pp. 18914–18919. ISSN: 1091-6490. DOI: 10.1073/pnas.0909892106. URL: http://dx.doi.org/10.1073/pnas.0909892106.

[BM11] Mohsen Bayati and Andrea Montanari. "The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing". In: *IEEE Transactions on Information Theory* 57.2 (Feb. 2011), pp. 764–785. ISSN: 1557-9654. DOI: 10.1109/tit.2010.2094817. URL: http://dx.doi.org/10.1109/TIT.2010.2094817.

[Efr11] Bradley Efron. "Tweedie's Formula and Selection Bias". In: *Journal of the American Statistical Association* 106.496 (2011), pp. 1602–1614. ISSN: 01621459. URL: http://www.jstor.org/stable/23239562 (visited on 06/27/2024).

[Mon11] Andrea Montanari. *Graphical Models Concepts in Compressed Sensing*. 2011. arXiv: 1011.4328 [cs.IT].

[Tal11] Michel Talagrand. *Mean field models for spin glasses. Volume I: Basic examples. 2nd revised and enlarged ed.* Jan. 2011. ISBN: 978-3-642-15201-6. DOI: 10.1007/978-3-642-15202-3.

[Bol12] Erwin Bolthausen. *An iterative construction of solutions of the TAP equations for the Sherrington-Kirkpatrick model*. 2012. arXiv: 1201.2891 [math.PR].

[Pan12] Dmitry Panchenko. *The Sherrington-Kirkpatrick Model*. Nov. 2012. ISBN: 978-1-4614-6288-0. DOI: 10.1007/978-1-4614-6289-7.

[DJM13] David L. Donoho, Adel Javanmard, and Andrea Montanari. *Information-Theoretically Optimal Compressed Sensing via Spatial Coupling and Approximate Message Passing*. 2013. arXiv: 1112.0708 [cs.IT]. URL: https://arxiv.org/abs/1112.0708.

[Eld13] Ronen Eldan. "Thin Shell Implies Spectral Gap Up to Polylog via a Stochastic Localization Scheme". In: *Geometric and Functional Analysis* 23.2 (Mar. 2013), pp. 532–569. ISSN: 1420-8970. DOI: 10.1007/s00039-013-0214-y. URL: http://dx.doi.org/10.1007/s00039-013-0214-y.

[MT13] Ryosuke Matsushita and Toshiyuki Tanaka. "Low-rank matrix reconstruction and clustering via approximate message passing". In: *Advances in Neural Information Processing Systems*. Ed. by C.J. Burges et al. Vol. 26. Curran Associates, Inc., 2013.

[DAM15] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. *Asymptotic Mutual Information for the Two-Groups Stochastic Block Model*. 2015. arXiv: 1507.08685 [cs.IT].

[Jeo+15] Charles Jeon et al. *Optimality of Large MIMO Detection via Approximate Message Passing*. 2015. arXiv: 1510.06095 [cs.IT].

[Soh+15] Jascha Sohl-Dickstein et al. *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*. 2015. arXiv: 1503.03585 [cs.LG]. URL: https://arxiv.org/abs/1503.03585.

[ZK16] Lenka Zdeborová and Florent Krzakala. "Statistical physics of inference: thresholds and algorithms". In: *Advances in Physics* 65.5 (Aug. 2016), pp. 453–552. ISSN: 1460-6976. DOI: 10.1080/00018732.2016.1211393. URL: http://dx.doi.org/10.1080/00018732.2016.1211393.

[LKZ17] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. "Constrained low-rank matrix estimation: phase transitions, approximate message passing and applications". In: *Journal of Statistical Mechanics: Theory and Experiment* 2017.7 (July 2017), p. 073403. ISSN: 1742-5468. DOI: 10.1088/1742-5468/aa7284. URL: http://dx.doi.org/10.1088/1742-5468/aa7284.

[Les+17] Thibault Lesieur et al. "Statistical and computational phase transitions in spiked tensor estimation". In: *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, June 2017. DOI: 10.1109/isit.2017.8006580. URL: http://dx.doi.org/10.1109/ISIT.2017.8006580.

