Journal of Statistical Physics, Vol. 10, No. 1, 1974

Gibbs and Markov Random Systems with Constraints

John Moussouris[1]

Received August 3, 1973

This paper concerns random systems made up out of a finite collection of elements. We are interested in how a fixed structure of interactions reflects on the assignment of probabilities to overall states. In particular, we consider two simple models of random systems: one generalizing the notion of "Gibbs ensemble" abstracted from statistical physics; the other, "Markov fields" derived from the idea of a Markov chain. We give background for these two types, review proofs that they are in fact identical for systems with nonzero probabilities, and explore the new behavior that arises with constraints. Finally, we discuss unsolved problems and make suggestions for further work.

KEY WORDS: Random system; Markov assumption; local Markov conditions; Gibbs potential; Gibbs–Markov equivalence; inversion formula for potentials; constraints; barriers and wells; limit representations; higher-order equations; strongly Markovian systems.

1. PRELIMINARIES

We describe a system by means of a finite graph G whose vertices Z represent components of the system called *sites*, and whose (undirected) edges represent mutual interactions. Each site $z \in Z$ can assume any of a range Ω_z of possible (*elementary*) *states*; this repertoire may vary from site to site. An assignment

[1] The Mathematical Institute and Merton College, Oxford, England. Present address: Project MAC, Massachusetts Institute of Technology, Cambridge, Massachusetts.

of values to all the sites in Z determines a *state* of the overall system: so the space of states is simply the Cartesian product $\Omega = \prod_{z \in Z} \Omega_z$. We also consider *substates* which assign values only to some subset of Z. For example, if $A \subseteq Z$, any list $\alpha \in \prod_{z \in A} \Omega_z$ is a substate whose *domain* is A (in general, lower case Greek letters $\alpha, \beta, \gamma, \delta, \epsilon, \ldots$ will be used for substates, and italic Roman capitals $A, B, C, D, E, \ldots$ for their corresponding domains). We may write $\omega = \alpha\beta$ to indicate that the state $\omega \in \Omega$ is built up out of substates α and β—or more precisely, that Z is the disjoint union of A and B, and the states agree on their respective domains of definition. We use "state" loosely to include "substate" or "elementary state," when context makes the meaning clear.

That two sites are connected by an edge of the graph indicates that they *interact*. The sites which interact with those of a given subset $A \subseteq Z$ compose the *environment* of A. Parts of the system which do not interact are called *independent*. In other words, two disjoint subsets of Z are independent when one is outside of the environment of the other.

Here we are concerned with probabilistic descriptions of the system. Thus to each state $\omega \in \Omega$ we assign a nonnegative likelihood $P(\omega)$. If α is a substate, we may also define a probability $P(\alpha\cdot) = \sum_\beta P(\alpha\beta)$, where the sum ranges over all possible β—that is, over all states $\alpha\beta \in \Omega$ which agree with α on its domain. In this notation the usual convention for normalization is $P(\cdot) = 1$. But we do not make this assumption, because it would just be a nuisance in proofs, and probabilities can always be normalized as the last step in a construction anyway. A triplet (G, Ω, P) will be called a *random system.*

How should the structure of interactions in the system represented by G be related to the probability distribution? Certainly the likelihood of a particular substate on some portion of the system A should change with different conditions on its environment. We make the simplifying assumption, however, that this probability does *not* depend on any subset B independent of A: That is, once the state of the environment is known, any further information about states outside of the environment does not effect our expectations about A. In particular, there are no "indirect" influences: If a site x interacts with y, and y with z, it is *not* true that x influences z, unless these sites are linked directly. This assumption resembles the definition of a Markov chain, in which the likelihood of a particular state at the present time depends on the state at the previous moment, but not on any earlier details of past history. In our situation, of course, the pattern of interdependences recorded in G is finite and undirected, and need not take the form of a temporal chain. Nevertheless, the analogy suggests calling systems which satisfy the above assumption "Markovian."

To make this notion precise, we specify how the likelihood of a substate

α is modified once a state γ has been fixed in some disjoint part of the system. Thus in the usual way we define the conditional probability α *given* γ to be

$$P(\alpha|\gamma) = P(\alpha\gamma\cdot)/P(\cdot\gamma\cdot)$$

i.e., the fractional chance that, among all the states that include γ, the substate α will hold as well (we have written $P(\cdot\gamma\cdot)$ rather than $P(\gamma\cdot)$ to emphasize that the range of the sum in the denominator includes all possible substates on the domain A). Now the Markov assumption becomes: If E is the environment of the domain A, and C is what remains independent of A outside of the environment, then

$$P(\alpha|\epsilon) = P(\alpha|\epsilon\gamma)$$

that is, the probability of α given ϵ is unchanged when γ is known as well.

There are several equivalent descriptions of a Markov system, some of which are intuitively or algebraically simpler than the above. Using the definition of conditional probabilities and multiplying out the denominators, we get

$$P(\alpha\epsilon\gamma)P(\cdot\epsilon\cdot) = P(\alpha\epsilon\cdot)P(\cdot\epsilon\gamma) \tag{M1}$$

This form has the advantage of being well defined even if P is allowed to vanish on some states. We note that $P(\alpha\epsilon\gamma)$ depends on α and γ through separate factors. Thus if α' and γ' are different substates on the same domains A and C, the summations can be eliminated as follows:

$$\begin{aligned}(\alpha\epsilon\gamma)(\cdot\epsilon\cdot)(\alpha'\epsilon\gamma')(\cdot\epsilon\cdot) &= (\alpha\epsilon\cdot)(\cdot\epsilon\gamma)(\alpha'\epsilon\cdot)(\cdot\epsilon\gamma') \\ &= (\alpha\epsilon\cdot)(\cdot\epsilon\gamma')(\alpha'\epsilon\cdot)(\cdot\epsilon\gamma) \\ &= (\alpha\epsilon\gamma')(\cdot\epsilon\cdot)(\alpha'\epsilon\gamma)(\cdot\epsilon\cdot)\end{aligned}$$

where the Ps have been omitted from in front of the parentheses. If we divide by $(\cdot\epsilon\cdot)^2$ and restore the Ps, we find

$$P(\alpha\epsilon\gamma)P(\alpha'\epsilon\gamma') = P(\alpha\epsilon\gamma')P(\alpha'\epsilon\gamma) \tag{M2}$$

Conversely, if (M2) holds, (M1) follows by summing over all possible states α' and γ'. This derivation is valid even when some of the probabilities are zero, for $P(\cdot\epsilon\cdot)$ cannot vanish unless $P(\alpha\epsilon\gamma) = 0$ for all α and γ, in which case both (M1) and (M2) are satisfied trivially.

Note that although the domain C in (M2) is the largest subset independent of A, the converse need not be true: the environment of C may be smaller than E (see example in Fig. 1). We remove this asymmetry in our final form for the *Markov conditions*: If A and B are independent domains, then

$$P(\alpha\sigma\beta)P(\alpha'\sigma\beta') = P(\alpha\sigma\beta')P(\alpha'\sigma\beta) \tag{M}$$

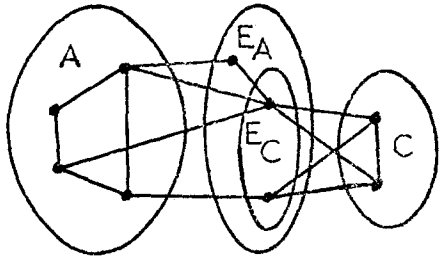

Fig. 1. The environment E_C is a proper subset of E_A.

