Quadratic Dynamical Systems (Preliminary Version)[†]

Yuri Rabinovich[‡] Alistair Sinclair[§] Avi Wigderson[¶]

Abstract

The main purpose of this paper is to promote the study of computational aspects, primarily the convergence rate, of nonlinear dynamical systems from a combinatorial perspective.

We identify the class of symmetric quadratic systems. Such systems have been widely used to model phenomena in the natural sciences, and also provide an appropriate framework for the study of genetic algorithms in combinatorial optimisation. We prove several fundamental general properties of these systems, including a characterisation of the set of fixed points to which the system converges.

We go on to give a detailed analysis of a quadratic system defined in a natural way on probability distributions over the set of matchings in a graph. In particular, we prove that convergence to the limit requires only polynomial time when the graph is a tree. This result demonstrates that such systems, though nonlinear, are amenable to quantitative analysis.

1 Introduction

1.1 Dynamical systems

Many natural phenomena can be described by dynamical systems in Euclidean space. In such a description there is a fixed set of types, one for each dimension, and a point in the system (often called a *state* or a *population*) specifies how many elements of each type exist at a given time instant. The system evolves under a fixed map f, which attempts to capture the underlying phenomenon. We mention below a few examples; for a detailed treatment see, e.g., [7].

In Physics, the system might describe the behaviour of gas molecules in a container. Here the types correspond to, say, velocity values, and the state p specifies how many molecules of each type there are at a certain point in time. The map f includes the Newtonian laws so as to produce from each state p a new state $f(p)$ in the next time step, under some assumption about the spatial distribution of the molecules [11]. (One can also easily define a continuous-time analogue of this system). This view of dynamics is quite general and can in principle be used to describe a wide range of physical phenomena.

In Biology, the types may be the genotypes of some species (specified by the values of their genes). A population p is then simply the number of individuals of each type. The map f determines the population in the next generation according to a fixed set of rules that includes the genetic outcome of mating, the survival capacity of different types, random mutations etc. [2]. A completely different biological example describes the ecological coexistence of different species. Here the types are species, p specifies the number of creatures of each type, and f describes the outcome of their interaction: for example, one species may be the food supply of the other, or compete with it for the same resource [7].

The basic object of study in any such system is the *trajectory* of a point p in the state space, i.e., the sequence of points $p, f(p), f(f(p)), \ldots$. There are many questions one can ask about these objects, each of which has given rise to a major research area: for example, there are geometric and topological studies of the structure of the limit points or sets (attractors) of trajectories and of the initial points converging to them; ergodic and measure-theoretic questions on the recurrence of visits of trajectories to subsets of the

---

[†]This is a slightly revised version of a paper that appeared in *Proceedings of the 33rd IEEE Symposium on Foundations of Computer Science* (1992), pp. 304–313.

[‡]Department of Computer Science, Hebrew University, Jerusalem, Israel 91904. E-mail: yuri@cs.huji.ac.il

[§]Department of Computer Science, University of Edinburgh, Scotland. Current address: International Computer Science Institute, 1947 Center Street, Berkeley CA 94704. E-mail: sinclair@icsi.berkeley.edu

[¶]Hebrew University, Jerusalem and Princeton University. E-mail: avi@cs.huji.ac.il. Supported in part by NSF Grant CCR–8920505

state space; "chaos"-type questions on the sensitivity of the trajectory and its limit to small perturbations in the initial point; and computational questions regarding the time it takes for transient behaviour to die out and the system to reach its limiting behaviour. This last type of question is of central importance both to understanding the phenomenon captured by the dynamical system and in computer simulations of the system, which are often performed as a means of realising complex limiting behaviour. Our main interest is in the last type of question, though when studying it one often needs to tackle the other questions as well.

The simplest dynamical systems arise from linear maps f. The ubiquitous example of a linear system is a Markov chain, viewed as acting on the simplex of probability distributions over types. (When f preserves the "size" of the population, it is often more convenient to take the state space to be just these probability distributions rather than all of Euclidean space; we shall adopt this view throughout.) It is well known that, if f is irreducible and aperiodic, there is a unique limit point (or stationary distribution) which attracts all initial points. Moreover, there is a beautiful spectral theory that explains how the second largest eigenvalue of the matrix representing f controls the rate of convergence of any trajectory to the limit. Although originally motivated by the study of natural phenomena, this theory has proved extremely useful in many mathematical disciplines. In particular, there has been a recent upsurge of interest in algorithmic applications of artificial Markov chains in Computer Science, spurred by the development of new tools for analysing the rate of convergence. Examples include probabilistic algorithms for approximate counting [6, 10], pseudo-random number generators [1] and "simulated annealing" type algorithms for combinatorial optimisation [9, 10, 15].

For nonlinear systems, in stark contrast, there are almost no global convergence rate bounds available (but see [3]). We now identify a class of nonlinear systems known as symmetric quadratic systems, and argue that they capture several natural and artificial phenomena for which obtaining convergence rate information is computationally important. Then we describe our results, which demonstrate that these systems are sometimes amenable to quantitative analysis.

1.2 Symmetric quadratic operators

We illustrate this class of systems by its oldest example, namely the (space homogeneous) kinetic gas model of Maxwell. Gas molecules interact by means of elastic collisions that change the distribution of their velocities. Thus the types are velocity vectors in $\mathbb{R}^3$, and when two types i, j collide they disappear and are replaced by a *random* pair k, l which preserves momentum and energy. (In classical mechanics the resulting pair would be uniquely determined, but statistical mechanics makes the outcome random by postulating a uniform distribution on all possible relative positions of the balls at collision.) Note that in this system the event that the pair i, j produces k, l has the same probability as its time reversal, in which k, l produces i, j. We call systems having this property *symmetric*. Moreover, it is clear that the map f defining this system is *quadratic*, i.e. $f(p)$ is a quadratic function of p, since outcomes depend only on pairwise collisions.

Many other physical theories assume pairwise interactions and hence define quadratic systems. Moreover, the outcomes of local interactions in physical systems are often defined in a symmetric fashion. In genetics, the feature of pairwise interaction (of parents) to produce the next generation is almost universal. Here symmetry is less common, but exists in simple genetic models where, for example, a child randomly picks each gene from either the father's or the mother's DNA (as can be seen by introducing a "complementary" child who makes the opposite choices).