[Per+18] Amelia Perry et al. "Message-Passing Algorithms for Synchronization Problems over Compact Groups". In: *Communications on Pure and Applied Mathematics* 71 (Apr. 2018), pp. 2275–2322. ISSN: 1097-0312. DOI: 10.1002/cpa.21750. URL: http://dx.doi.org/10.1002/cpa.21750.

[Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

[BB19] Roland Bauerschmidt and Thierry Bodineau. "A very simple proof of the LSI for high temperature spin systems". In: *Journal of Functional Analysis* 276.8 (Apr. 2019), pp. 2582–2588. ISSN: 0022-1236. DOI: 10.1016/j.jfa.2019.01.007. URL: http://dx.doi.org/10.1016/j.jfa.2019.01.007.

[Eld19] Ronen Eldan. *Taming correlations through entropy-efficient measure decompositions with applications to mean-field approximation*. 2019. arXiv: 1811.11530 [math.PR]. URL: https://arxiv.org/abs/1811.11530.

[MV19] Andrea Montanari and Ramji Venkataramanan. *Estimation of Low-Rank Matrices via Approximate Message Passing*. 2019. arXiv: 1711.01682 [math.ST].

[Son+19] Yang Song et al. *Sliced Score Matching: A Scalable Approach to Density and Score Estimation*. 2019. arXiv: 1905.07088 [cs.LG]. URL: https://arxiv.org/abs/1905.07088.

[FMM20] Zhou Fan, Song Mei, and Andrea Montanari. *TAP free energy, spin glasses, and variational inference*. 2020. arXiv: 1808.07890 [math.PR].

[HJA20] Jonathan Ho, Ajay Jain, and Pieter Abbeel. *Denoising Diffusion Probabilistic Models*. 2020. arXiv: 2006.11239 [cs.LG]. URL: https://arxiv.org/abs/2006.11239.

[SE20] Yang Song and Stefano Ermon. *Generative Modeling by Estimating Gradients of the Data Distribution*. 2020. arXiv: 1907.05600 [cs.LG]. URL: https://arxiv.org/abs/1907.05600.

[Ana+21] Nima Anari et al. *Entropic Independence I: Modified Log-Sobolev Inequalities for Fractionally Log-Concave Distributions and High-Temperature Ising Models*. 2021. arXiv: 2106.04105 [cs.DS]. URL: https://arxiv.org/abs/2106.04105.

[Fen+21] Oliver Y. Feng et al. *A unifying tutorial on Approximate Message Passing*. 2021. arXiv: 2105.02180 [math.ST].

[Son+21] Yang Song et al. *Score-Based Generative Modeling through Stochastic Differential Equations*. 2021. arXiv: 2011.13456 [cs.LG].

[AM22] T. Aspelmeier and M. A. Moore. "Free-energy barriers in the Sherrington-Kirkpatrick model". In: *Physical Review E* 105.3 (Mar. 2022). ISSN: 2470-0053. DOI: 10.1103/physreve.105.034138. URL: http://dx.doi.org/10.1103/PhysRevE.105.034138.

[Cam+22] Andrew Campbell et al. *A Continuous Time Framework for Discrete Denoising Models*. 2022. arXiv: 2205.14987 [stat.ML]. URL: https://arxiv.org/abs/2205.14987.

[Cel22] Michael Celentano. *Sudakov-Fernique post-AMP, and a new proof of the local convexity of the TAP free energy*. 2022. arXiv: 2208.09550 [math.PR].

[EKZ22] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni. "A spectral condition for spectral gap: fast mixing in high-temperature Ising models". In: *Probability Theory and Related Fields* 182.3 (2022), pp. 1035–1051. DOI: 10.1007/s00440-021-01085-x. URL: https://doi.org/10.1007/s00440-021-01085-x.

[NSZ22] Danny Nam, Allan Sly, and Lingfu Zhang. *Ising model on trees and factors of IID*. 2022. arXiv: 2012.09484 [math.PR]. URL: https://arxiv.org/abs/2012.09484.

