Multiple phases in stochastic dynamics: Geometry and probabilities

B. Gaveau*

Laboratoire Analyse et Physique Mathématique, 14 avenue Félix Faure, 75015 Paris, France

L. S. Schulman†

Physics Department, Clarkson University, Potsdam, New York 13699-5820, USA
and Max-Planck-Institut for the Physics of Complex Systems, Nöthnitzer Strasse 38, D-01187 Dresden, Germany

Stochastic dynamics is generated by a matrix of transition probabilities. Certain eigenvectors of this matrix provide observables, and when these are plotted in the appropriate multidimensional space the phases (in the sense of phase transitions) of the underlying system become manifest as extremal points. This geometrical construction, which we call an *observable representation of state space*, can allow hierarchical structure to be observed. It also provides a method for the calculation of the probability that an initial points ends in one or another asymptotic state.

I. INTRODUCTION

In a previous publication [1], we established the relation between phases, including metastable phases, and eigenvalue degeneracy, where the eigenvalues in question are in the spectrum of a matrix of transition probabilities. The context for this use of stochastic dynamics is an approach to non-equilibrium statistical mechanics based on the master equation [2]. In further work, these phases played a role in defining coarse grains in statistical mechanics [3] as well as in discerning community structure in a network [4].

In the present paper we re-examine the occurrence of *multiple phases*. For the case of $m+1$ phases, there turns out to be surprising simplicity in a certain m-dimensional space. This is a space in which the points of the state space are given coordinate values corresponding to the first m observables, where by *observable* we mean a slow eigenvector (in our convention, a *left* eigenvector) of the transition matrix. Remarkably, although one might expect no particular structure to emerge from this representation, for the case of a phase transition (i.e., eigenvalue degeneracy) the points form a simplex, which is to say, there is no more than the minimal number $(m+1)$ of extremal points for the convex hull of the set of state space points in this representation. We call this geometric structure the *observable representation of state space*, and it provides a practical method for the computation of the probability that an arbitrary initial state reaches one or another phase. This leads in turn to potential applications far removed from statistical mechanics. Thus one can have imperfectly defined classes of final states (i.e., they are similar but not exactly the same) for a complex random walk and be able to compute probabilities for arriving at each class. In fact one need not know the classes ahead of time. Moreover, this can be done for dynamics with relatively large state spaces. For a state space of cardinality N, the stochastic dynamics is generated by a matrix with N^2 elements—which may be daunting. Nevertheless, our method requires relatively little information about that matrix: the first few eigenvalues and eigenvectors. Hence, for sparse matrices, which characterize many random walks, these quantities can be computed.

A variation of the method also provides a striking diagrammatic representation of metastable phases, particularly when there is a hierarchical treelike structure, as occurs in spin glasses. One can see shorter and shorter lived (metastable) phases peel away (in reversed time) from those that are closer to the root of the tree. In previous work [5] we explored models of this sort, but in the present article there is fuller understanding and exploitation of the observables. We have also studied other features of the transition matrix spectrum, for example situations where the eigenvalues do *not* drop abruptly as the index increases. This should enhance the utility of this work in the spin glass context [6].

This paper has two principal sections. In Sec. II we develop the mathematical basis for the assertions just made, and in Sec. III we provide examples in which those assertions are realized. Because of the density of mathematical estimates, we begin Sec. II with an overview, which should allow an understanding of the examples without having to go through too many details. The remainder of that section is devoted to those details. Finally, in Sec. IV we discuss problems that may benefit from this treatment as well as mathematical issues, such as whether certain of our hypotheses might be weakened.

II. PHASE TRANSITIONS AND EXTREMALS IN THE SPACE OF OBSERVABLES

A. Overview

The states of the system we study are given by $x, y \in X$, and the system moves from state to state in discrete time according to transition probabilities given by a matrix R. For convenience in dealing with the eigenfunctions of R we define it as follows:

---

*Electronic address: gaveau@ccr.jussieu.fr
†Electronic address: schulman@clarkson.edu

$$R_{xy} = \text{``Pr}(x \leftarrow y)\text{''}$$

$$= \text{Pr}[\text{state at time } (t+1) \text{ is } x | \text{state at time } t \text{ is } y]. \tag{1}$$

For the processes we study there is a unique stationary distribution, $p_0(x)$, which satisfies $p_0 = Rp_0$, with the p_0 on the right of R. For this eigenvalue of R, $\lambda_0 \equiv 1$, the left eigenvector is simply $A_0(x) = 1$, corresponding to the conservation of probability, i.e., $\Sigma_{x \in X} R_{xy} = 1$, because from a state y you have to go *somewhere*.

The eigenvalues of R fall on or inside the unit circle [7,8] and we order them by decreasing magnitude: $\lambda_0 \equiv 1 \geqslant |\lambda_1| \geqslant |\lambda_2| \geqslant \cdots$. The corresponding right and left eigenvectors are respectively p_k and A_k, and satisfy

$$Rp_k = \lambda_k p_k, \quad A_k R = \lambda_k A_k, \quad k = 0, 1, \cdots. \tag{2}$$

Our story begins when several of R's eigenvalues are nearly degenerate with $\lambda_0 = 1$. As we showed in Ref. [1], this heralds a phase transition in the system, enabling the realization of an old dream relating eigenvalue degeneracy and phase transitions [9]. Our method of proof involved those left eigenvectors of R that correspond to the slowest eigenvalues (those nearest to 1). We now find that not only was this convenient for the proof, but it also provides a graphic illustration of phase structure along with the possibility of computing auxiliary quantities such as asymptotic probabilities and time dependence.

We suppose then that m of the eigenvalues (after λ_0) are very close to 1 and that λ_{m+1} is not. We focus on the *left* eigenvectors of R. If X has N states, then we can form an m by N array of quantities, $A_k(x)$, with $k = 1, \ldots, m$, and $x \in X$. Think of the m-tuple $(A_1(x), \ldots, A_m(x))$ as an m *vector*, with one such vector for each $x \in X$. All N of these can be plotted in $\mathbf{R}^m$ and we surround this set of points with a minimal convex surface, the convex hull. In general, such a surface can have many extremal points. We will show, however, that because of the eigenvalue conditions, the convex hull of this particular set of points has (essentially) just $m+1$ extrema, around each of which many, many points of X may cluster. These extremal points are what correspond to the phases. To see what they look like in a typical case, with $m = 2$, see Fig. 2. This is a plot of $A_1(x)$ vs $A_2(x)$. The vertices of the triangle that you see are actually composed of many points (shown in more detail in Fig. 3).

There is a straightforward intuitive way to understand this bunching. The phases are in a sense dynamically far from one another. That is, if you start in one phase you expect to stay there for a long while before going to any other phase. This means that there is a restricted dynamics within that phase that *nearly* conserves probability. The bunching of points in one phase (as we define it) means that for all points in that phase $A_k(x)$ has very nearly the same value (for every $k = 1, \ldots, m$). Let us see why that happens. Consider the eigenvalue equation for A_k applied t times, where t is small enough so λ_k^t is still close to one, so close that we will now treat it as unity. Then

$$A_k(y) \approx \sum_x A_k(x) R_{xy}^t. \tag{3}$$

Now restrict x and y to be such that R_{xy}^t is not small, so we would say x and y are in the same phase. Then for this restriction of R, Eq. (3) still holds, and we appear to have a number of eigenvectors of eigenvalue close unity. But we also have $1 \approx \Sigma_x R_{xy}^t$, because little probability escapes the phase. This last expression says that a constant on the phase plays the role of A_0 for the restricted time evolution. If we now make the further assumption that relaxation *within* the phase is relatively rapid, then other eigenvalues of the restricted R are significantly smaller than 1, and all the apparent eigenvectors A_k, as well as the constant pseudo-A_0, must in fact be proportional to one another. In other words, the A_k's ($k \leqslant m$) are constant on the phases.

The actual proof proceeds a bit differently. Its heart is Eq. (21), which is essentially a statement of the eigenvector property of the A's. In this equation, we write $p_y^t(x)$ in place of R_{xy}^t, since (as just observed) the latter is the probability that starting from y you reach x in t time steps.