The importance of understanding the evolution of natural systems of this kind, and in particular the computational issue of the rate at which they approach limiting behaviour, is well known. Further motivation for the study of these questions is provided by a family of artificial systems called genetic algorithms [8], which have recently gained acceptance as a general heuristic for solving combinatorial optimisation problems. The Appendix explains why similar questions about limit points and convergence rate arise when trying to understand these heuristics and estimate their efficiency.

1.3 Summary of results

We consider only discrete-time, finite-dimensional systems. For any symmetric quadratic system of this form, we prove that every trajectory that does not approach the boundary of the simplex converges to a fixed point; this is enough to imply global convergence of many naturally occurring systems. This is a strong property not shared by general, asymmetric systems. We go on to give a characterisation of all fixed points, and describe all the linear invariants of the system. Our main analytical tool is the fact that entropy is strictly increasing along any non-trivial trajectory. This is a discrete analogue of Boltzmann's

famous H-Theorem, which he proved when studying the kinetic gas model described above. Although such results, and sometimes much more, are known for certain special cases studied in Physics, our setting is more general and focuses on the combinatorial structure of the system.

Next we define a natural quadratic system on probability distributions over the matchings in a graph, motivated both by its inherent combinatorial interest and its application to algorithms for generating matchings at random from certain probability distributions. We give a complete analysis of this system when the underlying graph is a tree. The analysis is in two parts. The first part gives an explicit closed-form formula for the limit distribution in terms of simple parameters of the initial distribution. The second, and more significant part is a proof that the system always converges to its limit in time that depends only polynomially on the size of the tree. This is a strong result in view of the fact that the number of matchings in the tree is in general exponentially large. Moreover, it appears to be the first bound of its kind for a complex nonlinear system defined on combinatorial structures.

2 Definitions and basic properties

In this section we formally define the dynamical systems of interest to us and establish some fundamental general results about their behaviour.

2.1 Symmetric quadratic operators

For convenience we shall adopt throughout terminology appropriate to the genetic applications mentioned in the Introduction, though the systems we describe have more general applicability. We consider a quadratic operator, which we shall refer to as a "mating" operator, of the following form. Let $\mathcal{N}$ be a finite set of "types", with $|\mathcal{N}| = N$. For each quadruple $(i,j,k,l) \in \mathcal{N}^4$, let β_{ijkl} denote the probability that, in a mating between the two "parent" types i and j, the two types k and l are produced as "offspring". Since any mating must have a definite outcome, we always have $\sum_{k,l \in \mathcal{N}} \beta_{ijkl} = 1$ for all $i,j \in \mathcal{N}$.

We will adopt the view that the order of parents and offspring is not significant, so that β is symmetric in i and j, and in k and l, i.e., $\beta_{ijkl} = \beta_{jikl} = \beta_{ijlk}$. We also impose the further symmetry condition that $\beta_{ijkl} = \beta_{klij}$; this means that the mating operation is locally reversible. If β satisfies all the above symmetry conditions, we call it *symmetric*. As indicated in the Introduction, the operators encountered in many applications of interest are symmetric so this does not constitute a severe restriction. Moreover, as we shall see presently, symmetric operators have much useful structure which we will be able to exploit. Finally, we call β *aperiodic* if $\beta_{ijij} > 0$ for all $i,j \in \mathcal{N}$. Aperiodicity is a technical requirement, again satisfied in most applications, that simplifies the statement of our results.

Now let $p = (p_i)$ be a probability distribution on $\mathcal{N}$, which we shall often refer to as a "population." Given a symmetric β as above, we define the population $p \times p$ by

$$(p \times p)_l = \sum_{i,j,k \in \mathcal{N}} p_i p_j \beta_{ijkl} \qquad \forall l \in \mathcal{N}. \qquad (1)$$

(In the notation of the Introduction, the map f is defined by $f(p) = p \times p$.) It should be clear that these equations describe a mating process in which two parents are selected independently at random from the population p and mated to form random offspring according to the distribution specified by β. The operator $\times$ (or, equivalently, β) defines a (deterministic) quadratic mapping from the simplex of probability distributions over $\mathcal{N}$ to itself.

By repeated application of this operator, we can define a time-dependent process (or a dynamical system) in the obvious way: denote by $p(0)$ the population at time 0, and for all $t \in \mathbb{N}$ define $p(t+1) = p(t) \times p(t)$. We shall refer to the sequence of points $p(0), p(1), p(2), \ldots$ as the *trajectory* of the point $p(0)$.

Remark: We are focusing on systems described by *pairwise* interactions, as these capture the majority of natural examples. However, it should be clear that our results can be generalised in a straightforward manner to the case of m-ary interactions (in which m parents produce m offspring) for any fixed m, provided the notion of symmetry is suitably extended. □

2.2 Attractors (limit points): existence and structure

As explained earlier, it is our main aim here to study the evolution of $p(t)$ with time. The obvious first question to ask is whether $p(t)$ tends to some limit point π as $t \to \infty$. Clearly, since β is continuous any such population π must be a fixed point, i.e., $\pi \times \pi = \pi$; we call such a population *stationary*. In order to investigate this and other properties of our systems, we study the *entropy* of the population. For a probability distribution p on any set S, define the

entropy

$$H(p) = -\sum_{i \in S} p_i \lg p_i.$$

Theorem 1 *For a symmetric, aperiodic operator β, the entropy of any non-stationary population is strictly increased by mating; i.e., if p is not stationary then $H(p \times p) > H(p)$.*

Proof: It is helpful to work with probability distributions on the set $\mathcal{N}^2$ of ordered pairs of types. For any population p on $\mathcal{N}$, define the probability distribution $\widetilde{p}$ on $\mathcal{N}^2$ by $\widetilde{p}_{ij} = p_i p_j$. The relation between the entropies of p and $\widetilde{p}$ is easily seen to be

$$H(\widetilde{p}) = 2H(p). \tag{2}$$

We observe first that a single application of the quadratic operator β can be decomposed into two stages: a linear operation on the pair distribution $\widetilde{p}$, followed by a recomputation of the singleton probabilities. The first stage is defined by a linear operator B as follows:

$$(B\widetilde{p})_{kl} = \sum_{ij \in \mathcal{N}^2} \beta_{ijkl} \widetilde{p}_{ij} \qquad \forall\, kl \in \mathcal{N}^2.$$