[AMS23] A. El Kacimi Alaoui, Andrea Montanari, and Mark Sellke. "Sampling from Mean-Field Gibbs Measures via Diffusion Processes". In: 2023. URL: https://api.semanticscholar.org/CorpusID:264128362.

[CFM23] Michael Celentano, Zhou Fan, and Song Mei. *Local convexity of the TAP free energy and AMP convergence for Z2-synchronization*. 2023. arXiv: 2106.11428 [math.ST].

[Che+23] Sitan Chen et al. *Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions*. 2023. arXiv: 2209.11215 [cs.LG].

[Ghi+23] Davide Ghio et al. *Sampling with flows, diffusion and autoregressive neural networks: A spin-glass perspective*. 2023. arXiv: 2308.14085 [cond-mat.dis-nn].

[Mon23] Andrea Montanari. *Sampling, Diffusions, and Stochastic Localization*. 2023. arXiv: 2305.10690 [cs.LG].

[MW23] Andrea Montanari and Yuchen Wu. *Posterior Sampling from the Spiked Models via Diffusion Processes*. 2023. arXiv: 2304.11449 [math.ST].

[AMS24] Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. *Sampling from the Sherrington-Kirkpatrick Gibbs measure via algorithmic stochastic localization*. 2024. arXiv: 2203.05093 [math.PR].

A Background on Approximate Message Passing

A.1 Low-Rank Matrix Estimation

We consider the planted model introduced in Section 4.1 of [AMS24]. Namely, we have $\mathbf{x}_0, \mathbf{y}(t) \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{n\times n}$ such that

$$\mathbf{x}_0 \sim \bar{\nu}, \tag{A.1}$$
$$\mathbf{A} \sim \mu_{\text{pl}}(\cdot|\mathbf{x}_0), \tag{A.2}$$
$$\mathbf{y}(t) = t\mathbf{x}_0 + \mathbf{B}(t), \tag{A.3}$$

where $\mathbf{B}(t)$ is n-dimensional Brownian motion, independent of everything else so far, and $\bar{\nu}$ denotes the uniform distribution over $\{+1, -1\}^n$. As argued in [AMS24], the matrix $\mathbf{A}$ is described by the following spike mdoel:

$$\mathbf{A} = \frac{\beta}{n}\mathbf{x}_0\mathbf{x}_0^T + \mathbf{W}, \tag{A.4}$$

where $\mathbf{W} \sim \text{GOE}(n)$. We now wish to use an Approximate Message Passing algorithm to compute the posterior mean of $\mathbf{x}_0$ given the *observation process* described in (2.21). The problem has now effectively been reduced to a rank-one Wigner spike estimation problem, where the spike is uniformly distributed over $\{+1, -1\}^n$. This problem is also known as $\mathbb{Z}_2$-*synchronization* (see synchronization over compact groups [Per+18]).

We will now construct the AMP iterations, thus effectively 'recreating' Section 4 of [AMS24]. The general AMP procedure for estimating $\mathbf{x}_0$ is as follows ([Fen+21]):

$$\begin{aligned} x^k &:= g_k(z^k), \\ b_k &:= \frac{1}{n}\sum_{i=1}^n g_k'(z_i^k), \\ z^{k+1} &:= Ax^k + y - b_k x^{k-1}, \end{aligned} \tag{A.5}$$

where $(g_k)_{k=0}^\infty$ is a sequence of Lipschitz functions, and the initializers are $z^0 = x^{-1} = 0$.

