



## Inverse modeling of time-delayed interactions via the dynamic-entropy formalism

Elena Agliari <sup>1</sup>, Francesco Alemanno <sup>2</sup>, Adriano Barra,<sup>3,4,\*</sup> Michele Castellana,<sup>5</sup> Daniele Lotito,<sup>6</sup> and Matthieu Piel<sup>7</sup>

<sup>1</sup>*Dipartimento di Matematica, Sapienza Università di Roma, Roma, Italy*

<sup>2</sup>*Dipartimento di Matematica e Fisica, Università del Salento, Lecce, Italy*

<sup>3</sup>*Dipartimento di Scienze di Base e Applicate per l'Ingegneria, Sapienza Università di Roma, Roma, Italy*

<sup>4</sup>*Istituto Nazionale di Fisica Nucleare, Sezione di Lecce, Italy*

<sup>5</sup>*Laboratoire PhysicoChimie, Institut Curie, CNRS UMR168, Paris, France*

<sup>6</sup>*Dipartimento di Informatica, Università di Pisa, Pisa, Italy*

<sup>7</sup>*Laboratoire de Biologie Cellulaire et Cancer, Institut Curie, CNRS UMR168, Paris, France*



(Received 3 September 2023; revised 1 May 2024; accepted 15 July 2024; published 1 August 2024)

Although instantaneous interactions are unphysical, a large variety of maximum entropy statistical inference methods match the model-inferred and the empirically measured equal-time correlation functions. Focusing on collective motion of active units, this constraint is reasonable when the interaction timescale is much faster than that of the interacting units, as in starling flocks, yet it fails in a number of counterexamples, as in leukocyte coordination (where signaling proteins diffuse among two cells). Here, we relax this assumption and develop a path integral approach to maximum-entropy framework, which includes delay in signaling. Our method is able to infer the strength of couplings and fields, but also the time required by the couplings to completely transfer information among the units. We demonstrate the validity of our approach providing excellent results on synthetic datasets of non-Markovian trajectories generated by the Heisenberg-Kuramoto and Vicsek models equipped with delayed interactions. As a proof of concept, we also apply the method to experiments on dendritic migration, where matching equal-time correlations results in a significant information loss.

DOI: [10.1103/PhysRevE.110.024301](https://doi.org/10.1103/PhysRevE.110.024301)

The ability to move is a fundamental property of many living systems, ranging on different scales, from animals to cells [1]. As for the latter, advances in live imaging allowed scientists to collect big data for which statistical analysis has now become a robust and reliable tool [2]. These investigations suggest that cells often migrate in groups and communicate as they move. Establishing whether such interactions are present is, in many cases, of utmost importance to possibly control and anticipate the evolution of a system. However, answering this question is notoriously difficult, especially when unit-to-unit communication is not supported by the existence of physical bonds, as in juxtacrine interactions, but it is rather induced by some signaling pathways, as in paracrine interactions. To inspect these possible interactions, several methods have been introduced in the past few decades. Among these, inferential techniques have been designed to take as input some tracks of moving units, such as birds in flocks or migrating cells, and output the parameters which provide information about the motion and the existence of interactions among the units, and/or between the units and an external source, such as a predator for a bird flock [3] or a cancer cell producing signaling proteins for leukocytes [4].

In particular, approaches based on the maximum entropy (ME) principle [5] have been proven successful in a broad variety of contexts, revealing interaction patterns resulting

from amino-acid sequences in protein families [6], interaction structures of genetic networks [7], effective interactions in networks of neurons [8], etc. The state of the art of this inverse modeling still exhibits some limitations which make it unsatisfactory at work on units coordination for collective motion. Specifically, this approach usually forces the theoretical and experimental equal-time correlation functions to match, implicitly assuming instantaneous interactions among units; however, this assumption may not hold when the interaction is mediated by some chemicals which take some time to diffuse from the emitter to the receiver. For example, in cell-migration experiments, cell  $i$  may locally release a chemical compound along its migratory path, and cell  $j$  may cross the former path of  $i$  at a later time and thus feel a delayed interaction with  $i$  mediated by this compound. In this work, we introduce, develop, and test a ME framework that allows unveiling not only time-lagged effective interactions but also, and more generally, interactions that arise indirectly as cell  $i$  alters the environment of cell  $j$ .

Let us consider the motion of  $N$  particles (i.e., units) in a  $D$ -dimensional space and denote their trajectories by

$$\{\mathbf{x}_i^t\}, \quad 1 \leq i \leq N, \quad 1 \leq t \leq N_T + 1, \quad (1)$$

where  $\mathbf{x}_i^t$  is the  $D$ -dimensional position vector of particle  $i$  at time  $t$ , and the superscript index indicates instants of time separated by  $\delta t$ , for a total of  $N_T + 1$  temporal samples. The velocity  $\mathbf{v}_i^t$  and direction of motion  $\mathbf{s}_i^t$  of each particle  $i$  at time

\*Contact author: [adriano.barra@uniroma1.it](mailto:adriano.barra@uniroma1.it)

$t$  then read, respectively,

$$\mathbf{v}_i^t \equiv \frac{\mathbf{x}_i^{t+1} - \mathbf{x}_i^t}{\delta t}, \quad \mathbf{s}_i^t \equiv \frac{\mathbf{v}_i^t}{|\mathbf{v}_i^t|} \quad \text{and} \quad \mathbf{m}_t \equiv \frac{1}{N} \sum_{i=1}^N \mathbf{s}_i^t, \quad (2)$$

is the instantaneous mean alignment of the moving units at time  $t$  which shall also be referred to as the ‘‘magnetization,’’ while  $\mathbf{s}^t \equiv \{\mathbf{s}_1^t, \dots, \mathbf{s}_N^t\}$  is interpreted as a configuration (at time  $t$ ) of  $N$  Heisenberg spins. For practical purposes, we assume that the motion of these units may be temporally correlated up to a maximum temporal window of length  $\tau_M$ , with  $0 \leq \tau_M < N_T$ . Hereafter, we assume that the correct value for  $\tau_M$  is known a priori; later, once the necessary tools are developed, we will provide a means for determining  $\tau_M$  from the data in a self-consistent manner via the Akaike information criterion (AIC) [9].

The two natural observables we use to quantify this temporal correlation are the temporally averaged magnetization  $\mathbf{M}$ ,

$$\mathbf{M} \equiv \frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} \mathbf{m}_t, \quad (3)$$

and the two-point correlation function  $R$  with delay  $\tau$ ,

$$R(\tau) \equiv \frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} \mathbf{m}_t \cdot \mathbf{m}_{t-\tau}, \quad 1 \leq \tau \leq \tau_M. \quad (4)$$

We seek  $P(\mathbf{s})$  as the minimal probability measure such that the theoretical averages for the magnetization (3) and correlation function (4) match their empirical values and, accordingly, we obtain it following the ME principle [5,10,11]: we define the path entropy  $S$  as

$$S[P] = - \int \mathcal{D}\mathbf{s} P(\mathbf{s}) \log P(\mathbf{s}), \quad (5)$$

where  $\mathcal{D}\mathbf{s} \equiv \prod_{t=1}^{N_T} \prod_{i=1}^N ds_i^t$  and  $ds_i^t$  is the surface element of the  $(D-1)$ -dimensional sphere.

To solve this optimization task, we introduce the Lagrangian multipliers  $\mathbf{J} \equiv \{J_\tau\}_{\tau=1}^{\tau_M}$  and  $\mathbf{H} \equiv \{H_\ell\}_{\ell=1}^D$ , and obtain

$$P(\mathbf{s}) = \frac{1}{Z} \exp \left[ \sum_{t=\tau_M+1}^{N_T} \sum_{i=1}^N \mathbf{s}_i^t \cdot \left( \frac{1}{N} \sum_{\tau=1}^{\tau_M} J_\tau \sum_{j=1}^N \mathbf{s}_j^{t-\tau} + \mathbf{H} \right) \right], \quad (6)$$

where  $Z \equiv \int \mathcal{D}\mathbf{s} P[\mathbf{s}]$  is a normalization constant, see the Appendix for details.

Remarkably,  $P(\mathbf{s})$  can be looked at as a Boltzmann-Gibbs distribution and, consistently, the argument of its exponential can be interpreted as a Hamiltonian. Thus, the Lagrange multipliers  $\mathbf{H}$  and  $J_\tau$  play, respectively, as an effective external field and as an effective interaction-strength, such that the direction of the  $i$ th unit at time  $t$  tends to align with the direction of  $\mathbf{H}$  and to align (misalign) with the average unit’s direction at time  $t - \tau$  if  $J_\tau > 0$  ( $J_\tau < 0$ ). Having obtained an analytical expression for  $P(\mathbf{s})$ , we infer the parameters  $\mathbf{J}$  and  $\mathbf{H}$  from datasets.

Before proceeding we note that, in the following, the theoretical average over  $P(\mathbf{s})$  shall be denoted by brackets, while

the experimental average shall be denoted by the superscript  $E$  (e.g.,  $\langle \mathbf{m}_t \rangle$  and  $\mathbf{m}_t^E$ ).

As averaging over the measure  $P(\mathbf{s})$  is computationally cumbersome (as we have to integrate over  $N_T \times N$  coupled variables at once), *en route* toward an approach that can be routinely applied in a laboratory, hereafter we explore an approximation, whose consistency can be checked a posteriori. First, in Eq. (6) we replace the internal field with its empirical counterpart,

$$\frac{1}{N} \sum_{\tau=1}^{\tau_M} J_\tau \sum_{j=1}^N \mathbf{s}_j^{t-\tau} \rightarrow \sum_{\tau=1}^{\tau_M} J_\tau \mathbf{m}_{t-\tau}^E. \quad (7)$$

The resulting distribution is referred to as  $P_A$  to highlight that it is an approximation of  $P$ . Denoting by  $\langle \rangle_A$  the related average, we get

$$\begin{aligned} \langle \mathbf{m}_t \rangle_A &= \int \left( \prod_{t=\tau_M+1}^{N_T} \prod_{i=1}^N ds_i^t \right) P_A[\mathbf{s}] \frac{1}{N} \sum_{i=1}^N \mathbf{s}_i^t \\ &= \begin{cases} \mathcal{M}(\sum_{\tau=1}^{\tau_M} J_\tau \mathbf{m}_{t-\tau}^E + \mathbf{H}) & \text{if } \tau_M \leq t \leq N_T, \\ \mathbf{m}_t^E & \text{if } 1 \leq t \leq \tau_M, \end{cases} \end{aligned} \quad (8)$$

where

$$\mathcal{M}(\mathbf{x}) \equiv \frac{\mathbf{x} \mathcal{I}_1(|\mathbf{x}|)}{|\mathbf{x}| \mathcal{I}_0(|\mathbf{x}|)}, \quad (9)$$

and  $\mathcal{I}_0, \mathcal{I}_1$  are hyperbolic Bessel functions of order 0 and 1, respectively [12].

Next, we notice that  $P_A$  is factorized with respect to every variable  $\mathbf{s}_i^t$ , hence  $\langle \mathbf{m}_t \mathbf{m}_{t'} \rangle_A = \langle \mathbf{m}_t \rangle_A \langle \mathbf{m}_{t'} \rangle_A$ . This allows us to recast the constraints in the ME extremization as

$$\mathbf{M}^E = \frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} \langle \mathbf{m}_t \rangle_A, \quad (10)$$

$$R^E(\tau) = \frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} \langle \mathbf{m}_t \rangle_A \cdot \langle \mathbf{m}_{t-\tau} \rangle_A, \quad 1 \leq \tau \leq \tau_M. \quad (11)$$

The nonlinearities in  $\mathbf{J}$  and  $\mathbf{H}$  introduced by the operator  $\mathcal{M}$  make the direct evaluation of these parameters from Eqs. (10) and (11) nonfeasible. To circumvent this difficulty, we can note that these equations are simultaneously fulfilled by requiring that

$$\langle \mathbf{m}_t \rangle_A = \mathbf{m}_t^E, \quad \tau_M + 1 \leq t \leq N_T, \quad (12)$$

which, exploiting Eq. (8), becomes

$$\mathcal{M} \left( \sum_{\tau=1}^{\tau_M} J_\tau \mathbf{m}_{t-\tau}^E + \mathbf{H} \right) = \mathbf{m}_t^E, \quad \tau_M + 1 \leq t \leq N_T, \quad (13)$$

and, finally, by introducing the inverse function  $\mathbf{G}_t^E \equiv \mathcal{M}^{-1}(\mathbf{m}_t^E)$  we can shift the nonlinearity on the experimentally determined term and get

$$\sum_{\tau=1}^{\tau_M} J_\tau \mathbf{m}_{t-\tau}^E + \mathbf{H} = \mathbf{G}_t^E, \quad \tau_M + 1 \leq t \leq N_T. \quad (14)$$

Hence, the solution of the original integral over the past history of all the particles (5) is effectively approximated in terms of the solution of the linear system (14).