In the second stage, we compute $(p \times p)_l$ for each $l \in \mathcal{N}$ by

$$(p \times p)_l = \sum_{k \in \mathcal{N}} (B\widetilde{p})_{kl}.$$

It is immediate from the definition (1) of the operator $\times$ that these two stages do indeed correctly compute $p \times p$. We now proceed to show that, if p is not stationary, each of the two stages causes the entropy of the pair distribution to increase; by (2) this will imply the theorem. I.e., we will show that

$$2H(p) = H(\widetilde{p}) < H(B\widetilde{p}) \leq H(\widetilde{p \times p}) = 2H(p \times p).$$

Consider the first stage. Note that $B = (\beta_{ij,kl})$ is the transition matrix of a Markov chain on $\mathcal{N}^2$, since $\sum_{k,l \in \mathcal{N}} \beta_{ijkl} = 1$ for all $i, j \in \mathcal{N}$. By the assumption on the symmetry and aperiodicity of β, B is symmetric and aperiodic. Now partition the state space $\mathcal{N}^2$ into classes of communicating states, i.e., equivalence classes under the equivalence relation $\sim$ defined by $ij \sim kl$ iff $B^n_{ij,kl} > 0$ for some natural number n, where B^n is the nth power of B. Each of these classes behaves under B as an independent Markov chain, which is symmetric and aperiodic and thus has the uniform distribution as its unique stationary distribution. It is well known (see, e.g., [14, Section 5.6]) that such a Markov chain causes the entropy of any non-uniform probability distribution to increase strictly on every step. This in turn implies that the entropy of the entire distribution over $\mathcal{N}^2$ increases under B, i.e., $H(B\widetilde{p}) > H(\widetilde{p})$, unless $\widetilde{p}$ is uniform on all the classes.

We turn now to the second stage. The effect of this stage on the pair distribution is to transform it from $B\widetilde{p}$ into $\widetilde{p \times p}$. Note that this operation preserves the new marginals $q_l = \sum_k (B\widetilde{p})_{kl}$ but makes the components k, l independent, i.e., $(\widetilde{p \times p})_{kl} = q_k q_l$. It is a routine matter to check that this latter distribution has maximum entropy among all distributions on $\mathcal{N}^2$ with the given marginals q_l, so $H(\widetilde{p \times p}) \geq H(B\widetilde{p})$ as claimed.

Finally, we have seen that the increase in entropy in the first stage is strict unless $\widetilde{p}$ is uniform on each communicating class. But in this case we have $B\widetilde{p} = \widetilde{p}$, which implies by the definition of the second stage that $p \times p = p$, so p is stationary. Thus we get a strict increase in entropy unless p is stationary. □

Remark: As mentioned in the Introduction, Theorem 1 is a discrete version of a classical result in statistical mechanics known as Boltzmann's H-Theorem, which asserts that the entropy of an ideal gas increases monotonically with time (see, e.g., [11]). Here the interaction probabilities β_{ijkl} have values that describe the collisions of pairs of gas molecules. The H-Theorem is generally proved for continuous time systems and assumed to hold for discrete time without rigorous justification. Theorem 1 gives a rigorous proof in the discrete-time setting and makes clear the essential combinatorial properties required of β. □

Theorem 1 is a powerful analytical tool. We use it first to prove the following fundamental result, which states that every trajectory that does not approach the boundary of the simplex converges to some fixed point (stationary population). To make this statement precise, for distributions p, q on a finite set S define the variation distance

$$\|p - q\| = \tfrac{1}{2} \sum_{i \in S} |p_i - q_i| = \max_{A \subseteq S} |p(A) - q(A)|.$$

We will write $p(t) \to \pi$ to denote the fact that $\|p(t) - \pi\| \to 0$ as $t \to \infty$. We say that a trajectory $p(0), p(1), \ldots, p(t), \ldots$ *does not approach the boundary* if there exists some $\delta > 0$ (which may depend on $p(0)$) such that $p_i(t) \geq \delta$ for all $i \in \mathcal{N}$ and all sufficiently large t.

Theorem 2 *For any symmetric, aperiodic operator β and any initial population $p(0)$ such that the trajectory*

from $p(0)$ does not approach the boundary, there exists a stationary population π such that $p(t) \to \pi$.

Proof: Consider the trajectory $p(0), \ldots, p(t), \ldots$, and define the function $h(t) = H(p(t))$. By Theorem 1, $h(t)$ increases monotonically with t, and since it is evidently bounded above (by $\lg N$) it must tend to some finite limit, h say, as $t \to \infty$. Note that this is not sufficient to ensure that $p(t)$ itself converges. However, since the simplex is a compact set, there exists a convergent subsequence $\{p(t_j) : j = 0, 1, \ldots\}$ with some limit π. By our assumption that the trajectory does not approach the boundary, π must have full support (i.e., $\pi_i \geq \delta > 0$ for all $i \in \mathcal{N}$). Moreover, it is easy to verify using continuity of the functions H and $\times$ that π must be stationary and that $H(\pi) = h$.
We will show that $p(t) \to \pi$. To do this, we consider the *information divergence*, defined for distributions p, q on $\mathcal{N}$ by $D(p||q) = \sum_{i \in \mathcal{N}} p_i \lg(p_i/q_i)$; note that $D(p||q) \geq 0$, with equality iff $p = q$. By the well-known relationship (see, e.g., [4, page 58]) between variation distance and information divergence, it suffices to show that $D(p(t)||\pi)$ tends to zero as $t \to \infty$. Now for any $t \in \mathbb{N}$ we may write

$$D(p(t)||\pi) = \sum_{i \in \mathcal{N}} p_i(t) \lg p_i(t) - \sum_{i \in \mathcal{N}} (\lg \pi_i) p_i(t). \quad (3)$$