A.1.1 State Evolution & Limiting Covariance

Recalling Section 3 of [Fen+21], which covers general AMP implementations for our problem of interest, we have that the *state evolution* of our AMP is

$$\mu_{k+1} := \beta \cdot \mathbb{E}\left[X_0 \cdot g_k(\mu_0 X_0 + \sigma_k G)\right], \tag{A.6}$$
$$\sigma_{k+1}^2 := \mathbb{E}\left[g_k(\mu_k X_0 + \sigma_k G)^2\right], \tag{A.7}$$

where $X_0 \sim \bar{\nu}$ and $G \sim \mathcal{N}(0,1)$ and independent. Moreover, the iterations admit a *limiting covariance structure* $\Sigma_{k,\ell} \in \mathbb{R}^{k\times k}$ for $1 \leq \ell \leq k$:

$$\Sigma_{k+1,\ell+1} := \mathbb{E}\left[g_k(\mu_k X_0 + \sigma_k G) \cdot g_\ell(\mu_\ell X_0 + \sigma_\ell G)\right]. \tag{A.8}$$

Remark A.1. *Clearly, we will have that $\sigma_k^2 = \Sigma_{k,k}$.*

A.1.2 Bayes AMP

We now turn to the *choice of the denoising functions* g_k. In our problem, the prior of the signal of interest $\mathbf{x}_0$ is known, which allows us to use *Bayes Approximate Message Passing* (Section 3.3 [Fen+21], [DAM15][7]). The optimal choice is to denoise using the *Bayes estimator*

$$g_k(\mu_k X_0 + \sigma_k G) := \mathbb{E}[X_0|\mu_k X_0 + \sigma_k G]. \tag{A.9}$$

This in turn reduces our problem to denoising single channels based on observations corrupted by Gaussian noise. Our new state evolution is thus

$$\mu_{k+1} := \beta \cdot \mathbb{E}\left[X_0 \cdot \mathbb{E}[X_0|\mu_k X_0 + \sigma_k G]\right], \tag{A.10}$$

$$\sigma_{k+1}^2 := \mathbb{E}\left[\mathbb{E}[X_0|\mu_k X_0 + \sigma_k G]^2\right]. \tag{A.11}$$

Notice that in our case of $\mathbb{Z}_2 - \textit{synchronization}$, the Bayes estimator is given by

$$\mathbb{E}[X_0|\mu_k X_0 + \sigma_k G] = \tanh(\mu_k X_0 + \sigma_k G). \tag{A.12}$$

Thus,

$$\begin{aligned}
\mu_{k+1} &= \beta \cdot \mathbb{E}\left[X_0 \cdot \mathbb{E}[X_0|\mu_k X_0 + \sigma_k G]\right] \\
&= \beta \cdot \mathbb{E}\left[X_0 \cdot \tanh(\mu_k X_0 + \sigma_k G)\right] \\
&= \beta \cdot \mathbb{E}\left[\frac{1}{2}\tanh(\underbrace{\mu_k + \sigma_k G}_{:=\tilde{G}}) - \frac{1}{2}\tanh(-\mu_k + \sigma_k G)\right] \\
&= \beta \cdot \mathbb{E}\left[\frac{1}{2}\tanh(\tilde{G}) - \frac{1}{2}\tanh(-\tilde{G})\right] \\
&= \beta^2 \cdot \mathbb{E}\left[\tanh(\tilde{G})\right] \\
&= \beta \cdot \mathbb{E}\left[\tanh(\mu_k + \sigma_k G)\right] \\
&\overset{a)}{=} \beta \cdot \mathbb{E}\left[\tanh^2(\mu_k + \sigma_k G)\right] \\
&= \beta \cdot \mathbb{E}\left[\mathbb{E}[X_0|\mu_k X_0 + \sigma_k G]^2\right] \\
&= \beta \cdot \sigma_{k+1}^2.
\end{aligned} \tag{A.13}$$

where $a)$ is justified by Eq. (37) of [Fen+21].

Due to this surprisingly practical reduction, we then define the *signal-to-noise ratio* γ_k as follows (see also Eq. (38) [Fen+21]):

$$\begin{aligned}
\gamma_{k+1} &:= \left(\frac{\mu_{k+1}}{\sigma_{k+1}}\right)^2 \\
&= \beta^2 \cdot \sigma_{k+1}^2 \\
&= \beta^2 \cdot \mathbb{E}\left[\mathbb{E}[X_0|\beta\sigma_k^2 X_0 + \sigma_k G]^2\right] \\
&= \beta^2 \cdot \mathbb{E}\left[\mathbb{E}[X_0|\beta^2\sigma_k^2 X_0 + \beta\sigma_k G]^2\right] \\
&= \beta^2 \cdot \mathbb{E}\left[\tanh(\gamma_k + \sqrt{\gamma_k} G)\right],
\end{aligned} \tag{A.14}$$

effectively recovering Eq. (4.16) of [AMS24] for $t = 0$.

