Fourier Analysis of Iterative Algorithms

Chris Jones
Bocconi University, Milano, Italy

Lucas Pesenti
Bocconi University, Milano, Italy

Abstract

We study a general class of nonlinear iterative algorithms which includes power iteration, belief propagation and approximate message passing, and many forms of gradient descent. When the input is a random matrix with i.i.d. entries, we use Boolean Fourier analysis to analyze these algorithms as low-degree polynomials in the entries of the input matrix. Each symmetrized Fourier character represents all monomials with a certain shape as specified by a small graph, which we call a *Fourier diagram*.

We prove fundamental asymptotic properties of the Fourier diagrams: over the randomness of the input, all diagrams with cycles are negligible; the tree-shaped diagrams form a basis of *asymptotically independent Gaussian vectors*; and, when restricted to the trees, iterative algorithms exactly follow an idealized Gaussian dynamic. We use this to prove a state evolution formula, giving a "complete" asymptotic description of the algorithm's trajectory.

The restriction to tree-shaped monomials mirrors the assumption of the *cavity method*, a 40-year-old non-rigorous technique in statistical physics which has served as one of the most important techniques in the field. We demonstrate how to implement cavity method derivations by 1) restricting the iteration to its tree approximation, and 2) observing that heuristic cavity method-type arguments hold rigorously on the simplified iteration. Our proofs use combinatorial arguments similar to the trace method from random matrix theory.

Finally, we push the diagram analysis to a number of iterations that scales with the dimension n of the input matrix, proving that the tree approximation still holds for a simple variant of power iteration all the way up to $n^{\Omega(1)}$ iterations.

2012 ACM Subject Classification Mathematics of computing → Statistical paradigms; Mathematics of computing → Stochastic control and optimization; Mathematics of computing → Loopy belief propagation

Keywords and phrases Iterative Algorithms, Message-passing Algorithms, Random Matrix Theory

Digital Object Identifier 10.4230/LIPIcs.ICALP.2025.102

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: `https://arxiv.org/abs/2404.07881` [40]

1 Introduction

We study nonlinear iterative algorithms which take as input a matrix $A \in \mathbb{R}^{n \times n}$, maintain a vector state $x_t \in \mathbb{R}^n$, and at each step

1. either multiply the state by A, $x_{t+1} = Ax_t$.
2. or apply the same function $f_t : \mathbb{R}^{t+1} \to \mathbb{R}$ to each coordinate of the previous states, $x_{t+1} = f_t(x_t, \dots, x_0)$.

This class of algorithms has been coined *general first-order methods* (GFOM) [15, 64]. GFOM algorithms are a simple, widespread, practically efficient, and incredibly powerful computational model. Alternating linear and nonlinear steps can describe first-order optimization algorithms including power iteration and many types of gradient descent (see [15, 33]). This definition also captures belief propagation and other message-passing algorithms which


© Chris Jones and Lucas Pesenti;
licensed under Creative Commons License CC-BY 4.0
52nd International Colloquium on Automata, Languages, and Programming (ICALP 2025).
Editors: Keren Censor-Hillel, Fabrizio Grandoni, Joël Ouaknine, and Gabriele Puppis
Article No. 102; pp. 102:1–102:21


Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

play a central role not only in the design of Bayes-optimal algorithms for planted signal recovery [29], but also recently in average-case complexity theory for the optimization of random polynomials [4].

In machine learning and artificial intelligence, deep neural networks exhibit a similar structure which alternates multiplying weight matrices and applying nonlinear functions. Remarkably, viewed from this level of generality, the line blurs between neural networks and the gradient descent algorithms used to train them.

Despite the widespread use of GFOM and deep neural networks, developing a mathematical theory for these algorithms continues to be a major challenge. Thus far, it has been difficult to isolate mathematical theorems which describe key phenomena but avoid being too specific to any one setting, model, or algorithm. That being said, one effective theory has emerged at the interface of computer science, physics, and statistics for studying a class of nonlinear iterations known as Belief Propagation (BP) and Approximate Message Passing (AMP) algorithms. This theory is most developed for inputs A that are *dense random matrices with i.i.d entries*, also known as a *mean-field models* in physics, and which can be considered the simplest possible model of random data.

The analysis of BP and AMP algorithms in this setting can be summarized by the *state evolution* formula [25, 11]. This is an impressive "complete" description of the trajectory of the iterates $x_t \in \mathbb{R}^n$, in the limit $n \to \infty$. Specifically, state evolution defines a sequence of *scalar* random variables X_t such that for essentially *any* symmetric quantity of interest related to x_t, the expectation of a corresponding expression in X_t approximates the quantity with an error that goes to 0 as $n \to \infty$. This yields analytic formulas for quantities such as the loss function or objective value achieved by x_t, the norm of x_t, the correlation between x_s and x_t across iterations, or the fraction of x_t's coordinates which lie in the interval $[-1, +1]$. The ability to precisely analyze the trajectory and the fixed points of message-passing algorithms (through X_t with large t) has been key to their applications.

State evolution for BP/AMP iterations was originally predicted using a powerful and influential technique from statistical physics known as the *cavity method*. Variants of BP have been studied in physics as "non-linear dynamical systems" as far back as the work of Bethe [10], although the algorithmic perspective came into prominence only later. The cavity method and the related replica method were devised in the 1980s [68, 69, 59, 60], initially as a tool to compute thermodynamic properties of disordered systems, and later as a tool for analyzing belief propagation algorithms. Since their introduction, the cavity method and the replica method have served as two of the most fundamental tools in the statistical physics toolbox.

The deployment of these techniques has undoubtedly been a huge success; there are many survey articles offering various perspectives from different fields [87, 65, 46, 88, 31, 29, 89, 16]. However, the reality is that the situation is not as unified as the above picture would suggest, due to a major issue: *the physical methods are not mathematically rigorous.*

At present, there exists a significant gap between how results are established in the physical and mathematical literature. The two general types of results are: 1) simple non-rigorous arguments based on the cavity/replica method; 2) mathematically rigorous arguments that confirm the physical predictions, but with technically sophisticated proofs that can't closely follow the path of the physical reasoning. For example, the state evolution formula was first proven by Bolthausen [11] using a Gaussian conditioning technique which is fairly technically involved. Although many proofs have been found for predictions of the cavity and replica methods, none can be said to clearly explain the success of the physicists' techniques.

It has appeared that the physicists have some secret advantage yet unmatched by rigorous mathematics. Is there a simple and rigorous mathematical framework that explains why the assumptions made by physicists always seem to work?

1.1 Our contributions

We introduce a new method to analyze nonlinear iterations based on Fourier analysis, when the input to the algorithm is a random matrix with i.i.d entries. Our framework gives proofs that are able to closely follow heuristic physics derivations.

Our strategy is to replace the original iteration $(x_t)_{t\geq 0}$ by an idealized version $x_t \approx \widehat{x}_t$, which we call the *tree approximation* to x_t. The analysis then follows a two-step structure:

1. The tree approximation $\widehat{x}_t$ tracks the original iteration x_t up to a uniform $\widetilde{O}(n^{-\frac{1}{2}})$ entrywise error. Hence, any reasonable asymptotic result established on $(\widehat{x}_t)_{t\geq 0}$ (such as the joint distribution of their entries) automatically extends to $(x_t)_{t\geq 0}$.
2. Cavity method-type reasoning can be rigorously applied to the tree approximation. In cases where $\widehat{x}_t$ has already been analyzed in physics, one can essentially copy the heuristic physics derivation.

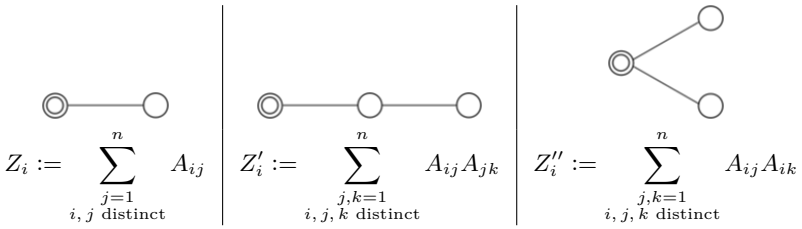
Analyzing $\widehat{x}_t$ is a significant simplification compared to the entire state x_t – in fact, we show that the former follows an explicit *Gaussian dynamic*. The simplification directly yields a *state evolution* formula for GFOM algorithms, as well as rigorous implementations of physics-based cavity method arguments (in the algorithmic or "replica symmetric" setting of the method). In other words, our new notion of tree approximation matches implicit assumptions of the cavity method and gives a way to justify them.