The first term in (3) is just $-h(t)$, while the second term is constant in time: this can be seen by noting that it is just the invariant function $\iota_\pi(p) \equiv \sum_{i \in \mathcal{N}} (\lg \pi_i) p_i$, defined in Theorem 4, evaluated at the point $p(t)$. But since π is the limit of the subsequence $p(t_j)$, by continuity the value of the invariant $\iota_\pi(p(t))$ must be $\iota_\pi(\pi) = -H(\pi) = -h$. Thus (3) yields

$$D(p(t)||\pi) = -h(t) + h \to 0,$$

as required. □

Remarks: (a) Theorem 2 is a very special property that depends crucially on the symmetry of the operator β. It is well known that asymmetric quadratic systems need not converge to a point but may exhibit more complex behaviour such as cycling [7].
(b) The above proof actually tells us rather more: it suffices to make the weaker assumption that some convergent subsequence has a limit with full support. Thus a trajectory can only fail to converge to a limit point if it approaches a connected set of equi-entropic stationary points on the boundary of the simplex. The analysis of trajectories close to the boundary is rather subtle and will be discussed in the full version of the paper.
(c) For many naturally occurring quadratic systems, it is possible to show that the trajectory from every initial population $p(0)$ with full support satisfies the condition in the above theorem, and hence converges to a limit point (with full support). This holds in particular for the quadratic system on matchings in graphs which is the subject of Sections 3 and 4 of this paper. □

A further corollary of Theorem 1 is the following elegant characterisation of the stationary populations.

Theorem 3 *For a symmetric, aperiodic operator β, a population p is stationary iff it satisfies*

$$p_i p_j = p_k p_l \qquad \forall ijkl \text{ with } \beta_{ijkl} > 0. \quad (4)$$

Proof: Recall from the proof of Theorem 1 that (4) is precisely the condition that the pair distribution $\widetilde{p}_{ij} = p_i p_j$ is a stationary distribution for the Markov chain B defined in the first stage. But this condition is in turn equivalent to stationarity of p for the quadratic process, since p itself changes if and only if $\widetilde{p}$ changes in the first stage; to see this, note that the second stage has no effect if $\widetilde{p}$ is unchanged in the first stage, and that it cannot undo any change in the first stage since it causes entropy to increase. □

Note that the characterisation of Theorem 3 is purely structural, i.e., it depends only on which of the β_{ijkl} are non-zero and not on their numerical values.

Theorem 3 confirms that in general there will be a continuum of stationary populations, corresponding to the set of solutions of the system of equations (4). Each stationary population will have a certain domain of attraction in the simplex. The question then arises of how to determine, for a given initial population satisfying the condition of Theorem 2, the particular stationary population which attracts it. The answer is to look for *invariants* of the operator β, i.e., functions ι on populations that satisfy $\iota(p \times p) = \iota(p)$ for all p.

The following family of linear invariants plays a crucial role in our systems. Recall that a population p has *full support* if $p_i > 0$ for all $i \in \mathcal{N}$.

Theorem 4 *Let π be any stationary population with full support. Then the function $\iota_\pi(p) = \sum_{l \in \mathcal{N}} (\lg \pi_l)\, p_l$ is an invariant.*

Proof: For each $l \in \mathcal{N}$ define the function $\Delta p_l(t) = p_l(t+1) - p_l(t)$. We may expand Δp_l as

$$\Delta p_l(t) = \sum_{ijk} \big(p_i(t) p_j(t) - p_k(t) p_l(t)\big) \beta_{ijkl}, \quad (5)$$

where we have used (1) to rewrite $p_l(t+1)$ and symmetry of β to rewrite $p_l(t)$.
Now let $\alpha_l = \lg \pi_l$, and note from Theorem 3, since π is stationary, that

$$\alpha_i + \alpha_j = \alpha_k + \alpha_l \qquad \forall ijkl \text{ with } \beta_{ijkl} > 0. \quad (6)$$

Now using (5) we may write

$$\iota_\pi(p \times p) - \iota_\pi(p) = \sum_{ijkl} \alpha_l (p_i p_j - p_k p_l) \beta_{ijkl}$$
$$= \tfrac{1}{4} \sum_{ijkl} (\alpha_k + \alpha_l - \alpha_i - \alpha_j)(p_i p_j - p_k p_l) \beta_{ijkl},$$

where we have used the symmetry properties of β. But (6) implies that every term in this sum is zero, so ι_π is indeed invariant. □

Remarks: (a) The invariants ι_π defined above are, up to scalings, the only linear invariants of the system, in the sense that any invariant of the form $\sum_{l \in \mathcal{N}} \alpha_l p_l$ defines a family of stationary populations (with full support) of the form $\pi_i = c^{\alpha_i}/Z$, where c is a positive constant and Z a normalising factor. We omit the proof.
(b) Equation (5) above indicates how to define an analogous system in continuous time: simply replace $\Delta p_l(t)$ on the left-hand side by the derivative dp_l/dt. This makes little essential difference to our results. □

The significance of the above invariants is expressed in our next theorem, which states that they completely determine the stationary population with full support to which a given initial population is attracted (assuming such a limit point exists).

Theorem 5 *Suppose $p(t) \to p^*$ and $q(t) \to q^*$, where p^* and q^* have full support. If $\iota_\pi(p(0)) = \iota_\pi(q(0))$ for all invariants ι_π as above, then $p^* = q^*$.*

Proof: We know in particular that $\iota_{p^*}(p(0)) = \iota_{p^*}(q(0))$, and hence by continuity of the operator $\times$ that $\iota_{p^*}(p^*) = \iota_{p^*}(q^*)$. Similarly, $\iota_{q^*}(p^*) = \iota_{q^*}(q^*)$. Writing out these two equalities yields

$$\sum\nolimits_l (\lg p_l^*) p_l^* = \sum\nolimits_l (\lg p_l^*) q_l^*$$
$$\sum\nolimits_l (\lg q_l^*) p_l^* = \sum\nolimits_l (\lg q_l^*) q_l^*$$

which can be subtracted to give

$$\sum\nolimits_l (\lg p_l^*/q_l^*) p_l^* = \sum\nolimits_l (\lg p_l^*/q_l^*) q_l^*.$$

But by Jensen's inequality applied to the convex function $x \lg x$, the left-hand side of this equation is always positive and the right-hand side always negative. Hence both sides must in fact equal zero, which happens iff $p^* = q^*$. □

Theorem 5 gives us a finite system of equations whose *unique* solution is the limit point of the trajectory of $p(0)$. The equations consist of (4) together with a finite basis for the set of linear equations $\iota_\pi(p) = c_\pi$, where the values c_π are determined by $p(0)$, i.e., $c_\pi = \iota_\pi(p(0))$.