[7] [DAM15] treats yet another similar, but slightly different case with X_0 being the output of a binary erasure channel.

A.1.3 Adding non-zero time t

As seen in A.5, the observation is actually made using two elements, unlinke in Section 3 of [Fen+21], [DAM15]. The situation more so resembles the case of [MW23], where we must take the observation process $\mathbf{y}(t) = t\mathbf{x}_0 + \mathbf{B}(t)$ into consideration. Addressed in Section 3 of [MW23], we wish to compute not just $\mathbb{E}[\mathbf{x}_0|\mathbf{A}]$, but rather $\mathbb{E}[\mathbf{x}_0|\mathbf{A}, \mathbf{y}(t)]$. We now slightly modify our state evolution by now considering the posterior mean with the observation process $\mathbf{y}(t)$. This could be considered as just having two different Gaussian channels to denoise X_0 ([DAM15]) uses the term *effective Gaussian scalar channels*. We will see below how to actually combine them for full efficiency.

$$
\begin{aligned}
\gamma_{k+1}(\beta, t) &= \left(\frac{\mu_{k+1}}{\sigma_{k+1}}\right)^2 \\
&= \beta^2 \sigma_{k+1}^2 \\
&= \beta^2 \cdot \mathbb{E}\left[\mathbb{E}[X_0|\beta\sigma_k^2 X_0 + \sigma_k G, tX_0 + B_k(t)]^2\right] \\
&= \beta^2 \cdot \mathbb{E}\left[\mathbb{E}[X_0|\beta^2\sigma_k^2 X_0 + \beta\sigma_k G, tX_0 + B_k(t)]^2\right] \\
&\overset{B(t)\text{ indep.}}{=} \beta^2 \cdot \mathbb{E}\left[\mathbb{E}\left[X_0|(\underbrace{\beta^2\sigma_k^2}_{\gamma(\beta,t)} + t)X_0 + \sqrt{\beta^2\sigma_k^2 + t}G\right]^2\right] \\
&= \beta^2 \cdot \mathbb{E}\left[\mathbb{E}\left[X_0|\left(\gamma_k(\beta,t) + t\right) X_0 + \sqrt{\gamma_k(\beta,t) + t}G\right]^2\right] \\
&= \beta^2 \cdot \mathbb{E}\left[\tanh(\gamma_k(\beta,t) + t + \sqrt{\gamma_k(\beta,t) + t}G)\right], \qquad \text{(A.15)}
\end{aligned}
$$

thus correctly recovering Eq. (4.16) of [AMS24].

We now propose a perhaps simpler change of variables

$$
\alpha_k(\beta, t) := \gamma_k(\beta, t) + t, \tag{A.16}
$$

and can now rewrite the state evolution finally as

$$
\begin{aligned}
\alpha_0(\beta, t) &:= t, \qquad \text{(A.17)} \\
\alpha_{k+1}(\beta, t) &:= \beta^2 \cdot \mathbb{E}\left[\tanh^2\left(\alpha_k(\beta,t) + \sqrt{\alpha_k(\beta,t)}G\right)\right] + t \\
&= \beta^2\left(1 - \mathrm{mmse}(\alpha_k(\beta,t))\right) + t. \qquad \text{(A.18)}
\end{aligned}
$$

Or equivalently,

$$
\gamma_{k+1}(\beta, t) = \beta^2\left(1 - \mathrm{mmse}(\gamma_k(\beta,t) + t)\right), \tag{A.19}
$$

which exactly recovers Eq. (4.20) of [AMS24], where the mmmse$(\cdot)$ function is the minimum mean square error:

$$
\mathrm{mmse} : \gamma \mapsto 1 - \mathbb{E}\left[\tanh^2\left(\gamma + \sqrt{\gamma}G\right)\right] \tag{A.20}
$$

