LECTURE NOTES ON ENTROPY AND MARKOV CHAINS

PIETRO CAPUTO

Abstract. Notes for lectures at UC Santa Barbara (Summer 2022) and UC Berkeley (Fall 2022). *In progress*

Contents

1. Setup and preliminaries

Throughout, Ω denotes a finite space, and $\mathcal{P}(\Omega)$ is the set of probability measures on Ω.

1.1. **Relative entropy.** Given $\nu, \mu \in \mathcal{P}(\Omega)$, the *relative entropy* of ν with respect to μ is defined as

$$H(\nu \,|\, \mu) = \sum_{\sigma \in \Omega} \nu(\sigma) \log\left(\frac{\nu(\sigma)}{\mu(\sigma)}\right),$$

with the conventions $0\log(0) = 0\log(0/0) = 0$. This definition makes sense whenever $\mu(\sigma) = 0$ implies $\nu(\sigma) = 0$, that is when ν is absolutely continuous with respect to μ. Otherwise, one defines $H(\nu \,|\, \mu) = +\infty$. Relative entropy is one of the most common and effective ways of measuring the "distance" between probability measures. It is also known as Kullback-Leibler divergence. If $H(X)$ denotes the Shannon entropy of a random variable X taking values in Ω and with distribution $\nu \in \mathcal{P}(\Omega)$ then

$$H(\nu \,|\, \mu) = -H(X) - \nu[\log \mu]. \tag{1.1}$$

Note that we are using only natural logarithms here, whereas Shannon entropy is often defined using base 2 logarithms. Recall that the total variation distance between $\mu, \nu \in \mathcal{P}(\Omega)$ is defined as

$$\|\mu - \nu\|_{\mathrm{TV}} = \tfrac{1}{2} \sum_{\sigma \in \Omega} |\mu(\sigma) - \nu(\sigma)|.$$

The following lemma summarizes the main properties of the relative entropy that we shall need.

Lemma 1.1. *For all $\nu, \mu \in \mathcal{P}(\Omega)$,*

1) *$H(\nu \,|\, \mu) \geqslant 0$ and $H(\nu \,|\, \mu) = 0 \iff \nu = \mu$.*
2) *Convexity: for any collection of probability measures $\{\nu_i\}$, and all $\alpha_i \geqslant 0$ with $\sum_i \alpha_i = 1$,*
$$H\Big(\sum_i \alpha_i \nu_i \,\Big|\, \mu\Big) \leqslant \sum_i \alpha_i H(\nu_i \,|\, \mu).$$
3) *Pinsker inequality:*
$$\|\mu - \nu\|_{\mathrm{TV}}^2 \leqslant \tfrac{1}{2} H(\nu \,|\, \mu). \tag{1.2}$$
4) *Variational principle:*
$$H(\nu \,|\, \mu) = \sup_{g:\Omega \mapsto \mathbb{R}} \{\nu[g] - \log \mu[e^g]\}. \tag{1.3}$$

Proof. Define $h(\sigma) = \nu(\sigma)/\mu(\sigma)$. Then $\mu[h] = 1$ and $H(\nu \,|\, \mu) = \mu[h \log h]$. By Jensen inequality and the convexity of $x \mapsto x \log x$, $x \geqslant 0$, $\mu[h \log h] \geqslant \mu[h] \log \mu[h]$, with equality if and only if h is constant. This proves 1.

To prove 2, if $h_i = \nu_i(\sigma)/\mu(\sigma)$, $h = \sum_i \alpha_i h_i$, then convexity of $x \mapsto x \log x$, $x \geqslant 0$, implies $h \log h \leqslant \sum_i \alpha_i h_i \log h_i$ and

$$H\Big(\sum_i \alpha_i \nu_i \,\Big|\, \mu\Big) = \mu[h \log h] \leqslant \sum_i \alpha_i \mu[h_i \log h_i] = \sum_i \alpha_i H(\nu_i \,|\, \mu).$$

To prove 3 one checks that $u \log u \geqslant u - 1$ for all $u \geqslant 0$, and that

$$3(u-1)^2 \leqslant (2u+4)(u \log u - u + 1), \qquad u \geqslant 0.$$

Taking the square root, applying this to $u = h(\sigma)$ and integrating one finds

$$\begin{aligned}\|\mu - \nu\|_{\mathrm{TV}} &= \tfrac{1}{2} \int |h-1|\, d\mu \leqslant \tfrac{1}{2\sqrt{3}} \int \sqrt{(2h+4)(h\log h - h + 1)}\, d\mu \\ &\leqslant \tfrac{1}{2\sqrt{3}} \left(\int (2h+4)\, d\mu\right)^{1/2} \left(\int (h \log h - h + 1)\, d\mu\right)^{1/2} = \tfrac{1}{\sqrt{2}} \sqrt{H(\nu \,|\, \mu)}.\end{aligned}$$

To prove the variational principle (1.3), note that for any $g : \Omega \mapsto \mathbb{R}$,

$$\nu[g] = H(\nu \,|\, \mu) + \mu[h \log(e^g/h)] \leqslant H(\nu \,|\, \mu) + \log \mu[e^g],$$

where we use the concavity of the logarithm and Jensen's inequality applied to $\nu = \mu h \in \mathcal{P}(\Omega)$. The inequality is saturated at $g = \log h$, which proves (1.3). ☐

1.2. **The Gibbs measure.** Consider a product space $\Omega = \Omega_1 \times \Omega_2 \cdots \times \Omega_n$, where Ω_i are finite sets. We fix a reference probability measure $\mu \in \mathcal{P}(\Omega)$. The measure μ is often referred to as the *Gibbs measure*. Some key examples to keep in mind for later applications are as follows.

Example 1.2 (Ising model). *Set $\Omega = \{-1, +1\}^n$. Fix $\beta \in \mathbb{R}$ and let $G = (V, E)$ be a graph with $V = [n] = \{1, \dots, n\}$. The Ising model $\mu = \mu_{G,\beta}$ is the Gibbs measure on Ω such that*

$$\mu(\sigma) \propto \exp\left(\beta \sum_{xy \in E} \sigma_x \sigma_y\right), \tag{1.4}$$

where $\sigma = (\sigma_x,\, x \in V) \in \Omega$ denotes a spin configuration.

Example 1.3 (Potts model). *Set $\Omega = \{1, \dots, q\}^n$, where $q \in \mathbb{N}$, $q \geqslant 2$. Fix $\beta \in \mathbb{R}$ and let $G = (V, E)$ be a graph with $V = [n]$. The Potts model $\mu = \mu_{G,\beta}$ is the Gibbs measure on Ω such that*

$$\mu(\sigma) \propto \exp\left(\beta \sum_{xy \in E} \mathbf{1}_{\sigma_x = \sigma_y}\right), \tag{1.5}$$

where $\sigma = (\sigma_x,\, x \in V) \in \Omega$. If $q = 2$ this is the Ising model with β replaced by $\beta/2$.

The models above are classical examples of spin systems from Statistical Mechanics, see the book [26] for background.

Example 1.4 (Proper colorings of a graph). *Set $\Omega = \{1, \dots, q\}^n$, where $q \in \mathbb{N}$, $q \geqslant 2$ and and let $G = (V, E)$ be a graph with $V = [n]$. Take μ as uniform probability measure over all proper colorings of the graph G with q colors. This can be obtained as the limit $\beta \to -\infty$ of* (1.5)*.*

Example 1.5 (Hard core model). *Set $\Omega = \{0, 1\}^n$, and let $G = (V, E)$ be a graph with $V = [n]$. For any $\lambda > 0$ define the hard core model with fugacity λ as the probability measure $\mu = \mu_{G,\lambda}$ such that*

$$\mu(\sigma) \propto \lambda^{\sum_{x \in V} \sigma_x} \mathbf{1}_{\sigma \in \Omega_0}$$

where $\Omega_0 = \{\sigma \in \Omega :\ \sigma_x \sigma_y = 0\,,\ \forall xy \in E\}$ is the set of independent sets of G.

Example 1.6 (Uniform permutations). *Set $\Omega = [n]^n$, and let μ be the uniform distribution on the symmetric group $S_n \subset \Omega$ of permutations of $[n]$, that is*

$$S_n = \{\sigma \in \Omega :\ \sigma_x \neq \sigma_y,\ \forall x, y \in [n],\ x \neq y\}.$$

We note that in the last three examples the measure μ is supported on a strict subset of configurations, that is the system has hard constraints.

1.3. **Conditional expectations.** The elements of Ω are referred to as spin configurations. Given $\sigma \in \Omega$, a subset $A \subset [n]$, $\sigma_A = (\sigma_x,\, x \in A)$ denotes the spin configuration restricted to A. The conditional distribution given the spins $\tau \in \Omega_{A^c}$ in A^c is denoted μ_A^τ:

$$\mu_A^\tau(\eta) = \mu(\sigma = \eta \,|\, \sigma_{A^c} = \tau) \mathbf{1}_{\eta_{A^c} = \tau}\,, \tag{1.6}$$

The definition (1.6) is interpreted as a probability measure on Ω and, with a slight abuse of notation, sometimes we think of it as a probability on Ω_A. Given a function $f : \Omega \mapsto \mathbb{R}$,

$\tau \in \Omega_{A^c}$, we write $\mu_A^\tau f$ for the expectation of f under μ_A^τ, and write $\mu_A f$ for the function $\Omega_{A^c} \ni \tau \mapsto \mu_A^\tau f$. Therefore, $\mu_A f$ satisfies

$$[\mu_A f](\tau) = \mu_A^\tau f = \sum_{\xi \in \Omega} \mu(\sigma_A = \xi_A \,|\, \sigma_{A^c} = \tau) f(\xi)\, \mathbf{1}_{\xi_{A^c} = \tau} \,.$$

We refer to the function $\mu_A f$ as the *conditional expectation*, and to the configuration τ in μ_A^τ as the *boundary condition*, or *pinning*. The elementary properties of conditional probabilities imply the relations

$$\mu_A \mu_B f = \mu_A f \,, \qquad B \subset A \subset [n],$$

valid for all all functions $f : \Omega \mapsto \mathbb{R}$. Another property that will be useful is the fact that if f is a probability density w.r.t. μ, that is $f \geqslant 0$ and $\mu(f) = 1$, then $\mu_A f$ is the probability density of the marginal of $\nu := f\mu$ on Ω_{A^c}. Indeed, setting $\eta_{A^c} = \tau$,

$$\begin{aligned}\sum_{\eta_A} \nu(\eta_A, \eta_{A^c}) &= \sum_{\eta_A} f(\eta_A, \tau)\mu(\eta_A, \tau) \\ &= \mu(\sigma_{A^c} = \tau) \sum_{\eta_A} f(\eta_A, \tau)\mu(\eta_A | \sigma_{A^c} = \tau) = \mu(\sigma_{A^c} = \tau)\mu_A^\tau f.\end{aligned}$$

1.4. **Variance, covariance, entropy.** The following functionals associated to μ are commonly used in our analysis. For functions $f, g : \Omega \mapsto \mathbb{R}$, any $A \subset [n]$ and boundary condition τ, the covariance of f, g with respect to μ_A^τ is defined as

$$\mathrm{Cov}_A^\tau(f, g) = \mu_A^\tau[fg] - \mu_A^\tau[f]\mu_A^\tau[g].$$

We also write $\mathrm{Cov}_A(f, g)$ for the function $\Omega_{A^c} \ni \tau \mapsto \mathrm{Cov}_A^\tau(f, g)$. Note that

$$\mu\left[\mathrm{Cov}_A(f, g)\right] = \mu\left[(f - \mu_A^\tau[f])(g - \mu_A^\tau[g])\right].$$

When $f = g$ we write $\mathrm{Var}_A^\tau f = \mathrm{Cov}_A^\tau(f, f)$ and $\mathrm{Var}_A f = \mathrm{Cov}_A(f, f)$ for the variance. For any nonnegative function $f : \Omega \mapsto \mathbb{R}_+$, the *entropy* of f with respect to μ_A^τ is defined as

$$\mathrm{Ent}_A^\tau f = \mu_A^\tau[f \log f] - \mu_A^\tau[f] \log \mu_A^\tau[f].$$

Again, we write $\mathrm{Ent}_A f$ for the function $\Omega_{A^c} \ni \tau \mapsto \mathrm{Ent}_A^\tau f$. Note that this function satisfies

$$\mu\left[\mathrm{Ent}_A f\right] = \mu\left[f \log\left(f/\mu_A[f]\right)\right].$$

When $A = [n]$ we simply write $\mathrm{Var} f, \mathrm{Ent} f$ for the functionals $\mathrm{Var}_{[n]} f, \mathrm{Ent}_{[n]} f$. Up to normalization, the entropy functional coincides with the relative entropy: if $\nu = f\mu/\mu[f]$ then

$$\mathrm{Ent} f = \mu[f]\, H(\nu \,|\, \mu),$$

or, equivalently, $\mathrm{Ent}(\nu/\mu) = H(\nu \,|\, \mu)$.