Here S is a *separator* of the domains A and B—that is, a subset that contains both of their environments (but may not be equal to either). We see that (M2) follows from (M) by specializing S to be the environment E of A. Conversely, all conditions (M) appear in disguise among the (M2): For, if $\sigma = \epsilon\delta$—i.e., S is bigger than the environment of A—simply take $\gamma = \delta\beta$ and $\gamma' = \delta\beta'$ in (M2) and retrieve (M). A random system that satisfies (M) for all α, σ, β will be called *Markovian.*

In the case of positive systems ($P > 0$) we may write (M) in terms of ratios:

$$P(\alpha'\sigma\beta)/P(\alpha\sigma\beta) = P(\alpha'\sigma\beta')/P(\alpha\sigma\beta')$$

obtaining another paraphrase of the Markov assumption: The relative chances of two alternative substates α' and α are unaffected by changes on domains independent of A. The fact that only relative probabilities are determined by the Markov conditions suggests referring all states to some arbitrary fixed state, which we will call the *null state* and designate o. Then the relative chance of α' and α may be written $P(\alpha'\epsilon o)/P(\alpha\epsilon o)$, to emphasize that it depends only on the state ϵ of the environment (the same symbol o will denote any substate of the reference state, when the domain is clear from the context).

The Markov conditions (M) are in general highly redundant. We single out a special class of them called *local Markov conditions*, in which states vary from the null state at single sites: If z_1 and z_2 are noninteracting sites, then

$$P(\zeta_1\sigma\zeta_2)P(o\sigma o) = P(\zeta_1\sigma o)P(o\sigma\zeta_2) \qquad \text{(LM)}$$

where ζ_1 and ζ_2 are states on z_1 and z_2, respectively. Assuming $P > 0$, it is easy to see that all the conditions (M) can be reconstructed from these local conditions. First check the case of the domains' being single sites:

$$(\zeta_1\sigma\zeta_2)(\zeta_1'\sigma\zeta_2') = \frac{(\zeta_1\sigma o)(o\sigma\zeta_2)(\zeta_1'\sigma o)(o\sigma\zeta_2')}{(o\sigma o)^2} = (\zeta_1\sigma\zeta_2')(\zeta_1'\sigma\zeta_2)$$

where we have omitted Ps and used the same arrangement of factors as in the derivation of (M2) from (M1). Next, suppose that (M) is true for all α, β such that $|\alpha| + |\beta| \leqslant n$ ($|\alpha|$ is the number of sites in A). Then if we enlarge α to $\alpha\zeta$ (defined on one more site),

$$\begin{aligned}(\alpha\zeta\sigma\beta)(\alpha'\zeta\sigma\beta')(\alpha'\zeta'\sigma\beta') &= (\alpha\zeta\sigma\beta')(\alpha'\zeta\sigma\beta)(\alpha'\zeta'\sigma\beta') \\ &= (\alpha\zeta\sigma\beta')(\alpha'\zeta\sigma\beta')(\alpha'\zeta'\sigma\beta)\end{aligned}$$

where we have used (M) on α and β for the first equality, and on ζ and β for the second. Dividing both sides by $(\alpha'\zeta\sigma\beta')$, we get

$$P(\alpha\zeta\sigma\beta)P(\alpha'\zeta'\sigma\beta') = P(\alpha\zeta\sigma\beta')P(\alpha'\zeta'\sigma\beta)$$

which is an instance of (M) with $|\alpha\zeta| + |\beta| = n + 1$. Hence by induction, positive Markov systems are exactly *local Markov systems* satisfying the local conditions.

Now the Markov conditions ought to reduce the amount of information needed to specify the distribution on a Markov system. In particular, if $\omega = \zeta_1\sigma\zeta_2$ is *active* (i.e., different from the null state) on at least two independent sites z_1 and z_2, then (LM) expresses $P(\omega)$ in terms of probabilities of states active on a smaller part of the system. We can continue this reduction until we arrive at states whose active portions contain no independent sites. Such totally interdependent subsets of the system are called *clans* (in other papers, they are called "cliques"; but graph-theoretic usage reserves this term for *maximal* complete subgraphs, whereas here we consider *any* subset which has no unlinked vertices, including complete subgraphs, single sites, and even the empty set). A state of the form κo ($\kappa, \kappa', \kappa'',\ldots$ will be used to denote active substrates whose domains are clans) is called a *clan state*. We have seen that once probabilities are assigned to the class of such states $\Lambda \subseteq \Omega$, the distribution over all the rest of the state space is fixed by local conditions.

So here is another interpretation of the Markov assumption: The probability of a particular overall excitation is compounded out of the likelihoods of activity in clans of the system. Although this result simplifies P assignment for a Markov system, it is not entirely satisfactory, because the relationship between $P(\omega)$ and its component $P(\kappa o)$s may be rather complicated (we will see what it is later on). More fundamentally, we have given only an extrinsic account of the way interactions reflect in the distribution—i.e., by means of relations between probabilities of states which differ on independent parts of the system. Now we take a different point of view, and describe the probability assignment in terms of a simpler underlying structure, in such a way that the Markov relations are automatically fulfilled. This simple structure is a "potential function," and we might be led to consider it by the familiar result about equilibria from statistical mechanics: The

probability of a given state of a system is proportional to $e^{-c\Phi}$, where c is a constant (depending on temperature) and Φ is the total potential energy of the state.

The usefulness of potentials in physics relies upon the additivity of energy: The total Φ is just the sum of terms from all the interacting parts of the system. In Markov systems the "units" of interaction are clans, so we seek to describe elementary terms as *clan potentials* defined by a real-valued function φ on the clan states. A potential is called *Gibbsian* when it is of the form

$$\Phi(\omega) = \sum_{\kappa \prec \omega} \varphi(\kappa)$$

where $\kappa \prec \omega$ means "κ is an *active* (clan) substate of ω." Thus if ω has the null value at some site $z \in Z$, there will be no contribution either from the singleton clan $\{z\}$, nor from any larger clan containing z. It seems arbitrary to exclude partially active clans, but this convention is exactly what is needed to remove the ambiguity in "reference level" of the various potential terms. As we shall see later, allowing contributions from arbitrary clan substates adds nothing to our description except redundancy. A random system in which $P = e^{-\Phi}$, where Φ is a Gibbsian potential, is called a *Gibbs system.*

We digress to clarify how our presentation differs from earlier ones in this field. The above definitions of Gibbs and Markov systems agree with those given by Hammersley and Clifford,[1] who extended these notions to arbitrary finite patterns of interaction. Other workers such as Dobrushin,[2] Avertinsev,[3] and Spitzer[4] were interested in modeling particle equilibria in a discrete space, so they took G to be a region of the simple cubic lattice. In this context the terminology used was "Gibbs ensembles" and "Markov fields," but we speak of "systems" to emphasize that the results apply to general arrangements of interrelated parts. In particular, there is no need to assume that interacting elements are close together, so we say that sites "interact" rather than "are neighbors," and use "environment" and "clan potential" rather than "boundary" and "local potential."

Also, note that we have made no restrictions on the elementary state spaces Ω_z. In the initial particle models Ω_z expressed occupancy at the corresponding site; and perhaps for this reason it has usually been assumed that the number of states available is finite or at least countable. In fact, the Gibbsian sum $\sum_{\kappa \prec \omega} \varphi(\kappa)$ is well defined regardless of the number of states, since it can have no more terms than there are clans in G. In the proofs given below of Gibbs–Markov equivalence, Ω_z can be arbitrary sets. Thus, for example, the kilowatt loads of generators in a power network would be acceptable elementary states. It is necessary, however, that some measure be available on each set Ω_z, in order to compute normalization constants or substate probabilities.

