Functional inequalities for Markov Chains and applications

Paolo Dai Pra
Università di Padova

Potsdam March 5-9, 2007
Fruehlingsschule

Outline

Functional inequalities in the context of Markov Chains may be interpreted as a method for estimating quantitatively the rate of convergence to equilibrium. The main purpose of these lectures is to give a self contained introduction to the field, mainly oriented to applications to specific models that, I hope, you will find sufficiently interesting. Some recent developments have made the treatment of many models quite accessible; so I chose to set up the basic notions to get as early as possible to the study of some of those models. I have nevertheless sacrificed to this choice many interesting aspects of functional inequalities, so the course has no hope to be exhaustive. Moreover, we do not mention a number of other approaches to estimate rates of convergence, such as the popular Stein's method (see [12]).

Contents

1 Preliminaries

1.1 Markov chains in discrete-time

Let S be a finite (or countable) set.

Definition 1. A sequence $(X_n)_{n\geq 0}$ of S-valued random variables is a *Markov chain* if for each $n \geq 1$ and $x_0, x_1, \dots x_{n+1} \in S$

$$\begin{aligned} P(X_{n+1} = x_{n+1} | X_n = x_n, \dots, X_1 = x_1, X_0 = x_0) \\ = P(X_{n+1} = x_{n+1} | X_n = x_n). \end{aligned}$$

In the case $P(X_{n+1} = x | X_n = y)$ does not depend on n, we say the chain is *time-homogeneous*. In this case the matrix P

$$P := (p_{yx})_{y,x\in S} \quad \text{with} \quad p_{yx} := P(X_{n+1} = x | X_n = y)$$

is called the *transition matrix*.

Note that

$$\begin{aligned} P(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) \\ = p_{x_{n-1}x_n} p_{x_{n-2}x_{n-1}} \cdots p_{x_0x_1} P(X_0 = x_0) \end{aligned}$$

In other words the initial distribution and the transition matrix determine the law of the process.

1.2 Markov chains in continuous-time

Let $(X_t)_{t\geq 0}$ be S-valued random variables. $\mathcal{F}_s := \sigma(X_s : s \leq t)$.

Definition 2. $(X_t)_{t\geq 0}$ is called a *Markov chain* if for every $x \in S$ and $0 \leq s \leq t$

$$P(X_t = x | \mathcal{F}_s) = P(X_t = x | X_s).$$

If $P(X_t = x | X_s = y)$ depends on s, t only through $t - s$ we say that the chain is *time-homogeneous*.

In this case, for $f : S \to \mathbb{R}$ and $t \geq 0$, define $S_t f : S \to \mathbb{R}$ by

$$S_t f(y) := E(f(X_t) | X_0 = y) = \sum_{x\in S} f(x) P(X_t = x | X_0 = y).$$

Proposition 1.1. *$(S_t)_{t\geq 0}$ is a semigroup, i.e. $S_0 = I$ and $S_{t+s} = S_t \circ S_s$.*

The proof is a simple exercise of use of the Markov property.

Note that S_t is a linear operator from $\mathbb{R}^S$ to itself, so it can be viewed as a matrix $((S_t)_{yx})_{y,x\in S}$, with

$$(S_t)_{yx} = P(X_t = x | X_0 = y)$$

If $t \mapsto S_t$ is continuous, it can be shown that:

$$\lim_{t\downarrow 0} \frac{S_t - I}{t}; +L \quad \text{exists}$$

and

$$S_t = e^{tL}$$

Note that for $x \neq y$

$$0 \leq P(X_t = x | X_0 = y) = (S_t)_{yx} = tL_{yx} + o(t) \;\Rightarrow\; L_{yx} \geq 0$$

while

$$\sum_{x\in S} L_{yx} = \lim_{t\downarrow 0} \sum_{x\in S} \frac{(S_t)_{yx} - \delta_{yx}}{t} = 0$$

and, therefore,

$$L_{yy} = -\sum_{x\neq y} L_{yx}.$$

This implies that

$$Lf(y) = \sum_{x\in S} L_{yx} f(x) = \sum_{y\neq x} L_{yx}[f(x) - f(y)].$$

The distribution of X_0 and the semigroup $(S_t)_{t\geq 0}$ identify the law of the process: for $0 < t_1 < t_2 < \cdots < t_n$ and $x_0, x_1, \ldots, x_n \in S$

$$\begin{aligned}&P(X_{t_n} = x_n, X_{t_{n-1}} = x_{n-1}, \ldots, X_0 = x_0)\\ &\qquad = (S_{t_n - t_{n-1}})_{x_{n-1}x_n} (S_{t_{n-1}-t_{n-2}})_{x_{n-2}x_{n-1}} \cdots (S_{t_1})_{x_0 x_1} P(X_0 = x_0).\end{aligned}$$

In particular, letting $\pi_t(x) := P(X_t = x)$, we have

$$\pi_t(x) = \sum_y P(X_t = x, X_0 = y) = \sum_y (S_t)_{yx} \pi_0(y)$$

and therefore

$$\pi_t = \pi_0 S_t \quad \Longleftrightarrow \quad \begin{cases} \dot{\pi}_t = \pi_t L \\ \pi_0 \end{cases}$$

Definition 3. A probability π on S is called a *stationary distribution* if for every $t \geq 0$ $\pi S_t = \pi$ or, equivalently, $\pi L = 0$

Fact 4. *If S is finite then at least one stationary measure exists. This is not necessarily true if S is countable.*

Definition 5. A Markov chain is said to be *irreducible* if for every $x, y \in S$, $x \neq y$, there exists a "path" $x = x_0, x_1, \ldots, x_n = y$ such that for every $k = 0, \ldots, n-1$, $x_k \neq x_{k+1}$ and $L_{x_k x_{k+1}} > 0$.