We define the tree approximation $\widehat{x}_t$ essentially as follows: we expand the entries of x_t as polynomials in the entries of the input matrix $A \in \mathbb{R}^{n\times n}$. If we represent the monomials (e.g. $A_{12}A_{23}A_{24}$) as graphs in the natural way, then $\widehat{x}_t$ consists of only the monomials appearing in x_t whose graph is a tree. Hence, we will show that the state of an iterative algorithm can be tightly approximated using the much smaller set of tree monomials.

1.1.1 Fourier diagrams

We view iterates of a GFOM with polynomial non-linearities as vector-valued polynomials in the entries of A. These polynomials have the symmetry that they are invariant under permutations of the row/column indices of A.

The polynomial representation can be visualized using *Fourier diagrams*, each of which is a small graph representing all the monomials with a given shape. For example, here are three Fourier diagrams along with the vectors associated with them.



$$Z_i := \sum_{\substack{j=1 \\ i,\, j \text{ distinct}}}^{n} A_{ij} \qquad Z'_i := \sum_{\substack{j,k=1 \\ i,\, j,\, k \text{ distinct}}}^{n} A_{ij}A_{jk} \qquad Z''_i := \sum_{\substack{j,k=1 \\ i,\, j,\, k \text{ distinct}}}^{n} A_{ij}A_{ik}$$

In general, a Fourier diagram is an undirected rooted multigraph $\alpha = (V(\alpha), E(\alpha))$ which represents the vector $Z_\alpha \in \mathbb{R}^n$ whose entries are:

$$Z_{\alpha,i} := \sum_{\substack{\text{injective } \varphi\colon V(\alpha)\to[n] \\ \varphi(\circledcirc)=i}} \prod_{\{u,v\}\in E(\alpha)} A_{\varphi(u)\varphi(v)}\,, \qquad \text{for all } i \in [n]\,. \tag{1}$$

We use $\circledcirc \in V(\alpha)$ to notate the root vertex.

The symmetry of the GFOM operations ensures that in the polynomial representation of an iterate x_t, all monomials corresponding to the same Fourier diagram come with the same coefficient. Therefore, any iterate x_t of a GFOM with polynomial non-linearities can be expressed as a linear combination of Fourier diagrams, in which case we say that it is written *in the Fourier diagram basis*.

We emphasize that these diagrams are constructed by summing over *injective* embeddings $\varphi\colon V(\alpha) \to [n]$, a crucial detail for the results that follow. The term "Fourier" reflects that this basis of polynomials is a symmetrized version of the standard Fourier basis from Boolean function analysis (see Section 6).

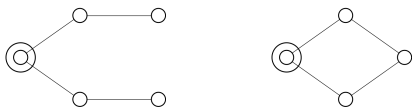
1.1.2 Asymptotic diagram analysis

It turns out that something special happens to the Fourier diagram basis in the limit $n \to \infty$, when A is a symmetric matrix with independent mean-0, variance-$\frac{1}{n}$ entries. Informally, the entries of the diagrams become mutually independent, and the following properties hold.

- The Fourier diagrams with cycles are negligible.
- The Fourier tree diagrams with one branch from the root are independent Gaussian vectors.
- The Fourier tree diagrams with several branches from the root are Hermite polynomials in the Gaussians represented by the branches.

Most importantly, **the only non-negligible contributions come from the trees**. Based on this classification, we define the *tree approximation* $\widehat{x}_t$ of an expression x_t written in the Fourier diagram basis to be obtained by discarding all diagrams with cycles.

The reason that cyclic Fourier diagrams are negligible is combinatorially intuitive: cyclic diagrams sum over fewer terms than tree-shaped diagrams. For example, the left diagram is a sum over $\approx n^4$ terms, each mean-zero with magnitude $\sim n^{-2}$, so the overall magnitude is $\Theta(1)$. The right diagram is a sum over $\approx n^3$ terms, again of magnitude n^{-2}, so the overall of order of the diagram is $\Theta(n^{-1/2})$.



We now state our main theorems. In all of them, we assume that A is a symmetric matrix with independent mean-0 variance-$\frac{1}{n}$ entries (see Assumption 2.1). First, we formalize the classification above by proving that all joint moments of the Fourier diagrams converge to those of the corresponding random variables in a Gaussian space.

▶ **Theorem 1** (Classification theorem). *For any $k \geq 0$ independent of n, for all connected Fourier diagrams $\alpha_1, \ldots, \alpha_k$ and $i_1, \ldots, i_k \in [n]$ (allowing repetitions in α_j and i_j),*

$$\mathbb{E}_A\left[\prod_{j=1}^{k} Z_{\alpha_j, i_j}\right] = \mathbb{E}\left[\prod_{j=1}^{k} Z^{\infty}_{\alpha_j, i_j}\right] + O(n^{-\frac{1}{2}}),$$

where for any connected Fourier diagram α and $i \in [n]$,

1. *$Z^{\infty}_{\alpha,i} = 0$, if α has a cycle.*
2. *$Z^{\infty}_{\alpha,i} \sim \mathcal{N}(0, |\mathrm{Aut}(\alpha)|)$ independently, if α is a tree whose root has degree 1.*
3. *$Z^{\infty}_{\alpha,i} = \prod_{\tau} h_{d_\tau}(Z^{\infty}_{\tau,i}\,; |\mathrm{Aut}(\tau)|)$ if α is a tree consisting of d_τ copies of each tree τ from case 2 merged at the root, where h_{d_τ} are the Hermite polynomials (defined in Section 2).*

Next, we prove that the tree approximation of a GFOM closely tracks the original iteration. This addresses the first of the two steps from the overview of our method.

▶ **Theorem 2** (Tree approximation of GFOMs). *Let t be a constant, $x_t \in \mathbb{R}^n$ be the state of a GFOM with polynomial non-linearities, and $\widehat{x}_t \in \mathbb{R}^n$ be the state obtained by performing the GFOM operations on only the tree diagrams. Then with high probability over A, we have $\|x_t - \widehat{x}_t\|_\infty = \widetilde{O}(n^{-\frac{1}{2}})$.*

The statement of this theorem exactly isolates a key and subtle point: not only are the cyclic diagrams negligible at time t, but they will never blow up to affect the state at any future time. The fact that "approximation errors do not propagate" is what gives us the ability to pass the algorithm to an asymptotic limit.[1]

The proof of Theorem 2 is intuitive. According to the diagram classification theorem, we can tease out the approximation error for x_t as the monomials with cycles, whereas the approximating quantity $\widehat{x}_t$ consists of the tree monomials. When a GFOM operation is applied, the cycles persist in all cyclic monomials, and hence they continue to be negligible.

As a direct consequence of these results, we deduce a very strong form of state evolution for all GFOM algorithms. The theorem below paints a nearly complete picture of the evolution of x_t as n independent trajectories of a single random variable X_t which is an "idealized Gaussian dynamic" in correspondence with $\widehat{x}_t$.

▶ **Theorem 3** (General state evolution). *Let t be a constant, $x_t \in \mathbb{R}^n$ be the state of a GFOM with polynomial non-linearities, and let X_t be the asymptotic state of x_t (Definition 15). Then:*

- **(i)** *For each $i \in [n]$, $(x_{0,i}, \dots, x_{t,i}) \xrightarrow{\mathrm{d}} (X_0, \dots, X_t)$. Furthermore, the coordinates' trajectories $\{(x_{0,i}, \dots, x_{t,i}) : i \in [n]\}$ are asymptotically independent.*
- **(ii)** *With high probability over A,*

$$\frac{1}{n}\sum_{i=1}^{n} x_{t,i} = \mathbb{E}[X_t] + \widetilde{O}(n^{-\frac{1}{2}})\,.$$

Quantities such as the norm of x_t can be computed using part (ii) along with one additional GFOM iteration that squares x_t componentwise. Without much extra work, Theorem 3 also encapsulates other key features of previous state evolution formulas including quantitative error bounds (similar to the main result of [75]) and universality (the main result of [6]).[2]

1.1.3 The cavity method

To explain the cavity method in one sentence, it allows you to assume that "loopy" message-passing algorithms on random graphs behave as if on a tree, gaining extra properties such as the independence of incoming messages. It turns out that the assumption of being on a tree matches the restriction to tree-shaped monomials in A, leading to a way to rigorously implement simple cavity method reasoning.

[1] This directly addresses a question raised in the seminal paper of Donoho, Maleki, and Montanari on approximate message passing [26, Section III.E].

[2] Similarly to [6], our technical analysis assumes that the nonlinearities in the GFOM are polynomial functions, but other works have been able to handle the larger class of *pseudo-Lipschitz* non-linearities. We do not find this assumption to be too restrictive since it is known in many cases that we can approximate the non-linearities by polynomials [38, Appendix B].

We formalize two types of cavity method arguments. For the first one, we introduce a combinatorial notion of asymptotic equality $\overset{\infty}{=}$ which can rigorously replace heuristic approximations in the cavity method.