2. PROOFS OF GIBBS–MARKOV EQUIVALENCE

One feature of a system in which $P = e^{-\Phi}$ is that $P(\omega)$ is strictly positive: Vanishing probabilities are excluded throughout this section. In this context it would be natural to define a function $\Phi = -\log P$ even without the prompting of statistical physics, simply as a device for converting the quadratic Markov conditions (M) into linear form:

$$\Phi(\alpha\sigma\beta) + \Phi(\alpha'\sigma\beta') = \Phi(\alpha\sigma\beta') + \Phi(\alpha'\sigma\beta) \qquad (\mathrm{M}\Phi)$$

It is not surprising that potentials additive on the "irreducible" states of the system satisfy these identities.

Theorem 1. Every Gibbs system is Markovian.

Proof. We show that any $\Phi(\omega) = \sum_{\kappa \prec \omega} \varphi(\kappa)$ satisfies (MΦ) by counting the contribution of clan substates on each side of the equation.

The key observation is that no clan can intersect both A and B, since otherwise there would be links between independent sets. Thus

$$\Phi(\alpha\sigma\beta) = \sum_{\kappa \prec \alpha\sigma} \varphi(\kappa) + \sum_{\kappa \prec \sigma\beta} \varphi(\kappa) - \sum_{\kappa \prec \sigma} \varphi(\kappa)$$

the last term preventing clan substates of σ from being counted twice. With the abbreviation $[\gamma] = \sum_{\kappa \prec \gamma} \varphi(\kappa)$, (M$\Phi$) becomes

$$\begin{aligned} &[\alpha\sigma] + [\sigma\beta] - [\sigma] + [\alpha'\sigma] + [\sigma\beta'] - [\sigma] \\ &\quad = [\alpha\sigma] + [\sigma\beta'] - [\sigma] + [\alpha'\sigma] + [\sigma\beta] - [\sigma] \quad \blacksquare \end{aligned}$$

The converse result is less trivial. We give three proofs: first, an elegant construction (due to Grimmett[5]) of an explicit formula for the clan potentials in terms of Φ; second, a proof by induction; and third, a clan-counting argument that ties together some ideas in this section and the next.

Theorem 2. Every positive Markov system is Gibbsian, with clan potentials uniquely determined by

$$\varphi(\kappa) = \sum_{\kappa' \prec \kappa} (-1)^{|\kappa| - |\kappa'|} \Phi(\kappa' o)$$

($|\kappa|$ is the number of sites in the domain K).

Proof 1. We start from a well-known inversion formula for arbitrary real functions on the subsets of some finite set:

$$F(A) = \sum_{B \subseteq A} G(B) \qquad \text{iff} \quad G(B) = \sum_{C \subseteq B} (-1)^{|B| - |C|} F(C)$$

or, equivalently,

$$F(A) = \sum_{B \subseteq A} \sum_{C \subseteq B} (-1)^{|B| - |C|} F(C)$$

where $B \subseteq A$ means that B is a subset of A, and $|B|$ is the number of elements in B. The second identity holds because $F(A)$ occurs only once in the sum, whereas any C with m fewer elements than A can be extended $\binom{m}{j}$ ways to Bs containing j of the missing elements; since $|B| - |C| = j$, the total coefficient of $F(C)$ is then $\sum_{j=0}^{m} (-1)^j \binom{m}{j} = (1-1)^m = 0$.

Application of this result gives

$$\Phi(\alpha o) = \sum_{\beta \prec \alpha} \Gamma(\beta) \qquad \text{iff} \quad \Gamma(\beta) = \sum_{\gamma \prec \beta} (-1)^{|\beta|-|\gamma|} \Phi(\gamma o)$$

where α is an active substate, and β and γ are substates of α, with $|\beta|$ and $|\gamma|$ active sites, respectively. Now we need only show that $\Gamma(\beta) = 0$ whenever B is not a clan: then we can identify $\varphi(\kappa) = \Gamma(\kappa)$ for clans K and get

$$\Phi(\alpha o) = \sum_{\kappa \prec \alpha} \varphi(\kappa) \qquad \text{iff} \quad \varphi(\kappa) = \sum_{\kappa' \prec \kappa} (-1)^{|\kappa|-|\kappa'|} \Phi(\kappa' o)$$

So suppose $\beta = \zeta_1 \sigma \zeta_2$, where z_1 and z_2 are independent sites. Then any $\gamma \prec \beta$ must be obtained from $\gamma' \prec \sigma$ by adjoining one, none, or both of the states ζ_1 and ζ_2. Hence

$$\Gamma(\beta) = \sum_{\gamma' \prec \sigma} (-1)^{|\beta|-|\gamma'|} [\Phi(\zeta_1 \gamma' \zeta_2 o) + \Phi(o \gamma' o o) - \Phi(\zeta_1 \gamma' o o) - \Phi(o \gamma' \zeta_2 o)]$$

The bracketed expression vanishes, by the local Markov conditions.

Proof 2. We induct on the number of sites in Z.

Note first that if some sites are fixed in the null state, the system on the remaining part of G is still Markovian, because all the conditions for $\Phi'(\omega') \equiv \Phi(\omega' o)$ are among the (MΦ) (with the domain of the o substate included in the separators).

Our result is trivial for $|Z| = 1$. Make the hypothesis for $|Z| < n$, and consider a system with n sites. If G is a complete graph, we have the trivial case in which there are no Markov conditions, and the formulas relating Φ and φ reduce to the set-theoretic result, since all subsets of Z are clans. So suppose that there are two independent sites z_1 and z_2 in G. By the local Markov conditions

$$\Phi(\zeta_1 \sigma \zeta_2) = \Phi(\zeta_1 \sigma o) + \Phi(o \sigma \zeta_2) - \Phi(o \sigma o)$$

Now define Φ_1, Φ_2, and Φ_3 to be the fields obtained by fixing z_1, z_2 and $z_1 z_2$, respectively, in the null state. Each of these yields a Markov system with less than n sites, so by the inductive hypothesis they have corresponding clan potentials φ_1, φ_2, and φ_3. Thus

$$\Phi(\zeta_1 \sigma \zeta_2) = \sum_{\kappa \prec \zeta_1 \sigma} \varphi_1(\kappa) + \sum_{\kappa \prec \sigma \zeta_2} \varphi_2(\kappa) - \sum_{\kappa \prec \sigma} \varphi_3(\kappa)$$

But then we may define φ such that $\Phi(\omega) = \sum_{\kappa \prec \omega} \varphi(\kappa)$, by

$$\varphi(\kappa) = \begin{cases} \varphi_1 + \varphi_2 - \varphi_3 & \text{if } \kappa \prec \sigma \\ \varphi_1 & \text{if } \zeta_1 \prec \kappa \\ \varphi_2 & \text{if } \zeta_2 \prec \kappa \end{cases}$$

Note that if $\varphi_1, \varphi_2, \varphi_3$ satisfy the inversion formula, so does φ, because the $\Phi_i(\kappa o)$ coincide for $i = 1, 2, 3$ whenever $\kappa \prec \sigma$.

Proof 3. This is a proof by clan counting.

We say that a term $m\Phi(\alpha o)$, m an integer, "contains m appearances of the clan state κ" if $\kappa \prec \alpha$; the number of appearances in an expression is the sum of those in individual terms. Call an equation that has the same clan appearances on both sides "balanced." We show that any balanced equation must hold true in a Markov system.

The conditions (LM) can be used to reduce any $P(\omega)$ to $P(\kappa o)$, and hence any $\Phi(\omega)$ to its component $\Phi(\kappa o)$. By the proof of Theorem 1, such a reduction preserves balance. Thus we may assume that the given balanced formula is composed entirely of $\Phi(\kappa o)$s. We claim that each $\Phi(\kappa o)$ occurs the same number of times on both sides, so the equation is true identically. For, let E be a minimal equation in which this claim fails. If κ^* is a maximal clan appearing in E, it appears only in $\Phi(\kappa^* o)$, so the number of these terms must be the same on both sides of E for κ^* to balance. But then we may cancel the $\Phi(\kappa^* o)$ to get a smaller equation E' that is still balanced, and thus must satisfy the claim, contradicting the choice of E.