B. Additional properties of the stochastic matrix

The matrix R is assumed to be irreducible. This implies that the eigenvalue 1 is unique and that its eigenvector, p_0 (the stationary distribution), is strictly positive

$$\lambda_1 \neq 1 \quad \text{and} \quad \sum_{y \in X} R_{xy} p_0(y) = p_0(x) > 0, \quad \forall \ x. \tag{4}$$

Since no detailed balance assumptions are made for R, it need not be diagonalizable nor have a spectral representation in terms of eigenvectors. Nevertheless, we will assume that for the eigenvalues that concern us (those near 1) each eigenvalue possesses one or more eigenvectors. The orthonormality condition for the eigenvectors, $\langle A_k | p_\ell \rangle = \delta_{k\ell}$, still leaves a single multiplicative factor for each pair (A_k, p_k). The stationary state p_0 is naturally normalized by $\Sigma p_0(x) = 1$, which fixes $A_0(x) = 1$ for all x. For the other A's, consistent with A_0, we normalize by the condition

$$\max_x |A_k(x)| = 1, \quad \forall \ k. \tag{5}$$

Our principal assumption is that for some integer m, $\lambda_1, \lambda_2, \ldots, \lambda_m$ are real and close to $\lambda_0 \equiv 1$. Specifically, this closeness is taken to mean that there exists a range of t (integer times) such that for some $\epsilon \ll 1$

$$1 - \lambda_k^t = O(\epsilon), \quad 1 \leqslant k \leqslant m. \tag{6}$$

In most of our development we further assume that the $|\lambda_k|$ for $k > m$ are much smaller than λ_m, that is,

$$|\lambda_k^t| \ll 1, \quad k \geqslant m + 1. \tag{7}$$

Remark. If R has eigenvalues near -1 (or other roots of unity), our arguments go through for R^2 [or other appropriate power of R, provided that power is not so high that Eq. (6) is violated]. See Sec. 8 of Ref. [10].

The spectral decomposition of R is written

$$R^t = |p_0\rangle\langle A_0| + \sum_{k=1}^{m} \lambda_k^t |p_k\rangle\langle A_k| + |\lambda_{m+1}^t |B^{(t)} \tag{8}$$

with

$$B^{(t)} = \sum_{k \geqslant m} \frac{\lambda_k^t}{|\lambda_{m+1}|^t} |p_k\rangle\langle A_k| \tag{9}$$

[11]. We assume that $\lambda_{m+1}^t B$ is uniformly small in the sense that for any subset $Y \subset X$ and any $x \in X$

$$|\lambda_{m+1}^t| \sum_{y \in Y} |B_{yx}^{(t)}| = O(\eta) \ll O(1). \tag{10}$$

Remark. Instead of the stochastic matrix "R" we sometimes use an alternative matrix, "W," which can be thought of as the generator of continuous time stochastic evolution. Schematically $W = (R-1)/\Delta t$. W will be called a *stochastic generator*. In fact it generates the usual master equation. The constraints on W are that its off-diagonal matrix elements must be non-negative and that its column sums must vanish. Its spectrum consists of zero and points to the left of the y axis in the complex plane. Eigenfunctions are unchanged from the R representation (since the matrices differ only by a multiple of the identity). The advantage of using W is that in producing randomly generated matrices one need not be concerned that the column sum of off-diagonal terms be less than 1.

C. The phases

Under the foregoing hypotheses and under a separation hypothesis "$\mathcal{S}$" to be introduced below, we will construct $m+1$ subsets, ultimately to be identified as the phases, $X^{(1)}, \ldots, X^{(m+1)}$, of X with the following properties:

(1) the sets, $X^{(1)}, \ldots, X^{(m+1)}$, are disjoint;

(2) on each $X^{(\ell)}$, the A_k are nearly constant, for $1 \leqslant \ell \leqslant m+1$ and $1 \leqslant k \leqslant m$;

(3) the complementary subset of $\cup_{j=1}^{m+1} X^{(j)}$ has small p_0 weight. Specifically

$$\sum_{y \in \cup X^{(j)}} p_0(y) \geqslant 1 - \frac{m\epsilon}{\delta} - O(\eta), \tag{11}$$

for a constant, δ such that ϵ/δ is small (and $\epsilon \equiv 1 - \lambda_m^t$); and

(4) The $X^{(1)}, \ldots, X^{(m+1)}$ are essentially unique in a sense to be described below.

Remark. Our phases are a bit larger than what are conventionally called phases and include states that rapidly transit to the usual phases. See Sec. II H.

D. Proof of the existence of the phases

For any pair $x, y \in X$, define

$$p_y^t(x) = R_{xy}^t. \tag{12}$$

$p_y^t(x)$ is the probability that a system in state y at time 0 is in x at time t.

Let $m \leqslant N - 1$. Consider the following geometric construction in $\mathbf{R}^m$: for any $y \in X$, form the vector

$\mathbf{A}(y) \equiv (A_1(y), \ldots, A_m(y)) \in \mathbf{R}^m$. This gives a set $\mathcal{A}$ of N vectors in $\mathbf{R}^m$. Let $\widehat{\mathcal{A}}$ be the convex hull of $\mathcal{A}$.

The first remark is that the vector $\mathbf{0} \equiv (0, 0, \ldots, 0)$ is in $\widehat{\mathcal{A}}$.

This follows from the orthogonality relation, $\langle A_k | p_0 \rangle = 0$, for $k \geqslant 1$. Thus

$$\sum_{y \in X} p_0(y) \mathbf{A}(y) = \mathbf{0}, \tag{13}$$

so that $\mathbf{0}$ is a convex combination of the vectors of $\mathcal{A}$.

As a consequence, one can find $m+1$ points y_ℓ^*, $1 \leqslant \ell \leqslant m+1$, such that the vectors

$$\mathbf{E}_\ell \equiv \mathbf{A}(y_\ell^*) \tag{14}$$

are extremal points of $\widehat{\mathcal{A}}$, and such that $\mathbf{0}$ is a convex combination of them [12,13]. There may be several ways to choose these points y_ℓ^*, but we shall prove that, in fact, the resulting vectors $\mathbf{E}_\ell$ ($1 \leqslant \ell \leqslant m+1$) are uniquely defined up to a small ambiguity to be stated later. By the selection of the $\mathbf{E}_\ell$, we can find μ_ℓ, $1 \leqslant \ell \leqslant m+1$, with $0 \leqslant \mu_\ell \leqslant 1$, $\Sigma_{\ell=1}^{m+1} \mu_\ell = 1$ such that

$$\Sigma \mu_\ell \mathbf{E}_\ell = \mathbf{0}. \tag{15}$$

We have found that there are subsets of these points that are separated from one another in a particular way, and we add to our assumptions the following "separation" hypothesis ($\mathcal{S}$).

Hypothesis $\mathcal{S}$. For each ℓ let

$$\phi_\ell = \min_{k=1,\ldots,m, \ k \neq \ell} \|\mathbf{E}_\ell - \mathbf{E}_k\|, \tag{16}$$

and define

$$\Phi \equiv \min_\ell \frac{\phi_\ell}{\|\mathbf{E}_\ell\|}. \tag{17}$$

Then our hypothesis is that the extrema y_ℓ^* can be selected so that [in addition to Eq. (15)] they satisfy the following:

$$1 - \lambda_m^t = \epsilon \ll \Phi \leqslant O(1) \tag{18}$$

(for an appropriate range of t).

We next observe that by definition

$$\lambda_k^t A_k(y_\ell^*) = \langle A_k | p_{y_\ell^*} \rangle, \tag{19}$$

so that for all ℓ

$$(\lambda_1^t A_1(y_\ell^*), \lambda_2^t A_2(y_\ell^*), \ldots, \lambda_m^t A_m(y_\ell^*)) = \sum_{y \in X} p_{y_\ell^*}^t(y) \mathbf{A}(y). \tag{20}$$

Because $0 \leqslant \lambda_k < 1$, $k \leqslant m$, the vector on the left side of Eq. (20) is in the convex set $\widehat{\mathcal{A}}$; moreover, its distance from the extremal vector $\mathbf{E}_\ell = (A_1(y_\ell^*), \ldots, A_m(y_\ell^*))$ is less than $1 - \lambda_m^t$. Since $\Sigma_y p_{y_\ell^*}^t(y) = 1$, we have from Eq. (20)

$$\mathbf{E}_\ell - (\lambda_1^t A_1(y_\ell^*), \lambda_2^t A_2(y_\ell^*), \ldots, \lambda_m^t A_m(y_\ell^*))$$
$$= \sum_y p_{y_\ell^*}^t(y)(\mathbf{E}_\ell - \mathbf{A}(y)). \tag{21}$$

The idea of the next few steps is as follows. For the case $m=1$ the above equation immediately yields the desired result. On the left you have something that is very small. On the right a sum of products, each factor of which is positive. Therefore, one or the other of these factors must be small. This means that if y can be reached from y^* (with moderate probability), then $A_1(y)$ cannot be much different from $A_1(y^*)$ [in this case E_1 is just $A_1(y^*)$]. For $m>1$ the positivity of $[E_\ell - A(y)]$ is not manifest, so that it is useful to change coordinates and take as origin the vertex of the cone based on E_ℓ. In this way all distances away from the vertex are positive in an appropriate coordinate system. Equation (21) then yields the near constancy of the A's on the phase.

