Regularization and decimation pseudolikelihood approaches to statistical inference inXYspin models

Regularization and decimation pseudolikelihood approaches to statistical inference in XY spin models

Payal Tyagi,[1] Alessia Marruzzo,[1] Andrea Pagnani,[2,3] Fabrizio Antenucci,[1] and Luca Leuzzi[1,4]

[1]*CNR-NANOTEC, Institute of Nanotechnology, Rome - Soft and Living Matter Laboratory, Piazzale Aldo Moro 5, I-00185 Rome, Italy*

[2]*Department of Applied Science and Technology and Center for Computational Sciences, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, Italy*

[3]*Human Genetics Foundation - Torino, Via Nizza 52, Torino, Italy*

[4]*Dipartimento di Fisica, Università "Sapienza", Piazzale Aldo Moro 5, I-00185 Rome, Italy*

(Received 24 March 2016; revised manuscript received 21 June 2016; published 15 July 2016)

We implement a pseudolikelihood approach with l_1 and l_2 regularizations as well as the recently introduced pseudolikelihood with decimation procedure to the inverse problem in continuous spin models on arbitrary networks, with arbitrarily disordered couplings. Performances of the approaches are tested against data produced by Monte Carlo numerical simulations and compared also to previously studied fully connected mean-field-based inference techniques. The results clearly show that the best network reconstruction is obtained through the decimation scheme, which also allows us to make the inference down to lower temperature regimes. Possible applications to phasor models for light propagation in random media are proposed and discussed.

DOI: 10.1103/PhysRevB.94.024203

I. INTRODUCTION

Given a data set and a model with some unknown parameters, the inverse problem aims to find the values of the model parameters that best fit the data. In this work, in which we focus on systems of interacting elements, the inverse problem concerns the statistical inference of the underlying interaction network and of its coupling coefficients from observed data on the dynamics of the system.

Inverse problems have relevant applications in physics [1–6], biology [7–12], neuroscience (e.g., [13–16]), ethology [17–19], and social sciences and finance (e.g., [20,21]), just to cite a few examples, and are becoming more and more important due to the increase in the number of data available from these fields.

A standard approach used in statistical inference is to predict the interaction couplings by maximizing the likelihood function. This technique, however, requires the evaluation of the partition function that, in the most general case, concerns a number of computations scaling exponentially with the system size. Boltzmann machine learning uses Monte Carlo sampling to compute the gradients of the log-likelihood looking for stationary points [22] but this method is computationally manageable only for small systems. A series of faster approximations, such as naive mean-field, independent-pair approximation [14,23], inversion of TAP equations [24,25], small correlations expansion [26], adaptive TAP [27], adaptive cluster expansion [5], or Bethe approximations [3,4], have, then, been developed. These techniques take as input means and correlations of observed variables and most of them assume a fully connected graph as underlying connectivity network, or expand around it by perturbative dilution. In most cases, network reconstruction turns out to be not accurate for small data sizes and/or when couplings are strong or, else, if the original interaction network is sparse.

A further method, substantially improving performances for small data, is the so-called pseudolikelihood method (PLM) [28]. In Ref. [2] Aurell and Ekeberg performed a comparison between PLM and some of the just mentioned mean-field-based algorithms on the pairwise interacting Ising-spin ($\sigma = \pm 1$) model, showing how PLM performs sensitively better, especially on sparse graphs and in the high-coupling limit, i.e., for low temperature.

In this work, we aim at performing statistical inference on a model whose interacting variables are continuous XY spins, i.e., $\sigma \equiv (\cos\phi, \sin\phi)$ with $\phi \in [0, 2\pi)$. The developed tools can, actually, be also straightforwardly applied to the p-clock model [29], where the phase ϕ takes discretely equispaced p values in the 2π interval, $\phi_a = a2\pi/p$, with $a = 0, 1, \ldots, p-1$. The p-clock model, else called the vector Potts model, gives a hierarchy of discretization of the XY model as p increases. For $p = 2$, one recovers the Ising model, for $p = 4$ the Ashkin-Teller model [30], for $p = 6$ the ice-type model [31,32], and the eight-vertex model [33–35] for $p = 8$. It turns out to be very useful also for numerical implementations of the continuous XY model. Recent analysis on the multibody XY model has shown that for a limited number of discrete phase values ($p \sim 16, 32$) the thermodynamic critical properties of the $p \to \infty$ XY limit are promptly recovered [36,37]. Our main motivation to study statistical inference is that this kind of model has recently turned out to be rather useful in describing the behavior of optical systems, including standard mode-locking lasers [36,38–40] and random lasers [37,41–44]. In particular, the inverse problem on the pairwise XY model analyzed here might be of help in recovering images from light propagated through random media.

This paper is organized as follows: In Sec. II we introduce the general model. In Sec. III we introduce the PLM with l_1 and l_2 regularizations, variants of the PLM respectively introduced in Refs. [45] and [2] for the inverse Ising problem, and the PLM with decimation [46]. Here, we analyze these techniques for continuous XY spins and we test them on thermalized data generated by exchange Monte Carlo numerical simulations of the original model dynamics. In Sec. IV and in Sec. V we present the results related to the PLM-l_1 and PLM-l_2, respectively. In Sec. VI the results related to the PLM with

decimation are reported and its performances are compared to the PLM-l_1, to the PLM-l_2, and to a variational mean-field method analyzed in Ref. [47]. In Sec. VII we discuss the model derivation in the framework of light transmission through random scattering media: we show how light propagation through random media can be mapped onto an XY model and that the inverse problem can, thus, be exploited to improve performances in imaging and focusing. In Sec. VIII, we outline conclusive remarks and perspectives.

II. THE LEADING xy MODEL

The leading model we are considering is defined, for a system of N angular XY variables, by the Hamiltonian

$$\mathcal{H} = -\sum_{ik}^{1,N} J_{ik}\cos(\phi_i - \phi_k). \tag{1}$$

The XY model is well known in statistical mechanics, displaying important physical insights, starting from the Berezinskii-Kosterlitz-Thouless transition in two dimensions [48,49] and moving to, e.g., the transition of liquid helium to its superfluid state [50], the roughening transition of the interface of a crystal in equilibrium with its vapor [51]. In the presence of disorder and frustration [52,53] the model has been adopted to describe synchronization problems such as the Kuramoto model [54] and in the theoretical modeling of Josephson junction arrays [55,56] and arrays of coupled lasers [57]. Besides several derivations and implementations of the model in quantum and classical physics, equilibrium or out of equilibrium, ordered or fully frustrated systems, Eq. (1), in its generic form, has found applications also in other fields, a rather fascinating example being the behavior of starling flocks [58–62]. Our interest in the XY model resides, though, in optics. Phasor and phase models with pairwise and multibody interaction terms can, indeed, describe the behavior of electromagnetic modes in both linear and nonlinear optical systems in the analysis of problems such as light propagation and lasing [38,63,64]. As couplings are strongly frustrated, these models turn out to be especially useful to the study of optical properties in random media [43,44], as in the noticeable case of random lasers [65–67], and they might as well be applied to linear scattering problems, e.g., propagation of waves in opaque systems or disordered fibers, as will be discussed in Sec. VII.

III. PSEUDOLIKELIHOOD MAXIMIZATION

The inverse problem consists of the reconstruction of the parameters J_{ik} of the Hamiltonian, Eq. (1). Given a set of M data configurations of N spins $\sigma = \{\cos\phi_i^{(\mu)}, \sin\phi_i^{(\mu)}\}$, $i = 1,\dots,N$ and $\mu = 1,\dots,M$, we want to *infer* the couplings:

$$\sigma \to J.$$

With this purpose in mind, in the rest of this section we implement the working equations for the techniques used. In order to test our methods, we generate the input data, i.e., the configurations, by Monte Carlo simulations of the model. The joint probability distribution of the N variables $\boldsymbol{\phi} \equiv \{\phi_1,\dots,\phi_N\}$ follows the Gibbs-Boltzmann distribution:

$$P(\boldsymbol{\phi}) = \frac{1}{Z}e^{-\beta\mathcal{H}(\boldsymbol{\phi})} \quad \text{where} \quad Z = \int\prod_{k=1}^{N} d\phi_k e^{-\beta\mathcal{H}(\boldsymbol{\phi})} \tag{2}$$

and where we also introduced the inverse temperature parameter, which is an externally tuneable parameter. For instance, in Sec. VII, we will see that $\beta = (2\Delta^2)^{-1}$, where Δ is the mean-squared displacement of the noise in the exact resolution of light transmission equations. In order to stick to the usual statistical inference notation, in the following we will rescale the couplings by a factor $\beta/2$: $\beta J_{ij}/2 \to J_{ij}$.