Now by the same counting argument used in Proof 1, we see that the expression

$$\varphi(\kappa) = \sum_{\kappa' \prec \kappa} (-1)^{|\kappa| - |\kappa'|} \Phi(\kappa' o)$$

has been cleverly designed to contain exactly one appearance of κ, and none of other clan states. Thus the equation

$$\Phi(\alpha o) = \sum_{\kappa \prec \alpha} \varphi(\kappa)$$

is balanced, and so must hold in a Markov system. ∎

Note that in all of these proofs only the local conditions were needed to derive the inversion formula, so with the help of Theorem 1 we have further demonstrations that every local Markov system is Markovian. Even these conditions (LM) are in general redundant. But the equations

$$\Phi(\alpha o) = \sum_{\kappa \prec \alpha} \sum_{\kappa' \prec \kappa} (-1)^{|\kappa| - |\kappa'|} \Phi(\kappa' o)$$

for A not a clan provide a sufficient set of independent relations on the

potential (if A is a clan, the equation holds identically). If there are a finite number N of states in Ω and k of them are clan states, there are exactly $N - k$ such relations bridging the gap from the $\varphi(\kappa)$ to $\Phi(\omega)$. Using these equations to verify the Markov property amounts to applying the inversion formula on clan substates and then checking whether potential sums predict the remaining probabilities correctly.

The inversion formula gives the *unique* value of $\varphi(\kappa)$ determined by the probability distribution. If a Gibbs potential were constructed which included extra contributions from partially active clans, the argument of Theorem 1 would still prove the system Markovian. But then the inversion formula could be applied to generate a potential in standard form, making the extra terms superfluous.

Using the definition of the potential, we may express the inversion formula directly in terms of probabilities:

$$\log P(\omega) = \sum_{\kappa' \prec \omega} \sum_{\kappa \prec \kappa'} (-1)^{|\kappa'| - |\kappa|} \log P(\kappa o)$$

Equivalently, we have the clan state decomposition formula promised in Section 1:

$$P(\omega) = \prod_{\kappa \prec \omega} P(\kappa o)^{n_{\omega\kappa}}, \qquad \text{where} \quad n_{\omega\kappa} = (-1)^{|\kappa|} \sum_{\kappa \prec \kappa' \prec \omega} (-1)^{|\kappa'|}$$

Note that this explicit expression was not needed in Proof 3 of the last theorem. We see in comparison that the formula

$$P(\omega) = \exp\left[-\sum_{\kappa \prec \omega} \varphi(\kappa) \right]$$

gives a much simpler reduction of the probability distribution to a function on the clan states.

But other than computational convenience, what is the significance of the Gibbs–Markov equivalence? Consider a random system of particles that interact only with their neighbors. We may interpret our proofs as follows: If influences propagate locally (in the sense of the Markov assumption), then $P = e^{-\Phi}$, where Φ is a sum of contact potentials. In other words, the canonical form of the probability distribution for particle equilibria depends only on (Markov) localization, and not upon any further details of the physics. Although this result is mathematically interesting, it does not displace any of the reasoning about ensembles in statistical mechanics. This reasoning is needed to establish that the $\log P$ potential corresponds with the physical measure of energy—that is, to connect statistics with dynamics. Once this connection has been established, the inversion formula for the local potentials may be useful in deducing from statistical measurements information about elementary forces in the system.

3. SYSTEMS WITH CONSTRAINTS

In some situations it may be convenient to impose constraints on a system so that nonzero probabilities are assigned only to some subset of the state space $\Omega^+ \subseteq \Omega$, which need not be a Cartesian product of elementary spaces. For example, these constraints might express the need for conservation of the particle number, total energy, or spin of the system. States in the region $\Omega^0 = \Omega - \Omega^+$ which violate the given requirements are *forbidden*—that is, assigned zero probabilities. In this context the Markov conditions (M) still make sense, so we may study their consequences even though $\log P$ is no longer well defined, and we do not have the convenience of an additive potential. Most of our results are pathological. But we establish criteria for a system to be expressible in terms of a suitable limit of Gibbs potentials. In this case a simple linear description is possible.

The first indication that earlier results go wrong is the existence of loca Markov systems that are not Markovian. Consider the graph in Fig. 2(a), where each of the three sites can have only two possible states: active (indicated by asterisks) and null. This is the simplest system with a *nonlocal* Markov condition, an example of which is given in Fig. 2(b). Note that none of the states in this equation differs at exactly two independent sites from another, so no pair can occur on the same side of a local condition. Thus if all the other states are forbidden, the (LM) have at least one vanishing probability on each side, and reduce to $0 = 0$, whether the distribution on Ω^+ is chosen to satisfy the global condition or not.

This example shows that the proofs of Theorem 2 cannot be expected to extend to the case with constraints, since they are all based solely on local conditions. Note, however, that a given system can be expressed as a continuous limit of positive Markov systems iff $P(\omega) = \exp[-\lim_i \sum_{\kappa \prec \omega} \varphi^i(\kappa)]$, since Gibbs–Markov equivalence holds at all the positive stages. Now one way such a $P(\omega)$ can vanish is that $\varphi^i(\kappa)$ can become infinitely large for some $\kappa \prec \omega$. Clan states that are energetically forbidden in this way are called *barriers*, and a Markov system in which all zero probabilities can be attributed to barriers is a *barrier system*. A necessary condition on such a system is that $P(\omega) = 0$ only when $P(\kappa\bullet) = 0$ for some (barrier) $\kappa \prec \omega$. In fact, this "barrier property" is also sufficient. For, define

$$\varphi^\rho(\kappa) = \begin{cases} \rho & \text{if } P(\kappa\bullet) = 0 \\ -\sum_{\kappa' \prec \kappa} (-1)^{|\kappa| - |\kappa'|} \log P(o\kappa') & \text{otherwise} \end{cases}$$

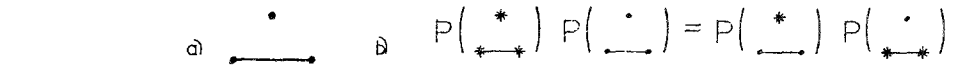

Fig. 2. The simplest local Markov system that is not Markovian.

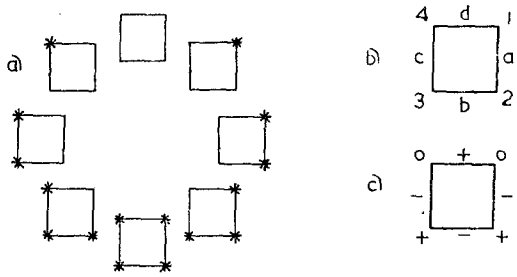

Fig. 3. A Markov system that violates the barrier property, but is expressible in terms of barriers and wells.

Then $\lim_{\rho\to\infty} \exp[-\sum_{\kappa\prec\omega} \varphi^\rho(\kappa)]$ will vanish whenever $P(\kappa\cdot) = 0$ for some $\kappa \prec \omega$, but will equal $P(\omega)$ whenever the probability is positive: for in this case $P(\alpha o) > 0$ for all substates α of ω, so $P(\omega)$ can be reduced to $P(\kappa o)$s just as in Theorem 2.