We stress that Eq. (12) constitutes a stricter condition than Eqs. (10) and (11), because it stems from approximation (7) that removed degrees of freedom. To avoid Eq. (14) being overconstrained, we include an additional and controllable term in it, i.e.,  $\sum_{\tau=1}^{\tau_M} J_\tau \mathbf{m}_{t-\tau}^E + \mathbf{H} + \sigma \boldsymbol{\epsilon}_t = \mathbf{G}_t^E$ , where  $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}_D, \mathbf{1}_D)$  and  $\sigma \geq 0$  is a further parameter to be determined: we have modified the constraint in such a way that the relation (14) does not hold deterministically. Thus, the related log-likelihood  $l_A$  reads

$$l_A(\mathbf{J}, \mathbf{H}, \sigma | s^E) \equiv -\frac{1}{2\sigma^2} \sum_{t=\tau_M+1}^{N_T} \left| \sum_{\tau=1}^{\tau_M} J_\tau \mathbf{m}_{t-\tau}^E + \mathbf{H} - \mathbf{G}_t^E \right|^2 - (N_T - \tau_M) \frac{D}{2} \log(2\pi\sigma^2). \quad (15)$$

We now obtain the optimal estimates for  $\mathbf{J}$ ,  $\mathbf{H}$ , and  $\sigma$ . Setting  $\nabla_{\mathbf{H}} l_A = 0$ , we obtain

$$\mathbf{H} = \langle \mathbf{G}_t^E \rangle_t - \sum_{\tau=1}^{\tau_M} J_\tau \langle \mathbf{m}_{t-\tau}^E \rangle_t, \quad (16)$$

where  $\langle \cdot \rangle_t \equiv \frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} (\cdot)$  denotes the time average, and, proceeding analogously for  $\nabla_{J_\tau} l_A = 0$ , we get

$$\mathbf{J} = \mathbf{A}^{-1} \cdot \mathbf{B}, \quad (17)$$

with

$$A_{\lambda\tau} \equiv \langle \mathbf{m}_{t-\tau}^E \cdot \mathbf{m}_{t-\lambda}^E \rangle_t - \langle \mathbf{m}_{t-\tau}^E \rangle_t \cdot \langle \mathbf{m}_{t-\lambda}^E \rangle_t, \quad (18)$$

$$B_\lambda \equiv \langle \mathbf{G}_t^E \cdot \mathbf{m}_{t-\lambda}^E \rangle_t - \langle \mathbf{G}_t^E \rangle_t \cdot \langle \mathbf{m}_{t-\lambda}^E \rangle_t. \quad (19)$$

Finally, setting  $\nabla_\sigma l_A = 0$ , we obtain

$$\sigma^2 = \frac{1}{D} \left\langle \left| \mathbf{H} + \sum_{\tau=1}^{\tau_M} J_\tau \mathbf{m}_{t-\tau}^E - \mathbf{G}_t^E \right|^2 \right\rangle_t. \quad (20)$$

As detailed in the Appendix, an estimate of the errors for the inferred parameters can be obtained by using the Fisher information for the log-likelihood [13]. In fact,

$$\text{Var}(\mathbf{J}, \mathbf{H}, \sigma) \approx \text{diag}(\mathbf{I}^{-1}), \quad (21)$$

where the approximation holds as long as the sample size is large enough and  $\mathbf{I}$  is the Fisher information matrix, which in the present setting reads

$$\mathbf{I} = \frac{N_T - \tau_M}{\sigma^2} \begin{pmatrix} \langle \mathbf{m}_{t-\tau}^E \cdot \mathbf{m}_{t-\lambda}^E \rangle_t & \langle \mathbf{m}_{t-\tau}^E \rangle_t & 0 \\ \langle \mathbf{m}_{t-\lambda}^E \rangle_t & \delta & \mathbf{0} \\ 0 & \mathbf{0} & 2D \end{pmatrix}.$$

So far in our analysis, we assumed that  $\tau_M$  is known, yet in most practical cases it is not. To estimate  $\tau_M$ , we rely on the AIC [9], which provides a tool for model selection, each model being characterised by a different value of  $\tau_M$ . In general, given a dataset and a model with a mean-squared error  $\epsilon^2$ ,  $n_p$  parameters, and  $n_o$  observations, the quantity

$$\text{AIC} \equiv \frac{2n_p}{n_o} + \log(\epsilon^2) \quad (22)$$

assesses the quality of such a model for the data available. In fact, the AIC favors models with a small error *and* a small number of parameters. In the analysis above, the parameters in Eq. (22) are

$$n_o = D(N_T - \tau_M), \quad n_p = \tau_M + D + 1, \quad \epsilon^2 = \frac{N_T}{N_T - \tau_M} \sigma^2,$$

thus

$$\text{AIC} = 2 \frac{\tau_M + D + 1}{D(N_T - \tau_M)} + \log \frac{N_T \sigma^2}{N_T - \tau_M}. \quad (23)$$

In the following, after performing inference for multiple values of  $\tau_M$  and determining  $\sigma^2$  from Eq. (20), we evaluate Eq. (23) for each inferred model (i.e., by varying  $\tau_M$ ) and we choose the effective  $\tau_M$  as that with the smallest value of AIC: the maximal (in modulus)  $\mathbf{J}$  within the time window  $0 \leq t \leq \tau_M$  is then interpreted as the best inferred coupling.

We test our framework on both synthetic and real datasets. The models used to generate the synthetic datasets are the Heisenberg-Kuramoto model (HKM) [4], whose Boltzmann-Gibbs distribution is precisely (6), and the (topological or metric [14]) Vicsek model (VM) [15], both equipped with a (tunable) delay in the interactions [16]. Each model has its own set of parameters, which we denote as *planted* parameters, while those returned by the ME extremization are labeled as *inferred*.

As a *conditio sine qua non* for further analyses, the validation of our ME method on the HKM is shown in Fig. 1 providing excellent results, even away from the Markovian limit  $\tau_M = 1$ .

Next, introducing the i.i.d. noise  $\boldsymbol{\eta}_k^t \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$  and using  $\Delta$  to account for the planted delay in signaling, we consider the VM in a 2D square lattice of size  $L$  with periodic boundary conditions, defined by the following dynamical equations [15,17]:

$$\mathbf{v}_k^{t+1} = v_0 \Gamma \left( \frac{J}{|n_k^t|} \sum_{j \in n_k^t} \frac{\mathbf{v}_j^{t-D_{ik}}}{v_0} + \mathbf{H} + \sqrt{\delta t} \boldsymbol{\eta}_k^t \right), \quad (24)$$

$$\mathbf{x}_k^{t+1} = \mathbf{x}_k^t + \delta t \mathbf{v}_k^{t+1} \quad k = 1, \dots, N, \quad (25)$$

where  $D_{ij} = (\Delta - 1)(1 - \delta_{ij}^j)$  while  $\mathbf{v}_i^t, \mathbf{x}_i^t$  are the velocity and position of the  $i$ th particle at time  $t$ , respectively, and two temporally consecutive frames are distant  $\delta t$ . In the topological implementation of the VM,  $n_k^t$  is the set of topological neighbors of the  $k$ th particle (including the  $k$ th particle itself) and  $|n_k^t| := n_c$  is the number of neighbors of the  $k$ th particle at time  $t$ . In the metric counterpart, deepened in the Appendix,  $n_k^t$  is the set of the closer units within an interaction range. As standard in the VM,  $\Gamma(\mathbf{v}) \equiv \mathbf{v}/|\mathbf{v}|$ , thus the velocity of each particle has constant magnitude  $v_0$ . Finally, the magnetization of the VM is given by

$$\mathbf{m}_t = \frac{1}{N} \sum_{k=1}^N \Gamma(\mathbf{v}_k^t), \quad (26)$$

which is the adaptation of the last expression in Eq. (2) to the case. In the validation of the ME inference for the VM we set  $H_y = 0$ : the parameters to estimate are thus  $J, H_x$  and the delay (whose inferred value we call  $\tau_{ME}$  as opposite to  $\Delta$  that is the planted value) as reported in Fig. 2. Further, to

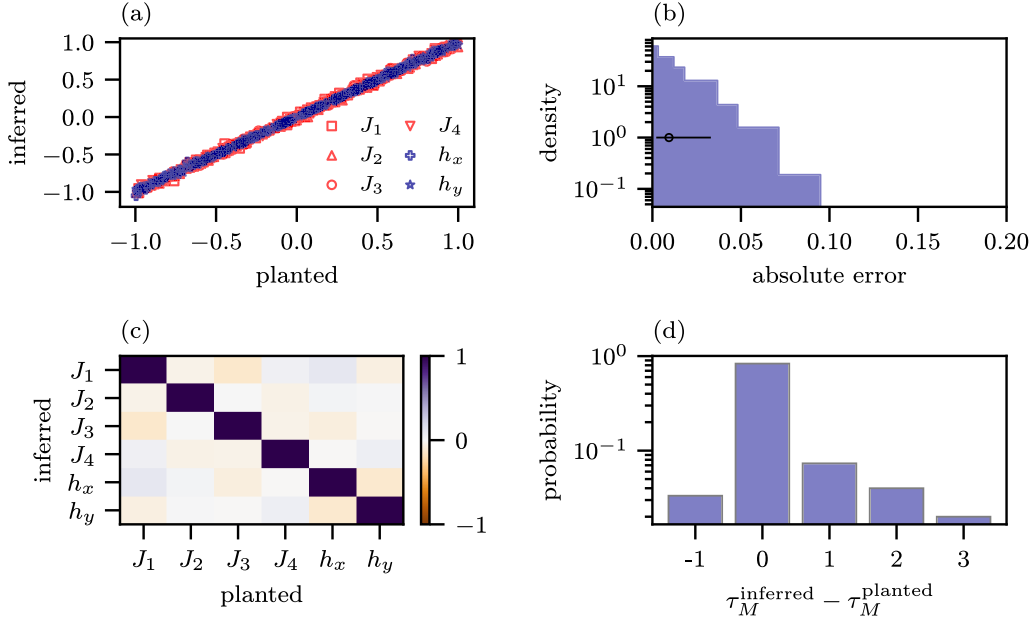


FIG. 1. Validation of our inference method with the HKM. Here  $D = 2$ ,  $\tau_M = 4$ ,  $\sigma = 0.1$ , and we drew 150 planted parameters  $J_1, \dots, J_{\tau_M}, H_x, H_y$  independently and identically from a uniform distribution over the range  $(-1, 1)$ . (a) Scatter plot of planted vs inferred parameters: they match along the diagonal as they should. (b) Histogram of the absolute error between  $\mathbf{J}, \mathbf{H}$  planted and  $\mathbf{J}, \mathbf{H}$  inferred: for each realisation of the process we calculate  $|\mathbf{J}^{\text{inferred}} - \mathbf{J}^{\text{planted}}|$  and  $|\mathbf{H}^{\text{inferred}} - \mathbf{H}^{\text{planted}}|$  and the related  $150 \times (D + \tau_M)$  components make up the sample; the black dot represents the average and the black horizontal line represents a confidence interval of 68%; the average error is  $\sim 10^{-2}$ , which is the expected amount of error given the size of the planted dataset. (c) Pearson correlation matrix between planted and inferred parameters: as expected distinct types of parameters, e.g.,  $\mathbf{H}$  and  $\mathbf{J}$ , are uncorrelated. (d) Histogram of  $\tau_M^{\text{inferred}} - \tau_M^{\text{planted}}$ : the peak of the distribution correctly returns the inferred delay as the planted one.

deepen the statistical properties of the inferred model, we also compare the lowest-order moments of some observables for trajectories generated by the inferred ME model and by the planted VM: in particular, we measure the magnetization  $\mathbf{M}$  and the correlation function  $R(\tau = \Delta)$  to check that their values are statistically compatible: Fig. 3 provides a positive test of the inference method for the topological VM (and the same holds also for the metric case, discussed in the Appendix).

Finally, to highlight the inconsistency that would result by neglecting interaction delays in real experiments, we consider datasets on dendritic migration in a chemokine gradient [10]: all the aspects of biological relevance will be discussed elsewhere, here we just focus on the methodology. Positions of all cells were recorded at (ordered) multiple times and we studied two regions of the experiment, the former (region A) where the gradient is low and the latter (region B) where the gradient is high, providing two datasets suitable to be addressed by our approach; see the Appendix for details. We estimate

$$\delta\text{AIC}(\tau_M) \equiv \text{AIC}(\tau_M) - \min_{\tau'_M} \text{AIC}(\tau'_M),$$

i.e., the difference between AIC for a given  $\tau_M$  and AIC for the optimal  $\tau_M$ , the inferred parameters  $J_\tau, \mathbf{H}, \sigma$  and the connected correlation function

$$C(\tau) \equiv \langle \mathbf{m}(s^t) \cdot \mathbf{m}(s^{t-\tau}) \rangle_t - \langle \mathbf{m}(s^t) \rangle_t \cdot \langle \mathbf{m}(s^{t-\tau}) \rangle_t \quad (27)$$

for the optimal value of  $\tau_M$ . These results are reported in Fig. 4 for both datasets A and B, highlighting marked differences between the prediction by the standard ME and our technique. Indeed, as shown in Fig. 4(a), the optimal  $\tau$  is 13

for region A and 14 for region B, and, as shown in Fig. 4(b), the corresponding couplings  $J_{\tau=13}$  and  $J_{\tau=14}$  are both positive, while in the  $\tau \rightarrow 0$  part of these panels the interactions are actually repulsive: accordingly the connected correlation functions provided in Fig. 4(d) are small but systematically different from zero for  $\tau < \tau_M$ , then they vanish. As expected the method correctly returns a higher field in region B, w.r.t. A, as shown in Fig. 4(c). A further last control test on this inference can be achieved simply by reshuffling the frames and feeding the ME method with this random permutation: in this case the coupling should disappear yet the field should be preserved and, indeed, this is the case, as shown in the Appendix (see Fig. 9).