More can be said about the general systems described above. However, having established some key facts, we leave further development of the general setting to the full paper [12] and instead show how the above ideas may be applied to a specific example.

3 Matchings in graphs

We define below a symmetric quadratic operator that acts on probability distributions over matchings in graphs. This operator is a natural one to choose as part of an algorithm for generating a random matching in a graph from a certain probability distribution. Applications of this include approximate counting of matchings and computation of the partition function of monomer-dimer systems [10], as well as a genetic algorithm for finding a large matching. This latter application is discussed in the Appendix.

Let $G = (V, E)$ be an undirected graph and $\mathcal{M}(G)$ the set of matchings in G. Given matchings $i, j \in \mathcal{M}(G)$, construct a random pair of offspring k and l according to the following scheme. Let $G'(i,j) = (V, E')$ be the subgraph of G whose edge set E' is the symmetric difference of the edge sets of i and j. Thus each connected component of $G'(i,j)$ is a path or a cycle whose edges belong alternately to i and to j. For each such component independently, flip a fair coin; if the coin comes up heads, make k agree with i and l agree with j on the component; if it comes up tails, reverse the roles of i and j. Also, include in both k and l all edges common to both i and j, and no other edges. (Thus the unions of the edge sets of i and j and of k and l, viewed as multisets, are equal.) It is a simple matter to check that the above scheme specifies a symmetric, aperiodic operator as defined in the previous section. We shall refer to this operator as β^{match}.

First let us consider the stationary populations under β^{match}. Using Theorem 3 we can obtain a simple

description of these. We call a population p on $\mathcal{M}(G)$ *normal* if it is of the form

$$p_i = \frac{1}{Z} \prod_{e \in i} \lambda_e,$$

where $\lambda_e \in \mathbb{R}^+$ for each edge e in G, and $Z = \sum_{i \in \mathcal{M}(G)} \prod_{e \in i} \lambda_e$ is a normalising factor.

Theorem 6 *Under the operator β^{match}, a population with full support is stationary iff it is normal.*

Proof: It is easy to check that any normal population is stationary. For the other direction, let p be a stationary population on $\mathcal{M}(G)$, and for each edge e of G define $\lambda_e = p_{\{e\}}/p_0$, where 0 denotes the empty matching and $\{e\}$ denotes the matching consisting of the single edge e. Since we are assuming that p has full support, each λ_e is positive and finite. Now we claim that, for each matching i, $p_i = p_0 \prod_{e \in i} \lambda_e$; note that this is enough to establish the theorem. We proceed by induction on the size of i. The claim is trivially true when the size of i is 0 or 1. Suppose $i = j \cup \{f\}$, where the edge f does not belong to matching j. Then it is easy to check that $\beta^{\text{match}}_{i0j\{f\}} > 0$, so Theorem 3 implies that $p_i p_0 = p_j p_{\{f\}} = \lambda_f p_0 p_j$. The claim now follows by applying the inductive hypothesis to j. This concludes the proof of the claim and the theorem. □

Now consider an arbitrary initial population $p(0)$ over $\mathcal{M}(G)$. We shall say that $p(0)$ *generates* $\mathcal{M}(G)$ if some point $p(t)$ on the trajectory from $p(0)$ has full support (i.e., if the trajectory from $p(0)$ enters the interior of the simplex). It is not hard to see that a sufficient condition for this is that $p(0)$ assigns positive weight to the empty matching in G and, for each edge e of G, to some matching containing e. Now it can be shown that no trajectory of the above form can approach the boundary of the simplex; the proof, which is left to the full paper, involves exhibiting, for each boundary facet of the simplex that contains a stationary point, an invariant that vanishes on the facet but is strictly positive on every interior point. Theorem 2 then ensures that every such trajectory converges to a stationary population with full support. Thus we see from Theorem 6 that, if $p(0)$ is any initial population that generates $\mathcal{M}(G)$, repeated application of β^{match} to $p(0)$ results asymptotically in a normal population.

Remark: With rather more work, it is possible to derive necessary and sufficient conditions for $p(0)$ to generate $\mathcal{M}(G)$. This question turns out to be intimately related to the structure of the matching polytope of G; we defer a detailed discussion, which is of independent interest, to the full paper [12]. □

We turn now to the problem of determining the normal population to which the trajectory of a given initial population $p(0)$ converges. Following the approach of the previous section, we identify a family of invariants of β^{match}. Let e be any edge of G, and define $q_e = \sum_{i \ni e} p_i$, the probability that a matching in population p contains e. It is easy to check that, for every $e \in E$, q_e is an invariant of β^{match}. This gives us a family of polynomial equations satisfied by the parameters λ_e of our normal population. Define the matching polynomial of a graph $G = (V, E)$ over variables $\overline{x} = (x_e)_{e \in E}$ by $g(G, \overline{x}) = \sum_{m \in \mathcal{M}(G)} \prod_{e \in m} x_e$. Then we may write the system of equations as follows:

$$q_e = \frac{\lambda_e g(G/e, \overline{\lambda})}{g(G, \overline{\lambda})} \qquad \forall e \in E, \tag{7}$$

where G/e denotes the graph obtained from G by removing the endpoints of e, and the values q_e are determined by the initial population $p(0)$.

It is not hard to show that the invariants $\{q_e\}_{e \in E}$ form a basis for the family of linear invariants ι_π defined in Theorem 4. Thus, by Theorem 5, they uniquely determine the stationary population. We summarise this discussion as follows.

