A unifying framework for mean-field theories of asymmetric kinetic Ising systems

Miguel Aguilera, S. Amin Moosavi & Hideaki Shimazaki

Kinetic Ising models are powerful tools for studying the non-equilibrium dynamics of complex systems. As their behavior is not tractable for large networks, many mean-field methods have been proposed for their analysis, each based on unique assumptions about the system’s temporal evolution. This disparity of approaches makes it challenging to systematically advance mean-field methods beyond previous contributions. Here, we propose a unifying framework for mean-field theories of asymmetric kinetic Ising systems from an information geometry perspective. The framework is built on Plefka expansions of a system around a simplified model obtained by an orthogonal projection to a sub-manifold of tractable probability distributions. This view not only unifies previous methods but also allows us to develop novel methods that, in contrast with traditional approaches, preserve the system’s correlations. We show that these new methods can outperform previous ones in predicting and assessing network properties near maximally fluctuating regimes.
Advances in high-throughput data acquisition technologies for very large biological and social systems are providing unprecedented possibilities to investigate their complex, non-equilibrium dynamics. For example, optical recordings from genetically modified neural populations make it possible to simultaneously monitor activities of the whole neural network of behaving C. elegans\textsuperscript{1} and zebrafish\textsuperscript{2}, as well as thousands of neurons in the mouse visual cortex\textsuperscript{3}. Such networks generally exhibit out-of-equilibrium dynamics\textsuperscript{4}, and are often found to self-organize near critical regimes at which their fluctuations are maximized\textsuperscript{5,6}. Evolution of such systems cannot be faithfully captured by methods assuming an asymptotic equilibrium state. Therefore, in general, there is a pressing demand for mathematical tools to study the dynamics of large-scale, non-equilibrium complex systems and to analyze high-dimensional datasets recorded from them.

The kinetic Ising model with asymmetric couplings is a prototypical model for studying such non-equilibrium dynamics in biological\textsuperscript{7,8} and social systems\textsuperscript{9}. It is described as a discrete-time Markov chain of interacting binary units, resembling the non-linear dynamics of recurrently connected neurons. The model exhibits non-equilibrium behavior when couplings are asymmetric or when model parameters are subject to rapid changes, ruling out quasi-static processes. These conditions induce a time reversal asymmetry in dynamical trajectories, leading to positive entropy production (the second law of thermodynamics) as revealed by the fluctuation theorem\textsuperscript{10–15} (see refs. 16,17 for reviews). This time-asymmetry is characteristic of non-equilibrium systems as it can only be displayed by systems in which energy dissipation takes place\textsuperscript{18}. In the case of symmetric connections and static parameters, the model converges to an equilibrium stationary state. Consequently, it is a generalization of its equilibrium counterpart known as the (equilibrium) Ising model\textsuperscript{19}.

The forward Ising problem refers to calculating statistical properties of the model, such as mean activation rates (mean magnetizations of spins) and correlations, given the parameters of the model. In contrast, inference of the model parameters from data is called the inverse Ising problem\textsuperscript{20}. In this regard, kinetic Ising models\textsuperscript{21,22} and their equilibrium counterparts\textsuperscript{23–25} have become popular tools for modeling and analyzing biological and social systems. In addition, they capture memory retrieval dynamics in classical associative networks. Namely, they are equivalent to the Boltzmann machine, extensively used in machine learning applications\textsuperscript{30}. Unfortunately, exact solutions of the forward and inverse problems often become computationally too expensive due to the combinatorial explosion of possible patterns in large, recurrent networks or the high volume of data, and applications of exact or sampling-based methods are limited in practice to a hundred of neurons\textsuperscript{5,25,26}. In consequence, analytical approximation methods are necessary for analysing large systems. In this endeavour, mean-field methods have emerged as powerful tools to track down otherwise intractable statistical quantities.

The standard mean-field approximations to study equilibrium Ising models are the classical naive mean-field (nMF) and the more accurate Thouless-Anderson-Palmer (TAP) approximations\textsuperscript{27}. These methods have also been employed to solve the inverse Ising problem\textsuperscript{28–31}. Plefka demonstrated that the nMF and TAP approximations for the equilibrium model can be derived using the power series expansion of the free energy around a model of independent spins, a method which is now referred to as the Plefka expansion\textsuperscript{32}. This expansion up to the first and second orders leads to the nMF and TAP mean-field approximations respectively. The Plefka expansion was later formalized by Tanaka and others in the framework of information geometry\textsuperscript{33–37}.

In non-equilibrium networks, however, the free energy is not directly defined, and thus it is not obvious how to apply the Plefka expansion. Kappen and Spanjers\textsuperscript{38} proposed an information geometric approach to mean-field solutions of the asymmetric Ising model with asynchronous dynamics. They showed that their second-order approximation for an asymmetric model in the stationary state is equivalent to the TAP approximation for equilibrium models. Later, Roudi and Hertz derived TAP equations for nonstationary states using a Legendre transformation of the generating functional of the set of trajectories of the model\textsuperscript{39}. Another study by Roudi and Hertz extended mean-field equations to provide expressions for the nonstationary delayed correlations assuming the presence of equal-time correlations at the previous step\textsuperscript{40}. Yet another interesting method proposed by Mézard and Sakellariou approximates the local fields by a Gaussian distribution according to the central limit theorem, yielding more accurate results for fully asymmetric networks\textsuperscript{41}. This method was later extended to include correlations at the previous time step, improving the results for symmetric couplings\textsuperscript{42}. More recently, Bachschmid-Romano et al. extended the path-integral methods in ref. 39 with Gaussian effective fields not only recovering ref. 41 for fully asymmetric networks but also proposing a method that better approximates mean rate dynamics by conserving autocorrelations of units. Although many choices of mean-field methods are available, the diversity of methods and assumptions makes it challenging to advance systematically over previous contributions.

Here, we propose a unified approach for mean-field approximations of the Ising model. While our method is applicable to symmetric and equilibrium models, we focus for generality on asymmetric kinetic Ising models. Our approach is defined as a family of Plefka expansions in an information geometric space. This approach allows us to unify and relate existing mean-field methods and to provide expressions for other statistics of the systems such as pairwise correlations. Furthermore, our approach can be extended beyond classical mean-field assumptions to propose novel approximations. Here, we introduce an approximation based on a pairwise model that better captures network correlations, and we show that it outperforms existing approximations of kinetic Ising models near a point of maximum fluctuations. We also provide a data-driven method to reconstruct and test if a system is near a phase transition by combining the forward and inverse Ising problems, and demonstrate that the proposed pairwise model more accurately estimates the system’s fluctuations and its sensitivity to parameter changes. These results confirm that our unified framework is a useful tool to develop methods to analyze large-scale, non-equilibrium biological and social dynamics operating near critical regimes. In addition, since the methods are directly applicable to Boltzmann machine learning, the geometrical framework introduced here is relevant in machine learning applications.