Maximizing Eq. (2), or its logarithm, called likelihood, with respect to the couplings would yield the most representative set of interaction values. However, in Eq. (2) one has to evaluate the partition function Z; i.e., one has to perform the integral over all possible configuration of N continuous variables. Let us, for a moment, reduce the range of values of the phases from infinite to 2 ($\phi = 0,\pi$). Even in this case, i.e., the simplest—Ising—case, one should compute the Z function over 2^N possible configurations at each step of the maximization procedure. It becomes readily intractable as N increases for any statistical mechanical problem that has no simple exact solution because the number of configurations grows exponentially with N. Increasing, furthermore, the number of values each ϕ can take, $2 \to p \to \infty$, the problem becomes intractable at lower and lower N. One has, therefore, to resort to an approximate procedure that takes into account only a fraction of all possible configurations and that is proved to be exact as the number of configurations tends to the total number of configurations. This is termed the pseudolikelihood method (PLM).

The main idea of the PLM is to work with the conditional probability distribution of one variable ϕ_i given all other variables, $\boldsymbol{\phi}_{\setminus i}$:

$$\begin{aligned} P(\phi_i|\boldsymbol{\phi}_{\setminus i}) &= \frac{1}{Z_i}\exp\left\{H_i^x(\boldsymbol{\phi}_{\setminus i})\cos\phi_i + H_i^y(\boldsymbol{\phi}_{\setminus i})\sin\phi_i\right\} \\ &= \frac{e^{H_i(\boldsymbol{\phi}_{\setminus i})\cos[\phi_i - \alpha_i(\boldsymbol{\phi}_{\setminus i})]}}{2\pi I_0(H_i)}, \end{aligned} \tag{3}$$

where H_i^x and H_i^y are defined as

$$H_i^x(\boldsymbol{\phi}_{\setminus i}) = \sum_{j(\neq i)} J_{ij}^R\cos\phi_j - \sum_{j(\neq i)} J_{ij}^I\sin\phi_j, \tag{4}$$

$$H_i^y(\boldsymbol{\phi}_{\setminus i}) = \sum_{j(\neq i)} J_{ij}^R\sin\phi_j + \sum_{j(\neq i)} J_{ij}^I\cos\phi_j; \tag{5}$$

$H_i = \sqrt{(H_i^x)^2 + (H_i^y)^2}$, $\alpha_i = \arctan H_i^y/H_i^x$, and we introduced the modified Bessel function of the first kind:

$$I_k(x) = \frac{1}{2\pi}\int_0^{2\pi} d\phi e^{x\cos\phi}\cos k\phi.$$

Given M observation samples $\boldsymbol{\phi}^{(\mu)} = \{\phi_1^\mu,\dots,\phi_N^\mu\}$, $\mu = 1,\dots,M$, the pseudo-log-likelihood for the variable i is given

by the logarithm of Eq. (3),

$$L_i = \frac{1}{M}\sum_{\mu=1}^{M} \ln P\left(\phi_i^{(\mu)}\middle|\boldsymbol{\phi}_{\setminus i}^{(\mu)}\right)$$

$$= \frac{1}{M}\sum_{\mu=1}^{M}\left[H_i^{(\mu)}\cos\left(\phi_i^{(\mu)}-\alpha_i^{(\mu)}\right) - \ln 2\pi\, I_0\left(H_i^{(\mu)}\right)\right]. \quad (6)$$

The underlying idea of PLM is that an approximation of the true parameters of the model is obtained for values that maximize the functions L_i. The specific maximization scheme differentiates the different techniques.

A. PLM with $l_{1,2}$ regularization

In order to prevent that some parameters grow indefinitely in the iterative inference procedure without converging, it is useful to add a regularizer, which prevents the maximization routine to move towards high values of J_{ij} and h_i. We will adopt an l_1 and l_2 regularization so that the pseudolikelihood function (PLF) at site i reads

$$\mathcal{L}_i = L_i - \lambda_p(||J^R||_p + ||J^I||_p), \quad (7)$$

where

$$||J||_p \equiv \left(\sum_{i\neq j}|J_{ij}|^p\right)^{1/p}$$

is the p norm of the coupling matrix J, and $\lambda_p > 0$. In the following we will only deal with $p = 1,2$. Note that the values of λ have to be chosen arbitrarily, but not too large, in order not to overcome L_i. Indeed, λ_p must tend to zero as $M \to \infty$.

All l_p norms with $p \geqslant 1$ avoid arbitrarily large parameter values. However, the l_1 norm is peculiar, in the sense that it forces small parameters to become exactly zero, still preserving the convexity of the inference problem. From this point of view, if we consider a finite number of configurations M, the l_2 norm always infers a fully connected model even if data are generated by a sparse one, while the l_1 norm allows for the inference of sparse topologies. l_p regularizations are particularly important in the case of undersampling (small M), where the likelihood, albeit globally concave, might develop flat directions which induces some parameters to become large due to bad sampling of relatively rare events.

The standard implementation of the PLM consists of maximizing each $\mathcal{L}_i$, for $i = 1 \ldots N$, separately. The expected values of the couplings are then

$$\{J_{ij}^*\}_{j\in\partial i} := \arg\max_{\{J_{ij}\}}[\mathcal{L}_i]. \quad (8)$$

In this way, we obtain two estimates for the coupling J_{ij}, one from maximization of $\mathcal{L}_i$, $J_{ij}^{(i)}$, and another one from $\mathcal{L}_j$, say $J_{ji}^{(j)}$. Since the original Hamiltonian of the XY model is Hermitian, we know that the real part of the couplings is symmetric while the imaginary part is skew-symmetric. The final estimate for J_{ij} can, then, be obtained averaging the two results:

$$J_{ij}^{\text{inferred}} = \frac{J_{ij}^{(i)} + \bar{J}_{ji}^{(j)}}{2}, \quad (9)$$

where with $\bar{J}$ we indicate the complex conjugate. It is worth noting that the pseudolikelihood L_i, Eq. (6), is characterized by the following properties: (i) the normalization term of Eq. (3) can be computed analytically at odds with the *full* likelihood case that in general requires a computational time which scales exponentially with the size of the systems; (ii) the l-regularized pseudolikelihoods defined in Eq. (7) are strictly concave (i.e., they have a single maximizer) [28]; (iii) it is consistent; i.e., if M samples are generated by a model $P(\phi|J*)$ the maximizer tends to $J*$ for $M \to \infty$ [68]. Note also that (iii) guarantees that $|J_{ij}^{(i)} - J_{ij}^{(j)}| \to 0$ for $M \to \infty$. In Secs. IV, V, and VI we report the results obtained and we analyze the performances of the PLM having taken the configurations from Monte Carlo simulations of models whose details are known.

B. PLM with decimation

Even though the PLM with $l_{1,2}$ regularization allows to make the inference towards the low-temperature region and in the low-sampling case with better performances than mean-field methods, in some situations some couplings are overestimated and not at all symmetric. Moreover, in the technique there is the bias of the l regularizer. Trying to overcome these problems, Decelle and Ricci-Tersenghi introduced a new method [46], known as PLM+decimation: the algorithm maximizes the sum of the L_i,

$$\mathcal{L} \equiv \frac{1}{N}\sum_{i=1}^{N} L_i, \quad (10)$$

and, then, it recursively sets to zero couplings which are estimated very small. We expect that as long as we are setting to zero couplings that are unnecessary to fit the data, there should be not much changing on $\mathcal{L}$. Keeping on with decimation, a point is reached where $\mathcal{L}$ decreases abruptly indicating that relevant couplings are being decimated and underfitting is taking place. Let us define by x the fraction of nondecimated couplings. To have a quantitative measure for the halt criterion of the decimation process, a tilted $\mathcal{L}$ is defined as

$$\mathcal{L}_t \equiv \mathcal{L} - x\mathcal{L}_{\max} - (1-x)\mathcal{L}_{\min}, \quad (11)$$

where

(1) $\mathcal{L}_{\min}$ is the pseudolikelihood of a model with independent variables. In the XY case: $\mathcal{L}_{\min} = -\ln 2\pi$.

(2) $\mathcal{L}_{\max}$ is the pseudolikelihood in the fully connected model and it is maximized over all the $N(N-1)/2$ possible couplings.