▶ **Definition 4** ($\overset{\infty}{=}$, informal version of [40, Definition 4.4]). *Let $x \overset{\infty}{=} y$ if $x - y$ is a sum of constantly many diagrams with cycles.*

As an application of this definition, we implement the cavity method argument that belief propagation and approximate message passing are asymptotically equivalent for dense random matrices. We refer to [40, Section 5] for some background and the definition of these message-passing iterations.

▶ **Theorem 5** (Equivalence of BP and AMP). *Let m_t^{BP} and m_t^{AMP} be the iterates of respectively the belief propagation and the approximate message passing iterations on the same input matrix A and the same polynomial non-linearities. Then with high probability over A, $\left\|m_t^{\mathrm{BP}} - m_t^{\mathrm{AMP}}\right\|_\infty = \widetilde{O}(n^{-\frac{1}{2}})$.*

We also use $\overset{\infty}{=}$ to prove a fundamental assumption of the cavity method for belief propagation iterations on dense models, namely that the messages incoming at a vertex are asymptotically independent.

▶ **Theorem 6** (Asymptotic independence of incoming messages). *Let m_t^{BP} be the iterates of a belief propagation iteration. For any $j \in [n]$, the incoming messages at j, $\{m_{i\to j}^t : i \in [n], i \neq j\}$, are asymptotically independent.*

The second way that we formalize the cavity method reasoning is through the idealized Gaussian dynamic X_t in Theorem 3. We recover the vanilla form of state evolution for approximate message passing, since in this case, X_t has a simple description.

▶ **Theorem 7** (Asymptotic state of AMP). *Consider the AMP iteration*

$$x_{t+1} = Af_t(x_t, \dots, x_0) - \frac{1}{n}\sum_{s=1}^{t}\sum_{i=1}^{n}\frac{\partial f_t}{\partial x_s}(x_{t,i}, \dots, x_{0,i})f_s(x_s, \dots, x_0)\,. \tag{2}$$

The asymptotic state of $(x_0, x_1, \dots)$ is a centered Gaussian vector $(X_0, X_1, \dots)$ with covariances given by the recurrence, for all s, t,

$$\mathbb{E}\left[X_s X_t\right] = \mathbb{E}\left[f_{s-1}(X_{s-1}, \dots, X_0)f_{t-1}(X_{t-1}, \dots, X_0)\right]\,.$$

The subtracted term in Equation 2 is called the *Onsager correction* which, as we show, is carefully designed to cancel out a backtracking term in the asymptotic tree space [40, Lemma 5.11].

We emphasize that Theorem 5 and Theorem 7 are *known*. They were originally predicted with the cavity method, then later confirmed by rigorous proofs (in [6] and [11, 7, 15], respectively). The main message about our proofs is the new and quite comprehensive perspective obtained through the tree approximation, providing a clear way in which GFOM algorithms on dense random inputs "can be assumed to occur on a tree".

Finally, we provide an exposition in [40, Section 5.5] of the breakthrough *iterative AMP* algorithm devised by Montanari to compute ground states of the Sherrington–Kirkpatrick model [62, 2, 3]. We explain from the diagram perspective how the algorithm is the optimal choice among algorithms which "extract" a Brownian motion from the input.

1.1.4 Taking the tree approximation farther

The asymptotic theory above applies to an iterative algorithm running for a *constant* number of iterations. Although this "short-time" setting is used in a large majority of previous works in this area, there is interest in extending the analysis to, say, $O(\log n)$ iterations, which may be enough to capture planted recovery from random initialization and distinct phases of learning algorithms [50].

Can we use the tree-like Fourier characters to analyze the long-time behavior? We show in [40, Section 6] that some care needs to be taken. First, we prove a positive result, that the tree approximation continues to hold for $n^{\Omega(1)}$ iterations for a simple belief propagation algorithm (debiased power iteration, or asymptotically equivalently, power iteration on the non-backtracking walk matrix),

$$x_0 = \vec{1}\,, \quad x_{t+1} = Ax_t - x_{t-1}\,. \tag{3}$$

▶ **Theorem 8.** *Generate $x_t \in \mathbb{R}^n$ from Equation 3 and let $\widehat{x}_t$ be the tree approximation to x_t. Then there exist universal constants $c, \delta > 0$ such that for all $t \le cn^\delta$,*

$$\|x_t - \widehat{x}_t\|_\infty \xrightarrow{\text{a.s.}} 0\,.$$

However, we also identify some problems with the technology which suggest that new ideas will be needed to completely capture the long-time setting. We observe that the asymptotic Gaussian classification theorem is no longer valid for diagrams of size $t \approx \log n$. Finally, we identify a further threshold at $t \approx \sqrt{n}$ iterations beyond which the tree approximation we use seems to break down.

1.1.5 Conclusion

We demonstrate that for iterative algorithms running on dense random inputs, trees are all you need. The tree-shaped Fourier diagrams form an asymptotic basis of independent Gaussian vectors associated to an arbitrary Wigner matrix. This basis seems extremely useful, and we are not aware of any previous works on it.

We note that from the outset, it is not at all clear how to find this basis. Individual monomials (i.e. Boolean Fourier characters) such as $A_{12}A_{23}A_{34}$ and $A_{12}A_{23}A_{13}$ have the same magnitude, and the asymptotic negligibility of the cyclic terms including $A_{12}A_{23}A_{13}$ only appears after summing up the total contribution of all monomials with the same shape. Furthermore, grouping terms in a different way does not identify our notion of tree approximation, such as by allowing repeated indices in Equation 1 (as done in [6, 38]). In this repeated-label representation, there is no clear notion of tree approximation of iterative algorithms (in fact, with this alternative definition, the iterates of a GFOM can always be represented *exactly* with trees!) or of the simplified Gaussian dynamic on trees, which is central to our approach.

As we show, the Fourier tree approximation leads to streamlined proofs of several arguments based on the cavity method. We believe that this framework has potential to generalize well beyond the Wigner case and to address outstanding open problems in the area – such as the long-time setting mentioned above.

1.2 Related work

1.2.1 Comparison with prior work

Our analysis is based on the recent "low-degree paradigm" in which algorithms are analyzed as low-degree multivariate polynomial functions of the input [49]. Several recent works have used a similar approach for iterative algorithms [6, 63, 38], with subtle but crucial differences to our work.

Bayati, Lelarge, and Montanari [6] decompose the AMP iterates into certain "nonreversing" labeled trees. They also observe that the Onsager correction corresponds to a backtracking term. Montanari and Wein [63, Section 4.2] introduce an orthogonal diagram basis (similar to our Fourier diagram basis) in their proof that no low-degree polynomial estimator can outperform AMP for rank-1 matrix estimation. Ivkov and Schramm [38] use a repeated-label representation to show that AMP can be robustified.

Diagrammatically, the main advantage of our method is the precise choice of the Fourier diagram basis. By summing over injective a.k.a. self-avoiding labelings φ in Equation 1, each diagram exactly describes all monomials with a given shape. When working with other polynomial basis, for example diagrams with repeated labels [6, 38] (see [40, Appendix A.3]), the key properties of the Fourier diagram basis (the family of asymptotically independent Gaussian vectors and the associated Gaussian dynamic) do not seem clearly visible. In particular, previous work does not show the tree approximation.

Our results stated above which are cavity method-based reproofs of existing results are Theorem 5, which essentially follows from [6, Proposition 3], and Theorem 7, which was first proven by Bayati and Montanari [7]. Notably, Bayati, Lelarge, and Montanari [6] use an approach based on the moment method as we do. Their proof is somewhat more technical, it does not use the Fourier diagram basis, and it is not able to clearly follow the simple cavity method argument that we reproduce in [40, Section 5.2.1].

We also compare our state evolution formula for GFOM in Theorem 3 with a state evolution formula for GFOM proven by [15]. They give a reduction from GFOM to AMP to derive a state evolution formula for GFOM. The corresponding description of the asymptotic state $X_0, \dots, X_t$ is inside a very compressed probability space generated by t Gaussians with a certain covariance structure.

Our description of the random variables $X_0, \dots, X_t$ (necessarily having the same distribution) has a simpler interpretation inside a larger probability space generated by $(Z_\sigma^\infty)_{\sigma \in \mathcal{S}}$. Both descriptions of the asymptotic state X_t are likely to be valuable for different purposes or explicit calculations. Our formulation of state evolution also includes the asymptotic independence of the trajectories of different coordinates.

1.2.2 Analyzing algorithms as low-degree polynomials

Our technical framework is adapted from the average-case analysis of *Sum-of-Squares* algorithms. The Sum-of-Squares algorithm is a powerful meta-algorithm for combinatorial optimization and statistical inference [73, 30]. Sum-of-Squares has been successfully analyzed on i.i.d. random inputs using *graph matrices*, which are a Fourier basis for matrix-valued functions of a random matrix A in the same way that our diagram basis is a basis for vector-valued functions of A.