The paper is organized as follows. First, we introduce the kinetic Ising model and its statistical properties of interest. Second, we introduce our framework for the Plefka approximation methods from a geometric perspective. To explain how it works, we derive the classical naive and TAP mean-field approximations under the proposed framework. Third, we show that our approach can unify other known mean-field approximation methods. We then propose a novel pairwise approximation under this framework. Finally, we compare different mean-field approximations in solving the forward and inverse Ising problems, as well as in performing the data-driven assessment of the system’s sensitivity. The last section is devoted to discussion.
Results

The kinetic Ising model. The kinetic Ising model is the least structured statistical model containing delayed pairwise interactions between its binary components (i.e., a maximum caliber model\(^\text{44}\)). The system consists of \(N\) interacting binary variables (down or up of Ising spins or inactive or active of neural units) \(s_i,t \in \{-1, +1\}, i = 1, 2, \ldots, N\), evolving in discrete-time steps \(t\) with parallel dynamics. Given the configuration of spins at \(t - 1\), \(s_{i,t-1} = \{s_{1,t-1}, s_{2,t-1}, \ldots, s_{N,t-1}\}\), spins \(s_i\) at time \(t\) are conditionally independent random variables, updated as a discrete-time Markov chain, following

\[
P(s_i|s_{-i,t-1}) = \prod_j e^{h_{i,j}} e^{2 \cosh h_{i,j}},
\]

\[
h_{i,t} = H_i + \sum_j J_{ij} s_{j,t-1}.
\]

The parameters \(H = \{H_i\}\) and \(J = \{J_{ij}\}\) represent local external fields at each spin and couplings between pairs of spins respectively. When the couplings are asymmetric (i.e., \(J_{ij} \neq J_{ji}\)), the system is away from equilibrium because the process is irreversible with respect to time. Given the probability mass function of the previous state \(P(s_{-i,t-1})\), the distribution of the current state is:

\[
P(s_i) = \sum_{s_{-i,t-1}} P(s_i|s_{-i,t-1})P(s_{-i,t-1}).
\]

This marginal distribution \(P(s_i)\) is not factorized (except at \(J = 0\)), but it rather exhibits a complex statistical structure, generally containing higher-order spin interactions. We can apply this equation recursively, e.g., decomposing \(P(s_{-i,t-1})\) in terms of the distribution \(P(s_{-i,t-2})\), and trace the evolution of the system from the initial distribution \(P(s_0)\).

In this article, we use variants of the Plefka expansion to calculate some statistical properties of the system. Namely, we investigate the average activation rates \(m_{i,t}\) and delayed correlations \(D_{i,t}\) given by

\[
m_{i,t} = \sum_{s_i} s_i P(s_i),
\]

\[
C_{i,k,t} = \sum_{s_i,s_k} s_i s_k P(s_i) - m_{i,t} m_{k,t},
\]

\[
D_{i,k,t} = \sum_{s_i,s_k} s_i s_k P(s_i,s_k) - m_{i,t} m_{k,t-1}.
\]

Note that \(m_i\) and \(D_{i}\) are sufficient statistics of the kinetic Ising model. Therefore, we will use them in solving the inverse Ising problem (see Methods). We additionally consider the equal-time correlations \(C_i\) as they are commonly used to describe neural systems, and are investigated by some of the mean-field approximations in the literature\(^\text{40}\). Calculation of these expectation values is analytically intractable and computationally very expensive for large networks, due to the combinatorial explosion of the number of possible states. To reduce this computational cost, we approximate the marginal probability distributions (Eq. (3)) by the Plefka expansion method that utilizes an alternative, tractable distribution.

Geometrical approach to mean-field approximation. Information geometry\(^\text{37,45,46}\) provides clear geometrical understanding of information-theoretic measures and probabilistic models\(^\text{15,47,48}\). Using the language of information geometry, we introduce our method for mean-field approximations of kinetic Ising systems.

---

**Fig. 1** A geometric view of the approximations based on Plefka expansions. The point \(P(s_i)\) is the marginal distribution of a kinetic Ising model at time \(t\). The submanifold \(Q_i\) is a set of tractable distributions, for example a manifold of independent models. The points in \(A\) correspond to a m-geodesic, that is a linear mixture of \(P(s_i)\) and \(Q(s_i)\) on \(Q_i\), where for independent \(Q_i\) all points on \(A\) share the same mean values \(m_i\).

Geometrically, \(A\) constitutes the m-projection from \(P(s_i)\) to \(Q_i\), defining \(Q(s_i)\) as the closest point in the submanifold \(Q_i\) to the point \(P(s_i)\). The Plefka expansion is defined by expanding an \(\alpha\)-dependent distribution \(P_\alpha(s_i)\) that satisfies \(P_{\alpha=0}(s_i) = Q(s_i)\) and \(P_{\alpha=1}(s_i) = P(s_i)\).

Let \(\mathcal{P}_t\) be the manifold of probability distributions at time \(t\) obtained from Eq. (3). Each point on the manifold corresponds with a set of parameter values. The manifold \(\mathcal{P}_t\) contains submanifolds \(Q_t\) of probability distributions with analytically tractable statistical properties (See Fig. 1). We use this tractable manifold, i.e., a reference model, to approximate a target point \(P(s_i|H, J)\) in the manifold \(\mathcal{P}_t\) and its statistical properties \(m_i, C_i, D_i\).

The simplest submanifold \(Q\) is the manifold of independent models, used in classical mean-field approximations to compute average activation rates. Each point on this submanifold corresponds to a distribution

\[
Q(s_i|\Theta_i) = \prod_t e^{s_i^t \Theta_i^t} / 2 \cosh \Theta_i^t,
\]

where \(\Theta_i = \{\Theta_{ij}\}\) is the vector of parameters that represents a point in \(Q_i\). This distribution does not include couplings between units, and its average activation rate is immediately given as \(m_{i,t} = \tanh \Theta_{i,t}\).

Our first goal is to find the average activation rates of the target distribution \(P(s_i|H, J)\). It turns out that they can be obtained from the independent model \(Q(s_i|\Theta_i)\) that minimizes the following Kullback-Leibler (KL) divergence from \(P(s_i)\):

\[
D(P||Q) = \sum_{s_i} P(s_i) \log \frac{P(s_i)}{Q(s_i)}
\]

The independent model \(Q(s_i|\Theta_i)\) that minimizes the KL divergence has activation rates \(m_i\) identical to those of the target distribution \(P(s_i)\) because the minimizing points \(\Theta_{i,t}\) satisfy (for \(i = 1, \ldots, N\))

\[
\frac{\partial D(P||Q)}{\partial \Theta_{i,t}} |_{\Theta = \Theta^\star} = - \sum_{s_i} s_i P(s_i) + \tanh \Theta_{i,t}^\star = m_{i,t}^P - m_{i,t}^Q = 0,
\]

where \(m_{i,t}^P\) and \(m_{i,t}^Q\) are respectively expectation values of \(s_i\) by \(P\)
are obtained by minimizing manifold \( Q \) orthogonal to the -ed line. This is known as an \( m \)-geodesic, and it is orthogonal to the e-flat manifold \( \mathcal{Q}_i \), constituting an m-projection to this manifold\(\text{36,47} \). Thus, the previous search of \( Q(s_t, \Theta^* \text{)} \) given \( P(s_t, H, J) \) is equivalent to finding the orthogonal projection point from \( P(s_t, H, J) \) to the manifold \( \mathcal{Q}_i \) of independent models\(\text{36,37} \).