At the first step, when $x = 1$, $\mathcal{L}$ takes value $\mathcal{L}_{\max}$ and $\mathcal{L}_t = 0$. On the last step, for a noninteracting graph, i.e., $x = 0$, $\mathcal{L}$ takes the value $\mathcal{L}_{\min}$ and, hence, again $\mathcal{L}_t = 0$. In the intermediate steps, during the decimation procedure, as x is decreasing from 1 to 0, one observes first that $\mathcal{L}_t$ increases linearly and, then, it displays an abrupt decrease indicating that from this point on relevant couplings are being decimated [46]. In Fig. 1 we give an instance of this behavior for the 2D short-range XY model with ordered couplings. We notice that the maximum point of $\mathcal{L}_t$ coincides with the minimum point of the reconstruction error, the latter defined as

$$\text{err}_J \equiv \sqrt{\frac{\sum_{i<j}\left(J_{ij}^{\text{inferred}} - J_{ij}^{\text{true}}\right)^2}{N(N-1)/2}}. \quad (12)$$

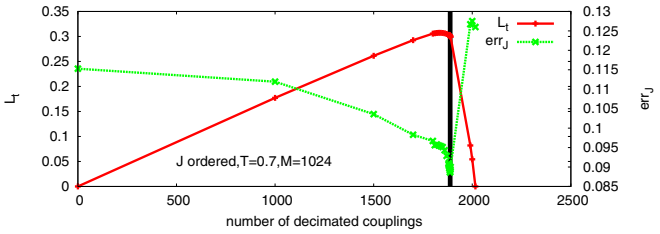

FIG. 1. The tilted likelihood $\mathcal{L}_t$ curve and the reconstruction error vs the number of decimated couplings for an ordered, real-valued J on the 2D XY model with $N = 64$ spins. The peak of $\mathcal{L}_t$ coincides with the dip of the error.

We stress that the $\mathcal{L}_t$ maximum is obtained ignoring the underlying graph, while the $\mathrm{err_J}$ minimum can be evaluated once the true graph has been reconstructed.

In the next sections we will show the results obtained on the XY model analyzing the performances of the regularization and of the decimation methods and comparing them also with a mean-field method [47].

IV. INFERRED COUPLINGS WITH PLM-l_1

In order to obtain the vector of couplings, $J_{ij}^{\mathrm{inferred}}$, the function $-\mathcal{L}_i$ is minimized through the vector of derivatives $\partial\mathcal{L}_i/\partial J_{ij}$. The process is repeated for all the couplings with an l_1 regularization obtaining then a fully connected adjacency matrix. The results here presented are obtained with $\lambda_1 = 0.02$ that turns out to be the optimal value to obtain a minimal reconstruction error and the best positive predictive value curve in a sensitive interval of temperature around the critical temperature also for a small number of data samples M. For the minimization we have used the MATLAB routine minFunc_2012 [69].

To produce the data by means of numerical Monte Carlo simulations a system with $N = 64$ spin variables is considered on a deterministic 2D lattice with periodic boundary conditions. Each spin has then connectivity 4; i.e., we expect to infer an adjacency matrix with $Nc = 256$ couplings different from zero. The dynamics of the simulated model is based on the Metropolis algorithm and parallel tempering [70] is used to speed up the thermalization of the system. The thermalization is tested looking at the average energy over logarithmic time windows and the acquisition of independent configurations starts only after the system is well thermalized.

For the values of the couplings we considered two cases: an ordered case, indicated in the figure as J ordered (e.g., left column of Fig. 2), where the couplings can take values $J_{ij} = 0, J$ (with $J = 1$), and a quenched disordered case, indicated in the figures as J disordered (e.g., right column of Fig. 2), where the couplings can also take negative values, i.e., $J_{ij} = 0, J, -J$, with a certain probability. The results here presented were obtained with bimodal distributed J's: $P(J_{ij} = J) = P(J_{ij} = -J) = 1/2$. The performances of the PLM appear not to depend on $P(J)$. We recall that in Sec. III we used the temperature-rescaled notation; i.e., J_{ij} stands for $J_{ij}/T/2$.

To analyze the performances of the PLM with l_1 regularization, in Fig. 2 the inferred couplings, J_{inf}^R, are shown on top of the original couplings, J_{true}^R. The top panels show the J_{inf}^R (red

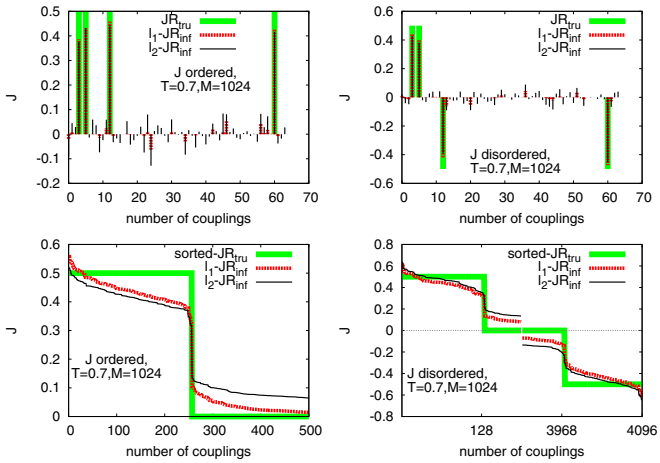

FIG. 2. Top panels: Instances of single-site-coupling reconstruction for the case of $N = 64$ XY spins on a 2D lattice with ordered J (left column) and bimodal distributed J (right column) for the PLM with l_1 and l_2 regularizations, with regularizers $\lambda_1 = 0.02$ and $\lambda_2 = 0.08$, respectively. Bottom panels: Sorted couplings.

dashed) and the J_{true}^R (green) for a given spin at temperature $T/J = 0.7$ and number of samples $M = 1024$. We display in the left column an instance for the ferromagnetic model and, in the right column, an instance of the bimodal frustrated couplings. In both cases PLM appears to reconstruct the correct couplings, though zero true couplings are sometimes given a small inferred nonzero value. In the bottom panels of Fig. 2, both the J_{inf}^R and the J_{true}^R are sorted in decreasing order and plotted on top of each other. We can clearly see that J_{inf}^R reproduces the expected step function. Even though the behavior above the step is not constant and the behavior below the step is not zero, the difference between inferred couplings corresponding to the set of nonzero couplings and inferred couplings corresponding to the set of zero couplings can be clearly distinguished and the number of existing couplings correctly inferred. In the bimodal distribution model, in particular, we note that the algorithm infers half positive and half negative couplings, as expected.

In order to analyze the effects of the number of samples and of the temperature regimes on the quality of the network reconstruction, we plot in Fig. 3 the reconstruction error, Eq. (12), as a function of temperature and data sample sizes. In the left panel we can appreciate that, as the data size is

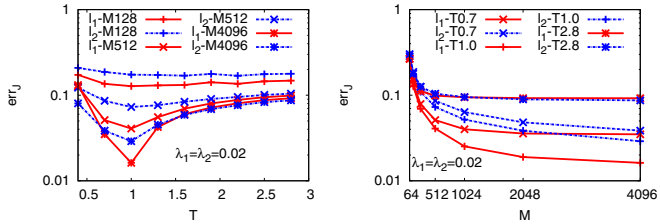

FIG. 3. Reconstruction error $\mathrm{err_J}$, cf. Eq. (12), plotted as a function of temperature (left) for three values of the number of samples M and as a function M (right) for three values of temperature in the ordered system, i.e., $J_{ij} = 0, 1$, for l_1 and l_2 regularizations with $\lambda_1 = 0.02$ and $\lambda_2 = 0.08$. The system size is $N = 64$.