2. Entropy and Markov chain mixing

2.1. **Markov chains and mixing time.** Let Ω_0 denote the support of μ. Let P be a stochastic matrix on Ω_0, that is $P = \{P(\sigma, \eta),\ \sigma, \eta \in \Omega_0\}$ with $P(\sigma, \eta) \geqslant 0$ and $\sum_{\eta \in \Omega_0} = 1$. We assume that P is *regular*, that is there exists $k \in \mathbb{N}$ such that $P^k(\sigma, \eta) > 0$ for all $\sigma, \eta \in \Omega_0$. We also assume that μ is stationary under P, that is

$$\sum_{\sigma \in \Omega_0} \mu(\sigma) P(\sigma, \eta) = \mu(\eta) \,, \qquad \eta \in \Omega_0 \,.$$

Under these assumptions it is a standard fact that μ is the unique stationary distribution and that

$$P^k(\sigma, \eta) \longrightarrow \mu(\eta) \,, \qquad k \to \infty, \quad \forall \sigma, \eta \in \Omega_0 \,.$$

We say that the pair (P, μ) is *reversible* if

$$\mu(\sigma)P(\sigma,\eta) = \mu(\eta)P(\eta,\sigma)\,, \qquad \sigma,\eta \in \Omega_0. \tag{2.1}$$

Clearly, if (P,μ) is reversible then $\mu P = \mu$ and therefore μ is stationary. Note that (2.1) is equivalent to the condition that P is self-adjoint in $L^2(\mu)$ (exercise). Most examples of Markov chains considered below are reversible. However, unless otherwise stated, we do not assume reversibility in what follows.

Example 2.1 (Glauber dynamics for the Ising model on a graph)**.** *Let $G = (V,E)$ be a graph with $V = [n]$, and let $\Omega_0 = \Omega = \{-1,+1\}^n$. Given $\sigma \in \Omega$, $x \in [n]$, define*

$$p_x(\sigma) = \frac{1}{2}\left(1 + \tanh\Big(\beta \sum_{y:\, xy\in E} \sigma_y\Big)\right)$$

Define $P(\sigma,\eta)$, $\sigma,\eta \in \Omega$, by

$$P(\sigma,\eta) = \frac{1}{n}\sum_{x\in[n]} \mathbf{1}_{\eta_y=\sigma_y\,\forall y\neq x}\left(p_x(\sigma)\mathbf{1}_{\eta_x=+1} + (1-p_x(\sigma))\mathbf{1}_{\eta_x=-1}\right)\,.$$

In words, we pick a vertex x uniformly at random, remove its spin σ_x and replace it by a random spin η_x such that $\eta_x = +1$ with probability $p_x(\sigma)$ and $\eta_x = -1$ with probability $1 - p_x(\sigma)$. The matrix P satisfies (2.1) *if μ is the Ising model defined in* (1.4) *(exercise). The chain is regular (exercise).*

Example 2.2 (Sampling proper colorings of a graph)**.** *Let $G = (V,E)$ be a graph with $V = [n]$, and let Ω_0 denote the set of all proper q-colorings of G, where $q \in \mathbb{N}$, $q \geqslant 2$. Given $\sigma \in \Omega_0$, $x \in [n]$, let $\omega_x(\sigma) \subset \{1,\ldots,q\}$ denote the set of colors $a_x \in \{1,\ldots,q\}$ such that $a_x \neq \sigma_y$ for all y neighboring x. Define P by*

$$P(\sigma,\eta) = \frac{1}{n}\sum_{x\in[n]} \frac{\mathbf{1}_{\eta_x\in\omega_x(\sigma),\,\eta_y=\sigma_y\,\forall y\neq x}}{|\omega_x(\sigma)|}\,.$$

In words, we pick a vertex uniformly at random, remove its color and replace it by uniformly random color among the ones that are compatible with the neighbors. The matrix P is symmetric and the stationary distribution μ is uniform over Ω_0. If q is sufficiently large then one can check that P is regular (exercise).

Example 2.3 (Interchange process on a graph)**.** *Let $G = (V,E)$ be a graph with $V = [n]$, and let $\Omega_0 = S_n \subset [n]^n$ denote the set of permutations of $[n]$. Define P by*

$$P(\sigma,\eta) = \frac{1}{2|E|}\sum_{xy\in E} \mathbf{1}_{\eta_z=\sigma_z,\,\forall z\notin xy}\,.$$

In words, we pick an edge uniformly at random, then with probability $1/2$ we swap the labels at xy and with probability $1/2$ we stay put. Note that for any $\sigma \in \Omega_0$, any edge $xy \in E$, there are exactly two configurations $\eta \in \Omega$ such that $\mathbf{1}_{\eta_z=\sigma_z,\forall z\notin xy}$, that is either $\eta = \sigma$ or η is σ swapped at xy. The matrix P is symmetric and the stationary distribution μ is uniform over Ω_0. If the graph G is connected, then one checks that the matrix P is regular (exercise).

The ε-*mixing time* $T_{\mathrm{mix}}(P,\varepsilon)$ of the chain is defined as

$$T_{\mathrm{mix}}(P,\varepsilon) = \max_{\sigma\in\Omega_0}\min\left\{k \in \mathbb{N} \,:\, \|P^k(\sigma,\cdot) - \mu\|_{\mathrm{TV}} \le \varepsilon\right\}.$$

We write $T_{\mathrm{mix}}(P) = T_{\mathrm{mix}}(P, 1/4)$, and call it the *mixing time* of the chain.

Lemma 2.4. *For any $k \in \mathbb{N}$,*

$$\max_{\sigma\in\Omega_0}\|P^k(\sigma,\cdot) - \mu\|_{\mathrm{TV}} \leqslant 2^{-\lfloor k/T_{\mathrm{mix}}(P)\rfloor}\,.$$

Proof. Define

$$d(k) = \max_{\sigma\in\Omega_0} \|P^k(\sigma,\cdot) - \mu\|_{\mathrm{TV}}\,, \qquad \bar d(k) = \max_{\sigma,\eta\in\Omega_0} \|P^k(\sigma,\cdot) - P^k(\eta,\cdot)\|_{\mathrm{TV}}\,.$$

Then $d(k) \leqslant \bar d(k) \leqslant 2d(k)$. Moreover, the Markov property implies $d(\ell k) \leqslant \bar d(k)^\ell$ for all $\ell, k \in \mathbb{N}$. The monotonicity $d(k+1) \leqslant d(k)$ then shows that

$$\begin{aligned} d(k) &\leqslant d(T_{\mathrm{mix}}\lfloor k/T_{\mathrm{mix}}(P)\rfloor) \leqslant \bar d(T_{\mathrm{mix}}\lfloor k/T_{\mathrm{mix}}(P)\rfloor) \\ &\leqslant (2d(T_{\mathrm{mix}}))^{\lfloor k/T_{\mathrm{mix}}(P)\rfloor} \leqslant 2^{-\lfloor k/T_{\mathrm{mix}}(P)\rfloor}\,. \end{aligned}$$

□

We refer to e.g. [37] ans [31] for more background on Markov chain mixing.

2.2. **Entropy contractions and mixing time.** A well established approach to the proof of upper bounds on the mixing time uses the following relations between mixing time and relative entropy contractions.

Definition 2.5. *A Markov chain with transition matrix P and stationary distribution μ has relative entropy contraction with constant $\delta \in (0,1)$ if for all distributions $\nu \in \mathcal{P}(\Omega_0)$,*

$$H(\nu P \,|\, \mu) \leqslant (1-\delta) H(\nu \,|\, \mu). \tag{2.2}$$

Lemma 2.6. *If a Markov chain with transition matrix P and stationary distribution μ has relative entropy contraction with constant $\delta \in (0,1)$, then its ε-mixing time satisfies*

$$T_{\mathrm{mix}}(P,\varepsilon) \leqslant 1 + \frac{1}{\delta}\left[\log\left(\tfrac{1}{2\varepsilon^2}\right) + \log\log\left(\tfrac{1}{\mu_*}\right)\right]\,. \tag{2.3}$$

where $\mu_* = \min_\sigma \mu(\sigma)$.

Proof. Pinsker's inequality (1.2) says that

$$\|\delta_\sigma P^k - \mu\|_{\mathrm{TV}}^2 \leqslant \tfrac12 H(\delta_\sigma P^k \,|\, \mu),$$

where $\delta_\sigma(\tau) = \mathbf{1}(\tau = \sigma)$ is the Dirac mass at σ. Iterating (2.2),

$$\|\delta_\sigma P^k - \mu\|_{\mathrm{TV}}^2 \leqslant \tfrac12 (1-\delta)^k H(\delta_\sigma \,|\, \mu).$$

Since $H(\delta_\sigma \,|\, \mu) = -\log\mu(\sigma)$ and $(1-\delta)^k \leqslant e^{-\delta k}$ we obtain

$$\max_\sigma \|\delta_\sigma P^k - \mu\|_{\mathrm{TV}} \leqslant \varepsilon,$$

as soon as k is an integer such that $k \geqslant \delta^{-1}\log[(2\varepsilon^2)^{-1}\log(1/\mu_*)]$. □

Remark 2.7. *If ν has density f with respect to μ, that is $\nu = f\mu$, then νP has density $P^* f$ with respect to μ, where P^* is the* adjoint *or* time-reversal *matrix $P^*(\sigma,\sigma') = \frac{\mu(\sigma')}{\mu(\sigma)} P(\sigma',\sigma)$. Thus, (2.2) is equivalent to*

$$\mathrm{Ent}(P^* f) \leqslant (1-\delta)\mathrm{Ent}(f), \tag{2.4}$$

for all $f \geqslant 0$ such that $\mu[f] = 1$. By homogeneity, this is equivalent to (2.4) for all $f \geqslant 0$. When P is reversible, that is when $P = P^*$, (2.2) is equivalent to $\mathrm{Ent}(Pf) \leqslant (1-\delta)\mathrm{Ent}(f)$ for all $f \geqslant 0$.

2.3. **Dirichlet forms, spectral gap.** Consider a stochastic matrix P with stationary distribution μ. The *Dirichlet form* associated to the pair (P,μ) is defined as

$$\mathcal{D}_P(f,g) = \langle f, (1-P)g\rangle,$$

where f,g are real functions, and $\langle f,g\rangle = \mu[fg]$ denotes the scalar product in $L^2(\mu)$. Since f is real we also have

$$\mathcal{D}_P(f,f) = \langle (1-Q)f, f\rangle = \frac{1}{2}\sum_{x,y}\mu(x)Q(x,y)(f(x)-f(y))^2,$$

where $Q = \frac{1}{2}(P+P^*)$. Moreover, if $P = P^*$ one has

$$\mathcal{D}_P(f,g) = \frac{1}{2}\sum_{x,y}\mu(x)P(x,y)(f(x)-f(y))(g(x)-g(y)),$$

for all f,g. The *spectral gap* $\mathrm{gap}(P)$ is defined as

$$\mathrm{gap}(P) = \inf_{f:\,\mathrm{Var}f\neq 0}\frac{\mathcal{D}_P(f,f)}{\mathrm{Var}f}.$$

The main use of spectral gap to bound convergence to equilibrium is as follows. Consider the continuous time kernel $K_t = e^{(P-1)t}$, $t \geqslant 0$. This is a stochastic matrix for each t, which defines the Markov chain at time t, that is νK_t is the distribution at time t when started at ν at time 0. A sample of the configuration at time t can be obtained by using the updates from P at the arrival times of a Poisson process with rate 1.

Lemma 2.8. *For any $\nu \in \mathcal{P}(\Omega_0)$, for any $t \geqslant 0$,*

$$\|\nu K_t - \mu\|_{\mathrm{TV}} \leqslant \frac{1}{2\sqrt{\mu_*}}\, e^{-\,\mathrm{gap}(P)\,t}.$$

Proof. Let $f_t = \nu K_t/\mu$, and observe that if $f = \nu/\mu$, then $f_t = K_t^* f$, where

$$K_t^* = e^{(P^*-1)t},$$

is the adjoint of K_t in $L^2(\mu)$. Then a simple computation shows that

$$\frac{d}{dt}\mathrm{Var}f_t = -2\mathcal{D}_P(f_t,f_t) \leqslant -2\,\mathrm{gap}(P)\,\mathrm{Var}f_t\,, \qquad t \geqslant 0.$$

Therefore, integrating,

$$\mathrm{Var}f_t \leqslant e^{-2\,\mathrm{gap}(P)\,t}\,\mathrm{Var}f\,, \qquad t \geqslant 0.$$

By convexity of $\mathrm{Var}(\cdot)$ one has that $\mathrm{Var}f$ is maximized when $\nu = \delta_{\sigma_*}$ with $\sigma_* \in \Omega$ such that $\mu(\sigma_*) = \mu_*$ and therefore $\mathrm{Var}f \leqslant \mathrm{Var}(\delta_{\sigma_*}/\mu) = \mu_*^{-1}(1-\mu_*) \leqslant \mu_*^{-1}$. Therefore, using Schwarz' inequality,

$$\|\nu K_t - \mu\|_{\mathrm{TV}} \leqslant \frac{1}{2}\sqrt{\mathrm{Var}f_t} \leqslant \frac{1}{2\sqrt{\mu_*}}\, e^{-\,\mathrm{gap}(P)\,t}.$$

∎

Note that $\mathrm{gap}(P) = \mathrm{gap}(P^*)$ is the smallest nonzero eigenvalue of the reversible matrix $1-(P+P^*)/2$ (exercise). In particular, when $P = P^*$ and all eigenvalues of P are nonnegative then $\mathrm{gap}(P) = 1-\lambda_2(P)$ where $\lambda_2(P)$ is the maximum eigenvalue of P that is less than 1. Note that if P is reversible but its eigenvalues are not all nonnegative one can always replace P by its lazy version $\hat{P} = (\mathbf{1}+P)/2$, to obtain a reversible chain with nonnegative eigenvalues.