The Plefka expansion. Although the m-projection provides the exact and unique average activation rates, its calculation in practice requires the complete distribution \( P(s_t) \). In the Plefka expansion, we relax the constraints of the m-projection, and introduce another set of more tractable distributions that passes only through \( P(s_t, H, J) \) and \( Q(s_t, \Theta^* \text{)} \) (the solid line in Fig. 1). This distribution is defined using a new conditional distribution introducing a parameter \( \alpha \) that connects a distribution on the manifold \( \mathcal{Q}_i \) with the original distribution \( P(s_t) \):

\[
P_a(s_t | s_{t-1}) = \frac{e^{s_t h_t(a)}}{2 \cosh h_t(a)},
\]

\[
h_{i,t}(a) = (1 - \alpha) \Theta_{i,t} + \alpha \left( H_i + \sum_j I_{ij} s_{j,t-1} \right).
\]

At \( \alpha = 0 \), \( P_a(s_t | s_{t-1}) = Q(s_t, \Theta_0) \), and \( \alpha \rightarrow 1 \) leads to \( P_a(s_t | s_{t-1}) = P(s_t | s_{t-1}) \). Using this alternative conditional distribution \( P_a(s_t | s_{t-1}) \), we construct an approximate marginal distribution \( P_a(s_t) \). Consequently, expectation values with respect to \( P_a(s_t) \) are functions of \( \alpha \). We thus write the statistics of the approximate system as \( m_i(\alpha), C_i(\alpha), \) and \( D_i(\alpha) \).

The Plefka expansions of these statistics are defined as the Taylor series expansion of these functions around \( \alpha = 0 \). In the case of the mean activation rate, the expansion up to the \( n \)-th order leads to:

\[
m_i(\alpha) = m_i(\alpha = 0) + \sum_{k=1}^{n} \frac{\alpha^k \delta^k m_i(\alpha = 0)}{k!} + \mathcal{O}(\alpha^{n+1}),
\]

where \( \mathcal{O}(\alpha^{n+1}) \) stands for the residual error of the approximation of order \( n + 1 \) and higher. For the \( n \)-th order approximation, we neglect the residual terms as \( \mathcal{O}(\alpha^{n+1}) \forall n \approx 0 \). Note that all coefficients of expansion are functions of \( \Theta_0 \). The mean-field approximation is computed by setting \( \alpha = 1 \) and finding the value of \( \Theta^*_i \) that satisfies Eq. (12). Since the original marginal distribution is recovered at \( \alpha = 1 \), the equality of Eq. (9) holds:

\[
m_i(\alpha = 1) = m_i(\alpha = 0). \tag{13}
\]

which should be solved with respect to the parameters \( \Theta_0 \). Since we neglected the terms higher than the \( n \)-th order, the solution may not lead to the exact projection, \( Q(s_t, \Theta^* \text{)} \). In this study, we investigate the first \( (n = 1) \) and second \( (n = 2) \) order approximations. Moreover we can apply the same expansion to approximate the correlations \( C_i \) and \( D_i \) using Eq. (10).

What is the difference between this approach and other mean-field methods? Conventionally, naive mean-field approximations are obtained by minimizing \( D(Q|P) \) as opposed to \( D(P|Q) \) (Eq. (8))\(\text{36,49} \). This approach is typically used in variational inference to construct a tractable approximate posterior in machine learning problems. Following the Bogolyubov inequality, minimizing this divergence is equivalent to minimizing the variational free energy. Geometrically, it comprises an e-projection of \( P(s_t, H, J) \) to the submanifold \( \mathcal{Q}_i \), which does not result in \( Q(s_t, \Theta^* \text{)} \). Namely, minimizing \( D(Q|P) \), as well as minimization of other \( a \)-divergences except for \( D(P|Q) \), introduces a bias in the estimation of the mean-field approximation\(\text{36,37} \). In contrast, if we consider the m-projection point that minimizes \( D(P|Q) \), we can approximate the exact value of \( m_i \), using Eq. (12) up to an arbitrary order.

In the subsequent sections we show that different approximations of the marginal distribution \( P(s_t) \) in Eq. (3) can be constructed by replacing \( P(s_t | s_{t-1}) \) with \( P_a(s_t | s_{t-1}) \) for different pairs \( i, \tau \) (here we will explore the cases of \( \tau = t \) and \( \tau = t-1 \)). More generally, we show in Supplementary Note 1 that this framework can be extended to a marginal path of arbitrary length \( k \), \( P(s_{t-k+1}, \ldots, s_t) \). In addition, we are not restricted to manifolds of independent models. The independent model is adopted as a reference model to approximate the average activation rate, but one can also more accurately approximate correlations using this method. In this vein, we can extend our framework to use reference manifolds \( \mathcal{Q}_{i-k+1, \ldots, i} \) (of models \( Q(s_{t-k+1}, \ldots, s_t) \)) that include interactions, e.g., pairwise couplings between elements at two different time points, to more accurately approximate the delayed correlations (see Supplementary Note 1). By systematically defining these reference distributions, we will provide a unified approach to derive Plefka approximations of \( m_i, C_i, \) and \( D_i \), including the one that utilizes a pairwise structure.

Plefka\( [t-1, t] \): expansion around independent models at times \( t-1 \) and \( t \). Before elaborating different mean-field approximations, we demonstrate our method by deriving the known results of the classical nMF and TAP approximations for the kinetic Ising model\(\text{38,39} \). In order to derive these classical mean-field equations, we make a Plefka expansion around the points \( \Theta^*_i \) and \( \Theta^*_{i-1} \) that are, respectively, obtained by orthogonal projection to the independent manifolds \( \mathcal{Q}_i \) and \( \mathcal{Q}_{i-1} \), computed as in Eq. (9). Here we should note that assuming an approximation where previous distributions (e.g., \( t-2, t-3, \ldots \)) are also independent yields exactly the same result. In this way, we derive the nMF and TAP equations of a model defined by a marginal probability distribution \( p^{t-1}_a \). Using Eqs. (3) and (10), we write

\[
P^{t-1}_a(s_t) = \sum_{s_{t-2}} P_a(s_t | s_{t-1}) P_a(s_{t-1} | s_{t-2}) P(s_{t-2}), \tag{14}
\]

where \( p^{t-1}_a(s_t) = Q(s_t) \) and the original distribution is recovered for \( P_a^{t-1}(s_t) = P(s_t) \).

Following Eq. (13), for the first order approximation we have

\[
\frac{\partial m_i(\alpha = 0)}{\partial \alpha} = 0.
\]

Since the derivative of the first order moment is

\[
\frac{\partial m_i(\alpha = 0)}{\partial \alpha} = \left(1 - m^2_j \right) \left(-\Theta_{i,t} + H_i + \sum_j I_{ij} m_{j,t-1} \right),
\]

by solving the equation, we find \( \Theta^*_{i,t} \approx H_i + \sum_j I_{ij} m_{j,t-1} \) that leads to the naive mean-field approximation:

\[
m_{i,t} \approx \tanh \left[ H_i + \sum_j I_{ij} m_{j,t-1} \right]. \tag{16}
\]

We apply the same expansion to approximate the correlations, expanding \( C_{ik}(\alpha) \) and \( D_{ik}(\alpha) \) around \( \alpha = 0 \) up to the first order.
using $\Theta_{il} = \Theta^*_l$. Then we obtain
\begin{align}
C_{ik,t} &\approx 0, \quad i \neq k, \quad \text{(17)} \\
D_{il,t} &\approx J_{il}(1-m_i^2_t)(1-m_l^2_{t-1}). \quad \text{(18)}
\end{align}

Detailed calculations are presented in Supplementary Note 2.