Note that the definition of barrier system depends upon a choice of reference state, because only active clans are candidates as barriers. But even if partially active clans are included, there are Markov systems that cannot be expressed in terms of barriers. Consider, for example, the four-site system with the eight states shown in Fig. 3(a) each having $\frac{1}{8}$ probability. If the other states are forbidden, it is easy to verify that the distribution is Markovian: for, note that states in Ω^+ that coincide on one of the diagonals (i.e., on a separator) are adjacent in the above octagon, and they differ only at a single site; so no nontrivial condition (M) can have two nonvanishing probabilities on the same side. On the other hand, none of the clan states satisfies $P(\kappa\cdot) = 0$, since every possible substate at each site or connected pair of sites occurs at least once among the Ω^+ states.

Nevertheless, it *is* possible to express the above system by means of a continuous limit of Gibbs potentials. Label the singletons and pair clans with numerals and letters as in Fig. 3(b), and then define

$$\varphi^\rho(\kappa) = \begin{cases} -\log\frac{1}{8} & \text{for } \kappa = 0 \\ 0 & \text{for } \kappa = 1, 4 \\ \rho & \text{for } \kappa = 2, 3, d \\ -\rho & \text{for } \kappa = a, b, c \end{cases}$$

Figure 3(c) abbreviates the $\pm\rho$ assignment. Note that any chain of active sites starting at 1 or 4 and curling around under the square encounters as many minuses as plusses, whereas other states have a majority of barrier substates. Thus $\exp[-\sum_{\kappa\prec\omega} \varphi^\rho(\kappa)]$ stays fixed at $\frac{1}{8}$ on Ω^+ states, but reduces

to $\frac{1}{8}e^{-\rho}$ on all states in Ω^0, which therefore have vanishing probability as $\rho \to \infty$.

In addition to the potential barriers $\varphi(\kappa) \to +\infty$, in this example we have introduced infinite potential "wells" $\varphi(\kappa) \to -\infty$. Now we can understand how $P(\omega)$ can vanish in a Gibbs–Markov limit without $P(\kappa\bullet) = 0$ for any $\kappa \prec \omega$: While an infinitely repulsive or forbidden configuration may exist in one part of the system, an infinitely attractive or favorable configuration may exist in another. The drawback of bottomless potential wells is that they can impart infinite energy to anything that falls in with them, so that the corresponding probabilities may diverge. Yet, as the above example shows, they may appear in such a way as to be completely masked by the repulsive potentials, manifesting themselves only as violations of the barrier property, while probabilities are bounded in the limit.

Are barriers and wells sufficient to describe *any* Markov system in terms of a limit of Gibbs potentials? Such a representation can be sought in two stages.

(a) Fill in zero probabilities with positive numbers, checking for values that are determined by the Markov conditions after each replacement. If there are no inconsistencies, this process yields a distribution P^+ that coincides with P on Ω^+, but is positive on the rest of Ω as well (e.g., $\exp[-\sum \varphi^0]$ above).

(b) Make use of any free parameters in the choice of the positive system to go to the limit in such a way that $P^+ \to P$ (e.g., the limit with respect to ρ above).

Stage (a) will fail only if some consequence of the Markov conditions is violated by the initial distribution on Ω^+. Now we have seen that substitution and cancellation of terms in (M) gives rise to "balanced" equations $\prod_{i=1}^n P(\omega_i) = \prod_{i=1}^n P(\omega_i')$, where $\{\omega_i\}$ and $\{\omega_i'\}$ contain the same clan states. Note, moreover, that stage (b) fails when one of these equations has some forbidden states ω_i^0 on one side, but is strictly positive on the other: For, then, any attempt to make some of the $P(\omega_i^0)$ go to zero forces the others to diverge. But if all the balanced conditions are satisfied, we expect that both stages of the limit construction will be feasible.

Extending the notion of a Gibbs ensemble to include the possibility of constraints, we call a random system which can be expressed in terms of a continuous limit of Gibbs potentials a *Gibbs–Markov system*. It is easy to see that the following holds.

Theorem 3. The balanced conditions (B) hold in a Gibbs–Markov system.

Proof. Each positive Markov system satisfies the balanced conditions, by Theorem 2. But then (B) hold in the limit as well, because the limit of a product is the product of the limits. ∎

The converse is less trivial. In order to make the two-stage algorithm sketched above precise, we formulate a linear framework. Regard the potential Φ as a vector Φ_ω indexed by the states $\omega \in \Omega$; clan potentials also form a vector φ_κ, indexed by clan states $\kappa \in \Lambda$. Let ζ be the matrix

$$\zeta_{\omega\kappa} = \begin{cases} 1 & \text{if } \kappa \prec \omega \\ 0 & \text{otherwise} \end{cases}$$

Then a potential Φ is Gibbsian iff $\Phi = \zeta\varphi$—that is, iff it is in the range of the transformation ζ. If ${\Phi_\omega}^+$ is defined on some subset $\Omega^+ \subseteq \Omega$, then it is a "restriction" of a Gibbs potential iff it is in the range of ζ^+, obtained from ζ by eliminating all the ωth rows for $\omega \notin \Omega^+$. Balanced conditions correspond to relations

$$\sum_\omega n_\omega \zeta_{\omega\kappa} = \sum_\omega m_\omega \zeta_{\omega\kappa} \qquad \text{or} \qquad (n - m)\zeta = 0$$

where n_ω and m_ω are lists of positive integers giving the exponents of $P(\omega)$ on the two sides of the equations. The heuristic arguments connecting conditions (B) with the two stages of limit construction are justified in the following lemmas (see appendix for proofs).

Lemma 1. Let M be a finite matrix of integers. Then v is in the range of M iff $nv = 0$ for every integer vector n such that $nM = 0$.

Lemma 2. Let M, N be finite matrices of integers. There exists an integer vector n^0 such that $Mn^0 = 0$ and $Nn^0 > 0$ iff there are no integer vectors m and $n \geqslant 0$ such that $mM + nN = 0$. (Here $>$ means that all components are strictly positive, and $\geqslant$ means that the components are nonnegative but not all zero.)

Now we are equipped to prove the following.

Theorem 4. A triplet (G, Ω, P) in which Ω is finite and P satisfies all the balanced conditions is a Gibbs–Markov system.

Proof. Decompose ζ into (ξ_0^+), where ζ^+ is indexed by the allowed states Ω^+ and ζ^0 by the forbidden ones Ω^0. The finiteness of Ω assures that ζ is a finite matrix, so the two lemmas apply. Take $M = \zeta^+$, $N = \zeta^0$, and $v = \Phi^+$, where ${\Phi_\omega}^+ = -\log P(\omega)$ for $\omega \in \Omega^+$.

(a) An integer vector n such that $n\zeta^+ = 0$ corresponds to a balanced condition, which would be violated if $n\Phi^+ \neq 0$. Hence by Lemma 1 there

exists a vector φ^+ satisfying $\Phi^+ = \zeta^+\varphi^+$. This clan potential defines an extension of P to a positive distribution on all of Ω.

(b) Integer vectors m and $n \geqslant 0$ such that $m\zeta^+ + n\zeta^0 = 0$ correspond to a balanced equation which has zero probabilities on one side, while the other side is strictly positive. Hence by Lemma 2 there exists an integer vector n^0 satisfying $\zeta^+ n^0 = 0$ and $\zeta^0 n^0 > 0$. But then if we define $\Phi^\rho = \zeta(\varphi^+ + \rho n^0)$, this potential stays fixed at Φ^+ on the states in Ω^+ regardless of any changes in the parameter ρ. Since, moreover, $\Phi_\omega{}^\rho \to \infty$ as $\rho \to \infty$ for $\omega \in \Omega^0$, we see that $\lim_{\rho\to\infty} \exp(-\Phi_\omega{}^\rho)$ is the desired limit representation.