Lemma 2.9. *Suppose that P is reversible. Then*

$$T_{\rm mix}(P,\varepsilon) \leqslant 1 + \frac{1}{{\rm gap}_*(P)} \left[\log\left(\tfrac{1}{2\varepsilon}\right) + \tfrac{1}{2}\log\left(\tfrac{1}{\mu_*}\right)\right], \tag{2.5}$$

where ${\rm gap}_*(P) = 1 - \max\{|\lambda| : \lambda \neq 1$ *eigenvalue of* $P\}$.

Proof. For any $k \in \mathbb{N}$, the spectral decomposition of P shows that ${\rm Var}(P_k f) \leqslant \lambda_*^{2k}{\rm Var} f$, for any f, where $\lambda_* = \max\{|\lambda| : \lambda \neq 1$ *eigenvalue of* $P\}$. Therefore taking $f = \nu/\mu$

$$\|\nu P^k - \mu\|_{\rm TV} \leqslant \frac{1}{2}\sqrt{{\rm Var}(P^k f)} \leqslant \frac{1}{2}\sqrt{{\rm Var} f}\,\lambda_*^k \leqslant \frac{e^{-\,{\rm gap}_*(P)\,k}}{2\sqrt{\mu_*}},$$

where we use $\lambda_* = 1 - {\rm gap}_*(P) \leqslant e^{-\,{\rm gap}_*(P)}$ and ${\rm Var} f \leqslant \mu_*^{-1}$. Thus, if k is an integer larger than $\frac{1}{{\rm gap}_*(P)}\left[\log\left(\frac{1}{2\varepsilon}\right) + \frac{1}{2}\log\left(\frac{1}{\mu_*}\right)\right]$, we have $\|\nu P^k - \mu\|_{\rm TV} \leqslant \varepsilon$. ☐

Clearly, when P is also positive semidefinite we also have ${\rm gap}_*(P) = {\rm gap}(P)$. For the reverse inequality, we note that when P is reversible, one can check that the mixing time is at least the inverse gap up to constants. More precisely, for all irreducible aperiodic chains, Th. 12.5 in [31] shows that

$$T_{\rm mix}(P,\varepsilon) \geqslant \left(\frac{1}{{\rm gap}_*(P)} - 1\right)\log\left(\tfrac{1}{2\varepsilon}\right). \tag{2.6}$$

This does not require reversibility. Indeed the proof goes as follows. Let f be such that $Pf = \lambda f$ with $\lambda \neq 1$ and let $\sigma \in \Omega$ be such that $|f(\sigma)| = \|f\|_\infty$. Note that $\mu(f) = 0$, since $\mu(f) = \mu(Pf) = \lambda\mu(f)$. Then

$$|\lambda|^k\|f\|_\infty = |\lambda^k f(\sigma)| = |P^k f(\sigma)| = |P^k f(\sigma) - \mu(f)| \leqslant 2\|f\|_\infty d(k).$$

Therefore

$$d(k) \geqslant \tfrac{1}{2}\,|\lambda|^k,$$

and $|\lambda|^{T_{\rm mix}(P,\varepsilon)} \leqslant 2\varepsilon$, which implies the claim.

The inequality (2.5) can be compared with (2.3). The latter is much better when δ is comparable with ${\rm gap}(P)$. Indeed, for all examples considered here $\log\log(1/\mu_*) \sim \log n$ and thus when $\delta, {\rm gap}(P)$ have the same order, going from (2.5) to (2.3) improves the bound by a factor $\frac{1}{n}\log n$.

In non-reversible cases the spectral gap is not necessarily useful. For instance one can find stochastic matrices P which mix much faster than the symmetrized $(P + P^*)/2$. The following is an example.

Example 2.10. *Consider a random sequence $Z = \{Z_1, Z_2, \dots\}$ of iid fair coins $\{0,1\}$, and fix an integer $L \in \mathbb{N}$. The Markov chian on $\Omega_L = \{0,1\}^L$ is defined as follows. Given an initial condition $\sigma = (\sigma_1, \dots, \sigma_L) \in \Omega_L$, define the sequence*

$$X_j = \begin{cases} \sigma_j & j \leqslant L \\ Z_{j-L} & j > L \end{cases}$$

and let Y^k denote the configuration $(X_{k+1}, \dots, X_{k+L})$. Then $Y^k \in \Omega_L$, $k = 0, \mathbb{1}, \dots$ is a sequence of random variables which defines a Markov chain on $\Omega = \{0,1\}^L$, with $Y^0 = \sigma$. The associated stochastic matrix P has uniform stationary distribution, it is non-reversible, and has all eigenvalues equal to zero except the trivial eigenvalue equal to 1 (exercise). Note that Y^k has uniform distribution on Ω_L as soon as $k \geqslant L$, so that $T_{\rm mix}(P,\varepsilon) \leqslant L$ for all $\varepsilon > 0$. However, the symmetrized matrix $\bar P = (P + P^*)/2$ has spectral gap proportional to $1/L^2$ (exercise) and therefore $T_{\rm mix}(\bar P,\varepsilon) \geqslant c_\varepsilon L^2$.

2.4. **Log-Sobolev inequalities.** The inequality (2.2) can be considered as a discrete time analogue of the so-called *modified log-Sobolev inequality* characterizing the relative entropy decay for continuous time Markov chains; see, e.g. [9]. Below we discuss some basic relations among (2.2), the standard log-Sobolev inequality and the modified log-Sobolev inequality.

Definition 2.11. *The pair (P,μ) is said to satisfy the (standard) log-Sobolev inequality (LSI) with constant α if for all $f \geqslant 0$:*

$$\mathcal{D}_P(\sqrt{f},\sqrt{f}) \geqslant \alpha \,\mathrm{Ent}_\mu f.$$

It is said to satisfy the modified log-Sobolev inequality (MLSI) with constant ϱ if for all $f \geqslant 0$:

$$\mathcal{D}_P(f,\log f) \geqslant \varrho \,\mathrm{Ent}_\mu f. \tag{2.7}$$

It is well known that the Log-Sobolev inequality characterizes the so-called hypercontractivity for the continuous time kernel $K_t = e^{(P-1)t}$, $t \geqslant 0$ (see [24, Theorem 3.5]), that is

$$\|K_t\|_{2\to\tau(t)} \leqslant 1\,, \qquad \tau(t) = 1 + e^{2\alpha t}\,, \tag{2.8}$$

where $\|K_t\|_{2\to p} = \sup_{\|f\|_2 \leqslant 1} \|K_t f\|_p$ and $\|\cdot\|_p$ denotes the $L^p(\mu)$ norm. In the reversible case one has (2.8) with $\tau(t) = 1 + e^{4\alpha t}$. The reverse implication, namely that (2.8) with $\tau(t) = 1 + e^{4ct}$ implies $\alpha \geqslant c$ always holds.

The modified Log-Sobolev inequality (2.7), on the other hand characterizes the exponential decay of the relative entropy in continuous time for the kernel K_t (see [24, Theorem 3.6]).

Lemma 2.12. *For any $\nu \in \mathcal{P}(\Omega_0)$, for any $t \geqslant 0$,*

$$H(\nu K_t \,|\, \mu) \leqslant H(\nu \,|\, \mu)\, e^{-\varrho\, t}. \tag{2.9}$$

Moreover, if (2.9) *holds for all $\nu \in \mathcal{P}(\Omega_0)$, $t \geqslant 0$, then* (2.7) *holds.*

Proof. We proceed as in the proof of Lemma 2.8. Let $f_t = \nu K_t/\mu = K_t^* f$, where $f = \nu/\mu$. Then

$$\frac{d}{dt}\, H(\nu K_t \,|\, \mu) = \frac{d}{dt}\mathrm{Ent}(K_t^* f) = -\mathcal{D}_P(K_t^* f, \log K_t^* f) \leqslant \, -\,\varrho \,\mathrm{Ent}(K_t^* f).$$

Therefore, integrating,

$$\mathrm{Ent}(K_t^* f) \leqslant e^{-\varrho\, t}\,\mathrm{Ent} f\,, \qquad t \geqslant 0.$$

To prove the converse, note that if (2.9) holds then

$$\mathrm{Ent}(K_t^* f) - \mathrm{Ent} f \leqslant (e^{-\varrho\, t} - 1)\,\mathrm{Ent} f\,, \qquad t > 0.$$

Dividing by t and passing to $t \to 0^+$, one finds (2.7). $\square$

Next, we observe that the bound (2.4) is stronger than the MLSI in (2.7).

Lemma 2.13. *If the relative entropy contraction* (2.2) *holds with constant δ, then* (2.7) *holds with the same constant $\varrho = \delta$.*

Proof. We show that (2.4) implies the MLSI with the same constant δ. By homogeneity, it is sufficient to restrict to the case $\mu[f] = 1$. Suppose that

$$\mathrm{Ent}(P^* f) \leqslant (1-\delta)\mathrm{Ent} f.$$

Since $\mu[P^* f] = \mu[f] = 1$, from the variational principle (1.3) it follows that for any $f \geqslant 0$ with $\mu[f] = 1$,

$$\mu[(P^* f)\log f] \leqslant \mu[(P^* f)\log(P^* f)] = \mathrm{Ent} P^* f.$$

Therefore,

$$\mathcal{D}_P(f, \log f) = \mu[((1 - P^*)f) \log f] \geqslant \mathrm{Ent} f - \mathrm{Ent} P^* f \geqslant \delta\, \mathrm{Ent} f.$$

□

It is well known that MLSI implies spectral gap with gap $\geqslant \varrho/2$.

Lemma 2.14. *For any Markov chain, $\varrho(P) \leqslant 2\,\mathrm{gap}(P)$.*

Proof. This is a standard linearization argument. Take $f_\varepsilon = 1 + \varepsilon g$ with $\mu[g] = 0$, and observe that $\mathrm{Ent} f_\varepsilon = \frac{\varepsilon^2}{2} \mathrm{Var} g + o(\varepsilon^2)$. Similarly, $\mathcal{D}_P(\log f_\varepsilon, f_\varepsilon) = \varepsilon^2\, \mathcal{D}_P(g, g) + o(\varepsilon^2)$. Since g is arbitrary with $\mu[g] = 0$, this implies $\varrho(P) \leqslant 2\,\mathrm{gap}(P)$. □

It is also known that the standard LSI with constant α implies the MLSI $\varrho = 2\alpha$, since $\mathcal{D}_P(f, \log f) \geqslant 2\mathcal{D}_P(\sqrt{f}, \sqrt{f})$ for all $f \geqslant 0$, and this can be improved to $\varrho = 4\alpha$ in the reversible case; see [24, Lemma 2.7]. Here we recall a result of Miclo [36] showing in what sense the LSI implies the discrete time entropy decay.

Lemma 2.15. *If the pair (P^*P, μ) satisfies the standard LSI with constant α, then the relative entropy contraction* (2.2) *holds for (P, μ) with constant $\varrho = \alpha$. In particular, when P is reversible and positive semidefinite, if (P, μ) satisfies the LSI with constant α, then for all $f \geqslant 0$:*

$$\mathrm{Ent}_\mu Pf \leqslant (1 - \alpha)\mathrm{Ent}_\mu f.$$

Proof. The first assertion is proved in [36, Proposition 6]. The second assertion follows from the first and the simple observation that if $P = P^*$ and P is positive semidefinite, then the LSI for (P, μ) implies the LSI for (P^*P, μ) with the same constant since $P^*P = P^2 \leqslant P$ as quadratic forms in $L^2(\mu)$ in this case. □

Summarizing, for reversible chains one has

$$4\alpha_{\mathrm{LSI}} \leqslant \varrho_{\mathrm{MLSI}}\,, \qquad \delta_{\mathrm{EC}} \leqslant \varrho_{\mathrm{MLSI}} \leqslant 2\,\mathrm{gap},$$

and $\alpha_{\mathrm{LSI}} \leqslant \delta_{\mathrm{EC}}$ when P is also positive semidefinite.