To obtain the second-order approximation, we need to solve
\[ \frac{\partial^2 m_{il}(\alpha=0)}{\partial \alpha^2} + \frac{1}{2} \frac{\partial^2 m_{il}(\alpha=0)}{\partial \alpha^2} = 0 \] from Eq. (13). Here the second-order derivative is given as
\begin{equation}
\frac{\partial^2 m_{il}(\alpha=0)}{\partial \alpha^2} \approx -2m_{il}(1-m^2_{il}) \sum_j J_{ij}^2 (1-m^2_{l,j-1}), \quad \text{(19)}
\end{equation}
where terms of the order higher than quadratic were neglected (see Supplementary Note 2 for further details). From these equations, we find $\Theta^*_l \approx H_l + \sum J_{ij}m_{ij,t-1} - m_{il} \sum_j J_{ij}^2 (1-m^2_{l,j-1})$ leading to the TAP equation:
\begin{equation}
m_{il,t} \approx \tanh \left[ H_l + \sum_j J_{ij}m_{ij,t-1} - m_{il} \sum_j J_{ij}^2 (1-m^2_{l,j-1}) \right]. \quad \text{(20)}
\end{equation}

Having $\Theta^*_l$, we can incorporate TAP approximations of the correlations by expanding $C_{ik,t}(\alpha)$ and $D_{il,t}(\alpha)$ (see Supplementary Note 2 for details) as:
\begin{align}
C_{ik,t} &\approx (1-m^2_{il}) (1-m^2_{k,t}) \sum_j J_{ij}^2 (1-m^2_{l,j-1}), \quad i \neq k, \quad \text{(21)} \\
D_{il,t} &\approx J_{il}(1-m^2_{il}) (1-m^2_{l,t-1}) (1+2J_{il}m_{il,t-1}). \quad \text{(22)}
\end{align}

In these approximations, Eqs. (16) and (20) of activation rates $m_i$ correspond to the classical nMF and TAP equations of the kinetic Ising model. The mean-field equations for the equal- and delayed correlations (Eqs. (17), (18), (21), and (22)) are novel contributions from applying the Plefka expansion to correlations.

Using the equations above, we can compute the approximate statistical properties of the system at $t$ ($m_i, C_{lk}, D_{l}$) from $m_i^{-1}$. Therefore, the system evolution is described by recursively computing $m_i$ from an initial state $m_0$ (for both transient and stationary dynamics), although approximation errors accumulate over the iterations. After we introduce a unified view of mean-field approximations in the subsequent sections, we will numerically examine approximation errors of these various methods in predicting statistical structure of the system.

Generalization of mean-field approximations. In the previous section, we described a Plefka expansion that uses a model containing independent units at time $t-1$ and $t$ to construct a marginal probability distribution $P_a^{t-1,t}(s_t)$. This is, however, not the only possible choice of approximation. As we mentioned above, other approximations have been introduced in the literature. In ref. [40] expressions are provided for the nonstationary delayed correlations $D_l$ as a function of $C_{lk}$. In ref. [41] an approximation is derived by assuming that units at state $s_{t-1}$ are independent while correlations of $s_t$ are preserved.

In the following sections, we show that various approximation methods, including those mentioned above, can be unified as Plefka expansions. Each method of the approximation corresponds to a specific choice of the submanifold $Q_{\alpha}$ at each time step. Fig. 2 shows the corresponding submanifolds $Q_{\alpha}$ of possible approximations, where gray lines represent interactions that are affected by $\alpha$ in the Plefka expansion. The mean-field approximations in the previous section were obtained by using the model represented in Fig. 2B, where the couplings at time $t-1$ and $t$ are affected by $\alpha$. Below, we present systematic applications of the Plefka expansions around other reference models in order to approximate the original distribution (Fig. 2C–E). By doing so, we not only unify the previously reported mean-field approximations but also provide novel solutions that can provide more precise approximations than known methods.

Plefka[$\alpha$]: expansion around an independent model at time $t$. For the Plefka[$\alpha$] approximation, explained above, the system becomes independent for $\alpha = 0$ at $t$ as well as $t - 1$. This leads to approximations of $m_i, C_{lk}, D_l$, being specified by $m_{i-1}$, while being independent of $C_{lk-1}$ and $D_{l-1}$. In ref. [40], the authors describe a mean-field approximation by performing new expansion over the classical nMF and TAP equations that takes into account previous correlations $C_{lk}$. Here, our framework allows us to obtain similar results by considering only a Plefka expansion over manifold $Q_{\alpha}$, while assuming that we know the properties of $P(s_{t-1})$ (Fig. 2C). Therefore, we denote this approximation as $P_a^{t}$ and consider
\begin{equation}
P_a^{t}(s_t) = \sum_{s_{t-1}} P_a(s_t|s_{t-1})P(s_{t-1}). \quad \text{(23)}
\end{equation}

In Supplementary Note 3 we derive the equations for this approximation. For the first order, we obtain
\begin{align}
m_{il,t} \approx \tanh \left[ H_l + \sum_j J_{ij}m_{ij,t-1} \right], \quad \text{(24)} \\
C_{ik,t} &\approx 0, \quad i \neq k, \quad \text{(25)} \\
D_{il,t} &\approx (1-m^2_{il}) \sum_j J_{ij}C_{jl,t-1}. \quad \text{(26)}
\end{align}

Note that Eqs. (24) and (25) are the same as the nMF Plefka[$\alpha$] equations. Equation (26) includes $C_{lk,t}$, being exactly the same result obtained in ref. [40], Eq. (4). The second-order approximations lead to:
\begin{align}
m_{il,t} \approx \tanh \left[ H_l + \sum_j J_{ij}m_{ij,t-1} - m_{il} \sum_j J_{ij}C_{jl,t-1} \right], \quad \text{(27)} \\
C_{ik,t} &\approx (1-m^2_{il})(1-m^2_{k,t}) \sum_j J_{ij}^2 C_{jl,t-1}, \quad i \neq k, \quad \text{(28)} \\
D_{il,t} &\approx (1-m^2_{il}) \sum_j J_{ij}C_{jl,t-1}(1+2J_{il}m_{il,t-1}). \quad \text{(29)}
\end{align}

All update rules include the effect of $C_{lk}$. We can see that if we use the covariance matrix of the independent model at $t - 1$, we recover the results of the Plefka[$\alpha$] approximation in the previous section. In contrast with ref. [40], we provide a novel approximation method that depends on previous correlations using a single expansion (instead of two subsequent expansions), and additionally present approximated equal-time correlations.