We now pick a particular ℓ, and consider the translated set $E_\ell - \mathcal{A}$ in R^m, i.e., the set of vectors $\{E_\ell - A(y) | y \in X\}$. Then, the vector $\mathbf{0}$ is extremal for the convex set $E_\ell - \widehat{\mathcal{A}}$. We prove the following lemma:

Lemma. Consider a finite set of vectors $w(y) \in R^m$, $y \in X$, such that $\mathbf{0}$ is an extremal point of the convex hull of those vectors. Let $w_0 \in R^m$ be a vector in the convex hull of the $\{w(y)\}$. Thus

$$w_0 = \sum_{y \in X} q(y) w(y), \qquad (22)$$

with $0 \leq q(y) \leq 1$, $\Sigma q(y) = 1$. We further assume that $\|w_0\| < \rho$ for some positive ρ, where the norm is the maximum norm, as in Eq. (5). Then,

(i) There exist m independent linear forms on R^m, $h_1, \dots, h_m$, such that $h_j(w(y)) \geq 0$, $h_j(w_0) < \rho$. Moreover, $h_j(w(y)) = 0$, $\forall j$, if and only if y is in the nonempty set, $\{y\}$, associated with the extremal point $\mathbf{0}$. These forms can be scaled so as to give the same distances that we now use in R^m.

(ii) Let a be a strictly positive real number. Define

$$X(a) = \{y \in X | h_j(w(y)) < a \text{ for all } j = 1, \dots, m\}. \quad (23)$$

Then

$$\sum_{y \notin X(a)} q(y) < \frac{m\rho}{a}. \qquad (24)$$

Proof. Assertion (i) comes from the fact that zero is an extremal point of the convex hull of the vectors $w(y)$ for $y \in X$. The linear forms are essentially a local coordinate system with the extremal at the vertex. Positive coordinates, not necessarily orthogonal can be defined for this cone. Since the scale of this coordinate system in general includes an arbitrary multiplicative constant, it can be taken to be the same as that of the original R^m. Therefore the coordinates of w_0 in this system remain of order ρ.

Assertion (ii): One has, by the hypotheses of the lemma

$$\rho > h_j(w_0) = \sum_{y \in X} q(y) h_j(w(y)) > \sum_{\{y \in X | h_j(w(y)) > a\}} a q(y).$$

Thus,

$$\sum_{y \notin X(a)} q(y) \leq \sum_{j=1}^m \left(\sum_{\{y \in X | h_j(w(y)) > a\}} q(y) \right) \leq \frac{m\rho}{a}. \quad (25)$$

This proves the lemma.

Remark. In practice, the "m" appearing in Eq. (25) may be a severe overestimate. This is because points in other phases will generally exceed a for *all* components of the linear forms.

Application of the lemma. For every ℓ, $1 \leq \ell \leq m+1$, we apply the lemma to the set of vectors and numbers

$$w(y) = E_\ell - A(y), \qquad (26)$$

$$w_0 = E_\ell - (\lambda_1^t A_1(y_\ell^*), \lambda_2^t A_2(y_\ell^*), \dots, \lambda_m^t A_m(y_\ell^*)), \quad (27)$$

$$q(y) = p_{y_\ell^*}^t(y), \qquad (28)$$

$$\rho = (1 - \lambda_m^t) \|E_\ell\|. \qquad (29)$$

These vectors and numbers satisfy the hypotheses of the lemma, because $\mathbf{0}$ is extremal for $E_\ell - \widehat{\mathcal{A}}$ [see Eq. (21)], $\|w_0\| < 1 - \lambda_m^t$, and $\|E_\ell\| \leq 1$. By the lemma, we can define for every ℓ, $1 \leq \ell \leq m+1$, m independent linear forms $h_{\ell,j}$, $j = 1, \dots, m$, such that if we define

$$X_\ell(a_\ell) = \{y \in X | h_{\ell,j}(E_\ell - A(y)) < a_\ell, \ j = 1, \dots, m\}, \quad (30)$$

for a positive real number, a_ℓ, then from Eq. (30) and the lemma

$$\sum_{y \in X_\ell(a_\ell)} p_{y_\ell^*}^t(y) \geq 1 - \frac{m}{a_\ell} (1 - \lambda_m^t) \|E_\ell\| \equiv 1 - \frac{m}{\delta_\ell} (1 - \lambda_m^t), \quad (31)$$

where $\delta_\ell = a_\ell / \|E_\ell\|$.

We now make use of hypothesis $\mathcal{S}$. Because $\epsilon \ll \Phi$ we can find δ_ℓ such that $\epsilon \ll \delta_\ell \ll \Phi$. This allows the probability of remaining *within* a phase to be large [as in Eq. (31)], while maintaining near constancy of the A_k's on that phase. Specifically, from Eq. (30) we have

$$|A_k(x) - A_k(x')| < a_\ell = \delta_\ell \|E_\ell\| \ll \Phi \|E_\ell\| \leq \Phi$$

$$\leq O(1), \quad \text{with } x, x' \in X^{(\ell)}. \quad (32)$$

We next establish that these phases nearly exhaust X in the sense of the probability measure p_0. The spectral decomposition of R gives

$$p_{y_\ell^*}^t(y) = R_{yy_\ell^*}^t = p_0(y) + \sum_{k=1}^m \lambda_k^t p_k(y) A_k(y_\ell^*) + |\lambda_{m+1}^t| B_{yy_\ell^*}^{(t)} \quad (33)$$

But

$$\sum \mu_\ell E_\ell = \mathbf{0}, \qquad (34)$$

with $0 \leq \mu_\ell \leq 1$, $\Sigma_{\ell=1}^{m+1} \mu_\ell = 1$. By the definition of E_ℓ this means

$$\sum_{\ell=1}^{m+1} \mu_\ell A_k(y_\ell^*) = 0, \quad 1 \le k \le m. \tag{35}$$

We next sum Eq. (33) for each ℓ with coefficient μ_ℓ to give

$$\sum_{\ell=1}^{m+1} \mu_\ell p_{y_\ell^*}^t(y) = p_0(y) + 0 + |\lambda_{m+1}^t| \sum_{\ell=1}^{m+1} \mu_\ell B_{yy_\ell^*}^{(t)} = p_0(y) + O(\eta). \tag{36}$$

Now sum over $y \in \cup_k X_k(a)$, i.e., all y *in* the phases, and interchange sums

$$\sum_{\ell=1}^{m+1} \mu_\ell \sum_{y \in \cup_k X_k(a_k)} p_{y_\ell^*}^t(y) = \sum_{y \in \cup_k X_k(a_k)} p_0(y) + O(\eta). \tag{37}$$

Then we deduce

$$\sum_{y \in \cup_k X_k(a_k)} p_{y_\ell^*}^t(y) = \sum_{y \in X_\ell(a_\ell)} p_{y_\ell^*}^t(y) + \sum_{y \in \cup_{k \ne \ell} X_k(a_k)} p_{y_\ell^*}^t(y)$$
$$\ge \sum_{y \in X_\ell(a_\ell)} p_{y_\ell^*}^t(y) \ge 1 - \frac{m}{\delta}\ell(1 - \lambda_m^t), \tag{38}$$

where the last step uses Eq. (31). By Eqs. (37) and (38) (and using $\delta = \min \delta_\ell$),

$$\sum_{y \in \cup X_k} p_0(y) = \sum_\ell \mu_\ell \sum_{y \in \cup X_k} p_{y_\ell^*}^t(y) + O(\eta) \tag{39}$$

$$\ge \sum_\ell \mu_\ell \sum_{y \in X_\ell} p_{y_\ell^*}^t(y) + O(\eta) \tag{40}$$

$$\ge \sum_\ell \mu_\ell \left[1 - \frac{m}{\delta}(1 - \lambda_m^t) \right] + O(\eta) \tag{41}$$

because $\Sigma \mu_\ell = 1$. Therefore, by Eq. (25)

$$\sum_{y \in \cup_\ell X_\ell(a_\ell)} p_0(y) \ge 1 - \frac{m}{\delta}(1 - \lambda_m^t) + O(\eta). \tag{42}$$

This proves statement (11) with $X^{(\ell)} = X_\ell(a_\ell)$.

On each phase $X^{(\ell)}$, one has by definition

$$0 \le h_{\ell,j}(\boldsymbol{A}(y_\ell^*) - \boldsymbol{A}(y)) < a_\ell, \quad j = 1, \dots, m. \tag{43}$$

This implies that the coordinates of the vector $\boldsymbol{A}(y_\ell^*) - \boldsymbol{A}(y)$, for $y \in X^{(\ell)}$, are $O(a_\ell)$ for $k = 1, \dots, m$, because the $h_{\ell,j}$ are linearly independent forms. This establishes Eq. (10).

Remark. If m is large (for example where many metastable phases are present) the factor $m\epsilon$ may not be smaller than 1. In that case there may not be the clean separation of phases that we discuss here.

