Review
The Observable Representation

L. S. Schulman (ORCID)

Physics Department, Clarkson University, Potsdam, NY 13699-5820, USA; schulman@clarkson.edu

check for updates

Abstract: The observable representation (OR) is an embedding of the space on which a stochastic dynamics is taking place into a low dimensional Euclidean space. The most significant feature of the OR is that it respects the dynamics. Examples are given in several areas: the definition of a phase transition (including metastable phases), random walks in which the OR recovers the original space, complex systems, systems in which the number of extrema exceed convenient viewing capacity, and systems in which successful features are displayed, but without the support of known theorems.

1. Introduction

The observable representation (OR) is an embedding of the space on which a stochastic dynamics is taking place into Euclidean n-space. The value of the integer n is determined by the dynamics as well as by the user. The most significant feature of the observable representation is that it respects the dynamics. Thus points close to one another in this embedding tend to be dynamically related. This often allows an appreciation of what's going on and insight into the relation of points to one another.

Much (but not all) of the work described here was done in collaboration with Bernard Gaveau at the University of Paris VI. Other collaborators are thanked in the acknowledgements.

In this article the observable representation will be abbreviated OR. The word "observable" is meant to bring to mind the slowest varying quantities, which correspond to the observables of a system [1].

As far as I know, this method is conceptually different from principal component analysis [2,3], although both methods allow visibility of relations among elements of networks, dynamical or not.

In Section 2, a precise definition as well as notation are given. In the following Section 3 various applications are described. For example, phase transitions (and the inclusion of metastable phases) are defined in this formalism. Next we discuss the reconstruction of spaces—remarkably the OR can look like the original space. In Section 4 we deal with an ansatz not supported by theorems, but which occasionally works. This comes up when the spectrum of the transition probabilities has a non-zero imaginary part in its large eigenvalues. A final section summarizes the previous material.

There is some new material in this article, although it is mostly a brief review. Section 3.6, Appendix A, and Figure 4 in Section 3.1 and the discussion in the text are all published for the first time. The examples in Section 4 are also new.

2. Abstract Definition of the OR

Let X be a space having N elements. These elements are typically coarse grains of some finer dynamics, for example Hamiltonian dynamics, which, because of the coarse graining introduce a probabilistic aspect to time evolution. However, in many applications this view is too restrictive and dissipative or intrinsically probabilistic dynamics is contemplated. (There are even extentions to quantum mechanics [4].)

On the space X a *stochastic* dynamics is defined, i.e., for two elements $x, y \in X$ there is a probability given that y goes to x:

$$R_{xy} = \Pr(x \leftarrow y). \tag{1}$$

Several points should be noted in this definition. First, the notation R has been introduced for the generator of the stochastic dynamics. (Sometimes this will be written R_{xy}, as above, sometimes $R(x, y)$.) Second (and here other authors may differ) the arrow in Equation (1) points from right to left [5]. Finally, there is an implicit time interval during which each step is taken, that is, each application of R is assumed to take the same time, say τ. In most situations τ will be taken to be 1, but for going to continuum dynamics (and the "master equation") it will be useful to let $\tau \to 0$.

As a result of R's probabilistic nature the state (the element $x \in X$) will not, in general, be definitely known and the most one can give is probabilities. That is, there will be a function $p(x, t)$ giving the probability that at time-t the system is in state x. Of course since the system is always *somewhere*, for each t, $\sum_{x \in X} p(x, t) = 1$. Finally, we can give the relation between $p(x, t)$ and $p(x, t + 1)$, namely

$$p(x, t + 1) = \sum_{y \in X} \Pr(x \leftarrow y) p(y, t) = \sum_y R_{xy} p(y, t). \tag{2}$$

R is thus the generator of time evolution, and it was to make Equation (2) look good that the right-to-left convention for R was chosen.

One property of R is immediately evident: $\sum_x R_{xy} = 1$, from any initial point y you have to go somewhere , and since $X = \{x\}$ is exhaustive the sum must give 1.

A matrix having non-negative elements and whose column sums are all 1 is called *stochastic*, so that our evolution matrix is stochastic. Stochastic matrices have an important spectral property: all eigenvalues are on or inside the unit circle. Proof can be found in [6] and other places.

A property that I'll often assume is *reducibility*. Abstractly this means that there is no permutation operator, P such that $P^T R P$ has a lower left non-trivial block that is zero. (P^T is the transpose of P.) If N (the cardinality of X) is small enough there is an easy computational way to find out if R is reducible: add the identity ($\mathbb{I}$) to R and replace every non-zero element of $R + \mathbb{I}$ by 1 (calling the result M). You now have a matrix, M, of ones and (possibly) zeros. Compute M^{N-1}, and if there are no zero elements of this matrix then the original R is irreducible; otherwise it is reducible (see [7]). In other words, to be irreducible every state must be able to be reached from any other state. Under these circumstances the leading right eigenvector is unique. Call it $p_0(x)$. Then $p_0(x) > 0$ for every x and it satisfies $R p_0 = p_0$, so it is the stationary probability distribution (often called the stationary *state*, which one could confuse with elements of X, but that is what people call it). By the way, it was clear that there would be a right eigenvector of R with eigenvalue 1, since we already had a left eigenvector with that property. That left eigenvector is all 1's, i.e., we define $A_0(x) \equiv 1$ (for all $x \in X$) and note that the condition on R that all column sums be one, can be written $A_0 = A_0 R$, confirming that the eigenvalue is 1.

Call the eigenvalues $\{\lambda_k\}$ with $\lambda_0 = 1$ and the rest ordered by descending absolute value. Thus $\lambda_0 \geq |\lambda_1| \geq |\lambda_2| \geq \dots$. It can happen that there are fewer than N eigenvectors (Jordan forms, [8]), but that situation will not be considered here. This also imposes an order on the left and right eigenvectors. Call the right eigenvectors p_k and the left eigenvectors A_k, for $k = 0, 1, \dots, N - 1$. The left and right eigenvectors may be different although you still have $\sum_x A_j^*(x) p_k(x) = \delta_{jk}$, the Kronecker delta. Complex eigenvalues are put in pairs, so that if λ_k is the first complex eigenvalue, then $\lambda_{k+1} = \lambda_k^*$, and so forth. (Complications can arise if the complex eigenvalues are also degenerate, but this will not be considered here.)

For complex eigenvalues not only do the eigenvalues λ_k and λ_{k+1} come in pairs, but so do the eigenvectors: thus $A_k^* = A_{k+1}$. This means that for k and $(k + 1)$ one can replace the true eigenvectors by $\tilde{A}_k = \Re(A_k)$ and $\tilde{A}_{k+1} = \Im(A_k)$. One can then define the "tilded" vectors that are real to replace the original eigenvectors and one has a full set $\{\tilde{A}_k\}$.

It is now possible to state what the OR is, although motivation and explanation of the word "observable" comes from the examples. Pick an integer m, with $m \leq N$, and consider for each $x \in X$ the m-component vector, $(A_1(x), A_2(x), \ldots, A_m(x)) \in \mathbb{R}^m$. This amounts to an embedding of the original space, X, in Euclidean m-space. (Or one could use the version with tildes, i.e., A replaced by $\tilde{A}$. See Section 4 for more on this subject.)

The *current*, an auxiliary quantity, can also be defined for a general stochastic process. For the stationary probability distribution p_0 we define

$$J(x, y) \equiv R(x, y) p_0(y) - R(y, x) p_0(x). \tag{3}$$

There is no summation convention above: J is defined for each x and y. Moreover, J satisfies Kirchoff's rules and the sum of all currents at each point is zero.

Detailed balance is the situation in which all currents are zero. In this case all eigenvalues of R are real. This is easy to see, since detailed balance implies that the matrix $C_{xy} \equiv \frac{1}{\sqrt{p_0(x)}} R(x, y) \sqrt{p_0(y)}$ is symmetric and since it is real, also Hermitian. This implies that all eigenvalues of C are real. Moreover, C has the same eigenvalues as R, with the right eigenvectors of R given as $\sqrt{p_0}$ times the eigenvectors of C.

Another feature that applies to all non-Jordan form matrices is the spectral expansion. One can write the matrix R in terms of its eigenvalues and left and right eigenvectors

$$R(x, y) = \sum_{k=0}^{N-1} \lambda_k p_k(x) A_k(y). \tag{4}$$

(Sometimes this is written $R(x, y) = \sum_{k=0}^{N-1} \lambda_k |k, x\rangle \langle k, y|$, although the bra-ket notation might obscure the fact that left and right eigenvectors can differ.)