The theory appears much more pristine in the current setting, so we hope that the current results can bring some new clarity to the technically challenging works on Sum-of-Squares. Many key ideas on graph matrices are present in a pioneering work by Barak

et al. which analyzes the Sum-of-Squares algorithm for the Planted Clique problem [5] (building on earlier work [20, 57, 35]). Analytical ideas were subsequently isolated by Ahn, Medarametla, and Potechin [1] and Potechin and Rajendran [71, 72] and developed in several more works [34, 74, 42, 41, 44, 43, 47]. Several recent works have made explicit connections between AMP, Sum-of-Squares, and low-degree polynomials [63, 38, 76, 77]. Another similar class of diagrammatic techniques are *tensor networks* [61, 48].

1.2.3 Statistical physics and the cavity method

The cavity and replica methods are widely used in statistical physics to compute the free energy, complexity, etc. of Gibbs distributions on large systems, or similarly to compute the satisfiability threshold, number of solutions, etc. for many non-convex random optimization problems. For an introduction to statistical physics methods in computer science, we recommend the surveys [55, 88, 31], the book [65], and the 40-year retrospective [16]. The cavity method is described in [58] and [65, Part V].

Rigorously verifying the predictions of the physical methods has been far from easy for mathematicians. To highlight some major landmarks in the literature over the past decades, tour-de-force proofs of the Parisi formula for the free energy of the SK model were developed by Talagrand [81, 82] and Panchenko [67]. Ding, Sly, and Sun [22, 21, 23] identified the satisfiability threshold for several random optimization problems including k-SAT with large k. Ding and Sun [24] and Huang [36] rigorously analyze the storage capacity of the Ising perceptron, assuming a numerical condition.

Note that the results above are strictly outside the regime of the current work. They require the replica method in "replica symmetry breaking" settings, whereas we study the simpler but related cavity method in the replica symmetric setting. k-SAT is also a sparse (a.k.a. dilute) model whereas our results are for dense (a.k.a. mean-field) models. Despite these differences, our results tantalizingly suggest that it may be possible to validate the physical techniques in a more direct and generic way than taken by current approaches.

Other authors have also directly considered the cavity assumption, albeit using a less combinatorial approach. Both proofs of the Parisi formula implement analytic forms of the cavity calculation ([82, Section 1.6] and [67, Section 3.5]). The cavity method can also be partially justified for sparse models in the replica symmetric regime using that the interactions are locally treelike with high probability [8, 18].

Diagrammatic methods are common in physics, and in fact they have been used in the vicinity of belief propagation even since a seminal 1977 paper by Thouless–Anderson–Palmer [83] which introduced the TAP equations. A version of the tree approximation actually appears briefly in their diagrammatic formula for the free energy of the SK model in Section 3. However, it has not been clear how or whether these arguments could be made rigorous, and to date, rigorous proofs have not directly followed these approaches.

1.2.4 Belief propagation and AMP

Belief propagation originates in computer science and statistics from Pearl [70]. In the current setting, we can view the underlying graphical model as the complete graph, with correlations between the variables induced by the random matrix A. State evolution was first predicted for BP algorithms in this setting by Kabashima [45] and Donoho–Maleki–Montanari [25]. Since the first rigorous proof of state evolution by Bolthausen [11], his Gaussian conditioning technique has been extended to prove state evolution for many variants of AMP [7, 39, 54, 78, 9, 79, 3, 80, 53, 29, 28, 32, 37].

A notably different proof of state evolution by Bayati, Lelarge, and Montanari [6] uses a moment-based approach which is closer to ours (see also follow-up proofs [17, 19, 84, 27]). These proofs and also ours show universality statements which the Bolthausen conditioning method cannot handle.

All of the above works restrict themselves to a constant number of iterations, although some recent papers push the analysis of AMP in some settings to a superconstant number of iterations [75, 13, 85, 86].

Very recently, [51, 50] managed to analyze $t = \widetilde{\Omega}(n)$ iterations of AMP in the spiked Wigner model. This last line of work is especially intriguing, given that our approach seems to break down at $t \approx \sqrt{n}$ [40, Section 6.1].

The perspective that we take is slightly different from most of these papers. Whereas previous works analyze the asymptotic *distribution* of the AMP iterates over the randomness of A, we give an explicit function $\widehat{x}_t$ which exactly satisfies a "Gaussian dynamics" and asymptotically approximates the iterates. This general approach provides more information and we hope that it has increased potential for generalization.

On first-order iterations which are not BP/AMP algorithms, a smaller number of physical analyses have been performed using the more general techniques of *dynamical mean field theory* [56]. We refer to the survey [31]. Most analyses rely on heuristic arguments, although some more recent works [14, 33, 52] prove rigorous results.

Finally, we note that the tree approximation bears similarities to the suppression of crossing partitions in free probability [66]. Unlike the traditional viewpoint of free probability, the combinatorial cancellations behind the tree approximation occur directly on the trajectory of random objects (the iterates of the algorithm), and not only for averaged quantities associated with them.

1.3 Organization of the paper

In the remaining of the paper, we give the key properties of the Fourier diagrams on a high level, delaying formal statements and proofs to the full version of the paper [40]. After some preliminaries in Section 2, we given an example in Section 3. In Section 4, we define the class of diagrams and describe their behavior both for fixed n and in the limit $n \to \infty$. In Section 5, we summarize how iterative algorithms behave asymptotically. In Section 6, we explain how the diagram basis can be derived from standard discrete Fourier analysis.

2 Preliminaries

To maintain generality, we specify the input (a random matrix) and the algorithm (a first-order iteration), but we do not specify an objective/energy function, and for this reason our results are in the flavor of random matrix theory. While the setting of this paper is a null model without any hidden signal, we expect that our techniques can also be applied to planted recovery problems. A concrete algorithmic application to keep in mind in the null model is the optimization of random degree-2 polynomials that we revisit in [40, Section 5.5].

Our results will apply universally to a Wigner random matrix model (they hold regardless of the specific choice of μ, μ_0 below).

▶ **Assumption 2.1** (Assumptions on matrix entries). *Let μ and μ_0 be two subgaussian*[3] *distributions on $\mathbb{R}$ such that $\mathbb{E}_{X\sim\mu}[X] = 0$ and $\mathbb{E}_{X\sim\mu}[X^2] = 1$.*

Let A be a random $n \times n$ symmetric matrix with independent entries (up to the symmetry) which are either $\sqrt{n}A_{ii} \sim \mu_0$ on the diagonal or $\sqrt{n}A_{ij} \sim \mu$ off the diagonal.

The subgaussian assumption on μ and μ_0 can be relaxed to require only the existence of the q-th moment of μ for some large enough constant $q \in \mathbb{N}$ that depends only on the number of iterations and the degree of the nonlinearities appearing in the algorithm. In this case, our statements of the form "$\|x_n - y_n\|_\infty = \widetilde{O}(n^{-1/2})$ with high probability"[4] weaken to "$\|x_n - y_n\|_\infty \xrightarrow{\text{a.s.}} 0$".

We will refer to the generalized (probabilist's) Hermite polynomials as $h_k(\,\cdot\,;\sigma^2)$, where h_k is the degree-k monic orthogonal polynomial for $\mathcal{N}(0,\sigma^2)$. If Z_i is an independent $\mathcal{N}(0,\sigma_i^2)$ random variable for all $i \in \mathcal{I}$, then $\left(\prod_{i\in\mathcal{I}} h_{k_i}(Z_i;\sigma_i^2)\right)_{k\in\mathbb{N}^{\mathcal{I}}}$ is an orthogonal basis for polynomials in $(Z_i)_{i\in\mathcal{I}}$ with respect to the expectation over $(Z_i)_{i\in\mathcal{I}}$.

3 Example of using diagrams

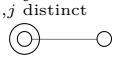
We show how to compute the vector $A(A\vec{1})^2$ in the diagram basis, where $\vec{1} \in \mathbb{R}^n$ denotes the all-ones vector and the square function is applied componentwise. Calculation with diagrams is a bit like a symbolic version of the trace method from random matrix theory [12].

For simplicity, we assume in this section that A satisfies Assumption 2.1 with $A_{ii} = 0$ for all $i \in [n]$.

We will use rooted multigraphs to represent vectors.[5] Multigraphs may include multiedges and self-loops. In our figures, the root will be drawn as a circled vertex ◎. The vector $\vec{1}$ will correspond to the singleton graph with one vertex (the root): ◎. Edges will correspond to A_{ij} terms.