To summarize, we revised the standard ME method to encompass scenarios where communication among units is slow: the dependence over the whole history of any particle has to be considered and the emerging dynamics does not have to be any longer Markovian. While from a gnoseological perspective such general setting should always be preferable to versions where equal-time correlators are constrained, when communication among units is reasonably faster than the fastest timescale of their dynamics, standard ME and our method return the same outcome. Whatever the route, in a nutshell, we use the knowledge stemming from empirical correlations to construct an effective Hamiltonian model that correctly reproduces the observed motion (e.g., its statistical characteristics, at some prescribed order, depending on how much empirical information we relied upon and thus how many Lagrange multipliers we used). The interactions that we infer via maximum entropy should however be seen as

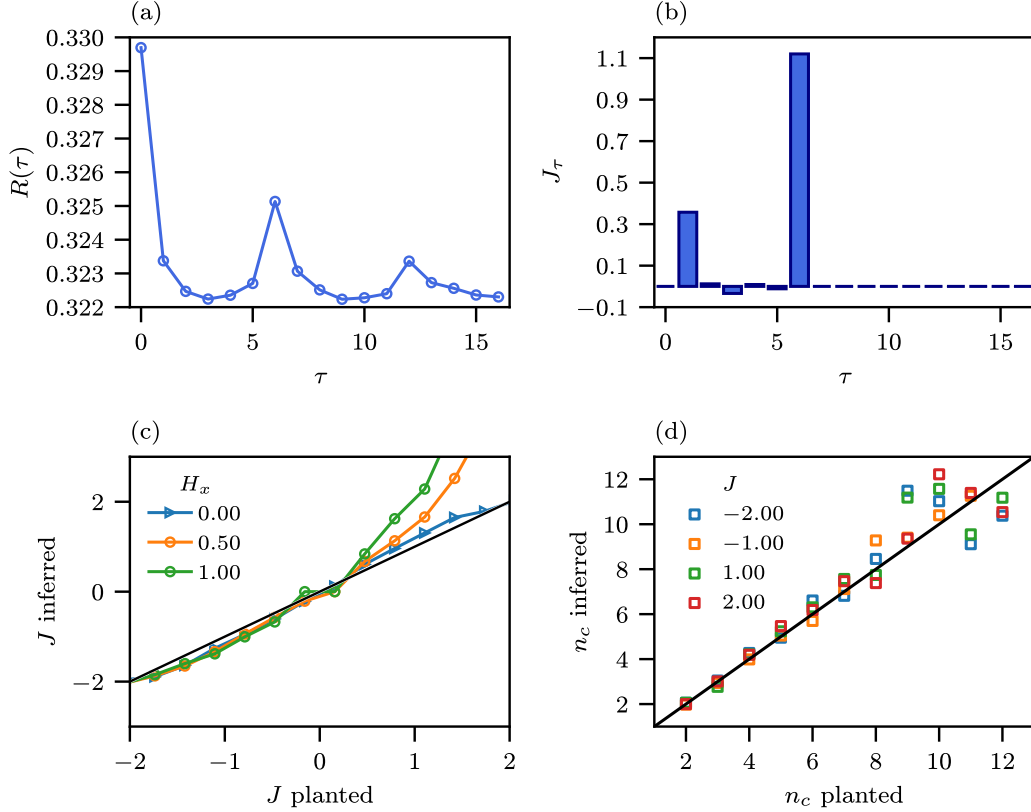


FIG. 2. Validation of our inference method with the topological VM, simulated in a periodic, two-dimensional square lattice with size  $L = 150$ ,  $N = 100$ ,  $\Delta = 6$ ,  $\delta t = 1$ ,  $v_0 = 1$ ,  $n_c = 4$ ,  $\mathbf{H} = (H_x, 0)$ .  $H_x$  and  $J$  are i.i.d. over the set  $[0, 2] \times [0, 2]$ . The analogous picture for the metric case is reported in the Appendix. (a) Example of the empirical correlation  $R(\tau)$  vs  $\tau$ . (b) Example of the (related) inferred coupling: note the presence of two peaks, the former (at  $\tau = 1$ ) is the Markovian self-interaction, the latter (at  $\tau = \Delta = 6$ ) is the non-Markovian contribution by nearest neighbors, in accordance with Eqs. (24) and (25). (c) Scatter plot of the inferred vs the planted values of the coupling: note that the larger the (external) field  $H_x$ , the greater the overestimation of the (inferred) coupling. (d) Scatter plot of the number of nearest neighbors  $n_c$  estimated as the ratio between the intensities of the coupling at  $\tau = 1$  (the Markovian self-interaction accounting for inertia) and at  $\tau = 6$  (the non-Markovian interaction with other units), i.e.,  $n_c = 1 + J_\Delta/J_1$ .

effective interactions as, in principle, we do not know the model that generated the dynamics. For instance, when analyzing trajectories generated by the Vicsek model, the details of the generating model do not enter in our analysis, nevertheless the inferred values of these interactions—despite being *effective*—carry valuable information about the motion (furthermore, an underlying model may not even exist, and in general it is now known).

#### ACKNOWLEDGMENTS

We acknowledge financial support from PNRR MUR (Grant No. PE0000013-FAIR), MAECI (Grant No. F85F21006230001), PRIN MUR (Grant No. 20229T9EAT), and Sapienza University of Rome (Grant No. RM12117A8590B3FA). The Authors are grateful to Andrea Cavagna, Irene Giardina, Theresa Jakuszeit, and Raphael Voituriez for enlightening discussions.

#### APPENDIX

In this Appendix, we present details of the mathematical framework developed for inferring delayed interactions among dynamical units, such as a group of migrating cells,

using only their trajectories. First, we present our approach to generalize the maximum entropy (ME) method, standardly based on matching equal-time correlation, to cope with delay in signaling, then we validate our method on synthetic and real datasets. As for the synthetic ones, we focus on trajectories generated by the Heisenberg-Kuramoto model (HKM) [4] with delay and on trajectories generated according to the Vicsek model (VM) [15], both in its topological as well as metric implementations; finally, we consider a real test-case focusing on the trajectories of dendritic cells migrating toward a chemoattractive source, namely in the presence of a chemokine gradient [10].

#### APPENDIX A: DYNAMIC MAXIMUM-ENTROPY FORMALISM

##### 1. Problem formulation

Let us consider the motion of  $N$  particles in a  $D$ -dimensional space, and denote by

$$\{\mathbf{x}'_i\}, \quad 1 \leq i \leq N, \quad 1 \leq t \leq N_T + 1 \quad (\text{A1})$$

the set of trajectories, where  $\mathbf{x}'_i$  is the position of the particle  $i$  at time  $t$  and the superscript index indicates instants of time separated by  $\delta t$ , for a total of  $N_T + 1$  temporal samples. By

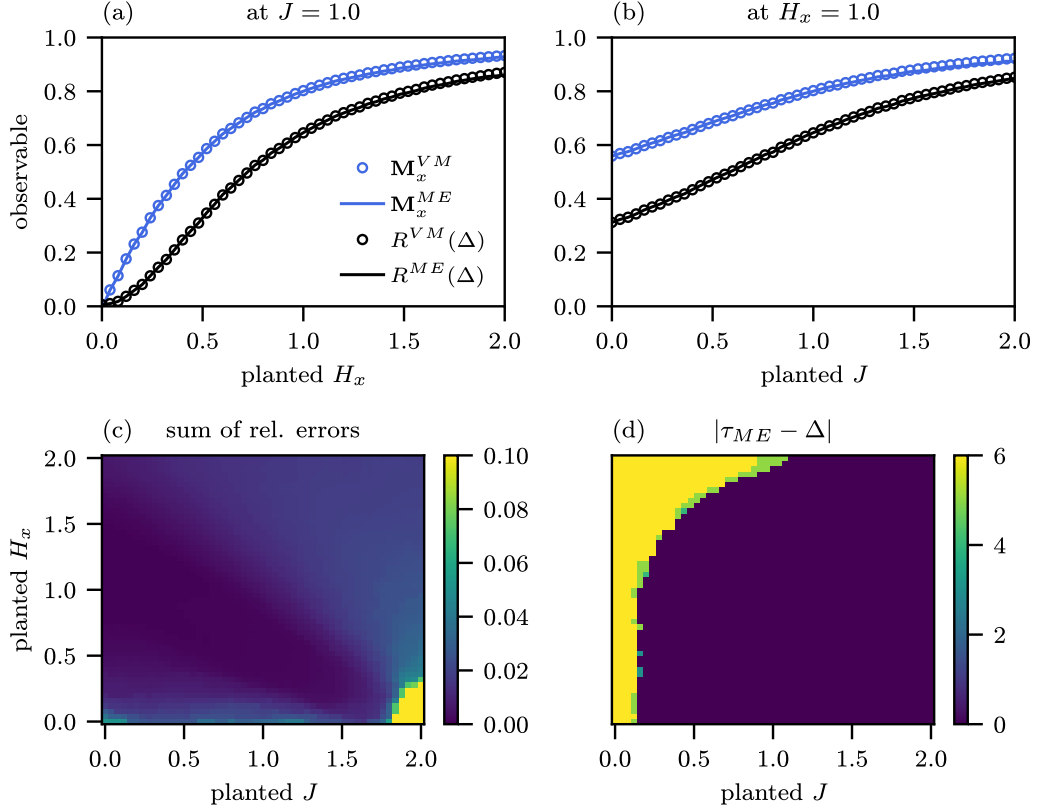


FIG. 3. Validation of our inference method with the topological VM whose setting is the same as in Fig. 2. The superscripts *V* and *ME* pertain to quantities calculated for the VM or inferred via the ME method, respectively. (a) Observables  $\mathbf{M}$  and  $R(\Delta)$  as functions of  $H_x$ , with  $J = 1.0$ . (b) Observables  $\mathbf{M}$  and  $R(\Delta)$  as functions of  $J$ , with  $H_x = 1.0$ . (c) Sum of the relative errors  $\frac{|M^{ME} - M^{VM}|}{|M^{VM}|} + \frac{|R^{VM}(\Delta) - R^{ME}(\Delta)|}{R^{VM}(\Delta)}$  as functions of  $H_x$  and  $J$ . (d) Absolute deviation between the inferred  $\tau_{ME} := \operatorname{argmax}_\tau |J_{ME}(\tau)|$  and the planted parameter  $\Delta$  as a function of  $J$  and  $H_x$ . The analogous figure for the metric case is reported in the Appendix.

using these collected positions  $\mathbf{x}_i^t$ , we can evaluate the velocity  $\mathbf{v}$  and direction of motion  $s$  of each particle  $i$  at time  $t$ :

$$\mathbf{v}_i^t \equiv \frac{\mathbf{x}_i^{t+1} - \mathbf{x}_i^t}{\delta t}, \quad s_i^t \equiv \frac{\mathbf{v}_i^t}{|\mathbf{v}_i^t|}, \quad 1 \leq i \leq N, \quad 1 \leq t \leq N_T. \quad (\text{A2})$$

Focusing on the collection of variables  $s \equiv \{s_i^t\}$ ,  $1 \leq i \leq N$ ,  $1 \leq t \leq N_T$ , a natural framework to determine a probability measure for these units is the path-integral formalism [10,11]: we define the path entropy  $S$  as

$$S[P] = - \int \mathcal{D}s P(s) \log P(s), \quad (\text{A3})$$

where  $\mathcal{D}s \equiv \prod_{i=1}^{N_T} \prod_{i=1}^N ds_i^t$  is the integration measure and each  $ds_i^t$  is a surface element of the  $(D-1)$ -dimensional sphere. The average alignment of the moving units at time  $t$  is given by

$$\mathbf{m}(s^t) \equiv \frac{1}{N} \sum_{i=1}^N s_i^t, \quad (\text{A4})$$

where

$$s^t \equiv \{s_i^t\}, \quad 1 \leq i \leq N. \quad (\text{A5})$$

The latter can be interpreted as a configuration of soft spins that evolve in time, thus, we will refer to  $\mathbf{m}(s^t)$  as the

magnetization at time  $t$ . Furthermore, to quantify the direction along which the units move on average, we introduce the temporally averaged magnetization  $\mathbf{M}$ :

$$\mathbf{M}(s) \equiv \frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} \mathbf{m}(s^t). \quad (\text{A6})$$

In our analysis, we assume that the motion of the units above can be temporally correlated up to a maximum temporal window whose length is  $\tau_M$  with  $0 \leq \tau_M < N_T$ . A natural observable to quantify this persistence is the two-point correlation function  $R$  with delay  $\tau$

$$R(\tau, s) \equiv \frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} \mathbf{m}(s^t) \cdot \mathbf{m}(s^{t-\tau}), \quad 1 \leq \tau \leq \tau_M. \quad (\text{A7})$$

For the moment we suppose that the correct value for  $\tau_M$  is known a priori; in the following sections, after having developed the necessary tools, we also provide a recipe for selecting  $\tau_M$  in a self-consistent manner for any dataset. Notice that the expression in Eq. (A7) implements a “mean-field” correlation as the direction of each unit  $i$  is related to the direction of any other unit up to  $\tau_M$  time steps before, regardless of their spatial distance.