3. Examples

We flesh out the abstract definition of the previous section with examples. The first set, dealing with phase transitions, is what motivated the original work, but the others followed with surprising relevance.

3.1. Phase Transitions

In the usual definition of phase transitions—a breakdown in analyticity—there is no such thing as supercooled water. For example, Isakov [9] proved that you cannot analytically continue the Ising model to a metastable phase (magnetization and magnetic field opposite to one another). However, if you look in handbooks [10] you can find the density of water at $-2\,\text{C}$. The problem is that the usual definition does not deal with non-uniform approximations. Once you let the number of particles or the volume go to infinity—which you must do to get a breakdown in analyticity—*somewhere* in this vast volume a critical droplet [11] will form and you no longer have a pure *metastable* phase. However, for real water in finite volumes, even macroscopic volumes, a critical droplet can take a long time—aeons—to form (for water you have to go to $-30\,\text{C}$ to get short survival times and to about $-40\,\text{C}$ for supercooling to cease [12]. This means that the time for formation of a critical droplet can be larger (in appropriate units) than the system size, and non-uniform limits are called for.

However, in stochastic dynamics a completely different definition is possible. As mentioned earlier, there is always an eigenvalue 1, and (for irreducible matrices) it is unique. What about the next few? A sign of a first order transition between two phases is if the next eigenvalue (λ_1) is very close to 1 and all others are much smaller. This applies also to metastable phases. Moreover the left eigenvector A_1 is nearly constant on states in X belonging to each phase, making its value a discriminator between the two phases; hence the term "observable." We will see shortly why all this is true.

It is simpler though to start with a Brownian motion example, namely a random walk on a potential surface. Take as the potential a surface with four minima, as in Figure 1. The space X is the 15-by-15 array of points on which the walk takes place. (So R is 225-by-225.) The stochastic process usually goes downhill (to lower values of the potential), but with some probability (dependent on the temperature) will go uphill. Step sizes are one and periodic boundary conditions are used. For low temperature it is clear what the result will be. The walker stays in one minimum for a long time, then every once in a while will move to one of the others (usually not crossing the center), where it also stays a long time.
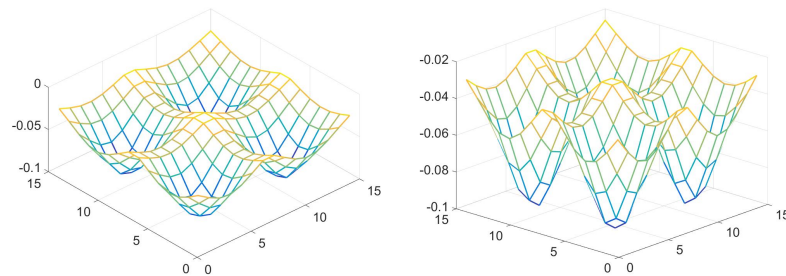

Figure 1. Two views of the same potential surface. The random walk prefers to move downhill, but occasionally (due to a finite temperature) will go to greater values of the potential.

The matrix R implements the rules just stated and can be diagonalized to yield eigenvalues and left eigenvectors. What is important is how close the eigenvalues are to 1, and for the example given the first five differences are $1 - \lambda_k = (0.000429, 0.000429, 0.000794, 0.1268, 0.1284)$ (for $k = 1, \ldots, 5$). Note the increase in $1 - \lambda_k$ at $k = 4$. As we shall see, this suggests a 3-dimensional plot of the first 3 (non-trivial) left eigenvectors. This is shown in Figure 2.
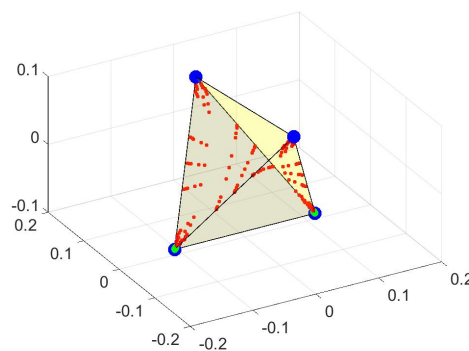

Figure 2. The OR for a random walk at low temperature in the 4-minima potential. This is a plot of the three first non-trivial left eigenvectors of the matrix R of transition probabilities, i.e., $A_1(x)$ vs. $A_2(x)$ vs. $A_3(x)$ for each $x \in X =$ space of positions. These are the three axes of the figure. For each $x \in X$ there is a certain probability of being found in the stationary state ($p_0(x)$). Different symbols are used for different categories of points, x. The largest probabilities are given the largest symbols, and these coincide with the extremal points (vertices) of the tetrahedron. Other points have smaller weight (values of $p_0(x)$) and are given smaller symbols (mostly red at this temperature, but some intermediate sizes, with green symbols, almost at the vertices). The convex hull of the points, which is the pictured tetrahedron, is shown as a guide to the eye. See the main text for further explanation.

This is a perfect tetrahedron with the points having the largest probability shown with the largest markers. These correspond to the minima of the potential. Points near them have slightly lower

probability while the peaks in Figure 1, while not zero, have very small probability. (For the example given, each of the minima has probability 0.07, while the peak in the center has probability 8×10^{-6}.)

The next pair of figures (Figure 3) require a bit of analysis, but lie behind the designation "observable representation" (OR). The upper figure shows the values of the three first (non-trivial) eigenvectors, and it is a mess. (The temperature is lower than for Figure 2 to accentuate features.) However, confining attention to only the "big" probability states, which in this case is all those greater than 6×10^{-8}, one has the lower graph. Remarkably, each eigenvector only takes a few values, and any triplet of values characterizes a minimum (or a phase). This is shown in Table 1. That table is like a code. Each eigenvector is nearly constant on a phase, but each phase has a unique characterization, using those constants. For this reason the leading left eigenvectors are called the "observables."

Table 1. Values of the left eigenvalues on the principle points of the 225 ($= 15^2$) of the space *X*. The phases are characterized by reading *down*. Thus phase 1, a vertex of the tetrahedron in Figure 2, is characterized by $A_1 = -0.0118$, $A_2 = 0.105$ and $A_3 = 0.819$.

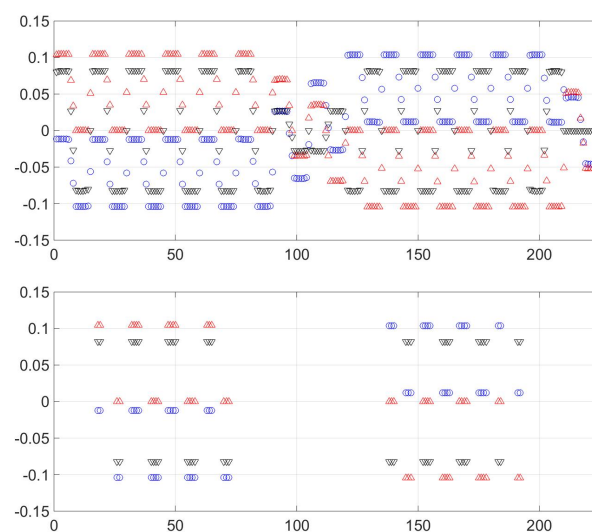| A_k, Symbol in Figure 3	Value on Phase 1	Value on Phase 2	Value on Phase 3	Value on Phase 4
1, circle	−0.0118	−0.104	0.104	0.0118
2, triangle, up	0.105	3×10^{-5}	1×10^{-5}	−0.105
3, triangle, down	0.819	−0.819	−0.819	0.0819

Figure 3. The first three left eigenvectors, pictured as blue circles, black downward pointing triangles and red upward pointing triangles. The upper figure shows all points, the lower shows only those with p_0 greater than 6×10^{-8}. The *x*-axis is a number assigned to each of the 225 points and can be considered arbitrary. As can be seen—also in Table 1—only 10 values of the left eigenvectors have significant weight. This allows each phase to be characterized by the unique *A* values that characterize that phase.

Note that there are points that are not at the vertices of the tetrahedron. These points have low probability and are not part of any phase. However, they have another interesting property. Suppose you give their *barycentric* coordinates (see Appendix B) with respect to the extremum points, namely those points that form the vertices of the tetrahedron. Barycentric coordinates must add to 1. However, so do probabilities. In fact, the barycentric coordinate with respect to any given extremum (phase) is just the probability that a walker starting from the given point ends up in that particular phase. (This is a theorem. See below.)