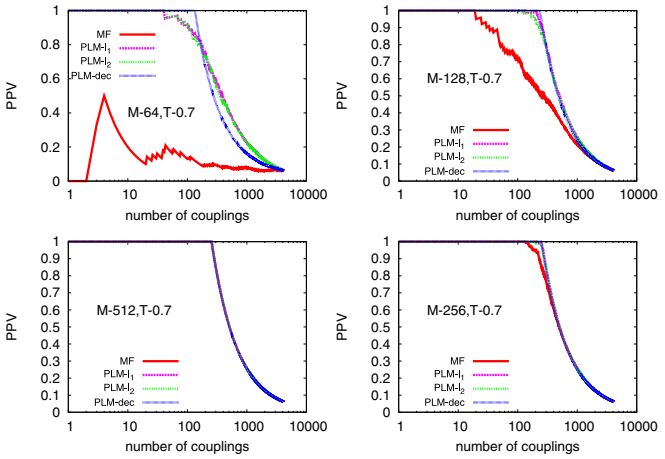

FIG. 4. Positive predictive value curves obtained with the four techniques: PLM with decimation, blue dotted line; PLM with l_1 regularization, magenta dashed line, $\lambda_1 = 0.02$; PLM with l_2 regularization, green dashed line, $\lambda_2 = 0.08$; and mean field, red full line. These results refer to real-valued ordered couplings with $N = 64$ spins on a 2D lattice. The temperature is here $T = 0.7$ while the four graphs refer to different sample sizes: $M = 64,128,256,512$, increasing clockwise.

large enough ($M \sim 512$), the error is seen to sharply rise at low temperature, up to one order of magnitude. Incidentally, in the ordered case, this increase occurs for $T < T_c \simeq 0.893$, which is the Kosterlitz-Thouless transition temperature of the $2XY$ model [71]. On the other hand, we can see that if only $M = 128$ samples are considered, $\mathrm{err_J}$ remains large almost independently from the working temperature. In the right plot of Fig. 3, $\mathrm{err_J}$ is plotted as a function of M for three different temperatures $T/J = 0.7, 1.0$, and 2.8. As we expect, $\mathrm{err_J}$ decreases as M increases, with approximately a $1/\sqrt{M}$ law [4] at $T = 1.0$. This decrease with M was observed also with mean-field inference techniques on the same model [47]. However, when T is too low, the reconstruction error decrease with M is very slow and for too high T, $\mathrm{err_J}$ stays larger than 0.1 even for $M = 4096$.

To better understand the performances of the algorithms, in Fig. 4 we show several positive predictive value (PPV) curves obtained for various values of M at $T = 0.7$. The positive predictive value curve displays how many times the inference method finds a true link of the original network as a function of the index of the vector of sorted absolute value of reconstructed couplings J_{ij}^{inf}. The index $n_{(ij)}$ represents the related spin couples (ij). The PPV curve is obtained as follows: first the values $|J_{ij}^{\mathrm{inf}}|$ are sorted in descending order and the spin pairs (ij) are ordered according to the sorting position of $|J_{ij}^{\mathrm{inf}}|$. Then, a cycle over the ordered set of pairs (ij), indexed by $n_{(ij)}$, is performed, comparing with the original network coupling J_{ij}^{true} and verifying whether it is zero or not. The positive predictive value is computed as

$$\mathrm{PPV}[n_{(ij)}] = \frac{\mathrm{PPV}[n_{(ij)} - 1](n_{ij} - 1) + 1 - \delta_{J_{ij}^{\mathrm{true}},0}}{n_{(ij)}}. \quad (13)$$

As far as $J_{ij}^{\mathrm{true}} \neq 0$, PPV $= 1$. As soon as the true coupling of a given (ij) couple in the sorted list is zero, the PPV curve departs from 1. In our case, where the connectivity per spin of the original system is $c = 4$ and there are $N = 64$ spins, we know that we will have 256 nonzero couplings. If the inverse problem is successful, hence, we expect a steep decrease of the PPV curve when $n_{ij} = 256$ is overcome. As M is large and/or temperature is not too far from the critical one, we are able to reconstruct correctly all the couplings present in the system (see bottom plots of Fig. 4). The quality of the positive predictive value for any M and T, though, depends on the technique employed. The results obtained with the PLM-l_1 is plotted in dashed magenta lines. As $M \geqslant 256$ the right network of $n = 64 \times 4 = 256$ couplings is recovered with this technique.

V. INFERRED COUPLINGS WITH PLM-l_2

A. XY model with real-valued couplings

We also carried out the l_2 regularization and compared its performances to the l_1. Even though the l_1 regularization is known to be better suited in the case of sparse graphs in the present study we have observed that minimization with l_2 regularization turns out to converge faster. As done for the previous case, in order to obtain the vector of couplings, $J_{ij}^{\mathrm{inferred}}$, the function $-\mathcal{L}_i$ is minimized through the vector of derivatives $\partial\mathcal{L}_i/\partial J_{ij}$. The results here presented are obtained with $\lambda_2 = 0.08$, which turns out to be optimal for what concerns PPV and $\mathrm{err_J}$, with respect to other values of λ_2. In Fig. 2, together with the results for l_1 regularization we show the inferred couplings obtained with the l_2 regularization and compare them to the couplings of the original network J_{true}^R. The top panels show the J_{inf}^R (thin black) and the J_{true}^R (green) for a given spin at temperature $T/J = 0.7$ and number of samples $M = 1024$. As in the case with the l_1, also the PLM with l_2 regularization appears to reconstruct the correct couplings, though zero couplings are always given an inferred nonzero value larger than in the l_1 case. In the bottom panels of Fig. 2, both the J_{inf}^R and the J_{true}^R are sorted in decreasing order and plotted on top of each other, and it can be observed that J_{inf}^R reproduces, in a smeared way, the expected step function and that it is not extremely worse that the l_1-inferred curve. Similarly, the plots in the right column of Fig. 2 show the results obtained for the case with bimodal disordered couplings, for the same working temperature and number of samples.

In Fig. 3 the reconstruction error, Eq. (12), is plotted as a function of T and M. For both l_1 and l_2 cases, the error presents a minimum at $T \simeq 1.0$ if M is not too small.

Comparing the two methods, the reconstruction error with the l_1 regularization turns out to be smaller than l_2 at all M when working around the *critical*-like temperature $T \sim 0.9$. In both regularizations, the larger the M, the smaller the error. However, $\mathrm{err_J}$ with l_1 regularization decreases more. For the other temperature regimes, the behaviors are qualitatively the same, including the fact that if only $M = 128$ samples are considered, $\mathrm{err_J}$ remains large independently of the working temperature; at $T = 2.8$ $\mathrm{err_J}$ tends asymptotically to a strictly nonzero (and comparably large) value in M; at lower $T = 0.7$ $\mathrm{err_J}$ also increases at any M, with respect to its minimal value.

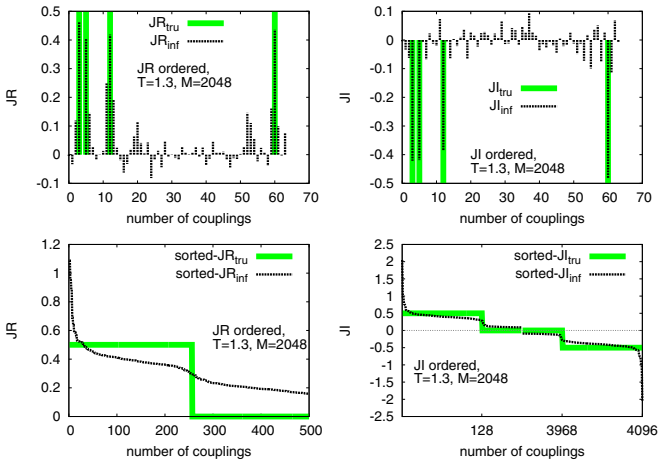

FIG. 5. Results related to the ordered complex XY model with $N = 64$ spins on a 2D lattice obtained with PLM and l_2 regularization. Top: Instances of single-site reconstruction for the real, JR (left column), and the imaginary, JI (right column), part of J_{ij}. Bottom: Sorted values of JR (left) and JI (right).

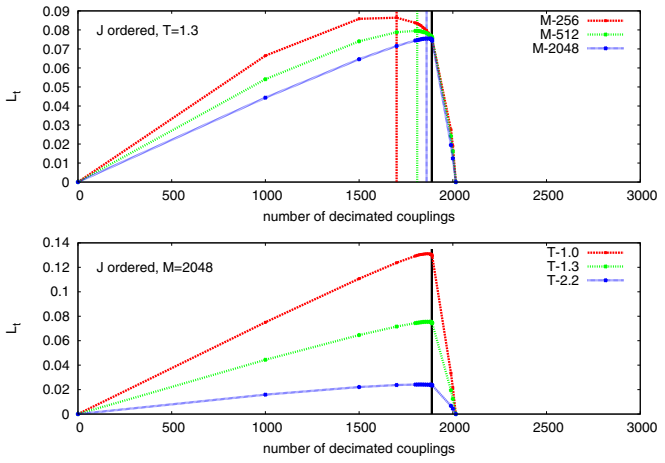

FIG. 6. Tilted pseudolikelihood, $\mathcal{L}_t$, plotted as a function of decimated couplings. Top: Different $\mathcal{L}_t$ curves obtained for different values of M plotted on top of each other. Here $T = 1.3$. The black line indicates the expected number of decimated couplings, $x^* = [N(N-1) - Nc]/2 = 1888$. As we can see, as M increases, the maximum point of $\mathcal{L}_t$ approaches x^*. Bottom: Different $\mathcal{L}_t$ curves obtained for different values of T with $M = 2048$. We can see that, with this value of M, no differences can be appreciated on the maximum points of the different $\mathcal{L}_t$ curves.