Plefka[$\alpha$]: expansion around an independent model at time $t - 1$. In ref. [41], a mean-field method is proposed by approximating the effective field $h_i$ as the sum of a large number of independent spins, approximated by a Gaussian distribution, yielding exact results for fully asymmetric networks in the thermodynamic limit. In our framework, we describe this approximation as an expansion around the projection point from $P(s_{t-1})$ to the submanifold $Q_{\alpha}$, using a model where only $s_{t-1}$ are independent (Fig. 2D). In this case (see Supplementary Note 4), the effective field $h_i$ at the submanifold is a sum of independent terms, which for large $N$ yields a Gaussian distribution.
To make the expansions more accurately approximate target statistics, we can consider a reference manifold composed of multiple time steps while maintaining some of the parameters in the system (see Supplementary Note 1). Motivated by this idea, here we propose new methods that directly approximate pairwise activities of the units by choosing a reference manifold that preserves a coupling term. Let us first consider the joint probability of any arbitrary pair of units at time \( t - 1 \) and \( t \) to compute the delayed correlations (Fig. 2E, left). Namely, we consider the joint probability of spins \( s_{i,t} \) and \( s_{j,t-1} \):

\[
P(s_{i,t}, s_{j,t-1}) = \sum_{s_{k,-2}} P(s_{i,t-1} | s_{j,t-2}) P(s_{j,t-2}),
\]

with \( s_{k,-2} \) containing all elements of \( s_{-1} \) except \( s_{j,t-1} \). As a reference manifold \( Q_{t-1} \), we consider the dependency among only the units \( i \) and \( k \):

\[
Q(s_{i,t} | s_{j,t-1}) = Q(s_{i,t} | s_{j,t-1}) Q(s_{j,t-1})
\]

\[
= \frac{e^{\theta_{i}(s_{j,t-1})} e^{\theta_{j}(s_{i,t})}}{2 \cosh \theta_{i}(s_{j,t-1}) \cosh \theta_{j}(s_{i,t})}
\]

where \( \theta_{i}(s_{j,t-1}) = \Theta_{i} + \Delta_{ij} s_{j,t-1} \). The orthogonal projection to \( Q_{t} \) is equivalent to minimizing the KL divergence \( D(P||Q) \) with respect to the parameters:

\[
\frac{\partial D(P||Q)}{\partial \Theta_{t}} \bigg|_{\theta_{i}=\theta_{j}=0} = m_{Q_{t}}^{P} - m_{Q_{t}}^{P} = 0,
\]

\[
\frac{\partial D(P||Q)}{\partial \Delta_{ij}} \bigg|_{\theta_{i}=\theta_{j}=0} = \langle s_{i,t} s_{j,t-1} \rangle_{Q^{*}} - \langle s_{i,t} s_{j,t-1} \rangle_{P} = 0.
\]
with
\[
\langle x \rangle_{Q^P} = \sum_{s_{t-1}} x \cdot Q(s_{t-1}|s_{t-2}, s_{t-2}) P(s_{t-1}).
\]
(39)

As in the previous approximations, \(P(s_{t-1}|s_{t-2})\) is connected to \(Q(s_{t-1}|s_{t-2})\) through an \(\alpha\)-dependent probability
\[
P_a(s_{t-1}|s_{t-2}) = \sum_{s_{t-1}} P_a(s_{t-1}|s_{t-2}) P(s_{t-1}|s_{t-2}),
\]
(40)
with conditional probabilities given by
\[
P_a(s_{t-1}|s_{t-2}) = \frac{e^{\theta_a(s_{t-1})}}{2 \cosh h_{t-1}(\alpha)},
\]
(41)
\[
h_{t-1}(\alpha) = (1 - \alpha)\theta_{t-1} + \alpha \left( H_t + \sum_j J_{ij} \right),
\]
(42)
\[
\Theta_{t-1} \approx H_t + \sum_j J_{ij} m_{j-1} - \sum_{mn} J_{mn} l_{mn} l_{mn},
\]
(43)
which directly leads to calculation of means and delayed correlations as:
\[
m_{t-1} \approx \sum_{s_{t-1}} \tanh \theta_a(s_{t-1}) q(s_{t-1}),
\]
(45)
\[
m_{t-1} \approx \tanh \Theta_{t-1},
\]
(46)
\[
D_{t-1} \approx \sum_{s_{t-1}} \tanh \theta_a(s_{t-1}) q(s_{t-1}) - m_{t-1} m_{t-1}.
\]
(47)

These results are related to previous work\(^\text{43}\) that included autocorrelations as one of the constraints to derive the Plefka approximation. Instead, here we provide a Plefka approximation that includes delayed correlations between any pair of units.

To compute the above approximations, we need to know \(C_{t-1}\) and \(C_{t-2}\). Here, we provide similar pairwise Plefka approximations for the pairwise distribution at time \(t\), \(P(s_{t-1}, s_{t-2})\). Since \(s_{t-1}, s_{t-2}\) are conditionally independent, we can construct a model in which first \(s_{t-2}\) is computed from \(s_{t-1}\), and then \(s_{t-1}\) is computed conditioned on \(s_{t-2}\) (Fig. 2E, right):
\[
Q(s_{t-1}|s_{t-2}) = Q(s_{t-1}|s_{t-2}) Q(s_{t-1}),
\]
(48)
\[
P_a^{(\alpha)}(s_{t-1}|s_{t-2}) = \sum_{s_{t-1}} P_a(s_{t-1}|s_{t-2}) P(s_{t-1}) P(s_{t-1}),
\]
(49)
with conditional probabilities given by
\[
P_a(s_{t-1}|s_{t-2}, s_{t-2}) = \frac{e^{\theta_a(s_{t-1})}}{2 \cosh h_{t-1}(\alpha)}
\]
(50)
\[
h_{t-1}(\alpha) = (1 - \alpha)\Theta_{t-1} + \alpha \left( H_t + \sum_j J_{ij} \right),
\]
(51)
Here \(\theta_{t-1}\) is a function of \(s_{t-1}\) that accounts for equal-time correlations between \(s_{t-1}\) and \(s_{t-2}\). Computed similarly to delayed correlations, the second-order approximation yields (see Supplementary Note 5):
\[
\Theta_{t-1} \approx H_t + \sum_j J_{ij} m_{j-1} - \sum_{mn} J_{mn} l_{mn} l_{mn},
\]
(52)
\[
\Theta_{t-1} \approx H_t + \sum_j J_{ij} m_{j-1} - \sum_{mn} J_{mn} l_{mn} l_{mn},
\]
(53)

Using these expressions, approximate equal-time correlations are given as
\[
C_{t-1} \approx \sum_{s_{t-1}} \tanh \theta_a(s_{t-1}) q(s_{t-1}) - m_{t-1} m_{t-1}.
\]
(54)
Note that the approximation of equal-time correlations may not be symmetric for \(C_{t-1}\) and \(C_{t-1}\). In the results of this paper we use the average of the two.

**Comparison of the different approximations.** In the subsequent sections, we compare the family of Plefka approximation methods described above by testing their performance in the forward and inverse Ising problems. More specifically, we compare the second-order approximations of Plefka\([t - 1, t]\) and Plefka\([t]\), the first order approximation of Plefka\([t - 1]\), and the second-order pairwise approximation of Plefka\([t]\). We define an Ising model as an asymmetric version of the kinetic Sherrington-Kirkpatrick (SK) model, setting its parameters around the equivalent of a ferromagnetic phase transition in the equilibrium SK model. External fields \(H_t\) are sampled from independent uniform distributions \(U(-\beta H_0, \beta H_0)\), \(H_0 = 0.5\), whereas coupling terms \(J_{ij}\) are sampled from independent Gaussian distributions \(N(\beta J_0, \beta J_0)\), \(J_0 = 1\), \(\beta = 0.1\), where \(\beta\) is a scaling parameter (i.e., an inverse temperature).