Following the ME principle [5], we seek  $P$  as the minimal probability measure whose average magnetization (A6) and

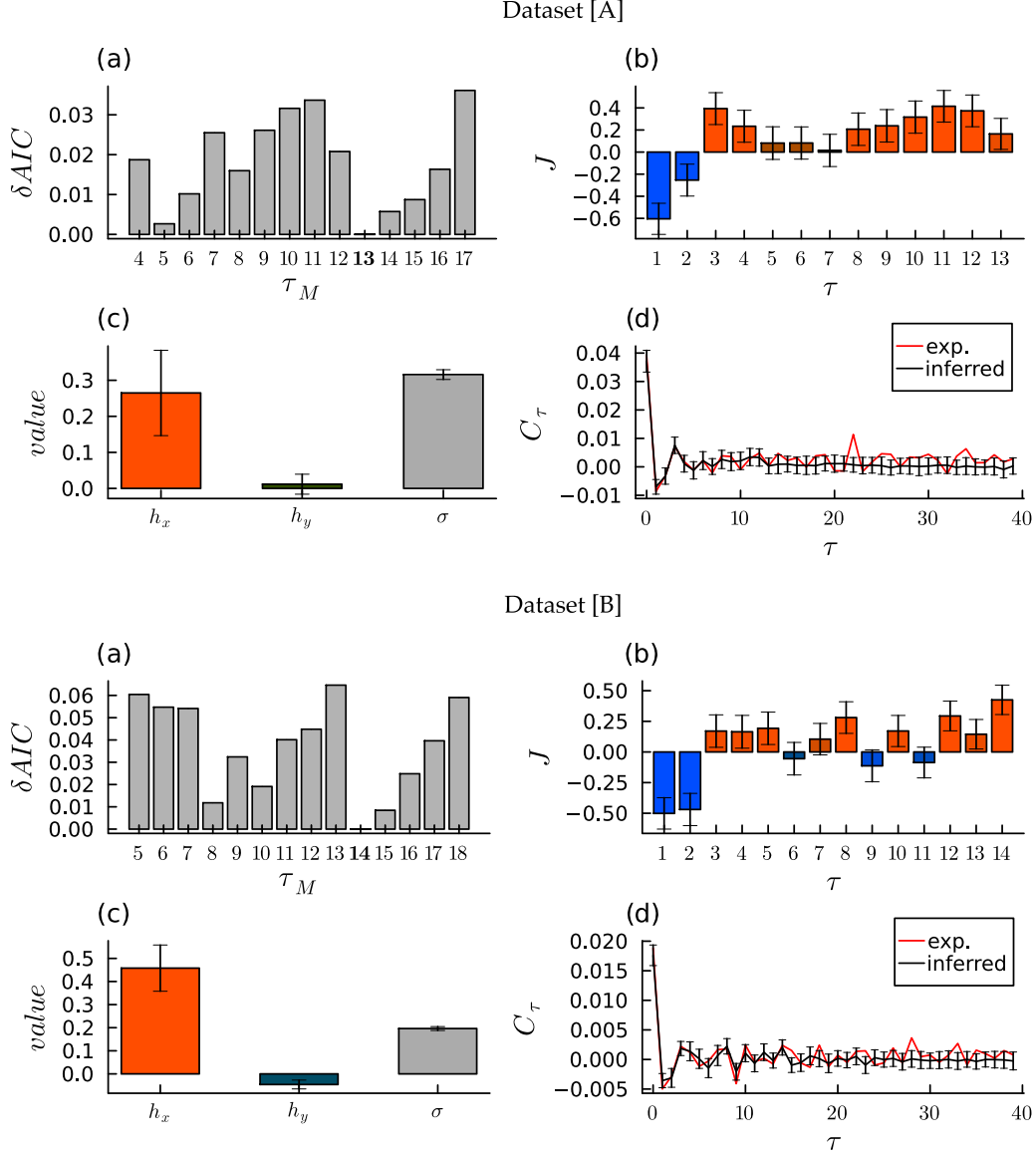


FIG. 4. Inference for the chemokine-gradient dataset ( $\delta t = 2$  min), for both regions A and B. (a) Difference between the AIC and its minimal value, as a function of the total number of interactions  $\tau_M$ . (b) Inferred values of the delayed interaction  $J_\tau$  and corresponding errors. (c) Inferred values of the two components of the field,  $H_x$  and  $H_y$ , and inferred value of  $\sigma$ . (d) Correlation function  $C(\tau)$ ; see Eq. (27).

two-point correlation function (A7) match their respective empirical values:

$$\max_P S[P], \quad (\text{A8})$$

subject to

$$\int \mathcal{D}s P(s) \mathbf{M}(s) = \mathbf{M}^E, \quad (\text{A9})$$

$$\int \mathcal{D}s P(s) R(\tau, s) = R^E(\tau), \quad 1 \leq \tau \leq \tau_M, \quad (\text{A10})$$

$$\int \mathcal{D}s P(s) = 1, \quad (\text{A11})$$

where  $\mathbf{M}^E$  and  $R^E$  are the empirical values of the average magnetization and two-point correlation, respectively.

To solve Eq. (A8), we use the method of Lagrangian multipliers. The Lagrangian  $S^*$  reads

$$\begin{aligned} S^*[P, \mathbf{J}, \mathbf{H}] = & -\frac{1}{N(N_T - \tau_M)} \int \mathcal{D}s P[s] \log P[s] \\ & + \sum_{\tau=1}^{\tau_M} J_\tau \left( \int \mathcal{D}s P[s] R(\tau, s) - R^E(\tau) \right) \\ & + \mathbf{H} \cdot \left( \int \mathcal{D}s P(s) \mathbf{M}(s) - \mathbf{M}^E \right) \\ & + \frac{\zeta}{N(N_T - \tau_M)} \left( \int \mathcal{D}s P[s] - 1 \right), \quad (\text{A12}) \end{aligned}$$

where  $\mathbf{J} \equiv \{J_\tau\}_{\tau=1}^{\tau_M}$ ,  $\mathbf{H} \equiv \{H_\ell\}_{\ell=1}^D$ ,  $\zeta$  play as Lagrangian multipliers and the normalization factor  $\frac{1}{N(N_T - \tau_M)}$  ensures that  $S^*$  is an intensive quantity with respect to  $N$  and  $N_T$ , and that

the final expression for  $P(\mathbf{s})$  is well defined (*vide infra*). The Lagrangian can be easily extremized with respect to  $P$ :

$$\frac{\delta S^*}{\delta P} = -\frac{1}{N(N_T - \tau_M)} (\log P(\mathbf{s}) + 1 - \zeta) + \sum_{\tau=1}^{\tau_M} J_\tau R(\tau, \mathbf{s}) + \mathbf{H} \cdot \mathbf{M}(\mathbf{s}) = 0, \quad (\text{A13})$$

yielding

$$P(\mathbf{s}) = \frac{1}{Z} \exp \left\{ N(N_T - \tau_M) \left[ \sum_{\tau=1}^{\tau_M} J_\tau R(\tau, \mathbf{s}) + \mathbf{H} \cdot \mathbf{M}(\mathbf{s}) \right] \right\}, \quad (\text{A14})$$

where the Lagrangian multipliers  $\mathbf{J}, \mathbf{H}$  are determined implicitly by Eqs. (A9) to (A11), and  $Z \equiv \exp\{(1 - \zeta)/[N(N_T - \tau_M)]\}$ . By substituting Eqs. (A6) and (A7) into Eq. (A14), we get

$$P(\mathbf{s}) = \frac{1}{Z} \exp \left[ N \sum_{t=\tau_M+1}^{N_T} \mathbf{m}(s^t) \cdot \left( \sum_{\tau=1}^{\tau_M} J_\tau \mathbf{m}(s^{t-\tau}) + \mathbf{H} \right) \right]. \quad (\text{A15})$$

Finally, by plugging Eq. (A4) into Eq. (A15), we obtain

$$P(\mathbf{s}) = \frac{1}{Z} \exp \left[ \sum_{t=\tau_M+1}^{N_T} \sum_{i=1}^N s_i^t \cdot \left( \frac{1}{N} \sum_{\tau=1}^{\tau_M} J_\tau \sum_{j=1}^N s_j^{t-\tau} + \mathbf{H} \right) \right], \quad (\text{A16})$$

where the normalization constant  $Z$  reads

$$Z = \int \mathcal{D}\mathbf{s} \exp \left[ \sum_{t=\tau_M+1}^{N_T} \sum_{i=1}^N s_i^t \cdot \left( \frac{1}{N} \sum_{\tau=1}^{\tau_M} J_\tau \sum_{j=1}^N s_j^{t-\tau} + \mathbf{H} \right) \right]. \quad (\text{A17})$$

We emphasize that the resulting  $P(\mathbf{s})$  can be interpreted as a Boltzmann-Gibbs distribution whose exponential defines an effective energy function that, consistently with thermodynamic principles, scales linearly with the number of degrees of freedom  $N$  and  $N_T$ . More generally, the structure of  $P(\mathbf{s})$  suggests that the direction of the  $i$ th particle at time  $t$  tends to be aligned with the direction given by  $\mathbf{H}$ —which plays as an external field—and that, if  $J_\tau > 0$  (resp.  $J_\tau < 0$ ), the direction of the  $i$ th particle at time  $t$  tends to be aligned (resp. misaligned) with the average unit's direction at time  $(t - \tau)$ —which plays as an internal field.

## 2. Solution method

In this section we present our strategy to obtain an estimate for the parameters  $\mathbf{J}$  and  $\mathbf{H}$ , starting from a sample of experimental data. We first recast the problem into a maximum likelihood setting, which allows us to implement a simplification in the expression for  $P(\mathbf{s})$ , yielding to an approximated expression denoted as  $P_A(\mathbf{s})$ . This approximation is made necessary by the prohibitive difficulty of averaging over the probability measure (A16), since it involves handling simultaneously all the correlated variables over their past history, as standard in path-integral formulations. In fact,

the idea is to replace the mean expectation of the internal field produced by the peer units and appearing in the exponent of the Boltzmann-Gibbs representation (A16), with its empirical evaluation. As we will show, this makes the measure  $P_A(\mathbf{s})$  factorized in such a way that we can recast the Lagrangian constraints in an approximated linear system that is feasible for a straightforward evaluation of the parameters  $\mathbf{J}$  and  $\mathbf{H}$ .

Let us implement the plan. Suppose that we experimentally observe a system of  $N$  particles at regular time intervals for a total of  $N_T + 1$  timepoints, then, according to Eq. (A2), we can evaluate the velocity  $\mathbf{v}$  and the direction  $\mathbf{s}$  for each particle and time point, overall collecting the dataset:

$$\mathbf{s}^E = \{\mathbf{s}_i^{E,t} \mid 1 \leq i \leq N, \quad 1 \leq t \leq N_T\}, \quad (\text{A18})$$

hence, we can calculate

$$\mathbf{m}_t^E \equiv \mathbf{m}^E(\mathbf{s}^{E,t}) = \frac{1}{N} \sum_{i=1}^N \mathbf{s}_i^{E,t}, \quad (\text{A19})$$

$$\mathbf{M}^E \equiv \mathbf{M}^E(\mathbf{s}^E) = \frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} \mathbf{m}^E(\mathbf{s}^{E,t}), \quad (\text{A20})$$

$$R^E(\tau) = \frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} \mathbf{m}^E(\mathbf{s}^{E,t}) \cdot \mathbf{m}^E(\mathbf{s}^{E,t-\tau}), \quad (\text{A21})$$

where the superscript E highlights that the quantity is evaluated by experimental data and in the last definition we wrote  $R^E(\tau) \equiv R^E(\tau, \mathbf{s}^E)$  to lighten notation. To estimate the parameters  $\mathbf{J}$  and  $\mathbf{H}$  appearing in Eq. (A16), and thus build a model for the dataset  $\mathbf{s}^E$ , we can use the method of maximum likelihood [18]. The log-likelihood  $l$  of the model (A16) given the dataset  $\mathbf{s}^E$  is

$$l(\mathbf{J}, \mathbf{H} | \mathbf{s}^E) \equiv \log P(\mathbf{s}^E | \mathbf{J}, \mathbf{H}) = \sum_{t=\tau_M+1}^{N_T} \sum_{i=1}^N s_i^{E,t} \cdot \left( \frac{1}{N} \sum_{\tau=1}^{\tau_M} J_\tau \sum_{j=1}^N s_j^{E,t-\tau} + \mathbf{H} \right) - \log \int \mathcal{D}\mathbf{s} e^{[\sum_{t=\tau_M+1}^{N_T} \sum_{i=1}^N s_i^t \cdot (\frac{1}{N} \sum_{\tau=1}^{\tau_M} J_\tau \sum_{j=1}^N s_j^{t-\tau} + \mathbf{H})]}. \quad (\text{A22})$$

The values of  $\mathbf{J}$  and  $\mathbf{H}$  that maximize  $l(\mathbf{J}, \mathbf{H} | \mathbf{s}^E)$  represent the maximum likelihood estimate for the parameters. We therefore derive  $l(\mathbf{J}, \mathbf{H} | \mathbf{s}^E)$  in Eq. (A22) and solve for

$$\frac{\partial l}{\partial \mathbf{J}} = \mathbf{0}, \quad \frac{\partial l}{\partial \mathbf{H}} = \mathbf{0}. \quad (\text{A23})$$

However, such a direct approach requires the evaluation of high-dimensional multi-variable integrals, like the one appearing in the last term of Eq. (A22), which entail several numerical issues as for the stability of the solution and the computational time (e.g., in our experimental datasets the number of correlated variables, for which we have to perform integrals, is  $\mathcal{O}(10^5)$ , being  $N_T = 150$  and  $N = 300$ ). Thus, we resort to an approximation: First, in Eq. (A16) we replace the internal field, acting on  $s_i^t$ , with its empirical



counterpart

$$\frac{1}{N} \sum_{\tau=1}^{\tau_M} J_{\tau} \sum_{j=1}^N s_j^{t-\tau} \rightarrow \sum_{\tau=1}^{\tau_M} J_{\tau} m_{t-\tau}^E, \quad (\text{A24})$$

along the lines of the mean-field approximation in statistical mechanics.