In Fig. 4 we show the PPV curves obtained for various values of M and T. As M is large ($M \geqslant 256$) we are able to reconstruct correctly all the couplings present in the system (see bottom plots), yielding a PPV equal to 1 up to the first $Nc = 256$ couplings. Results are plotted at $T = 0.7$, with increasing number of samples $M = 64, 128, 256$, and 512 (clockwise). The PPV score improves as M and results with the two regularizations being very similar.

B. XY model with complex-valued couplings

For the complex XY we have to contemporarily infer two apart coupling matrices, J^R_{ij} and J^I_{ij}. As before, a system of $N = 64$ spins is considered on a 2D lattice. For the couplings we have considered both ordered and bimodal disordered cases and the results shown are obtained with PLM with l_2 regularization. In Fig. 5, a single row of the matrix J (top) and the whole sorted couplings (bottom) are displayed for the ordered model (same legend as in Fig. 2) for the real, J^R (left column), and the imaginary part, J^I.

VI. PLM WITH DECIMATION

For the ordered real-valued XY model we show in Fig. 6, top panel, the outcome on the tilted pseudolikelihood, $\mathcal{L}_t$ Eq. (11), of the progressive decimation: from a fully connected lattice down to a noninteracting lattice. The figure shows the behavior of $\mathcal{L}_t$ for three different data sizes M. A clear data size dependence of the maximum point of $\mathcal{L}_t$, signaling the most likely value for decimation, is shown. For small M the most likely number of couplings is overestimated and for increasing M it tends to the true value, as displayed in Fig. 7. In the bottom panel of Fig. 6 we display instead different $\mathcal{L}_t$ curves obtained for three different values of T. Even though the values of $\mathcal{L}_t$ decrease with increasing temperature, the value of the most likely number of decimated couplings appears to be quite independent of T with $M = 2048$ number of samples. In Fig. 8 we eventually display the tilted pseudolikelihood for a 2D network with complex-valued ordered couplings, where the decimation of the real and imaginary coupling matrices proceeds in parallel, that is, when a real coupling is small enough to be decimated its imaginary part is also decimated, and vice versa. One can see that though the apart errors for the real and imaginary parts are different in absolute values, they display the same dip, to be compared with the maximum point of $\mathcal{L}_t$.

Once the most likely network has been identified through the decimation procedure, we perform the same analysis displayed in Fig. 9 for ordered and then quenched disordered real-valued couplings, and in Fig. 10 for complex-valued ordered couplings. In comparison to the results shown in Secs. IV and V, the PLM with decimation leads to rather cleaner results. In Figs. 4 and 11 we compare the performances

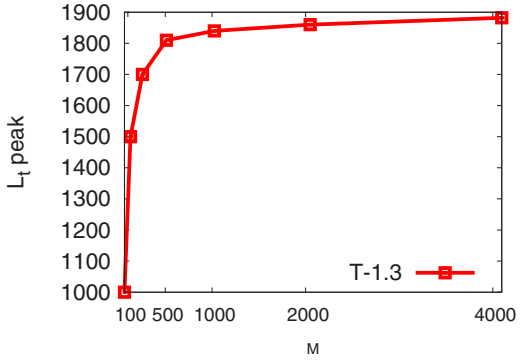

FIG. 7. Number of most likely decimated couplings, estimated by the maximum point of $\mathcal{L}_t$, as a function of the number of samples M. We can clearly see that the maximum point of $\mathcal{L}_t$ tends toward $x^* = 1888$, which is the correct expected number of zero couplings in the system.

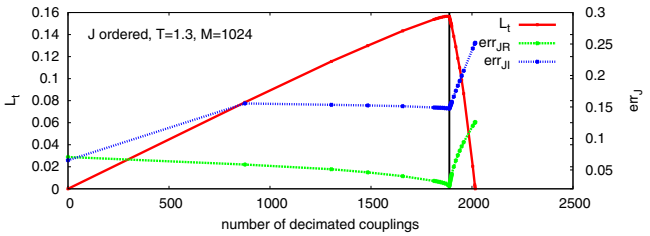

FIG. 8. Tilted pseudolikelihood, $\mathcal{L}_t$, plotted with the reconstruction errors for the XY model with $N = 64$ spins on a 2D lattice. These results refer to the case of ordered and complex-valued couplings. The full (red) line indicates $\mathcal{L}_t$. The dashed (green) and the dotted (blue) lines show the reconstruction errors [Eq. (12)] obtained for the real and the imaginary couplings, respectively. We can see that both err_{JR} and err_{JI} have a minimum at $x^* = 1888$.

of the PLM with decimation with respect to the ones of the PLM with l_1 and l_2 regularization. These techniques are compared to a mean-field technique previously implemented on the same XY systems [47].

For what concerns the network of connecting links, in Fig. 4 we compare the PPV curves. The results refer to the case of ordered and real-valued couplings, but similar behaviors were obtained for the other cases analyzed. The four graphs are related to different sample sizes, with M increasing clockwise. When M is high enough ($M = 8N = 512$), all techniques reproduce the true network. For lower values of M, instead, the performances of the various methods differ substantially. At any M and temperature, the PPV curves produced by PLM with l_1 and l_2 regularization and with decimation drastically overcome those yielded by means of the mean-field technique. When $M = N = 64$, however, the network is neatly reconstructed only through the PLM with decimation while the PLM with regularizations and the mean-field method fails to exactly reconstruct the network.

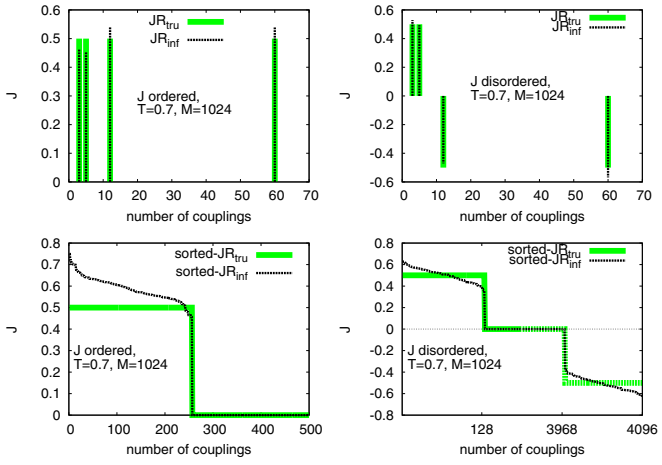

FIG. 9. XY model on a 2D lattice with $N = 64$ sites and real-valued couplings. The graphs show the inferred (dashed black lines) and true couplings (full green lines) plotted on top of each other. The left and right columns refer to the cases of ordered and bimodal disordered couplings, respectively. Top panels: Single-site reconstruction, i.e., one row of the matrix J. Bottom panels: Couplings are plotted sorted in descending order.

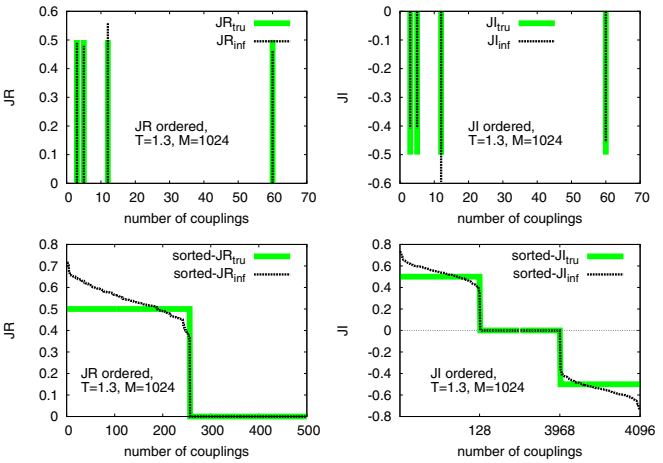

FIG. 10. XY model on a 2D lattice with $N = 64$ sites and ordered complex-valued couplings. The inferred and true couplings are plotted on top of each other. The left and right columns show the real and imaginary parts, respectively, of the couplings. Top panels refer to a single-site reconstruction, i.e., one row of the matrix J. Bottom panels report the couplings sorted in descending order.

