Entropy production in nonlinear recombination models

PIETRO CAPUTO[1] and ALISTAIR SINCLAIR[2]

[1]*Dipartimento di Matematica, Università Roma Tre, Italy.*
E-mail: caputo@mat.uniroma3.it; url: http://www.mat.uniroma3.it/users/caputo/
[2]*Computer Science Division, Soda Hall, University of California Berkeley, CA 94720-1776, USA.*
E-mail: sinclair@cs.berkeley.edu; url: https://www.cs.berkeley.edu/~sinclair/

We study the convergence to equilibrium of a class of nonlinear recombination models. In analogy with Boltzmann's H-theorem from kinetic theory, and in contrast with previous analysis of these models, convergence is measured in terms of relative entropy. The problem is formulated within a general framework that we refer to as Reversible Quadratic Systems. Our main result is a tight quantitative estimate for the entropy production functional. Along the way, we establish some new entropy inequalities generalizing Shearer's and related inequalities.

Keywords: Boltzmann equation; entropy; functional inequalities; nonlinear equations; population dynamics

1. Introduction

Recombination models based on random mating have a wide range of applications in the natural sciences and play an important role in the analysis of genetic algorithms. The following nonlinear system is a commonly studied model in population genetics [22]. Let $\Omega = X_1 \times \cdots \times X_n$ denote the set of sequences of length n, such that the ith element of the sequence takes values in a given finite space X_i. A sequence $\sigma \in \Omega$ is written as $\sigma = (\sigma_i, i \in [n])$, where $[n] = \{1, \ldots, n\}$ is the set of *loci*, or *sites*, and $\sigma_i \in X_i$ for all $i \in [n]$. Given a subset $A \subset [n]$, and a $\sigma \in \Omega$ we write σ_A for the A-component of σ, that is, the subsequence $(\sigma_i, i \in A)$. If $(\sigma, \eta) \in \Omega \times \Omega$ is a pair of sequences, the *recombination* at A consists in exchanging the A-component of σ with the A-component of η. This defines the map

$$(\sigma, \eta) \mapsto (\eta_A \sigma_{A^c}, \sigma_A \eta_{A^c}).$$

If the original pair (σ, η) is obtained by sampling independently from a probability measure p on Ω, then the sequence $\eta_A \sigma_{A^c}$ is distributed according to $p_A \otimes p_{A^c}$, the product measure obtained from the marginals of p on A and A^c. By choosing the set A at random according to some probability distribution ν one obtains the quadratic dynamical system

$$p \mapsto \Psi[p] := \sum_A \nu(A)(p_A \otimes p_{A^c}). \tag{1.1}$$

The discrete time evolution of the initial distribution p is then defined by iteration of the map Ψ, namely $p^{(k)} = \Psi[p^{(k-1)}]$, $k \in \mathbb{N}$, $p^{(0)} = p$.

Analogously, in continuous time one has the quadratic differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t} p_t = \sum_A \nu(A)(p_{t,A} \otimes p_{t,A^c} - p_t), \qquad t \geq 0, \tag{1.2}$$

with the initial condition $p_0 = p$. Here p_t is the probability measure describing the state at time t, and $p_{t,A}$ denotes its marginal on A.

When $A = \{i\}$ is a single site we write $p_{t,i}$ for the marginal at i. It is not difficult to see that the map (1.1) preserves the single site marginals, so that $p_{t,i} = p_{0,i}$ for all $t \geq 0$ and for all $i \in [n]$. The study of this model goes back to the pioneering work of Geiringer [16]; see also [1,24,27,28] for more recent accounts.

As emphasized in [27,28] this model is a special case of the much larger class of so-called "quadratic dynamical systems", which provides a rich family of discrete analogues of Boltzmann's equation from statistical physics [30]. It is a classical result that, under an obvious non-degeneracy assumption on the distribution ν, the system converges to the stationary state given by the product of the marginals of the initial state p; that is, if $p_i = p_{0,i}$ denotes the marginal of p at site i, then

$$\pi = \otimes_i p_i \tag{1.3}$$

is the equilibrium distribution and one has convergence in distribution: $p^{(k)} \to \pi$, $k \to \infty$, and $p_t \to \pi$, $t \to \infty$. We shall be interested in the speed of convergence to equilibrium in both continuous and discrete time.

We consider the following natural examples of the distribution ν:

(1) Single site recombination: $\nu(A) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(A = \{i\})$;
(2) One-point crossover: $\nu(A) = \frac{1}{n+1} \sum_{i=0}^n \mathbf{1}(A = J_i)$, where $J_0 = \varnothing$, $J_i = \{1, \ldots, i\}$, $i \geq 1$;
(3) Uniform crossover: $\nu(A) = \frac{1}{2^n}$, for all $A \subset [n]$;
(4) The Bernoulli(q) model: for some $q \in [0, \frac{1}{2}]$, $\nu(A) = q^{|A|}(1-q)^{n-|A|}$.

The Bernoulli(q) model is a generalization of the uniform crossover model. In principle our method can be applied to other generalizations, such as the k-crossover model or the so-called Poisson model. However, to keep this work at a reasonable length, in what follows we will restrict attention to the models listed above. The first example generates a simple linear evolution, but the other choices produce genuinely nonlinear processes.

Tight estimates on the speed of convergence in total variation norm of the associated discrete time processes were obtained in [27]. While the first example reduces to a standard coupon collecting argument, implying that the system mixes in $\Theta(n \log n)$ steps, the other cases require a finer coupling analysis. In particular, it is shown in [27] that one-point crossover mixes in $\Theta(n \log n)$ steps, while uniform crossover mixes in $\Theta(\log n)$ steps. Further results on convergence to equilibrium in total variation norm together with the analysis of the quasi-stationary measure were recently obtained in [24].

In this paper, we focus on convergence to equilibrium in terms of *entropy*. Here we recall that the *relative entropy* of p with respect to μ, for two probability measures p, μ on Ω, is given by

$$H(p \mid \mu) = \sum_{\sigma \in \Omega} p(\sigma) \log\big(p(\sigma)/\mu(\sigma)\big),$$

with $H(p \mid \mu) = +\infty$ if there exists $\sigma \in \Omega$ with $\mu(\sigma) = 0$ and $p(\sigma) \neq 0$. Recall also Pinsker's inequality, asserting that

$$\|p - \mu\| \leq \sqrt{\frac{1}{2} H(p \mid \mu)}, \tag{1.4}$$

where $\|p - \mu\| = \frac{1}{2} \sum_{\sigma \in \Omega} |p(\sigma) - \mu(\sigma)|$ denotes the total variation distance. Now, for any probability measure p on Ω, the convergence $p_t \to \pi$ implies that the relative entropy $H(p_t \mid \pi)$ satisfies $H(p_t \mid \pi) \to 0$, $t \to \infty$. It is then natural to study the rate of exponential decay of relative entropy or, equivalently, the existence of an inequality of the form

$$\frac{\mathrm{d}}{\mathrm{d}t} H(p_t \mid \pi) \leq -\delta H(p_t \mid \pi), \tag{1.5}$$

for some $\delta > 0$ independent of t. The bound (1.5) is often referred to as an *entropy production* estimate.

Our main motivation for the study of entropy production estimates in nonlinear recombination models comes from the analogy with kinetic theory [8]. In recent years, the study of the validity of a bound of the type (1.5) for the Boltzmann equation and related quadratic evolutions has been at the heart of many spectacular developments in kinetic theory; see, for example, [11] for a survey. For the Boltzmann equation, a famous conjecture of Cercignani asserting that a positive rate δ exists for all initial conditions has been shown to be true only for some special choices of the collision kernels, while for the more common choices of the collision kernel one may have a zero entropy production rate for suitable initial conditions [31]. It remains a very active area of research to establish the exponential decay of relative entropy under suitable moment and smoothness assumptions on the initial distribution. Similarly, for the Boltzmann–Kac equation it is known that there cannot be exponential decay in relative entropy for all initial states, and it is a major open problem to establish exponential decay of relative entropy for initial distributions satisfying suitable moment conditions; see [7]. On the other hand, in our finite dimensional setting one expects that any initial distribution decays exponentially fast to equilibrium in terms of relative entropy. It is then of interest to investigate how the rate of decay depends on the initial distribution and on the parameter n determining the dimension of the problem. Actually, we shall see that one can find a positive rate $\delta = \delta(n)$ such that (1.5) holds uniformly for all initial distributions, and we shall determine the precise behavior of this δ as a function of n up to a constant factor.

We discuss the problem of entropy production in the framework of what we refer to as Reversible Quadratic Systems. These provide a generalized class of natural nonlinear dynamics converging to an arbitrary target distribution (not necessarily of product form), such as the Ising model on a graph or other random combinatorial structures. This may be viewed as a nonlinear analog of the standard Markov Chain Monte Carlo paradigm; see Section 3 below.

To the best of our knowledge, this is the first work dealing systematically with decay to equilibrium in the sense of relative entropy for nonlinear evolution of discrete combinatorial structures, and as such it is likely to provide a convenient stage for testing some of the results and conjectures formulated in the continuous setting of kinetic theory. The hope is that this alternative setting allows one to isolate those features of entropy decay that are less model specific.

We emphasize also that convergence to equilibrium with respect to relative entropy can in general be quite different from convergence in total variation. For example, from (1.4) one knows that relative entropy cannot decay faster than the square of the total variation norm, but a comparison of Theorem 1.2 below with the main results in [27] shows that, in the case of recombinations with uniform crossover, the decay can be considerably slower. See, for example, [13] for related phenomena in the kinetic theory literature.

Finally, let us add that, following an important idea of M. Kac [20], it should be possible to interpret the nonlinear models considered here as the single particle marginal of a linear system of N particles interacting in a mean field fashion, in the limit as $N \to \infty$. Establishing this picture rigorously can be a very challenging problem; see, for example, [7,18,25] for a survey of recent interesting developments on the so-called Kac program in kinetic theory. A crucial part of this program consists in obtaining a sharp control of the decay to equilibrium for the linear system in terms of relative entropy. In forthcoming work, we plan to investigate a family of discrete combinatorial versions of this program.

To describe our main results, it is convenient to reformulate the entropy production estimate in terms of suitable functional inequalities. Let π be a product measure on Ω of the form (1.3). For any nonnegative function $f : \Omega \mapsto [0, \infty)$, define the entropy functional

$$\text{Ent}(f) = \pi[f \log f] - \pi[f] \log \pi[f].$$

For any $A \subset [n]$, let f_A denote the function

$$f_A(\sigma) = \sum_{\eta \in \Omega} \pi(\eta) f(\sigma_A \eta_{A^c});$$

note that f_A depends only on σ_A. When $A = \{i\}$ for some $i \in [n]$, we simply write f_i for $f_{\{i\}}$. Let $\mathbb{S}_\pi$ denote the set of all $f : \Omega \mapsto [0, \infty)$ such that $\pi[f] = 1$ and $f_i = 1$ for all $i \in [n]$. Notice that $\mathbb{S}_\pi$ is precisely the set of all functions f of the form $f = p/\pi$, where p is any probability measure on Ω satisfying (1.3); that is, f is the density of p with respect to π and p has the same marginals as π. Moreover, $f_A f_{A^c}$ is the density of $p_A \otimes p_{A^c}$ with respect to π. Given a distribution ν over subsets of $[n]$ we call $\delta(\pi, \nu)$ the largest constant $\delta \geq 0$ such that the inequality

$$\sum_A \nu(A)\pi \left[(f_A f_{A^c} - f) \log \frac{f_A f_{A^c}}{f} \right] \geq \delta \, \text{Ent}(f) \tag{1.6}$$

holds for all $f \in \mathbb{S}_\pi$. As we shall see, inequality (1.6) coincides with (1.5) when $f = p_t/\pi$ is the density of p_t with respect to π. Define also

$$\delta(\nu) = \inf_\pi \delta(\pi, \nu),$$

where the infimum is taken over all product measures on $X_1 \times \cdots \times X_n$ and over all possible underlying finite spaces X_i.

Theorem 1.1. *The recombination models defined above satisfy the following bounds*:

(1) *Single site recombination*: $\frac{2}{n} + O(n^{-2}) \geq \delta(\nu) \geq \frac{1}{n-1}$;

(2) *One-point crossover*: $\frac{4}{n} + O(n^{-2}) \geq \delta(\nu) \geq \frac{1}{n+1}$;

(3) *Uniform crossover*: $\frac{4}{n} + O(n^{-2}) \geq \delta(\nu) \geq \frac{1-2^{-n+1}}{n-1}$;

(4) Bernoulli(q) *model*: $\frac{4(1-(1-q/2)^n)}{n} + O(n^{-2}) \geq \delta(\nu) \geq \frac{1-(1-q)^n-q^n}{n-1}$.

Inequality (1.6) can be seen as a nonlinear version of a logarithmic Sobolev inequality; see, for example, [12] for background on logarithmic Sobolev inequalities in the usual Markov chain setting. Remarkable works have been devoted to the study of functional inequalities of the form (1.6) in the Boltzmann equation literature; see, for example, [5,31] and [11,30] for an overview. In our combinatorial setting, we shall establish the entropy production inequalities of Theorem 1.1 by proving a new set of inequalities for the entropy functional of a product measure. More precisely, we shall first observe (see Lemma 4.2), that $\delta(\pi, \nu) \geq \kappa(\pi, \nu)$, where $\kappa(\pi, \nu)$ denotes the largest possible constant $\kappa > 0$ such that the inequality

$$\sum_A \nu(A)\big(\mathrm{Ent}(f_A) + \mathrm{Ent}(f_{A^c})\big) \leq (1-\kappa)\,\mathrm{Ent}(f) \tag{1.7}$$

holds for all $f \in \mathbb{S}_\pi$. Define

$$\kappa(\nu) = \inf_\pi \kappa(\pi, \nu),$$

where the infimum is taken over all product measures on $X_1 \times \cdots \times X_n$ and over all choices of the spaces X_i. Then, since $\delta(\nu) \geq \kappa(\nu)$, we can obtain the lower bounds in Theorem 1.1 by computing κ. The following result, which may be of independent interest, does this for all of the above recombination models.

Theorem 1.2. *For the recombination models defined above, the constant $\kappa(\nu)$ satisfies*:

(1) *Single site recombination*: $\kappa(\nu) = \frac{1}{n-1}$;

(2) *One-point crossover*: $\kappa(\nu) = \frac{1}{n+1}$;

(3) *Uniform crossover*: $\kappa(\nu) = \frac{1-2^{-n+1}}{n-1}$;

(4) Bernoulli(q) *model*: $\kappa(\nu) = \frac{1-(1-q)^n-q^n}{n-1}$.

Inequality (1.7) expresses a generalized subadditivity property of the entropy functional for a product measure. As discussed in Section 4 below, it can be understood as a nontrivial refinement of Shearer's inequality for Shannon's entropy [9]. We refer also to [2,15,23] for other interesting extensions and applications of Shearer's inequality. Our setting is somewhat non-standard because of the restriction to $f \in \mathbb{S}_\pi$. It is important to note that without it there cannot be a positive constant κ in (1.7). Indeed, by taking f of the form $\prod_i f_i$ for some nontrivial single site marginals f_i, one has $\mathrm{Ent}(f_A) + \mathrm{Ent}(f_{A^c}) = \mathrm{Ent}(f)$ for any A so that (1.7) can only hold with $\kappa = 0$.