Example 2.16. *Consider the trivial Markov chain $P(\sigma, \eta) = \mu(\eta)$ for all $\sigma, \eta \in \Omega$. Then $P_0 f = \mu[f]$ and therefore $\delta_{\mathrm{EC}}(P_0) = \mathrm{gap}(P_0) = 1$. It follows that $\varrho_{\mathrm{MLSI}}(P_0) \geqslant 1$, but exact value is not explicitly known. On the other hand it is known that*

$$\alpha_{\mathrm{LSI}}(P_0) = \frac{1 - 2\mu_*}{\log\left(\frac{1}{\mu_*} - 1\right)}\,, \tag{2.10}$$

see [24, Theorem A.2]*, and the minimum is attained at the indicator function of the configuration σ_* such that $\mu(\sigma_*) = \mu_*$. As a special case, the simple random walk on the complete graph K_n has $\alpha_{\mathrm{LSI}} \sim \log n$.*

We remark that (2.10) shows that for any Markov chain P,

$$\alpha_{\mathrm{LSI}}(P_0)\mathrm{Ent} f \leqslant \mathrm{Var}\sqrt{f} \leqslant \mathrm{gap}(P)^{-1}\mathcal{D}_P(\sqrt{f}, \sqrt{f})\,.$$

Therefore,

$$\alpha_{\mathrm{LSI}}(P) \geqslant \frac{(1 - 2\mu_*)\,\mathrm{gap}(P)}{\log\left(\frac{1}{\mu_*} - 1\right)}\,.$$

3. Entropy decompositions and Block dynamics

3.1. **Heat bath chains.** We now consider a class of Markov chains where at each step a random region $A \subset [n]$ is chosen according to some distribution α_A and the spins $\sigma_A = (\sigma_x, x \in A)$ are updated by taking a sample from the conditional distribution μ_A^τ, where $\tau = \sigma_{A^c}$ is the current configuration outside of A. This type of chain is usually called a *heat bath dynamics* or *Gibbs sampler*. More formally, given nonnegative weightes α_A such that $\sum_{A\subset[n]} \alpha_A = 1$, we write $P = P_\alpha$ where

$$P_\alpha(\sigma, \eta) = \sum_{A\subset V} \alpha_A \, \mu_A^{\sigma_{A^c}}(\eta),$$

or in a more compact form,

$$P_\alpha f = \sum_{A\subset V} \alpha_A \, \mu_A f$$

We call the associated Markov chain the *α-weighted block dynamics*. Some examples:

- When $\alpha_A = \frac{1}{n} \mathbf{1}_{|A|=1}$ this is a single site Gibbs sampler, usually called the *Glauber dynamics*. Example 2.1 coincides with the Glauber dynamics for the Ising model, while Example 2.2 is the Glauber dynamics for proper colorings.
- Consider the Potts model on a bipartite graph. The even/odd chain is obtained by taking $\alpha_E = \alpha_O = 1/2$ where $E \subset [n]$ is the subset of even vertices and $O \subset [n]$ is the subset of odd vertices.
- In the case of Example 2.3 instead we have binary blocks (edges) with weight given by $\alpha_A = \frac{1}{|E|} \mathbf{1}_{A\in E}$.

Exercise 3.1. *For any distribution α as above, the matrix P_α defines a reversible Markov chain with invariant measure μ and Dirichlet form given by*

$$\mathcal{D}_{P_\alpha}(f, g) = \langle f, (1 - P_\alpha)g\rangle = \sum_{A\subset V} \alpha_A \mu \left[\mathrm{Cov}_A(f, g)\right].$$

In particular, P_α is positive semi-definite for any α.

By convexity, it follows that

$$\mathrm{Ent} P_\alpha f \leqslant \sum_{A\subset V} \alpha_A \, \mathrm{Ent} \mu_A f. \tag{3.1}$$

If we are interested in controlling the entropy contraction of P_α we can look for an upper bound on the right hand side in (3.1). More precisely, if

$$\sum_{A\subset V} \alpha_A \, \mathrm{Ent} \mu_A f \leqslant (1 - \delta) \mathrm{Ent} f,$$

then the relative entropy contraction (2.2) for $P = P_\alpha$ holds with constant δ.

3.2. **Useful decompositions.** The following decompositions are commonly used.

Lemma 3.2. *For all $A \subset [n]$,*

$$\mathrm{Ent} f = \mu \left[\mathrm{Ent}_A f\right] + \mathrm{Ent}\, \mu_A[f]. \tag{3.2}$$

More generally, for any $A_1 \subset \cdots \subset A_k \subset [n]$,

$$\mathrm{Ent}_{A_k} f = \sum_{i=1}^{k} \mu_{A_k} \left[\mathrm{Ent}_{A_i} \mu_{A_{i-1}} f\right], \tag{3.3}$$

where we set $A_0 = \emptyset$ and $\mu_{A_0} f = f$. The same decompositions apply to the variance functional, that is when Ent *is replaced by* Var.

Proof. Note that (3.2) is a special case of (3.3) when $k = 2$, $A_1 = A$, $A_2 = [n]$. To prove (3.3) we start by noting that

$$\begin{aligned}\mathrm{Ent}_{A_k} f &= \mu_{A_k}\left[f \log(f/\mu_{A_k} f)\right] \\ &= \mu_{A_k}\left[f \log(f/\mu_{A_{k-1}} f)\right] + \mu_{A_k}\left[f \log(\mu_{A_{k-1}} f/\mu_{A_k} f)\right] \\ &= \mu_{A_k}\left[\mathrm{Ent}_{A_{k-1}} f\right] + \mathrm{Ent}_{A_k} \mu_{A_{k-1}} f\,. \end{aligned} \tag{3.4}$$

We may apply this decomposition again to the first term in (3.4) to obtain

$$\mathrm{Ent}_{A_k} f = \mu_{A_k}\left[\mathrm{Ent}_{A_{k-2}} f\right] + \mu_{A_k}\left[\mathrm{Ent}_{A_{k-1}} \mu_{A_{k-2}} f\right] + \mathrm{Ent}_{A_k} \mu_{A_{k-1}} f\,.$$

Iterating, noting that $\mu_{A_k}\left[\mathrm{Ent}_{A_0} f\right] = 0$, we obtain (3.3). A similar argument proves the statement for the variance. □

4. Product measures

When there is no interaction the spin system takes the form of a product measure. It is instructive to see the form taken by our inequalities in this simple case.

4.1. **Subadditivity and tensorization.** When μ is a product measure, the functionals Var and Ent satisfy a family of simple inequalities. Suppose $\mu = \otimes_{x\in[n]}\mu_x$ is a product measure for some $\mu_x \in \mathcal{P}(\Omega_x)$. Then, for any $\nu \in \mathcal{P}(\Omega)$, letting $\nu_x \in \mathcal{P}(\Omega_x)$ denote the marginal of ν on Ω_x, and taking $g = \sum_{x\in[n]} \log(\nu_x/\mu_x)$ in the variational principle (1.3), one obtains the *subadditivity* of relative entropy

$$H(\nu \,|\, \mu) \geqslant \nu[g] - \log \mu\left[e^g\right] = \sum_{x\in[n]} \nu_x\left[\log(\nu_x/\mu_x)\right] = \sum_{x\in[n]} H(\nu_x \,|\, \mu_x)\,. \tag{4.1}$$

In terms of the entropy functional Ent, this says that for any $f : \Omega \mapsto \mathbb{R}_+$, defining $f_x = \mu_{[n]\setminus\{x\}} f$, one has the subadditivity statement for product measures

$$\sum_{x\in[n]} \mathrm{Ent} f_x \leqslant \mathrm{Ent} f. \tag{4.2}$$

We note that using (1.1) and translating (4.1) one obtains the familiar subadditivity of the Shannon entropy of an arbitrary random vector $X = (X_1, \dots, X_n)$, that is

$$H(X) \leqslant \sum_{i=1}^n H(X_i).$$

Using the decomposition (3.2) for $A = \{x\}$ and averaging over $x \in [n]$ one obtains

$$\mathrm{Ent} f = \frac{1}{n}\sum_{x\in[n]} \mu\left[\mathrm{Ent}_{[n]\setminus\{x\}} f\right] + \frac{1}{n}\sum_{x\in[n]} \mathrm{Ent} f_x\,.$$

Therefore, subadditivity in the form (4.2) is equivalent to

$$\mathrm{Ent} f \leqslant \frac{1}{n-1}\sum_{x\in[n]} \mu\left[\mathrm{Ent}_{[n]\setminus\{x\}} f\right]\,, \tag{4.3}$$

for all $f : \Omega \mapsto \mathbb{R}_+$. Applying the same bound with $[n]$ replaced by $[n] \setminus \{x\}$, and then iterating, shows that

$$\begin{aligned}\mathrm{Ent} f &\leqslant \frac{1}{(n-1)(n-2)} \sum_{x_1\in[n]}\sum_{x_2\in[n]\setminus\{x_1\}} \mu\left[\mathrm{Ent}_{[n]\setminus\{x_1,x_2\}} f\right] \\ &\leqslant \frac{1}{(n-1)!}\sum_{x_1\in[n]}\sum_{x_2\in[n]\setminus\{x_1\}} \cdots \sum_{x_{n-1}\in[n]\setminus\{x_1,\dots,x_{n-2}\}} \mu\left[\mathrm{Ent}_{[n]\setminus\{x_1,\dots,x_{n-1}\}} f\right]\,.\end{aligned}$$

Equivalently,

$$\mathrm{Ent} f \leqslant \sum_{x\in[n]} \mu\left[\mathrm{Ent}_{\{x\}} f\right] \,, \tag{4.4}$$

which is known as the *tensorization* property of entropy for product measures.

4.2. **Shearer inequality.** The inequalities (4.2), (4.3) and (4.4) are special cases of the following much more general estimate.

Lemma 4.1 (Shearer inequality)**.** *Suppose $\mu = \otimes_{x\in[n]}\mu_x$ is a product measure for some $\mu_x \in \mathcal{P}(\Omega_x)$, $x \in [n]$. For all $\alpha = \{\alpha_A,\, A \subset [n]\}$, such that $\alpha_A \geqslant 0$, for all $f : \Omega \mapsto \mathbb{R}_+$,*

$$\gamma(\alpha)\,\mathrm{Ent} f \leqslant \sum_{A\subset[n]} \alpha_A\, \mu\left[\mathrm{Ent}_A f\right],$$

where $\gamma(\alpha) = \min_{x\in[n]} \sum_{A\ni x} \alpha_A$. The same inequality applies to the variance functional, that is when Ent *is replaced by* Var*.*

Proof. For any A, define $A_i = \{x \in A\,,\ x \leqslant i\}$ and $A_i^- = \{x \in A\,,\ x < i\}$. Then (3.3) implies

$$\mathrm{Ent}_A f = \sum_{i\in A} \mu_A \left[\mathrm{Ent}_{A_i}\mu_{A_i^-} f\right].$$

Since μ is a product measure one has $\mu_{A_i} = \mu_{\{i\}} \otimes \mu_{A_i^-}$, and therefore

$$\mu_A \left[\mathrm{Ent}_{A_i}\mu_{A_i^-} f\right] = \mu_A \left[\mathrm{Ent}_{\{i\}}\mu_{A_i^-} f\right]. \tag{4.5}$$

Next, we claim that

$$\mu \left[\mathrm{Ent}_{\{i\}}\mu_{A_i^-} f\right] \geqslant \mu\left[\mathrm{Ent}_{\{i\}}\mu_{[i-1]} f\right], \tag{4.6}$$

where $[i-1] = \{1,\dots,i-1\}$ for $i \geqslant 1$ and $[i-1] = \emptyset$ for $i = 1$. The inequality is actually a consequence of the more general statement that whenever $U, V \subset [n]$, with $U \cap V = \emptyset$ and $\mu_U\mu_V = \mu_V\mu_U$ then for all $f : \Omega \mapsto \mathbb{R}_+$ one has

$$\mu\left[\mathrm{Ent}_U \mu_V f\right] \leqslant \mu\left[\mathrm{Ent}_U f\right]. \tag{4.7}$$

Note that (4.6) follows by applying (4.7) with $U = \{i\}$, $V = [i-1] \setminus A_i^-$ and f replaced by $\mu_{A_i^-} f$. To prove (4.7), we write

$$\begin{aligned}\mu\left[\mathrm{Ent}_U \mu_V f\right] &= \mu\left[(\mu_V f)\log(\mu_V f/\mu_U\mu_V f)\right]\\ &= \mu\left[(\mu_V f)\log(\mu_V f/\mu_V\mu_U f)\right]\\ &= \mu\left[f\log(\mu_V f/\mu_V\mu_U f)\right].\end{aligned}$$

Taking $g = \log(\mu_V f/\mu_V\mu_U f)$, $\nu = f\mu_U$, and observing that $\mu_U[e^g] = 1$, the variational principle (1.3) shows that

$$\mu_U\left[f\log(\mu_V f/\mu_V\mu_U f)\right] = \nu[g] \leqslant H(\nu \,|\, \mu_U) = \mu_U\left[f\log(f/\mu_U f)\right].$$

Integrating one concludes $\mu\left[\mathrm{Ent}_U \mu_V f\right] \leqslant \mu\left[\mathrm{Ent}_U f\right]$. This ends the proof of (4.7). To conclude the lemma, observe that by (3.3) we know that

$$\mathrm{Ent} f = \sum_{i\in[n]} \mu\left[\mathrm{Ent}_{[i]}\mu_{[i-1]} f\right] = \sum_{i\in[n]} \mu\left[\mathrm{Ent}_{\{i\}}\mu_{[i-1]} f\right].$$

Therefore, summing over A in (4.5) one obtains

$$\sum_{A\subset[n]} \alpha_A\, \mu\left[\mathrm{Ent}_A f\right] \geqslant \sum_{A\subset[n]} \alpha_A \sum_{i\in A} \mu\left[\mathrm{Ent}_{\{i\}}\mu_{[i-1]} f\right] \geqslant \gamma(\alpha)\,\mathrm{Ent} f.$$

The proof of the same statement for the variance is left as an exercise. □

Exercise 4.2. *Prove the estimate of Lemma 4.1 for the variance functional.*

4.3. **Shearer inequality for sub-modular functions.** The statement of Lemma 4.1 can be extended as follows. Let $\mathcal{P}_n$ denote the set of all subsets of $[n] = \{1, \dots, n\}$. A function $h : \mathcal{P}_n \mapsto \mathbb{R}$ is called monotone if $h(A) \leqslant h(B)$ whenever $A \subset B$ and it is called *sub-modular* if for all $A, B \in \mathcal{P}_n$,

$$h(A) + h(B) \geqslant h(A \cap B) + h(A \cup B).$$

Exercise 4.3. *Fix $f : \Omega \mapsto \mathbb{R}_+$. Show that both $h(A) = \mu\left[\mathrm{Ent}_A f\right]$ and $h(A) = \mu\left[\mathrm{Var}_A f\right]$ are monotone and sub-modular.*

Lemma 4.4. *For any monotone and sub-modular function h such that $h(\emptyset) = 0$, for any choice of non-negative weights $\alpha = \{\alpha_A,\, A \subset [n]\}$:*

$$\gamma(\alpha) h([n]) \leqslant \sum_A \alpha_A h(A),$$

where $\gamma(\alpha) = \min_{i \in [n]} \sum_{A \ni i} \alpha_A$.