Among the latter, the performances of l_1 and l_2 regularizations are comparable, and both of them are sensitively better than the mean-field PPV curve. For $M = 2N = 128$, the PLM with l_1 and l_2 regularization is almost optimal, almost equal to the PPV provided by decimation, yet better than mean-field PPV. For $M = 4N = 256$, the PLM with both regularizations and with decimation provides optimal PPV curves, as opposed to the mean-field approach. We further stress that the PLM method with decimation is able to clearly infer the network of interaction even when $M \sim N$ signaling that it could be considered a valid tool also in the undersampling regime. In Fig. 11 we compare the behavior of the reconstruction error. It can be observed that for all temperatures close to and above the critical-line temperature, $T \sim 0.9$, independently of the sample sizes, the reconstruction error, err_J (plotted here

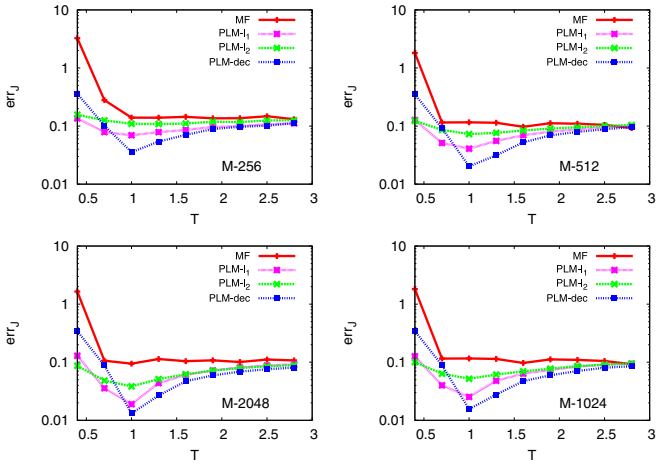

FIG. 11. Variation of reconstruction error, err_J, with respect to temperature as obtained with the four different techniques, see Fig. 4, for four different sample sizes: clockwise from top $M = 256, 512, 1024, 2048$.

in log scale), obtained with the PLM+decimation is always smaller than that one obtained with the other techniques. The temperature behavior of $\mathrm{err_J}$ agrees with the one already observed for Ising spins in [72] and for *XY* spins in [47] with a mean-field approach: $\mathrm{err_J}$ displays a minimum around $T \simeq 0.9$ and then it increases for lower T. However, the error obtained with the PLM with decimation is sensitively smaller than the error estimated by the other methods. Indeed, the only regime in which PLM+regularization methods work better is for the lowest temperatures, where, though, $\mathrm{err_J}$ turns out to be always larger than 0.1, i.e., one order of magnitude larger than the reconstruction error provided by the decimation in the critical region.

VII. A PROPAGATING WAVE MODEL

We briefly mention a derivation of the model as a proxy for the propagation of light through random linear media. This allows us to cast problems such as image reconstruction and focusing into inverse problems in statistical mechanics. Indeed, such problems necessarily require an estimate of the transmission matrix of the light through the medium and such a matrix turns out to play the role of an interaction matrix between modes in the statistical mechanical mapping, as will be clarified hereafter. Scattering of light is held responsible for obstructing our view and making objects opaque. Light rays, once that they enter the material, only exit after getting scattered multiple times within the material. In such a disordered medium, both the direction and the phase of the propagating waves are random. Transmitted light yields a disordered interference pattern typically having low intensity, random phase, and almost no resolution, called a speckle. Nevertheless, in recent years it has been realized that disorder is rather a blessing in disguise [73–75]. Several experiments have made it possible to control the behavior of light and other optical processes in a given random disordered medium, by exploiting, e.g., the tools developed for wave-front shaping to control the propagation of light and to engineer the confinement of light [76,77]. Applying the inverse problem techniques to a faithful statistical mechanical representation of light propagating through random media would contribute to increase the quality of image reconstruction and focusing. Indeed, in optical problems, many data acquisition techniques are easily available and effective techniques are continuously developed to reconstruct images from the output light pattern across a random medium once the transmission matrix is known. The bottleneck is, though, the reconstruction of the effective transmission matrix of the random medium [78–81].

In a linear dielectric medium, light propagation can be described through a part of the scattering matrix, the transmission matrix $\mathbb{T}$, linking the outgoing to the incoming fields. Consider the case in which there are N_I incoming channels and N_O outgoing ones; we can indicate with $E_k^{\mathrm{in,out}}$ the input/output electromagnetic field phasors of channel k. In the most general case, i.e., without making any particular assumptions on the field polarizations, each light mode and its polarization state can be represented by means of the 4-dimensional Stokes vector. Each t_{ki} element of $\mathbb{T}$, thus, is a 4×4 Müller matrix. If, on the other hand, we know that the source is polarized and the observation is made on the same polarization, one can use a scalar model and adopt the Jones calculus [78,79,82]:

$$E_k^{\mathrm{out}} = \sum_{i=1}^{N_I} t_{ki} E_i^{\mathrm{in}}, \quad \forall k = 1, \ldots, N_O. \tag{14}$$

We recall that the elements of the transmission matrix are random complex coefficients [78]. For the case of completely unpolarized modes, we can also use a scalar model similar to Eq. (14), but whose variables are the intensities $I_k^{\mathrm{in,out}} = |E_k^{\mathrm{in,out}}|^2$ of the outgoing/incoming fields, rather than the fields themselves.

In the following, for simplicity, we will consider Eq. (14) as our starting point, where E_k^{out}, E_i^{in}, and t_{ki} are all complex scalars. If Eq. (14) holds for any k, we can write

$$\int \prod_{k=1}^{N_O} dE_k^{\mathrm{out}} \prod_{k=1}^{N_O} \delta\left(E_k^{\mathrm{out}} - \sum_{j=1}^{N_I} t_{kj} E_j^{\mathrm{in}} \right) = 1. \tag{15}$$

Observed data are a noisy representation of the true values of the fields. Therefore, in inference problems it is statistically more meaningful to take that noise into account in a probabilistic way, rather than looking at the precise solutions of the exact equations (whose parameters are unknown). To this aim we can introduce Gaussian distributions whose limits for zero variance are the Dirac deltas in Eq. (15). Moreover, we move to consider the ensemble of all possible solutions of Eq. (14) at given $\mathbb{T}$, looking at all configurations of input fields. We, thus, define the function

$$Z \equiv \int_{\mathcal{S}_{\mathrm{in}}} \prod_{j=1}^{N_I} dE_j^{\mathrm{in}} \int_{\mathcal{S}_{\mathrm{out}}} \prod_{k=1}^{N_O} dE_k^{\mathrm{out}} \times \prod_{k=1}^{N_O} \frac{1}{\sqrt{2\pi\Delta^2}} \exp\left\{ -\frac{1}{2\Delta^2} \left| E_k^{\mathrm{out}} - \sum_{j=1}^{N_I} t_{kj} E_j^{\mathrm{in}} \right|^2 \right\}. \tag{16}$$

We stress that the integral of Eq. (16) is not exactly a Gaussian integral. Indeed, starting from Eq. (15), two constraints on the electromagnetic field intensities must be taken into account.

The space of solutions is delimited by the total power $\mathcal{P}$ received by system, i.e., $\mathcal{S}_{\mathrm{in}} : \{E^{\mathrm{in}} | \sum_k I_k^{\mathrm{in}} = \mathcal{P}\}$, also implying a constraint on the total amount of energy that is transmitted through the medium, i.e., $\mathcal{S}_{\mathrm{out}} : \{E^{\mathrm{out}} | \sum_k I_k^{\mathrm{out}} = c\mathcal{P}\}$, where the attenuation factor $c < 1$ accounts for total losses. As we will see more in details in the following, being interested in inferring the transmission matrix through the PLM, we can omit to explicitly include these terms since they do not depend on $\mathbb{T}$ not adding any information on the gradients with respect to the elements of $\mathbb{T}$.