As mentioned earlier, the lower bounds in Theorem 1.1 are a consequence of Theorem 1.2 and the fact that $\delta(\nu) \geq \kappa(\nu)$. The upper bounds on the other hand will follow by exhibiting an explicit test function in (1.6).

Finally, we turn to the consequences of Theorem 1.1 for the convergence to equilibrium of both the continuous time and the discrete time evolutions. The following result shows that in all

models considered above one has exponential decay in relative entropy with rate $\kappa(\nu)$ and that, if we insist on uniformity in the initial state p, this decay rate is optimal up to a constant factor.

Corollary 1.3. *Consider the recombination model with distribution ν, and let $\kappa(\nu)$ be as in Theorem 1.2. For any initial state p, let π denote the associated product measure (1.3). Then, in continuous time one has*

$$H(p_t \mid \pi) \leq e^{-\kappa(\nu)t} H(p \mid \pi), \qquad t \geq 0. \tag{1.8}$$

Similarly, in discrete time

$$H\big(p^{(k)} \mid \pi\big) \leq \big(1 - \kappa(\nu)\big)^k H(p \mid \pi), \qquad k \in \mathbb{N}. \tag{1.9}$$

Moreover, there exists an initial state p such that, if $\gamma(\nu)$ denotes the upper bound on $\delta(\nu)$ appearing in Theorem 1.1, then

$$\frac{\mathrm{d}}{\mathrm{d}t} H(p_t \mid \pi)\bigg|_{t=0^+} \geq -\gamma(\nu) H(p \mid \pi), \qquad H\big(p^{(1)} \mid \pi\big) \geq \big(1 - \gamma(\nu)\big) H(p \mid \pi). \tag{1.10}$$

These bounds are tight in the sense that there exists a constant $C > 0$, independent of n, such that $\gamma(\nu) \leq C\kappa(\nu)$ for all models (1)–(3). *In the case of the* Bernoulli(q) *model, the bound $\gamma(\nu) \leq C\kappa(\nu)$ holds provided $q \geq n^{-2}$.*

The proofs of Theorem 1.1, Theorem 1.2 and Corollary 1.3 are deferred to Section 4. In Section 2, we formulate the problem of entropy production estimates in the much more general setting of reversible quadratic systems. In Section 3, we present some examples of reversible quadratic systems, including the recombination models defined above.

2. Reversible quadratic systems

Following [28], we introduce a general framework, which includes the recombination models as special cases. We shall work in continuous time, but as we will see a translation to the discrete time setting is immediate.[1]

2.1. Setup

Let $\mathbb{X}$ be a finite space, and call $\mathbb{P}(\mathbb{X})$ the set of probability measures on $\mathbb{X}$. We refer to, for example, [26] for background on Markov chains. A *reversible quadratic system (RQS)* on $\mathbb{X}$ is defined by a pair $(\mathbb{G}, \mu)$ where

---

[1][28] worked in discrete time, and also restricted attention to *symmetric* quadratic systems, which are reversible quadratic systems with μ uniform.

(1) $\mathbb{G}$ is the infinitesimal generator of a Markov chain with state space $\mathbb{X} \times \mathbb{X}$, that is, $\mathbb{G}(\sigma, \sigma'; \tau, \tau') \geq 0$ for all $(\sigma, \sigma') \neq (\tau, \tau')$ and

$$\mathbb{G}(\sigma, \sigma'; \sigma, \sigma') = - \sum_{(\tau, \tau') \neq (\sigma, \sigma')} \mathbb{G}(\sigma, \sigma'; \tau, \tau'). \tag{2.1}$$

(2) $\mu \in \mathbb{P}(\mathbb{X})$ satisfies $\mu(\sigma) > 0$ for all $\sigma \in \mathbb{X}$ and is such that $\mu \otimes \mu$ is reversible for $\mathbb{G}$, that is, for all $\sigma, \sigma', \tau, \tau' \in \mathbb{X}$ one has

$$\mu(\sigma)\mu(\sigma')\mathbb{G}(\sigma, \sigma'; \tau, \tau') = \mu(\tau)\mu(\tau')\mathbb{G}(\tau, \tau'; \sigma, \sigma'). \tag{2.2}$$

Moreover, we assume that $\mathbb{G}$ has the symmetry

$$\mathbb{G}(\sigma, \sigma'; \tau, \tau') = \mathbb{G}(\sigma', \sigma; \tau', \tau), \tag{2.3}$$

for all $\sigma, \sigma', \tau, \tau' \in \mathbb{X}$.

The quantity $\mathbb{G}(\sigma, \sigma'; \tau, \tau')$ is interpreted as the rate at which the pair (τ, τ') is produced from a collision (or "mating") between σ and σ'. Below, $\mathbb{G}$ will be mostly of the form $\mathbb{G} = Q - \mathbf{1}$, where Q is a Markov kernel on $\mathbb{X} \times \mathbb{X}$ with reversible measure $\mu \otimes \mu$. It is important to note that the Markov chain on the product space $\mathbb{X} \times \mathbb{X}$ defining the RQS is not assumed to be irreducible, and therefore μ can be taken to be any of the possibly many measures such that $\mu \otimes \mu$ is reversible for $\mathbb{G}$. Indeed, in most cases to be considered below, $\mathbb{G}$ will not be irreducible and we shall use that freedom in selecting μ.

The dynamics of the system are specified by the equation

$$\frac{\mathrm{d}}{\mathrm{d}t} p_t(\tau) = \sum_{\tau' \in \mathbb{X}} \Phi[p_t](\tau, \tau'), \tag{2.4}$$

where we define, for any $p \in \mathbb{P}(\mathbb{X})$,

$$\Phi[p](\tau, \tau') := \sum_{\sigma, \sigma' \in \mathbb{X}} p(\sigma)p(\sigma')\mathbb{G}(\sigma, \sigma'; \tau, \tau').$$

We consider equation (2.4) with the initial condition $p_0 = p$ for some given $p \in \mathbb{P}(\mathbb{X})$. By (2.1), for any fixed σ, σ' one has $\sum_{(\tau, \tau')} \mathbb{G}(\sigma, \sigma'; \tau, \tau') = 0$, which implies the conservation law

$$\sum_{\tau \in \mathbb{X}} p_t(\tau) = 1, \tag{2.5}$$

for all $t \geq 0$. Thus equation (2.4) gives a well defined evolution of the state of the system. In general, there are many other conservation laws in the system (2.4); see part 3 of Proposition 2.3 below. Existence and uniqueness of the solution of (2.4) for any $p \in \mathbb{P}(\mathbb{X})$ can be established in a standard way; see, for example, [1].

From reversibility (2.2), setting $f(\sigma) = p(\sigma)/\mu(\sigma)$ one finds

$$\Phi[p](\tau, \tau') = \sum_{\sigma, \sigma' \in \mathbb{X}} \mu(\tau)\mu(\tau')\mathbb{G}(\tau, \tau'; \sigma, \sigma')[f(\sigma)f(\sigma') - f(\tau)f(\tau')]. \tag{2.6}$$

Therefore, in terms of $f_t(\sigma) = p_t(\sigma)/\mu(\sigma)$, equation (2.4) becomes

$$\frac{\mathrm{d}}{\mathrm{d}t} f_t(\tau) = \sum_{\sigma, \sigma', \tau' \in \mathbb{X}} \mu(\tau') \mathbb{G}(\tau, \tau'; \sigma, \sigma') \left[f_t(\sigma) f_t(\sigma') - f_t(\tau) f_t(\tau') \right].$$

Remark 2.1. RQS include linear evolutions associated to Markov chains as a special case. For instance, if $\mathbb{G}_0$ is a Markov generator with state space $\mathbb{X}$, with reversible measure μ, then the expression

$$\mathbb{G}(\sigma, \sigma'; \tau, \tau') = \mathbb{G}_0(\sigma, \tau) \mathbf{1}(\sigma' = \tau') + \mathbb{G}_0(\sigma', \tau') \mathbf{1}(\sigma = \tau),$$

defines a RQS $(\mathbb{G}, \mu)$. In this case the evolution is linear, and coincides with the Markov chain generated by $\mathbb{G}_0$, that is, $p_t = p e^{t\mathbb{G}_0}$.

Remark 2.2. Given the generator $\mathbb{G}$, let $S(\tau, \tau'; \sigma, \sigma') = \frac{1}{2}(\mathbb{G}(\tau, \tau'; \sigma, \sigma') + \mathbb{G}(\tau', \tau; \sigma, \sigma'))$. Note that this does not define a Markov generator, since off-diagonal elements need not be non-negative. However, it is not hard to check that

$$\mathbb{G}_{\mathrm{sym}}(\tau, \tau'; \sigma, \sigma')$$
$$= \begin{cases} S(\tau, \tau'; \sigma, \sigma') & \text{if } (\sigma, \sigma') \neq (\tau', \tau), (\tau, \tau'), \\ \mathbb{G}(\tau, \tau'; \tau', \tau) & \text{if } (\sigma, \sigma') = (\tau', \tau), \\ S(\tau, \tau'; \tau, \tau') + S(\tau, \tau'; \tau', \tau) - \mathbb{G}(\tau, \tau'; \tau', \tau) & \text{if } (\sigma, \sigma') = (\tau, \tau') \end{cases}$$

does define a Markov generator, i.e., for all $\tau, \tau' \in \mathbb{X}$ one has $\sum_{\sigma, \sigma'} \mathbb{G}_{\mathrm{sym}}(\tau, \tau'; \sigma, \sigma') = 0$, and $\mathbb{G}_{\mathrm{sym}}(\tau, \tau'; \sigma, \sigma') \geq 0$ if $(\sigma, \sigma') \neq (\tau, \tau')$. From (2.6) and (2.3), it is not hard to see that $\Phi[p](\tau, \tau')$ is unchanged if $\mathbb{G}$ is replaced by $\mathbb{G}_{\mathrm{sym}}$. In particular, we may and will assume without loss of generality, that the generator $\mathbb{G}$ satisfies the symmetry

$$\mathbb{G}(\tau, \tau'; \sigma, \sigma') = \mathbb{G}(\tau', \tau; \sigma, \sigma'), \qquad \text{if } (\sigma, \sigma') \neq (\tau', \tau), (\tau, \tau'). \tag{2.7}$$

By reversibility (2.2), from (2.7) one has the further symmetry

$$\mathbb{G}(\sigma, \sigma'; \tau, \tau') = \mathbb{G}(\sigma, \sigma'; \tau', \tau), \qquad \text{if } (\sigma, \sigma') \neq (\tau', \tau), (\tau, \tau').$$

In most examples below, $\mathbb{G}$ has the form $\mathbb{G} = Q - \mathbf{1}$ for some Markov kernel Q. In this case the symmetry (2.7) is equivalent to $Q(\sigma, \sigma'; \tau, \tau') = Q(\sigma, \sigma'; \tau', \tau)$, for all $(\sigma, \sigma') \neq (\tau', \tau), (\tau, \tau')$.

2.2. Entropy and stationary states

Reversible quadratic systems satisfy an analogue of Boltzmann's H theorem from kinetic theory, which we summarize in Proposition 2.3 below. Call $\rho \in \mathbb{P}(\mathbb{X})$ *stationary* if $\Phi[\rho] = 0$. Note that by (2.6) the reference measure μ is stationary. Let $\mathbb{P}_+(\mathbb{X})$ denote the set of $\rho \in \mathbb{P}(\mathbb{X})$ such that $\rho(\sigma) > 0$ for all $\sigma \in \mathbb{X}$. For any RQS $(\mathbb{G}, \mu)$, one has the following facts.

Proposition 2.3.

(1) *For any initial state $p \in \mathbb{P}(\mathbb{X})$,*

$$\frac{\mathrm{d}}{\mathrm{d}t} H(p_t \mid \mu) = -D(f_t, f_t), \tag{2.8}$$

where $f_t(\sigma) = p_t(\sigma)/\mu(\sigma)$ and, for any $f, g : \mathbb{X} \mapsto [0, \infty)$:

$$D(f, g) := \frac{1}{4} \sum_{\sigma, \sigma', \tau, \tau' \in \mathbb{X}} \mu(\tau)\mu(\tau')\mathbb{G}(\tau, \tau'; \sigma, \sigma')$$

$$\times \left[f(\sigma)f(\sigma') - f(\tau)f(\tau') \right] \log \frac{g(\sigma)g(\sigma')}{g(\tau)g(\tau')}.$$

In particular, $D(f_t, f_t) \geq 0$ for all $t \geq 0$ and $D(f_t, f_t) = 0$ iff p_t is stationary.

(2) *$\rho \in \mathbb{P}(\mathbb{X})$ is stationary iff $f(\sigma) = \rho(\sigma)/\mu(\sigma)$ satisfies*

$$f(\sigma)f(\sigma') = f(\tau)f(\tau'),$$

for all $\sigma, \sigma', \tau, \tau' \in \mathbb{X}$ such that $\mathbb{G}(\tau, \tau'; \sigma, \sigma') > 0$. In particular, $\rho \otimes \rho$ is reversible for $\mathbb{G}$ iff ρ is stationary.

(3) *If $\rho \in \mathbb{P}_+(\mathbb{X})$ is stationary, then for any initial condition $p \in \mathbb{P}(\mathbb{X})$, one has*

$$\frac{\mathrm{d}}{\mathrm{d}t} \sum_{\sigma \in \mathbb{X}} p_t(\sigma) \log\big(\rho(\sigma)/\mu(\sigma)\big) = 0.$$

Proof. Differentiating and using (2.5) one has

$$\frac{\mathrm{d}}{\mathrm{d}t} H(p_t \mid \mu) = \sum_{\tau \in \mathbb{X}} \left[\frac{\mathrm{d}}{\mathrm{d}t} p_t(\tau) \right] \log \frac{p_t(\tau)}{\mu(\tau)}$$

$$= \sum_{\sigma, \sigma', \tau, \tau' \in \mathbb{X}} \mu(\tau)\mu(\tau')\mathbb{G}(\tau, \tau'; \sigma, \sigma')\left[f_t(\sigma)f_t(\sigma') - f_t(\tau)f_t(\tau') \right] \log f_t(\tau).$$

By the symmetry (2.7), one can replace $\log f_t(\tau)$ above with $\log(f_t(\tau)f_t(\tau'))$ at the cost of a factor $1/2$. Finally, by reversibility one obtains (2.8). Clearly, $D(f, f) \geq 0$ and $D(f, f) = 0$ iff f satisfies $f(\sigma)f(\sigma') = f(\tau)f(\tau')$, whenever $\mathbb{G}(\tau, \tau'; \sigma, \sigma') > 0$. From this and (2.6) it follows that $D(f_t, f_t) = 0$ iff p_t is stationary. This proves part 1.