There can also be surprises in the OR. By "surprise" I mean a property that I do not know how to prove. It turns out that the tetrahedral form of the OR is retained even for higher temperatures, even though the theorems that have been proved until now do not establish this. Then—suddenly—the figure becomes something altogether different. This is shown in Figure 4. The change in the figures represents a relative temperature change $((T_{\text{ball}} - T_{\text{tetrahedron}})/T_{\text{tetrahedron}})$ of order 10^{-9}. Up until $T_{\text{tetrahedron}}$ the OR is recognizably a tetrahedron. From T_{ball} on it remains a ball. There is no recognizable change in the probability distribution nor in the spectrum (which no longer guarantees anything), but the change takes place in as short a temperature interval as I have patience to compute.
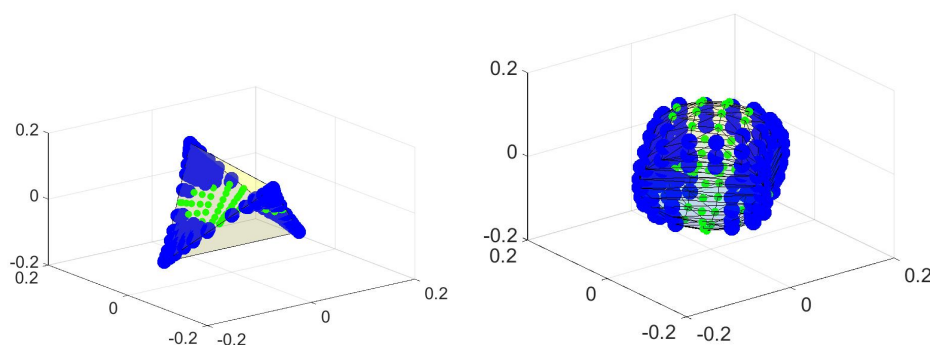

Figure 4. A sharp transition, whose explanation eludes me. On the left is the OR (a plot of $(A_1(x), A_2(x), A_3(x))$ for all $x \in X$) for the same potential that is pictured in Figure 1, but at a higher temperature ($T_{\text{tetrahedron}} \gg$ the temperature used in Figure 2). At this higher temperature I know of no reason for the points to form a tetrahedron, but they do. The probabilities of non-vertex points is higher (so the large symbols are far more common) but the shape of the convex hull is still a tetrahedron. Then, with a tiny fraction of temperature change (10^{-9}), the figure changes completely—roughly to a ball. The probabilities (p_0) of individual points do not change (significantly), but the OR becomes dramatically different.

Phase transitions, whether between stable or metastable phases, have the same properties. The general notion follows from the eigenvalues of the matrix, R, of transition probabilities. The criterion is that the first few non-trivial eigenvalues must be close to 1, while those following them drop off quickly. (For now I consider the case of detailed balance, so all eigenvalues are real.) Suppose $1 - \lambda_k \ll 1$ for $k = 1, \ldots, n$ while $1 - \lambda_k$ for $k > n$ is much larger. Then the OR is a simplex of $n + 1$ extrema in Euclidean n-space. Moreover, points deeply in the interior of the simplex (having barycentric coordinates with no nearly-1 component) have much lower probability (p_0 is small for those points). A way to look at this is to use the spectral expansion. Let T (a time) be sufficiently large that λ_k^T is small for $k > n$ but λ_k^T is still close to 1 for $k \le n$. Then

$$R^T(x, y) = \sum_{k=0}^{N-1} \lambda_k^T p_k(x) A_k(y) \approx \sum_{k=0}^{n} \lambda_k^T p_k(x) A_k(y) \approx \sum_{k=0}^{n} p_k(x) A_k(y). \tag{5}$$

On this time scale the dynamics is fully equilibrated within each phase, and all that is left is the hopping between phases.

A proof of the simplex property is a bit complicated and for a full proof the reader is referred to [13]. (In that paper "m" is what is here called $n + 1$.) For $n = 1$ the proof is simpler although even in this low dimension some preparation is needed. First, $\sum_x A_1(x) p_0(x) = 0$ by the basic orthogonality relation. Since (as usual) we assume irreducibility, we know that $p_0(x)$ is strictly positive. It follows that there is a positive maximum value of A_1, call it $A_M > 0$ and (one of) the point(s) where it takes that value x_M. Similarly there is a negative minimum value, call it $A_m < 0$ and define x_m analogously. Now take as your initial point x_M. In other words $p(x, 0) = \delta_{x, x_M}$. Evolve it for T time

steps. Clearly almost all points remain in the relevant phase, which follows from the spectral expansion. What we do know is that

$$A_M \lambda_1^T = \sum A_1(x) R^T(x,y) \delta_{y,x_M},$$ (6)

since each application of R to the left leaves A_1 unchanged except for building up factors of λ_1, and $A_1(x_M) = A_M$ by definition. Call $\sum_y R^T(x,y) \delta_{y,x_M}$, which is the probability distribution after T time steps starting from x_M, $p_M(x,T)$. Similarly the same can be said for x_m and its corresponding p_m. It follows that

$$\lambda_1^T = \sum \frac{A_1(x)}{A_M} p_M(x,T) = \sum \frac{A_1(x)}{A_m} p_m(x,T),$$ (7)

From the equalities $1 = \sum_x p_M(x,T) = \sum_x p_m(x,T)$ we subtract Equation (7) to obtain

$$1 - \lambda_1^T = \sum \left[1 - \frac{A_1(x)}{A_M} \right] p_M(x,T) = \sum \left[1 - \frac{A_1(x)}{A_m} \right] p_m(x,T),$$ (8)

The quantity on the left is nearly zero. The contribution from eigenvalues beyond $k = 1$ is also negligible. It follows that to get the sums to be zero one or the other term in the respective products must be near zero. In other words, if $p_M(x,T)$ is not zero, i.e., if you are the phase associated with x_M then $A_1(x)$ is going to be very close to A_M. The same is true for A_m; both are (nearly) constant on their respective phases.

With this result we are able to define phases. As usual there is a certain arbitrariness involved. Suppose we pick an a and define the phase P_μ to mean $|A_1(x) - A_\mu| < a$ (where μ is either M or m). The quantity a is small, but not too small, since as we shall see it appears in a significant denominator.

Our next goal is to establish that there is little probability outside the phases; that is $\sum_{x \notin \cup P_\mu} p_0(x)$ is small. Define P'_M to be the complement of P_M. Then

$$\sum_{x \in P'_M} p_M^T(x) < \frac{1}{a} \sum_{x \in P'_M} p_M^T \left[1 - \frac{A_1(x)}{A_M} \right] < \frac{1 - \lambda_1^T}{a}.$$ (9)

And since $1 = \sum_{x \in X} p_M^T(x) = \sum_{x \in P_M} p_M^T(x) + \sum_{x \in P'_M} p_M^T(x)$ it follows that

$$\sum_{x \in P_M} p_M^T(x) > 1 - \frac{1 - \lambda_1^T}{a}.$$ (10)

The same relations hold with m (minimum) replacing M (maximum). Note that λ_1 may not be all that close to unity in some applications (for example in Figure 4) and (relatively) high probability x's may occur at some distance from the extremum.