An immediate consequence of Theorems 3 and 4 is the following result.

Corollary. A Gibbs–Markov system can always be expressed in the form $P(\omega) = \lim_{\rho\to\infty} \exp(-\Phi_\omega{}^\rho)$, where $\Phi^\rho = \zeta(\varphi^+ + \rho n^0)$. The integer components of n^0 record the relative heights of potential barriers and depths of potential wells needed to exclude or reinstate ωs in accordance with constraints.

To get more detailed information on Gibbs–Markov systems, we must investigate the balanced conditions (B). First of all, we note that the number of terms n must be the same on each side, because the o clan appears once in each $P(\omega_i)$ (or, equivalently, because the Markov conditions are invariant under changes in normalization). Thus the balanced equations can be classified according to their *order n*. The first few orders have simple structure, as follows.

$n = 1$. $P(\omega) = P(\omega')$. Since the same point states appear in ω and ω', these states must be identical, so the first-order conditions are all trivial equalities.

$n = 2$. $P(\omega_1)P(\omega_2) = P(\omega_1')P(\omega_2')$. One of the consequences of Theorem 2 is that an equation balanced on clans active with respect to a particular null state remains balanced under a change of reference. Thus we may suppose that, say, ω_2 is defined as the null state. Then active clans appear only once on the left side, as substates of ω_1. Thus if the active parts of ω_1' and ω_2' are α and β, respectively, the domains A and B must be disjoint, so $\omega_1 = \alpha o \beta$ and $P(\alpha o\beta)P(o) = P(\alpha o)P(o\beta)$. Moreover, A and B are independent, since otherwise there is a 2-clan on the left which does not appear on the right. Hence the above condition is simply an instance of (M). If we now restore the original reference state, we remain within (M), so the second-order equations are exactly the Markov conditions.

$n = 3$. $P(\omega_1)P(\omega_2)P(\omega_3) = P(\omega_1')P(\omega_2')P(\omega_3')$. Let θ be the substate made up of all those active elementary states that appear more than once among the ω_i (there can only be one multiple state at each particular site, so

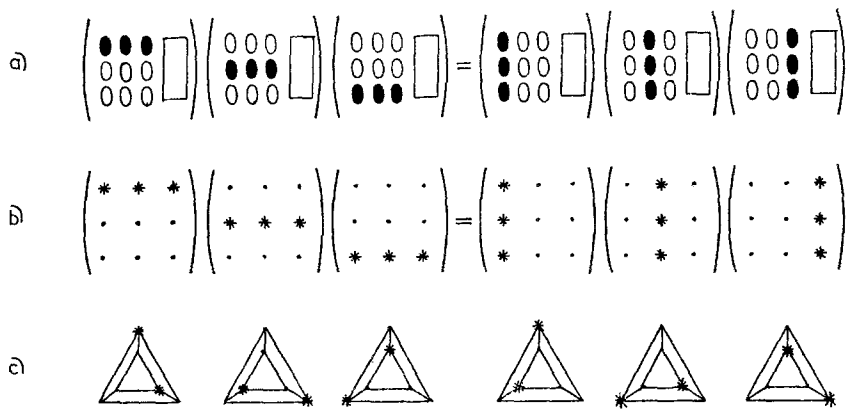

Fig. 4. Third-order conditions and examples of Markov systems that cannot be expressed as Gibbs limits in any form.

θ is well defined). Then by changing the reference state to θ_0, we induce a third-order equation with clan states appearing only once on each side: $P(\alpha_1 o)P(\alpha_2 o)P(\alpha_3 o) = P(\beta_1 o)P(\beta_2 o)P(\beta_3 o)$ where the α_i are active substates with disjoint domains (similarly the β_j). Note that $\bigcup_i A_i = \bigcup_i B_i$ is required to balance 1-clans. Now let γ_{ij} be the substate that coincides with α_i and β_j where they are the same; then $\alpha_i = \gamma_{i1}\gamma_{i2}\gamma_{i3}$ and $\beta_j = \gamma_{1j}\gamma_{2j}\gamma_{3j}$. The condition can be pictured as in Fig. 4(a), where the 3×3 array corresponds to the domains C_{ij} of the γ_{ij}, and the darkened rows and columns represent the activity of the α and β substates, respectively. The blank box is the part of the system which is fixed in the null state. Domains in the same row or column must be independent, since otherwise there would be an active 2-clan on one side that does not appear on the other; domains on different lines can interact arbitrarily. Note, conversely, that this pattern of interaction assures that the equation is balanced. By restoring the original reference state, we derive the general third-order condition from one of the above form. Equations of order $n \geqslant 4$ are much more complicated.

Now when $P > 0$ all the balanced conditions can be derived from the second order (LM) by substitution and cancellation. But if there are constraints, cancellations may involve division by zero, so the higher-order conditions may go wrong. Consider, for example, the third-order equation in Fig. 4(b), obtained by taking the C_{ij} as single points in a 3×3 array. We assume that sites in the same row or column are unconnected, but all the other edges in this nine-point graph are present. Recall that a pair of states on one side of a Markov condition must differ on two (nonempty) independent sets. But any pair of the above six states differ on a *connected* set (an hexagonal graph for states on the same side of the equation; a square on opposite

sides). Hence if all the remaining states are in Ω^0, the Markov conditions will reduce to $0 = 0$, regardless of whether the six probabilities in the illustration are chosen to satisfy the cubic identity or not.

In fact, the domains on which the six states differ are all cycles, so they remain connected even if a diagonal of elements (forming a triangle) is deleted. Thus a six-point graph with Ω^+ states as shown in Fig. 4(c) gives an even simpler Markov system that cannot be expressed in terms of a limit of Gibbs potentials.

One might hope at least to set an upper bound to the order of conditions needed to assure that all the balanced equations hold. But it is easy to generalize the above example to a system that is well behaved at all orders $k < n$ but fails at arbitrarily large n. Simply take an $n \times n$ grid of sites, independent as before along rows and columns, but interacting everywhere else. Letting α_i be the ith row state and β_j the jth column, we have that $\prod_{i=1}^{n} P(\alpha_i) = \prod_{j=1}^{n} P(\beta_j)$ is balanced. Now consider any identity of order $k < n$ which has only row or column states on one side, say the left. Then exactly nk point clans appear on that side, so nk must appear on the right. Since, however, no 2-clans appear on the left, all the states on the right must be active on complete rows or columns as well (a domain with more than n sites would contain at least one connected pair). But then if a particular row α_i occurs on the left, it is also on the right: otherwise, in order that every point clan in α_i be included, all the β_j must occur on the right, contradicting $k < n$. By induction, the only lower-order equations containing just α_is and β_js on one side are trivial identities. So if the other states are assigned zero probabilities, all conditions of order less than n are satisfied, even if the nth-order one is arranged to fail.

4. SUGGESTIONS FOR FURTHER WORK

We have seen that there are graphs which require that an arbitrarily large number of balanced conditions be checked to verify that a Gibbs–Markov representation of the system exists. But for any particular graph there should be a finite set of relations from which all of (B) can be generated by multiplication and substitution (without cancellation). Is there a practical algorithm for deriving these generators from the structure G, Ω of the system? The two-stage construction of limit representations gives a procedure for checking balanced conditions, and linear programming techniques can be applied directly to the matrices ζ^+ and ζ^0 to ascertain whether there are clan potential solutions. But can we make use of counting arguments (like those on the form of $n = 1, 2, 3$ conditions) to organize a more economical or intelligible approach?