Part 2 follows in the same way since if p is stationary then $p_t = p$ for all t.

Finally, for part 3, reasoning as above, if $g(\sigma) = \rho(\sigma)/\mu(\sigma)$, $\rho \in \mathbb{P}_+(\mathbb{X})$, then

$$\frac{\mathrm{d}}{\mathrm{d}t} \sum_{\tau \in \mathbb{X}} p_t(\tau) \log\big(g(\tau)\big) = -D(f_t, g).$$

Since ρ is stationary, one has that $D(f_t, g) = 0$ by part 2. $\qquad\square$

Let us remark that by Proposition 2.3, for any initial condition $p \in \mathbb{P}(\mathbb{X})$, one has that $H(p_t \mid \mu)$ is monotone non-increasing, and therefore has a limit. This alone does not imply that p_t converges. In fact, at this level of generality it may be hard to give a complete characterization of the limit points of p_t as $t \to \infty$, as the initial state p varies in $\mathbb{P}(\mathbb{X})$. However, by compactness one has that along some subsequence $t_n \to \infty$, $p_{t_n} \to \rho$ for some $\rho \in \mathbb{P}(\mathbb{X})$. As already observed in [28], Theorem 2, if one knows that ρ is stationary and has full support, that is, $\rho \in \mathbb{P}_+(\mathbb{X})$, then it follows by Proposition 2.3 that $p_t \to \rho$, $t \to \infty$. To see this, observe that

$$H(p_t \mid \rho) = H(p_t \mid \mu) - p_t\big[\log(\rho/\mu)\big].$$

From Proposition 2.3 part 3, $p_t[\log(\rho/\mu)]$ is independent of t and therefore equals $H(\rho \mid \mu)$. On the other hand one must also have that $H(p_t \mid \mu)$ decreases to $H(\rho \mid \mu)$, and therefore $H(p_t \mid \rho) \to 0$, which implies $p_t \to \rho$ by (1.4).

For all the examples of RQS to be discussed below we shall not need to appeal to the above abstract argument. In fact, for any initial distribution $p \in \mathbb{P}(\mathbb{X})$ we shall always be able to identify an explicit limit point $\mu \in \mathbb{P}_+(\mathbb{X})$ such that $H(p_t \mid \mu) \to 0$, $t \to \infty$. This in turn implies the convergence $p_t \to \mu$ by (1.4). To quantify the convergence of relative entropy, we proceed as follows.

Definition 2.4. Given the RQS $(\mathbb{G}, \mu)$ on $\mathbb{X}$, we define $\mathbb{F} = \mathbb{F}(\mathbb{G}, \mu)$ as the set of functions $f : \mathbb{X} \mapsto [0, \infty)$ such that $\mu[f] = 1$, and

$$\mu\big[f \log(\rho/\mu)\big] = \mu\big[\log(\rho/\mu)\big],$$

for all stationary $\rho \in \mathbb{P}_+(\mathbb{X})$. Moreover, we say that the RQS satisfies the *entropy production estimate* with constant $\delta > 0$ if for all $f \in \mathbb{F}$ one has

$$D(f, f) \geq \delta \operatorname{Ent}(f), \tag{2.9}$$

where $\operatorname{Ent}(f) = \mu[f \log f] - \mu[f] \log \mu[f]$.

Proposition 2.5. *Suppose the RQS $(\mathbb{G}, \mu)$ satisfies the entropy production bound with constant $\delta > 0$. Then any initial state $p \in \mathbb{P}(\mathbb{X})$ such that $p/\mu \in \mathbb{F}$ satisfies*

$$H(p_t \mid \mu) \leq e^{-\delta t} H(p \mid \mu),$$

for all $t \geq 0$.

Proof. For any $f \geq 0$ such that $\mu[f] = 1$ one has $H(f\mu \mid \mu) = \operatorname{Ent}(f)$. Thus, by Proposition 2.3 part 1, it suffices to show that (2.9) holds for all functions $f_t = p_t/\mu$, $t \geq 0$. Since by assumption $p/\mu \in \mathbb{F}$ one has $p[\log(\rho/\mu)] = \mu[\log(\rho/\mu)]$, and by Proposition 2.3 part 3, the same holds for p_t, for all $t \geq 0$. In particular $f_t \in \mathbb{F}$, for all $t \geq 0$. $\qquad\square$

In practical applications, our approach may be summarized as follows. Given the initial $p \in \mathbb{P}(\mathbb{X})$, the goal will be to find a stationary measure $\mu \in \mathbb{P}_+(\mathbb{X})$ such that $p[\log(\rho/\mu)] =$

$\mu[\log(\rho/\mu)]$, for all stationary $\rho \in \mathbb{P}_+(\mathbb{X})$ and then to prove the entropy production bound of Definition 2.4 for the RQS with that choice of μ. Notice that in this framework the rate of decay δ may depend on the initial value $p \in \mathbb{P}(\mathbb{X})$ but only through the selected equilibrium point $\mu \in \mathbb{P}_+(\mathbb{X})$.

The functional inequality (2.9) can be interpreted as a nonlinear version of the logarithmic Sobolev inequality for Markov chains. In our setup, one can actually write it in the form of a classical logarithmic Sobolev inequality in the product space $\mathbb{X} \times \mathbb{X}$, as follows. For any $f, g : \mathbb{X} \mapsto [0, \infty)$, write $F(\sigma, \sigma') = f(\sigma) f(\sigma')$ and $G(\sigma, \sigma') = g(\sigma) g(\sigma')$. Then, as in the proof of Proposition 2.3 one has

$$D(f, g) = -\frac{1}{2}\widehat{\mu}\big[(\mathbb{G}F) \log G\big], \tag{2.10}$$

where we use the notation $\widehat{\mu} = \mu \otimes \mu$, and $\mathbb{G}F$ denotes the usual linear action of $\mathbb{G}$ on F. In particular, (2.9) now becomes

$$-\widehat{\mu}\big[(\mathbb{G}F) \log F\big] \geq \delta \operatorname{Ent}_{\widehat{\mu}}(F), \tag{2.11}$$

where $\operatorname{Ent}_{\widehat{\mu}}(F) = \widehat{\mu}[F \log F] - \widehat{\mu}[F] \log(\widehat{\mu}[F]) = 2 \operatorname{Ent}(f)$. The inequality (2.11) is often referred to in the Markov chain literature as a "modified log-Sobolev inequality"; see, for example, [3,10]. An important difference to keep in mind here with respect to the usual Markov chain setup is that we do not assume irreducibility, and therefore (2.11) in general cannot hold for all F. Indeed, in our setting we require this to hold only for F of the form $F(\sigma, \sigma') = f(\sigma) f(\sigma')$, where $f \in \mathbb{F}$.

2.3. Linearized problem and spectral gap

As in kinetic theory, see, for example, [30], to gain insight into the functional inequality (2.9) it is natural to investigate the linearized problem for near-to-equilibrium densities, that is, $f = 1 + \varepsilon \phi$ for some $\phi : \mathbb{X} \mapsto \mathbb{R}$ such that $\mu[\phi] = 0$ with small $\varepsilon > 0$.

Lemma 2.6. *Let $f = 1 + \varepsilon \phi$ for some $\phi : \mathbb{X} \mapsto \mathbb{R}$ such that $\mu[\phi] = 0$. Then, as $\varepsilon \to 0$ one has*

$$\lim_{\varepsilon \to 0} \varepsilon^{-2} \operatorname{Ent}(f) = \frac{1}{2}\mu\big[\phi^2\big], \qquad \lim_{\varepsilon \to 0} \varepsilon^{-2} D(f, f) = -\mu\big[(\Gamma \phi)\phi\big],$$

where Γ is the linear operator defined by

$$\Gamma(\tau, \sigma) = \sum_{\sigma', \tau'} \mu(\tau')\big[\mathbb{G}(\tau, \tau'; \sigma, \sigma') + \mathbb{G}(\tau, \tau'; \sigma', \sigma)\big].$$

Proof. By expanding in the parameter ε and retaining only terms up to order ε^2 it is not hard to check that $\varepsilon^{-2} \operatorname{Ent}(f) \to \frac{1}{2}\mu[\phi^2]$, as $\varepsilon \to 0$. Similarly, neglecting terms of order $o(\varepsilon^2)$, $D(f, f)$

is given by

$$-\varepsilon^2 \sum_{\sigma,\sigma',\tau,\tau'} \mu(\tau)\mu(\tau')\mathbb{G}(\tau,\tau';\sigma,\sigma')\phi(\tau)[\phi(\sigma)+\phi(\sigma')].$$

The latter expression equals $-\varepsilon^2\mu[(\Gamma\phi)\phi]$. □

Thus, the linearized version of inequality (2.9) states that

$$-\mu[(\Gamma\phi)\phi] \geq \frac{\delta}{2}\mu[\phi^2]. \qquad (2.12)$$

We note that if $\psi = \log(\rho/\mu)$ for some stationary $\rho \in \mathbb{P}_+(\mathbb{X})$, then, as in the proof of Proposition 2.3, one finds $\Gamma\psi = 0$. This is not in contradiction with (2.12). Indeed, the condition $f = 1 + \varepsilon\phi \in \mathbb{F}$ is equivalent to requiring that ϕ is orthogonal in $L^2(\mu)$ to the constant functions and to all conserved quantities $\psi = \log(\rho/\mu)$ for stationary $\rho \in \mathbb{P}_+(\mathbb{X})$. Thus it is meaningful to study (2.12) for all ϕ restricted to this class. This may be interpreted as a spectral gap bound, as follows.

Suppose that $\mathbb{G} = Q - \mathbf{1}$, where Q is a Markov kernel on $\mathbb{X} \times \mathbb{X}$. Then it is not hard to see that

$$\Gamma(\tau,\sigma) = 2K(\tau,\sigma) - \delta(\sigma,\tau) - \mu(\sigma), \qquad (2.13)$$

where we introduce the Markov kernel

$$K(\tau,\sigma) := \frac{1}{2}\sum_{\sigma',\tau'} \mu(\tau')[Q(\tau,\tau';\sigma,\sigma') + Q(\tau,\tau';\sigma',\sigma)]. \qquad (2.14)$$

The kernel K is reversible with respect to μ. Moreover, if $\psi = \log(\rho/\mu)$ for a stationary $\rho \in \mathbb{P}_+(\mathbb{X})$, then from $\Gamma\psi = 0$ and (2.13) we obtain the eigenvalue equation

$$K\bar{\psi} = \frac{1}{2}\bar{\psi},$$

where $\bar{\psi} := \psi - \mu[\psi]$. Inequality (2.12) is then equivalent to

$$\mu[(K\phi)\phi] \leq \left(\frac{1}{2} - \frac{\delta}{4}\right)\mu[\phi^2], \qquad (2.15)$$

for all ϕ orthogonal in $L^2(\mu)$ to the constant functions and to the conserved quantities $\psi = \log(\rho/\mu)$ as above. From the obvious inequality $-\mu[(\Gamma\phi)\phi] \geq 0$, one obtains for free that, apart from the trivial eigenvalue 1, all eigenvalues of K must be at most $\frac{1}{2}$. Thus, (2.15) says that, besides the eigenvalues associated to the conserved quantities, all other eigenvalues of K are at most $\frac{1}{2} - \frac{\delta}{4}$. In some special cases, one can fully diagonalize the operator K and compute the optimal constant δ in (2.15). It should be noted that, while the entropy production bound (2.9) always implies the spectral gap (2.15) via Lemma 2.6, the converse is of course not true.

2.4. Continuous vs. discrete time

If $\mathbb{G} = Q - 1$, where Q is a Markov kernel on $\mathbb{X} \times \mathbb{X}$, then one can define the discrete time RQS as follows. Set

$$\Psi[p](\tau) = \sum_{\sigma, \sigma', \tau'} p(\sigma) p(\sigma') Q(\sigma, \sigma', \tau, \tau').$$

This defines a map $\Psi : \mathbb{P}(\mathbb{X}) \mapsto \mathbb{P}(\mathbb{X})$, whose kth iterate $p^{(k)} = \Psi[p^{(k-1)}]$ describes the state of the system after k steps, with initial state $p^{(0)} = p$. The entropy production estimate (2.9) now takes the form

$$H(p^{(k)} \mid \mu) \le (1 - \delta) H(p^{(k-1)} \mid \mu), \tag{2.16}$$

for all $k \in \mathbb{N}$. Moreover, the same general remarks about stationary states and convergence to equilibrium apply in the discrete time setting; see also [28].

However, we caution the reader that, in contrast with the case of linear evolution associated to a Markov chain, here there is a more pronounced difference between the discrete time evolution and the continuous time evolution. To briefly address this point, let us regard the operation $\Psi[p]$ as the product $p \odot p$, where we define, for all $p, q \in \mathbb{P}(\mathbb{X})$ the new probability on $\mathbb{X} \times \mathbb{X}$

$$(p \odot q)(\tau, \tau') = \sum_{\sigma, \sigma'} p(\sigma) q(\sigma') Q(\sigma, \sigma', \tau, \tau').$$

Using the symmetry (2.7), one sees that the product $p \odot q$ is commutative. However, it is not in general associative. This is the main source of difficulty in the explicit construction of the continuous time evolution p_t, in contrast with the simple iterations $p^{(k)}$ of the discrete time process. The form $p \odot q$ is the analogue of the Wild convolution product in the Boltzmann equation literature. A solution of the continuous time system can be constructed using suitable sums over so-called "McKean trees", which encode the various ways of taking products, such as $p \odot (p \odot (p \odot p))$ or $(p \odot p) \odot (p \odot p)$ and so on. This construction builds on the pioneering work of Wild and McKean; see [6] and references therein. Our results below will not make use of this method; they will be based only on functional inequalities of the form (2.9) or (2.16), and will not make any significant distinction between discrete and continuous time evolution.

3. Main examples

We now turn to concrete examples of RQS.

3.1. Binary uniform crossover

We begin with the simplest possible example. Let $\mathbb{X} = \{0, 1\}^n$ for some fixed integer n, and

$$Q(\sigma, \sigma'; \tau, \tau') = \frac{1}{2^n} \sum_{A \subset [n]} \mathbf{1}(\tau = \sigma'_A \sigma_{A^c}, \tau' = \sigma_A \sigma'_{A^c}). \tag{3.1}$$

In words, we move from (σ, σ') to (τ, τ') under a uniform crossover, that is, (τ, τ') is obtained from (σ, σ') by picking a uniformly random $A \subset [n]$ and swapping the A-components σ_A and σ'_A, leaving the rest unchanged. Then $\mathbb{G} = Q - \mathbf{1}$ defines the generator of the RQS. For the reference measure μ, we may choose any product of Bernoulli probability measures. Indeed, (2.2) holds if $\mu = \bigotimes_{i=1} \mu_i$, with μ_i arbitrary Bernoulli distributions. Now, given any initial state $p \in \mathbb{P}(\mathbb{X})$, equation (2.4) can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t} p_t = \frac{1}{2^n} \sum_A (p_{t,A} \otimes p_{t,A^c} - p_t), \qquad t \geq 0,$$

with the initial condition $p_0 = p$. This is the model described in (1.2) in the special case where the single site spaces X_i all coincide with $\{0, 1\}$ and the distribution ν is uniform. The marginals p_i of p are preserved by the evolution, that is, $p_{t,i} = p_i$ for all i, t, and thus the natural candidate for convergence of p_t as $t \to \infty$ is the product measure $\pi = \bigotimes_{i=1} p_i$; see also Lemma 3.1 below. Our results imply that the entropy production bound (2.9) holds with $\delta \geq 1/n$, independent of the initial state $p \in \mathbb{P}(\mathbb{X})$. In particular,

$$H(p_t \mid \pi) \leq H(p \mid \pi) e^{-t/n}$$

for any initial distribution $p \in \mathbb{P}(\mathbb{X})$ and any n. As we shall see in the proof of Corollary 1.3 the $1/n$ behavior of the constant cannot be improved if we require uniformity of the decay rate as a function of the initial state p. In this example, the linear operator K from (2.14) can be fully diagonalized, and one finds that (2.15) holds with $\delta = 1$. In particular, this is an example of a system with uniformly positive spectral gap but with a vanishing rate (as $n \to \infty$) in the exponential decay of relative entropy.