E. Proof of the uniqueness of the phases

We start from Eq. (2) for A_k and R^t,

$$\lambda_k^t A_k = A_k R^t, \quad 1 \le k \le m. \tag{44}$$

The right hand side of this equation can be split by summing over each phase $X^{(\ell)}$ and over the set of points of X outside of any phase

$$\lambda_k^t A_k(y) = \sum_{\ell=1}^{m+1} \left(\sum_{x \in X^{(\ell)}} A_k(x) R_{xy}^t \right) + \sum_{x \notin \cup_\ell X^{(\ell)}} A_k(x) R_{xy}^t \tag{45}$$

We next estimate each sum in Eq. (45). One has $|A_k(x)| \le 1$, by normalization. Moreover,

$$\sum_{x \notin \cup_\ell X^{(\ell)}} R_{xy}^t \le K \frac{1 - \lambda_m^t}{\delta_\ell} + O(\eta), \tag{46}$$

where K is $O(1)$, as will be shown below in Sec. II G, and in particular Eq. (61). Choose a point $x^{(\ell)} \in X^{(\ell)}$ for $1 \le \ell \le m+1$. One has

$$\sum_{x \in X^{(\ell)}} A_k(x) R_{xy}^t = A_k(x^{(\ell)}) \sum_{x \in X^{(\ell)}} R_{xy}^t$$
$$+ \sum_{x \in X^{(\ell)}} (A_k(x) - A_k(x^{(\ell)})) R_{xy}^t. \tag{47}$$

The last sum in Eq. (47) is $O(a_\ell) \Sigma_{x \in X^{(\ell)}} R_{xy}^t$, using Eq. (32), which is in turn $O(a_\ell)$, since $\Sigma_{x \in X^{(\ell)}} R_{xy}^t < 1$. Therefore, Eq. (45) becomes

$$A_k(y) = \sum_{\ell=1}^{m+1} \left(\frac{1}{\lambda_k^t} \sum_{x \in X^{(\ell)}} R_{xy}^t \right) A_k(x^{(\ell)}) + O(a_\ell) + O(\eta), \tag{48}$$

with $a \equiv \max_{1 \le \ell \le m+1} a_\ell \ll O(1)$. Moreover, $\lambda_k^t = 1 + O(\epsilon)$, so that finally

$$A_k(y) = \sum_{\ell=1}^{m+1} q_\ell(y) A_k(x^{(\ell)}) + O(\eta) + O(a) \tag{49}$$

with

$$q_\ell(y) = \sum_{x \in X^{(\ell)}} R_{xy}^t. \tag{50}$$

$q_\ell(y)$ is the probability that, starting from $y \in X$, the system is in phase $X^{(\ell)}$ at time $t \gg 1$ [14]. The system of Eq. (49) is a vector equation

$$\boldsymbol{A}(y) = \sum_{\ell=1}^{m+1} q_\ell(y) \boldsymbol{A}(x^{(\ell)}) + O(a). \tag{51}$$

Moreover, one has

$$\sum_{\ell=1}^{m+1} q_\ell(y) = \sum_{\ell=1}^{m+1} \sum_{x \in X^{(\ell)}} R_{xy}^t = 1 - \sum_{x \notin \cup X^{(\ell)}} R_{xy}^t = 1 + O(\epsilon) \tag{52}$$

because of Eq. (46).

Thus, Eq. (51) says that the vector $\boldsymbol{A}(y)$ for any y is in the convex hull generated by the $m+1$ vectors $\boldsymbol{A}(x^{(\ell)})$, up to $O(a)$ corrections. Therefore, up to $O(a)$, there can be no extremal points except those already among the $\{x^{(\ell)}\}$, since otherwise Eq. (51) would be an expression for one extremal point in terms of the others. This implies that the phases are unique [up to $O(a)$].

F. Barycentric coordinates

Equation (51) says that the vector $A(y)$ has barycentric coordinates $q_\ell(y)$ defined by Eq. (50), with respect to the vectors $A(x^{(\ell)})$ in each phase $X^{(\ell)}$, where it does not matter which $x^{(\ell)} \in X^{(\ell)}$ is chosen, up to errors of order a. Moreover, by Eq. (50), $q_\ell(y)$ is the probability that, starting from y, the state of the system is in phase $X^{(\ell)}$ at time t (where t is the time scale used to define the phases).

But this means that one can calculate these probabilities $q_\ell(y)$ in a geometric manner. We need to calculate the first m left eigenvectors $A_1(y), \dots, A_m(y)$ (after the trivial A_0). This is sufficient to define the phases by our construction. Using these phases the $q_\ell(y)$ are the barycentric coordinates of A with respect to the phases. Thus, the spectral properties and convex hull construction provide a way to calculate the probability of reaching classes of intermediate asymptotic (time scale t) states, that is to say, the phases.

G. Weight estimates inside and outside the phases

We wish to establish Eq. (46).

(1) For each k, $1 \le k \le m+1$, the following relation is satisfied by the points inside the phase:

$$\sum_{x \in X^{(k)}} p^t_{y^*_k}(x) \ge 1 - \frac{1-\lambda^t_m}{\delta} \qquad (53)$$

[cf. Eq. (25)].

(2) We now consider arbitrary x (not necessarily in one of the phases). By the spectral decomposition

$$p^t_{y^*_k}(x) = p_0(x) + \sum_{\ell=1}^m \lambda^t_\ell p_\ell(x) A_\ell(y^*_k) + \lambda^t_{m+1} B^{(t)}_{xy^*_k},$$

$$1 \le k \le m+1. \qquad (54)$$

With the notation

$$\alpha_{n\ell} \equiv A_{n-1}(y^*_\ell), \quad p'_k \equiv p_{k-1}, \quad P_j(x) \equiv p^t_{y^*_j}(x), \qquad (55)$$

Eq. (54) becomes

$$P_k(x) = \sum_{\ell=1}^{m+1} p'_\ell(x) \alpha_{\ell k} + O(\epsilon) + O(\eta), \quad 1 \le k \le m+1. \qquad (56)$$

This is a linear system of $m+1$ equations for the $p_k(x)$, and as a consequence, one can solve for those quantities

$$p'_k(x) = \sum_{\ell=1}^{m+1} P_\ell(x) c_{\ell k} + O(\epsilon) + O(\eta) \qquad (57)$$

with $c = \alpha^{-1}$. That is, the first $m+1$ right eigenvector are expressible in terms of the distributions within each phase. This is the generalization of Eq. (3.7) of Ref. [1].

Remark. From this relation we can also see that each p_k (of the first $m+1$) right eigenvector is proportional to p_0 on each phase; that is, $p_k(x) \approx (\text{const}) \times p_0(x)$ for $\{x\}$ that constitute a single phase. First recall that each $p^t_{y^*_\ell}(x)$ looks locally

like p_0 on its phase, because t is assumed large enough that local equilibration is complete (that is the smallness of $\lambda^t_{m+1} B$). On the other hand, outside phase ℓ each $p^t_{y^*_\ell}(x)$ is zero, since starting from the extremal on a time scale such that λ^t_m is still close to unity there is little escape. Therefore, by Eq. (57), each p_k (or p'_k) is simply given as a sum of such $p^t_{y^*_\ell}(x)[\text{or } P_\ell(x)]$.

(3) Next take the points x *outside* all phases and write again, for any $y^\dagger$ (inside or outside of any phase)

$$p^t_{y^\dagger}(x) = p_0(x) + \sum_{\ell=1}^m \lambda^t_\ell p_\ell(x) A_\ell(y^\dagger) + \lambda^t_{m+1} B^{(t)}_{xy^\dagger}. \qquad (58)$$

In Eq. (58) replace the $p_k(x)$'s by their values given in Eq. (57)

$$p^t_{y^\dagger}(x) = \sum_{k=0}^m \sum_{\ell=1}^{m+1} \lambda^t_k A_k(y^\dagger) c_{k\ell} p^t_{y^\dagger_\ell}(x) + \lambda^t_{m+1} C^{(t)}, \qquad (59)$$

where $C^{(t)}$ is a combination of the various $B^{(t)}$ and is assumed to be bounded. But from Eq. (31)

$$\sum_{x \notin X^{(k)}} p^t_{y^*_k}(x) \le \frac{m}{\delta}(1 - \lambda^t_m), \qquad (60)$$

so from Eq. (59)

$$\sum_{x \notin \cup_k X^{(k)}} p^t_{y^\dagger}(x) \le K \frac{1 - \lambda^t_m}{\delta} + O(\eta), \qquad (61)$$

where K depends on the $c_{k\ell}$ and thus only on the geometry of the $A(y^*_k)$ and is therefore $O(1)$. Therefore at time t, the probability of starting from $y^\dagger$ (for any $y^\dagger$) to be outside of all phases is small, when t is such that

$$\frac{1 - \lambda^t_m}{\delta} \ll O(1), \quad \text{and} \quad |\lambda^t_{m+1}||B^{(t)}| \ll O(1). \qquad (62)$$

(4) We know that

$$\sum_{x \notin \cup_k X^{(k)}} p_0(x) \le \frac{m}{\delta}(1 - \lambda^t_m) + O(\eta). \qquad (63)$$

For this last estimate, the left hand side does not depend on t, so that the right hand side can be taken a value of t such that Eq. (62) is valid.