Taking the same number of incoming and outgoing channels, $N_I = N_O = N/2$, and ordering the input fields in the first $N/2$ mode indices and the output fields in the last $N/2$ indices, we can drop the "in" and "out" superscripts and formally write

Z as a partition function

$$Z = \int_{\mathcal{S}} \prod_{j=1}^{N} dE_j \left(\frac{1}{\sqrt{2\pi\,\Delta^2}} \right)^{N/2} \exp\left\{ -\frac{\mathcal{H}[\{E\};\mathbb{T}]}{2\Delta^2} \right\}, \quad (17)$$

$$\begin{aligned} \mathcal{H}[\{E\};\mathbb{T}] &= -\sum_{k=1}^{N/2} \sum_{j=N/2+1}^{N} [E_j^* t_{jk} E_k + E_j t_{kj}^* E_k^*] \\ &\quad + \sum_{j=N/2+1}^{N} |E_j|^2 + \sum_{k,l}^{1,N/2} E_k U_{kl} E_l^* \\ &= -\sum_{nm}^{1,N} E_n J_{nm} E_m^*, \end{aligned} \quad (18)$$

where $\mathcal{H}$ is a real-valued function by construction, we have introduced the effective input-input coupling matrix

$$U_{kl} \equiv \sum_{j=N/2+1}^{N} t_{lj}^* t_{jk}, \quad (19)$$

and the whole interaction matrix reads ($\mathbb{T} \equiv \{t_{jk}\}$)

$$\mathbb{J} \equiv \left(\begin{array}{c|c} -\mathbb{U} & \mathbb{T} \\ \hline \mathbb{T}^\dagger & -\mathbb{I} \end{array} \right). \quad (20)$$

Determining the electromagnetic complex amplitude configurations that minimize the *cost function* $\mathcal{H}$, Eq. (18), means to maximize the overall distribution peaked around the solutions of the transmission Eq. (14). As the variance $\Delta^2 \to 0$, eventually, the initial set of Eq. (14) are recovered. The $\mathcal{H}$ function, thus, plays the role of a Hamiltonian and Δ^2 the role of a noise-inducing temperature. The exact numerical problem corresponds to the zero-temperature limit of the statistical mechanical problem. Working with real data, though, which are noisy, with a finite statistical uncertainty, a finite "temperature" allows for a better representation of the ensemble of solutions to the sets of equations of continuous variables.

Now, we can express every phasor in Eq. (17) as $E_k = A_k e^{\imath\phi_k}$. As a working hypothesis we will consider the intensities A_k^2 as either homogeneous or as *quenched* with respect to phases. The first condition occurs, for instance, to the input intensities $|E_k^{\text{in}}|^2$ produced by a phase-only spatial light modulator (SLM) with homogeneous illumination [80]. With *quenched* here we mean, instead, that the intensity of each mode is the same for every solution of Eq. (14) at fixed $\mathbb{T}$. We stress that including intensities in the model does not preclude the inference analysis but it is out of the focus of the present work and will be considered elsewhere.

If all intensities are uniform in input and in output, this amounts to a constant rescaling for each one of the four sectors of matrix $\mathbb{J}$ in Eq. (20) that will not change the properties of the matrices. For instance, if the original transmission matrix is unitary, it will be the rescaled one and the matrix $\mathbb{U}$ will be diagonal. Otherwise, if intensities are *quenched*, i.e., they can be considered as constants in Eq. (14), they are inhomogeneous with respect to phases. The generic Hamiltonian element will, therefore, rescale as

$$E_n^* J_{nm} E_m = J_{nm} A_n A_m e^{\imath(\phi_n - \phi_m)} \to J_{nm} e^{\imath(\phi_n - \phi_m)}$$

and the properties of the original J_{nm} components are not conserved in the rescaled one. In particular, we have no argument, anymore, to possibly set the rescaled $U_{nm} \propto \delta_{nm}$ in any case. Eventually, we end up with the complex couplings XY model, whose real-valued Hamiltonian is written as

$$\begin{aligned} \mathcal{H} &= -\frac{1}{2} \sum_{nm} J_{nm} e^{-\imath(\phi_n - \phi_m)} + \text{c.c.} \\ &= -\frac{1}{2} \sum_{nm} \left[J_{nm}^R \cos(\phi_n - \phi_m) + J_{nm}^I \sin(\phi_n - \phi_m) \right], \end{aligned} \quad (21)$$

where J_{nm}^R and J_{nm}^I are the real and imaginary parts of J_{nm}. $\mathbb{J}$ being Hermitian, $J_{nm}^R = J_{mn}^R$ is symmetric and $J_{nm}^I = -J_{mn}^I$ is skew-symmetric. Using data on phase configurations, such as those produced by spatial light modulators, or on any vectorial Potts representation of intensity pixels, such as those recorded on standard CCD cameras on output [73,74,78–81], one can thus apply the techniques presented in this work to infer the values of the effective transmission matrix elements.

VIII. CONCLUSIONS

Different statistical inference methods have been applied to the inverse problem of the XY model. After a short review of techniques based on pseudolikelihood and their formal generalization to the model we have tested their performances against data generated by means of Monte Carlo numerical simulations of known instances with diluted, sparse interactions.

The main outcome is that the best performances are obtained by means of the pseudolikelihood method combined with decimation. Putting to zero (i.e., decimating) very weak bonds, this technique turns out to be very precise for problems whose real underlying interaction network is sparse; i.e., the number of couplings per variable does not scale with number of variables. The PLM+decimation method is compared to the PLM+regularization method, with both l_1 and l_2 regularization, and to a mean-field-based method. The behavior of the quality of the network reconstruction is analyzed by looking at the overall sorted couplings, at the single-site couplings, comparing them with the real network, at the positive predictive value curves, and at the reconstruction errors in all approaches. In the PLM+decimation method, moreover, the identification of the number of decimated bonds at which the tilted pseudolikelihood is maximum allows for a precise estimate of the total number of bonds. Concerning this technique, it is also shown that the network with the most likely number of bonds is also the one of least reconstruction error, where not only the prediction of the presence of a bond is taken into account but also its value.

The behavior of the inference quality in temperature and in the size of data samples is also investigated, basically confirming the low-T behavior hinted at by Nguyen and Berg [72] for the Ising model. In temperature, in particular,

the reconstruction error curve displays a minimum at a low temperature, close to the critical point in those cases in which a critical behavior occurs, and a sharp increase as temperature goes to zero. The decimation method, once again, appears to enhance this minimum of the reconstruction error of almost an order of magnitude with respect to other methods.

The techniques displayed and the results obtained in this work can be of use in any of the many systems whose theoretical representation is given by Eq. (1) or Eq. (21), some of which are recalled in Sec. II. In particular, a possible application can be the field of light wave propagation through random media and the corresponding problem of the reconstruction of an object seen through an opaque medium or a disordered optical fiber [73–80].

ACKNOWLEDGMENTS

We thank Paolo Cazzato, Aurelienne Decelle, Milena De Giorgi, Federico Ricci Tersenghi, and Daniele Sanvitto for fruitful discussions. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 694925).