3.2. General recombination model

The binary uniform crossover model can be readily generalized to the case where the uniform choice of A in (3.1) is replaced by a given distribution ν on subsets of $[n]$, and the state space is taken as $\mathbb{X} = X_1 \times \cdots \times X_n$ with arbitrary finite single site spaces X_i. Again $\mathbb{G} = Q - \mathbf{1}$, where now

$$Q(\sigma, \sigma; \tau, \tau') = \sum_{A \subset [n]} \nu(A) \mathbf{1}(\tau = \sigma'_A \sigma_{A^c}, \tau' = \sigma_A \sigma'_{A^c}).$$

The system equation (2.4) then coincides with (1.2). As above it is not hard to check that (2.2) is satisfied by any product measure $\mu = \bigotimes \mu_i$, with μ_i an arbitrary probability measure on X_i. To define the RQS $(\mathbb{G}, \mu)$, we fix one such μ, with $\mu \in \mathbb{P}_+(\mathbb{X})$. Say that ν is *nondegenerate* if for any $i, j \in [n]$ there is a positive probability that the random set with distribution ν separates i and j.

Lemma 3.1. *If ν is nondegenerate, then $\rho \in \mathbb{P}(\mathbb{X})$ is stationary iff ρ has the product form $\rho = \bigotimes_{i=1}^n \rho_i$ for some probability measures ρ_i on X_i.*

Proof. From Proposition 2.3, it follows that $\rho \in \mathbb{P}(\mathbb{X})$ is stationary iff it satisfies

$$\rho(\sigma)\rho(\sigma') = \rho(\sigma'_A \sigma_{A^c})\rho(\sigma_A \sigma'_{A^c}),$$

for all $\sigma, \sigma' \in \mathbb{X}$ and all $A \subset [n]$ such that $\nu(A) > 0$. In particular, any product measure $\rho = \bigotimes_{i=1}^n \rho_i$ is stationary. To prove the converse, notice that summing over $\sigma' \in \mathbb{X}$ in the above equation one has that

$$\rho(\sigma) = \rho_A(\sigma_A)\rho_{A^c}(\sigma_{A^c}), \tag{3.2}$$

for all $\sigma \in \mathbb{X}$ and all $A \subset [n]$ such that $\nu(A) > 0$, where ρ_A denotes the marginal of ρ on A. We prove by induction that for any set $B \subset [n]$ one has $\rho_B(\sigma_B) = \prod_{i \in B} \rho_i(\sigma_i)$. Plainly, this is true for all sets $B \subset [n]$ with $|B| = 1$. Suppose that it is true for all sets $B \subset [n]$ with $1 \leq |B| \leq k$. Take $B' \subset [n]$ with $|B'| = k+1$, and choose $i, j \in B'$. By the nondegeneracy assumption, there is a set A such that $i \in A$, $j \in A^c$, and $\nu(A) > 0$. Applying (3.2) with this choice of A and taking the marginal over B', one finds $\rho_{B'}(\sigma_{B'}) = \rho_{A \cap B'}(\sigma_{A \cap B'})\rho_{A^c \cap B'}(\sigma_{A^c \cap B'})$. Since $1 \leq |A^c \cap B'| \leq k$ and $1 \leq |A \cap B'| \leq k$ we may apply the inductive assumption to conclude. $\qquad\square$

Clearly, all models discussed in the introduction satisfy the nondegeneracy assumption. Since the marginals p_i of p are preserved by the evolution, the natural candidate for convergence of p_t is the product measure $\pi = \bigotimes_{i=1}^n p_i$. As highlighted in Corollary 1.3, our analysis will show that, for any initial state $p \in \mathbb{P}(\mathbb{X})$,

$$H(p_t \mid \pi) \leq H(p \mid \pi)e^{-\kappa(\nu)t}$$

for all $t \geq 0$, where the constant $\kappa(\nu)$ is as specified in Theorem 1.2.

We remark that we may always pretend that π has full support, and take π itself as the reference measure μ. Indeed, this is equivalent to the condition that p_i has full support on X_i for all $i \in [n]$, and if that is not the case then we may simply replace the X_i with subspaces $\bar{X}_i$ such that p_i has full support on $\bar{X}_i$ and work with the RQS $(\mathbb{G}, \pi)$ within the space $\bar{\mathbb{X}} = \bar{X}_1 \times \cdots \times \bar{X}_n$ instead of $\mathbb{X}$; this ensures that now $\pi \in \mathbb{P}_+(\bar{\mathbb{X}})$.

In the remainder of this section, we briefly consider some more general examples of RQS that go beyond our recombination examples. In particular, these generalized models will admit stationary measures that are not product measures. While our quantitative results on entropy production do not so far extend to these cases, they serve as examples of natural open questions in this area.

3.3. Nonlinear stochastic Ising model

As a canonical example of a generalization that admits nontrivial correlations in the equilibrium state, we introduce a natural nonlinear dynamics on the Ising model on a finite graph $G = (V, E)$, $V = [n]$. The Ising model is the probability measure $\mu = \mu_{G,\beta,\mathbf{h}}$ on $\mathbb{X} = \{-1, 1\}^V$, given by

$$\mu(\sigma) = \frac{1}{Z}e^{-\beta H(\sigma) + \sum_{i \in V} h_i \sigma_i}, \qquad H(\sigma) = -\sum_{ij \in E} \sigma_i \sigma_j. \tag{3.3}$$

Here $\mathbf{h} = \{h_i\}_{i \in V} \in [-\infty, \infty]^V$ are the so-called "external fields", $\beta \in \mathbb{R}$ is a parameter (the "inverse temperature"), and $Z = Z_{G,\beta,\mathbf{h}}$ is the normalizing factor, or "partition function." Infinite values of external fields encode so-called "boundary conditions": if $h_i = +\infty$ (respectively, $-\infty$) then one has a $+$ (respectively, $-$) boundary condition at site i. Clearly, $\mu \in \mathbb{P}_+(\mathbb{X})$ iff all external fields are finite. Below, we fix one such μ as reference measure. We set $\mathbb{G} = Q - \mathbf{1}$, with the Markov kernel Q defined by $Q = \sum_A \nu(A) Q_A$, where ν is a given distribution over subsets and for any $A \subset [n]$,

$$
\begin{aligned}
Q_A(\sigma, \sigma'; \tau, \tau') &= \alpha_A(\sigma, \sigma') \mathbf{1}(\tau = \sigma'_A \sigma_{A^c}, \tau' = \sigma_A \sigma'_{A^c}) \\
&\quad + (1 - \alpha_A(\sigma, \sigma')) \mathbf{1}(\tau = \sigma, \tau' = \sigma'), \\
\alpha_A(\sigma, \sigma') &:= \frac{\mu(\sigma'_A \sigma_{A^c}) \mu(\sigma_A \sigma'_{A^c})}{\mu(\sigma'_A \sigma_{A^c}) \mu(\sigma_A \sigma'_{A^c}) + \mu(\sigma) \mu(\sigma')}.
\end{aligned} \tag{3.4}
$$

Notice that $\alpha_A(\sigma, \sigma')$ can be written as the conditional $\mu \otimes \mu$ probability of the pair $(\sigma'_A \sigma_{A^c}, \sigma_A \sigma'_{A^c})$ given the occurrence of either $(\sigma'_A \sigma_{A^c}, \sigma_A \sigma'_{A^c})$ or (σ, σ'). It satisfies

$$
\alpha_A(\sigma, \sigma') = \frac{1}{1 + e^{\beta \phi_A(\sigma, \sigma')}}, \qquad \phi_A(\sigma, \sigma') = \sum_{i \in A, j \in A^c} (\sigma_i - \sigma'_i)(\sigma_j - \sigma'_j) \mathbf{1}(ij \in E).
$$

In particular, α_A is independent of the external fields $\mathbf{h}$. It is easily checked that for each A, one has the reversibility

$$
\mu(\tau) \mu(\tau') Q_A(\tau, \tau'; \sigma, \sigma') = \mu(\sigma) \mu(\sigma') Q_A(\sigma, \sigma'; \tau, \tau'), \tag{3.5}
$$

for all $\sigma, \sigma', \tau, \tau' \in \mathbb{X}$. Therefore $(\mathbb{G}, \mu)$ defines a RQS. Since the kernel Q is independent of the external fields $\mathbf{h}$, the RQS here is determined by the distribution ν, the parameter β and the graph G (and not by $\mathbf{h}$). Moreover, any Ising measure of the form (3.3) satisfies (3.5) and therefore it is stationary for the RQS. A particularly interesting choice is the single site update $\nu(A) = \frac{1}{n} \mathbf{1}(|A| = 1)$, which can be interpreted as a nonlinear version of the usual Ising Gibbs sampler, or Glauber dynamics; see, for example, [21] for an introduction. One can prove the following characterization of the stationary distributions.

Lemma 3.2. *Fix a graph G with n vertices, and $\beta \in \mathbb{R}$. Assume $\nu(A) = \frac{1}{n} \mathbf{1}(|A| = 1)$. Let $\mu \in \mathbb{P}_+(\mathbb{X})$ be as in (3.3) with arbitrary external fields. A distribution $\rho \in \mathbb{P}(\mathbb{X})$ is stationary for the RQS $(\mathbb{G}, \mu)$ if and only if ρ is of the form (3.3) for some choice of $\mathbf{h}$.*

Proof. We have seen that any ρ of the form (3.3) satisfies (3.5) and it is therefore stationary. To prove the converse, observe that by Proposition 2.3 part 2, one has that any stationary ρ must satisfy

$$
\frac{\rho(\sigma) \rho(\sigma')}{\mu(\sigma) \mu(\sigma')} = \frac{\rho(\sigma'_{\{i\}} \sigma_{[n] \setminus \{i\}}) \rho(\sigma_{\{i\}} \sigma'_{[n] \setminus \{i\}})}{\mu(\sigma'_{\{i\}} \sigma_{[n] \setminus \{i\}}) \mu(\sigma_{\{i\}} \sigma'_{[n] \setminus \{i\}})}, \tag{3.6}
$$

for all sites $i \in [n]$, and for all $\sigma, \sigma' \in \mathbb{X}$. Suppose first that $\rho \in \mathbb{P}_+(\mathbb{X})$. We will show that, if $f(\sigma) = \rho(\sigma)/\mu(\sigma)$, then there exist $\varphi_1, \ldots, \varphi_n$ with $\varphi_j : \{-1, +1\} \mapsto (0, \infty)$ such that

$$f(\sigma) = f(+) \prod_{j=1}^{n} \varphi_j(\sigma_j), \tag{3.7}$$

where $+$ denotes the configuration with all spins equal to $+1$. To prove (3.7), for any $\sigma \in \mathbb{X}$, $j \in [n]$, let $\sigma[j] \in \mathbb{X}$ denote the configuration equal to $+1$ for all sites in $[j] = \{1, \ldots, j\}$ and equal to σ on $\{j + 1, \ldots, n\}$. Since $\rho, \mu \in \mathbb{P}_+(\mathbb{X})$, if $\sigma[0] = \sigma$, one has

$$f(\sigma) = f(+) \prod_{j=1}^{n} \frac{f(\sigma[j-1])}{f(\sigma[j])}.$$

Taking $i = j$, $\sigma = \sigma[j-1]$ and $\sigma' = +$ in (3.6), and setting $\varphi_j(\sigma_j) = \frac{f(+^{j,\sigma_j})}{f(+)}$, where $+^{j,\sigma_j}$ denotes the configuration equal to σ_j at j and equal to $+1$ elsewhere, one has

$$\frac{f(\sigma[j-1])}{f(\sigma[j])} = \varphi_j(\sigma_j), \qquad j = 1, \ldots, n.$$

This proves (3.7). Since φ_j is a positive function of a single spin, it can be written as $\varphi_j(\sigma_j) = e^{h_j \sigma_j + c_j}$ for some real numbers h_j, c_j. Therefore, from (3.7) we obtain $\rho(\sigma) = \text{const} \times \mu(\sigma) e^{\sum_{j \in V} \sigma_j h_j}$. This ends the proof for $\rho \in \mathbb{P}_+(\mathbb{X})$.

If ρ does not have full support, then there must exist $A \subset [n]$ and a vector $v_A = (v_i)_{i \in A} \in \{-1, +1\}^A$ such that for all $\tau \in \mathbb{X}$,

$$\rho(\tau) = \mathbf{1}(\tau_A = v_A) \rho'(\tau_{A^c}), \tag{3.8}$$

for some probability ρ' on $\{-1, +1\}^{A^c}$ with full support. (Here A is the set of vertices with boundary conditions in ρ.) This can be seen as follows. It is not hard to check that if $\rho \in \mathbb{P}(\mathbb{X}) \setminus \mathbb{P}_+(\mathbb{X})$, then there exists $\sigma \in \mathbb{X}$, $i \in [n]$ such that $\rho(\sigma) > 0$ and $\rho(\sigma^i) = 0$. Then, taking σ' equal to $-\sigma_i$ at i and arbitrary otherwise, from (3.6) one sees that $\rho(\tau) = 0$ for all $\tau \in \mathbb{X}$ such that $\tau_i = -\sigma_i$. Now, if $\rho(\tau') > 0$ for all $\tau' \in \mathbb{X}$ such that $\tau'_i = \sigma_i$, then (3.8) holds with $A = \{i\}$ and $v_i = \sigma_i$. Otherwise, restricting attention to configurations with the ith spin equal to σ_i, one can repeat the above reasoning, and the desired conclusion follows by recursion.