H. Basins of attraction

As noted above, our phases are not only the usual states in a phase, but also include the short-time-scale basins of attraction for those phases. In this context "short-time-scale" means $O(\eta)$, that is, at worst, the next slower mode after λ_m. A particular example of this can occur when the extremum point itself is not in what one usually calls the phase but only gets there in $O(\eta)$. In this behavior, the extremal is like the points that are not uniquely identified with any phase and have nonzero probabilities for going to several. (See below, Fig. 2, for an example of this.) The only difference between these intermediate points and a not-in-the-phase extremal is
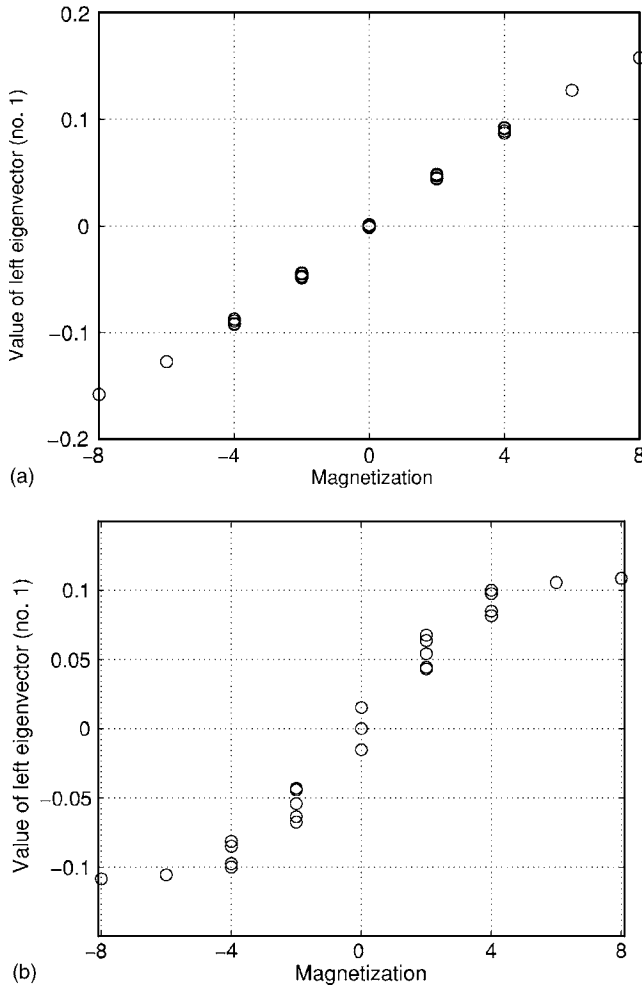
FIG. 1. Ising model stochastic dynamics. A_1 vs magnetization. On the top is shown $T > 1$, on the bottom $T < 1$. In this figure an L^2 normalization is used for A_k (see Ref. [5]).

that the barycentric coordinates for such an extremal have a single 1, with the other entries zeros. The distinction between basin-of-attraction states and those conventionally assigned to the phase will lie in the size dependence of $p_0(x)$—which assumes that one is in a conventional context, where a thermodynamic limit is contemplated.

An illustration of an extremal that does not lie in the usual phase can be found in our article on the definition of coarse grains [3]. There we analyzed a one-dimensional Ising model. Although there is no conventional phase transition in this system, there is a marked difference in system behavior for temperature above and below the value $T = 1$ (in the units we use there) even for moderate values of the number of spins on the ring. In that paper we plotted the left eigenvector, $A_1(x)$, as a function of magnetization. That is, $x = (\sigma_1, \dots, \sigma_N)$ is a spin configuration, with each $\sigma_k = \pm 1$, and the magnetization is $\Sigma_k \sigma_k$. Two plots, one below $T = 1$ and one above are shown in Fig. 1. In Ref. [3] we made the point that the magnetization emerges from the slow eigenvalues (λ near 1) in a natural way. Surprisingly (to us) this turned out to be more marked *above* $T = 1$ than below. The close relation of A_1 to magnetization is evident in the $T > 1$ figure. Below

$T = 1$ there was a bunching of values near the maximum (and symmetrically about the minimum) [15]. The maximum of A_1, which we use to label the phase, occurs for the state with maximum magnetization, which, using analyticity-related definitions, is not part of the phase in the thermodynamic limit [16]. Nevertheless, this maximum of A_1 (for $T < 1$) differs little from $A_1(x)$ for other points, x.

I. Detailed balance for the principal portions of R^t

As above, assume that the first $m+1$ eigenvectors of R are real and that in the spectral expansion the rest of R is small. Then R itself nearly satisfies detailed balance. That is, for the truncated R, $J_{xy} = R^t_{xy} p_0(y) - R^t_{yx} p_0(x)$ is small.

The truncated spectral expansion for R is [cf. Eq. (33)] $\bar{R}^t_{xy} = p_0(x) + \Sigma^m_{k=1} \lambda^t_k p_k(x) A_k(y)$, from which we have

$$\bar{J}_{xy} = \bar{R}^t_{xy} p_0(y) - \bar{R}^t_{yx} p_0(x)$$

$$= \sum_{k=1}^{m} \lambda^t_k [p_k(x) A_k(y) p_0(y) - p_k(y) A_k(x) p_0(x)]. \quad (64)$$

Case 1. x and y are in the same phase. Up to $O(\eta)$, $A_k(y) = A_k(x)$, so that every term in the sum contains differences, $p_k(x) p_0(y) - p_k(y) p_0(x)$. However, by the Remark following Eq. (57), on any particular phase $p_k(x)$ is proportional to $p_0(x)$, so these differences are zero.

Case 2. they are not in the same phase: then R^t_{xy}, which is $p^t_y(x)$, is close to zero (on the order of ϵ). Therefore $\bar{J}_{xy}$ is also zero.

The foregoing observation must be used with caution. It only asserts that $\bar{J}_{xy}$ is small on scale of η or of ϵ. However, for times, τ, such that $1 - \lambda^\tau_m$ is *not* small, this conclusion does not hold.

Remark. Although detailed balance implies that all eigenvalues of R are real, the converse is not true. (Thus the mere fact that all eigenvalues in the truncated R are real does not already imply detailed balance.) From numerical exploration we have indeed found that there is a correlation between $\Sigma_{xy} |J_{xy}|$ and $\Sigma |\text{Im} \lambda_k|$ (for matrices generated in a certain random way); nevertheless it does happen that $\Sigma |\text{Im} \lambda_k| = 0$, while $\Sigma_{xy} |J_{xy}|$ is not [17].

J. A reduced stochastic process

Until now we have focused on a time scale "t" such that $1 - \lambda^t_m \ll 1$. Now take a time, T, such that this is not the case. For such T, we can drop $O(\eta)$ and $O(\epsilon)$ terms in our representation of R, but nevertheless need to retain $1 - \lambda^T_m$. This also implies that for $k = 1, \dots, m$, $A_k(y)$ can be replaced by $A_k(y^*_j)$ for $y \in X^{(j)}$. Furthermore, for these longer times, points not in any phase can be dropped, since they have long before made their way to one or another phase.

Start with the standard spectral representation [dropping $O(\eta)$ and $O(\epsilon)$]

$$R^T_{xy} = \sum_{k=0}^{m} \lambda^T_k p_k(x) A_k(y^*_\ell), \quad \text{for } y \in X^{(\ell)}. \quad (65)$$

Take $x \in X^{(k)}$ and do a standard coarse graining [3]

$$\widetilde{R}(k,j) = \sum_{x\in X^{(k)}, y\in X^{(j)}} R^T_{xy} \frac{p_0(y)}{\mu(j)}$$

$$= \sum_{\ell=1}^{m+1} \lambda_{\ell-1}^T \left(\sum_{x\in X^{(k)}} p'_\ell(x) \right) \left(\sum_{y\in X^{(j)}} \alpha_{\ell j} \frac{p_0(y)}{\mu(j)} \right), \quad (66)$$

where $\mu(j) \equiv \Sigma_{x\in X^{(j)}} p_0(x)$. (N.B. $\mu(k)$, the measure, and μ_k, the barycentric coordinate, are not the same.) Recall from the observation following Eq. (57) that $p'_n(x)$ is proportional to $p_0(x)$ in each phase. Define the proportionality constant by $p_\ell(x) = p_{\ell k} p_0(x)$ [or $p'_\ell(x) = p'_{\ell k} p_0(x)$]. Doing the sums, Eq. (66) becomes

$$\widetilde{R}(k,j) = \sum_{\ell=1}^{m+1} \lambda_{\ell-1}^T p'_{\ell k} \mu(k) \left(\alpha_{\ell j} \frac{\mu(j)}{\mu(j)} \right) = \sum_{\ell=1}^{m+1} \lambda_{\ell-1}^T p'_{\ell k} \mu(k) \alpha_{\ell j},$$

$$(67)$$

The fact that $\Sigma_j \widetilde{R}(k,j) = 1$ follows immediately by summing over x in Eq. (65).