Exercise 4.5. *Prove Lemma 4.4.*

5. Approximate versions in the non-product case

5.1. **Approximate tensorization and approximate subadditivity.** It is natural to apply an approximate version of the above arguments to weakly interacting Gibbs measures, e.g. high temperature Potts models or high q colorings, in which case the distribution μ is not product but somewhat close to a product measure. One can indeed show that such systems with weak interactions satisfy an *approximate tensorization* of the form

$$\mathrm{Ent} f \leqslant C \sum_{x \in [n]} \mu\left[\mathrm{Ent}_{\{x\}} f\right], \tag{5.1}$$

where $C \geqslant 1$ is a constant. This recursive approach was initiated in the 90's with the works of Martinelli, Olivieri, Stroock, H.T. Yau, Zegarlinski [44, 43, 42, 32, 34]. Broadly speaking, the main results of these works can be summarized with the statement that for spin systems on $\mathbb{Z}^d$, with finite or compact spin space, if the system satisfies the *strong spatial mixing condition*, then the approximate tensorization (5.1) holds, see also [13, 35] for related results.

A related problem is the validity of an approximate version of the inequality (4.2), that is the *approximate subadditivity* statement

$$\sum_{x \in [n]} \mathrm{Ent} f_x \leqslant C\, \mathrm{Ent} f, \tag{5.2}$$

where $f_x = \mu_{[n] \setminus \{x\}} f$ as above and $C \geqslant 1$ is a constant. In the case $C = 1$, as we have seen in Section 4.1, (5.2) and (5.1) are equivalent. However, in the general case they are not, and one cannot be deduced from the other even by modifying the constants C involved. A theorem of Carlen and Cordero shows that (5.2) for all $f \geqslant 0$ is equivalent to

$$\mu\left[\prod_{x \in V} \varphi_x(\sigma_x)\right] \leqslant \prod_{x \in V} \mu\left[\varphi_x(\sigma_x)^C\right]^{1/C}, \tag{5.3}$$

for any collection of functions $\varphi_x : V \mapsto \mathbb{R}_+$, where C is the same constant appearing in (5.2). This follows by an application of the variational principle (1.3) for entropy, see [18, Theorem 2.1]. It is interesting to recall a result of Carlen, Lieb, Loss [16] establishing that if μ is the uniform measure on the sphere

$$S^{n-1} = \{\sigma = (\sigma_1, \dots, \sigma_n) \in [-1, 1]^n : \sum_i \sigma_i^2 = 1\},$$

then (5.2) holds with $C = 2$, and the constant 2 is optimal. We refer to Section 7 below for a discussion of the analogous problem for permutations. In Section 6 we show that under a suitable weak dependence assumption a spin system satisfies the approximate subadditivity (5.2) with $C = O(1)$.

5.2. **Block factorization and mixing time of block dynamics.** For any collection of weights α we define C_α as the optimal constant $C_\alpha > 0$ such that for all $f \geqslant 0$,

$$\gamma(\alpha)\,\mathrm{Ent} f \leqslant C_\alpha \sum_{A\subset[n]} \alpha_A\,\mu\,[\mathrm{Ent}_A f]\,, \tag{5.4}$$

where $\gamma(\alpha) = \min_{x\in[n]} \sum_{A\ni x} \alpha_A$. From Lemma 4.1 we know that $C_\alpha = 1$ if μ is a product measure.

Remark 5.1. *For any Gibbs measure μ, and for any α with $\gamma(\alpha) > 0$, the constant C_α must be at least 1. Indeed, let f be a function of σ_x only, where x is such that $\gamma(\alpha) = \sum_{A\ni x} \alpha_A$. Then $\mu\,[\mathrm{Ent}_A f] = 0$ if $A \not\ni x$, and $\mu\,[\mathrm{Ent}_A f] \leqslant \mathrm{Ent} f$ if $A \ni x$. Therefore for this function one has $\sum_{A\subset[n]} \alpha_A\,\mu\,[\mathrm{Ent}_A f] \leqslant \gamma(\alpha)\mathrm{Ent} f$, which implies $C_\alpha \geqslant 1$.*

Remark 5.2. *Note that if $\alpha_A = \frac{1}{n}\mathbf{1}_{|A|=1}$ then C_α is the best constant C in the approximate tensorization statement* (5.1)*, while taking $\alpha_A = \frac{1}{n}\mathbf{1}_{|A|=n-1}$ we obtain that*

$$C_{\mathrm{subadd}} = n - \frac{n-1}{C_\alpha}$$

is the best constant C in the approximate subadditivity statement (5.2)*.*

Definition 5.3. *We say that μ satisfies the* block factorization estimate *with constant C, for short $BF(C)$ if $C_\alpha \leqslant C$ for all weights α.*

Note that $BF(1)$ holds when μ is a product measure.

Lemma 5.4. *For all weights α, for all $f \geqslant 0$,*

$$\mathrm{Ent}(P_\alpha f) \leqslant (1 - \tfrac{\gamma(\alpha)}{C_\alpha})\mathrm{Ent}(f)\,, \quad \textit{and} \quad T_{\mathrm{mix}}(P_\alpha) = O(\tfrac{C_\alpha}{\gamma(\alpha)})\log\log(1/\mu_*)).$$

In particular, if $BF(C)$ holds, then $T_{\mathrm{mix}}(P_\alpha) = O(C/\gamma(\alpha))\log\log(1/\mu_*))$ for all weights α.

Proof. By definition of C_α,

$$\textstyle\sum_\Lambda \alpha_\Lambda \mu[\mathrm{Ent}_\Lambda(f)] \geqslant \frac{\gamma(\alpha)}{C_\alpha}\,\mathrm{Ent}(f).$$

By convexity and Lemma 3.2

$$\begin{aligned}\mathrm{Ent}(P_\alpha f) &\leqslant \textstyle\sum_\Lambda \alpha_\Lambda\,\mu[\mathrm{Ent}(\mu_\Lambda(f))] \\ &= \mathrm{Ent}(f) - \textstyle\sum_\Lambda \alpha_\Lambda \mu[\mathrm{Ent}_\Lambda(f)] \leqslant (1 - \frac{\gamma(\alpha)}{C_\alpha})\mathrm{Ent}(f).\end{aligned}$$

□

Exercise 5.5. *For any α, prove that* $\mathrm{gap}(P_\alpha) \leqslant \gamma(\alpha)$*. Thus, using* (2.6) *one has that the bound on $T_{\mathrm{mix}}(P_\alpha)$ is tight up to $O(C_\alpha \log\log(1/\mu_*)))$.*

6. Block factorizations under spectral independence

The concept of entropy or variance factorization is at the heart of many results on rapid mixing for Markov chains, see e.g. [33, 5, 19, 23]. Roughly speaking these works obtained bounds on spectral gaps or mixing times for the Glauber dynamics (single site updates) by using suitable recursive arguments based on repeated two-blocks factorizations. Here we investigate much more general factorizations in order to obtain bounds on arbitrary block dynamics.

The notion of $BF(C)$ presented above was introduced in [14], where we proved that for any spin system on a graph $G \subset \mathbb{Z}^d$, $d \geqslant 1$, if the *strong spatial mixing* (SSM) holds, then

$BF(C)$ holds for some constant $C \geqslant 1$. The results were then extended to obtain optimal mixing for the Swendsen-Wang dynamics on $\mathbb{Z}^d$ in [6].

The block factorization result of [14] was recently refined in [7], where we consider arbitrary graphs with bounded degrees and assume only *spectral independence*, a property weaker than SSM. Spectral independence was recently introduced in [2]. Under the spectral independence assumption, in [7] a tight mixing time bound was obtained for the Swendsen-Wang dynamics of the ferromagnetic Potts model on arbitrary graphs with bounded degree. We shall review the main results of [7].

6.1. **Spectral independence.** The following matrix captures the pairwise influence between vertices. For a pair of vertices $x, y \in [n]$ and a pair of single spin values $a, a' \in X$, it measures the influence of the spin a at x on a' at y. We define $\mathbb{X} = \{(x, a),\ x \in [n],\ a \in X\}$ as the set of possible pairs of sites and single spin values, that is $\mathbb{X} = [n] \times X$.

Definition 6.1 (Influence matrix [2])**.** *The* ALO influence matrix $J \in \mathbb{R}^{\mathbb{X}\times\mathbb{X}}$ *is defined by* $J(x, a; x, a') = 0$ *for all* x, a, a' *and*

$$J(x, a; y, a') = \mu(\sigma_y = a' \,|\, \sigma_x = a) - \mu(\sigma_y = a') \qquad \text{for } x \neq y.$$

For a boundary condition $\tau \in X^U$ *on some set of vertices* U *(also referred to as a pinning on* U*), the matrix* J^τ *is defined as above when* μ *is replaced by* $\mu^\tau := \mu(\cdot \,|\, \sigma_U = \tau)$*, that the Gibbs measure* μ *conditioned on having the spin configuration* τ *on the set* U*.*

For any pinning τ, all row sums of J^τ vanish, and therefore the matrix J^τ has always the eigenvalue zero. We denote by $\lambda_{\max}(J^\tau) \geqslant 0$ the maximal eigenvalue of J^τ.

Definition 6.2 (Spectral independence [2])**.** *We say that a spin system is* η-spectrally independent *if for all possible pinnings* τ *we have* $\lambda_{\max}(J^\tau) \leqslant \eta$*.*

Given $b > 0$, we say that the Gibbs measure μ is b-marginally bounded if for any pinning τ and for any $a \in X$ such that $\mu^\tau(\sigma_x = a) > 0$ one has $\mu^\tau(\sigma_x = a) \geqslant b$. All systems we consider here are b-marginally bounded for some $b > 0$ uniformly in n, and have an underlying interaction graph with bounded degrees, except for the uniform random permutation model in Example 2.3 which has $b = 1/n$ and mean field type interaction (degree of size n), therefore the following theorem is not interesting in that case. See however Section 7 for specific results in the permutation model case. The following was proved in [7].

Theorem 6.3. *Consider a Gibbs measure* μ *on a graph* $G = ([n], E)$ *with maximum degree* Δ*, and assume that* μ *is* b*-marginally bounded for some* $b > 0$*. If* μ *is* η*-spectrally independent, then* $BF(C)$ *holds with constant* $C = C(\eta, \Delta, b)$*.*

Before getting to the proof of Theorem 6.3 we establish some preliminary facts.

6.2. **Approximate subadditivity and uniform factorizations under spectral independence.** The following theorem was obtained in [7]; see also [21] for an earlier similar result. We use the notation $\mathrm{Av}_{|U|=\ell}$ to denote the uniform average over all subsets $U \subset V := [n]$ such that $|U| = \ell$.

Theorem 6.4. *If the spin system is* η*-spectrally independent and* b*-marginally bounded then there exists a constant* $C = O(1 + \frac{\eta}{b})$ *such that for any* $\ell = \{1, \ldots, n-1\}$ *and for all* $f \geqslant 0$*:*

$$\frac{n}{\ell}\, \mathrm{Av}_{|U|=\ell}\, \mathrm{Ent}(\mu_{V\setminus U} f) \leqslant C\, \mathrm{Ent} f. \tag{6.1}$$

In particular, (5.2) *holds with* $C = O(1 + \frac{\eta}{b})$*. Moreover, for any* $\theta \in (0, 1]$*, there exists* $C = \left(\frac{1}{\theta}\right)^{O(\frac{\eta}{b})}$ *such that for* $\ell = \lceil \theta n \rceil$*:*

$$\frac{\ell}{n}\, \mathrm{Ent} f \leqslant C\, \mathrm{Av}_{|\Lambda|=\ell}\, \mu\left[\mathrm{Ent}_\Lambda f\right]. \tag{6.2}$$

Note that (6.2) is equivalent to saying that for all α of the form $\alpha_\Lambda = \binom{n}{\ell}^{-1}\mathbf{1}_{|\Lambda|=\ell}$, for some ℓ, one has $C_\alpha = C$.