The vector $A\vec{1}$ will be represented by the graph consisting of a single edge, with one of the endpoints being the root:

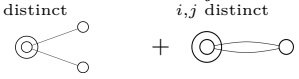
$$(A\vec{1})_i = \sum_{j=1}^{n} A_{ij} = \sum_{\substack{j=1 \\ i,j \text{ distinct}}}^{n} A_{ij} + \equiv$$



where the second equality uses the assumption that A has zero diagonal. Now to apply the square function componentwise, we can decompose:

$$(A\vec{1})_i^2 = \sum_{\substack{j,k=1 \\ i,j,k \text{ distinct}}}^{n} A_{ij}A_{ik} + \sum_{\substack{j=1 \\ i,j \text{ distinct}}}^{n} A_{ij}^2$$

$$\equiv \quad + \quad$$



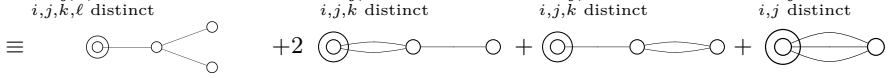
Moving on, we apply A to this representation by casing on whether the new index i matches one of the previous indices. We group terms together using the symmetry of A and the fact that $A_{ii} = 0$.

[3] A distribution μ on $\mathbb{R}$ is *subgaussian* if there exists a constant $C > 0$ such that for all $q \in \mathbb{N}$, $\mathbb{E}_{X\sim\mu}[|X|^q] \le C^q q^{q/2}$.

[4] We say a sequence of events $(A_n)_{n\ge 0}$ occurs with high probability if $\Pr(A_n) \ge 1 - 1/\text{poly}(n)$.

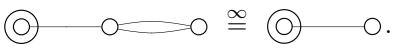
[5] Graphs with multiple roots can be used to represent matrices and tensors, although we will not need those here.

$$(A(A\vec{1})^2)_i = \sum_{\substack{j,k,\ell=1 \\ i,j,k,\ell \text{ distinct}}}^{n} A_{ij}A_{jk}A_{j\ell} + 2\sum_{\substack{j,k=1 \\ i,j,k \text{ distinct}}}^{n} A_{ij}^2 A_{jk} + \sum_{\substack{j,k=1 \\ i,j,k \text{ distinct}}}^{n} A_{ij}A_{jk}^2 + \sum_{\substack{j=1 \\ i,j \text{ distinct}}}^{n} A_{ij}^3$$

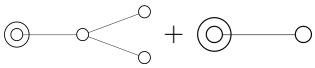


This is the non-asymptotic Fourier diagram representation of $A(A\vec{1})^2$.

In the limit $n \to \infty$, only some of these terms contribute to the *asymptotic* Fourier diagram basis representation. Asymptotically, *hanging* double edges can be removed from a diagram[6], so that the third diagram in the representation above satisfies, as $n \to \infty$,



The second and fourth diagrams in the representation of $A(A\vec{1})^2$ have entries on the scale $O(n^{-1/2})$ and so they will be dropped from the asymptotic diagram representation. In total,

$A(A\vec{1})^2 \overset{\infty}{=}$ 

We will show that as $n \to \infty$, the left diagram becomes a Gaussian vector with independent entries of variance 2, and the right diagram becomes a Gaussian vector with independent entries of variance 1. In fact, these $2n$ entries are asymptotically jointly independent. It can be verified numerically that approximately for large n, $A(A\vec{1})^2$ matches the sum of these two random vectors, the histogram of each vector's entries is Gaussian, and the vectors are approximately orthogonal.

4 Properties of the diagram basis

▶ **Definition 9.** *A* Fourier diagram *is an unlabeled undirected multigraph $\alpha = (V(\alpha), E(\alpha))$ with a special vertex labeled ◎ which we call the root. No vertices may be isolated except for the root. We let $\mathcal{A}$ be the set of all Fourier diagrams.*

▶ **Definition 10** (Z_α)**.** *For a Fourier diagram $\alpha \in \mathcal{A}$ with root ◎, define the vector $Z_\alpha \in \mathbb{R}^n$ by*

$$Z_{\alpha,i} = \sum_{\substack{\varphi: V(\alpha)\to[n] \\ \varphi \text{ injective} \\ \varphi(◎)=i}} \prod_{\{u,v\}\in E(\alpha)} A_{\varphi(u)\varphi(v)}\,, \qquad \text{for all } i \in [n]\,.$$

Among all Fourier diagrams, the ones corresponding to trees play a special role. They will constitute the *asymptotic Fourier diagram basis.*

▶ **Definition 11** ($\mathcal{S}$ and $\mathcal{T}$)**.** *Let $\mathcal{S}$ be the set of unlabeled rooted trees such that the root has exactly one subtree (i.e. the root has degree 1). Let $\mathcal{T}$ be the set of all unlabeled rooted trees (non-empty, but allowing the singleton).*

▶ **Definition 12** (Proper Fourier diagram)**.** *A proper Fourier diagram is a Fourier diagram with no multiedges or self-loops (i.e. a rooted simple graph).*

[6] To be convinced of this, the reader can think of the case where the entries of A are uniform $\pm\frac{1}{\sqrt{n}}$.

For *proper* Fourier diagrams $\alpha \in \mathcal{A}$, the following properties of Z_α hold non-asymptotically i.e. for arbitrary n:

- **(i)** Z_α is a multilinear polynomial in the entries of A with degree $|E(\alpha)|$ (or $Z_\alpha = 0$ when $|V(\alpha)| > n$).
- **(ii)** Z_α has the symmetry that $Z_{\alpha,i}(A) = Z_{\alpha,\pi(i)}(\pi(A))$ for all permutations $\pi \in S_n$, where π acts on A by permuting the rows and columns simultaneously.
- **(iii)** For each $i \in [n]$, the set $\{Z_{\alpha,i} : \text{proper Fourier diagram } \alpha \in \mathcal{A}\}$ is orthogonal with respect to the expectation over A.
- **(iv)** In fact, Z_α is a symmetrized multilinear Fourier character (see Section 6). This implies the previous properties and it shows that the proper diagrams are an orthogonal basis for a class of symmetric functions of A.

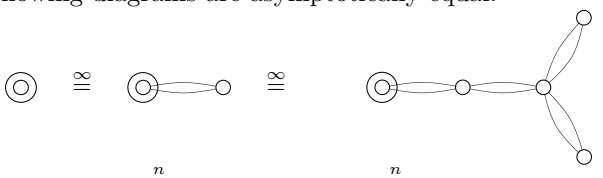
We represent the algorithmic state as a Fourier diagram expression of the form $x = \sum_{\alpha\in\mathcal{A}} c_\alpha Z_\alpha$. To multiply together or apply algorithmic operations on a diagram expression, we case on which indices repeat, like in the example in Section 3. See [40, Appendix A.2] for a formal derivation of these rules.

Now we turn to the asymptotic properties. The constant-size tree diagrams $(Z_\tau)_{\tau\in\mathcal{T}}$ exhibit the following key properties in the limit $n \to \infty$ and with respect to the randomness of A.

- **(i)** The coordinates of $Z_\tau \in \mathbb{R}^n$ for any $\tau \in \mathcal{T}$ are asymptotically independent and identically distributed.
- **(ii)** The random variables $Z_{\sigma,1}$ for $\sigma \in \mathcal{S}$ (the tree diagrams with one subtree) are asymptotically independent Gaussians with variance $|\mathrm{Aut}(\sigma)|$, where $\mathrm{Aut}(\sigma)$ are the graph automorphisms of σ which fix the root.
- **(iii)** The random variable $Z_{\tau,1}$ for $\tau \in \mathcal{T}$ (the tree diagrams with multiple subtrees) is asymptotically equal to the multivariate Hermite polynomial $\prod_{\sigma\in\mathcal{S}} h_{d_\sigma}(Z_{\sigma,1}; |\mathrm{Aut}(\sigma)|)$ where d_σ is the number of children of the root whose subtree (including the root) equals $\sigma \in \mathcal{S}$.

The remaining Fourier diagrams not in $\mathcal{T}$ can be understood using the further asymptotic properties:

- **(iv)** For any diagram $\alpha \in \mathcal{A}$, if α has a *hanging double edge* i.e. a double edge with one non-root endpoint of degree exactly 2, letting α_0 be the diagram with the hanging double edge and hanging vertex removed, then Z_α is asymptotically equal to Z_{α_0}. For example, the following diagrams are asymptotically equal:



$$1 \quad \approx \quad \sum_{\substack{j=1\\ i\neq j}}^{n} A_{ij}^2 \quad \approx \quad \sum_{\substack{j,k,\ell,m=1\\ i,j,k,\ell,m \text{ distinct}}}^{n} A_{ij}^2 A_{jk}^2 A_{k\ell}^2 A_{km}^2$$

- **(v)** For any *connected* $\alpha \in \mathcal{A}$, if removing the hanging trees of double edges from α creates a diagram in $\mathcal{T}$, then by the previous property, Z_α is asymptotically equal to that diagram. If the result is not in $\mathcal{T}$, then Z_α is asymptotically negligible.
- **(vi)** The disconnected diagrams have only a minor and negligible role in the algorithms that we consider. See [40, Section 4.2] for the description of these random variables.