Once (3.8) is established, one can repeat the argument given in the case of $\rho \in \mathbb{P}_+(\mathbb{X})$, by restricting to the subspace of $\sigma \in \mathbb{X}$ such that $\sigma_A = v_A$. On this set the measure ρ has full support and one obtains again (3.7), this time with $f(+)$ replaced by $f(\xi)$ where ξ equals v_A on A and $+$ on A^c, with $\sigma[j]$ interpolating from σ to ξ, and with the product restricted to $j \in A^c$. It follows that ρ has the form (3.3), with $h_i \in \mathbb{R}$ for $i \in A^c$, $h_i = +\infty$ for $i \in A$ such that $v_i = +1$, and $h_i = -\infty$ for $i \in A$ such that $v_i = -1$. $\qquad \square$

We remark that Lemma 3.2 assumes that $\mu \in \mathbb{P}_+(\mathbb{X})$, i.e. that the corresponding external fields are all finite. However, in order to analyze the Ising model with boundary conditions one may

wish to take some of the external fields to have positive or negative infinite values. In this case, one can define the RQS $(\mathbb{G}, \mu)$ as above with the difference that now the state space $\mathbb{X}$ is replaced by the restricted space of $\sigma \in \mathbb{X}$ that are aligned with the boundary conditions at those sites which have infinite external field, so that μ has full support when restricted to this set. Then it is not hard to check that Lemma 3.2 continues to hold, with the same proof.

Since marginals are conserved by the evolution, for any $p \in \mathbb{P}(\mathbb{X})$ we expect that $p_t \to \mu$, where $\mu = \mu(p)$ is the equilibrium measure (3.3) with external fields $\mathbf{h}$ such that the marginals μ_i coincide with the marginals p_i. It is standard to check that such a choice always exists. Equivalently, the measure $\mu = \mu(p)$ is characterized by the condition

$$\mu\big[\log(\rho/\mu)\big] = p\big[\log(\rho/\mu)\big],$$

for all stationary $\rho \in \mathbb{P}_+(\mathbb{X})$, since by Lemma 3.2 one has that $\log(\rho/\mu) = \sum_i a_i \sigma_i$ for arbitrary coefficients a_i. Recall that this is the condition appearing in Definition 2.4.

In analogy with results on entropy decay for the Glauber dynamics (see, e.g., [4] and references therein), we propose the following conjecture concerning convergence to equilibrium.

Conjecture 3.3. *Assume $\nu(A) = \frac{1}{n}\mathbf{1}(|A| = 1)$. There exists a constant $c > 0$ such that for any graph G with n vertices and maximum degree Δ, for any $\beta \in \mathbb{R}$ with $|\beta| \le c/\Delta$, and for any choice of external fields $\mathbf{h}$, the RQS $(\mathbb{G}, \mu)$ with $\mu = \mu_{G,\beta,\mathbf{h}}$ satisfies the entropy production estimate of Definition 2.4 with constant $\delta = c/n$.*

Note that if $\beta = 0$, then $\alpha_A \equiv \frac{1}{2}$, then we are back to the single site recombination model, for which Conjecture 3.3 holds by Corollary 1.3. We turn now to the description of some possible variants of the nonlinear stochastic Ising model.

3.3.1. *Foldings*

A further stochastic Ising model is obtained as follows. Given a pair $(\sigma, \sigma') \in \mathbb{X} \times \mathbb{X}$, let $B = B(\sigma, \sigma')$ denote the set of vertices where they agree: $B = \{i \in [n] : \sigma_i = \sigma'_i\}$. Then define the kernel

$$Q\big(\sigma, \sigma'; \tau, \tau'\big) = \frac{1}{\mathcal{Z}(\sigma, \sigma')}\mu(\sigma_B \tau_{B^c})\mu\big(\sigma_B \tau'_{B^c}\big)\mathbf{1}\big(\tau'_{B^c} = -\tau_{B^c}\big), \tag{3.9}$$

where $\mathcal{Z}(\sigma, \sigma')$ is the normalizing constant

$$\mathcal{Z}\big(\sigma, \sigma'\big) = \sum_{\tau \in \{-1, +1\}^{B^c}} \mu(\sigma_B \tau_{B^c})\mu\big(\sigma_B(-\tau_{B^c})\big).$$

It is not hard to check that $\mathbb{G} = Q - \mathbf{1}$ defines a RQS with the required properties for any measure $\mu = \mu_{G,\beta,\mathbf{h}} \in \mathbb{P}_+(\mathbb{X})$ as in (3.3). Again Q does not depend on the external field $\mathbf{h}$ and all choices of $\mathbf{h}$ produce a valid stationary state. In the case $\beta = 0$, one has that (3.9) coincides with the binary uniform crossover example (3.1) above. We expect that for small β an estimate as in Conjecture 3.3 should hold for this model as well. The kernel Q in (3.9) is an example of a "folding" transformation in the terminology introduced in [29].

3.3.2. *Adding a dissipative term*

The previous models are conservative in the sense that single site marginals are constant in time. One can obtain a non-conservative evolution by adding a dissipative term as follows. The added terms can be interpreted as mutation operators in the context of genetic algorithms; see, for example, [17].

Fix a graph G with vertex set $V = [n]$, $\beta \in \mathbb{R}$ and a set of external fields $\mathbf{h}$, and let $\mu = \mu_{G,\beta,\mathbf{h}}$ be the associated Ising Gibbs measure. Suppose that $W(\sigma; \tau)$ is the usual Glauber dynamics kernel for μ, that is,

$$W(\sigma; \tau) = \frac{1}{n} \sum_{i=1}^{n} \left[\mu\left(\sigma^i \mid \{\sigma\} \cup \{\sigma^i\}\right) \mathbf{1}\left(\tau = \sigma^i\right) + \mu\left(\sigma \mid \{\sigma\} \cup \{\sigma^i\}\right) \mathbf{1}\left(\tau = \sigma\right) \right],$$

where again, for any $\sigma \in \mathbb{X}$, $\sigma^i \in \mathbb{X}$ is obtained from σ by reversing the spin at i. Define

$$\widetilde{\mathbb{G}}\left(\sigma, \sigma'; \tau, \tau'\right) = \left(W(\sigma; \tau) - \mathbf{1}(\sigma = \tau)\right) \mathbf{1}\left(\sigma' = \tau'\right) + \left(W\left(\sigma'; \tau'\right) - \mathbf{1}\left(\sigma' = \tau'\right)\right) \mathbf{1}(\sigma = \tau).$$

As in Remark 2.1, the RQS $(\widetilde{\mathbb{G}}, \mu)$ defines a linear evolution, namely the Glauber dynamics. If $\mathbb{G}$ is the generator of the RQS $(\mathbb{G}, \mu)$ introduced in (3.4), then one can define a new nonlinear RQS $(\mathbb{G}', \mu)$ with generator $\mathbb{G}' = \mathbb{G} + \widetilde{\mathbb{G}}$. It is not hard to check that $(\mathbb{G}', \mu)$ satisfies the required properties. Moreover, in this case one has the following quantitative convergence results.

Theorem 3.4. *Fix the graph G with n vertices, $\beta \in \mathbb{R}$, and a set of external fields $\mathbf{h}$. Let p_t denote the evolution according to the RQS $(\mathbb{G}', \mu)$ defined above with $\mu = \mu_{G,\beta,\mathbf{h}}$.*

(1) *There exists $c(\beta, G) > 0$ independent of $\mathbf{h}$ such that for all $p \in \mathbb{P}(\mathbb{X})$, and all $t \geq 0$:*

$$H(p_t \mid \mu) \leq e^{-c(\beta,G)t} H(p \mid \mu).$$

(2) *There exists a constant $c > 0$ independent of G, β and $\mathbf{h}$ such that for all $\beta \in \mathbb{R}$ with $|\beta| \leq c/\Delta$, where Δ is the maximal degree of G, and for all $p \in \mathbb{P}(\mathbb{X})$, and all $t \geq 0$:*

$$H(p_t \mid \mu) \leq e^{-ct/n} H(p \mid \mu).$$

The main difference between the two estimates above is that the first is valid for any β and involves a constant $c(\beta, G)$ that may be exponentially small as a function of the graph G (see the proof below for an explicit expression), while the second is a bound with decay rate of order $1/n$ that is valid only at sufficiently high temperature. In any case, Theorem 3.4 shows that in contrast with the conservative case, all initial states $p \in \mathbb{P}(\mathbb{X})$ converge to the same equilibrium point μ. Indeed, in this case the kernel W, which depends on the external fields, drives the system towards the equilibrium μ.

Proof of Theorem 3.4. For the first estimate, it is sufficient to prove that (2.9) holds for all $f : \mathbb{X} \mapsto [0, \infty)$ with $\delta = c(\beta, G) > 0$. Notice that we do not restrict to any particular class of functions here. With the notation of (2.10), we have

$$D(f, f) = -\frac{1}{2}\widehat{\mu}\left[(\mathbb{G}F + \widetilde{\mathbb{G}}F) \log F\right] \geq -\frac{1}{2}\widehat{\mu}\left[(\widetilde{\mathbb{G}}F) \log F\right],$$

where we use the fact that $-\widehat{\mu}[(\mathbb{G}F)\log F] \geq 0$, which follows from the representation in Proposition 2.3, part 1. Since the operator $\widetilde{\mathbb{G}}$ corresponds to two independent Glauber dynamics on $\mathbb{X} \times \mathbb{X}$, from Proposition 1.1 and Lemma 2.2 in [4] one has the modified logarithmic Sobolev inequality

$$-\widehat{\mu}\big[(\widetilde{\mathbb{G}}F)\log F\big] \geq c(\beta, G)\widehat{\mu}[F\log F],$$

where $c(\beta, G) = n^{-1}e^{-6|\beta||E|}$ and $|E|$ is the number of edges in G. This proves part 1 since $\widehat{\mu}[F\log F] = 2\operatorname{Ent}(f)$.

To prove part 2, we use the same argument, but now observe that for some absolute constant $c > 0$, for $|\beta| \leq c/\Delta$, by Corollary 2.3 in [4] one has

$$-\widehat{\mu}\big[(\widetilde{\mathbb{G}}F)\log F\big] \geq \frac{c}{n}\widehat{\mu}[F\log F]. \qquad \square$$

3.4. Further examples

Many more examples of RQS can be constructed by using suitable Markov chain generators in the product space $\mathbb{X} \times \mathbb{X}$. This allows one to construct natural and possibly useful nonlinear versions of various familiar stochastic processes such as random walks or card shuffling. We refer to [28] for an interesting application to matchings in graphs.

Our discussion in this paper is limited to the quadratic case where two independent samples from a given population interact to produce a new population. However, it is not difficult to generalize the setting by considering more than just two samples from the starting population. One may then obtain higher order nonlinear equations. For instance, a cubic version of the recombination process (1.2) would take the form

$$\frac{\mathrm{d}}{\mathrm{d}t}p_t = \sum_{A,B,C} \nu(A, B, C)(p_{t,A} \otimes p_{t,B} \otimes p_{t,C} - p_t), \qquad t \geq 0,$$

where A, B, C form a partition of $[n]$, and ν is a probability over such partitions. Generalized recombination models of this kind have been recently considered in [1]. Further generalizations (which are not necessarily even mass-preserving) can be found in the field of "mass action kinetics" introduced in [14,19], which remains a very active area today.

4. Entropy production estimate for recombinations

In this section, we prove Theorem 1.1, Theorem 1.2 and Corollary 1.3. Most of the work goes into proving the lower bounds on the constant $\delta(\nu)$ in Theorem 1.1. The first step consists in reducing this problem to the more tractable problem of controlling the constant $\kappa(\nu)$ in Theorem 1.2.

Consider the RQS $(\mathbb{G}, \mu)$ defined by $\mathbb{G} = Q - 1$ where

$$Q\big(\sigma, \sigma; \tau, \tau'\big) = \sum_A \nu(A)\mathbf{1}\big(\tau = \sigma'_A\sigma_{A^c}, \tau' = \sigma_A\sigma'_{A^c}\big),$$

for some distribution ν on subsets of $[n]$, and $\mu = \bigotimes_{i=1}^{n} \mu_i$ an arbitrary product measure on the product space $\mathbb{X} = X_1 \times \cdots \times X_n$, where X_i are given finite spaces. Assume that $\mu(\sigma) > 0$ for all $\sigma \in \mathbb{X}$.

For any $A \subset [n]$, let $\mathbb{X}_A = \prod_{i \in A} X_i$, and for any $f : \mathbb{X} \mapsto [0, \infty)$, write f_A for the function $f_A : \mathbb{X}_A \mapsto [0, \infty)$ defined by

$$f_A(\sigma_A) = \sum_{\sigma' \in \mathbb{X}} \mu(\sigma') f(\sigma_A \sigma'_{A^c}). \tag{4.1}$$

Notice that, if f is a density with respect to μ (i.e., $\mu[f] = 1$) then f_A is the density of the marginal of $f\mu$ on $\mathbb{X}_A$. It will often be convenient to regard f_A as a function on the whole space $\mathbb{X}$ simply by setting $f_A(\sigma) = f_A(\sigma_A)$. If $p \in \mathbb{P}(\mathbb{X})$, we write p_A for the marginal on $\mathbb{X}_A$. Thus, if $f = p/\mu$, then $f_A f_{A^c}$ denotes the density of $p_A \otimes p_{A^c}$ with respect to μ. When $A = \varnothing$, we set $f_\varnothing = 1$. When $A = \{i\}$ is a singleton, we simply write $f_i = f_{\{i\}}$ for the single site marginal.

Lemma 4.1. *Fix an arbitrary distribution ν on subsets of $[n]$. Let $\mathbb{S}_\mu$ denote the set of $f : \mathbb{X} \mapsto [0, \infty)$ such that $\mu[f] = 1$ and $f_i = 1$ for all $i = 1, \dots, n$. Suppose that there exists $\delta > 0$ such that*

$$\sum_A \nu(A) \mu[f_A f_{A^c} \log f] \leq (1 - \delta) \operatorname{Ent}(f) \tag{4.2}$$

for all $f \in \mathbb{S}_\mu$. Then the RQS $(\mathbb{G}, \mu)$ satisfies the entropy production estimate with constant δ, as defined in Definition 2.4.

Proof. Rewrite the functional $D(f, f)$ as

$$D(f, f) = -\sum_A \nu(A) \sum_{\tau, \tau' \in \mathbb{X}} \mu(\tau) \mu(\tau') \big(f(\tau'_A \tau_{A^c}) f(\tau_A \tau'_{A^c}) - f(\tau) f(\tau') \big) \log f(\tau).$$

With the notation f_A for the marginal densities, and using the product structure of μ, the above expression becomes

$$D(f, f) = \operatorname{Ent}(f) - \sum_A \nu(A) \mu[f_A f_{A^c} \log f]. \tag{4.3}$$

To conclude the proof it remains to show that $\mathbb{F} \subset \mathbb{S}_\mu$, where $\mathbb{F}$ is the set of functions from Definition 2.4. Indeed, suppose that $f \in \mathbb{F}$. Then for any stationary $\rho \in \mathbb{P}_+(\mathbb{X})$ one has $\mu[f \log(\rho/\mu)] = \mu[\log(\rho/\mu)]$. Let us show that $f_i = 1$ for all $i \in [n]$. Take $\rho \in \mathbb{P}_+(\mathbb{X})$ of the form $\rho = \bigotimes_{i=1}^{n} \rho_i$ for some probability measures ρ_i on X_i. Then ρ is stationary. Choosing $\rho_j = \mu_j$ for all $j \neq i$, one has

$$\mu\big[f_i \log(\rho_i/\mu_i)\big] = \mu\big[\log(\rho_i/\mu_i)\big].$$

Since the ρ_i's are arbitrary, it follows that $f_i = 1$ for all i. $\qquad\square$

We remark that since the distribution ν in Lemma 4.1 is arbitrary the inclusion $\mathbb{F} \subset \mathbb{S}_\mu$ used in the previous proof may be strict. However, by Lemma 3.1, if ν is nondegenerate then one has that $\mathbb{F} = \mathbb{S}_\mu$ since the only stationary measures are of product form.