We next show that barycentric coordinates can serve as probabilities. For arbitrary y we look at $A_1(y)$. The fundamental equation for left eigenvectors can be written

$$\lambda_1^T A_1(y) = \sum_\mu \sum_{x \in P_\mu} A_1(x) R^T(x,y) + \sum_{x \notin \cup P_\mu} A_1(x) R^T(x,y).$$ (11)

At this point we need to give some normalization to A_1, since until now the only demand was $\sum_x A_1^* p_k = \delta_{1k}$. There are two natural normalizations. One can set $\max_x |A_k| = 1$ for all k, and this is the normalization we here adopt. A second normalization, particularly when detailed balance obtains (which is not demanded for what we now prove) is L^2 normalization, which will be described in detail in the next subsection. Either way we have $|A_1(x)| \leq 1$ and therefore

$$\left| \sum_{x \notin \cup P_\mu} A_1(x) R^T(x,y) \right| \leq \sum_{x \notin \cup P_\mu} R^T(x,y) \leq \frac{1 - \lambda_1^T}{a}.$$ (12)

Next A_1 within a given phase is replaced by A_μ for that phase with an error of order a. It follows that

$$\lambda_1^T A_1(y) = \sum_\mu A_\mu \sum_{x \in P_\mu} R^T(x,y) + O(a) + O\left((1-\lambda_1^T)/a\right). \tag{13}$$

Two points should be noted about this equation. First, up to order $1 - \lambda_1^T$, λ_1^T is simply 1, and this factor already appears on the right hand side, divided by a, a number equal to or less than one. So set the λ^T on the left to one. Consider the sum $\sum_{x \in P_\mu} R^T(x,y)$: it is just the probability that a point starting at y ends up in phase μ. We give it a name: $q_\mu(y) = \sum_{x \in P_\mu} R^T(x,y)$. Equation (13) becomes

$$A_1(y) = \sum_\mu A_\mu q_\mu(y) + O(a) + O((1-\lambda_1^T)/a). \tag{14}$$

Now $A_1(y)$ is the position of the point y in the (one-dimensional) OR. What Equation (14) says is that up to $O(a) + O((1-\lambda_1^T)/a)$, $A_1(y)$ and be expressed as a sum of the extremal points, $\{A_\mu\}$; that is, the quantities $q_\mu(y)$ are *barycentric* coordinates. Or to restate it, to a(n often) good approximation, the probability to enter one or another phase is given by the barycentric coordinates in the OR.

As for the usual OR these results generalize to a multidimensional OR. See [13] for details. To summarize: When there are m eigenvalues that are real and close to 1, with the other $(N - m - 1)$ much smaller in absolute value, then the OR is a simplex. Phases are defined through proximity to the extrema of the simplex and the barycentric coordinates of points within the simplex are the probabilities that starting from that point you will end in one or another phase. The proof relied on the fact that there is an intermediate time at which points within a phase have come to equilibrium, but the overall distribution of phase occupation will depend on the initial conditions.

3.2. Reconstructing Spaces

Now consider the opposite case: the eigenvalues crowd around 1; they do not drop off suddenly. Under these conditions you do not expect a simplex nor weights (values of p_0) that are concentrated in phases; in fact there may not be phases.

Brownian motion on various configurations without a potential [14] is an example of this situation. Consider Brownian motion on N points arrayed in a circle. The transition probabilities for this can be taken to be all 1/2's and zeros. The 1/2's are just above and just below the diagonal, as well as in positions $(N, 1)$ and $(1, N)$. This matrix can be diagonalized exactly and its eigenvalues are $1, \cos(2\pi/N), \sin(2\pi/N), \ldots, \cos(2\pi k/N), \sin(2\pi k/N), \ldots$. The first two non-trivial (i.e., not all ones) left eigenvectors can be taken to be $\cos(2\pi/N)$ and $\sin(2\pi/N)$. Now the punch line: plot A_2 versus A_1 and you get Figure 5.

How about a random walk on a figure Y; that is, the walker can get to the ends of the segments (no periodic boundary conditions) but in the middle there is one place where there is a choice of which branch to go on. I will not give a figure, but the OR is a Y. This continues to more complicated figures. As an extreme example, I used the computer to create a random walk on a fractal built of triangles. (This was computer generated so the fractal had 5 levels instead of going on indefinitely.) What is illustrated in Figure 6 is *not* the original fractal, but the OR that results from the stochastic matrix for a walk on the fractal. The eigenfunctions themselves (see [14]) give every appearance of being a mess, but somehow they conspire to produce the given figure. A second example of the power of this method is the Petersen graph. This is a graph with 10 vertices in which each vertex has 3 edges emerging from it. It has the property that *it cannot be embedded in a plane.* See Figure 7. It is interesting to see what the standard prescription for the OR does to this. The random walk again goes from vertex to vertex. Each line has several steps to make the connections visible. From the pair of graphs in Figure 8 it is clear that *three* dimensions are required in order to embed this figure.
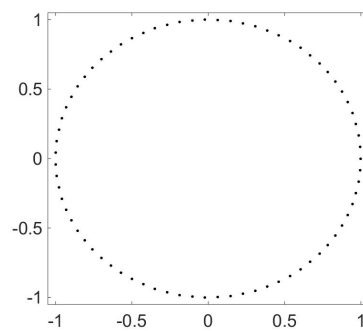
Figure 5. OR for a random walk on a circle of 100 points. The random walk has probability 1/2 of taking a step of size 1 in either direction on a circle. The figure though is not a plot of the original circle; rather it is the values taken by $(A_1(x), A_2(x))$ for each $x \in X$, where X is the discrete set of positions on a circle.
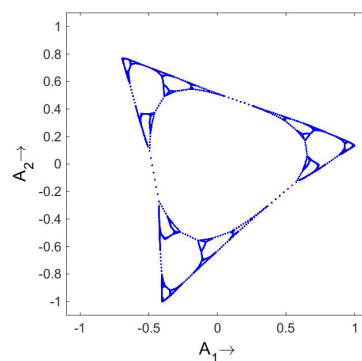

Figure 6. The OR for a random walk on a fractal. In this case the space X is a discrete representation of a fractal. When travelling on a straight portion of the fractal the walker has probability 1/2 of going in either direction. When the walker comes to a split there would be three choices, go back or go to one of the branches of the fractal. These are each given probability 1/3. As in the previous figure, what is illustrated is *not* the original fractal (X), but the OR, a plot of $(A_1(x), A_2(x))$ for each $x \in X$.
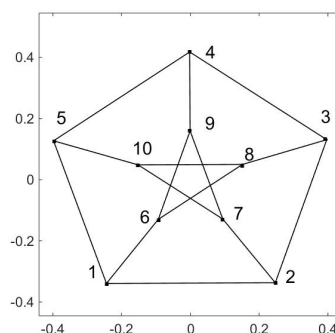

Figure 7. The Petersen graph. Note that there is no intersection in the interior pentagon; e.g., the line 6–8 does *not* touch the line 7–10. This graph cannot be embedded in a plane.
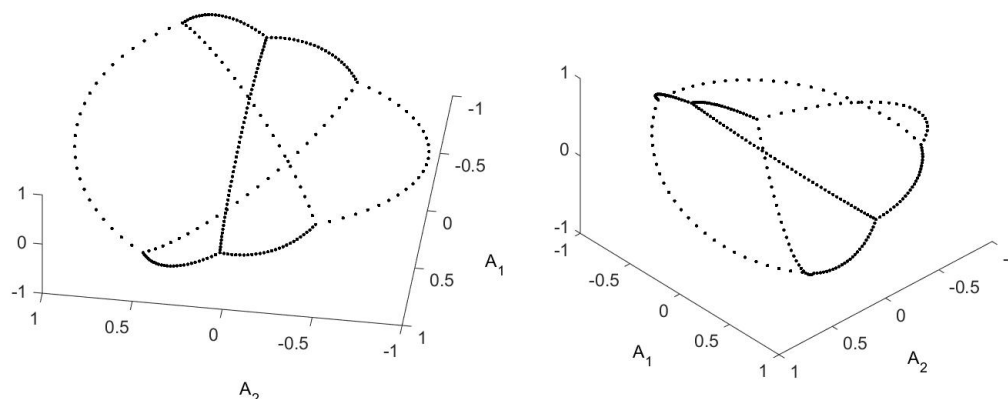
Figure 8. The OR for the Petersen graph. Two views are presented to emphasize that the OR also has the property that it needs three dimensions for an embedding.

This works in higher dimension and even in "mixed dimension" (although strictly speaking we are always in dimension zero, since our space X is finite). As one illustration we show a random walk which is mostly on a square, but which can wander off onto a line. As in previous examples, what is shown in Figure 9 is the OR, not the original line plus square—although they look the same.
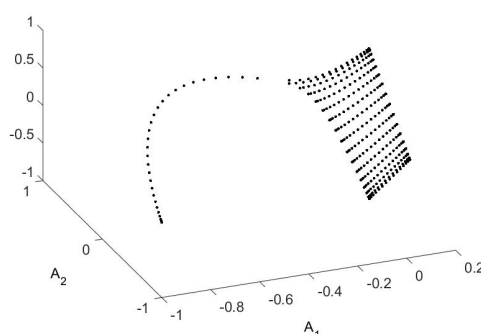

Figure 9. OR for a random walk on a square with "one-dimensional" path attached to one corner.

For other examples see [14].