A.2 Coupling the CDMA AMP as an Observation Process

We can now apply the same logic, but with our procedurally denoised observation process $\mathbf{y}(t)$, explained in Section 4.2.1.

$$\begin{aligned}
\gamma_{k+1}(\beta, t) &= \beta^2 \cdot \mathbb{E}\left[\mathbb{E}\left[X_0 | \beta^2\sigma_k^2 X_0 + \beta\sigma_k G, \mathbf{y}(t)\right]^2\right] \\
&\overset{a)}{=} \beta^2 \cdot \mathbb{E}\left[\mathbb{E}\left[X_0 | \beta^2\sigma_k^2 X_0 + \beta\sigma_k G, tX_0 + \sqrt{t}\tau_k Z\right]^2\right] \\
&\overset{b)}{=} \beta^2 \cdot \mathbb{E}\left[\mathbb{E}\left[X_0 | \beta^2\sigma_k^2 X_0 + \beta\sigma_k G, t\tau_k^{-2} X_0 + \sqrt{t}\tau_k^{-1} Z\right]^2\right] \\
&\overset{c)}{=} \beta^2 \cdot \mathbb{E}\left[\mathbb{E}\left[X_0 | (\beta^2\sigma_k^2 + t\tau_k^{-2}) X_0 + \sqrt{\beta^2\sigma_k^2 + t\tau_k^{-2}} G\right]^2\right] \qquad \text{(A.21)}
\end{aligned}$$

where $a)$ holds by asymptotic MMSE optimality of AMP and the decoupling principle ([Jeo+15], [MT06]), $b)$ holds by Proposition A.1 and $c)$ holds by independence of the driving Brownian motions. We have thus successfully recovered (4.15).

Proposition A.1. *Let X be a Rademacher random variable, and $Y := \sqrt{a}X + Z$, $\tilde{Y} := aX + \sqrt{a}Z$ be two scaled Gaussian observations of X, for $a \in \mathbb{R}_{>0}$ and $Z \sim \mathcal{N}(0, 1)$. Then, the Bayes estimators coincide.*

Moreover, if X is a random n-vector of i.i.d. Rademacher random variables, and $Y = \sqrt{Q}X + Z$, $Y' = QX + \sqrt{Q}Z$ are Gaussian observations for a PSD $n \times n$ matrix Q and $Z \sim \mathcal{N}(0, \mathbf{I}_n)$, then the Bayes estimators coincide as well.

Proof. The Bayes estimators for the scalar case are given by the formula

$$\begin{aligned}
\mathbb{E}[X|Y = y] &= \frac{e^{-\frac{(y-\sqrt{a})^2}{2}} - e^{-\frac{(y+\sqrt{a})^2}{2}}}{e^{-\frac{(y-\sqrt{a})^2}{2}} + e^{-\frac{(y+\sqrt{a})^2}{2}}}, \\
\mathbb{E}[X|Y' = y'] &= \frac{e^{-\frac{(y'-a)^2}{2a}} - e^{-\frac{(y'+a)^2}{2a}}}{e^{-\frac{(y'-a)^2}{2a}} + e^{-\frac{(y'+a)^2}{2a}}}, \qquad \text{(A.22)}
\end{aligned}$$

both of which reduce to

$$\begin{aligned}
\mathbb{E}[X|Y = y] &= \tanh\left(\sqrt{a}y\right), \\
\mathbb{E}[X|Y' = y'] &= \tanh\left(y'\right). \qquad \text{(A.23)}
\end{aligned}$$

The claim follows by $y' = \sqrt{a}y$.

For the vector case, the Bayes estimators analogously read

$$\begin{aligned}
\mathbb{E}[X|Y] &= \tanh\left(\sqrt{Q}^{-1} Y\right), \\
\mathbb{E}[X|Y] &= \tanh\left(Q^{-1} Y'\right), \qquad \text{(A.24)}
\end{aligned}$$

where $\tanh(\cdot)$ is applied element-wise. The claim follows by $Y' = \sqrt{Q}Y$. □