Notice that (4.2) is equivalent to

$$\sum_A \nu(A)\mu\left[(f_A f_{A^c} - f)\log\frac{f_A f_{A^c}}{f}\right] \geq \delta \operatorname{Ent}(f). \tag{4.4}$$

This follows from the fact that for any A one has

$$\mu\left[(f_A f_{A^c} - f)\log(f_A f_{A^c})\right] = 0.$$

Thus, the largest constant δ such that (4.2) holds for all $f \in \mathbb{S}_\mu$ is precisely the constant $\delta(\pi, \nu)$ appearing in inequality (1.6) (with π replaced by μ).

The following observation allows us to reduce (4.4) to a more tractable expression. Suppose $f = p/\mu$ for some $p \in \mathbb{P}(\mathbb{X})$. Then

$$\mu[f_A f_{A^c} \log f] = -(p_A \otimes p_{A^c})\left[\log\left(\frac{p_A \otimes p_{A^c}}{p}\right)\right] + (p_A \otimes p_{A^c})\left[\log\left(\frac{p_A \otimes p_{A^c}}{\mu}\right)\right].$$

Rearranging and using $\mu = \mu_A \otimes \mu_{A^c}$ one has the identity

$$\mu[f_A f_{A^c} \log f] = -H(p_A \otimes p_{A^c} \mid p) + H(p_A \mid \mu_A) + H(p_{A^c} \mid \mu_{A^c}).$$

In particular,

$$\mu[f_A f_{A^c} \log f] \leq H(p_A \mid \mu_A) + H(p_{A^c} \mid \mu_{A^c})$$
$$= \mu[f_A \log f_A] + \mu[f_{A^c} \log f_{A^c}] = \operatorname{Ent}(f_A) + \operatorname{Ent}(f_{A^c}).$$

Since the above holds for arbitrary product measures μ, with the notation of Lemma 4.1 we have obtained the following criterion.

Lemma 4.2. *Suppose that, for all $f \in \mathbb{S}_\mu$,*

$$\sum_A \nu(A)\big(\operatorname{Ent}(f_A) + \operatorname{Ent}(f_{A^c})\big) \leq (1 - \kappa)\operatorname{Ent}(f), \tag{4.5}$$

with some $\kappa > 0$. Then (4.2) holds with constant $\delta = \kappa$. In particular, if $\delta(\nu)$ and $\kappa(\nu)$ are the constants introduced in Theorem 1.1 and Theorem 1.2, then for any distribution ν,

$$\delta(\nu) \geq \kappa(\nu).$$

We turn now to the analysis of the constant $\kappa(\nu)$. We start by recalling some preliminary inequalities.

4.1. Tensorization and Shearer-type inequalities

We recall that for any probability measure μ, for any $A \subset [n]$, one has the decomposition

$$\text{Ent}(f) = \text{Ent}\big(\mu[f \mid A]\big) + \mu\big[\text{Ent}(f \mid A)\big], \tag{4.6}$$

where $\text{Ent}(f \mid A) = \mu[f \log(f/\mu[f \mid A]) \mid A]$ denotes the entropy of f with respect to the conditional probability measure $\mu[\cdot \mid A]$, obtained by conditioning μ on a given realization of the variables $\sigma_A \in \mathbb{X}_A$. The decomposition (4.6) is obtained by adding and subtracting $\mu[f \log \mu[f \mid A]]$ from $\text{Ent}(f)$. We note that if μ is a product measure, then $\mu[f \mid A]$ coincides with f_A defined in (4.1). Let $\mathbb{A}$ be an arbitrary family of subsets $A \subset [n]$. Let $\deg_k(\mathbb{A})$ denote the degree of a site k in $\mathbb{A}$, i.e., the number of subsets $A \in \mathbb{A}$ such that $A \ni k$, and set

$$n_-(\mathbb{A}) = \min\big\{\deg_k(\mathbb{A}), k \in [n]\big\}, \qquad n_+(\mathbb{A}) = \max\big\{\deg_k(\mathbb{A}), k \in [n]\big\},$$

for the minimal and maximal degrees, respectively.

Proposition 4.3. *Let μ be a product measure. For any family of subsets $\mathbb{A}$ and any function $f \geq 0$,*

$$n_-(\mathbb{A}) \text{Ent}(f) \leq \sum_{A \in \mathbb{A}} \mu\big[\text{Ent}\big(f \mid A^c\big)\big]. \tag{4.7}$$

Equivalently, for any $\mathbb{A}$ and any function $f \geq 0$,

$$\sum_{A \in \mathbb{A}} \text{Ent}(f_A) \leq n_+(\mathbb{A}) \text{Ent}(f). \tag{4.8}$$

Proof. The equivalence of (4.7) and (4.8) follows from (4.6) by passing from $\mathbb{A}$ to the complementary set of subsets $\bar{\mathbb{A}} = \{A^c, A \in \mathbb{A}\}$. We prove (4.8) as a consequence of the classical Shearer estimate for Shannon entropy [9]. Suppose first that $\mathbb{A}$ is a *regular cover* of $[n]$, i.e., the union of $A \in \mathbb{A}$ is $[n]$ and the degrees $n(\mathbb{A}) := \deg_k(\mathbb{A})$ are independent of $k \in [n]$. By homogeneity, we may assume $\mu[f] = 1$. Call $Z = (Z_1, \ldots, Z_n)$ the random variable with probability distribution $f\mu$, so that $f_A \mu_A$ is the law of the marginal $Z_A = (Z_i, i \in A)$; see (4.1). For any A, the Shannon entropy $H(Z_A)$ of Z_A satisfies

$$\begin{aligned}
H(Z_A) &= -\sum_{\sigma_A} f_A(\sigma_A)\mu_A(\sigma_A) \log\big(f_A(\sigma_A)\mu_A(\sigma_A)\big) \\
&= -\text{Ent}(f_A) - \sum_{\sigma_A} \sum_{i \in A} f_A(\sigma_A)\mu_A(\sigma_A) \log\big(\mu_i(\sigma_i)\big) \\
&= -\text{Ent}(f_A) - \sum_{i \in A} \sum_{\sigma_i} f_i(\sigma_i)\mu_i(\sigma_i) \log\big(\mu_i(\sigma_i)\big) \\
&= -\text{Ent}(f_A) + \sum_{i \in A} H(Z_i) + \sum_{i \in A} \mu[f_i \log f_i].
\end{aligned}$$

Equivalently,

$$\sum_{i\in A} H(Z_i) - H(Z_A) = \mathrm{Ent}(f_A) - \sum_{i\in A} \mathrm{Ent}(f_i). \tag{4.9}$$

Shearer's estimate for the Shannon entropy states that

$$n(\mathbb{A})H(Z) \leq \sum_{A\in\mathbb{A}} H(Z_A); \tag{4.10}$$

see, for example, [9] or [2] for a proof. Therefore, summing over $A \in \mathbb{A}$ in (4.9) and using (4.10),

$$\sum_{A\in\mathbb{A}} \mathrm{Ent}(f_A) - n(\mathbb{A}) \sum_{i\in[n]} \mathrm{Ent}(f_i) \leq n(\mathbb{A}) \sum_{i\in[n]} H(Z_i) - n(\mathbb{A})H(Z). \tag{4.11}$$

Applying (4.9) with $A = [n]$ to the right hand side of (4.11), one obtains (4.8).

Suppose now that $\mathbb{A}$ is arbitrary, so that it need not cover $[n]$ or have uniform degrees. Then one can add singleton sets to $\mathbb{A}$ until one obtains a regular cover $\mathbb{A}'$ such that $n_+(\mathbb{A}) = n(\mathbb{A}')$. It then follows that

$$\sum_{A\in\mathbb{A}} \mathrm{Ent}(f_A) \leq \sum_{A\in\mathbb{A}'} \mathrm{Ent}(f_A) \leq n(\mathbb{A}') \mathrm{Ent}(f) = n_+(\mathbb{A}) \mathrm{Ent}(f).$$

This ends the proof of (4.8). $\qquad\square$

By applying Proposition 4.3 to the dyadic cover $\mathbb{A} = \{A, A^c\}$ one obtains that, for any $A \subset [n]$ and any nonnegative function f:

$$\mathrm{Ent}(f) \leq \mu\big[\mathrm{Ent}(f \mid A)\big] + \mu\big[\mathrm{Ent}(f \mid A^c)\big]; \tag{4.12}$$

$$\mathrm{Ent}(f_A) + \mathrm{Ent}(f_{A^c}) \leq \mathrm{Ent}(f). \tag{4.13}$$

The bounds (4.12) and (4.13) express respectively, the well-known tensorization and subadditivity properties of a product measure.

4.2. Proof of Theorem 1.2

Let us point out first that a simple application of Proposition 4.3 is not sufficient to prove Theorem 1.2. For instance, consider the uniform crossover model where $\nu(A) = 2^{-n}$ for any $A \subset [n]$. Then the left-hand side of (4.5) becomes

$$2^{-n+1} \sum_A \mathrm{Ent}(f_A).$$

If $f \in \mathbb{S}_\mu$, then

$$\mathrm{Ent}(f_i) = \mu[f_i \log f_i] = 0,$$

and, from Proposition 4.3 applied to the k-set cover $\mathbb{A}_k = \{A, |A| = k\}$, with $n_\pm(\mathbb{A}) = \binom{n-1}{k-1}$ one has

$$\sum_{A \in \mathbb{A}_k} \mathrm{Ent}(f_A) \le \binom{n-1}{k-1} \mathrm{Ent}(f),$$

for all $k = 2, \ldots, n$. These observations show that

$$2^{-n+1} \sum_A \mathrm{Ent}(f_A) \le 2^{-n+1} \mathrm{Ent}(f) \sum_{k=2}^n \binom{n-1}{k-1} = \left(1 - 2^{-n+1}\right) \mathrm{Ent}(f).$$

It follows that (4.5) holds with $\kappa = 2^{-n+1}$, a very poor estimate compared with the constant $\kappa(\nu) = (1 - 2^{-n+1})/(n-1)$ appearing in Theorem 1.2.

The key ingredient in our analysis is a suitable refinement of Proposition 4.3. Various improved versions of Shearer bounds have been proved recently; see [2,23]. Our refinements below (see Lemmas 4.5 and 4.6) appear to be new, and are inspired by the work of Balister and Bollobás [2], who emphasized the role played by the sub-modularity property of entropy. Recall that a map $A \mapsto h(A)$, $A \subset [n]$ is called *sub-modular* if for all $A, B \subset [n]$

$$h(A) + h(B) \ge h(A \cap B) + h(A \cup B). \tag{4.14}$$

We shall need the following simple lemma.

Lemma 4.4. *For every nonnegative function f and any product measure μ, the map*

$$A \mapsto h(A) = -\mathrm{Ent}(f_A)$$

is sub-modular, and $h(\varnothing) = 0$.

Proof. This follows from (4.9), and the fact that

$$H(Z_A) + H(Z_B) \ge H(Z_{A \cap B}) + H(Z_{A \cup B}), \tag{4.15}$$

for all $A, B \subset [n]$, and any random variable $Z = (Z_1, \ldots, Z_n)$. The proof of (4.15) is standard; see, for example, [2]. $\qquad\square$

We now proceed with the proof of Theorem 1.2, beginning with the case of uniform crossover.

4.2.1. *Uniform crossover*

Our analysis hinges on the following improved Shearer bound, which sharpens inequality (4.8) in Proposition 4.3.

Lemma 4.5. *For every $n \ge 2$, for any sub-modular map h with $h(\varnothing) = 0$,*

$$\sum_{A \subset [n]} h(A) \ge \frac{(n-2)2^{n-1} + 1}{n-1} h([n]) + \frac{2^{n-1} - 1}{n-1} \sum_{i=1}^n h(\{i\}). \tag{4.16}$$

Proof. For any $1 \leq k \leq n$, define $\mathbb{A}_k = \{A \subset [n] : |A| = k\}$, and

$$\varphi_k = \sum_{A \in \mathbb{A}_k} h(A).$$

Let us show that, for all $2 \leq k \leq n-1$,

$$\varphi_k \geq \frac{1}{k}\binom{n}{k-1}\varphi_n + \frac{n-k}{k}\varphi_{k-1}. \qquad (4.17)$$

Write

$$\varphi_k = \frac{1}{k} \sum_{A' \in \mathbb{A}_{k-1}} \sum_{\substack{A \in \mathbb{A}_k: \\ A' \subset A}} h(A).$$

Let us first prove that, for any $A' \in \mathbb{A}_{k-1}$, one has

$$\sum_{\substack{A \in \mathbb{A}_k: \\ A' \subset A}} h(A) \geq h([n]) + (n-k)h(A'). \qquad (4.18)$$

Assume, without loss of generality, that $A' = \{1, \ldots, k-1\}$. The sum above then becomes

$$\sum_{\substack{A \in \mathbb{A}_k: \\ A' \subset A}} h(A) = \sum_{j=k}^{n} h(A' \cup \{j\}).$$

From sub-modularity (4.14), one has

$$h(A' \cup \{k\}) + h(A' \cup \{k+1\}) \geq h(A') + h(A' \cup \{k, k+1\}).$$

Set $U_\ell = A' \cup \{k, \ldots, k+\ell\}$, for $\ell \in \{0, \ldots, n-k\}$. Then, recursively:

$$\sum_{j=k}^{k+\ell} h(A' \cup \{j\}) \geq \ell h(A') + h(U_\ell).$$

Setting $\ell = n-k$ proves (4.18), since $U_{n-k} = [n]$. Next, using (4.18) and noting that $h([n]) = \varphi_n$, one has (4.17).