To summarize the properties, given a sum x of connected diagrams, by removing the hanging double trees, and then removing all diagrams not in $\mathcal{T}$, the expression admits an *asymptotic* Fourier diagram basis representation of the form

$$x \overset{\infty}{=} \sum_{\tau \in \mathcal{T}} c_\tau Z_\tau \,, \tag{4}$$

for some coefficients $c_\tau \in \mathbb{R}$ independent of n and A. We call this the *tree approximation* to x. Note that all tree diagrams have order 1 variance regardless of their size, which can be counter-intuitive.

5 Asymptotic state evolution

The main appeal of the tree approximation in Equation 4 is that when restricted to the tree-shaped diagrams, the GFOM operations have a very simple interpretation: they implement an idealized *Gaussian dynamics* which we describe now.

The idealized GFOM moves through an "asymptotic Gaussian probability space" which is naturally the one corresponding to the $n \to \infty$ limit of the diagrams. Based on the diagram classification, this consists of an infinite family of independent Gaussian vectors $(Z_\sigma)_{\sigma \in \mathcal{S}}$. However, due to symmetry, all of the coordinates follow the same dynamic, so we can compress the representation of the dynamic down to a one-dimensional random variable X_t which is the asymptotic distribution of each coordinate $x_{t,i}$. We call X_t the *asymptotic state* of x_t.

For example, Approximate Message Passing (AMP) is a special type of GFOM whose iterates are asymptotically Gaussian i.e. X_t is a Gaussian random variable for all t (in general GFOMs, although X_t is defined in terms of Gaussians, it is not necessarily Gaussian).

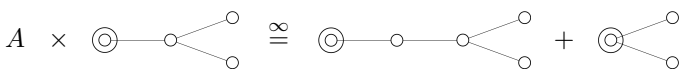
The algorithmic operations restricted to the trees and the corresponding evolution of the asymptotic state X_t are as follows. Two important operations on a tree-shaped diagram are extending/contracting the root by one edge.

▶ **Definition 13** (+ and − operators). *We define* $+ : \mathcal{T} \to \mathcal{S}$ *and* $- : \mathcal{S} \to \mathcal{T}$ *by:*

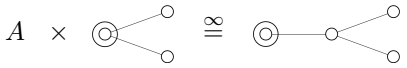
- *If* $\tau \in \mathcal{T}$*, let* τ^+ *be the diagram obtained by extending the root by one edge (i.e. adding one new vertex and one edge connecting it to the root of* τ*, and re-rooting* τ^+ *at this new vertex).*
- *If* $\tau \in \mathcal{S}$*, let* τ^- *be the diagram obtained by contracting the root by one edge (i.e. removing the root vertex and the unique edge from it, and re-rooting* τ^- *at the endpoint of that edge).*

Recall that the possible operations of a GFOM are either multiplying the state by A or applying a function componentwise. The effect of these two operations on the tree-shaped diagrams are:

- If $\sigma \in \mathcal{S}$, then AZ_σ is asymptotically the sum of the diagrams σ^+ and σ^- obtained by respectively extending and contracting the root by one edge. For example,



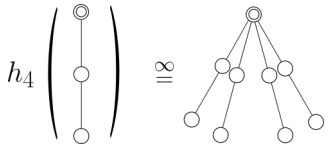
If $\tau \in \mathcal{T} \setminus \mathcal{S}$, then AZ_τ is asymptotically only the τ^+ term. For example,



- From the classification of diagrams, if $\tau \in \mathcal{T}$ consists of d_σ copies of $\sigma \in \mathcal{S}$, then

$$\prod_{\sigma \in \mathcal{S}} h_{d_\sigma}(Z_\sigma; |\mathrm{Aut}(\sigma)|) \overset{\infty}{=} Z_\tau \,. \tag{5}$$

Therefore, to compute $f(Z_\sigma : \sigma \in \mathcal{S})$, we expand f in the Hermite polynomial basis associated to $\mathcal{S}$, and apply the rule Equation 5 to all the terms. For example,



These operations correspond to the following Gaussian dynamic.

▶ **Definition 14** (Asymptotic Gaussian space, Ω). *Let $(Z_\sigma^\infty)_{\sigma \in \mathcal{S}}$ be a set of independent centered (one-dimensional) Gaussian random variables with variances $\mathrm{Var}(Z_\sigma^\infty) = |\mathrm{Aut}(\sigma)|$.*

If $\tau \in \mathcal{T}$ can be decomposed as d_σ copies of each $\sigma \in \mathcal{S}$ branching from the root, we define

$$Z_\tau^\infty = \prod_{\sigma \in \mathcal{S}} h_{d_\sigma}(Z_\sigma^\infty; |\mathrm{Aut}(\sigma)|) \,.$$

We call asymptotic states *the elements in the linear span of $(Z_\tau^\infty)_{\tau \in \mathcal{T}}$. We can view them both as polynomials in the formal variables $(Z_\sigma^\infty)_{\sigma \in \mathcal{S}}$ and as real-valued random variables. The set of asymptotic states is denoted Ω.*

▶ **Definition 15** (Asymptotic state). *If $x \in \mathbb{R}^n$ is such that $x \overset{\infty}{=} \sum_{\tau \in \mathcal{T}} c_\tau Z_\tau$, we define the* asymptotic state *of x by*

$$X = \sum_{\tau \in \mathcal{T}} c_\tau Z_\tau^\infty \,.$$

The state evolution of the algorithm can now be described concisely as:

- If x_t has asymptotic state X_t, then the asymptotic state of Ax_t is $X_t^+ + X_t^-$. Here we extend the $+$ and $-$ operators linearly to sums of Z_τ or Z_τ^∞ (let $Z_\tau^- = (Z_\tau^\infty)^- = 0$ if $\tau \in \mathcal{T} \setminus \mathcal{S}$).
- If $x_{t-1}, \ldots, x_0$ have asymptotic states $X_{t-1}, \ldots, X_0$ and f is any polynomial, then the asymptotic state of $f(x_{t-1}, \ldots, x_0)$ is $f(X_{t-1}, \ldots, X_0)$.

6 Perspective: equivariant Fourier analysis

The Fourier diagrams form an orthogonal basis that can be derived in a mechanical way using *symmetrization*.

We can use Fourier analysis to express a function or algorithm with respect to a natural basis. The unsymmetrized underlying analytical space consists of functions of the n^2 entries of A. Since the entries of A are independent, the associated Fourier basis is the product basis for the different entries. When $A \in \{-1, 1\}^{n \times n}$ is a Rademacher random matrix, the Fourier characters are the multilinear monomials in A. An arbitrary function $f : \{-1, 1\}^{n \times n} \to \mathbb{R}$ is then expressed as

$$f(A) = \sum_{\alpha \subseteq [n] \times [n]} c_\alpha \prod_{(i,j) \in \alpha} A_{ij} \,,$$

where c_α are the Fourier coefficients of f. When A is a symmetric matrix with zero diagonal, we only need Fourier characters for the top half of A, and the basis simplifies to $\alpha \subseteq \binom{[n]}{2}$. That is, the possible α can be interpreted combinatorially as graphs on the vertex set $[n]$.

An observation that allows us to significantly simplify the representation is that many of the Fourier coefficients are equal for S_n*-equivariant* algorithms. A function $f : \mathbb{R}^{n\times n} \to \mathbb{R}$ is S_n-equivariant if it satisfies $f(\pi(A)) = f(A)$ or if $f : \mathbb{R}^{n\times n} \to \mathbb{R}^n$ satisfies $f(\pi(A)) = \pi(f(A))$ where π acts on A by permuting the rows and columns simultaneously. For scalar-valued functions, considering the action of S_n on the vertex set of the Fourier characters $[n]$, any two Fourier characters α, β which are in the same orbit will have the same Fourier coefficient. Equivalently, if α and β are isomorphic as graphs, then their Fourier coefficients are the same. By grouping together all isomorphic Fourier characters, we obtain the symmetry-reduced representation defining the Fourier diagram basis,

$$f(A) = \sum_{\text{nonisomorphic } \alpha \subseteq \binom{[n]}{2}} c_\alpha \left(\sum_{\text{injective } \varphi : V(\alpha) \to [n]} \prod_{\{u,v\} \in \alpha} A_{\varphi(u)\varphi(v)} \right).$$

Thus by construction, the diagrams are an orthogonal basis for symmetric low-degree polynomials of A. We use this to derive some simple facts in [40, Appendix A.1]. Asymptotic independence of the Gaussian diagrams can be predicted based on the fact that the diagrams are an *orthogonal* basis, and orthogonal Gaussians are independent (thus we expect a set of independent Gaussians to appear from other types of i.i.d. inputs as well).