The stochastic matrix $\widetilde{R}$ therefore describes the transitions between phases on a much-elongated time scale.

III. ILLUSTRATIONS

In this section we illustrate the general principles with specific numerical examples.

A. Multiple phases with relatively rapid internal relaxation

In Fig. 2 we show a three phase situation.

For simplicity we work with the matrix W, discussed in the Remark just before Sec. II C. The form of the matrix corresponding to this figure is schematically

$$\widetilde{W} = \begin{pmatrix} W_1 & \epsilon & \epsilon & \alpha \\ \epsilon & W_2 & \epsilon & \alpha \\ \epsilon & \epsilon & W_3 & \alpha \\ \epsilon & \epsilon & \epsilon & 0 \end{pmatrix} + \epsilon \cdot \text{random},$$

$$W = \widetilde{W} - \text{diag}\left(\sum \widetilde{W} \right). \quad (68)$$

Thus three random matrices are produced and weakly coupled to one-another: "ϵ" is a generic small matrix and need not be the same matrix for each appearance in $\widetilde{W}$. Then additional states are added to the state space (those appearing after W_3). These have large one-way couplings to the other states ("α," a generic not-small matrix which again need not be the same in each of its appearances in $\widetilde{W}$) plus small probabilities of return. Next small transition probabilities are added to be sure the matrix is ergodic. Finally the actual stochastic generating matrix, W, is computed from $\widetilde{W}$ by forming column sums and subtracting those sums on the diagonal (where for A an $n\times n$ matrix, ΣA is the n-component object $\Sigma_x A_{xy}$ and "diag" is a diagonal $n\times n$ matrix with its argument on the diagonal and zeros elsewhere).

Because this is a three-phase system (by construction) it is sufficient to take $A(x)$ to be two-dimensional, i.e., we plot
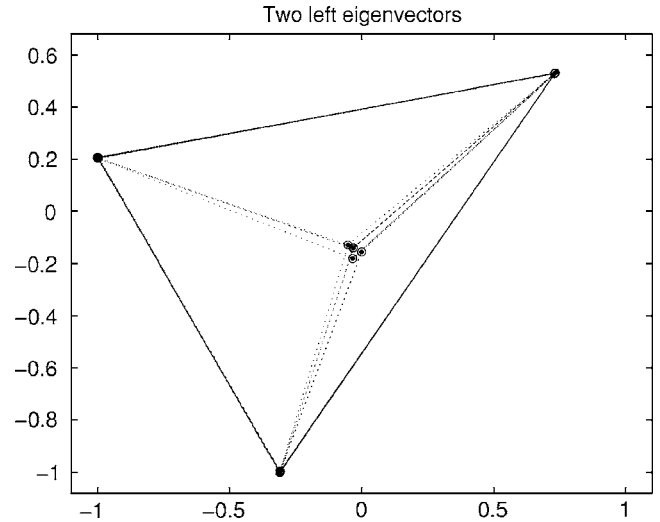

FIG. 2. Plot using the first two left eigenvectors (A_1 and A_2) of the transition matrix, R, for a three-phase system. A circle is placed at each point $[A_1(x), A_2(x)]$ for each of the N states, x, in X. The lines connecting the circles are for visualization. The matrix R is generated by combining four blocks, 3 of which are random matrices, the fourth essentially zero. Then a bit of noise is added throughout, with bigger terms for migration out of the fourth block. Finally the diagonal is adjusted to make the matrix stochastic. This leads to a pair of eigenvalues near 1. This plot using the first two eigenvectors shows the extremal points to be clustering in three regions, corresponding to the phases. The points not at the extremals represent the fourth block, all of which head toward one or another phase under the dynamics. For the particular matrix chosen, they are about as likely to end in one phase as another. For all eigenvector plots, the quantities plotted are pure numbers whose scale is set by our normalization convention, discussed in Sec. II B.

only $A_1(x)$ vs $A_2(x)$ for all $x \in X$. This is Fig. 2. The vertices of the triangle consist of the large number of points (dimensions of the spaces associated with W_k, $k=1,2,3$), while near the middle of the triangle are the points associated with the additional dimensions in the lower right hand corner of W. Because the form of α was approximately the same for all of them, they are near one another. If any of these points is expressed in barycentric coordinates with respect to the vertices, the coefficients give the probability that starting from this point one reaches the respective phase (vertex). To see that the vertices are actually blurred a bit, we plot in Fig. 3 a close-up of one of the vertices. This is the *same* matrix as in Fig. 2.

In the next figures we do the same exercise but with four phases, that is, Eq. (68) is modified by putting in a fourth block, W_4. Figure 4 shows the real part of the spectrum of W. The spectrum is similar to the three-phase case, and as can be seen from the numbers, one does not require extremes of magnitude, large or small, to get useful information from the geometrical construction. By construction, the W of Fig. 4 has three eigenvalues near the stationary one (0), leading to four phases. Figure 5 shows the convex hull of the points $A(x)$. If one cuts off the plot in too low a dimension one gets what is seen in Fig. 6. Here only A_1 and A_2 are plotted and as can be seen there are four rather than three extrema. This is
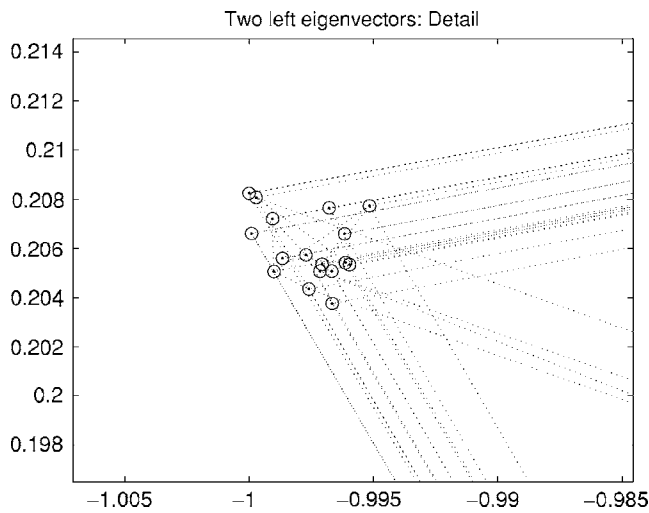
FIG. 3. Detail of the upper left vertex in Fig. 2. In actuality the points in each phase cluster closely together and more than one extremal might, in principle, occur. The precision is limited by the non-negligible magnitudes of the quantities $1-\lambda_2$ and λ_3.

an illustration our need to have λ_{m+1} much smaller than those preceding it. If this condition is not satisfied, more extrema appear. This is what Sec. II E was all about.

B. Hierarchical phases, no sharp cutoff in eigenvalue; simplified spin glasses

For two classes of phenomena we do not expect the eigenvalues to drop off suddenly, as discussed in connection with first order phase transitions. For spin glasses there is expected to be a hierarchical sequence of metastable states. For critical points the eigenvalues should have a power law dropoff near the stationary state.

For hierarchical structures, already studied by us in Ref. [5], we do a variant of the geometrical construction just displayed. The overall W matrix has the following form:
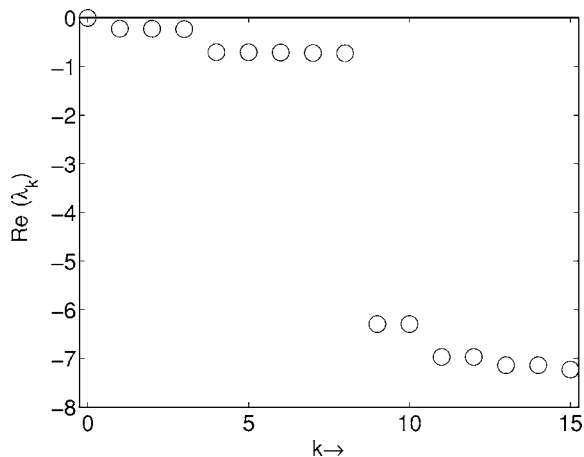
FIG. 4. The first few eigenvalues of W for the four phase system. For those eigenvalues having an imaginary part (which is not the case for the first four), only the real part is shown.
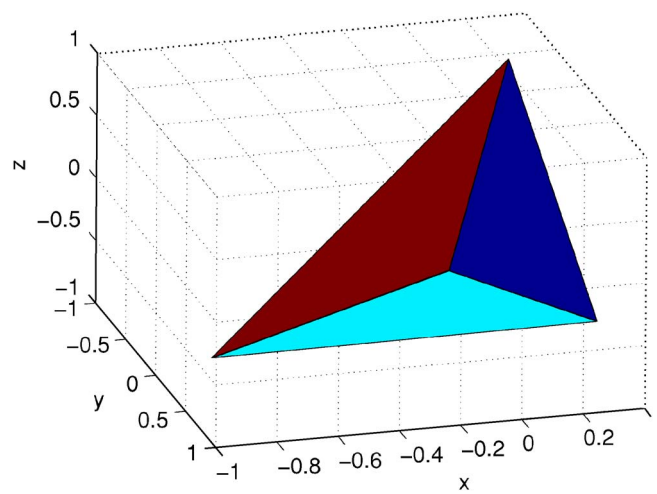