In particular, what special properties does the transformation ζ have by

virtue of being an incidence matrix for the relation $\prec$ of clan inclusion? We have already proved in Theorem 2 that ζ has a left inverse μ given by

$$\mu_{\kappa\omega} = \begin{cases} (-1)^{|\kappa|-|\omega|} & \text{if } \omega \prec \kappa \\ 0 & \text{otherwise} \end{cases}$$

Note that $\zeta\mu = \pi$ is a projection operator from arbitrary distributions onto the subspace of Gibbs potentials: Its components

$$\pi_{\omega\omega'} = \sum_{\kappa} \zeta_{\omega\kappa}\mu_{\kappa\omega} = (-1)^{|\omega'|} \sum_{\omega' \prec \kappa \prec \omega} (-1)^{|\kappa|} \qquad \text{for } \omega' \prec \omega$$

coincide with the exponents in the formula for decomposing $P(\omega)$ into its clan state probabilities. The matrix μ is obtained from the "zeta function" ζ by a process resembling Möbius inversion (see Hall,[6] Chapter 3). In fact, ζ is not a square matrix, and the relation $\prec$, though transitive and antisymmetric, fails as a partial ordering because it is not reflexive. But it *is* reflexive on the subset $\Lambda \subseteq \Omega$ of clan states which indexes the domain of ζ, so $\mu\zeta = 1$ *does* provide the usual recursive definition of a Möbius function:

$$\mu_{\kappa\kappa} = 1 \qquad \text{and} \qquad \mu_{\kappa\kappa'} = -\sum_{\kappa' \prec \prec \kappa'' \prec \kappa} \mu_{\kappa\kappa''} \qquad \text{for } \kappa' \prec \kappa$$

(where $\kappa' \ll \kappa''$ means that κ' is a strict substate of κ''). If the state space is restricted to $\Omega' \subseteq \Omega$, this inversion can still be carried out, provided that the domain index set is also cut down to $\Lambda' = \Lambda \cap \Omega'$. Are the resulting μ' and projection π' of any use in determining whether a restricted distribution can be extended to the whole state space?

Another approach is to formulate alternative types of random systems that have simpler behavior under constraints than Markov systems. Actually, the examples in Section 3 rely on some rather peculiar quirks of the Markov conditions. In particular, the reason the barrier property may fail is that, although $P'(\alpha) \equiv P(\alpha o)$ is Markovian on the subset A of Z, the distribution $P''(\alpha) = P(\alpha \cdot)$ need not satisfy (M). In other words, a subsystem of a Markov system need not be Markovian! From another point of view, we have insisted that $P(\alpha|\epsilon) = P(\alpha|\epsilon\gamma)$ whenever E is the environment of A, and C is independent; but if only *part* of E is known, then the likelihood of α *will* in general change with knowledge of states independent of A (perhaps because influences may creep in through undetermined parts of the environment).

These peculiarities are excluded in a *strongly Markovian* system, for which any subsystem is Markovian: that is, the distribution satisfies the conditions

$$P(\alpha\sigma \cdot \beta)P(\alpha'\sigma \cdot \beta') = P(\alpha\sigma \cdot \beta')P(\alpha'\sigma \cdot \beta) \qquad \text{(SM)}$$

whenever A and B are independent (the dot indicates summation on an

Fig. 5. Graphs for the simplest Gibbs systems that are not (a) strongly Markovian and (b) expressible in terms of pair potentials.

arbitrary part of Z). Constraints cause no difficulty in such systems, because the barrier property always holds. For, if $P(\omega) = 0$ and $\omega = \alpha\sigma\beta$, where A and B are independent, then (as in ordinary Markov systems)

$$P(\alpha\sigma\beta)P(\cdot\sigma\cdot) = P(\alpha\sigma\cdot)P(\cdot\sigma\beta)$$

implies that $P(\alpha\sigma\cdot) = 0$ or $P(\cdot\sigma\beta) = 0$. But now if the forbidden substates has independent parts, we can use (SM) to continue the reduction until we arrive at a clan $\kappa \prec \omega$ such that $P(\kappa\cdot) = 0$. Note also that there is no problem about dependence of barriers on the choice of null state: in a strongly Markovian context, (SM) implies that $P(\omega) = \prod_{\kappa \prec \omega} P(\kappa\cdot)^{n_{\omega\kappa}}$, so we may define $\Phi(\kappa\cdot) = -\log P(\kappa\cdot)$ for $P(\kappa\cdot) > 0$ and use the usual inversion formula for the clan potentials without any mention of a reference state at all.

Unfortunately, many Markov systems are not strongly Markovian. Consider, the instance, the three-point graph shown in Fig. 5(a). We define

$$\varphi(\kappa) = \begin{cases} -a & \text{for } \kappa = 12 \text{ active} \\ -b & \text{for } \kappa = 23 \text{ active} \\ 0 & \text{otherwise} \end{cases}$$

Then the strong condition

$$[(13) + (123)][(0) + (2)] = [(1) + (12)][(3) + (23)]$$

(where the numbers indicate active sites, and summation occurs at site 2) becomes $(1 + e^a e^b)(1 + 1) = (1 + e^a)(1 + e^b)$. This reduces to $(1 - e^a) \times (1 - e^b) = 0$; and it is easy to see in general that a given Gibbs potential on this graph yields a strongly Markovian system iff $\varphi(\kappa)$ vanishes on one of the pair clans 12 or 23. Is there a simple characterization of (SM) systems that holds on larger graphs as well?

Another special kind of system that may be of interest is determined by a *pair potential*—that is, a Gibbs potential which has no contributions from clans larger than two sites. The need for n-body interactions with $n > 2$ does not seem to arise in physics: known exchanges can be analyzed in terms of two-body potentials. Is there any Markovlike (extrinsic) property of the probability distribution that corresponds to this requirement? There does not seem to be any connection with the strongly Markovian systems: The previous example was given by a pair potential, but violates (SM); conversely, the four-site system in Fig. 5(b), where $\varphi(\kappa) = a$ on the triangle and zero

elsewhere, is strongly Markovian, but cannot be expressed in terms of pair potentials.

Finally, we consider how the notions of Gibbs–Markov systems may be extended to infinite or continuous domains. Define $\Phi^\omega(A)$ to be $\Phi(\alpha o)$, where α is the substate of ω on the domain A—that is, $\Phi^\omega(A)$ is the potential resulting from "turning on" ω on some part of the system A. Then the Markov conditions become

$$\Phi^\omega(A \cup B) = \Phi^\omega(A) + \Phi^\omega(B) - \Phi^\omega(A \cap B) \qquad (\Phi)$$

whenever A and B are *separable* in the sense that their nonoverlapping parts $A - B$ and $B - A$ are independent. If we ignore the latter condition, Φ^ω resembles a finitely additive set function. Thus if (Φ) held for all subsets of a finite set, $\Phi^\omega(A)$ would just be the sum of $\Phi^\omega(\{a\})$ for all $a \in A$ (assuming that $\Phi^\omega(\varnothing)$ is normalized to zero). In our case we may say that Φ^ω is additive "with respect to the relation of separability." This relation is monotonic in the sense that any pair of subsets A' and B' of separable sets A and B is separable as well. The essential content of Theorem 2 and the inversion formula is that a set function additive with respect to a monotonic relation R can be expressed as a (particular) linear combination of values on sets "irreducible" under R; the proofs depend on the discreteness of Z. But is it possible to find an analogous expression $\Phi = \zeta\varphi$ for Φ additive with respect to relations on more general (continuous) domains?

APPENDIX

First we consider two propositions about arbitrary real matrices; then we derive Lemmas 1 and 2 by introducing restrictions to integral values. For ease in writing, vectors are denoted by lower case letters; matrices, by capitals; matrix multiplication, by juxtaposition of the corresponding symbols, where it is understood that vectors appearing on the right of products are column vectors, while those appearing on the left are rows. In particular, the inner product $\sum_i w_i v_i$ is abbreviated wv.