The above discussion was for Boolean matrices with $A_{ij} \sim \{\pm 1\}$. The natural generalization expresses polynomials in the basis of orthogonal polynomials for the entries A_{ij} (e.g. the Hermite polynomials when the $A_{ij} \sim \mathcal{N}(0, 1/n)$ [63, Section 3.2]).

Our results show that for the first-order algorithms we consider, only the multilinear part of the basis matters (i.e. the orthogonal polynomials which are degree 0 or 1 in each variable): up to negligible error, we can approximate $A_{ij}^2 \approx \frac{1}{n}$ and $A_{ij}^k \approx 0$ for $k \geq 3$. We use the monomial basis[7] to represent higher-degree polynomials instead of higher-degree orthogonal polynomials in order to simplify the presentation (except for the degree-2 orthogonal polynomial $A_{ij}^2 - \frac{1}{n}$ which expresses some error terms).

References

1. Kwangjun Ahn, Dhruv Medarametla, and Aaron Potechin. Graph matrices: Norm bounds and applications. *arXiv preprint arXiv:1604.03423*, 2020.
2. Ahmed El Alaoui and Andrea Montanari. Algorithmic Thresholds in Mean Field Spin Glasses. *arXiv preprint arXiv:2009.11481*, 2020.
3. Ahmed El Alaoui, Andrea Montanari, and Mark Sellke. Optimization of mean-field spin glasses. *Annals of Probability*, 49(6):2922–2960, 2021. doi:10.1214/21-AOP1519.
4. Antonio Auffinger, Andrea Montanari, and Eliran Subag. Optimization of Random High-Dimensional Functions: Structure and Algorithms. In *Spin Glass Theory and Far Beyond*, chapter 29, pages 609–633. World Scientific, 2023.
5. Boaz Barak, Samuel B. Hopkins, Jonathan A. Kelner, Pravesh K. Kothari, Ankur Moitra, and Aaron Potechin. A Nearly Tight Sum-of-Squares Lower Bound for the Planted Clique Problem. *SIAM Journal on Computing*, 48(2):687–735, 2019. doi:10.1137/17M1138236.

[7] The monomial "basis" is a misnomer in the cases when A_{ij} satisfies a polynomial identity such as $A_{ij}^2 = \frac{1}{n}$. In these cases, representation as a sum of diagrams is not unique. Our expressions should be interpreted as giving explicit sums of diagrams.

6 Mohsen Bayati, Marc Lelarge, and Andrea Montanari. Universality in polytope phase transitions and message passing algorithms. *Annals of Applied Probability*, 25(2):753–822, 2015.

7 Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011. doi:10.1109/TIT.2010.2094817.

8 Mohsen Bayati and Chandra Nair. A rigorous proof of the cavity method for counting matchings. In *Proceedings of the 44th Annual Allerton Conference on Communication, Control and Computing*, 2006.

9 Raphael Berthier, Andrea Montanari, and Phan-Minh Nguyen. State evolution for approximate message passing with non-separable functions. *Information and Inference: A Journal of the IMA*, 9(1):33–79, 2020.

10 Hans A. Bethe. Statistical theory of superlattices. *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, 150(871):552–575, 1935.

11 Erwin Bolthausen. An Iterative Construction of Solutions of the TAP Equations for the Sherrington–Kirkpatrick Model. *Communications in Mathematical Physics*, 325(1):333–366, 2014.

12 Charles Bordenave. Lecture notes on random matrix theory, 2019.

13 Collin Cademartori and Cynthia Rush. A non-asymptotic analysis of generalized approximate message passing algorithms with right rotationally invariant designs. *arXiv preprint arXiv:2302.00088*, 2023.

14 Michael Celentano, Chen Cheng, and Andrea Montanari. The high-dimensional asymptotics of first order methods with random data. *arXiv preprint arXiv:2112.07572*, 2021.

15 Michael Celentano, Andrea Montanari, and Yuchen Wu. The estimation error of general first order methods. In *Conference on Learning Theory, COLT 2020*, pages 1078–1141. PMLR, 2020. URL: http://proceedings.mlr.press/v125/celentano20a.html.

16 Patrick Charbonneau, Enzo Marinari, Giorgio Parisi, Federico Ricci-Tersenghi, Gabriele Sicuro, Francesco Zamponi, and Marc Mézard. *Spin Glass Theory and Far Beyond: Replica Symmetry Breaking after 40 Years*. World Scientific, 2023.

17 Wei-Kuo Chen and Wai-Kit Lam. Universality of approximate message passing algorithms. *Electronic Journal of Probability*, 26:1–44, 2021.

18 Amin Coja-Oghlan, Florent Krzakala, Will Perkins, and Lenka Zdeborová. Information-theoretic thresholds from the cavity method. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017*, pages 146–157. ACM, 2017. doi:10.1145/3055399.3055420.

19 Amir Dembo and Reza Gheissari. Diffusions interacting through a random matrix: universality via stochastic Taylor expansion. *Probability Theory and Related Fields*, 180:1057–1097, 2021.

20 Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *Conference on Learning Theory, COLT 2015*, pages 523–562. PMLR, 2015. URL: http://proceedings.mlr.press/v40/Deshpande15.html.

21 Jian Ding, Allan Sly, and Nike Sun. Satisfiability threshold for random regular NAE-SAT. In *Proceedings of the forty-sixth Annual ACM Symposium on Theory of Computing, STOC 2014*, pages 814–822, 2014. doi:10.1145/2591796.2591862.

22 Jian Ding, Allan Sly, and Nike Sun. Maximum independent sets on random regular graphs. *Acta Mathematica*, 217(2):263–340, 2016.

23 Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. *Annals of Mathematics*, 196(1):1–388, 2022.

24 Jian Ding and Nike Sun. Capacity lower bound for the Ising perceptron. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019*, pages 816–827, 2019. doi:10.1145/3313276.3316383.

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

26 David L. Donoho, Arian Maleki, and Andrea Montanari. Message passing algorithms for compressed sensing: I. motivation and construction. In *IEEE Information Theory Workshop on Information Theory, ITW 2010*, pages 1–5. IEEE, 2010.

27 Rishabh Dudeja, Yue M. Lu, and Subhabrata Sen. Universality of approximate message passing with semirandom matrices. *Annals of Probability*, 51(5):1616–1683, 2023.

28 Zhou Fan. Approximate message passing algorithms for rotationally invariant matrices. *Annals of Statistics*, 50(1):197–224, 2022.

29 Oliver Y. Feng, Ramji Venkataramanan, Cynthia Rush, and Richard J. Samworth. A Unifying Tutorial on Approximate Message Passing. *Foundations and Trends in Machine Learning*, 15(4):335–536, 2022. doi:10.1561/2200000092.

30 Noah Fleming, Pravesh Kothari, and Toniann Pitassi. Semialgebraic proofs and efficient algorithm design. *Foundations and Trends in Theoretical Computer Science*, 14(1-2):1–221, 2019. doi:10.1561/0400000086.

31 Marylou Gabrié. Mean-field inference methods for neural networks. *Journal of Physics A: Mathematical and Theoretical*, 53(22):223002, 2020.

32 Cédric Gerbelot and Raphaël Berthier. Graph-based approximate message passing iterations. *Information and Inference: A Journal of the IMA*, 12(4):2562–2628, 2023.

33 Cedric Gerbelot, Emanuele Troiani, Francesca Mignacco, Florent Krzakala, and Lenka Zdeborová. Rigorous dynamical mean field theory for stochastic gradient descent methods. *arXiv preprint arXiv:2210.06591*, 2022. doi:10.48550/arXiv.2210.06591.

34 Mrinalkanti Ghosh, Fernando Granha Jeronimo, Chris Jones, Aaron Potechin, and Goutham Rajendran. Sum-of-squares lower bounds for Sherrington-Kirkpatrick via planted affine planes. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020*, pages 954–965. IEEE, 2020. doi:10.1109/FOCS46700.2020.00093.

35 Samuel B. Hopkins, Pravesh K. Kothari, Aaron Potechin, Prasad Raghavendra, and Tselil Schramm. On the Integrality Gap of Degree-4 Sum of Squares for Planted Clique. *ACM Transactions on Algorithms*, 14(3):1–31, 2018. doi:10.1145/3178538.

36 Brice Huang. Capacity threshold for the Ising perceptron. *arXiv preprint arXiv:2404.18902*, 2024.

37 Brice Huang and Mark Sellke. Optimization Algorithms for Multi-Species Spherical Spin Glasses. *arXiv preprint arXiv:2308.09672*, 2023.

38 Misha Ivkov and Tselil Schramm. Semidefinite programs simulate approximate message passing robustly. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024*, pages 348–357. ACM, 2024. doi:10.1145/3618260.3649713.

39 Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2):115–144, 2013.