3.3. Rationale

In general, proximity in the OR reflects dynamical proximity. This was seen for phase transitions, for all sorts of Brownian motion and through the barycentric relation, which does not require detailed balance. On the other hand, if there is detailed balance a precise relation can be found.

We first generalize notation: Let $p_y(x, t)$ be the probability distribution for points that were at the point y at $t = 0$. (This is a clear generalization of $p_M(x, t)$ or $p_m(x, t)$.) Clearly $p_y(x, t) = R^t(x, y)$.

For the case of detailed balance ($R(x, y)p_0(y) = R(y, x)p_0(x)$) we earlier saw that $\sqrt{p_0}$ played an important role. We now formalize that. Let σ be a diagonal N-by-N matrix with $\sigma(x, x) = \sqrt{p_0(x)}$. Then as observed earlier $C \equiv \sigma^{-1}R\sigma$ is symmetric and by virtue of its reality, Hermitian. As such, C has a complete set of eigenvalues and eigenvectors with

$$C(x, y) = \sum \lambda_k \psi_k(x) \psi_k^*(y) , \tag{15}$$

with the same set of λ's as for R and with $\{\psi_k\}$ orthonormal, i.e., $\sum \psi_k(x)^* \psi_j(x) = \delta_{kj}$. Since $p_k = \sigma \psi_k$ and $A_k = \sigma^{-1}\psi_k^*$ it is clear that this normalization of eigenvectors is different from setting the maximum of $|A_k|$ to be unity.

The probability distribution $p_y(x, t)$ can be thought of as a cloud of values; for example in the case of a phase transition if t is such that λ_1^t is close to 1, but $|\lambda_2|^t$ is near zero, then $p_y(x, t)$ will either be an entire phase, or a weighted sum of the two of them, and will exclude (have small value at) points in between.

We define a distance between two such clouds

$$D(x, y) = \sum_u \frac{|p_x(u, t) - p_y(u, t)|}{\sqrt{p_0(u)}}, \tag{16}$$

Note that

$$\frac{p_x(u, t)}{\sqrt{p_0(u)}} = \frac{1}{\sqrt{p_0(u)}} \left(\sigma C^t \sigma^{-1} \right)_{ux} = \sum_\alpha \lambda_\alpha^t \psi_\alpha(u) \frac{\psi_\alpha^\dagger(x)}{\sqrt{p_0(x)}} = \sum_\alpha \lambda_\alpha^t \psi_\alpha(u) A_\alpha(x), \tag{17}$$

The function $D(x, y)$ of Equation (16) is of the form of an L^1 norm. The L^1 norm is always equal to or greater than the L^2 norm, and therefore

$$D(x, y) \geq \left(\sum_u \frac{|p_x(u, t) - p_y(u, t)|}{\sqrt{p_0(u)}} \right)^{1/2} = \sqrt{\sum_u \left| \sum_\alpha \psi_\alpha(u) \lambda_\alpha^t \left[A_\alpha(x) - A_\alpha(y) \right] \right|^2}. \tag{18}$$

This expression is of the form $\left[\sum_u \sum_\alpha \sum_{\alpha'} \psi_\alpha(u) \psi_{\alpha'}^*(u) c_\alpha c_{\alpha'}^* \right]^{1/2}$, which because of the orthonormality of the ψ's is $\left[\sum_\alpha |c_\alpha|^2 \right]^{1/2}$. It follows that

$$D(x, y; t) \geq \sqrt{\sum_\alpha |\lambda_\alpha|^t |A_\alpha(x) - A_\alpha(y)|^2}. \tag{19}$$

The sum on the right can be truncated at the mth eigenvalue, preserving the direction of the inequality. Since the magnitudes of the eigenvalues decrease monotonically (with index) we obtain, finally,

$$D(x, y; t)/|\lambda_m|^t \geq \sqrt{\sum_{\alpha=1}^m |A_\alpha(x) - A_\alpha(y)|^2}. \tag{20}$$

The quantity on the right is the distance in the OR. Thus, two points x and y which are adjacent dynamically, meaning their clouds overlap, will be spatially adjacent in the OR.

3.4. Spin Glass

The spin glass has Hamiltonian $H = -\sum_{\langle ij \rangle} J_{ij}\sigma_i\sigma_j$ and has served as a model for many phenomena, from glass to the brain. The coupling constants, $\{J_{ij}\}$, are typically random but quenched (meaning they are fixed for each calculation); they are usually taken to be either ± 1 or with values in a normal distribution. The spins, $\{\sigma_i\}$ are ± 1. The pairs $\langle ij \rangle$ are usually nearest neighbors on a lattice of some sort. A model of the model is a mean field version known as the Sherrington-Kirkpatrick model (SK). The Hamiltonian is

$$H = -\frac{1}{\sqrt{2N}} \sum J_{ij}\sigma_i\sigma_j, \tag{21}$$

where the sum is now over all i and j (different from each other). The states are now $x \equiv (\sigma_1, \sigma_2, \ldots, \sigma_N)$ and the energy is $E(x) = H(x)$. The transition probability for x and x' differing by the value of a single spin is

$$R(x', x) = \begin{cases} \frac{1}{N} \exp\left(-\frac{[E(x')-E(x)]}{T}\right) & \text{if } E(x') > E(x) \\ \frac{1}{N} & \text{otherwise} \end{cases}, \tag{22}$$

Then the diagonal of R is adjusted to make all column sums unity. This transition probability matrix guarantees that the stationary probability distribution is just $p_0(x) = \exp(-E(x)/T)/Z$ with $Z = \sum_{x \in X} \exp(-E(x)/T)$.

The SK model spin glass [15] offers a mixture of the previous situations. Yes, there are phases, but the spectrum of eigenvalues has many points near 1. This is because there are *many* phases, most of them metastable. The idea is that the system passes slowly through ever deeper (in energy) metastable phases on its way to the stationary probability distribution.

To diagonalize R you first need to fix a matrix, J (the set $\{J_{ij}\}$), of couplings; in this case each coupling was ± 1. For N spins there are 2^N different states, so that R is a 2^N-by-2^N matrix. For the computer used (and the limited skill of the user) this meant that for most calculations N was 12 or 13. Nevertheless, insight into the coordinate space and the progression of metastable phases could be gained.

First, each $x = (\sigma_1, \ldots, \sigma_N)$ would naturally be associated with a point on the unit N-cube, a not very helpful association. A measure of distance could be hamming distance or Euclidean distance, neither of which would describe the underlying dynamics, which also depends on the matrix J. With the OR there is an embedding that respects the dynamics, so that proximity means that one state can easily become the other.

Having more than 4 phases means that three dimensions are not adequate to fully visualize the geometry. Nevertheless, one can find the convex hull in more than three dimensions and mark the relevant state (the vertices) on a 3-dimensional plot. (Actually, barring the possibility that this article will be published with a 3-dimensional printer, what you will see is a 2-dimensional plot of a 3-dimensional object.) For a particular choice of J and T this is shown in Figure 10.

What does the OR tell you about the dynamics? It provides extrema and a distance scale. By the procedure mentioned earlier (cf. the paragraph between Equations (8) and (9)) this allows us to define phases. (Further verification is that for the example in Figure 10, 40% of the points lie outside all phases, and yet their *combined* probability is about 2×10^{-5}). What about the flow between phases? For that it suffices to start a bunch of points in a phase and see where they end. This leads to an effective $\tilde{R}$ between phases. In this case there is an element of randomness due to the routes of various points from the phases. (You could also add the transition probabilities from the each point in a given phase.) The diagram of transitions (with the randomness) is illustrated in Figure 11.
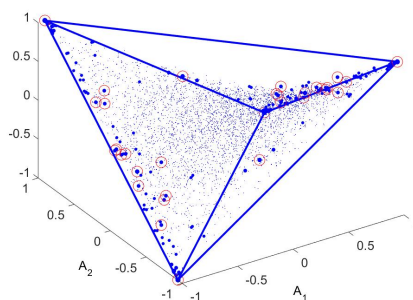

Figure 10. Symbol size increases with increasing p_0 (hence lower energy). Circles indicate local minima in energy. The convex hull (a tetrahedron) has been outlined. Many of the minima are vertices (extrema) in a higher dimension convex hull; this has been checked to dimension seven.
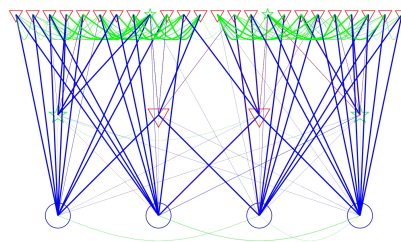
Figure 11. Transitions between phases. The phases are sorted (based on measured lifetimes) into 3 categories, long-, medium- and short-lived. On the lowest level are the longest-lived, etc. The symbol and color correspond to the amount of inflow. A phase's left-right position follows the A_1 values of that phase's nucleus among phases at the same lifetime level. Lines indicate transition probabilities, with line width a monotonic function of transition probability. Blue lines are flow from higher (shorter lifetime) levels to lower, red go oppositely, and green represents lateral transitions.