FIG. 5. (Color online) Convex hull of the set of points $A(y)$ for $y \in X$. This is for a case of four phases and the figure formed in $\boldsymbol{R}^3$ is a tetrahedron.

$$W = \begin{pmatrix} W_1 & \epsilon & \epsilon \\ \epsilon & W_2 & \epsilon \\ \epsilon & \epsilon & W_3 \end{pmatrix},$$

with each W_k of the form $W_k = \begin{pmatrix} w_1 & \delta \\ \delta & w_2 \end{pmatrix}$,　(69)

and $\epsilon \ll \delta \ll 1$. For this structure it is instructive to introduce *time* into the picture. The vectors to be plotted are $\boldsymbol{A}^{(t)}(x) \equiv (A_1(x)\lambda_1^t, \ldots, A_m(x)\lambda_m^t)$. We have built the hierarchy to have six phases; on a medium time scale three pairs of them decay into a common branch, subsequent to which the three branches merge into a single trunk. Since we cannot image the five-dimensional structure, we take the projection of this
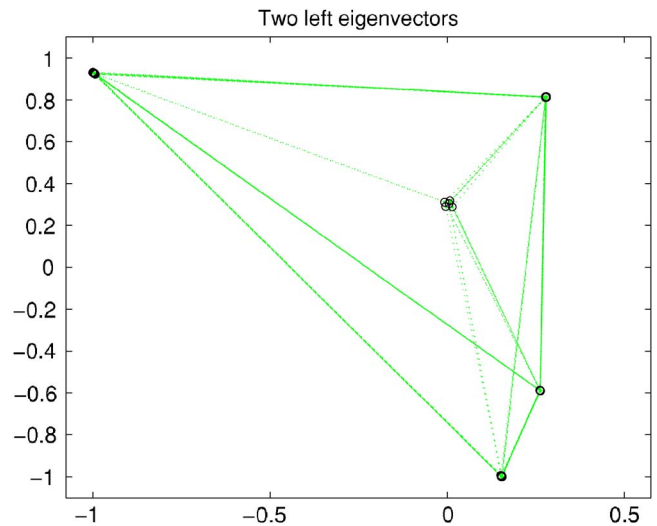

FIG. 6. (Color online) For the four-phase case, if one plots only in two dimensions one does not see only three extremal points. The fourth apparently sticks out of the triangle formed by the others (although it did not have to) and in a third dimension actually forms a node of a tetrahedron (Fig. 5).

FIG. 9. Probability distribution for the stationary state of a walk on the landscape shown in Fig. 8.
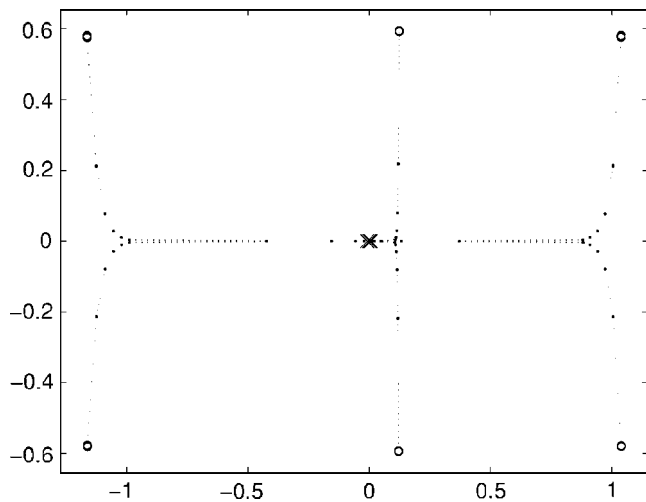
FIG. 7. Phases at successively later times for a hierarchical stochastic matrix. As explained in the text this is a two-dimensional projection of the five-dimensional plot of eigenvectors multiplied by eigenvalue to the power t. On the shortest time scale there are 6 metastable phases (circles); subsequently they merge into three and finally into a single stationary state. The axes represent particular linear combinations of eigenvectors and lack physical dimensions (but have a scale determined by the normalization of Sec. II B).

motion (as a function of time) on a particular plane. This is shown in Fig. 7, where the circles represent the original phases and the "$\times$" is the final state, $(0,0)$.

C. Asymptotic probabilities

Consider a random walk on the landscape shown in Fig. 8. The stationary state is shown in Fig. 9. There are clearly four regions of attraction, which we identify as the "phases" discussed in this paper. The spectrum of the (225 by 225) generator of the stochastic dynamics is $[0, \exp(-16.0), \exp(-15.3), \exp(-14.8), \exp(+1.2),\ldots]$, so

that this satisfies the conditions for having four well-demarcated phases, which in this case represent regions of attraction. Finally, in Fig. 10 we show how the methods of this paper can be used to calculate the probability that from a given initial condition one will arrive at one or another asymptotic state. Each circle in the graph (which is a three-dimensional plot of a tetrahedron; cf. Fig. 5) represents a point on the 15 by 15 lattice and its location in the plot, when expressed in barycentric coordinates (positive numbers that add to 1) with respect to the extremals, gives its probability of reaching a particular phase. In the graph we do not identify the particular circles, but the same computer program that generated the graph can easily provide a table of probabilities for each initial condition.

IV. PROSPECTS

The transition matrix for a stochastic process gives rise to *observables*, namely its slowest left eigenvectors [in our con-
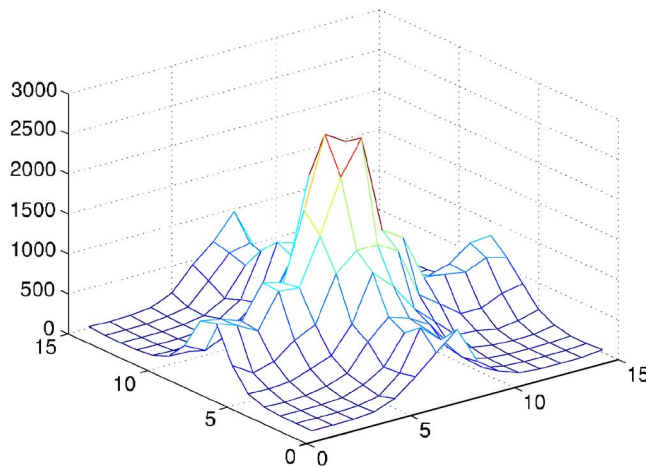

FIG. 8. (Color online) Landscape for a random walk on a 15 by 15 lattice. The distance unit on the lattice is arbitrary and the scale of the potential chosen so as to give a dynamical spectrum illustrating our representation.
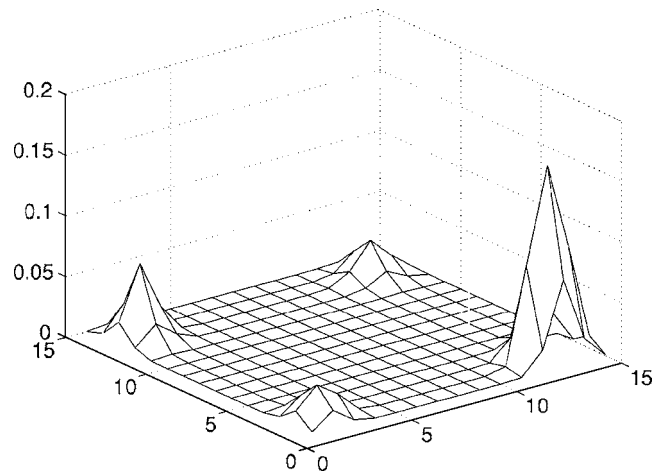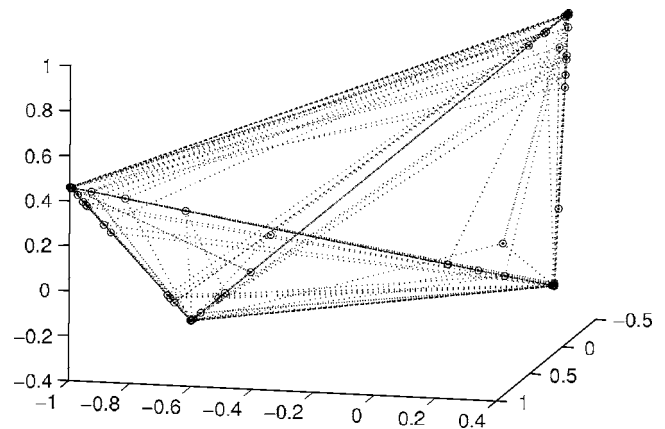