[1] S. Cocco and R. Monasson, Phys. Rev. Lett. **106**, 090601 (2011).
[2] E. Aurell and M. Ekeberg, Phys. Rev. Lett. **108**, 090201 (2012).
[3] F. Ricci-Tersenghi, J. Stat. Mech.: Theory Exp. (2012) P08015.
[4] H. C. Nguyen and J. Berg, J. Stat. Mech.: Theory Exp. (2012) P03004.
[5] S. Cocco and R. Monasson, J. Stat. Phys. **147**, 252 (2012).
[6] J. Raymond and F. Ricci-Tersenghi, Phys. Rev. E **87**, 052111 (2013).
[7] T. Mora, A. M. Walczak, W. Bialek, C. G. Callan, and G. J. Curtis, Proc. Natl. Acad. Sci. U.S.A. **107**, 5405 (2010).
[8] S. Balakrishnan, H. Kamisetty, J. G. Carbonell, S.-I. Lee, and C. J. Langmead, Proteins: Struct., Funct., Bioinf. **79**, 1061 (2011).
[9] M. Ekeberg, C. Lövkvist, Y. Lan, M. Weigt, and E. Aurell, Phys. Rev. E **87**, 012707 (2013).
[10] C. Feinauer, M. J. Skwark, A. Pagnani, and E. Aurell, PLoS Comput. Biol. **10**, e1003847 (2014).
[11] E. De Leonardis, B. Lutz, S. Ratz, S. Cocco, R. Monasson, A. Schug, and W. Weigt, Nucleic Acids Res. **43**, 10444 (2015).
[12] L. Asti, G. Uguzzoni, P. Marcatili, and A. Pagnani, PLoS Comput. Biol. **12**, e1004870 (2016).
[13] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Nature (London) **440**, 1007 (2006).
[14] Y. Roudi, E. Aurell, and J. Hertz, Front. Comput. Neurosci. **3**, 22 (2009).
[15] J. Tyrcha, Y. Roudi, M. Marsili, and J. Hertz, J. Stat. Mech.: Theory Exp. (2012) P03005.
[16] G. Tkacik, T. Mora, O. Marre, D. Amodei, S. E. Palmer, M. J. Berry, II, and W. Bialek, Proc. Natl. Acad. Sci. U.S.A. **112**, 11508 (2015).
[17] Y. Katz, K. Tunstrom, C. C. Ioannou, C. Huepe, and I. D. Couzin, Proc. Natl. Acad. Sci. U.S.A. **108**, 18720 (2011).
[18] W. Bialek, A. Cavagna, I. Giardina, T. Mora, E. Silvestri, M. Viale, and A. M. Walczak, Proc. Natl. Acad. Sci. U.S.A. **109**, 4786 (2012).
[19] W. Bialek, A. Cavagna, I. Giardina, T. Mora, O. Pohl, E. Silvestri, M. Viale, and A. M. Walczak, Proc. Natl. Acad. Sci. U.S.A. **111**, 7212 (2014).
[20] I. Mastromatteo, E. Zarinelli, and M. Marsili, J. Stat. Mech.: Theory Exp. (2012) P03011.
[21] S. Yamanaka, M. Ohzeki, and A. Decelle, J. Phys. Soc. Jpn. **84**, 024801 (2015).
[22] K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (MIT, Cambridge, Massachusetts, 2012).
[23] Y. Roudi, J. Tyrcha, and J. Hertz, Phys. Rev. E **79**, 051915 (2009).
[24] H. J. Kappen and F. B. Rodríguez, Neural Comput. **10**, 1137 (1998).
[25] T. Tanaka, Phys. Rev. E **58**, 2302 (1998).
[26] V. Sessak and R. Monasson, J. Phys. A **42**, 055001 (2009).
[27] M. Opper and D. Saad, *Advanced Mean Field Methods: Theory and Practice* (MIT, Cambridge, Massachusetts, 2001).
[28] P. Ravikumar, M. J. Wainwright, and J. D. Lafferty, Ann. Stat. **38**, 1287 (2010).
[29] R. Potts, Proc. Cambridge Philos. Soc. **48**, 106 (1952).
[30] J. Ashkin and E. Teller, Phys. Rev. **64**, 178 (1943).
[31] L. Pauling, J. Am. Chem. Soc. **57**, 2680 (1935).
[32] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, London, 1982).
[33] B. Sutherland, J. Math. Phys. **11**, 3183 (1970).
[34] C. Fan and F. Y. Wu, Phys. Rev. B **2**, 723 (1970).
[35] R. Baxter, Phys. Rev. Lett. **26**, 832 (1971).
[36] A. Marruzzo and L. Leuzzi, Phys. Rev. B **91**, 054201 (2015).
[37] A. Marruzzo and L. Leuzzi, Phys. Rev. B **93**, 094206 (2016).
[38] A. Gordon and B. Fischer, Phys. Rev. Lett. **89**, 103901 (2002).
[39] O. Gat, A. Gordon, and B. Fischer, Phys. Rev. E **70**, 046108 (2004).
[40] L. Angelani, C. Conti, L. Prignano, G. Ruocco, and F. Zamponi, Phys. Rev. B **76**, 064202 (2007).
[41] L. Angelani, C. Conti, G. Ruocco, and F. Zamponi, Phys. Rev. Lett. **96**, 065702 (2006).
[42] L. Leuzzi, C. Conti, V. Folli, L. Angelani, and G. Ruocco, Phys. Rev. Lett. **102**, 083901 (2009).
[43] F. Antenucci, C. Conti, A. Crisanti, and L. Leuzzi, Phys. Rev. Lett. **114**, 043901 (2015).
[44] F. Antenucci, A. Crisanti, and L. Leuzzi, Phys. Rev. A **91**, 053816 (2015).
[45] Edited by B. Schölkopf, J. C. Platt, and T. Hoffman, *Advances in Neural Information Processing Systems* (MIT, Cambridge, Massachusetts, 2006).
[46] A. Decelle and F. Ricci-Tersenghi, Phys. Rev. Lett. **112**, 070603 (2014).
[47] P. Tyagi, A. Pagnani, F. Antenucci, M. Ibañez Berganza, and L. Leuzzi, J. Stat. Mech.: Theory Exp. (2015) P05031.
[48] V. L. Berezinskii, Zh. Eksp. Teor. Fiz. **59**, 907 (1970) [Sov. Phys. JETP **32**, 493 (1971)] .
[49] J. Kosterlitz and D. Thouless, J. Phys. C **5**, L124 (1972).
[50] E. Brézin, J. Phys. (France) **43**, 15 (1982).
[51] J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, Cambridge, 1996).

[52] J. Villain, J. Phys. C **10**, 1717 (1977); **10**, 4793 (1977).
[53] E. Fradkin, B. A. Huberman, and S. H. Shenker, Phys. Rev. B **18**, 4789 (1978).
[54] Y. Kuramoto, *International Symposium on Mathematical Problems in Theoretical Physics: January 23–29, 1975, Kyoto University, Kyoto/Japan*, edited by H. Araki (Springer, Berlin, Heidelberg, 1975) pp. 420–422.
[55] S. Teitel and C. Jayaprakash, Phys. Rev. B **27**, 598 (1983).
[56] S. Teitel and C. Jayaprakash, Phys. Rev. Lett. **51**, 1999 (1983).
[57] M. Nixon, E. Ronen, A. A. Friesem, and N. Davidson, Phys. Rev. Lett. **110**, 184102 (2013).
[58] C. Reynolds, Computer Graphics **21**, 25 (1987).
[59] J. L. Deneubourg and S. Goss, Ethology, Ecology, Evolution **1**, 295 (1989).
[60] A. Huth and C. Wissel, in *Biological Motion*, edited by W. Alt and E. Hoffmann (Springer-Verlag, Berlin, 1990), pp. 577–590.
[61] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, Phys. Rev. Lett. **75**, 1226 (1995).
[62] A. Cavagna, S. M. Duarte Queirós, I. Giardina, F. Stefanini, and M. Viale, Proc. R. Soc. London B **280**, 20122484 (2013).
[63] F. Antenucci, M. Ibáñez Berganza, and L. Leuzzi, Phys. Rev. A **91**, 043811 (2015).
[64] F. Antenucci, M. Ibañez Berganza, and L. Leuzzi, Phys. Rev. B **92**, 014204 (2015).
[65] D. S. Wiersma, Nat. Phys. **4**, 359 (2008).
[66] J. Andreasen, A. A. Asatryan, L. C. Botten, B. A. Byrne, H. Cao, L. Ge, L. Labonté, P. Sebbah, A. D. Stone, H. E. Türeci, and C. Vanneste, Adv. Opt. Photonics **3**, 88 (2011).
[67] F. Antenucci, A. Crisanti, M. Ibañez Berganza, A. Marruzzo, and L. Leuzzi, Philos. Mag. **96**, 704 (2016).
[68] J. Besag, J. R. Stat. Soc. D **24**, 179 (1975).
[69] M. Schmidt, minFunc: Unconstrained differentiable multivariate optimization in MATLAB, http://www.cs.ubc.ca/~schmidtm/Software/minFunc.html.
[70] D. J. Earl and M. W. Deem, Phys. Chem. Chem. Phys. **7**, 3910 (2005).
[71] P. Olsson and P. Minnhagen, Phys. Scr. **43**, 203 (1992).
[72] H. C. Nguyen and J. Berg, Phys. Rev. Lett. **109**, 050602 (2012).
[73] I. Vellekoop and A. Mosk, Opt. Lett. **32**, 2309 (2007).
[74] I. Vellekoop, E. van Putten, A. Lagendijk, and A. Mosk, Opt. Express **16**, 67 (2008).
[75] I. M. Vellekoop and A. P. Mosk, Phys. Rev. Lett. **101**, 120601 (2008).
[76] H. Yilmaz, W. L. Vos, and A. P. Mosk, Biomed. Opt. Express **4**, 1759 (2013).
[77] F. Riboli, N. Caselli, S. Vignolini *et al.*, Nat. Mater. **13**, 720 (2014).
[78] S. M. Popoff, G. Lerosey, R. Carminati, M. Fink, A. C. Boccara, and S. Gigan, Phys. Rev. Lett. **104**, 100601 (2010).
[79] D. Akbulut, T. J. Huisman, E. G. van Putten, W. L. Vos, and A. P. Mosk, Opt. Express **19**, 4017 (2011).
[80] S. M. Popoff, G. Lerosy, M. Fink, A. C. Boccara, and S. Gigan, New J. Phys. **13**, 123021 (2011).
[81] A. Dremeau, A. Liutkus, D. Martina, O. Katz, C. Schulke, F. Krzakala, S. Gigan, and L. Daudet, Opt. Express **23**, 11898 (2015).
[82] J. W. Goodman, *Statistical Optics* (John Wiley & Sons, Hoboken, New Jersey, 1985).