Generally, mean-field methods are suitable for approximating properties of systems with small fluctuations. However, there is evidence that many biological systems operate in critical, highly fluctuating regimes\(^\text{56}\). In order to examine different approximations in such a biologically plausible yet challenging situation, we select the model parameters around a phase transition point displaying large fluctuations.

To find such conditions, we employed path-integral methods to solve the asymmetric SK model (Supplementary Note 6). We find that the stationary solution of the asymmetric model displays for our choice of parameters a non-equilibrium analogue of a critical point for a ferromagnetic phase transition, which takes place at \(\beta = 1.1108\) in thermodynamic limit (see Supplementary Note 6, Supplementary Fig. 1). The uniformly distributed bias
terms $H$ shift the phase transition point from $\beta = 1$ obtained at $H = 0$. By simulation of the finite size systems, we confirmed that the maximum fluctuations in the model are found near the theoretical $\beta_*$, which shows maximal covariance values (see Supplementary Note 6, Supplementary Fig. 2).

Fluctuations of a system are generally expected to be maximized at a critical phase transition. In addition, entropy production (a signature of time irreversibility) has been suggested as an indicator of phase transitions. For example, it presents a peak at the transition point of a continuous phase transition in a non-equilibrium Curie-Weiss Ising model with oscillatory fields and some instances of mean-field majority vote models. We found that the entropy production of the kinetic Ising system is also maximized around $\beta_*$ (discussed later, see also Methods for its derivation).

**Forward Ising problem.** We examine the performance of the different Plefka expansions in predicting the evolution of an asymmetric SK model of size $N = 512$ with random $H$ and $J$. To study the nonstationary transient dynamics of the model, we start from $s_0 = 1$ (all elements set to 1 at $t = 0$) and recursively update its state for $T = 128$ steps. We repeated this stochastic simulation for $R = 10^6$ trials for 21 values of $\beta$ in the range $[0.7\beta_*, 1.3\beta_*, \beta_*, \beta_*]$, which shows maximal covariance values (see Fig. 3G–I). The model parameters are estimated by the Boltzmann learning method under the maximum likelihood principle: $H$ and $J$ are updated to minimize the differences between the average rates $m_i$ or delayed correlations $D_i$ of the original data and the model approximations, which can significantly reduce computational time (see Methods). While Boltzmann learning requires to compute the likelihood of every point in a trajectory and every trial ($RT$ calculations) each iteration, we can estimate the gradient at each iteration in a one-shot computation by applying the Plefka approximations (Methods). At $\beta = \beta_*$, we observe that the classical Plefka approximation adds significant offset values to the fields $H$ and couplings $J$. In contrast, Plefka and Plefka2 approximations are all precise in estimating the values of $H$ and $J$.

Fig. 4C, D shows the mean squared error $\epsilon_H$, $\epsilon_J$ for bias terms and couplings between the original model and the inferred values for different $\beta$. In this case, errors are large in the estimation of $J$ for Plefka and $t = 1, t$ steps. In comparison, Plefka and Plefka2 work equally well even in the high fluctuation regime ($\beta = \beta_*$). Since the inverse Ising problem is solved by applying approximation one single time step (per iteration), it is not as challenging as the forward problem that can accumulate errors by recursively applying the approximations. Therefore, different approximations other than the classical mean-field Plefka perform equally well in this case.

**Phase transition reconstruction.** We have shown how different methods perform in computing the behavior of the system (forward problem) and inferring the parameters of a given network from its activation data (inverse problem). Combining the two, we can ask how well the methods explored here can reconstruct the behavior of a system from data, potentially exploring behaviors under different conditions than the recorded data.

First, in Fig. 5A–C we examine how the different approximation methods approximate fluctuations (equal-time and time-delayed covariances) and the entropy production (see Methods) at $t = 128$ after solving the forward problem by recursively applying the approximations for the 128 steps. As we mentioned above, the asymmetric SK model explored here presents maximum fluctuations and maximum entropy production around $\beta = \beta_*$ (Supplementary Note 6, Supplementary Fig. 2). However, we see that Plefka and Plefka2 cannot reproduce the behavior of correlations $C_i$ and $D_i$ of the original SK model around the transition point. Plefka and Plefka2 show much better performance in capturing the behavior of $C_i$ and $D_i$ in the phase transition, although Plefka overestimates both correlations. Additionally, all the methods capture the phase transition in entropy production, though Plefka2 overestimates its value around $\beta_*$ and Plefka2 is more precise than the other methods.

Next, we combine the forward and inverse Ising problem and try to reproduce the transition in the asymmetric SK model in the models inferred from the data. We first take the values of $H, J$ from solving the inverse problem from the data sampled at $\beta = \beta_*$ and next we solve again the forward problem with those estimated
parameters rescaled by a new inverse temperature $\tilde{\beta}$. The results for the correlations (Fig. 5D, E) show that in this case Plefka$[t-1]$ works badly, not being able to capture the transition. Plefka$[t-1]$ shows similar performances as in the forward problem, and Plefka$[t]$ and Plefka2$[t]$ have a similar behavior, underestimating fluctuations slightly. When we analyze entropy production of the system (Fig. 5F), we find that Plefka2$[t]$ exhibits better performance with a high precision, with Plefka$[t]$ underestimating it, and Plefka$[t-1, t]$ not capturing the phase transition. Overall, the results above suggest that Plefka2$[t]$ is preferable for these kinds of problems.
better suited to identify non-equilibrium phase transitions in models reconstructed from experimental data.

**Discussion**

We have proposed a framework that unifies different mean-field approximations of the evolving statistical properties of non-equilibrium Ising models. This allows us to derive approximations premised on specific assumptions about the correlation structure of the system previously proposed in the literature. Furthermore, using our framework we derive a new approximation (Plefka2[t]) using atypical assumptions for mean-field methods, i.e., the maintenance of pairwise correlations in the system. This new pairwise approximation outperforms existing ones for approximating the behavior of an asymmetric SK model near the non-equilibrium equivalent of a ferromagnetic phase transition (see Supplementary Note 6), where classical mean-field approximations face problems. This shows that the proposed transition (see Supplementary Note 6), where classical mean-field approximations face problems. This shows that the proposed transition (see Supplementary Note 6), where classical mean-field approximations face problems. This shows that the proposed transition (see Supplementary Note 6), where classical mean-field approximations face problems. This shows that the proposed transition (see Supplementary Note 6), where classical mean-field approximations face problems. This shows that the proposed transition (see Supplementary Note 6), where classical mean-field approximations face problems. This shows that the proposed transition (see Supplementary Note 6), where classical mean-field approximations face problems.

Augmenting accuracy by increasing parameters often involves a computational cost. As a practical guideline for using each method, Supplementary Note 7 compares their precision and computation time in the forward and inverse problems (see also Supplementary Figs. 3 and 4).