The figure can be considered a test—or a reflection—of the hierarchical model. That is the picture of phases flowing into ever deeper metastable phases. Probabilistically there should be flow in both directions, but in a larger setup the flow to less likely phases would be quite small. In the illustration some of the wrong-direction flow happens to be large but this can be attributed to random paths. It is interesting though that, up to randomness, the OR distances (and resulting phases) reproduce the hierarchical model.

3.5. Foraging

How do animals look for food? Lots of different strategies, all of them reasonably successful, or else the species would cease to exist. In recent work [16] Wosniack calculated foraging behavior for an animal that was far-ranging and with food sources that were renewable. She used exact positions (with machine precision), but when it came to analyze the data coarse graining was performed. However, at this stage all she had was a single path. It was long enough to come back to the same position several times, so it was a candidate for the method of Appendix A. (See that section for details.) Food sources were at particular positions in this (periodic) universe, and each food source gave an extremum in the OR. In the simulation she used 5 places where food could be found. This meant there would be 5 extrema. How to picture them?

The key was a mapping back to the original two-dimensional space. For details see the original article [16], but with this additional piece of information (and using k-means) the location of the food sources could be established. See Figure 12.
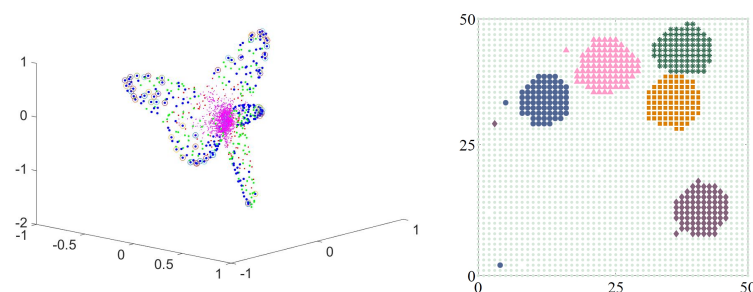

Figure 12. On the left is the OR for foraging with 5 extrema. It is possible to see that there are 5 extrema, but there isn't much precision in their location. However, using k-means plus the mapping to the original two-dimensional space (of the forager) the locations of the food sources could be recovered (right hand figure).

3.6. Complex Systems

Complexity has yet to be defined, but people will agree that linguistics, ecology and protein networks can qualify. An example of an application to each field is given in [17]. Occasionally the embedding is easy to understand and has the appearance of one of those I've already dealt with. However, there are also situations where deductions require a bit of knowledge of the system. Even in our earlier examples if the number of extrema exceeded four, other tricks had to be used for visualization. Still, as one can see from the spin glass example, it is possible to find structure (e.g., hierarchical properties) in the system.

In the present article I'll give an example of an OR that I do not understand. The results are significant, the system is arguably complex, but I do not understand what's going on. Sometimes the OR brings surprises, and sometimes you (or I, anyway) do not know why the figure picks out special points.

I took the text of Melville's novel, Moby Dick ("Call me Ishmael. ... "). The question was if letter 337 was an "a" what would letter 338 be? In other words I got the probability of ordered pairs, what is most likely to follow each letter. To make things easily doable I made several simplifications. I got rid of commas and other markings such as quotation marks, changed semicolons to periods and all capital letters to lower case, so in the end there were only 28 characters, the Roman alphabet (sans accents) plus a period and a space. Then I counted all ordered pairs and these became the first version of the transition matrix. However, the columns sums were not all the same. Thus, if $R(i,j) = \text{const} \cdot \Pr(\text{letter } \#j \text{ goes to letter } \#i)$ then the sums, $\text{sum}_j = \sum_i R(i,j)$, vary. Since all these numbers should be one, further measures must be taken—and they are not unique. In MATLAB$^{\text{TM}}$ notation, you would have choices:

`R=R/max(sum(R));` followed by `R=R+diag(1-sum(R));` or

`R = R*diag(1./sum(R));` and many others.

If you choose the first method, you divide R by the maximum of its column sums and make it up in the diagonal terms. In the second scheme you divide each column by its sum. I prefer the first method because for equilibrium systems it gives the Boltzmann distribution as it is stationary probability distribution. Also, in the present instance scheme #1 gives interesting results, Scheme #2, as far as I know, does not.

Those "interesting results" are illustrated in Figure 13. You can see that there is a well-formed tetrahedron, with the extremal points indicated. All the other points are gathered in that fuzzy image near the vertex at point #19. Now the first few eigenvalues are $1, 0.9972, 0.9952, 0.9947, 0.9924, 0.9618, 0.9587)$ so there is no sudden dropoff after $\lambda_3 = 0.9947$, but there is still a tetrahedron. However, the real surprise comes when the identity of the vertices are revealed. They are the letters "jqxz" (not ordered). These are the most rare letters in the book. One can also look at the 4-dimensional convex hull formed by the first 4 non-trivial eigenvectors, and one gets "jkqvxz", again rare letters. (There are 6 letters, rather than 5, because the convex hull is not a perfect simplex.) Why in 3 dimensions one gets a neat tetrahedron, and why the rare letters are at the vertices, I do not know. The OR is full of surprises, and this one has me baffled.
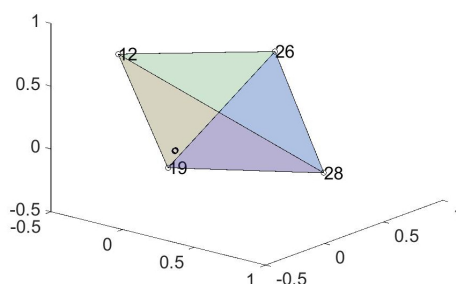
Figure 13. Letter pairings from the book, "Moby Dick," with the transition matrix defined by $R(k, j) = \Pr(\text{sybmol } j \text{ is followd by symbol } k)$. The "symbols" are the 26 letters of the Roman alphabet, plus a period and space. The letters corresponding to the points 12, 19, 26 and 28 are j, q, x and z (the first two on the list are a space and a period).

4. Currents and Complex Eigenvectors

Current was defined in Section 2 as $J(x, y) \equiv R(x, y)p_0(y) - R(y, x)p_0(x)$ and it was there proved that if the current at all points in X is zero then detailed balance prevails and all eigenvalues of R are real. However, currents often do occur and eigenvalues can be complex. For example, in the results of Section 3.6 (Moby Dick) the first few eigenvalues are real, but the 9th and 10th (counting 1 as the zeroth, and numbered in decreasing magnitude) have non-zero imaginary parts, as do the pairs $(21, 22)$ and $(24, 25)$. Complex eigenvalues mean complex eigenvectors.

Remark: The occurrence of non-zero current has been suggested [18] as a basis for evaluating complexity. The current has a topological structure and is thus richer than the assignment of a single number as a measure of complexity. Clearly all detailed balance systems are (by a definition based on currents) not complex, so that only non-equilibrium systems are candidates for complexity. However, a general definition has yet to be offered.

As currents increase (in any particular model), so does the imaginary part of the eigenvalues. In addition—in my experience—it also moves up the eigenvalues. That is, at first it is the smallest eigenvalues that have imaginary parts, but as the current increases larger and larger eigenvalues acquire imaginary parts. Finally this process reaches the eigenvectors (which have imaginary parts once the eigenvalues do) that we wish to plot; that is, the leading left eigenvectors are complex.