40 Chris Jones and Lucas Pesenti. Fourier Analysis of Iterative Algorithms. *arXiv preprint arXiv:2404.07881*, 2024.

41 Chris Jones and Aaron Potechin. Almost-orthogonal bases for inner product polynomials. In *Proceedings of the 13th Conference on Innovations in Theoretical Computer Science, ITCS 2022*, volume 215, pages 89:1–89:21, 2022. doi:10.4230/LIPICS.ITCS.2022.89.

42 Chris Jones, Aaron Potechin, Goutham Rajendran, Madhur Tulsiani, and Jeff Xu. Sum-of-Squares Lower Bounds for Sparse Independent Set. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021*, pages 406–416. IEEE, 2021. doi:10.1109/FOCS52979.2021.00048.

43 Chris Jones, Aaron Potechin, Goutham Rajendran, and Jeff Xu. Sum-of-Squares Lower Bounds for Densest k-Subgraph. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023*, pages 84–95. ACM, 2023. doi:10.1145/3564246.3585221.

44 Christopher Jones. *Symmetrized Fourier Analysis of Convex Relaxations for Combinatorial Optimization Problems*. PhD thesis, The University of Chicago, 2022.

45 Yoshiyuki Kabashima. A CDMA multiuser detection algorithm on the basis of belief propagation. *Journal of Physics A: Mathematical and General*, 36(43):11111, 2003.

46 Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT press, 2009.

47 Pravesh K. Kothari, Aaron Potechin, and Jeff Xu. Sum-of-Squares Lower Bounds for Independent Set on Ultra-Sparse Random Graphs. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024*, pages 1923–1934. ACM, 2024. doi:10.1145/3618260.3649703.

48 Dmitriy Kunisky, Cristopher Moore, and Alexander S. Wein. Tensor cumulants for statistical inference on invariant distributions. In *65th Annual Symposium on Foundations of Computer Science, FOCS 2024*, 2024.

49 Dmitriy Kunisky, Alexander S. Wein, and Afonso S. Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2019.

50 Gen Li, Wei Fan, and Yuting Wei. Approximate message passing from random initialization with applications to $\mathbb{Z}_2$-synchronization. *Proceedings of the National Academy of Sciences*, 120(31):e2302930120, 2023.

51 Gen Li and Yuting Wei. A non-asymptotic framework for approximate message passing in spiked models. *arXiv preprint arXiv:2208.03313*, 2022. doi:10.48550/arXiv.2208.03313.

52 Tengyuan Liang, Subhabrata Sen, and Pragya Sur. High-dimensional Asymptotics of Langevin Dynamics in Spiked Matrix Models. *Information and Inference: A Journal of the IMA*, 12(4):2720–2752, 2023.

53 Yue M. Lu. Householder Dice: A Matrix-Free Algorithm for Simulating Dynamics on Gaussian and Random Orthogonal Ensembles. *IEEE Transactions on Information Theory*, 67(12):8264–8272, 2021. doi:10.1109/TIT.2021.3114351.

54 Yanting Ma, Cynthia Rush, and Dror Baron. Analysis of approximate message passing with a class of non-separable denoisers. In *IEEE International Symposium on Information Theory, ISIT 2017*, pages 231–235. IEEE, 2017. doi:10.1109/ISIT.2017.8006524.

55 Olivier C. Martin, Rémi Monasson, and Riccardo Zecchina. Statistical mechanics methods and phase transitions in optimization problems. *Theoretical computer science*, 265(1-2):3–67, 2001. doi:10.1016/S0304-3975(01)00149-9.

56 Paul C. Martin, Eric D. Siggia, and Harvey A. Rose. Statistical dynamics of classical systems. *Physical Review A*, 8(1):423, 1973.

57 Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares Lower Bounds for Planted Clique. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing, STOC 2015*, pages 87–96, 2015. doi:10.1145/2746539.2746600.

58 Marc Mézard and Giorgio Parisi. The cavity method at zero temperature. *Journal of Statistical Physics*, 111:1–34, 2003.

59 Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. SK Model: The Replica Solution without Replicas. *Europhysics Letters*, 1(2):77, 1986. doi:10.1209/0295-5075/1/2/006.

60 Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific, 1987.

61 Ankur Moitra and Alexander S. Wein. Spectral methods from tensor networks. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019*, pages 926–937, 2019. doi:10.1145/3313276.3316357.

62 Andrea Montanari. Optimization of the Sherrington-Kirkpatrick Hamiltonian. In *60th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2019*, pages 1417–1433. IEEE, 2019. doi:10.1109/FOCS.2019.00087.

63 Andrea Montanari and Alexander S. Wein. Equivalence of approximate message passing and low-degree polynomials in rank-one matrix estimation. *arXiv preprint arXiv:2212.06996*, 2022.

64 Andrea Montanari and Yuchen Wu. Statistically optimal first order algorithms: A proof via orthogonalization. *arXiv preprint arXiv:2201.05101*, 2022.

65 Marc Mézard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, 2009.

66 Alexandru Nica and Roland Speicher. *Lectures on the Combinatorics of Free Probability*. Cambridge University Press, 2006.

67 Dmitry Panchenko. *The Sherrington–Kirkpatrick model*. Springer Science & Business Media, 2013.

68 Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754, 1979.

69 Giorgio Parisi. A sequence of approximated solutions to the SK model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115, 1980.

70 Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

71 Aaron Potechin and Goutham Rajendran. Machinery for Proving Sum-of-Squares Lower Bounds on Certification Problems. *arXiv preprint arXiv:2011.04253*, 2020. `arXiv:2011.04253`.

72 Aaron Potechin and Goutham Rajendran. Sub-exponential time Sum-of-Squares lower bounds for Principal Components Analysis. In *Advances in Neural Information Processing Systems, NeurIPS 2022*, volume 35, pages 35724–35740, 2022.

73 Prasad Raghavendra, Tselil Schramm, and David Steurer. High dimensional estimation via sum-of-squares proofs. In *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*, pages 3389–3423. World Scientific, 2018.

74 Goutham Rajendran and Madhur Tulsiani. Concentration of polynomial random matrices via Efron-Stein inequalities. In *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023*, pages 3614–3653. SIAM, 2023. `doi:10.1137/1.9781611977554.CH138`.

75 Cynthia Rush and Ramji Venkataramanan. Finite sample analysis of approximate message passing algorithms. *IEEE Transactions on Information Theory*, 64(11):7264–7286, 2018. `doi:10.1109/TIT.2018.2816681`.

76 Juspreet Singh Sandhu and Jonathan Shi. Sum-of-Squares & Gaussian Processes I: Certification. *arXiv preprint arXiv:2401.14383*, 2024.

77 Juspreet Singh Sandhu and Jonathan Shi. Sum-of-Squares & Gaussian Processes II: Rounding. *To appear*, 2024.

78 Keigo Takeuchi. Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements. *IEEE Transactions on Information Theory*, 66(1):368–386, 2019. `doi:10.1109/TIT.2019.2947058`.

79 Keigo Takeuchi. A unified framework of state evolution for message-passing algorithms. In *IEEE International Symposium on Information Theory, ISIT 2019*, pages 151–155. IEEE, 2019. `doi:10.1109/ISIT.2019.8849321`.

80 Keigo Takeuchi. Bayes-optimal convolutional AMP. *IEEE Transactions on Information Theory*, 67(7):4405–4428, 2021. `doi:10.1109/TIT.2021.3077471`.

81 Michel Talagrand. The Parisi formula. *Annals of Mathematics*, pages 221–263, 2006.

82 Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54. Springer Science & Business Media, 2010.

83 David J. Thouless, Philip W. Anderson, and Robert G. Palmer. Solution of 'Solvable model of a spin glass'. *Philosophical Magazine*, 35(3):593–601, 1977.

84 Tianhao Wang, Xinyi Zhong, and Zhou Fan. Universality of approximate message passing algorithms and tensor networks. *arXiv preprint arXiv:2206.13037*, 2022. `doi:10.48550/arXiv.2206.13037`.

85 Yuchen Wu and Kangjie Zhou. Lower Bounds for the Convergence of Tensor Power Iteration on Random Overcomplete Models. In *Conference on Learning Theory, COLT 2023*, volume 195, pages 3783–3820. PMLR, 2023. URL: `https://proceedings.mlr.press/v195/wu23b.html`.

86 Yuchen Wu and Kangjie Zhou. Sharp Analysis of Power Iteration for Tensor PCA. *arXiv preprint arXiv:2401.01047*, 2024. doi:10.48550/arXiv.2401.01047.

87 Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, 8(236-239):0018–9448, 2003.

88 Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.

89 Qiuyun Zou and Hongwen Yang. A Concise Tutorial on Approximate Message Passing. *arXiv preprint arXiv:2201.07487*, 2022. arXiv:2201.07487.