With the substitution (A24), pairs like  $s_i^t s_j^{t-\tau}$  no longer appear in Eq. (A22), that is, in a statistical-mechanics jargon, the original two-body model has been recast into a one-body model. Also, with this passage, the dependence displayed by the probability distribution on the first tranche of trajectories, i.e.,  $s_i^t$ ,  $t = 1, \dots, \tau_M$  is lost, but it can be restored, at least formally, by introducing some constraints that set the missing variables to their experimental values, in such a way that  $P_A(s)$  remains defined over the whole set of variables:

$$P(s) \rightarrow P_A(s) \equiv \frac{1}{Z_A} \prod_{i=1}^N \left\{ \prod_{t=\tau_M+1}^{N_T} e^{[s_i^t \cdot (\sum_{\tau=1}^{\tau_M} J_{\tau} m_{t-\tau}^E + \mathbf{H})]} \right\} \cdot \left[ \prod_{t=1}^{\tau_M} \delta(s_i^t - s_i^{E,t}) \right], \quad (\text{A25})$$

with

$$Z_A \equiv \int \left( \prod_{t=\tau_M+1}^{N_T} \prod_{i=1}^N ds_i^t \right) \exp \left[ s_i^t \cdot \left( \sum_{\tau=1}^{\tau_M} J_{\tau} m_{t-\tau}^E + \mathbf{H} \right) \right], \quad (\text{A26})$$

where the subscript A stands for *approximated*. In this way, we are still able to evaluate all the momenta of the variables  $s$ , including those involving  $s_i^t$ ,  $t = 1, \dots, \tau_M$ .

We denote by  $\langle \rangle_A$  the average performed with respect to  $P_A$ , i.e., for the generic observable  $F(s)$ ,

$$\langle F(s) \rangle_A \equiv \int \mathcal{D}s P_A[s] F(s), \quad (\text{A27})$$

and note that

$$\langle m_t \rangle_A = \int \left( \prod_{t=\tau_M+1}^{N_T} \prod_{i=1}^N ds_i^t \right) P_A(s) \left( \frac{1}{N} \sum_{i=1}^N s_i^t \right) = \begin{cases} \mathcal{M}(\sum_{\tau=1}^{\tau_M} J_{\tau} m_{t-\tau}^E + \mathbf{H}) & \tau_M + 1 \leq t \leq N_T, \\ m_t^E & 1 \leq t \leq \tau_M, \end{cases} \quad (\text{A28})$$

where the vector function  $\mathcal{M}$  is defined as

$$\mathcal{M}(\mathbf{x}) \equiv \frac{\mathbf{x} \mathcal{I}_1(|\mathbf{x}|)}{|\mathbf{x}| \mathcal{I}_0(|\mathbf{x}|)}, \quad (\text{A29})$$

and  $\mathcal{I}_0, \mathcal{I}_1$  are hyperbolic Bessel functions (also known as modified Bessel function of the first kind) of order 0 and 1, respectively [12]. In addition, we note that  $P_A$  is factorized with respect to every variable  $s_i^t$ , thus any average of products of magnetizations at different times  $t$  and  $t'$  equals a product of averages:

$$\langle m_t m_{t'} \rangle_A = \langle m_t \rangle_A \langle m_{t'} \rangle_A. \quad (\text{A30})$$

The approximation (A25) and the property (A30) allow us to rearrange Eqs. (A9) and (A10) in the following form:

$$\frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} \langle m_t \rangle_A = \mathbf{M}^E = \frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} m_t^E, \quad (\text{A31})$$

$$\begin{aligned} \frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} \langle m_t \rangle_A \cdot \langle m_{t-\tau} \rangle_A &= R^E(\tau) \\ &= \frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} m_t^E \cdot m_{t-\tau}^E, \quad 1 \leq \tau \leq \tau_M, \end{aligned} \quad (\text{A32})$$

where we dropped the explicit dependence on  $s^{E,t}$  to lighten the notation. These equations are simultaneously fulfilled by requiring that

$$\langle m_t \rangle_A = m_t^E, \quad \tau_M + 1 \leq t \leq N_T. \quad (\text{A33})$$

In fact, given that Eq. (A33) is automatically satisfied for  $1 \leq t \leq \tau_M$ , if Eq. (A33) holds, then Eqs. (A31) and (A32) are satisfied too.

We stress that (A33) is only a *sufficient* condition for Eqs. (A31) and (A32) to hold since this condition introduces stricter constraints than those stemming from Eqs. (A31) and (A32). However, by moving from Eqs. (A31) and (A32) to Eq. (A33) we retain the very same parameters, that are  $\mathbf{H}$  and  $\mathbf{J}$ . Specifically, in this problem, the amount of data available  $N_T \times N$  is typically much larger than the number of parameters  $D + \tau_M$ . Thus, this stricter condition does not imply any overfitting (we are introducing constraints and not parameters) which, however, could be controlled by tuning  $\tau_M$ .

Exploiting Eq. (A28), Eq. (A33) can be rewritten as

$$\mathcal{M} \left( \sum_{\tau=1}^{\tau_M} J_{\tau} m_{t-\tau}^E + \mathbf{H} \right) = m_t^E, \quad \tau_M + 1 \leq t \leq N_T. \quad (\text{A34})$$

Since  $\mathcal{M}$  is a monotone function, its inverse function is well defined and we can rewrite Eq. (A34) as

$$\mathbf{G}_t^E \equiv \mathcal{M}^{-1}(m_t^E), \quad (\text{A35})$$

$$\sum_{\tau=1}^{\tau_M} J_{\tau} m_{t-\tau}^E + \mathbf{H} = \mathbf{G}_t^E, \quad \tau_M + 1 \leq t \leq N_T. \quad (\text{A36})$$

We observe that the operator  $\mathbf{G}_t^E$  is linear in the variables  $\mathbf{J}, \mathbf{H}$ : this property turns out to be crucial for the following developments. Let us underline that Eq. (A36) constitutes a linear system, whose solution provides us with the parameters  $\mathbf{J}, \mathbf{H}$ , but this system involves more constraints than parameters. More precisely, as a result of Eq. (A33), the number of parameters is still  $D + \tau_M$ , and the number of constraints is now  $(N_T - \tau_M) \times D$ . We can therefore enrich the model by stating that the relation (A36) does not hold deterministically and, thus, introduce a source of noise represented by a standard Gaussian variable tuned by the additional parameter  $\sigma$ . This stochastic term turns the linear system (A36) into an autoregressive model (A37) of order  $\tau_M$

$$\sum_{\tau=1}^{\tau_M} J_{\tau} m_{t-\tau}^E + \mathbf{H} + \sigma \epsilon_t = \mathbf{G}_t^E, \quad \tau_M + 1 \leq t \leq N_T, \quad (\text{A37})$$

$$\epsilon_t \sim \mathcal{N}(\mathbf{0}_D, \mathbf{1}_D), \quad \sigma \geq 0. \quad (\text{A38})$$

In other words, given the parameters  $\mathbf{J}$ ,  $\mathbf{H}$ , and  $\sigma$ , the magnetization  $\mathbf{m}_t$  at time  $t$  depends on the magnetization exhibited in the previous  $\tau_M$  times according to the evolution rule:

$$\mathbf{m}_t = \mathcal{M} \left[ \mathbf{H} + \sum_{\tau=1}^{\tau_M} J_\tau \mathbf{m}_{t-\tau} + \sigma \boldsymbol{\epsilon}_t \right]. \quad (\text{A39})$$

Now, the log-likelihood  $l_A(\mathbf{J}, \mathbf{H}, \sigma | \mathbf{s}^E)$  related to the autoregressive model is introduced as

$$-\frac{1}{2\sigma^2} \sum_{t=\tau_M+1}^{N_T} \left| \sum_{\tau=1}^{\tau_M} J_\tau \mathbf{m}_{t-\tau}^E + \mathbf{H} - \mathbf{G}_t^E \right|^2 - (N_T - \tau_M) \frac{D}{2} \log(2\pi\sigma^2), \quad (\text{A40})$$

and, in the next section, we will extremize  $l_A(\mathbf{J}, \mathbf{H}, \sigma | \mathbf{s}^E)$  to obtain an estimate for  $\mathbf{J}$ ,  $\mathbf{H}$ , and  $\sigma$ .

#### a. Analytical solution for the estimated parameters

As customary in maximum likelihood estimations, we evaluate the gradient of the log-likelihood  $l_A(\mathbf{J}, \mathbf{H}, \sigma | \mathbf{s}^E)$  and set it to zero to estimate  $\mathbf{J}$ ,  $\mathbf{H}$ , and  $\sigma$ . To this goal, we first compute the derivatives of Eq. (A40),

$$\nabla_{\mathbf{H}} l_A = -\frac{N_T - \tau_M}{\sigma^2} \left\langle \mathbf{H} + \sum_{\tau=1}^{\tau_M} J_\tau \mathbf{m}_{t-\tau}^E - \mathbf{G}_t^E \right\rangle_t, \quad (\text{A41})$$

$$\nabla_{J_\tau} l_A = -\frac{N_T - \tau_M}{\sigma^2} \left\langle \left( \mathbf{H} + \sum_{\lambda=1}^{\tau_M} J_\lambda \mathbf{m}_{t-\lambda}^E - \mathbf{G}_t^E \right) \cdot \mathbf{m}_{t-\tau}^E \right\rangle_t, \quad (\text{A42})$$

$$\nabla_{\sigma} l_A = \frac{N_T - \tau_M}{\sigma^3} \left\langle \left| \sum_{\tau=1}^{\tau_M} J_\tau \mathbf{m}_{t-\tau}^E + \mathbf{H} - \mathbf{G}_t^E \right|^2 \right\rangle_t - (N_T - \tau_M) \frac{D}{\sigma}, \quad (\text{A43})$$

where

$$\langle \cdot \rangle_t \equiv \frac{1}{N_T - \tau_M} \sum_{t=\tau_M+1}^{N_T} \quad (\text{A44})$$

denotes the time average.

Setting Eq. (A41) to zero and solving for  $\mathbf{H}$ , we get

$$\mathbf{H} = \langle \mathbf{G}_t^E \rangle_t - \sum_{\tau=1}^{\tau_M} J_\tau \langle \mathbf{m}_{t-\tau}^E \rangle_t. \quad (\text{A45})$$

Proceeding along the same lines, from Eq. (A42), for  $\mathbf{J}$  we obtain

$$\sum_{\tau=1}^{\tau_M} A_{\lambda\tau} J_\tau = B_\lambda \Leftrightarrow \mathbf{J} \equiv \mathbf{A}^{-1} \cdot \mathbf{B}, \quad (\text{A46})$$

where the  $\tau_M \times \tau_M$  matrix  $A_{\lambda\tau}$  and the  $\tau_M$ -dimensional vector  $B_\lambda$  are

$$A_{\lambda\tau} \equiv \langle \mathbf{m}_{t-\tau}^E \cdot \mathbf{m}_{t-\lambda}^E \rangle_t - \langle \mathbf{m}_{t-\tau}^E \rangle_t \cdot \langle \mathbf{m}_{t-\lambda}^E \rangle_t, \quad (\text{A47})$$

$$B_\lambda \equiv \langle \mathbf{G}_t^E \cdot \mathbf{m}_{t-\lambda}^E \rangle_t - \langle \mathbf{G}_t^E \rangle_t \cdot \langle \mathbf{m}_{t-\lambda}^E \rangle_t. \quad (\text{A48})$$

Finally, setting  $\nabla_{\sigma} l_A = 0$ , we obtain the variance  $\sigma^2$  of the autoregressive process (A39):

$$\sigma^2 = \frac{1}{D} \left\langle \left| \mathbf{H} + \sum_{\tau=1}^{\tau_M} J_\tau \mathbf{m}_{t-\tau}^E - \mathbf{G}_t^E \right|^2 \right\rangle_t. \quad (\text{A49})$$

#### b. Errors on the inferred measurements

To estimate the errors for the inferred parameters in the maximum likelihood estimation, we exploit the Fisher information matrix (see, e.g., Ref. [13]). We recall that, given some dataset  $\mathbf{X}$  where each observation  $\mathbf{x}_i$  is assumed to be identically and independently distributed according to a true underlying distribution, and being  $f_\theta(\mathbf{x})$  a model probability density function parametrized by  $\theta$ , we can write the log-likelihood function as  $l(\theta | \mathbf{x}) = \sum_{i=1}^n \log f_\theta(\mathbf{x}_i)$ , where  $n$  is the sample size. Then, the (empirical) Fisher information matrix  $\mathbf{I}$  has elements given by

$$\mathbf{I}_{\theta_a, \theta_b} \equiv - \int \exp[l(\theta | \mathbf{x})] \frac{\partial^2 l(\theta | \mathbf{x})}{\partial \theta_a \partial \theta_b} d\mathbf{x}, \quad (\text{A50})$$

namely its elements correspond to the expectation of the elements of the Hessian matrix of the log-likelihood. Let us suppose that the true parameter is  $\theta_0$ , and that the maximum-likelihood estimate of  $\theta_0$  is  $\theta^* = \text{argmax}_\theta l(\theta | \mathbf{x})$ . Then, one can prove that  $\theta^* \sim \mathcal{N}(\theta_0, \mathbf{I}^{-1}(\theta_0))$ . Since when the sample size approaches infinity, the maximum-likelihood estimate approaches the true parameter (this is also known as the consistency property), we can write that the covariance matrix for the estimated parameters is just

$$\boldsymbol{\Sigma}_{\theta^*} \underset{n \gg 1}{\approx} \mathbf{I}^{-1}(\theta^*). \quad (\text{A51})$$

We will use this result to provide the error estimates for our model. By Eq. (A40) the matrix elements of the Fisher information read

$$I_{H_a, H_b} = \frac{N_T - \tau_M}{\sigma^2} \delta_{ab}, \quad (\text{A52})$$

$$I_{J_\tau, H} = \frac{N_T - \tau_M}{\sigma^2} \langle \mathbf{m}_{t-\tau}^E \rangle_t, \quad (\text{A53})$$

$$I_{J_\tau, J_\lambda} = \frac{N_T - \tau_M}{\sigma^2} \langle \mathbf{m}_{t-\tau}^E \cdot \mathbf{m}_{t-\lambda}^E \rangle_t, \quad (\text{A54})$$

$$I_{\sigma, \sigma} = \frac{N_T - \tau_M}{\sigma^2} 2D, \quad (\text{A55})$$

$$I_{\sigma, J_\tau} = I_{\sigma, H} = 0, \quad (\text{A56})$$

and Eq. (A50) becomes

$$\mathbf{I} = \frac{N_T - \tau_M}{\sigma^2} \begin{pmatrix} \langle \mathbf{m}_{t-\tau}^E \cdot \mathbf{m}_{t-\lambda}^E \rangle_t & \langle \mathbf{m}_{t-\tau}^E \rangle_t & \mathbf{0} \\ \langle \mathbf{m}_{t-\lambda}^E \rangle_t & \delta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 2D \end{pmatrix}. \quad (\text{A57})$$

Finally, by substituting Eq. (A57) into Eq. (A51), we obtain the error estimates for  $\mathbf{J}$ ,  $\mathbf{H}$ , and  $\sigma$ :

$$\text{Var}(\mathbf{J}, \mathbf{H}, \sigma) = \text{diag}(\boldsymbol{\Sigma}_{\theta^*}). \quad (\text{A58})$$

In the section dedicated to the model validation, we use this relation to test the goodness of our method on synthetic datasets.