At this point, we are left without theorems. Barycentric coordinates, simplices, all those results only hold for real eigenvectors. What to do? Gaveau and I have managed some theorems about high powers of R, but have not yet found them to be practical. Also Nicholson et al. [19] have proposed another operator, call it B, with dynamics close to that of R, which does satisfy detailed balance [20]. However, the dynamics of B may be close, but in my opinion, not close enough. So the short answer to my earlier question, like much else in the use of the OR, is "try." What happens if a left eigenvector that you plan to plot is complex, take its real and imaginary parts and plot them. This actually does not change the number of left eigenvectors plotted, since there is another left eigenvector that is the complex conjugate of the one you're plotting. (In case that complex eigenvector is the last one to be plotted its real part is usually taken.) Sometimes this works to give interesting results, sometimes not.

An example is the permutation operator. Except for a small amount of random noise that I impose (in order that λ_0 be the only eigenvalue of absolute value 1) this matrix takes 2 to 1, 3 to 2, $\dots$, and 1 to N (with $N = 30$ in this case). This gives exactly the same representation as does the random walk on the circle, which has detailed balance. So taking real and imaginary parts can work in some cases. See Figure 14.
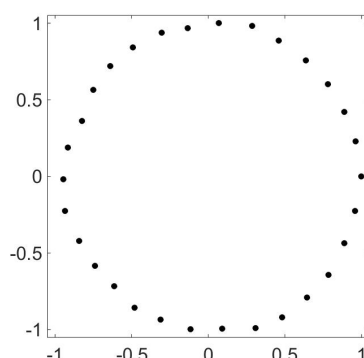
Figure 14. Modified OR for a permutation, which has complex roots. It is modified in the sense that instead of plotting real eigenvectors against each other, the plot is of the real and imaginary parts of the first eigenvector.

Another case in which taking the real and imaginary part works is a walk in a potential. On the left of Figure 15 is the OR for a walk in a potential with 3 minima. For this walk [21] it is found that the extrema of the triangle are are the points numbered 196, 522 and 724 (the stochastic matrix is 961-by-961 ($= 31^2$). In the right hand figure I've added a term that allows a travel shortcut between the minima, a matrix connecting the three extrema. This makes the eigenvalue complex and induces a current. The right hand figure of Figure 15 uses the real and imaginary parts of the first non-trivial eigenvector. As you can see, it is (essentially) the same diagram. This does not always happen, but in this case it did.
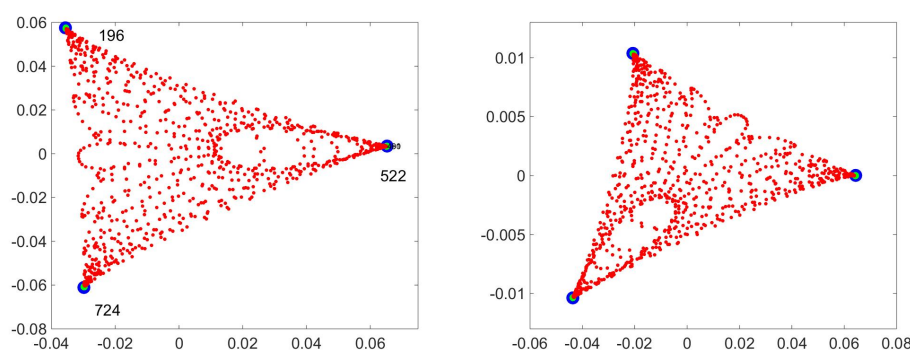

Figure 15. On the left is the OR for a three minima potential whose eigenvalues are real. On the right, I have added to the transition probability a travel shortcut between the minima, making the eigenvalues (and eigenvectors) complex. Up to re-orientation there is no difference between the figures.

5. Summary and Conclusions

The observable representation (OR) is a dynamics-respecting embedding of the space of states into Euclidean *m*-space, where typically *m* is two or three. A "state" is a recognizable configuration of a system, typically a coarse grain. "Dynamics respecting" means that typically the easier it is for a given state to become another state under the dynamics, the closer they will be in the OR. The original motivation for this work concerned metastable phases [22]. Conventional definitions of phase transitions involving analyticity fail when applied to metastable phases, leading Bernard Gaveau and myself to a different definition, one depending on the spectrum of the matrix of transition probabilities. The matrix, called *R* in the present article, plays a central role in the definition of the OR. Call its eigenvalues $\lambda_0, \lambda_1, \ldots$ in descending value (which may need further definition; see below). Then $\lambda_0 = 1$, and near degeneracy of λ_1 may be the hallmark of a second phase, i.e., there may be two collections of states of *X* representing two phases, one of which may be metastable. This can

easily be generalized to more phases although embedding more than 4 phases may require more than three dimensions.

The original work on phases can be summarized as follows. If there are $m + 1$ phases there need to be m eigenvalues nearly degenerate with 1, and $|\lambda_{m+1}| \ll \lambda_m$. The OR in this case becomes (to a good approximation) a simplex with the barycentric coordinates of internal points the probability that with that point as an initial value the system goes to one or another phase.

However, the OR turns out to have meaning well beyond the definition of phases. For a random walk without potential it can reproduce the space on which the walk takes place. If there are more extrema than can be imaged in three dimensions there can be ways to pick out the overall structure. In addition, there are surprises, for example in the linguistic analysis of the book, Moby Dick, for reasons unknown (to me) the OR is a tetrahedron with the rare letters the extrema.

Finally there is a special class of images for which we do not have theorems but which sometimes work. I refer to the situation when some of the large eigenvalues are complex. In this case I took the real and imaginary parts and a useful image emerged. The ordering of the eigenvalues is also affected and usually they are ordered by absolute value, with an arbitrary choice within one particular absolute value.

To summarize, if it is easy to do, do it. For some cases it is known that the OR gives useful information, for some case it manages to do so despite the absence of theorems. Various special cases have been studied, but for a stochastic process that does not fit in one of the categories studied, there may or may not be useful information in the plot.

Appendix A. How to Handle (Some) Large Spaces

It often happens that a state space is large. For example in the foraging case, even coarse graining, still leaves a 50-by-50 state space. Thus "X" is of cardinality 2500 and would require a transition probability matrix that is 2500-by-2500. Or one could try to create an OR for the Jensen evolutionary process ("Tangled Nature," see, e.g., [23]) where the state space can be infinite.

In practice it is generally easier to trace a particular path through the space than to write down an entire transition matrix. In this section I will describe how to use such a path to recover an approximation to the transition matrix. Generally it involves fewer points than the entire state space.

Suppose one is given a specific random walk, $x(t)$, $t = 1, \ldots, T$. The nature of the space in which x takes its values is irrelevant, and we replace it by integers. Thus $x(1)$ is just 1. If $x(2)$ is different it is called 2; otherwise it is also 1. The total state space *of interest in the present construction* is the set of unique values taken by $\{x(t) \mid t = 1, \ldots, T\}$. The actual state space may in principle be infinite, but this is not a problem. Let the number of unique sites be called N, so that "R," the set of approximate transition probabilities, is an $N \times N$ matrix.

Define an intermediate quantity $S(a, b)$ to be the number of times (out of all $T - 1$ time steps) that b goes to a, where a and b are integers (i.e., $x(t) = b$ and $x(t + 1) = a$). The column sums of S are of course not in general 1, but that is less serious than the fact that S may be reducible. A way to cure this is to introduce an artificial step: Let $x(T + 1) = x(1)$ and add this contribution to S.

As indicated in Section 3.6 there are various ways to make S stochastic. Let s be its column sums and $\widetilde{s^{-1}}$ the diagonal matrix made from the inverses of s. In this section we multiply the matrix S by $\widetilde{s^{-1}}$. (In MATLAB$^{\text{TM}}$ notation the sum is `sum(S)` and the method used here is `R=S*diag(1./sum(S))`, where `diag` is the diagonal matrix whose (possibly) non-zero elements are the argument of `diag`.)

It must be emphasized that this is an uncontrolled approximation (at least as far as the author knows). It works when it works. What is shown in [24] is that sometimes it does work (as in the case

of foraging) and sometimes it does not. The latter situation does not lead to publication, which may cause a biased view.