Proposition 1. A vector v is in the range of M iff $wv = 0$ for every w such that $wM = 0$.

Proof. Define the *orthogonal complement* of a subspace X to be

$$X^\perp = \{y: \quad yx = 0 \quad \text{for every} \quad x \in X\}$$

The proposition follows from the familiar result that $X = X^{\perp\perp}$ in any finite-dimensional vector space. In particular, v is in the range or column space $R(M)$ iff it is perpendicular to every vector in the orthogonal complement $R(M)^\perp$. ∎

For our second result we need a different notion of complement, to take into account the positivity requirements on vectors. Suppose C is a closed convex cone in a finite vector space (see Hall,[6] Chapter 8). Then the *dual* C^* defined by

$$C^* = \{y: \quad yx \geqslant 0 \quad \text{for every} \quad x \in C\}$$

is itself a convex cone. Corresponding to the orthogonality results, we have $C = C^{**}$, and the following special case (due to Farkas): $y = xA$, where $x \geqslant 0$, iff $yw \geqslant 0$ for every w such that $Aw \geqslant 0$.

Now we are equipped to prove the following.

Proposition 2. There exists a vector w such that $Mw = 0$ and $Nw > 0$ iff there are no w^1 and $w^2 \geqslant 0$ such that $w^1M + w^2N = 0$. (Recall that $>$ means that all components are strictly positive; $\geqslant$ means that they are non-negative and not all zero.)

Proof. Let $\mathscr{R}$ be the collection of rows of N. We seek a w satisfying $Mw = 0$ and $rw > 0$ for all $r \in \mathscr{R}$. Suppose $\mathscr{R}'$ is a maximal (possibly empty) subset of $\mathscr{R}$ for which a corresponding w exists. We show that either $\mathscr{R}' = \mathscr{R}$, or else there are vectors w^1 and $w^2 \geqslant 0$ satisfying $w^1M + w^2N = 0$.

For suppose there is some $r^0 \in \mathscr{R} - \mathscr{R}'$. Define N' to be the matrix obtained from N by striking out all such r^0—that is, N' consists exactly of the rows in $\mathscr{R}'$. Then there exist vectors w^0 satisfying $Mw^0 = 0$ and $N'w^0 > 0$, but for all such w^0, $r^0w^0 \leqslant 0$. In fact, $r^0w \leqslant 0$ for all w such that $Mw = 0$ and $N'w \geqslant 0$ (even if some of the latter components vanish). For any such w can be expressed as $\lim_{\lambda \to 0^+}(w + \lambda w^0)$, where $N'(w + \lambda w^0) > 0$ for all $\lambda > 0$, so $r^0w = \lim_{\lambda \to 0^+} r^0(w + \lambda w^0) \leqslant 0$. Define

$$A = \begin{pmatrix} M \\ -M \\ N' \end{pmatrix}$$

Then $Aw \geqslant 0$ iff $Mw = 0$ and $N'w \geqslant 0$. Hence by Farkas' theorem $-r^0 = xA$ or $x^1M + (x^2N' + r^0) = 0$, where $x^2 \geqslant 0$. This identity provides the desired values for w^1 and w^2.

Conversely, suppose that w^1 and w^2 exist. Then $Mw = 0$ implies $w^1Mw + w^2Nw = w^2Nw = 0$, contradicting $Nw > 0$ and $w^2 \geqslant 0$. ∎

In order to introduce restriction to integer values, we use the following additional result.

Proposition 3. Let M be a finite rational matrix. If $wM = 0$, then $w = \sum_k \lambda_k n^k$, where the λ_k are real numbers and the n^k are integer vectors such that $n^kM = 0$.

Proof. Induct on the number of rows in M.

The result is obvious if M vanishes or consists of a single row. Hence we may assume that some element, say M_{11}, is nonzero, so that w_1 can be expressed as a rational combination of the other components of w: $w_1 = -(1/M_{11}) \sum_{i=2}^{n} w_i M_{i1}$. Then $wM = 0$ becomes $\sum_{i=2}^{n} w_i[M_{ij} - (M_{i1}M_{1j}/M_{11})] = 0$. Regard the expression in brackets as a matrix M' with one less row than M, and let w' be $(w_2, w_3, \ldots, w_n)$. By the inductive hypothesis, $w'M' = 0$ implies $w' = \sum_k \lambda_k n'^k$, where n'^k satisfies $n'^k M' = 0$. Now we define $n_1'^k = -(1/M_{11}) \sum_{i=2}^{n} n_i'^k M_{i1}$, and note that $\sum_{i=1}^{n} n_i'^k M_{i1} = 0$ and $w = \sum_k \lambda_k n'^k$. Since M is rational, all the $n_i'^k$ are rational, so we may multiply out the denominators (transfer them to the λ_k) to get the required integral n^k. ∎

Lemma 1. Let M be a finite matrix of integers. Then v is in the range of M iff $nv = 0$ for every integer vector n such that $nM = 0$.

Proof. In light of Proposition 1, we need only suppose that there is some real w such that $wM = 0$ but $wv \neq 0$. By Proposition 3, $w = \sum_k \lambda_k n^k$, where $n^k M = 0$, and at least one of these integer vectors must satisfy $n^k v \neq 0$.

Lemma 2. Let M and N be finite matrices of integers. There exists an integer vector n^0 such that $Mn^0 = 0$ and $Nn^0 > 0$ iff there are no integer vectors m and $n \geqslant 0$ such that $mM + nN = 0$.

Proof. We use Proposition 3 to show that the statements in this lemma and in Proposition 2 are equivalent when M and N are integral.

Suppose there is a real vector w such that $Mw = 0$ and $Nw > \delta > 0$. Then $w = \sum_k \lambda_k n^k$, where $Mn^k = 0$. If the λ_k are irrational, there are rational λ_k' such that $|\lambda_k - \lambda_k'|$ is arbitrarily small. Since all the n^k are bounded, there is therefore a rational $w' = \sum_k \lambda_k' n^k$ for which $Nw' > \delta/2 > 0$. The required n^0 is obtained by multiplying out denominators.

On the other hand, suppose there are real vectors w^1 and $w^2 \geqslant 0$ such that $w^1 M + w^2 N = 0$. We may assume that w^2 is in fact strictly positive, since the zero components and corresponding rows of N can simply be ignored. Again, $w^1 = \sum_k \lambda_k m^k$ and $w^2 = \sum_k \lambda_k n^k$, where $m^k M + n^k N = 0$. We choose rational λ_k' sufficiently close to λ_k—this time so that w'^2 remains strictly positive—and then multiply denominators to get the desired integral vectors.

ACKNOWLEDGMENTS

I am grateful to Dr. J. M. Hammersley for introducing me to this field, and providing the ideas upon which this paper is based. I would also like to thank Dr. D. J. A. Welsh for his encouragement and suggestions.

REFERENCES

1. J. M. Hammersley and P. E. Clifford, "Markov fields on finite graphs and lattices," unpublished (1971).

2. R. L. Dobrushin, "The description of a random field by means of its conditional probabilities, and conditions of its regularity," *Th. Prob. & Appl.* [English transl. of *Teoriia Veroiatn.*] **13**:197 (1968).
3. M. B. Averintsev, "On a method of describing complete parameter fields," *Problemy Peredaci Informatsii* **6**:100 (1970).
4. F. Spitzer, "Markov random fields and Gibbs ensembles," *Am. Math. Month.* **78**:142 (1971).
5. G. R. Grimmett, "A theorem about random fields," *Bull. London Math. Soc.* **5**(13):81 (1973).
6. M. Hall, *Combinatorial Theory*, Blaisdell (1967).