Asides from its theoretical implications, our unified framework offers analysis tools for diverse data-driven research fields. In neuroscience, it has been popular to study the activity of ensembles of neurons by inferring an equilibrium Ising model with homogeneous (fixed) parameters or inhomogeneous (time-dependent) parameters from empirical data. Extended analyses based on the equilibrium model have reported that neurons operate near a critical regime. However, studies of non-equilibrium dynamics in neural spike trains are scarce partly due to the lack of systematic methods for analysing large-scale non-equilibrium data from neurons exhibiting large fluctuations. The proposed pairwise model Plefka2[t] is suitable for simulating such network activities, being more accurate than previous methods in predicting the network evolution at criticality (Fig. 3) and in testing if the system is near the maximally fluctuating regime (Fig. 5). In particular, application of our methods for computing entropy production in non-equilibrium systems could provide tools for characterizing the non-equilibrium dynamics of neural systems.

In summary, a unified framework of mean-field theories offers a systematic way to construct suitable mean-field methods in accordance with the statistical properties of the systems researchers wish to uncover. This is expected to foster a variety of tools to analyze large-scale non-equilibrium systems in physical, biological, and social systems.

**Methods**

**Boltzmann learning in the inverse Ising problem.** Let $S_i = \{S_{i,t}, S_{i,t-1}, \ldots, S_{i,1}\}$ for $t = 1, \ldots, T$ be observed states of a process described by Eq. (1) at the $r$-th trial ($r = 1, \ldots, R$). We also define $r_{i,t}$ to represent the processes from all trials. The inverse Ising problem consists in inferring the external fields $H$ and couplings $J_j$ of the system. These parameters can be estimated by maximizing the log-likelihood $f(S_{i,1})$ of the observed states under the model:

$$f(S_{i,1}) = \log \prod_{r=1}^R \prod_{t=1}^{T} P(S_{i,t}|S_{i,t-1}) = \sum_{t=1}^{T} \sum_{j=1}^{N} \left( S_{i,j} J_{i,j} - \log \sinh 2 J_{i,j} \right).$$ (55)
with \( h_{ijr}^x = H_{ij} + \sum J_{ijs}^x \). The learning steps are obtained as:

\[
\frac{\partial (S_{i}^{\tau r})}{\partial H_i} = RT \langle S_{i}^{\tau r} \rangle_{r_{\tau}} - \langle \tanh h_{ijr}^x \rangle_{r_{\tau}},
\]

(56)

\[
\frac{\partial \langle S_{i}^{\tau r} \rangle}{\partial J_{ij}} = RT \langle S_{i}^{\tau r} S_{j}^{\tau r} \rangle_{r_{\tau}} - \langle \tanh h_{ijr}^x \rangle_{r_{\tau}}^2,
\]

(57)

where \( \langle \cdot \rangle \) denotes average over trials. We solve the inverse Ising problem by applying these equations as a gradient ascent rule adjusting the second terms of Eqs. (56) and (57) to obtain the best fit between the empirical and modeled data.

References

59 (https://doi.org/10.5281/zenodo.4357634).

The source code for implementing the methods and results in this work is available on Zenodo https://zenodo.org/record/4318985 (https://doi.org/10.5281/zenodo.4318985).

NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-021-20890-5

Data availability

The datasets generated and analysed in this study are available under CC BY license at Zenodo https://zenodo.org/record/4318985 (https://doi.org/10.5281/zenodo.4318985).

Code availability

The source code for implementing the methods and results in this work is available under GPL license at GitHub https://github.com/MiguelAguilera/kinetic-Plefka-expansions (https://doi.org/10.5281/zenodo.4357634).

Received: 19 May 2020; Accepted: 29 December 2020; Published online: 19 February 2021

References

1. Nguyen, J. P. et al. Whole-brain calcium imaging with cellular resolution in freely behaving Caenorhabditis elegans. Proc. Natl Acad. Sci. USA 113, E1074 (2016).

2. Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M. & Keller, P. J. Whole-brain functional imaging at cellular resolution using light-sheet microscopy. Nat. Methods 10, 413 (2013).

3. Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M. & Harris, K. D. High-dimensional geometry of population responses in visual cortex. Nature https://doi.org/10.1038/s41586-019-1346-5 (2019).

4. Nicolò, L. & Prigogine, I. Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations. 1st edn. (Wiley, New York, 1977).

5. Tkáčik, G. et al. Thermodynamics and signatures of criticality in a network of neurons. Proc. Natl Acad. Sci. USA 112, 11508 (2015).

6. Mora, T., Deny, S. & Marre, O. Dynamical criticality in the collective activity of a population of retinal neurons. Phys. Rev. Lett. 114, 078105 (2015).

7. Hertz, J., Roudi, Y. & Tyrcha, J. Ising model for inferring network structure from spike data. In Principles of Neural Coding, 527–546 (CRC Press, 2013).

8. Roudi, Y., Dunn, B. & Hertz, J. Multi-neuronal activity and functional connectivity in cell assemblies. Curr. Opin. Neurobiol. 32, 38 (2015).

9. Bouchaud, J. P. Crises and collective socio-economic phenomena: simple models and challenges. J. Stat. Phys. 151, 867 (2013).

10. Evans, D. J., Cohen, E. D. & Morris, G. P. Implications of second law violations in strong imaging states. Phys. Rev. Lett. 71, 2401 (1993).

11. Jarzynski, C. Nonequilibrium equality for free energy differences. Phys. Rev. Lett. 78, 2690 (1997).

12. Crooks, G. E. Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. J. Stat. Phys. 90, 1481 (1998).

13. Crooks, G. E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. Phys. Rev. E 60, 2721 (1999).

14. Lebowitz, J. L. & Spohn, H. A Gallavotti-Cohen-type symmetry in the large deviation functional for stochastic dynamics. J. Stat. Phys. 95, 333 (1999).

15. Crooks, G. E. Entropy production in the forward and reverse processes: a unified framework for the entropy production and the stochastic interaction based on information geometry. Phys. Rev. Res. 2, 033048 (2020).

16. Evans, D. J. & Searles, D. J. The fluctuation theorem. Adv. Phys. 51, 1529 (2002).

17. Seifert, U. Stochastic thermodynamics, fluctuation theorems and molecular machines. Rep. Prog. Phys. 75, 126001 (2012).

18. Gaspard, P. Time Asymmetry in Nonequilibrium Statistical Mechanics. In Special Volume in Memory of Ilya Prigogine, 83–133 (John Wiley, Sons, Ltd, 2007).

19. Salinas, S. R. A. The Ising Model. In Introduction to Statistical Physics, Graduate Texts in Contemporary Physics (ed. Salinas, S. R. A.) 257–276 (Springer New York, New York, NY, 2001).

20. Ackley, D. H., Hinton, G. E. & Sejnowski, T. J. A learning algorithm for Boltzmann machines. Cogn. Sci. 9, 147 (1985).

21. Witoelar, A. & Roudi, Y. Neural network reconstruction using kinetic Ising models with memory. BMC Neurosci. 12, P274 (2011).

22. Donner, C. & Opper, M. Inverse Ising problem in continuous time: a latent variable approach. Phys. Rev. E 96, 062104 (2017).

23. Schneiderman, E., Berry, M. J., Segev, R. & Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. Nature 440, 1007 (2006).