One way to see the efficacy of this strategy is to produce a path using a definite set of transition probabilities, to be called R_{orig}. Then use the method described above to see how effective the recovery is. In [24] it is found that there are several major factors in the accuracy of the recovery. The first is T, the length of the path. A measure of accuracy is the largest (in absolute value) discrepancy, i.e., $\max_x |R(x) - R_{\mathrm{orig}}(x)|$, and this is found to decrease as $T^{-1/2}$. See Figure A1. Then there is the size of the matrix. Obviously the larger the matrix, the more steps are needed. However, if additional steps are given the discrepancy actually decreases. Roughly, for N up to $N = 2048$ and the time growing linearly in N, the decrease in accuracy (for an N-by-N matrix) went roughly like Discrepancy $\sim N^{-0.4}$. Finally, there is the possible existence of metastable states. The path $x(t)$ may get stuck in one particular metastable state and not explore others, even for large T. This was tested in the potential illustrated in Figure A2 and the results of the procedure described above for various times (T) are also shown. Note that even for the shorter time, when two of the metastable states are not explored at all, the pictures of the states that *were* visited is fairly accurate.
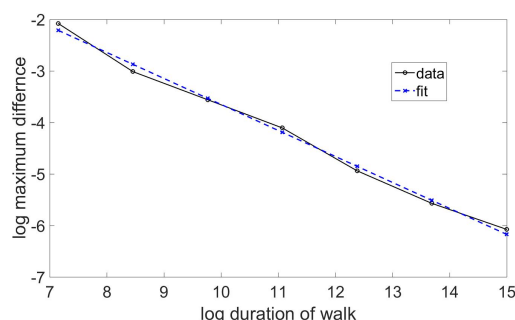

Figure A1. Dependence of $\max_x |R(x) - R_{\mathrm{orig}}(x)|$ on length of the path for a system without metastable states.
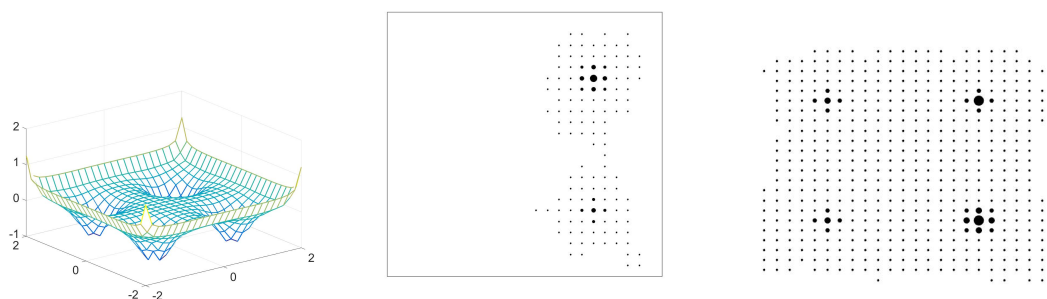

Figure A2. On the left is a potential with 4 minima, two of them slightly deeper than the others. The other two figures show the occupation: the size of a dot is proportional to the weight of the stationary distribution of the matrix R for the given path length. For the middle figure $T \sim 2 \times 10^5$ and a boundary is shown to emphasize the regions *not* visited. For the image on the right much longer times are used.

Appendix B. Barycentric Coordinates

Let $(y_1, \ldots, y_{n+1})$ be the vertices of a simplex in n-dimensional Euclidean space. (Each y_k is itself an n-vector.) Then a vector x contained in the simplex, which is convex, can be written

$$x = \sum_k \mu_k y_k, \tag{B1}$$

with each $\mu_k \geq 0$ and $\sum_k \mu_k = 1$. If a point lies outside the simplex it can still be written in the form (B1), but at least one of the μ_k's will be negative.

In one dimension, $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ are just points on a line and any other point can be written $\boldsymbol{x} = \mu_1 \boldsymbol{y}_1 + \mu_2 \boldsymbol{y}_2$ with $\mu_k = \frac{x - y_{k'}}{y_k - y_{k'}}$ and $k' \neq k$. In the general case one can use $\mu_{n+1} = 1 - \sum_1^n \mu_k$ to obtain n equations in n unknowns (the remaining μ_k's) with non-vanishing matrix because the simplex has strictly positive volume in n-space.

References and Notes

1. Landau, L.D.; Lifshitz, E.M. Course of theoretical physics. In *Statistical Physics*, 3rd ed.; Part 1; Pergamon Press: Oxford, UK, 1980; Volume 5, pp. 13–14.
2. Pearson, K. On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **1901**, *2*, 559–572. [CrossRef]
3. See the Wikipedia Entry for Principal Component Analysis. Available online: https://en.wikipedia.org$\penalty\z@$/wiki/Principal_component_analysis (accessed on 14 March 2019).
4. Macieszczak, K.; Guță, M.; Lesanovsky, I.; Garrahan, J.P. Towards a theory of metastability in open quantum dynamics. *Phys. Rev. Lett.* **2016**, *116*, 240404. [CrossRef] [PubMed]
5. Reversing the arrow in (1) or (2) implies that left eigenvectors become right and right left. The OR in that case would be focused on *right* eigenvectors rather than left.
6. Gantmacher, F.R. *Matrix Theory*; Chelsea: New York, NY, USA, 1960; this is a two-volume set, translated from the Russian by Hirsh, K.A., The relevant material is in Volume II, Chapter XIII.
7. Horn, R.A.; Johnson, C.R. *Matrix Analysis*; Cambridge University Press: Cambridge, UK, 1985.
8. Schulman, L.S.; Luck, J.M.; Mehta, A. Spectral properties of zero temperature dynamics in a model of a compacting granular column. *J. Stat. Phys.* **2012**, *146*, 924–954. [CrossRef]
9. Isakov, S.N. Nonanalytic features of the first order phase transition in the ising model. *Commun. Math. Phys.* **1984**, *95*, 427–443. [CrossRef]
10. Hodgman, C.; Weast, R.C.; Shankland, R.S.; Selby, S.M. *Handbook of Chemistry and Physics*, 44th ed.; The Chemical Rubber Publishing Co.: Cleveland, OH, USA, 1963; p. 2198.
11. A critical droplet is a piece of the stable phase that is large enough to grow, rather than shrink.
12. Debenedetti, P.G.; Stanley, H.E. Supercooled and glassy water. *Phys. Today* **2003**, *56*, 40–46. [CrossRef]
13. Gaveau, B.; Schulman, L.S. Multiple phases in stochastic dynamics: Geometry and probabilities. *Phys. Rev. E* **2006**, *73*, 036124. [CrossRef] [PubMed]
14. Gaveau, B.; Schulman, L.S.; Schulman, L.J. Imaging geometry through dynamics: The observable representation. *J. Phys. A* **2006**, *39*, 10307–10321. [CrossRef]
15. Schulman, L.S. Mean field spin glass in the observable representation. *Phys. Rev. Lett.* **2007**, *98*, 257202. [CrossRef] [PubMed]
16. Wosniack, M.E.; Santos, M.C.; Raposo, E.P.; Viswanathan, G.M.; da Luz, M.G.E.; Schulman, L.S. Identifying dynamical structures in the physical space of stochastic processes. *EPL* **2019**, *125*, 20004. [CrossRef]
17. Schulman, L.S.; Bagrow, J.P.; Gaveau, B. Visualizing relations using the "observable representation". *Adv. Comp. Syst.* **2011**, *14*, 829–851. [CrossRef]
18. Schulman, L.S.; Gaveau, B. Complex systems under stochastic dynamics. In *Anomalies and Strange Behavior in Physics: Challenging the Conventional*; Istituto Italiano per gli Studi Filosofici Palazzo Serra di Cassano: Napoli, Italy, 2003; pp. 805–818.
19. Nicholson, S.B.; Schulman, L.S.; Kim, E. Deciphering interactions of complex systems that do not satisfy detailed balance. *Phys. Lett. A* **2013**, *377*, 1810–1813. [CrossRef]
20. The matrix B is defined by $B_{xy} = R_{xy} - J_{xy}/2p_0(y)$. This satisfies detailed balance, whereas if $J \neq 0$, R does not.
21. The potential is $\propto 1/$distance and the temperature is low. The lattice is 31 by 31.
22. Often "phases" are called "states," where the word "state" is being used differently. If X is the collection of "recognizable configurations," (what we have defined as states) and $x, y \in X$, then the meaning of phase is a collection of points $\{x\} \subset X$. Nevertheless, the usage "metastable state" is sufficiently common that confusion can arise. In this article states are individual points of X, while phases are collections of them.

23. Hall, M.; Christensen, K.; di Collobiano, S.A.; Jensen, H.J. Time-dependent extinction rate and species abundance in a tangled-nature model of biological evolution. *Phys. Rev. E* **2002**, *66*, 011904. [CrossRef] [PubMed]

24. Schulman, L.S. Transition matrix from a random walk. *arXiv* **2016**, arXiv:1605.04282v3.