Theorem 7 *Let $p(0)$ be any initial population that generates $\mathcal{M}(G)$. Then the equations (7) have a unique solution $\{\lambda_e\}_{e \in E}$ which determines the normal population corresponding to $p(0)$.* □

For general graphs, the solution to these equations is not a rational function of the quantities q_e. However, as we shall see in the next section, in the case that the underlying graph G is a tree the equations do have a simple closed-form solution. More significantly, we shall also show in this case that the quadratic system defined by β^{match} converges to a normal population very rapidly, in time bounded by a low-degree polynomial in the size of the tree.

4 Matchings in trees

In this section, we present a complete analysis of the mating operator β^{match} in the case where the underlying graph G is a tree T.

4.1 The stationary population

First, we give an explicit formula for the stationary population corresponding to any initial population which generates $\mathcal{M}(T)$. This formula will be in terms of the invariants q_e and r_v, where for each vertex $v \in V$ we define $r_v = 1 - \sum_{e=(v,u)} q_e$, the probability that v is *not* covered by a matching.

Theorem 8 *Let T be a tree and $p(0)$ an initial population that generates $\mathcal{M}(T)$. Then $p(t)$ converges to the normal population π defined by*

$$\lambda_e = \frac{q_e(1-q_e)}{r_v r_u},$$

for each edge $e = (v,u)$ of T, where the values of q_e and r_v are determined by $p(0)$. The population π satisfies

$$\pi_i = \frac{\prod_{e \in i} q_e \prod_{v \notin i} r_v}{\prod_{e \notin i}(1-q_e)} \qquad \forall i \in \mathcal{M}(T).$$

(The notation $v \notin i$ indicates that vertex v is not covered by the matching i.) □

Theorem 8 is readily verified by considering the structure of the matching polynomial; the details are left for the full paper.

4.2 Rate of convergence

In order to analyse the rate of convergence of β^{match} on $\mathcal{M}(T)$, we split the process conceptually into a sequence of stages. Each stage is associated with a partition of $\mathcal{M}(T)$ into equivalence classes, with the partition becoming successively finer. During each stage, we analyse the time for the probability distribution over its classes to converge. The crucial observation is that this process on equivalence classes can be viewed as a *linear* system, which can then be tackled using established techniques plus some additional insight.

Let v be a leaf of T, and $e = (u,v)$ the edge adjacent to v. Let T' be the subtree obtained by removing e and v from T. Now partition the set $\mathcal{M}(T)$ into equivalence classes by placing two matchings in the same class iff they agree on T'; thus each class contains at most two matchings. Note that there is a 1-1 correspondence between equivalence classes and matchings in T'. To fix notation, let $[i]$ denote the equivalence class of matching $i \in \mathcal{M}(T')$, and if i contains two elements denote them i^+ and i^- according to whether or not they contain e. Also, for any population p on $\mathcal{M}(T)$ define the population p' on $\mathcal{M}(T')$ by $p'_i = \sum_{j \in [i]} p_j$.

Our first claim is that the operator β^{match} acting on the equivalence classes of $\mathcal{M}(T)$ induces a similar mating process on $\mathcal{M}(T')$.

Lemma 9 *For any population p on $\mathcal{M}(T)$, we have $(p \times p)' = p' \times p'$.*

Proof (sketch): Since the mating operator is bilinear, and the operation $p \mapsto p'$ is linear, it suffices to show that, for any two matchings $m_1, m_2 \in \mathcal{M}(T)$, if p_1 and p_2 are populations concentrated at m_1, m_2 respectively then $(p_1 \times p_2)' = p'_1 \times p'_2$, where we have extended the definition of $\times$ in the obvious way to allow parents to be taken from different populations. This fact is straightforward to check by a simple case analysis depending on whether one, both or neither of m_1, m_2 contains e. □

In the light of Lemma 9, we can understand the evolution of the mating process on $\mathcal{M}(T)$ inductively as follows. First we view the process as working on equivalence classes: by Lemma 9, this is equivalent to a mating process on matchings in the smaller tree T', so we may argue inductively about its behaviour. Then, assuming that the process on equivalence classes has reached stationarity, we may reason about the effects of mating *within* the classes. As we shall see in a moment, this latter process is much more tractable than the original quadratic system.

Let us call a population p on $\mathcal{M}(T)$ *pseudo-stationary* if the induced population p' on $\mathcal{M}(T')$ is stationary. Our next observation is that, if we start with a pseudo-stationary population, the mating operator on $\mathcal{M}(T)$ may be viewed as a linear system. In fact, it defines a Markov chain on the space $\mathcal{M}(T)^+$ of those matchings in T that include the edge e.

Lemma 10 *Let p be any pseudo-stationary population on $\mathcal{M}(T)$. Then the probability distribution on matchings in $\mathcal{M}(T)^+$ behaves under β^{match} as a time-homogeneous Markov chain with transition probabilities*

$$\alpha_{i^+ l^+} = \sum_{jk} p'_j \beta'_{ijkl} \qquad \forall i^+, l^+ \in \mathcal{M}(T)^+ \tag{8}$$

where β'_{ijkl} defines the mating operator in T'.

Proof (sketch): Let $l^+ \in \mathcal{M}(T)$. By definition of the system (and ignoring the possibility that some matching appearing in the sum may not be defined), we may write $(p \times p)_l$ as

$$\sum_{[i][j][k]} \big(p_{i^+} p_{j^+} \beta_{i^+ j^+ k^+ l^+} \;+\; p_{i^-} p_{j^+} \beta_{i^- j^+ k^- l^+} \;+\; p_{i^+} p_{j^-} \beta_{i^+ j^- k^- l^+} \big) \tag{9}$$

Consider first the case where both $[i]$ and $[j]$ consist of two elements. It is not hard to check that in this case $\beta_{i+j+k+l+} = 2\beta_{i-j+k-l+} = 2\beta_{i+j-k-l+} = \beta'_{ijkl}$. Using these relationships and the fact that $p_{i-} = p'_i - p_{i+}$, we may write the corresponding term in the sum (9) as $\frac{1}{2}(p'_i\,\beta'_{ijkl})p_{j+} + \frac{1}{2}(p'_j\,\beta'_{ijkl})p_{i+}$. The case where $[i]$ consists of two elements and $[j]$ of one contributes only a single term to (9), which may be written as $p_{i+}p_{j-}\beta_{i+j-k-l+} = (p'_j\,\beta'_{ijkl})p_{i+}$. The remaining case ($[i]$ consists of one element and $[j]$ of two) is symmetrical.
Bringing all these terms together we obtain

$$(p \times p)_{l+} = \sum_{i+}\left(\sum_{jk} p'_j\,\beta'_{ijkl}\right)p_{i+} = \sum_{i+}\alpha_{i+l+}p_{i+}. \tag{10}$$