We articulate the proof in two steps. The first step defines the recursive scheme, which allows one to go from a local inequality to a global one; see Lemma 6.7. The second step is a control of the local inequality; see Lemma 6.8. The main line of attack is inspired by several recent papers, among which [1, 3, 25, 22, 21]. This local to global procedure is reminiscent of the recursive approach developed in [15, 12, 11], where similar ideas were used to derive spectral gap estimates for a class of conservative spin systems. The argument here seems to be more robust and, unlike the one in [15, 12, 11], it does not rely on symmetries of the underlying measures.

6.3. **Setting up the recursion.** If $U \subset V$, and $\tau = \tau_U$ a configuration of spins on U, recall that we use notation $\mu^\tau = \mu(\cdot \,|\, \sigma_U = \tau)$ for the conditional distribution $\mu_{V\setminus U}$ when the spins on U are given by τ. Moreover, we write $\mu^{\tau,x} = \mu(\cdot \,|\, \tau \cup \sigma_x)$ if we additionally condition on the spin σ_x at vertex $x \notin U$ and similarly for $\mu^{\tau,x,y} = \mu(\cdot \,|\, \tau \cup \sigma_x \cup \sigma_y)$ for $x, y \notin U$, so that e.g. the expression $\mu^\tau\left[\mathrm{Ent}_{\mu^{\tau,x,y}} f\right]$ indicates the entropy of f with respect to $\mu(\cdot \,|\, \tau \cup \sigma_x \cup \sigma_y)$, that is

$$\mathrm{Ent}_{\mu^{\tau,x,y}} f = \mu^{\tau,x,y}[f \log(f/\mu^{\tau,x,y}(f))] \,,$$

averaged over the two spins σ_x, σ_y sampled according to μ^τ. Define the constants α_k, $k = 0, \dots, n-2$, as the largest numbers such that the inequalities

$$(1+\alpha_k)\,\mathrm{Av}_{\,x\notin U}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) \leqslant \mathrm{Av}_{\,x,y\notin U}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f))\,, \tag{6.3}$$

hold for all $k = 0, \dots, n-2$, for all $U \subset [n]$ with $|U| = k$, for all configurations τ on U and for all functions $f \geqslant 0$. The symbol $\mathrm{Av}_{\,x\notin U}$ denotes the uniform average over all $n-k$ vertices $x \notin U$, and $\mathrm{Av}_{\,x,y\notin U}$ stands for the uniform average over all $(n-k)(n-k-1)$ pairs (x, y) with $x, y \notin U$ and $x \neq y$. We refer to (6.3) as the local inequality, since for each choice of x, y, the distributions involved are concerned with the spins at two vertices only.

Exercise 6.5. *If μ is a product measure, show that*

$$\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) \geqslant \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) + \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,y}(f)). \tag{6.4}$$

Remark 6.6. *Fix $x, y \notin U$. Using $\mu^{\tau,x} f = \mu^{\tau,x}\mu^{\tau,x,y} f$, from Lemma 3.2 we have the decomposition*

$$\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) = \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)) + \mu^\tau\left[\mathrm{Ent}_{\mu^{\tau,x}}(\mu^{\tau,x,y}(f))\right].$$

In particular, $\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) \geqslant \mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f))$ and therefore (6.3) is always true with $\alpha_k = 0$. If μ is a product measure then the subadditivity (6.4) implies the validity of (6.3) with $\alpha_k = 1$ for all $k = 0, \dots, n-2$. In the general case one has $\alpha_k \in [0, 1]$

The recursion is based on the following statement, which rephrases [21, Theorem 5.4].

Lemma 6.7. *Let α_k, $k = 0, \dots, n-2$, be defined by (6.3). Then, for all functions $f \geqslant 0$,*

$$\mathrm{Av}_{\,|U|=j}\mathrm{Ent}(\mu_{V\setminus U} f) \leqslant (1-\kappa_j)\mathrm{Ent}(f), \qquad j = 1, \dots, n-1, \tag{6.5}$$

where

$$\kappa_j = \frac{\sum_{i=j}^{n-1}\Gamma_i}{\sum_{i=0}^{n-1}\Gamma_i}\,, \qquad \Gamma_i = \prod_{k=0}^{i-1}\alpha_k\,, \quad \Gamma_0 = 1.$$

Proof. The claim (6.5) follows from the fact that for all $k = 1, \dots, n-1$:

$$\mathrm{Av}_{\,|U|=k}\,\mathrm{Ent}(\mu_{V\setminus U} f) \leqslant \delta_k\,\mathrm{Av}_{\,|U|=k+1}\,\mathrm{Ent}(\mu_{V\setminus U} f)\,, \qquad \delta_k = \frac{\sum_{i=0}^{k-1}\Gamma_i}{\sum_{i=0}^{k}\Gamma_i}, \tag{6.6}$$

since $\mathrm{Av}_{\,|U|=n}\,\mathrm{Ent}(\mu_{V\setminus U} f) = \mathrm{Ent}(f)$, and $\delta_j\delta_{j+1}\cdots\delta_{n-1} = (1-\kappa_j)$.

To prove (6.6), note that it holds for $k=1$ with $\delta_1 = 1/(1+\alpha_0) = \Gamma_0/(\Gamma_0+\Gamma_1)$ by the assumption (6.3) at $\tau=\emptyset$. Next, we suppose it holds for $0<k-1<n-1$ and show it for k. For any $|U|=k+1$ and $U'\subset U$ with $|U'|=k-1$, setting $\{x,y\}=U\setminus U'$ and letting $\tau=\tau_{U'}$ be the configuration on U', as in Lemma 3.2 we have the decomposition

$$\begin{aligned}\mathrm{Ent}(\mu_{V\setminus U}f) &= \mathrm{Ent}(\mu(\mu_{V\setminus U}f \mid \tau_{U'})) + \mu\left[\mathrm{Ent}(\mu_{V\setminus U}f \mid \tau_{U'})\right]\\ &= \mathrm{Ent}(\mu_{V\setminus U'}f) + \mu\left[\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}f)\right].\end{aligned}$$

Averaging we obtain

$$\begin{aligned}\mathrm{Av}_{|U|=k+1}\mathrm{Ent}(\mu_{V\setminus U}f) =\ &\mathrm{Av}_{|U'|=k-1}\mathrm{Ent}(\mu_{V\setminus U'}f)\\ &+\ \mathrm{Av}_{|U'|=k-1}\,\mathrm{Av}_{x,y\notin U'}\mu\left[\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}f)\right].\end{aligned}$$

In the same way

$$\begin{aligned}\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu_{V\setminus U}f) =\ &\mathrm{Av}_{|U'|=k-1}\mathrm{Ent}(\mu_{V\setminus U'}f)\\ &+\ \mathrm{Av}_{|U'|=k-1}\,\mathrm{Av}_{x\notin U'}\mu\left[\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}f)\right].\end{aligned}$$

From (6.3),

$$\begin{aligned}&\mathrm{Av}_{|U|=k+1}\mathrm{Ent}(\mu_{V\setminus U}f) -\ \mathrm{Av}_{|U'|=k-1}\mathrm{Ent}(\mu_{V\setminus U'}f)\\ &\qquad\geqslant (1+\alpha_{k-1})\,\mathrm{Av}_{|U'|=k-1}\,\mathrm{Av}_{x\notin U'}\mu\left[\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}f)\right]\\ &\qquad= (1+\alpha_{k-1})\left[\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu_{V\setminus U}f) -\ \mathrm{Av}_{|U'|=k-1}\mathrm{Ent}(\mu_{V\setminus U'}f)\right].\end{aligned}$$

Therefore,

$$\mathrm{Av}_{|U|=k+1}\mathrm{Ent}(\mu_{V\setminus U}f) \geqslant (1+\alpha_{k-1})\,\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu_{V\setminus U}f) - \alpha_{k-1}\,\mathrm{Av}_{|U'|=k-1}\mathrm{Ent}(\mu_{V\setminus U'}f).$$

By the inductive assumption (6.6) at $k-1$ we have

$$\begin{aligned}\mathrm{Av}_{|U|=k+1}\mathrm{Ent}(\mu_{V\setminus U}f) &\geqslant (1+\alpha_{k-1}-\alpha_{k-1}\delta_{k-1})\,\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu_{V\setminus U}f)\\ &= \delta_k^{-1}\,\mathrm{Av}_{|U|=k}\mathrm{Ent}(\mu_{V\setminus U}f).\end{aligned}$$

□

6.4. **Estimating the local coefficients.** The next step is an estimate on the coefficients α_k appearing in (6.3).

Lemma 6.8. *If the spin system is η-spectrally independent and b-marginally bounded then the local inequality* (6.3) *holds with*

$$\alpha_k \geqslant 1 - \frac{2\eta}{b(n-k-1)}.$$

Proof. Fix $U\subset V$, $|U|=k\leqslant n-2$ and $\tau=\tau_U$. We may assume $\mu^\tau(f)=1$, which implies $\mu^\tau(\mu^{\tau,x,y}(f)) = \mu^\tau(\mu^{\tau,x}(f)) = 1$ for all $x,y\notin U$. For simplicity, we write $\mathrm{Av}_{x,y}$ and Av_x for the averages $\mathrm{Av}_{x,y\notin U}$ and $\mathrm{Av}_{x\notin U}$. Observe that

$$\begin{aligned}&\mathrm{Av}_{x,y}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) - 2\,\mathrm{Av}_x\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f))\\ &\quad=\ \mathrm{Av}_{x,y}\,\mu^\tau\left[\mu^{\tau,x,y}(f)\log\mu^{\tau,x,y}(f) - \mu^{\tau,x}(f)\log\mu^{\tau,x}(f) - \mu^{\tau,y}(f)\log\mu^{\tau,y}(f)\right]\\ &\quad=\ \mathrm{Av}_{x,y}\,\mu^\tau\left[\mu^{\tau,x,y}(f)\log\frac{\mu^{\tau,x,y}(f)}{\mu^{\tau,x}(f)\mu^{\tau,y}(f)}\right].\end{aligned}$$

Using $a\log(a/b)\geqslant a-b$ for all $a,b\geqslant 0$,

$$\begin{aligned}&\mathrm{Av}_{x,y}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f)) - 2\,\mathrm{Av}_x\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f))\\ &\quad\geqslant 1 -\ \mathrm{Av}_{x,y}\,\mu^\tau\left[\mu^{\tau,x}(f)\mu^{\tau,y}(f)\right]\\ &\quad= -\ \mathrm{Av}_{x,y}\,\mu^\tau\left[(\mu^{\tau,x}(f)-1)(\mu^{\tau,y}(f)-1)\right].\end{aligned}\tag{6.7}$$

We may rewrite

$$\begin{aligned}&\mathrm{Av}_{x,y}\,\mu^\tau\left[(\mu^{\tau,x}(f)-1)(\mu^{\tau,y}(f)-1)\right]\\&\qquad=\frac{1}{n-k-1}\sum_{(x,a)}\nu(x,a)\varphi(x,a)[J^\tau\varphi](x,a),\end{aligned}\tag{6.8}$$

where

$$\varphi(x,a)=\mu^\tau(f\,|\,\sigma_x=a)-1=[\mu^{\tau,x}(f)](a)-1,$$

$\mathbb{X}$ is the set of all pairs (x,a) where $x\in V\setminus U$ (if U is the set where $\tau=\tau_U$ is specified) and $a\in[q]$, ν denotes the probability measure on $\mathbb{X}$ obtained by setting

$$\nu(x,a)=\frac{1}{n-k}\,\mu^\tau(\sigma_x=a),$$

and $J^\tau:\mathbb{X}\times\mathbb{X}\mapsto\mathbb{R}$ denotes the influence matrix from Definition 6.1. Note that in the derivation of (6.8) we have used the fact that for each fixed $y\notin U$ one has

$$\sum_{a'\in[q]}\nu(y,a')\varphi(y,a')=\frac{1}{n-k}\,\mu^\tau(\mu^{\tau,y}(f)-1)=0.$$

Observe that J^τ is self-adjoint in $L^2(\mathbb{X},\nu)$:

$$\nu(x,a)J^\tau(x,a;y,a')=\nu(y,a')J^\tau(y,a';x,a).$$

In particular, its eigenvalues are real. Let $\eta\geqslant 0$ denote its largest eigenvalue (the eigenvalue zero always exists since all row sums of J^τ vanish). Letting $\langle\cdot,\cdot\rangle$ denote the scalar product in $L^2(\mathbb{X},\nu)$ we have $\langle\psi,J^\tau\psi\rangle\leqslant\eta\langle\psi,\psi\rangle$ for all $\psi\in L^2(\mathbb{X},\nu)$. Therefore,

$$\begin{aligned}&\mathrm{Av}_{x,y}\,\mu^\tau\left[(\mu^{\tau,x}(f)-1)(\mu^{\tau,y}(f)-1)\right]\\&\quad=\frac{1}{n-k-1}\langle\varphi,J^\tau\varphi\rangle\leqslant\frac{\eta}{n-k-1}\langle\varphi,\varphi\rangle\\&\quad=\frac{\eta}{n-k-1}\,\mathrm{Av}_x\,\mu^\tau\left[(\mu^{\tau,x}(f)-1)^2\right]=\frac{\eta}{n-k-1}\,\mathrm{Av}_x\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f)).\end{aligned}$$