FIG. 10. Observable representation, in $\boldsymbol{R}^3$, of the states for the walk on the landscape shown in Fig. 8. Each circle represents a point on the 15 by 15 lattice and its position within the tetrahedron (when expressed in barycentric coordinates with respect to the extremals) gives the probability of starting at that point and arriving at one or another extremal. The plot is very much like that shown in Fig. 5, but includes interior points. (Fig. 5 shows only the convex hull.)

vention $R_{xy} = \Pr(x \leftarrow y)$]. For each x in the state space of the process one can form a vector $(A_1(x), \ldots, A_m(x))$ for integer m, with A_k the left eigenvectors. Depending on the spectrum of R, the space of these vectors can provide a graphic demonstration of the phases (in the sense of phase transitions) of this process. We call a plot of the points of the state space using a collection of slow (left) eigenvectors an *observable-representation of state space.*

We have shown how when there is a hierarchical structure of phases that structure becomes manifest in the space of observables. Our model for demonstrating this is artificially constructed, but we expect that for systems of greater intrinsic interest the same features seen here should emerge. Thus, spin glass models could be considered, for example the Sherrington-Kirkpatrick model. Even for local spin glasses, although the state space grows large quite rapidly, our method requires little information from the transition matrix, R. For example, a two-dimensional 4 by 4 spin glass would involve a 2^{16} by 2^{16} matrix ($2^{16} = 65\ 536$), but it is a *sparse* matrix, and all we would want to know would be the first few eigenvectors, which is quite feasible. Now 4×4 may not be much of a lattice, but the knowledge gained from the corresponding observables would immediately give information for the longest possible time scales. For the mean field Sherrington-Kirkpatrick model one should be able to do even better.

In the more traditional arena of stochastic processes, our geometric construction allows one to read off the probabilities of an initial point reaching any of various asymptotic states, even when one does not have prior knowledge of what those states are. We gave a simple example of a random walk on a multiwell landscape, but other examples easily come to mind.

At the mathematical level, we believe our assumptions are stronger than they need to be. It is likely that hypothesis $\mathcal{S}$ can be replaced by something weaker. We already have preliminary results on this point in low dimension. Another place where we have an implicit assumption (although we did not emphasize it at the time) is in the proof of assertion (1) of the lemma, where we make a generic assumption about the geometry of the linear forms: specifically that angles in the effective coordinate system are not such that large values of h could be generated from small distances.

Cases where one has eigenvalues near one but there is not a sharp dropoff after one particular eigenvalue are of great importance in physical applications. Certainly for spin glasses, although they are expected to show the hierarchical structure discussed above, they would have a collection of time scales of decreasing size (local relaxation times) with no cutoff at a particular value. Also of interest is the case of critical phenomena. Here, absent hierarchical structure, we do not expect a small number of extrema to dominate. One also has other properties, for example the divergence of spatial correlation lengths, that on the face of it do not appear to be directly related to the dynamics. Nevertheless, as shown in Ref. [3], much of this structure can be recovered from the eigenvectors, so that a dynamical characterization of this kind of transition, as we have done here for first order transitions, may well be possible.

Nevertheless, even where there is no phase transition, the observable representation can provide an image of the state
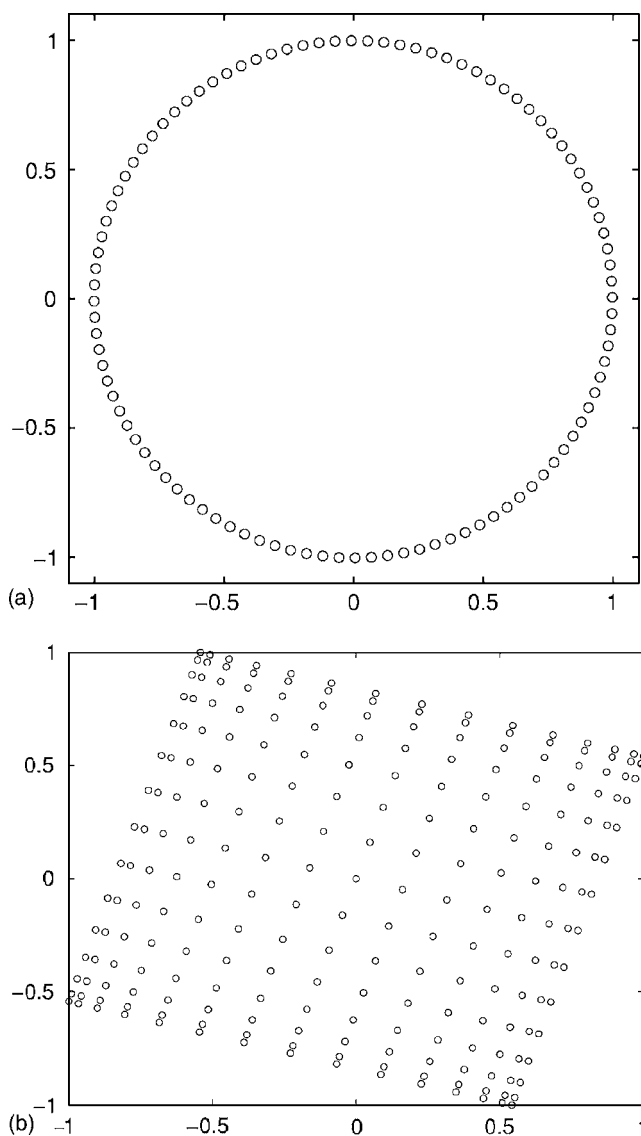

FIG. 11. Observable representation, in $\mathbf{R}^2$, of the states for Brownian motion. The circle is for the one-dimensional case of a walk on a ring. The rectangular figure is for a two-dimensional random walk with nonperiodic boundary conditions.

space. For the case of Brownian motion on a ring, the transition matrix simply has constants just off the diagonal, as well as in the corners to provide periodicity. We showed in Ref. [3] how R can recover spatial structure, but a plot of $A_1(x)$ vs $A_2(x)$, as in Fig. 11, is even more direct. As can be seen, this immediately gives the coordinate space ring. For two dimensions we show (in the same figure) a slight variant in which we have reflecting rather than periodic boundary conditions.

[1] B. Gaveau and L. S. Schulman, J. Math. Phys. **39**, 1517 (1998).

[2] B. Gaveau and L. S. Schulman, J. Math. Phys. **37**, 3897 (1996).

[3] L. S. Schulman and B. Gaveau, Found. Phys. **31**, 713 (2001).

[4] L. S. Schulman and B. Gaveau, Bull. Sci. Math. **129**, 631 (2005).

[5] B. Gaveau, A. Lesne, and L. S. Schulman, Phys. Lett. A **258**, 222 (1999).

[6] G. Biroli and J. Kurchan, Phys. Rev. E **64**, 016101 (2001).

[7] F. R. Gantmacher, *The Theory of Matrices* (Chelsea, New York, 1959), Vol. 1.

[8] F. R. Gantmacher, *The Theory of Matrices* (Chelsea, New York, 1959), Vol. 2.

[9] M. Kac, in *Fundamental Problems in Statistical Mechanics II*, edited by E. G. D. Cohen (North-Holland, Amsterdam, 1968), pp. 71–105.

[10] B. Gaveau and L. S. Schulman, J. Stat. Phys. **110**, 1317 (2003).

[11] In Eq. (9), $B^{(t)}$ is represented by a spectral decomposition. This is for convenience only. No harm would come if a Jordan form would be needed; all we require is the smallness of $B^{(t)}$.

[12] T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

[13] H. Moulin and F. Fogelman-Soulié, *La convexité dans les mathématiques de la décision* (Hermann, Paris, 1979).

[14] Note that the q in Eq. (28) corresponds to $y=y_\ell^*$ for some ℓ, while in Eq. (50) it can refer to any point in X.

[15] For the one-dimensional Ising model (used to generate our figure), with single-spin-flip dynamics, one can show that in the high temperature limit $MW=-2\exp(-4\beta J)M+O(1-\exp[-4\beta J])$, with M the magnetization vector, J the coupling constant, and β the inverse temperature. For lower temperatures additional terms become important.

[16] The present discussion is at a heuristic level, since in one dimension, by the analyticity definition, there is no phase transition. However, the features observed here should persist in two dimensions and indeed in any phase transition of this sort. Note too, that for us metastable phases stand on an equal footing with stable ones and that our formalism abandons the thermodynamic limit. Rather, as should be evident here and in Ref. [1], a phase transition is an asymptotic concept, so that we have no difficulty speaking of a phase transition in the one-dimensional model.

[17] Here is a 3-by-3 example of a stochastic generating matrix (W, not R) that provides a counterexample: $W_{12}=0.486$, $W_{13}=0.457$, $W_{21}=0.231$, $W_{23}=0.019$, $W_{31}=0.607$, $W_{32}=0.762$ (with appropriate diagonal to satisfy $\Sigma_x W_{xy}=0$). This has eigenvalues 0, -1.20 and -1.36, all real. The calculated positive current is 0.458 times the permutation matrix with 1 in the (1,3) position.