### c. Effective number of delayed interactions

So far, in our analysis, we assumed that the parameter  $\tau_M$  is a known constant. In most practical inference scenarios, where one needs to find from scratch the best model for a given dataset, the number of the effective delayed interactions  $\tau_M$  that must be considered is unknown: in what follows, we will provide a way to estimate it.

To tackle this problem, we rely on the Akaike information criterion (AIC) [9], which is used in the literature to probe the flexibility of an autoregressive model. In fact, given a dataset and a model with a mean-squared error  $\epsilon^2$ ,  $n_p$  parameters and  $n_o$  observations, the quantity

$$\text{AIC} \equiv \frac{2n_p}{n_o} + \log(\epsilon^2) \quad (\text{A59})$$

estimates the prediction error of the model. Thus, having developed several models for a certain dataset, the AIC estimates the quality of each model relative to all others, hence providing a tool for model selection. Note that, by definition, the AIC favours those models yielding a small error *and* a small number of parameters.

In the analysis above, the parameters appearing in Eq. (A59) are

$$n_o = D(N_T - \tau_M), \quad n_p = \tau_M + D + 1, \quad \epsilon^2 = \frac{N_T}{N_T - \tau_M} \sigma^2. \quad (\text{A60})$$

Thus,

$$\text{AIC} = 2 \frac{\tau_M + D + 1}{D(N_T - \tau_M)} + \log \left( \frac{N_T \sigma^2}{N_T - \tau_M} \right). \quad (\text{A61})$$

To summarize, the protocol followed in the next sections is the following: we will perform inference for multiple values of  $\tau_M$  and we solve for the fields and the couplings via Eqs. (A45), (A46). We further determine  $\sigma^2$  from Eq. (A49), then we will evaluate Eq. (A61) for each of the inferred models, and we will select the value of  $\tau_M$  that implies the smallest estimate of AIC. Finally, the maximal  $J_\tau$  in the selected time window  $\tau \in [0, \tau_M]$  turns out to be the best estimate (inferred value) of the actual coupling (planted value).

## APPENDIX B: MODEL VALIDATION

In the remainder of this Appendix, we test our inference method on synthetic and real datasets: we consider two synthetic datasets, generated from the HKM and from the VM, and a biological dataset, collected via time-lapse microscopy on dendritic cell migration in a chemoattractant field (see Ref. [4] for details).

In general, we first run the inference protocol on such datasets and obtain the parameters  $\mathbf{J}$  and  $\mathbf{H}$ , which we call *inferred* parameters, and then we evaluate the goodness of the inference process. As for the HKM and the VM, previously addressed, a validation is possible by comparing the inferred parameters and those set in the model under consideration and denoted as *planted* parameters. The HKM is the direct

formulation of the inverse problem developed in this paper, thus, in this setting, *inferred* and *planted* parameters must match even quantitatively (and their agreement is a *conditio sine qua non* to move to more challenging scenarios); in fact, in this case, inference is excellent. Conversely, for the VM, the inference process can be harder as this model is not the direct model of the delayed ME approach we developed, and, to investigate the problem meticulously, we face two versions of the delayed VM, namely, the topological and the metric ones. In both cases we have a set of units whose dynamics is non-Markovian and influenced by the trajectories of a subset of peers, made of *i.* the closest  $n_c$  units (regardless of their reciprocal distance) in the former version, or *ii.* the closest units within a certain interaction range  $r$  (regardless of their number) in the latter version.

Finally, the analysis of the biological datasets aims to be a proof of concept to highlight the non-Markovianity of biological dynamics: this is why we consider for the test study the migration of dendritic cells as the latter are expected to perceive signaling proteins that must diffuse from the emitter cell to the receiver one, thus, these interactions are natural prototypes of delayed interactions.

We anticipate that, beyond successfully testing our delayed-inference protocol on all the synthetic datasets, we obtained a remarkable result in inspecting the biological ones: the trajectory of a given dendritic cell at time  $t$  turns out to be influenced by those of the other cells for a past temporal window spanning back up to  $\tau_M \sim \mathcal{O}(10)$  time-steps, thus resulting in severe violation of the Markovian assumption.

### 1. Synthetic dataset: Heisenberg-Kuramoto model

The HKM with delayed interactions can be defined by interpreting the probability distribution (A16), with normalization given by Eq. (A17), as the Gibbs measure  $\exp(-\mathcal{H}(s|\mathbf{J}, \mathbf{H}))$  of the Hamiltonian

$$\mathcal{H}(s|\mathbf{J}, \mathbf{H}) = \sum_{i=1}^N s_i^t \cdot \left( \frac{1}{N} \sum_{\tau=1}^{\tau_M} J_\tau \sum_{j=1}^N s_j^{t-\tau} + \mathbf{H} \right), \quad (\text{B1})$$

where the dynamical variables are the spins  $s$ , while the parameters  $\mathbf{J}$  and  $\mathbf{H}$  represent, respectively, the interactions perceived with a delay  $\tau$  (back in the history up to  $\tau = \tau_M$ ) and the external field.

To setup the validation of our inferential protocol on the HKM we perform the following steps:

(1) We set  $D = 2$  and  $\tau_M = 4$ , and draw the planted parameters  $J_1, \dots, J_{\tau_M}, H_x, H_y$  independently and identically from  $U(-1, 1)$ , where  $U(a, b)$  denotes the uniform probability distribution between  $a$  and  $b$ .

(2) With the parameters just set, we sample the trajectories from the Boltzmann-Gibbs distribution related to the Hamiltonian (B1) from  $t = 0$  to  $t = 5 \times 10^3$ , and store the magnetizations from  $t = 1 \times 10^3$  to  $t = 5 \times 10^3$  to consider equilibrated snapshots.

(a) The simplest (despite rather lengthy) way of generating trajectories is to create the configuration  $s^t$  by collecting all the vectors  $s^{t-1}, \dots, s^{t-\tau_M}$  and inserting them into the Hamiltonian (B1). Then we generate, through its Boltzmann-Gibbs weight, the configuration  $s^t$  and so on for  $s^{t+1}$ ;

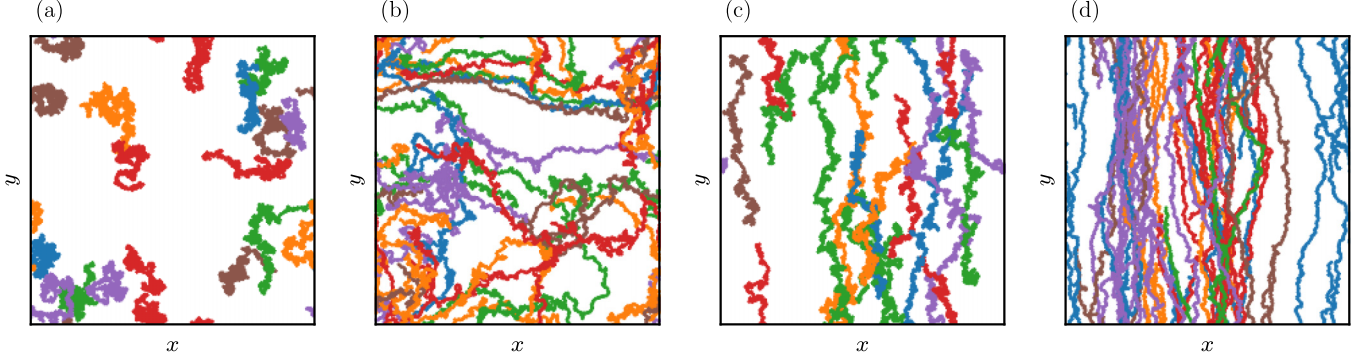


FIG. 5. Examples of trajectories generated by the VM. Simulations were implemented by letting  $N = 100$  particles free to move in a 2D square of linear size  $L = 150$  with periodic boundary conditions. Further, we set  $\delta t = 1$ ,  $\Delta = 6$  and  $n_c = 4$ . (a) Trajectories sampled at  $J = 0$ ,  $H_x = H_y = 0$  (as in a standard random walk). (b) Trajectories sampled at  $J > 0$ ,  $H_x = H_y = 0$  (as in a correlated random walk). (c) Trajectories sampled at  $J = 0$ ,  $H_x = 0$ ,  $H_y > 0$  (as in a drifted random walk). (d) Trajectories sampled at  $J > 0$  and  $H_x = 0$ ,  $H_y > 0$  (as in a correlated and drifted random walk).

(b) However, as we just need their averages (i.e.,  $\mathbf{m}_t$  as a function of  $\mathbf{m}_{t-1}, \dots, \mathbf{m}_{t-\tau_M}$ ), an alternative (faster) route consists in using the coarse-grained evolution coded by Eq. (A39) and iterate the latter (that is equivalent to the previous one as the model is mean-field).

Note that trajectories are not sampled *as a whole*, rather gradually built up time step by time step, such that their generation is not a unique high-dimensional sample from  $P(\mathbf{s})$ , rather a stochastic process with memory.

(3) We use the stored magnetizations to carry out the inference process according to our formulation [see Eq. (A45) for the fields, Eq. (A46) for the couplings, and Eq. (A49) for the variance], obtaining the estimates of  $J, H_x, H_y, J_1, \dots, J_{\tau_M}, \sigma$ .  $\tau_M$  is found by using the Akaike information criterion [see Eq. (A61) and its related discussion].

(4) We store the corresponding pairs of *inferred* and *planted* parameters.

This process is repeated 150 times and the collected data are summarized in Fig. 1. In particular, Fig. 1(a) shows that the planted and inferred parameters are arranged along the diagonal  $x = y$ , implying a quantitatively excellent estimate for parameters. In addition, Fig. 1(d) shows that the inference framework is able to exactly detect the extent of the time window  $[0, \tau_M]$  most of the times; although in some cases this is slightly overestimated, we verified that the magnitude of the extra interactions  $J_{\tau > \tau_{\text{planted}}}^{\text{inferred}}$  is statistically compatible with zero within the error on the planted data. This can be seen from Fig. 1(b): as  $J_{\tau > \tau_{\text{planted}}}^{\text{planted}} \equiv 0$ , if  $J_{\tau > \tau_{\text{planted}}}^{\text{inferred}} \neq 0$ , then its magnitude corresponds to the absolute error and in these cases we reveal only small errors [ $\sim \mathcal{O}(10^{-2})$ ].

## 2. Synthetic dataset: Vicsek model

### a. Topological Vicsek model with delay

In this section, we consider the topological version of the VM [15,17], equipped with a delay  $\Delta$  in signaling among different units. Such a model is defined by the following

dynamical equations:

$$D_{ij} = (\Delta - 1)(1 - \delta_{ij}^j), \quad (\text{B2})$$

$$\mathbf{v}_k^{t+1} = v_0 \Gamma \left( \frac{J}{|n_k^t|} \sum_{j \in n_k^t} \frac{\mathbf{v}_j^{t-D_{jk}}}{v_0} + \mathbf{H} + \sqrt{\delta t} \boldsymbol{\eta}_k^t \right), \quad (\text{B3})$$

$$\mathbf{x}_k^{t+1} = \mathbf{x}_k^t + \delta t \mathbf{v}_k^{t+1} \quad k = 1, \dots, N, \quad (\text{B4})$$

where  $\mathbf{v}_i^t, \mathbf{x}_i^t$  are the velocity and the position of the  $i$ th particle at time  $t$ , respectively.

Note that  $n_c$  and  $\{n_1^t, \dots, n_{n_c}^t\}$  are, respectively, the total number of topological neighbors and the set of topological neighbors of the  $i$ th particle, including the  $i$ th particle, at time  $t$ , and that, by selecting  $\Delta = 1$ , the Markovian limit of the standard VM is recovered. The normalization operator  $\Gamma$ ,

$$\Gamma(\mathbf{v}) \equiv \frac{\mathbf{v}}{|\mathbf{v}|}, \quad (\text{B5})$$

ensures that the magnitude of the velocity of each unit is kept constant. Further,  $\mathbf{H}$  is the external field applied to each particle, and  $\boldsymbol{\eta}_k^t \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$  is an i.i.d. noise. In this model, the particles are self-propelled, i.e.,  $|\mathbf{v}_j^t| = v_0$ . Note that, as discussed above, we denoted the planted parameter  $\mathbf{H}$  with the same symbol as the magnetic field in the ME distribution (A15), but they do not need to coincide. Finally, the magnetization of the VM is given by the expression

$$\mathbf{m}_t = \frac{1}{N v_0} \sum_{k=1}^N \mathbf{v}_k^t, \quad (\text{B6})$$

which is analogous to Eq. (A4).