Iterating (4.17), we arrive at

$$\varphi_k \geq c(k, n)\varphi_n + d(k, n)\varphi_1, \qquad (4.19)$$

where

$$d(k, n) = \frac{(n-2)!}{k!(n-k-1)!},$$

$$c(k, n) = \left[\frac{1}{k}\binom{n}{k-1} + \frac{(n-k)}{k(k-1)}\binom{n}{k-2} + \cdots + \frac{n(n-3)!}{k!(n-k-1)!} \right].$$

Note that (4.19) holds for all $k = 2, \ldots, n-1$. It can be extended to $k = 1, \ldots, n$ as well, provided we set $c(1, n) = 0$, $c(n, n) = 1$, $d(n, n) = 0$ and $d(1, n) = 1$. Thus, the claim (4.16) follows once we prove

$$\sum_{k=1}^{n} c(k, n) = \frac{(n-2)2^{n-1} + 1}{n-1} \tag{4.20}$$

and

$$\sum_{k=1}^{n} d(k, n) = \frac{2^{n-1} - 1}{n-1}. \tag{4.21}$$

The identity (4.21) follows from

$$\sum_{k=1}^{n-1} \frac{(n-2)!}{k!(n-k-1)!} = \frac{1}{n-1} \sum_{k=1}^{n-1} \binom{n-1}{k} = \frac{(2^{n-1} - 1)}{n-1}.$$

To check (4.20), notice that if $h(A) = \mathbf{1}(A \neq \varnothing)$, then (4.18) and (4.17) are both identities. Therefore, (4.19) is an identity as well, for all $k = 1, \ldots, n$ (with the above definitions of $c(k, n)$ and $d(k, n)$). Since here $\varphi_k = \binom{n}{k}$, summing over $k = 1, \ldots, n$ in (4.19) one has

$$2^n - 1 = \sum_{k=1}^{n} c(k, n) + n \frac{2^{n-1} - 1}{n-1},$$

which is equivalent to (4.20). This ends the proof of (4.16). $\qquad \square$

From Lemmas 4.5 and 4.4,

$$\sum_{A \subset [n]} \mathrm{Ent}(f_A) \leq \frac{(n-2)2^{n-1} + 1}{n-1} \mathrm{Ent}(f) + \frac{2^{n-1} - 1}{n-1} \sum_{i=1}^{n} \mathrm{Ent}(f_i). \tag{4.22}$$

Since $f \in \mathbb{S}_\mu$, one has $\mathrm{Ent}(f_i) = 0$ for all i, and therefore

$$2^{-n} \sum_{A \subset [n]} \left(\mathrm{Ent}(f_A) + \mathrm{Ent}(f_{A^c}) \right) \leq \frac{(n-2) + 2^{-n+1}}{n-1} \mathrm{Ent}(f),$$

which implies the lower bound $\kappa(\nu) \geq (1 - 2^{-n+1})/(n-1)$.

To prove the upper bound, we argue as follows. Suppose $X_1 = \cdots = X_n = X$ so that $\mathbb{X} = X^n$, and suppose μ is a product of identical probability measures μ_0 on X. Let Z_0 denote a random variable with values in X with distribution μ_0, and call Z the random variable $(Z_0, \ldots, Z_0)$ with values in $\mathbb{X}$, that is, Z consists of n identical copies of Z_0. Clearly,

$$H(Z_A) = H(Z_0) = -\sum_{x \in X} \mu_0(x) \log \mu_0(x), \tag{4.23}$$

for any $A \subset [n]$, $A \neq \varnothing$. Next, let f denote the probability density of Z with respect to μ, that is, $f\mu$ is the law of Z. Notice that $f \in \mathbb{S}_\mu$, since $f_i = 1$ for all i. By (4.9) one has, for all $A \subset [n]$,

$$\text{Ent}(f_A) = \big(|A| - 1\big)\mathbf{1}(A \neq \varnothing)H(Z_0). \tag{4.24}$$

It follows that

$$\sum_{A \subset [n]} \text{Ent}(f_A) = \big[(n-2)2^{n-1} + 1\big]H(Z_0).$$

Since $\text{Ent}(f) = (n-1)H(Z_0)$ we see that this choice of f saturates the bound (4.22). This concludes the proof of Theorem 1.2 for uniform crossover.

4.2.2. *The* Bernoulli(q) *model*

We need the following extension of Lemma 4.5.

Lemma 4.6. *For every* $\gamma \in (0, \infty)$, *for any sub-modular map h with $h(\varnothing) = 0$,*

$$\sum_{A \subset [n]} \gamma^{|A|}h(A) \geq \frac{(1+\gamma)^{n-1}[\gamma(n-1)-1]+1}{n-1}h\big([n]\big) + \frac{(1+\gamma)^{n-1}-1}{n-1}\sum_{i=1}^{n}h\big(\{i\}\big). \tag{4.25}$$

Proof. The left-hand side of (4.25) coincides with $\sum_{k=1}^{n} \hat{\varphi}_k$, where $\hat{\varphi}_k := \gamma^k\varphi_k$ and φ_k was defined in Lemma 4.5. From (4.19), it follows that

$$\hat{\varphi}_k \geq \gamma^k c(k, n)\varphi_n + \gamma^k d(k, n)\varphi_1.$$

Thus, (4.25) will follow if we can prove

$$\sum_{k=1}^{n} \gamma^k c(k, n) = \frac{(1+\gamma)^{n-1}[\gamma(n-1)-1]+1}{n-1}$$

and

$$\sum_{k=1}^{n} \gamma^k d(k, n) = \frac{(1+\gamma)^{n-1}-1}{n-1}.$$

It is not hard to check that these identities follow in the same way as (4.20)–(4.21). $\qquad\square$

Next, observe that for $\nu(A) = q^{|A|}(1-q)^{n-|A|}$, one has

$$\sum_{A \subset [n]} \nu(A)\big[\text{Ent}(f_A) + \text{Ent}(f_{A^c})\big]$$

$$= (1-q)^n \sum_{A \subset [n]} \left(\frac{q}{1-q}\right)^{|A|} \text{Ent}(f_A) + q^n \sum_{A \subset [n]} \left(\frac{q}{1-q}\right)^{-|A|} \text{Ent}(f_A).$$

We estimate each sum above using Lemma 4.6 with $h(A) = -\text{Ent}(f_A)$, once with $\gamma = \frac{q}{1-q}$ and once with $\gamma = (\frac{q}{1-q})^{-1}$. This yields

$$\sum_{A \subset [n]} v(A)\big[\text{Ent}(f_A) + \text{Ent}(f_{A^c})\big] \leq \Delta(q,n) \text{Ent}(f), \tag{4.26}$$

where

$$\Delta(q,n) = (1-q)^n \frac{(1-q)^{1-n}[\frac{q}{1-q}(n-1)-1]+1}{n-1} + q^n \frac{q^{1-n}[\frac{1-q}{q}(n-1)-1]+1}{n-1}$$

$$= 1 - \frac{1-q^n-(1-q)^n}{n-1}.$$

This implies the lower bound

$$\kappa(v) \geq \frac{1-q^n-(1-q)^n}{n-1}.$$

To prove the upper bound, we argue as in (4.23)–(4.24). Using the same function f defined there, one has

$$\sum_{A \subset [n]} v(A)\big(\text{Ent}(f_A) + \text{Ent}(f_{A^c})\big)$$

$$= \sum_{A \subset [n]} q^{|A|}(1-q)^{n-|A|}\big[\big(|A|-1\big)\mathbf{1}(A \neq \varnothing) + \big(|A^c|-1\big)\mathbf{1}\big(A \neq [n]\big)\big]H(Z_0)$$

$$= \big(n-2+v([n])+v(\varnothing)\big)H(Z_0) = (n-1)H(Z_0)\bigg[1 - \frac{1-q^n-(1-q)^n}{n-1}\bigg].$$

Since $\text{Ent}(f) = (n-1)H(Z_0)$, this choice of f saturates the bound (4.26). This proves Theorem 1.2 for the Bernoulli(q) model.

4.2.3. *Single site recombination*

When $v(A) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(A = \{i\})$, using $\text{Ent}(f_i) = 0$ for all $i \in [n]$, (4.5) becomes

$$\frac{1}{n}\sum_{i=1}^n \mu[f_{\{i\}^c} \log f_{\{i\}^c}] \leq (1-\kappa)\text{Ent}(f). \tag{4.27}$$

We note that an application of Proposition 4.3 with the $(n-1)$-cover $\mathbb{A}_{n-1} = \{\{i\}^c, i \in [n]\}$ gives the inequality (4.27) with constant $\kappa = 1/n$.

To prove that it can be strengthened to $\kappa = 1/(n-1)$, we observe that the left hand side of (4.27) can be written

$$\frac{1}{n}\sum_{i=1}^n \mu[f_{\{i\}^c} \log f_{\{i\}^c}] = -\frac{1}{n}\varphi_{n-1},$$

where φ_k is defined in the proof of Lemma 4.5 with $h(A) = -\text{Ent}(f_A)$. From (4.19),

$$\frac{1}{n} \sum_{i=1}^{n} \mu[f_{\{i\}^c} \log f_{\{i\}^c}] \leq \frac{1}{n} c(n-1, n) \, \text{Ent}(f).$$

The coefficient $c(n-1, n)$ can be computed as in the proof of Lemma 4.5, and one finds $c(n-1, n) = n - \frac{n}{n-1}$. Thus, (4.27) holds with $\kappa = \frac{1}{n-1}$, which proves the lower bound $\kappa(\nu) \geq \frac{1}{n-1}$.

The upper bound follows as in (4.23)–(4.24). Indeed, with that choice of f and μ one has

$$\frac{1}{n} \sum_{i=1}^{n} \mu[f_{\{i\}^c} \log f_{\{i\}^c}] = (n-2) H(Z_0) = \frac{n-2}{n-1} \, \text{Ent}(f).$$

This concludes the proof of Theorem 1.2 for single site recombinations.

4.2.4. *One-point crossover*

Here

$$\nu(A) = \frac{1}{n+1} \sum_{i=0}^{n} \mathbf{1}(A = J_i),$$

where $J_0 = \varnothing$ and $J_i = \{1, \ldots, i\}$, $i \geq 1$. Fix $f \geq 0$ and define, for $i = 1, \ldots, n$:

$$u(i) = \text{Ent}(f_{J_i}), \qquad \bar{u}(i) = \text{Ent}(f_{J_{i-1}^c}).$$

Notice that $J_i \cap J_{i-1}^c = \{i\}$, and $J_i \cup J_{i-1}^c = [n]$. Thus, from Lemma 4.4 one has

$$u(i) + \bar{u}(i) \leq \text{Ent}(f_i) + \text{Ent}(f).$$

Therefore,

$$\sum_{A \subset [n]} \nu(A)\big[\text{Ent}(f_A) + \text{Ent}(f_{A^c})\big] = \frac{1}{n+1} \sum_{i=1}^{n} \big(u(i) + \bar{u}(i)\big)$$

$$\leq \frac{n}{n+1} \, \text{Ent}(f) + \frac{1}{n+1} \sum_{i=1}^{n} \text{Ent}(f_i). \tag{4.28}$$

If $f \in \mathbb{S}_\mu$ one has $\text{Ent}(f_i) = 0$ for all $i \in [n]$, and (4.28) proves the desired upper bound. To prove that (4.28) is optimal, notice that with the argument in (4.23)–(4.24) one obtains $u(i) + \bar{u}(i) = \text{Ent}(f)$ for every $i = 1, \ldots, n$, and therefore (4.28) is an identity for this choice of f. This proves Theorem 1.2 for the one-point crossover model, and thus concludes the proof of the theorem.

4.3. Proof of Theorem 1.1

From Lemma 4.2 and Theorem 1.2 we have already obtained the desired lower bounds $\delta(\nu) \geq \kappa(\nu)$. The upper bounds on $\delta(\nu)$ are based on the following estimate.

Proposition 4.7. *Let ν be one of the four recombination distributions, and set $\mathbb{X} = \{0, 1\}^n$. Let $w = w(n) = 2^{-n}$, and let μ be the product of Bernoulli(w) probability measures. Define*

$$\Delta_\nu = \sum_A \nu(A)\left(2^{-|A|} + 2^{-|A^c|}\right). \tag{4.29}$$

Then,

$$\delta(\mu, \nu) \leq \frac{4(1 - \Delta_\nu)}{n} + O\left(n^{-2}\right).$$

Let us first check that Proposition 4.7 implies the upper bounds on $\delta(\nu)$ announced in Theorem 1.1. Indeed, in the case of single site recombination $\nu(A) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(A = \{i\})$, one has

$$\Delta_\nu = \frac{1}{2} + O\left(2^{-n}\right), \tag{4.30}$$

and therefore $\delta(\nu) \leq \delta(\mu, \nu) \leq \frac{2}{n} + O(n^{-2})$. In the case of one-point crossover, one finds easily that $\Delta_\nu = O(1/n)$, and therefore $\delta(\nu) \leq \frac{4}{n} + O(n^{-2})$. For uniform crossover $\nu(A) = 2^{-n}$, one has that Δ_ν is exponentially small, and thus again $\delta(\nu) \leq \frac{4}{n} + O(n^{-2})$. Finally, for the Bernoulli(q) model, one finds

$$\Delta_\nu = \sum_{k=0}^n \binom{n}{k} q^k (1-q)^{n-k} \left(2^{-k} + 2^{-n+k}\right) = \left(1 - \frac{q}{2}\right)^n + \frac{1}{2^n}(1+q)^n. \tag{4.31}$$

Since $q \leq 1/2$, $\Delta_\nu = (1 - q/2)^n + O((3/4)^n)$, which yields the claimed upper bound on $\delta(\nu)$.