Since q_e is invariant we have $\sum_{l+}(p \times p)_{l+} = \sum_{i+} p_{i+}$, so (10), suitably normalised, does indeed define a Markov chain as claimed. □

Suppose now that p is a pseudo-stationary population on $\mathcal{M}(T)$. Then clearly $p'_i = \sum_{j\in[i]}\pi_j$, where π is the normal population corresponding to p. Now it is easy to verify the following:

Proposition 11 *The Markov chain defined in Lemma 10 is reversible and converges to the stationary distribution $\widehat{\pi}_{l+} = \pi_{l+}/q_e$ over $\mathcal{M}(T)^+$.* □

We now address the question of the rate of convergence of the above Markov chain. Let $\Delta(t)$ denote the variation distance of the distribution at time t from $\widehat{\pi}$ maximised over initial states, and for $\epsilon \in (0,1)$ define

$$\tau(\epsilon) = \min\{t : \Delta(t') \le \epsilon \text{ for all } t' \ge t\}. \tag{11}$$

A bound on $\tau(\epsilon)$ may be obtained by viewing the chain as a network whose vertices are the states and which has an oriented edge of capacity $C(i^+,l^+) = \widehat{\pi}_{i+}\alpha_{i+l+}$ between vertices i^+ and l^+. Suppose it is possible to route $\widehat{\pi}_{a+}\widehat{\pi}_{b+}$ units of a commodity (a^+,b^+) from a^+ to b^+, for each pair of distinct states a^+ and b^+ simultaneously, in such a way that no oriented edge (i^+,l^+) carries more than $\rho C(i^+,l^+)$ units of flow in total and the length of any flow-carrying path is at most ℓ. Then Corollary 6′ of [17] provides the following bound (see also [5, 10]):

Theorem 12 [17] *The quantity $\tau(\epsilon)$ satisfies*

$$\tau(\epsilon) \le \rho\ell\left(\ln\widehat{\pi}_{\min}^{-1} + \ln\epsilon^{-1}\right),$$

where $\widehat{\pi}_{\min} = \min_{i+}\widehat{\pi}_{i+}$. □

The key to rapid convergence is the construction of a good flow in our Markov chain:

Theorem 13 *For the Markov chain of Lemma 10, there exists a flow with $\rho \le 2q_e^{-1}$ and $\ell = 2$.* □

Proof: Note from (8) that we may view each edge (i^+,l^+) of the Markov chain as a set of parallel edges, one for each $j \in \mathcal{M}(T')$: the edge associated with j has capacity $\widehat{\pi}_{i+}p'_j\beta'_{ijkl}$ and corresponds to mating i^+ with j^+ and producing l^+ as an offspring. We refer to this edge as the j-edge from i^+ to l^+. We will describe our flow in terms of these edges.
Let a^+,b^+ be any pair of distinct states in $\mathcal{M}(T)^+$, and consider the set $\mathcal{P}_{ab}$ of paths of length 2 from a^+ to b^+ of the following form: the first edge of the path is any b-edge leaving a^+, to c^+ say, and the second edge is the unique $\overline{c}$-edge from c^+ to b^+, where $\overline{c}$ is defined by $c \cup \overline{c} = a \cup b$ (viewed as multisets). We distribute the $\widehat{\pi}_{a+}\widehat{\pi}_{b+}$ units of flow from a^+ to b^+ evenly over all 2^r paths in $\mathcal{P}_{ab}$, where r is the number of non-trivial connected components in $a^+ \cup b^+$.
Now consider any j-edge e from i^+ to l^+. We may assume that the equivalence class $[j]$ contains two elements since in the above flow we have only used j-edges for which $j^+ \in \mathcal{M}(T)^+$. Let us compute the flow through e. Clearly only two flow-carrying paths use e; for if e is the first edge of a path then this path must belong to $\mathcal{P}_{ij}$, while if it is the second edge of a path then this path must belong to $\mathcal{P}_{ml}$, where m is the complement of l in the multiset $i \cup j$. Suppose e is the first edge of a path from i^+ to j^+. The flow along this path is precisely $2^{-r}\widehat{\pi}_{i+}\widehat{\pi}_{j+}$, where r is the number of non-trivial connected components in $i^+ \cup j^+$. Dividing this by the capacity of e we get

$$\frac{2^{-r}\widehat{\pi}_{i+}\widehat{\pi}_{j+}}{\widehat{\pi}_{i+}p'_j\beta'_{ijkl}} = \frac{2^{-r}\pi_{j+}/q_e}{(\pi_{j+}+\pi_{j-})2^{-r}} \le \frac{1}{q_e},$$

where we have used the fact that $\beta'_{ijkl} = 2^{-r}$. By a symmetrical argument, the ratio of flow to capacity in the case where e is the second edge on a path can be similarly bounded. Hence $\rho \le 2q_e^{-1}$. Evidently all flow-carrying paths have length 2 so $\ell = 2$. □

Finally, we may combine the above results to derive a bound on the rate of convergence of the original quadratic mating process. Define the quantity $\tau^{\text{match}}(\epsilon)$ in analogous fashion to $\tau(\epsilon)$ in (11) but using the variation distance of $p(t)$ from the corresponding stationary population π defined in Theorem 8.

Theorem 14 *Let T be a tree with n vertices, and $p(0)$ any population that generates $\mathcal{M}(T)$. Then $\tau^{\text{match}}(\epsilon)$ satisfies*

$$\tau^{\text{match}}(\epsilon) = \mathrm{O}\Big((n\lg n + n\lg\pi_{\min}^{-1} + \lg\epsilon^{-1})q_{\min}^{-1}\Big),$$

where $q_{\min} = \min_e q_e$ and $\pi_{\min} = \min_{i\in\mathcal{M}(T)}\pi_i$.