Recalling (6.7) we have shown

$$\begin{aligned}&\mathrm{Av}_{x,y}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x,y}(f))-2\,\mathrm{Av}_x\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f))\\&\qquad\geqslant-\frac{\eta}{n-k-1}\,\mathrm{Av}_x\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f)).\end{aligned}\tag{6.9}$$

Next, observe that for every fixed $x\notin U$, setting $h^\tau(\sigma_x)=[\mu^{\tau,x}(f)](\sigma_x)$:

$$\begin{aligned}\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f))&=\sum_a\mu^\tau(\sigma_x=a)(h^\tau(a)-1)^2\\&\leqslant\frac{1}{b}\left(\sum_a\mu^\tau(\sigma_x=a)|h^\tau(a)-1|\right)^2\end{aligned}$$

where $b=\min_{x\notin U}\min_a\mu^\tau(\sigma_x=a)$, with the minimum over a restricted to spin values that are allowed at x, that is such that $\mu^\tau(\sigma_x=a)>0$, and we have used $\sum_i a_i^2\leqslant(\sum_i a_i)^2$ for all $a_i\geqslant 0$. Pinsker's inequality shows that

$$\sum_a\mu^\tau(\sigma_x=a)|h^\tau(a)-1|\leqslant\sqrt{2\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f))}.$$

It follows that

$$\mathrm{Var}_{\mu^\tau}(\mu^{\tau,x}(f))\leqslant\frac{2}{b}\,\mathrm{Ent}_{\mu^\tau}(\mu^{\tau,x}(f)).\tag{6.10}$$

Inserting (6.10) into (6.9) concludes the proof. ∎

6.5. **Proof of Theorem 6.4.** From Lemma 6.7, we see that (6.1) holds with $C = \frac{n}{\ell}(1-\kappa_\ell)$. From Lemma 6.8 if follows that

$$\alpha_k \geqslant \max\{1 - R/(n-k-1), 0\}, \qquad R = \lceil 2\eta/b \rceil.$$

Using this bound in the definition of the coefficients κ_ℓ and rearranging, see Section 2.2 of [21], it is not hard to see that for any $1 \leqslant \ell \leqslant n-1$:

$$\kappa_\ell \geqslant \frac{(n-\ell-1)\cdots(n-\ell-R)}{(n-1)\cdots(n-R)}. \tag{6.11}$$

In particular,

$$\frac{n}{\ell}(1-\kappa_\ell) \leqslant \frac{n}{\ell}\left(1 - \frac{(n-\ell-1)\cdots(n-\ell-R)}{(n-1)\cdots(n-R)}\right).$$

Remarkably, the expression in the right hand side above is decreasing with ℓ, and therefore it is always less than $R+1$, its value at $\ell = 1$. This shows that (6.1) holds with $C \leqslant R+1 = O(1+\frac{\eta}{b})$.

To prove (6.2), we start with the decomposition

$$\mathrm{Av}_{|\Lambda|=\ell}\, \mu\,[\mathrm{Ent}_\Lambda f] = \mathrm{Ent}(f) - \mathrm{Av}_{|U|=n-\ell}\, \mathrm{Ent}\left[\mu_{V\setminus U} f\right],$$

which follows from Lemma 3.2. Therefore Lemma 6.7 implies that (6.2) holds with $C = \frac{\ell}{n\,\kappa_{n-\ell}}$. Using (6.11) we see that

$$\frac{\ell}{n\,\kappa_{n-\ell}} \leqslant \frac{(n-1)\cdots(n-R)}{(\ell-1)\cdots(\ell-R)}.$$

In particular, if $\ell = \lceil \theta n \rceil$ with $\theta \in (0,1]$ fixed, then for all sufficiently large n one has $\frac{\ell}{n\,\kappa_{n-\ell}} \leqslant (\frac{1}{\theta})^{O(R)}$. This ends the proof of Theorem 6.4.

6.6. **Proof of Theorem 6.3.** We briefly sketch the main ideas involved in the proof.

1) Note that G is k-partite for some $k \leqslant \Delta+1$ and let $V_1, \dots, V_k$ be disjoint independent sets such that $[n] = \cup_{i=1}^k V_i$. The following lemma reduces the general BF problem to a factorization into the independent sets.

   Lemma 6.9. *If μ satisfies*

   $$\mathrm{Ent} f \leqslant C\,\frac{1}{k}\sum_{i=1}^{k} \mu\,[\mathrm{Ent}_{V_i} f], \tag{6.12}$$

   then BF(C) holds with the same constant C.

2) Use Theorem 6.3 to obtain the ℓ-uniform block factorization (UBF) with $\ell = \lfloor \theta n \rfloor$:

   $$\frac{\ell}{n}\,\mathrm{Ent} f \leqslant C_{\mathrm{UBF}}\,\frac{1}{\binom{n}{\ell}}\sum_{|S|=\ell} \mu\,[\mathrm{Ent}_S f],$$

   where $C_{\mathrm{UBF}} = C_{\mathrm{UBF}}(\eta, b, \theta, \Delta)$.

3) Prove that if θ is small enough, $\theta \leqslant \theta_0(\Delta)$, then the $\lfloor \theta n \rfloor$-UBF implies the k-partite factorization (6.12) with constant $C = C(\theta, \Delta)$. This is a delicate part of the argument, see [7] for the details.

6.7. **Some applications.** We turn to some applications of Theorem 6.3.

6.7.1. *Ising and Potts models.* Consider the Ising Gibbs measure from Example 1.2, in the ferromagnetic case, that is $\beta \geqslant 0$. We define

$$\beta_c(\Delta) := \log\left(\frac{\Delta}{\Delta - 2}\right).$$

This parameter is known to be the threshold of the uniqueness/non-uniqueness phase transition on the Δ-regular tree, see Mossel and Sly [38] and references therein for more details.

Theorem 6.10. *For any $\Delta \geqslant 3$, $0 \leqslant \beta < \beta_c(\Delta)$, there exists a constant $C = C(\beta, \Delta) > 0$ depending only on β, Δ such that the ferromagnetic Ising model on any graph of maximum degree Δ satisfies the block factorization $BF(C)$. In particular, any α-weighted block dynamics has mixing time $T_{\rm mix} = O(\gamma(\alpha)^{-1} \log n)$.*

In the case of Glauber dynamics $\alpha_A = \frac{1}{n}\mathbf{1}_{|A|=1}$ the above theorem predicts the correct $O(n \log n)$ mixing time bound, which was already established in [38]. In light of Lemma 5.4 and Theorem 6.3, to prove Theorem 6.10 all we need to check is that under these assumptions there exists $\eta > 0$ such that the ferromagnetic Ising model on any graph of maximum degree Δ is η-spectrally independent. For a proof of this fact we refer to [20]. In fact, these results up to the critical threshold hold for antiferromagnetic Ising model as well and for the hard-core model in Example 1.5. Again it suffices to prove the spectral independence statement and the rest follows from Lemma 5.4 and Theorem 6.3. We refer to [2, 21] and references therein for the validity of the spectral independence in these cases. The results can also be extended to the ferromagnetic Potts model, that is Example 1.3, with $q \geqslant 3$. However, here the spectral independence is known to hold for sufficiently high temperature only (not up to the critical tree threshold as in the case $q = 2$), see [7, Theorem 4.13] for more details.

6.7.2. *Swendsen-Wang dynamics.* In [8, 7] we extended Theorem 6.10 to obtain optimal mixing bounds for the Swendsen-Wang (SW) dynamics. The latter is defined as follows. Let μ be the Potts distribution on G with configuration space $\Omega = [q]^V$. The SW dynamics takes a spin configuration, transforms it into a "joint" spin-edge configuration, performs a step in the joint space, and then drops the edges to obtain a new Potts configuration. Formally, from a Potts configuration $\sigma_t \in [q]^V$, a transition $\sigma_t \to \sigma_{t+1}$ of the SW dynamics is defined as follows:

1. Let $M_t = M(\sigma_t)$ denote the set of monochromatic edges in σ_t.
2. Independently for each edge $e \in M_t$, keep e with probability $p = 1 - \exp(-\beta)$ and remove e with probability $1 - p$. Let $A_t \subset M_t$ denote the resulting subset.
3. In the subgraph (V, A_t), independently for each connected component C (including isolated vertices), choose a spin s_C uniformly at random from $[q]$ and assign to each vertex in C the spin s_C. This spin assignment defines σ_{t+1}.

We refer to these works for more background and for the details of the proof of the following statement.

Theorem 6.11. *In all cases covered by Theorem 6.10, the Swendsen-Wang dynamics has mixing time $T_{\rm mix} = \Theta(\log n)$. The same conclusions hold for the ferromagnetic Potts model with $q \geqslant 3$, provided β is sufficiently small.*

The proof is based on the concept of *spin-edge factorization* of entropy described as follows. Consider the "joint" Edwards-Sokal distribution for G with parameters $p \in [0, 1]$ and integer $q \geq 2$. This is the probability measure ν on $\Omega_J = \Omega \times \{0, 1\}^E$, the set of "joint" spin-edge configurations (σ, A) consisting of a spin assignment to the vertices $\sigma \in \Omega$ and a subset of edges $A \subset E$, such that

$$\nu(\sigma, A) = \frac{1}{Z_{\rm J}} p^{|A|}(1 - p)^{|E|-|A|}\mathbf{1}(\sigma \sim A),$$

where $\sigma \sim A$ means that $A \subset M(\sigma)$ (i.e., every edge in A is monochromatic in σ) and Z_J is the corresponding normalizing constant or partition function. When $p = 1 - e^{-\beta}$, the "spin marginal" of ν is precisely the Potts distribution μ, while the "edge marginal" of ν corresponds to the random-cluster measure. For a fixed configuration $\sigma \in \Omega$ and subset of edges $A \subset E$, $\mathrm{Ent}_\nu(f \,|\, \sigma)$ and $\mathrm{Ent}_\nu(f \,|\, A)$ denote the entropy of f with respect to the conditional measures $\nu(\cdot \,|\, \sigma)$ and $\nu(\cdot \,|\, A)$, respectively. More precisely, for a given $\sigma \in \Omega$, $\nu(\cdot \,|\, \sigma)$ is the measure ν conditioned on the event that the spin configuration is equal to σ, and for a given $A \subset E$, $\nu(\cdot \,|\, A)$ is the measure ν conditioned on the event that the edge configuration is equal to A. In this way, $\mathrm{Ent}_\nu(f \,|\, \sigma)$ and $\mathrm{Ent}_\nu(f \,|\, A)$ are functions of σ and A, respectively, and $\nu\,[\mathrm{Ent}_\nu(f \,|\, \sigma)]$, $\nu\,[\mathrm{Ent}_\nu(f \,|\, A)]$ denote the corresponding expectations with respect to ν. We say that ν satisfies the spin-edge factorization of entropy with constant C if for all $f : \Omega_J \mapsto \mathbb{R}_+$

$$\mathrm{Ent}_\nu(f) \leqslant C\,\left(\nu\,[\mathrm{Ent}_\nu(f \,|\, \sigma)] + \nu[\mathrm{Ent}_\nu(f \,|\, A)]\right). \tag{6.13}$$

It is possible to show that once (6.13) is available with $C = O(1)$ then the mixing time of SW is $O(\log n)$. Thus, the proof of Theorem 6.11 is reduced to proving (6.13). On the other hand, as shown in [6, 7] one can prove that $BF(C)$, $C = O(1)$, implies the spin-edge factorization of entropy, with possibly different constant from $C' = O(1)$.

6.7.3. *Colorings.* Finally, let us consider proper colorings of a graph as in Example 1.4. An application of Theorem 6.3 in this case yields the following statement.

Theorem 6.12. *If $q > (\frac{11}{6} - \epsilon_0)\Delta$, where $\epsilon_0 \approx 10^{-5} > 0$ is a fixed constant, there exists a constant $C = C(q, \Delta)$ depending only on q, Δ, such that the uniform distribution over q-colorings of any graph of maximum degree Δ satisfies the block factorization $BF(C)$. In particular, any α-weighted block dynamics has mixing time $T_{\mathrm{mix}} = O(\gamma(\alpha)^{-1} \log n)$.*

The reason for the condition $q > (\frac{11}{6} - \epsilon_0)\Delta$ is that under this assumption it is known that an auxiliary dynamics on proper q-colorings, known as the *flip dynamics*, has good contraction properties. The latter, in turn, can be shown to imply spectral independence, and therefore the above theorem follows again from Lemma 5.4 and Theorem 6.3. We refer to [7, Section 4] for the details. It is a major open problem in the field to obtain such results under less restrictive conditions on q.

7. Entropy inequalities for permutations

Consider the uniform distribution over the symmetric group S_n, as in Example (1.6). With A. Bristiel we prove the following sharp ℓ-UBF factorization with explicit constant in this case, see [10].