Proof of Proposition 4.7. Let $B(u) \in \mathbb{P}(\mathbb{X})$ denote the product of independent Bernoulli with parameter $u \in [0, 1]$, so that $\mu = B(w)$, and define $f = p/\mu$, where

$$p = w^2 B(1) + (1 - w)^2 B(0) + 2w(1 - w) B\left(\frac{1}{2}\right). \tag{4.32}$$

It is easily checked that p and μ have the same marginals, that is, $p_i = \mu_i$, so that $f_i = 1$ for all $i \in [n]$. Then p can be written as

$$p(\sigma) = a\mathbf{1}(\sigma \equiv 1) + b\mathbf{1}(\sigma \equiv 0) + c\mathbf{1}(\sigma \not\equiv 1 \text{ and } \sigma \not\equiv 0),$$

where $a = w^2 + 2w(1-w)2^{-n}$, $b = (1-w)^2 + 2w(1-w)2^{-n}$, and $c = 2w(1-w)2^{-n}$. The relative entropy is given by

$$\text{Ent}(f) = a \log\left(\frac{a}{w^n}\right) + b \log\left(\frac{b}{(1-w)^n}\right) + c \sum_{k=1}^{n-1} \binom{n}{k} \log\left(\frac{c}{w^k(1-w)^{n-k}}\right).$$

Using $c\sum_{k=1}^{n-1}\binom{n}{k} = 2w(1-w)(1-2^{-n+1})$ and $c\sum_{k=1}^{n-1}\binom{n}{k}k = nw(1-w)(1-2^{-n+1})$,

$$
\begin{aligned}
\text{Ent}(f) = {} & a\log\left(\frac{a}{w^n}\right) + b\log\left(\frac{b}{(1-w)^n}\right) \\
& + 2w(1-w)\left(1-2^{-n+1}\right)\log\left(\frac{c}{(1-w)^n}\right) \\
& + nw(1-w)\left(1-2^{-n+1}\right)\log\left(\frac{1-w}{w}\right).
\end{aligned}
\tag{4.33}
$$

Using $w = 2^{-n}$, one finds

$$
a\log\left(\frac{a}{w^n}\right) = O\left(2^{-n}\right); \qquad b\log\left(\frac{b}{(1-w)^n}\right) = n2^{-n} + O\left(2^{-n}\right);
$$

$$
2w\log\left(\frac{c}{(1-w)^n}\right) = -4n2^{-n}\log 2 + O\left(2^{-n}\right);
$$

$$
nw\log\left(\frac{1-w}{w}\right) = n^2 2^{-n}\log 2 + O\left(2^{-n}\right).
$$

Therefore,

$$
\text{Ent}(f) = n^2 2^{-n}\log 2 - 4n2^{-n}\log 2 + n2^{-n} + O\left(2^{-n}\right).
\tag{4.34}
$$

Next, we compute the entropy production $D(f, f)$. Define

$$
\widetilde{p}(\sigma) = \sum_A \nu(A)(p_A \otimes p_{A^c}).
$$

As in (4.3), we write

$$
D(f, f) = \text{Ent}(f) - 2^{-n}\sum_A \mu[f_A f_{A^c}\log f] = \text{Ent}(f) - \widetilde{p}[\log(p/\mu)].
\tag{4.35}
$$

Observe that

$$
\widetilde{p}[\log(p/\mu)] = \alpha_n\log\left(\frac{a}{w^n}\right) + \alpha_0\log\left(\frac{b}{(1-w)^n}\right) + \sum_{k=1}^{n-1}\alpha_k\log\left(\frac{c}{w^k(1-w)^{n-k}}\right),
$$

where $\alpha_k = \sum_{\sigma:|\sigma|=k}\widetilde{p}(\sigma)$ and $|\sigma|$ denotes the number of 1's in σ. From the conservation of marginals, one has that $\sum_{k=1}^n k\alpha_k = nw$. Therefore,

$$
\begin{aligned}
\widetilde{p}[\log(p/\mu)] = {} & \alpha_n\log\left(\frac{a}{w^n}\right) + \alpha_0\log\left(\frac{b}{(1-w)^n}\right) \\
& + (1-\alpha_0-\alpha_n)\log\left(\frac{c}{(1-w)^n}\right) + (nw - n\alpha_n)\log\left(\frac{1-w}{w}\right).
\end{aligned}
$$

From (4.33) and (4.35), we obtain

$$D(f, f) = (a - \alpha_n) \log\left(\frac{a}{w^n}\right) + (b - \alpha_0) \log\left(\frac{b}{(1 - w)^n}\right)$$

$$+ \beta \log\left(\frac{c}{(1 - w)^n}\right) + \gamma \log\left(\frac{1 - w}{w}\right),$$

(4.36)

where the new coefficients β, γ are given by

$$\beta = 2w(1 - w)\left(1 - 2^{-n+1}\right) - (1 - \alpha_0 - \alpha_n), \qquad \gamma = n\alpha_n - nw\left(2^{-n+1}(1 - w) + w\right).$$

In order to estimate the ratio $D(f, f)/\mathrm{Ent}(f)$, we need to control α_0 and α_n. These can be computed as follows. For any $A \subset [n]$, $u, v \in [0, 1]$, write

$$B(u)_A \otimes B(v)_{A^c} = B(u_A v_{A^c}),$$

where $B(u_A v_{A^c})$ denotes the product of independent Bernoulli measures such that the sites in A have parameter u and the sites in A^c have parameter v. It follows that

$$\widetilde{p} = \sum_A v(A)\big\{w^4 B(1) + (1 - w)^4 B(0) + 4w^2(1 - w)^2 B(1/2) + w^2(1 - w)^2 B(1_A 0_{A^c})$$

$$+ w^2(1 - w)^2 B(0_A 1_{A^c}) + 2w(1 - w)^3 B\big(0_A(1/2)_{A^c}\big) + 2w(1 - w)^3 B\big((1/2)_A 0_{A^c}\big)$$

$$+ 2w^3(1 - w) B\big(1_A(1/2)_{A^c}\big) + 2w^3(1 - w) B\big((1/2)_A 1_{A^c}\big)\big\}.$$

From this expression, it is not hard to check that

$$\alpha_n = O\big(w^2\big), \qquad \alpha_0 = 1 - 4w + 2w\Delta_v + O\big(w^2\big),$$

where Δ_v is given by (4.29). Moreover, $\beta = -2w + 2w\Delta_v + O(w^2)$ and $\gamma = O(nw^2)$. Therefore,

$$(a - \alpha_n) \log\left(\frac{a}{w^n}\right) = O\big(2^{-n}\big); \qquad (b - \alpha_0) \log\left(\frac{b}{(1 - w)^n}\right) = O\big(2^{-n}\big);$$

$$\beta \log\left(\frac{c}{(1 - w)^n}\right) = -4(1 - \Delta_v)n2^{-n} \log 2 + O\big(2^{-n}\big); \qquad \gamma \log\left(\frac{1 - w}{w}\right) = O\big(2^{-n}\big).$$

From (4.34) and (4.36), it follows that

$$\delta(\mu, v) \leq \frac{D(f, f)}{\mathrm{Ent}(f)} = \frac{4(1 - \Delta_v)}{n} + O\big(n^{-2}\big).$$

This concludes the proof of Proposition 4.7 and of Theorem 1.1.   □

4.4. Proof of Corollary 1.3

Let ν be one of the four recombination distributions and fix an arbitrary initial state $p \in \mathbb{P}(\mathbb{X})$. Let $\pi = \otimes_{i=1}^{n} p_i$ be the associated product measure. If π fails to satisfy $\pi(\sigma) > 0$ for all $\sigma \in \mathbb{X}$ we can always redefine $\mathbb{X}$ so that this holds. Thus, without loss of generality, we assume that $\pi \in \mathbb{P}_+(\mathbb{X})$, and we can work within the general framework developed in Section 2 with $\mu = \pi$. From Proposition 2.5 and the observations in Lemma 4.1 and Lemma 4.2, we know that

$$H(p_t \mid \pi) \leq e^{-\kappa(\nu)t} H(p \mid \pi)$$

for all $t \geq 0$, where $\kappa(\nu)$ is the constant computed in Theorem 1.2. This proves the upper bound (1.8). We turn to the corresponding statement in discrete time. The following lemma proves the desired upper bound (1.9). Let $p^{(k)} = \Psi[p^{(k-1)}]$, $k \in \mathbb{N}$, $p^{(0)} = p$, where

$$\Psi[p] = \sum_{A} \nu(A)(p_A \otimes p_{A^c}).$$

Lemma 4.8. *For any initial state $p \in \mathbb{P}(\mathbb{X})$,*

$$H\big(p^{(k)} \mid \pi\big) \leq \big(1 - \kappa(\nu)\big) H\big(p^{(k-1)} \mid \pi\big),$$

for all $k \in \mathbb{N}$, where $\kappa(\nu)$ is the constant in Theorem 1.2 and $\pi = \otimes_{i=1}^{n} p_i$.

Proof. Since the map Ψ preserves the marginals, it suffices to show that, for any p with the same marginals as π, one has

$$H\big(\Psi[p] \mid \pi\big) \leq \big(1 - \kappa(\nu)\big) H(p \mid \pi), \tag{4.37}$$

Set $f = p/\pi$ and $f_\Psi = \Psi[p]/\pi$. Now, $f_\Psi = \sum_A \nu(A) f_A f_{A^c}$. Convexity of the function $x \mapsto x \log x$, $x \geq 0$, shows that

$$\pi[f_\Psi \log f_\Psi] \leq \sum_{A} \nu(A) \pi[f_A f_{A^c} \log f_A f_{A^c}] = \sum_{A} \nu(A)\big[\mathrm{Ent}(f_A) + \mathrm{Ent}(f_{A^c})\big].$$

Therefore, by Theorem 1.2,

$$\pi[f_\Psi \log f_\Psi] \leq \big(1 - \kappa(\nu)\big) \pi[f \log f],$$

which coincides with (4.37). $\qquad\square$

To conclude the proof of Corollary 1.3 we need to check the bounds in (1.10). To this end, take $\mathbb{X} = \{0, 1\}^n$ and p the probability measure defined in (4.32). If $f = p/\pi$, then the proof of Proposition 4.7 shows that

$$D(f, f) \leq \gamma(\nu) \mathrm{Ent}(f), \qquad \gamma(\nu) := \left(\frac{4}{n}(1 - \Delta_\nu) + O\big(n^{-2}\big)\right). \tag{4.38}$$

By Proposition 2.3 part 1, one has

$$\frac{\mathrm{d}}{\mathrm{d}t} H(p_t \mid \pi)\Big|_{t=0^+} = -D(f, f) \geq -\gamma(\nu) H(p \mid \pi).$$

On the other hand, in discrete time one has

$$H\big(p^{(1)} \mid \pi\big) = \pi[f_\Psi \log f_\Psi].$$

From the variational principle for relative entropy,

$$\pi[f_\Psi \log f_\Psi] \geq \pi[f_\Psi \log f] = \sum_A \nu(A)\pi[f_A f_{A^c} \log f]. \tag{4.39}$$

From (4.5), (4.39) and (4.38) it follows that

$$H\big(p^{(1)} \mid \pi\big) \geq H(p \mid \pi) - D(f, f) \geq \big(1 - \gamma(\nu)\big) H(p \mid \pi),$$

which proves (1.10). Finally, the estimate

$$\gamma(\nu) \leq C\kappa(\nu)$$

follows easily from (4.30)–(4.31). Because of the $O(n^{-2})$ correction in $\gamma(\nu)$ one needs to require $q \geq n^{-2}$ in the case of the Bernoulli(q) model.

Acknowledgements

References

[1] Baake, E., Baake, M. and Salamat, M. (2016). The general recombination equation in continuous time and its solution. *Discrete Contin*. *Dyn*. *Syst*. **36** 63–95. MR3369214

[2] Balister, P. and Bollobás, B. (2012). Projections, entropy and sumsets. *Combinatorica* **32** 125–141. MR2927635

[3] Bobkov, S.G. and Tetali, P. (2006). Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret*. *Probab*. **19** 289–336. MR2283379

[4] Caputo, P., Menz, G. and Tetali, P. (2015). Approximate tensorization of entropy at high temperature. *Ann*. *Fac*. *Sci*. *Toulouse Math*. (6) **24** 691–716. MR3434252

[5] Carlen, E.A. and Carvalho, M.C. (1992). Strict entropy production bounds and stability of the rate of convergence to equilibrium for the Boltzmann equation. *J. Stat*. *Phys*. **67** 575–608. MR1171145

[6] Carlen, E.A., Carvalho, M.C. and Gabetta, E. (2000). Central limit theorem for Maxwellian molecules and truncation of the Wild expansion. *Comm*. *Pure Appl*. *Math*. **53** 370–397. MR1725612

[7] Carlen, E.A., Carvalho, M.C., Le Roux, J., Loss, M. and Villani, C. (2010). Entropy and chaos in the Kac model. *Kinet*. *Relat*. *Models* **3** 85–122. MR2580955

[8] Cercignani, C. (1988). *The Boltzmann Equation and Its Applications*. *Applied Mathematical Sciences* **67**. New York: Springer. MR1313028

[9] Chung, F.R.K., Graham, R.L., Frankl, P. and Shearer, J.B. (1986). Some intersection theorems for ordered sets and graphs. *J. Combin. Theory Ser. A* **43** 23–37. MR0859293

[10] Dai Pra, P., Paganoni, A.M. and Posta, G. (2002). Entropy inequalities for unbounded spin systems. *Ann. Probab.* **30** 1959–1976. MR1944012

[11] Desvillettes, L., Mouhot, C. and Villani, C. (2011). Celebrating Cercignani's conjecture for the Boltzmann equation. *Kinet. Relat. Models* **4** 277–294. MR2765747

[12] Diaconis, P. and Saloff-Coste, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** 695–750. MR1410112

[13] Dolera, E., Gabetta, E. and Regazzini, E. (2009). Reaching the best possible rate of convergence to equilibrium for solutions of Kac's equation via central limit theorem. *Ann. Appl. Probab.* **19** 186–209. MR2498676

[14] Feinberg, M. (1972). On chemical kinetics of a certain class. *Arch. Ration. Mech. Anal.* **46** 1–41. MR0378591

[15] Friedgut, E. (2004). Hypergraphs, entropy, and inequalities. *Amer. Math. Monthly* **111** 749–760. MR2104047

[16] Geiringer, H. (1944). On the probability theory of linkage in Mendelian heredity. *Ann. Math. Stat.* **15** 25–57. MR0010384

[17] Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading MA: Addison-Wesley.

[18] Hauray, M. and Mischler, S. (2014). On Kac's chaos and related problems. *J. Funct. Anal.* **266** 6055–6157. MR3188710

[19] Horn, F. and Jackson, R. (1972). General mass action kinetics. *Arch. Ration. Mech. Anal.* **47** 81–116. MR0400923

[20] Kac, M. (1956). Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, 1954–1955, *Vol. III* 171–197. Univ. California Press, Berkeley and Los Angeles. MR0084985

[21] Levin, D.A., Peres, Y. and Wilmer, E.L. (2009). *Markov Chains and Mixing Times*. Providence, RI: Amer. Math. Soc. With a chapter by James G. Propp and David B. Wilson. MR2466937

[22] Lyubich, Y.I. (1992). *Mathematical Structures in Population Genetics*. *Biomathematics* **22**. Berlin: Springer. Translated from the 1983 Russian original by D. Vulis and A. Karpov. MR1224676

[23] Madiman, M. and Tetali, P. (2010). Information inequalities for joint distributions, with interpretations and applications. *IEEE Trans. Inform. Theory* **56** 2699–2713. MR2683430

[24] Martínez, S. (2017). A probabilistic analysis of a discrete-time evolution in recombination. *Adv. in Appl. Math.* **91** 115–136. MR3673582

[25] Mischler, S. and Mouhot, C. (2013). Kac's program in kinetic theory. *Invent. Math.* **193** 1–147. MR3069113

[26] Norris, J.R. (1998). *Markov Chains*. *Cambridge Series in Statistical and Probabilistic Mathematics* **2**. Cambridge: Cambridge Univ. Press. Reprint of 1997 original. MR1600720

[27] Rabani, Y., Rabinovich, Y. and Sinclair, A. (1998). A computational view of population genetics. *Random Structures Algorithms* **12** 313–334. MR1639744

[28] Rabinovich, Y., Sinclair, A. and Wigderson, A. (1992). Quadratic dynamical systems. In *Foundations of Computer Science*, 1992. *Proceedings*, 33*rd Annual Symposium on* 304–313. IEEE.

[29] van den Berg, J. and Gandolfi, A. (2013). BK-type inequalities and generalized random-cluster representations. *Probab. Theory Related Fields* **157** 157–181. MR3101843

[30] Villani, C. (2002). A review of mathematical topics in collisional kinetic theory. In *Handbook of Mathematical Fluid Dynamics*, *Vol. I* 71–305. Amsterdam: North-Holland. MR1942465

[31] Villani, C. (2003). Cercignani's conjecture is sometimes true and always almost true. *Comm. Math. Phys.* **234** 455–490. MR1964379

Received October 2016 and revised May 2017