24. Cocco, S., Leibler, S. & Monasson, R. Neuronal couplings inferred by efficient inverse statistical physics methods. Proc. Natl Acad. Sci. USA 106, 14058 (2009).

25. Shimazaki, H., Amari, S.-i, Brown, E. N. & Grün, S. State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data. PLoS Comput. Biol. 8, e1002385 (2012).

26. Tyrcha, J., Roudi, Y., Marsili, M. & Hertz, J. The effect of nonstationarity on models inferred from neural data. J. Stat. Mech. 2013, P03005 (2013).

27. Thouless, D. J., Anderson, P. W. & Palmer, R. G. Solution of a spin glass. Philos. Mag. 35, 593 (1977).

28. Kappen, H. J. & Rodríguez, F. B. Efficient learning in Boltzmann machines using linear response theory. Neural Comput. 10, 1137 (1998).

29. Roudi, Y., Aurell, E. & Hertz, J. A. Statistical physics of pairwise probability models. Front. Comput. Neurosci. https://doi.org/10.3389/neuro.02.2009 (2009).

30. Roudi, Y., Tyrcha, J. & Hertz, J. Ising model for neural data: model quality and approximate methods for extracting functional connectivity. Phys. Rev. E 79, 051915 (2009).

31. Doerr, C., Obermayer, K. & Shimazaki, H. Approximate inference for time-varying interactions and macroscopic dynamics of neural populations. PLoS Comput. Biol. 13, e1005309 (2017).

32. Plefka, T. Convergence condition of the TAP equation for the infinite-range Ising spin glass model. J. Phys. A 15, 1971 (1982).

33. Tanaka, T. Mean-field theory of Boltzmann machine learning. Phys. Rev. E 58, 2302 (1998).

34. Tanaka, T. A theory of mean field approximation. In Advances in Neural Information Processing Systems, 351–357 (1999).

35. Bhattacharyya, C. & Keerthi, S. S. Information geometry and Plefkaas mean-field theory. J. Phys. A 33, 1507 (2000).

36. Hertz, J. Information Geometry of Mean-Field Approximation. In Advanced mean field method: Theory and practice 351–360 (MIT press, 2001).
37. Amari, S., Ikeda, S. & Shimokawa, H. Information Geometry of Alpha-Projection in Mean Field Approximation. In Advanced Mean Field Methods: Theory and Practice (MIT Press, 2001).
38. Kappen, H. J. & Spanjers, J. J. Mean field theory for asymmetric neural networks. Phys. Rev. E 61, 5658 (2000).
39. Roudi, Y. & Hertz, J. Dynamical TAP equations for non-equilibrium Ising spin glasses. J. Stat. Mech. 2011, P03031 (2011).
40. Zbaida, Y. & Hertz, J. Mean field theory for nonequilibrium network reconstruction. Phys. Rev. Lett. 106, 048702 (2011).
41. Mézard, M. & Sakellariou, J. Exact mean-field inference in asymmetric kinetic Ising systems. J. Stat. Mech. 2011, L07001 (2011).
42. Mahmoudi, H. & Saad, D. Generalized mean field approximation for parallel dynamics of the Ising model. J. Stat. Mech. 2014, P07001 (2014).
43. Bachschmid-Romano, L., Battistin, G., Oppo, M. & Roudi, Y. Variational perturbation and extended Pefkak approaches to dynamics on random networks: the case of the kinetic Ising model. J. Phys. A 49, 434003 (2016).
44. Press, S., Ghosh, K., Lee, J. & Dill, K. A. Principles of maximum entropy and maximum calibre in statistical physics. Rev. Mod. Phys. 85, 1115 (2013).
45. Amari, S. & Nagaoka, H. Methods of information geometry. Vol. 191 (American Mathematical Soc., 2007).
46. Amari, S. Information geometry and its applications, Vol. 194 (Springer, 2016).
47. Amari, S., Kurata, K. & Nagaoka, H. Information geometry of Boltzmann machines. IEEE Trans. Neural Netw. 3, 260 (1992).
48. Oizumi, M., Tsuchiya, N. & Amari, S.-I. Uniﬁcation of neural learning machines. Rev. E 72, 057103 (2005).
49. Noa, C. F., Harunari, P. E., de Oliveira, M. & Fiore, C. Entropy production as a tool for characterizing nonequilibrium phase transitions. Phys. Rev. E 100, 012104 (2019).
50. Zhang, Y. & Barato, A. C. Critical behavior of entropy production and equilibrium steady states of maximum entropy spike trains. Entropy 21, 884 (2019).
51. Cofre, R., Videla, L. & Rosas, F. An introduction to the nonequilibrium steady states of maximum entropy spike trains. Entropy 21, 884 (2019).
52. Lynn, C. W., Cornblath, E. J., Papadopoulos, L., Bertolero, M. A. & Bassett, D. S. Non-equilibrium dynamics and entropy production in the human brain. Preprint at arXiv: 2005.02526 (2020).
53. Schnakenberg, J. Network theory of microscopic and macroscopic behavior of master equation systems. Rev. Mod. Phys. 48, 571 (1976).
54. Aguilera, M. A unifying framework for mean field theories of asymmetric kinetic Ising systems [Dataset]. Zenodo https://doi.org/10.5281/zenodo.4318983 (2020).
55. Aguilera, M. A unifying framework for mean field theories of asymmetric kinetic Ising systems [Code]. GitHub https://doi.org/10.5281/zenodo.4357634 (2020).

Acknowledgements

We thank Yasser Roudi and Masanao Igarashi for valuable comments and discussions on this manuscript. This work was supported in part by the Cooperative Intelligence Joint Research between Kyoto University and Honda Research Institute Japan, MEXT/ISPS KAKENHI Grant Number JP 20K11709, and the grant of Joint Research by the National Institutes of Natural Sciences (NINS Program No. 0112005). M.A. was funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 892715 and the University of the Basque Country UPV/EHU post-doctoral training program grant Espidoc1717, and supported in part by the Basque Government project IT 1228-19 and project Outonomy PID2019-104576GB-I00 by the Spanish Ministry of Science and Innovation.

Author contributions

M.A., S.A.M., and H.S. designed and reviewed research; M.A. contributed analytical and numerical results; M.A., S.A.M., and H.S. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-20890-5.

Correspondence and requests for materials should be addressed to M.A.

Peer review information Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements

We thank Yasser Roudi and Masanao Igarashi for valuable comments and discussions on this manuscript. This work was supported in part by the Cooperative Intelligence Joint Research between Kyoto University and Honda Research Institute Japan, MEXT/ISPS KAKENHI Grant Number JP 20K11709, and the grant of Joint Research by the National Institutes of Natural Sciences (NINS Program No. 0112005). M.A. was funded by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 892715 and the University of the Basque Country UPV/EHU post-doctoral training program grant Espidoc1717, and supported in part by the Basque Government project IT 1228-19 and project Outonomy PID2019-104576GB-I00 by the Spanish Ministry of Science and Innovation.

Author contributions

M.A., S.A.M., and H.S. designed and reviewed research; M.A. contributed analytical and numerical results; M.A., S.A.M., and H.S. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-021-20890-5.

Correspondence and requests for materials should be addressed to M.A.

Peer review information Nature Communications thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.