Examples of trajectories generated by this VM are shown in Fig. 5.

The validation procedure for the ME inference in the case of the VM is different from that of the HKM: in the case of the VM we have  $J, H_x$  as free parameters. Despite these parameters play the same role of the coupling and the field in the HKM, the way they enter in the update rule of the units (B2)–(B4) is different from that of the ME model:

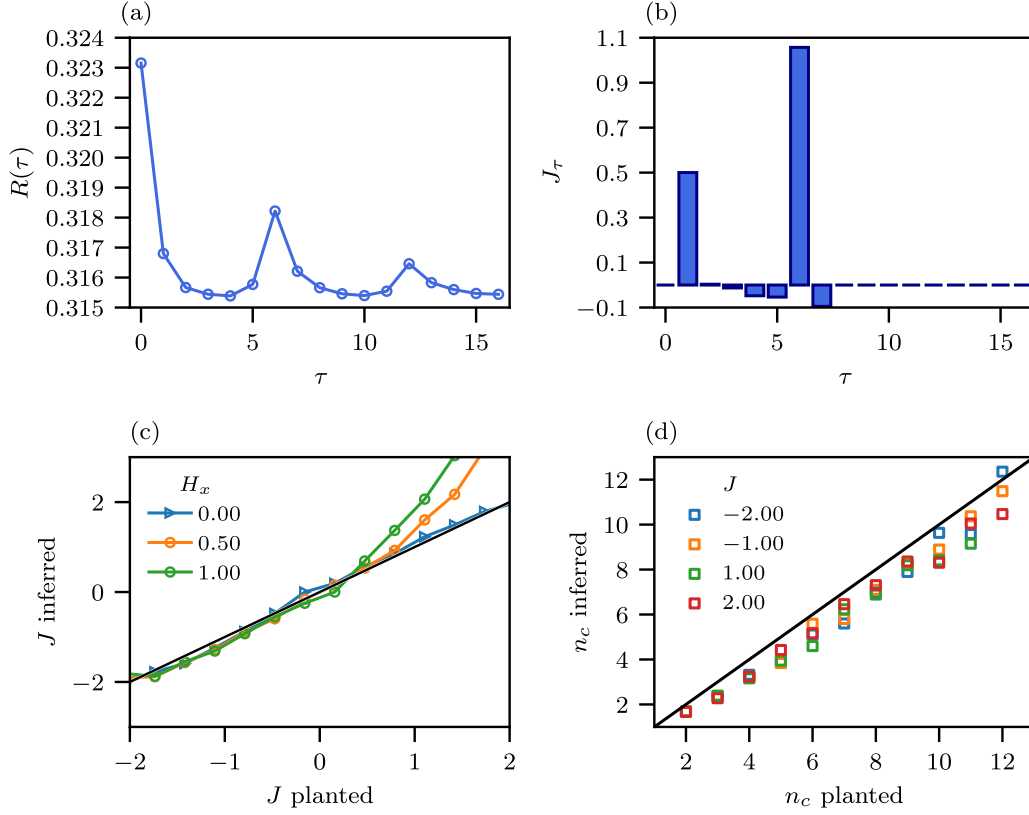


FIG. 6. Validation of our inference method with the metric Vicsek model (VM). The parameter setting is the same as in Fig. 2. (a) Example of the empirical correlation  $R(\tau)$  vs  $\tau$ . (b) Example of the (related) inferred coupling: note the presence of two peaks as in the topological case. (c) Scatter plot of the inferred vs the planted values of the coupling, for several values of the (external) field  $H_x$ , as explained in the legend. (d) Scatter plot of the nearest neighbors  $n_c$  where the inferred estimate is obtained as  $n_c = 1 + \frac{J_{\tau=\Delta}}{J_{\tau=1}}$ .

for the VM we lack a Hamiltonian formulation and direct and inverse problems are not *diagonal*, therefore a straight comparison between planted and inferred parameters is not expected to hold sharply. Nevertheless, the estimate of  $J_{\text{inferred}}$  and  $H_{\text{inferred}}$  is still very informative about the real values of the delayed couplings. In fact, as shown in Fig. 2(c), for  $J_{\text{planted}} = 0$  the algorithm correctly returns  $J_{\text{inferred}} = 0$ , then, for values of  $J_{\text{planted}} > 0$  ( $J_{\text{planted}} < 0$ ) the inference protocol returns positive (negative) values of  $J_{\text{inferred}}$ ; we also notice slight overestimates in the positive branch as the external field gets larger and larger and this should be ascribed to an intrinsic limitation of the present protocol.

As a sideline, we also stress that, since in our mean-field framework all units contribute equally, it is possible to obtain an estimate of the number of nearest neighbors  $n_c$ . Indeed, by a glance at Fig. 2(b) we see that there are mainly two contributions to the coupling, namely  $J_{(\tau=1)}$  and  $J_{(\tau=\Delta)}$ : the former is due to the inertia of the unit mass and this contribution is intrinsically Markovian, while the latter is due to the contribution of all the other units within  $n_c$ ; see Eqs. (B2)–(B4). Hence, we propose the formula  $n_c \approx 1 + J_{\tau=\Delta}/J_{\tau=1}$  to estimate the number of nearest neighbors and this looks like a reasonable one; see Fig. 2(d).

Beyond these evidences, a standard validation method in these nondiagonal problems is to compare the lowest-order moments of some observables evaluated from the original trajectories generated via the VM with those sampled from

the ME once the optimal values for  $\mathbf{J}$  and  $\mathbf{H}$  are learned. In particular, we measure the magnetization  $\mathbf{M}$  and the correlation function  $R(\tau = \Delta)$  in both models, and check if their values are statistically compatible. Indeed, they are as shown in Fig. 3.

Furthermore, we also inspect the Markovian limit, namely the standard topological VM, where we set  $\Delta = 1$  in Eq. (B2). In this case the inference protocol correctly returns that the trajectories are Markovian: we expect that only  $J_{(\tau=1)}$  is significantly different from zero among the inferred couplings  $J_\tau$  for all  $\tau \in (1, \dots, \tau_M)$  of the delayed ME protocol: this prediction is confirmed too (data not shown).

To summarize, in general, the validation procedure is carried out by means of the following algorithm:

- (1) Draw  $H_x$  from  $U(0, 2)$  and  $J$  from  $U(-2, +2)$ .
- (2) Run the dynamics (B2)–(B4) from  $t = 0$  to  $t = 2 \times 10^3$ , storing the magnetization at each step (excluding the first  $10^3$  that are used for equilibration).
- (3) Perform the ME inference process, using the magnetizations defined in Eq. (B6) as input and infer the couplings  $\mathbf{J}$  and the field  $\mathbf{H}$ .
- (4) Setting the parameters according to the previous step, simulate the ME model from  $t = 0$  to  $t = 2 \times 10^3$ , collecting the magnetization at each step (excluding the first  $10^3$  that are used for equilibration).
- (5) For both frameworks, evaluate the moments  $\mathbf{M}$  and  $R(\tau)$ , and calculate their relative difference.

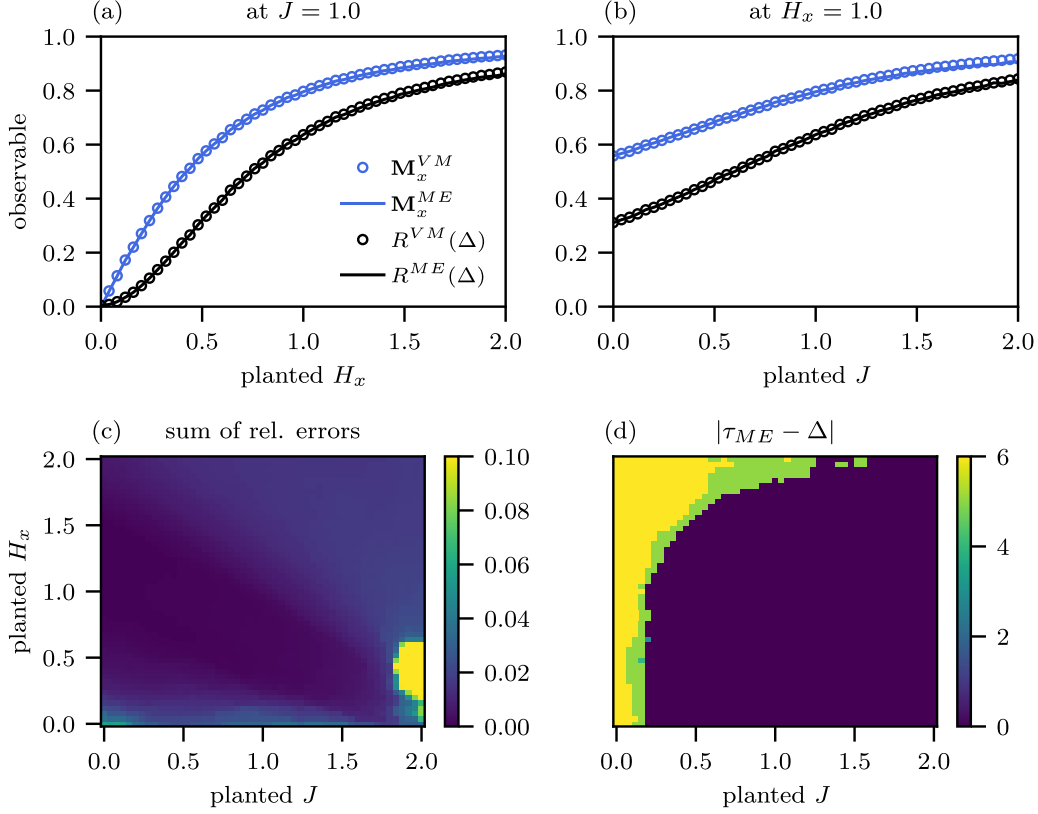


FIG. 7. Validation of our inference method with the metric Vicsek model (VM). The parameter setting is the same as in Fig. 2. The superscripts  $V$  and  $ME$  pertain to quantities calculated for the VM or inferred via the ME method, respectively. (a) Observables  $M$  and  $R(\Delta)$  as functions of  $H_x$ , for both models, with  $J = 1.0$ . (b) Observables  $M$  and  $R(\Delta)$  as functions of  $J$ , for both models, with  $H_x = 1.0$ . (c) Sum of the relative errors  $\frac{|M_{ME} - M_V|}{|M_V|} + \frac{|R_V(\Delta) - R_{ME}(\Delta)|}{R_V(\Delta)}$  as functions of  $H_x$  and  $J$ . (d) Absolute deviation between  $\tau_{ME} := \operatorname{argmax}_\tau |J_{ME}(\tau)|$  and the planted parameter  $\Delta$  as a function of  $J$  and  $H_x$ .

### b. Metric Vicsek model with delay

The solely difference between the metric version of the VM and its topological counterpart is that, for the former, the peers that contribute to the internal field for the velocity update of

the  $k$ th unit [see Eqs. (B2)–(B4)] are those units within a given radius and we fix the interaction range to  $r_{nc} := \sqrt{\frac{n_c L^2}{\pi N}}$ . The interaction range has been defined as a function of  $n_c$  to obtain comparable results for the two implementations.

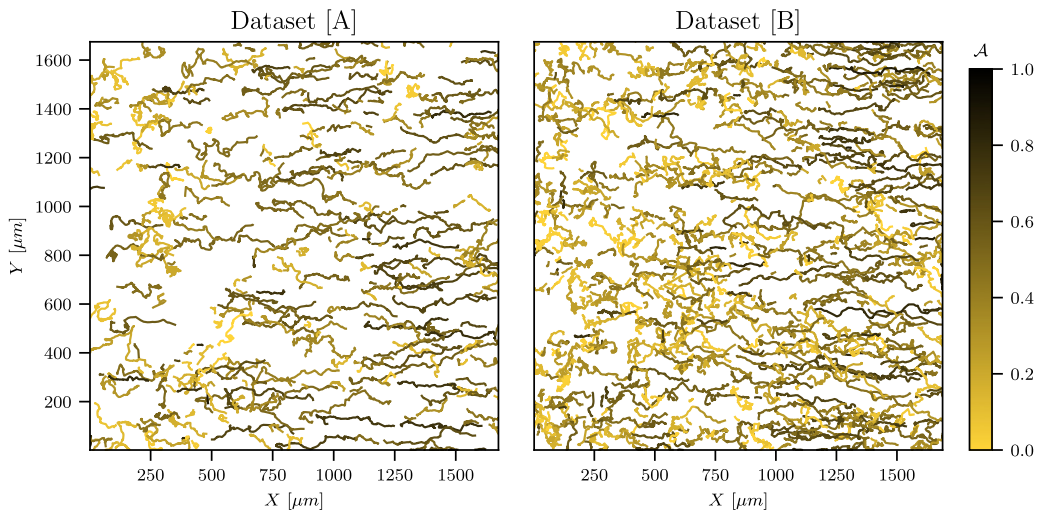


FIG. 8. Trajectories of each cell through the whole observation time. For each trajectory a parameter that we call anisotropy  $\mathcal{A} = 2 \frac{\max(|\Delta X_{\max}|, |\Delta Y_{\max}|)}{|\Delta X_{\max}| + |\Delta Y_{\max}|} - 1 \in [0, 1]$  has been measured and used to paint tracks with the most anisotropy in black, and the least anisotropic tracks in yellow. Datasets corresponding to region A and region B are represented, respectively, in the left and in the right panel.