Theorem 7.1. *For any $\ell = 1, \ldots, n$, for any $f : S_n \mapsto \mathbb{R}_+$*

$$\frac{\ell}{n}\,\mathrm{Ent} f \leqslant \frac{K(n,\ell)}{\binom{n}{\ell}} \sum_{|A|=\ell} \mu\,[\mathrm{Ent}_A f]\,, \qquad K(n,\ell) = \frac{\ell\,\log(n!)}{n\,\log(\ell!)}.$$

The inequality is saturated at any multiple of a Dirac mass at a single permutation, and there are no other extremal functions.

Note that $\ell = 1$ is trivial since fixing all labels except x determines the label at x. Similarly, the case $\ell = n$ is trivial with $K(n, n) = 1$. Using Lemma 3.2, the above theorem is equivalent to the statement that for any $\ell = 1, \ldots, n$, for any $f : S_n \mapsto \mathbb{R}_+$

$$\frac{1}{\binom{n}{\ell}} \sum_{|A|=\ell} \mathrm{Ent}\,\mu_A f \leqslant \left(1 - \frac{\log(\ell!)}{\log(n!)}\right) \mathrm{Ent} f.$$

In particular, taking $\ell = n-1$ shows that for all $f : \mathcal{S}_n \mapsto \mathbb{R}_+$, setting $f_x = \mu_{[n]\setminus\{x\}}$,

$$\sum_{x\in[n]} \mathrm{Ent}\, f_x \leqslant \frac{n\log n}{\log(n!)}\,\mathrm{Ent} f, \tag{7.1}$$

Note that this establishes the subadditivity (5.2) with constant C given by

$$\frac{n\log n}{\log(n!)} \sim 1 + \frac{1}{\log n}.$$

A previous bound by Carlen, Lieb, and Loss has 2 instead of $\frac{n\log n}{\log(n!)}$, see [17]. Proving the optimal bound (7.1) had been an open problem for many years, see Section 7.1 below.

The proof is obtained by using a suitable recursive technique, based on the fact that when a single label σ_x is fixed then the conditional distribution is uniform over permutations of $[n-1]$. The usual decomposition gives

$$\mathrm{Ent} f = \frac{1}{n}\sum_{x\in[n]} \mu\left[\mathrm{Ent}_{[n]\setminus\{x\}} f\right] + \frac{1}{n}\sum_{x\in[n]} \mathrm{Ent} f_x\,,$$

and the first term can be estimated with the inductive hypothesis. The second term however requires work. A similar approach was used in the case $\ell = 2$, that is Random Transpositions, see Lee,Yau [30] for the LSI, and Goel and Guo,Quastel for the MLSI [28, 27]. Note however that none of these works obtains an exact constant as in our estimate. This shows that there is an advantage in considering block factorization constants instead of LSI or MLSI.

7.1. **A permanent bound using entropy subadditivity for permutations.** As a combinatorial application of this result we mention the following sharp upper bound on the permanent of a matrix with arbitrary nonnegative entries, which was independently conjectured by the author, by Carlen, Lieb, Loss [17] and by Samorodnitsky [40]. Let $A = (a_{i,j})$ denote an $n\times n$ matrix, and write

$$\mathrm{perm}(A) = \sum_{\sigma\in\mathcal{S}_n}\prod_{i=1}^{n} a_{i,\sigma_i},$$

for the permanent of A.

Theorem 7.2. *For any $p \geqslant 1$, for any $n\times n$ nonnegative matrix A,*

$$\mathrm{perm}(A) \leqslant \max\left\{1, \frac{n!}{n^{n/p}}\right\}\prod_{i=1}^{n}\|R_i\|_p, \tag{7.2}$$

where R_i denotes the $i-th$ row of A and $\|\cdot\|_p$ denotes the ℓ_p-norm of a vector, and equality is uniquely achieved at either the identity matrix or the all-1 matrix (up to permutation of rows and multiplication by a scalar).

Note that $p_c := \frac{n\log n}{\log(n!)}$ is the value at which the increasing function

$$p \mapsto \frac{n!}{n^{n/p}}$$

takes the value 1, and that the values 1 and $\frac{n!}{n^{n/p}}$ correspond to the case where A is the identity matrix or A is the all-1 matrix respectively, and thus (7.2) is optimal.

The proof is based on the sharp subadditivity result (7.1). The main observation is that (7.1), as we discussed above, is equivalent to (5.3) with $C = p_c$. Consider now the matrix $A = (a_{x,y})$ such that $a_{x,y} = \varphi_x(y)$. The left hand side in (5.3) equals $(1/n!)\mathrm{perm}(A)$, while for every x:

$$\mu\left[\varphi_x(\sigma_x)^{p_c}\right]^{1/p_c} = n^{-1/p_c}\|R_x\|_{p_c} = \left(\frac{1}{n!}\right)^{1/n}\|R_x\|_{p_c},$$

where R_x denotes the x-th row of A. Therefore (5.3) proves the theorem at $p = p_c$. As already noted in [40, Lemma 1], this is sufficient to prove the desired statement for all $p \geqslant 1$.

The use of entropy to prove upper bounds on the permanent goes back to [41, 39]. We refer to [29, 4] for further varations of the Bregman-Minc theorem.

8. Open problems

Problem 8.1. *Can one remove the assumption of bounded degree in Theorem 6.3, that is prove BF for Gibbs measures on arbitrary graphs assuming spectral independence only ? For example, consider the sub-critical Curie-Weiss model (Ising model on complete graph in the uniqueness regime).*

Problem 8.2. *Theorem 7.1 computes, in the case of uniform permutations, the optimal constant C_α as defined in* (5.4) *for all α of the form $\alpha_A = \binom{n}{\ell}^{-1}\mathbf{1}_{|A|=\ell}$. What can be said about other weights α ? What is a relevant version of Shearer inequality for uniformly random permutations ?*

References

[1] Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1198–1211, 2020. 17

[2] N. Anari, K. Liu, and S. Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1319–1330, 2020. 16, 21

[3] Nima Anari, Kuikui Liu, Shayan Oveis Gharan, and Cynthia Vinzant. Log-concave polynomials ii: high-dimensional walks and an fpras for counting bases of a matroid. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 1–12, 2019. 17

[4] Nima Anari and Alireza Rezaei. A tight analysis of Bethe approximation for permanent. *SIAM Journal on Computing*, (0):FOCS19–81, 2021. 24

[5] Lorenzo Bertini, Nicoletta Cancrini, and Filippo Cesi. The spectral gap for a Glauber-type dynamics in a continuous gas. In *Annales de l'IHP Probabilités et statistiques*, volume 38, pages 91–108, 2002. 15

[6] A. Blanca, P. Caputo, D. Parisi, A. Sinclair, and E. Vigoda. Entropy decay in the Swendsen-Wang dynamics. In *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*, 2021. 16, 22

[7] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of markov chains: Coupling, spectral independence, and entropy factorization. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 3670–3692. SIAM, 2022. 16, 20, 21, 22

[8] Antonio Blanca, Pietro Caputo, Daniel Parisi, Alistair Sinclair, and Eric Vigoda. Entropy decay in the Swendsen–Wang dynamics on Z^d. *The Annals of Applied Probability*, 32(2):1018–1057, 2022. 21

[9] Sergey G. Bobkov and Prasad Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.*, 19(2):289–336, 2006. 9

[10] Alexandre Bristiel and Pietro Caputo. Entropy inequalities for random walks and permutations. *arXiv preprint arXiv:2109.06009*, 2021. 22

[11] P. Caputo. Spectral gap inequalities in product spaces with conservation laws. *Stochastic analysis on large scale interacting systems*, 39:53–88, 2004. 17

[12] P. Caputo and F. Martinelli. Relaxation time of anisotropic simple exclusion processes and quantum Heisenberg models. *The Annals of Applied Probability*, 13(2):691–721, 2003. 17

[13] Pietro Caputo, Georg Menz, and Prasad Tetali. Approximate tensorization of entropy at high temperature. In *Annales de la Faculté des sciences de Toulouse: Mathématiques*, volume 24, pages 691–716, 2015. 14

[14] Pietro Caputo and Daniel Parisi. Block factorization of the relative entropy via spatial mixing. *Communications in Mathematical Physics*, 388(2):793–818, 2021. 15, 16

[15] E. A. Carlen, M. C. Carvalho, and M. Loss. Determination of the spectral gap for Kac's master equation and related stochastic evolution. *Acta Mathematica*, 191:1–54, 2003. 17

[16] E. A. Carlen, E. H. Lieb, and M. Loss. A sharp analog of Young's inequality on S^N and related entropy inequalities. *The Journal of Geometric Analysis*, 14:487–520, 2004. 14

[17] Eric Carlen, Elliott H Lieb, and Michael Loss. An inequality of Hadamard type for permanents. *Methods and Applications of Analysis*, 13(1):1–18, 2006. 23

[18] Eric A Carlen and Dario Cordero-Erausquin. Subadditivity of the entropy and its relation to brascamp–lieb type inequalities. *Geometric and Functional Analysis*, 19(2):373–405, 2009. 14

[19] Filippo Cesi. Quasi-factorization of the entropy and logarithmic Sobolev inequalities for Gibbs random fields. *Probab. Theory Related Fields*, 120(4):569–584, 2001. 15

[20] Z. Chen, K. Liu, and E. Vigoda. Rapid mixing of Glauber dynamics up to uniqueness via contraction. In *Proceedings of the 61st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 1307–1318, 2020. 21

[21] Z. Chen, K. Liu, and E. Vigoda. Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. In *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*, 2021. 16, 17, 20, 21

[22] M. Cryan, H. Guo, and G. Mousa. Modified log-Sobolev inequalities for strongly log-concave distributions. *The Annals of Probability*, 49(1):506–525, 2021. 17

[23] Paolo Dai Pra, Anna Maria Paganoni, and Gustavo Posta. Entropy inequalities for unbounded spin systems. *Ann. Probab.*, 30(4):1959–1976, 2002. 15

[24] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *The Annals of Applied Probability*, 9(3):695–750, 1996. 9, 10

[25] Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 974–985. IEEE, 2017. 17

[26] Sacha Friedli and Yvan Velenik. *Statistical mechanics of lattice systems: a concrete mathematical introduction*. Cambridge University Press, 2017. 3

[27] Fuqing Gao and Jeremy Quastel. Exponential decay of entropy in the random transposition and Bernoulli-Laplace models. *Ann. Appl. Probab.*, 13(4):1591–1600, 2003. 23

[28] Sharad Goel. Modified logarithmic Sobolev inequalities for some models of random walk. *Stochastic Process. Appl.*, 114(1):51–79, 2004. 23

[29] Leonid Gurvits and Alex Samorodnitsky. Bounds on the permanent and some applications. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 90–99. IEEE, 2014. 24

[30] Tzong-Yow Lee and Horng-Tzer Yau. Logarithmic Sobolev inequality for some models of random walks. *The Annals of Probability*, 26(4):1855–1873, 1998. 23

[31] D. A. Levin and Y. Peres. *Markov chains and mixing times*. American Mathematical Society, 2017. 6, 8

[32] Sheng Lin Lu and Horng-Tzer Yau. Spectral gap and logarithmic Sobolev inequality for Kawasaki and Glauber dynamics. *Comm. Math. Phys.*, 156(2):399–433, 1993. 14

[33] Fabio Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics (Saint-Flour, 1997)*, volume 1717 of *Lecture Notes in Math.*, pages 93–191. Springer, Berlin, 1999. 15

[34] Fabio Martinelli and Enzo Olivieri. Approach to equilibrium of Glauber dynamics in the one phase region. II. The general case. *Comm. Math. Phys.*, 161(3):487–514, 1994. 14

[35] Katalin Marton. An inequality for relative entropy and logarithmic Sobolev inequalities in Euclidean spaces. *J. Funct. Anal.*, 264(1):34–61, 2013. 14

[36] Laurent Miclo. Remarques sur l?hypercontractivité et l?évolution de l?entropie pour des chaînes de markov finies. pages 136–167, 1997. 10

[37] Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.*, 1(3):x+121, 2006. 6

[38] E. Mossel and A. Sly. Exact thresholds for Ising–Gibbs samplers on general graphs. *The Annals of Probability*, 41(1):294–328, 2013. 21

[39] Jaikumar Radhakrishnan. An entropy proof of Bregman's theorem. *Journal of combinatorial theory, Series A*, 77(1):161–164, 1997. 24

[40] Alex Samorodnitsky. An upper bound for permanents of nonnegative matrices. *Journal of Combinatorial Theory, Series A*, 115(2):279–292, 2008. 23, 24

[41] Alexander Schrijver. A short proof of Minc's conjecture. *Journal of combinatorial theory, Series A*, 25(1):80–83, 1978. 24

[42] Daniel W. Stroock and Boguslaw Zegarliński. The logarithmic Sobolev inequality for continuous spin systems on a lattice. *J. Funct. Anal.*, 104(2):299–326, 1992. 14

[43] Daniel W. Stroock and Boguslaw Zegarliński. The logarithmic Sobolev inequality for discrete spin systems on a lattice. *Comm. Math. Phys.*, 149(1):175–193, 1992. 14

[44] Boguslaw Zegarlinski. Dobrushin uniqueness theorem and logarithmic Sobolev inequalities. *J. Funct. Anal.*, 105(1):77–111, 1992. 14

Department of Mathematics and Physics, Roma Tre University, Largo San Murialdo 1, 00146 Roma, Italy.
E-mail address: pietro.caputo@uniroma3.it