Proof (sketch): By Lemma 9 we may assume inductively that, after sufficiently many steps, the population p' is close to stationarity. Lemma 10 now allows us to view the system as a Markov chain on $\mathcal{M}(T)^+$. Strictly speaking, since p' is not exactly stationary the transition probabilities appearing in (8) are not in fact constant but are perturbed by a small time-varying amount; however, this error may be effectively bounded. Now we apply Theorem 12 together with the flow bound of Theorem 13 to obtain a bound on the rate of convergence of this Markov chain. Clearly, once the Markov chain is close to its stationary distribution, near-stationarity of p' implies that the population on $\mathcal{M}(T)$ is also almost stationary. The details are left for the full paper. □

Theorem 14 is a surprisingly strong result. It says that the time required for the mating process to approach stationarity depends on the size of the tree only as a low degree polynomial, despite the fact that the number of types $|\mathcal{M}(T)|$ is in general exponentially large in n. This is apparently the first analysis of its kind for a combinatorially non-trivial problem. Moreover, although this is a somewhat simplified setting, we conjecture that the analytical tools introduced in this paper are applicable to other systems.

Acknowledgements

We thank Eric Carlen, Persi Diaconis, Jeff Kahn, Jeff Lagarias, Han La Poutré, Noam Nisan, Nick Pippenger, Dirk Vertigan and Benjamin Weiss for enlightening discussions on the ideas presented here.

References

[1] M. Ajtai, J. Komlós and E. Szemerédi. Deterministic simulation in LOGSPACE. *Proceedings of 19th ACM STOC* (1987), p. 132.

[2] E. Akin. *The geometry of population genetics.* Lecture Notes in Biomathematics, Vol. 31, Springer-Verlag, Berlin, 1979.

[3] E. Carlen and M. Carvalho. Strict entropy production bounds and stability of the rate of convergence to equilibrium for the Boltzmann equation. Preprint, 1992.

[4] I. Csiszár and J. Körner. *Information Theory: Coding Theorems for Discrete Memoryless Systems.* Academic Press, 1981.

[5] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of Markov chains. *Annals of Applied Probability* **1** (1991), pp. 36–61.

[6] M. Dyer, A. Frieze and R. Kannan. A random polynomial time algorithm for approximating the volume of convex bodies. *Proceedings of 21st ACM STOC* (1989), pp. 375–381.

[7] M. Hirsch and S. Smale. *Differential equations, dynamical systems, and linear algebra.* Academic Press, New York, 1974.

[8] J. Holland. *Adaptation in natural and artificial systems.* MIT Press, 1992.

[9] M. Jerrum. Large cliques elude the Metropolis process. *Random Structures and Algorithms* **3** (1992), pp. 347–359.

[10] M. Jerrum and A. Sinclair. Approximating the permanent. *SIAM Journal on Computing* **18** (1989), pp. 1149–1178.

[11] O. Penrose. *Foundations of statistical mechanics.* Pergamon Press, Oxford, 1970.

[12] Y. Rabinovich, A. Sinclair and A. Wigderson. Quadratic dynamical systems. Full version, in preparation.

[13] Y. Rabinovich and A. Wigderson. Analysis of a simple genetic algorithm. *Proceedings of the 4th International Conference on Genetic Algorithms* 1991, pp. 215–221.

[14] A. Rényi. *Foundations of probability.* Holden-Day, San Francisco, 1970.

[15] G. Sasaki and B. Hajek. The time complexity of maximum matching by simulated annealing. *Journal of the ACM* **35** (1988), pp. 387–403.

[16] H. Siegelmann. Personal communication.

[17] A. Sinclair. Improved bounds for mixing rates of Markov chains and multicommodity flow. Report ECS-LFCS-91-178, Dept. of Computer Science, University of Edinburgh, 1991. To appear in *Combinatorics, Probability & Computing.*

Appendix: Genetic algorithms

Genetic algorithms provide a heuristic paradigm for solving combinatorial optimisation problems based on a biological analogy. Suppose we are searching for an optimal (or near-optimal) solution amongst a set of feasible solutions; as a concrete example, consider the problem of finding a matching of maximum cardinality amongst all matchings in a graph. The algorithm

starts with an initial population over simple solutions that is easily constructed; in our example, this might be the uniform distribution on all matchings that consist of a single edge. The population evolves in discrete steps under the application of two operators, called *mating* and *fitness*. Mating is designed to increase variety in the population, while fitness is designed to increase its average quality. In our example, the quality of a matching is simply its size. Usually a third operator, known as *mutation*, is added, but for the sake of simplicity we shall ignore this here.

In mating, random pairs of solutions give birth to random children; the operator β^{match} discussed in Sections 3 and 4 is a natural choice in our example. Note that the average size of a matching is an invariant of β^{match}, so mating itself does not improve the population. In fitness, solutions of low quality are discarded or their number reduced; for example, all matchings below a certain size might be removed from the population. The overall genetic algorithm repeatedly applies the mating and fitness operators alternately, in the hope that the population will quickly converge to one whose average quality is high.

Despite the apparent success of this approach in several experimental applications, few rigorous theoretical results exist to explain and motivate the experimental results or guide the design of the algorithms. One of the few such rigorous results was obtained in [13], albeit in the context of a highly artificial optimisation problem. One lesson to be learned from that work was that the interaction between the mating and fitness operators is much easier to analyse if mating is applied enough times for the system to converge to a (near-)stationary population before the fitness operator is applied; this is a reasonable strategy with mating operators which, like β^{match}, preserve or do not substantially reduce the average quality of solutions. The reason for this is clear: in general, the effect on the population of a single step of the mating operator is rather hard to quantify, but after a sufficiently long sequence of mating operations the population is close to a stationary population, which may be well understood. Interestingly, we have recently learned that this idea has led to improved performance of genetic algorithms in some experiments [16]. Given that the effect of the fitness operator is usually relatively straightforward to describe, we contend that the principal challenge in analysing the performance of genetic algorithms of the above kind lies in estimating the rate of convergence to a stationary population under the mating operator. Since the mating operator is quadratic and almost always symmetric, this is precisely the central question that is addressed in the present paper.