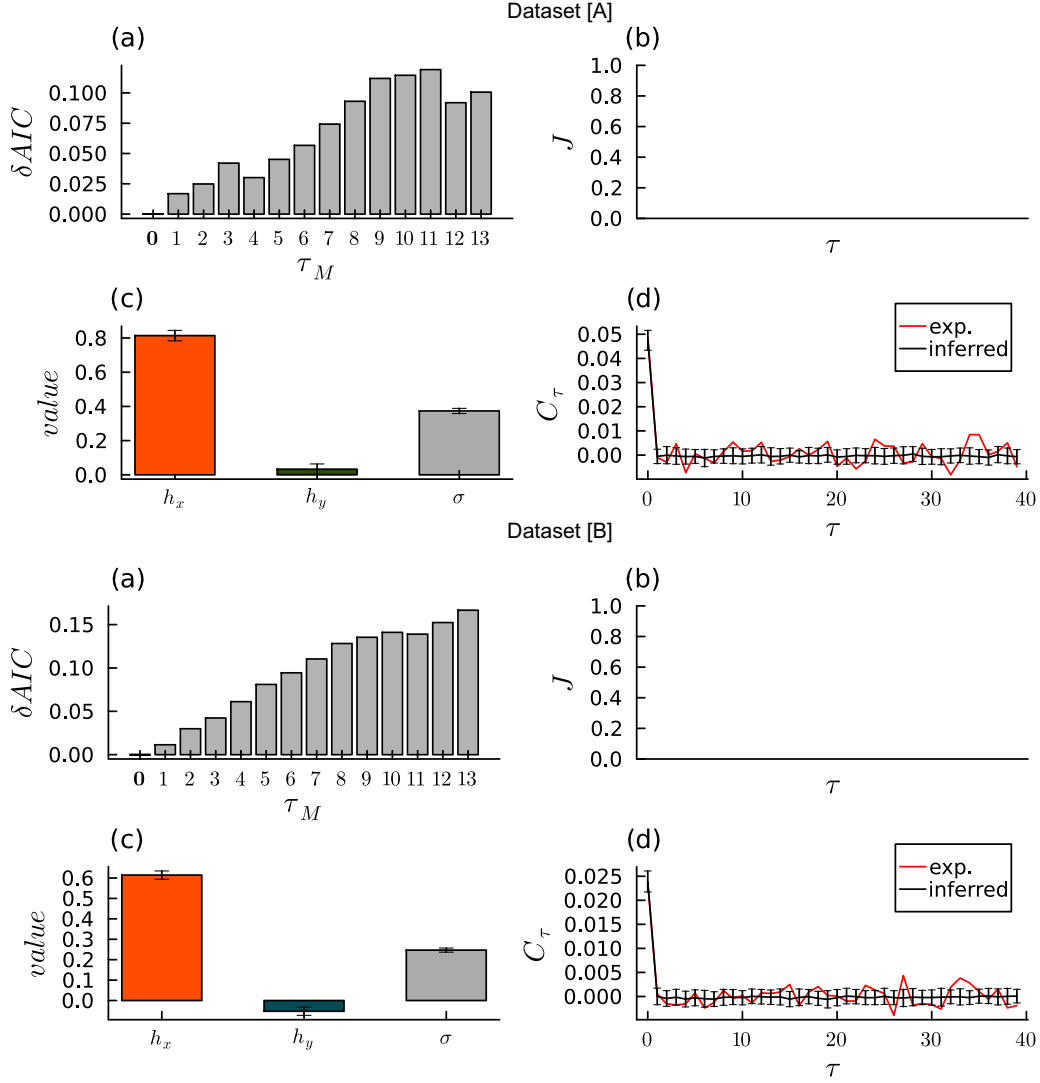


FIG. 9. Test on the validity of the delayed inference procedure: we reshuffled trajectories in the chemokine-gradient dataset, for both regions A and B, and gave this random dataset back to the inferential protocol to inspect its outcomes. (a) Difference between the AIC and its minimal value, as a function of the total number of interactions  $\tau_M$ ; note that this time the optimal  $\tau_M$  is zero, in agreement with a Bernoullian process. (b) Inferred values of the delayed interaction  $J_\tau$  and corresponding errors: note that, as expected, in this case there is no inferred coupling. (c) Inferred values for the two components of the field,  $H_x$  and  $H_y$ , and inferred value of  $\sigma$ . (d) Correlation function  $C(\tau)$ , see Eq. (B7), from the experimental data and from the ME model. As expected, while permutation invariance over the frames destroyed the temporal correlation of the motion (induced by the couplings) it did not erase the gradient of the chemoattractant, that is still well inferred.

Apart from this criterion for neighbor selection, the dynamical evolution is the same as in the previous case.

As in the topological counterpart, we perform inference of both coupling and field. Results are collected in Fig. 6 and a comparison of the lowest order statistics, e.g.,  $M_x^{VM}$  versus  $M_x^{ME}$  and  $R^{VM}$  versus  $R^{ME}$ , is presented in Fig. 7. As in the topological case, we can also inspect the number of units contributing to the internal field by the formula  $n_c \sim 1 + J_{(\tau=\Delta)}/J_{(\tau=1)}$ , yet, in this metric case, there is a (very mild) systematic underestimate of  $n_c$  beyond the over-estimate of the coupling in the presence of large field, as in the previous case [see Figs. 6(c) and 6(d)]. Finally, mirroring the previous scenario, we also inspected in this metric situation if the Markovian limit [i.e.,  $\Delta = 1$  in Eqs. (B2)–(B4)] is captured by our ME delayed inference protocol and this test turns out

again to be successful as the algorithm return  $J_{(\tau=1)} \neq 0$  only, as it should (data not shown).

### 3. Real dataset: Dendritic cells in a chemokine gradient

Having successfully tested our delayed ME inference on heterogeneous synthetic datasets, we now apply it to an experimental dataset on cell migration as a proof of concept. To be specific, in this section we consider a two-dimensional tracking experiment where leukocytes move in the presence of chemokines (i.e., signaling proteins for white cells), which guide cell migration by acting as a chemoattractant, see Fig. 8 (the higher values of the gradient are in the right side of the pictures that is where cells are migrating). Some chemokines control cells of the immune system during

immune-surveillance processes: for example, they may direct lymphocytes to lymph nodes so that they can control the invasion of pathogens, by interacting with antigen-presenting cells residing in these tissues: we consider this phenomenology as the most suitable to be investigated with this new inferential methodology. This is because, from one side, there is a vast network of signaling proteins (released by other cells) affecting immune dynamics [19], in such a way that these leukocytes have to integrate all these diffusing signals to decide the next direction of motion and this may result in a delayed interaction because the diffusion of these proteins is slow if compared for instance with electric signals that are instead used by neurons to communicate or the electromagnetic field, the light, used by flocks. From the other side, there is a vast literature that implements Markov processes as a natural starting point to model immunodynamics, see, e.g., Refs. [10,20–24], thus confirming—or otherwise disproving—that their motion can be effectively described by a Markov chain is an important question in theoretical immunology.

The system under investigation has been described in Ref. [10]: we collected ordered time series of the positions of dendritic cells migrating via chemotaxis toward a chemoattractant (i.e., dendritic cells perceive, and follow, a cytokine gradient released by the chemoattractant). Data was collected for regions A and B, thus yielding two distinct datasets; see Fig. 8.

We apply the inferential ME framework to datasets A and B separately. For the inferred ME model, we will show

$$\delta\text{AIC}(\tau_M) \equiv \text{AIC}(\tau_M) - \min_{\tau'_M} \text{AIC}(\tau'_M),$$

i.e., the difference between AIC for a given  $\tau_M$  and AIC for the optimal value of  $\tau_M$ , as well as the inferred parameters  $J_\tau$ ,

$H_x$ ,  $H_y$ , and  $\sigma$  and the correlation function

$$C(\tau) \equiv \langle \mathbf{m}(s^t) \cdot \mathbf{m}(s^{t-\tau}) \rangle_t - \langle \mathbf{m}(s^t) \rangle_t \cdot \langle \mathbf{m}(s^{t-\tau}) \rangle_t, \quad (\text{B7})$$

with  $1 \leq \tau \leq \tau_M$ , for the optimal value of  $\tau_M$ . These results are shown in Fig. 4 for both datasets A and B.

The results obtained for the two datasets are consistent and the unique picture that emerges can be streamlined as follows: the dynamics of a given leukocyte at a given time is influenced by the past actions of its peers for a very long time window (in these cases  $\tau_M \sim 13, 14$ ), hence the trajectories that these cells paint while migrating are far from the Markovian limit. Moreover, along the dynamics, the inferred coupling are (mainly) negative in the recent past ( $\tau = 1, 2$ ) and (essentially) positive in the further past ( $\tau \geq 3$ ) and the AIC criterion selects  $\tau_M \sim 13, 14$ . We speculate that a cell, during migration, may integrate signals it receives in time and, from this perspective, the present research seems to suggest that—at least in the present setting—dendritic cells may integrate signals for a temporal time-window of  $\mathcal{O}(10)$  time steps, but clearly a dedicated study has to deepen this aspect in a forthcoming paper.

Finally, as an additional test on the inference protocol on these real datasets, we also kept the collected ordered magnetizations of the various frames and we reshuffled them to produce two new fake datasets (one per region) that we used to test the temporal-delayed protocol: results, reported in Fig. 9, correctly return the same external field (as that is a one-point correlation information that is not destroyed by the permutation of frames), while it selects as the optimal delay  $\tau_M = 0$  (that is correct as shuffling rules out any temporal correlation) and, coherently, a null value for the coupling.

- 
- [1] A. Biewener and S. Patek, *Animal Locomotion* (Oxford University Press, Oxford, UK, 2018).
- [2] E. Meijering, O. Dzyubachyk, and I. Smal, Methods for cell and particle tracking, *Methods Enzymol.* **504**, 183 (2012).
- [3] A. Procaccini *et al.*, Understanding cell fate control by continuous single-cell quantification, *Anim. Behav.* **82**, 759 (2011).
- [4] F. Alemanno *et al.*, Quantifying heterogeneity to drug response in cancer–stroma kinetics, *Proc. Natl. Acad. Sci. USA* **120**, e2122352120 (2023).
- [5] E. T. Jaynes, Information theory and statistical mechanics, *Phys. Rev.* **106**, 620 (1957).
- [6] F. Seno, A. Trovato, J. R. Banavar, and A. Maritan, Maximum entropy approach for deducing amino acid interactions in proteins, *Phys. Rev. Lett.* **100**, 078102 (2008).
- [7] T. R. Lezon *et al.*, Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns, *Proc. Natl. Acad. Sci. USA* **103**, 19033 (2006).
- [8] W. Bialek *et al.*, Statistical mechanics for natural flocks of birds, *Proc. Natl. Acad. Sci. USA* **109**, 4786 (2012).
- [9] H. T. Banks and M. L. Joyner, AIC under the framework of least squares estimation, *Appl. Math. Lett.* **74**, 33 (2017).
- [10] E. Agliari *et al.*, A statistical inference approach to reconstruct intercellular interactions in cell migration experiments, *Sci. Adv.* **6**, eaay2103 (2020).
- [11] A. Cavagna, I. Giardina, F. Ginelli, T. Mora, D. Piovani, R. Tavarone, and A. M. Walczak, Dynamical maximum entropy approach to flocking, *Phys. Rev. E* **89**, 042707 (2014).
- [12] *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, edited by M. Abramowitz and I. A. Stegun (National Bureau of Standards, 1972).
- [13] B. R. Frieden, *Science from Fisher Information: A Unification* (Cambridge University Press, Cambridge, UK, 2004).
- [14] L. Barberis and V. E. Alban, Evidence of a robust universality class in the critical behavior of self-propelled agents: Metric versus topological interactions, *Phys. Rev. E* **89**, 012139 (2014).
- [15] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, Novel type of phase transition in a system of self-driven particles, *Phys. Rev. Lett.* **75**, 1226 (1995).
- [16] D. Geiß, K. Kroy, and V. Holubec, Signal propagation and linear response in the delay Vicsek model, *Phys. Rev. E* **106**, 054612 (2022).
- [17] F. Ginelli and H. Chaté, Relevance of metric-free interactions in flocking phenomena, *Phys. Rev. Lett.* **105**, 168103 (2010).
- [18] D. R. Cox, *Principles of Statistical Inference* (Cambridge University Press, Cambridge, UK, 2011).
- [19] K. Althoff, J. Degerman, C. Wahlby, T. Thorlin, J. Fajjerson, P. S. Eriksson, and T. Gustavsson, Time-lapse microscopy and



- classification of *in vitro* cell migration using hidden Markov modeling, in *Proceedings of the IEEE International Conference Acoustics Speech and Signal Processing* (IEEE, Piscataway, NJ, 2006), Vol. 5, p. 346.
- [20] R. T. Tranquillo and D. A. Lauffenburger, Stochastic model of leukocyte chemosensory movement, *J. Math. Biol.* **25**, 229 (1987).
- [21] E. Gavagnin and C. A. Yates, Stochastic and deterministic modelling of cell migration, *Handbook Stat.* **39**, 37 (2018).
- [22] A. K. Abbas, A. H. Lichtman, and S. Pillai, *Basic Immunology: Functions and Disorders of the Immune System* (Elsevier, Amsterdam, 2019).
- [23] E. Agliari, E. Biselli, A. De Ninno, G. Schiavoni, L. Gabriele, A. Gerardino, F. Mattei, A. Barra, and L. Businaro, Cancer-driven dynamics of immune cells in a microfluidic environment, *Sci. Rep.* **4**, 6639 (2014).
- [24] E. Biselli *et al.*, Organs on chip approach: A tool to evaluate cancer-immune cells interactions, *Sci. Rep.* **7**, 12737 (2017).