Theorem 6. *(Ergodic Theorem). For irreducible chains there exists at most one stationary distribution π. When it exists (in particular for S finite), $\pi(x) > 0$ for every $x \in S$, and for every probability π_0*

$$\pi = \lim_{n\to+\infty} \pi_0 S_t.$$

1.3 Simulation and graphical contruction

Simulation of finite, continuous-time Markov chains involves simulation of random numbers with exponential distribution.

We recall that a *Poisson process* of intensity λ is a random countable subset $N = \{X_1, X_2, \ldots\}$ of $(0, +\infty)$ such that the random variables $(X_k - X_{k-1})_{k\geq 1}$ (with $X_0 = 0$) are i.i.d. $Exp(\lambda)$.

Let $S_2 := \{(x, y) \in S \times S : x \neq y\}$.

For each $(x, y) \in S_2$, let N_{xy} be a Poisson process of intensity L_{xy}. Assume the following conditions hold.

i. For each x fixed, the Poisson processes $(N_{xy})_{y\neq x}$ are independent.

ii. Let I, J be two bounded intervals in $(0, +\infty)$ with $I \cap J = \emptyset$. Then the two families of random variables

$$\{|N_{xy} \cap I| : (x, y) \in S_2\} \quad \text{and} \quad \{|N_{xy} \cap J| : (x, y) \in S_2\}$$

are independent.

iii. Let $s > 0$, and $\theta_s N_{xy}$ be defined by: $t \in \theta_s N_{xy} \iff t + s \in N_{xy}$. Then

$$\{\theta_s N_{xy} : (x, y) \in S_2\} \quad \text{and} \quad \{N_{xy} : (x, y) \in S_2\}$$

are equally distributed.

In particular the conditions above hold when the Poisson processes N_{xy} are all independent. This, however, may be not a convenient choice for simulation.

Then we consider the following *pathwise* updating rule: whenever $t \in N_{xy}$ and $X_{t^-} = x$, then $X_t = y$.

Proposition 1.2. *Under conditions i., ii., iii., the process so constructed is Markov with generator L.*

Let $\omega := (N_{xy})_{(x,y)\in S_2} \in \Omega$, where Ω is provided with the probability P induced by the Poisson processes. ω contains "all the randomness" in the dynamics: if $X_0 = x$ then $X_t = \varphi(t, \omega, x)$, where $\varphi : \mathbb{R}^+ \times \Omega \times S \to S$ is a deterministic function.

Moreover, the following *cocycle* property holds

$$\varphi(t + s, \omega, x) = \varphi(t, \theta_s \omega, \varphi(s, \omega, x)).$$

With this description we are realizing on the same probability space Markov processes with generator L starting from every initial point $x \in S$. In other words $\varphi(t, \omega, \cdot)$ is a Markov process on S^S such that every *one point motion* $t \mapsto \varphi(t, \omega, x)$ is Markov with generator L.

1.4 Samplers. Perfect sampling. Monotonicity.

In many situations Markov chains are used as *samplers*. Let π be a probability on S. The aim is to sample from π, i.e. generate random points with distribution π. For S finite but large, direct sampling from π may be unfeasible. As later examples will show, it is much cheaper to set up an irreducible Markov chain $(X_t)_{t\geq 0}$ with stationary distribution π: for t large X_t is nearly distributed as π, providing an *approximate sample*.

How good this approximation is it depends on how rapidly $\pi_0 S_t$ converges to π. This will be the object of most of the remaining part of these lectures.

In recent years various methods have also been designed to sample *exactly* from π using Markov chains. One of these methods, known as "coupling from the past", reads as follows.

Let $\omega = (N_{xy})_{(x,y)\in S_2}$ as in the previous section, and $\overline{\omega} = (\overline{N}_{xy})_{(x,y)\in S_2}$ be its time-reversed, i.e. $-t \in \overline{N}_{xy} \iff t \in N_{xy}$. Similarly to above, for $t > 0$, we can define the backward time shift θ_{-t} by:

$$s \in \theta_{-t}\overline{N}_{xy} \iff s - t \in \overline{N}_{xy} \; (\iff t - s \in N_{xy}).$$

By simple invariant properties of the Poisson processes, $\theta_{-t}\overline{N}_{xy} \cap [0,t]$ and $N_{xy} \cap [0,t]$ have the same distribution. Moreover the family of point processes $\{\theta_{-t}\overline{N}_{xy} : (x,y) \in S_2\}$ obey conditions i., ii., iii. for the graphical construction restricted to the time-interval $[0,t]$. Thus, the process $s \mapsto \varphi(s, \theta_{-t}\overline{\omega}, x)$ is in $[0,t]$ a Markov process with generator L.

Theorem 7. *Let $T = \inf\{t \geq 0 : \varphi(t, \theta_{-t}\overline{\omega}, \cdot)$ is constant$\}$. Assume $P(T < +\infty) = 1$. Then $\varphi(T, \theta_{-T}\overline{\omega}, x) \sim \pi$ for every $x \in S$.*

Remark 1.1.
1. *The assumption $P(T < +\infty)$ is satisfied when the N_{xy} are all independent, as well as in many other cases, but it does not follow from i., ii., iii.*
2. *The Theorem above provides an explicit algorithm for perfect simulation.*

In general, for S large, it is not feasible to compute $\varphi(t, \theta_{-t}\overline{\omega}, x)$ for *each* $x \in S$.

The algorithm above becomes interesting under the following conditions.

a. S has a partial order, with a maximal and a minimal element: $\underline{x} \leq x \leq \overline{x}$ for every $x \in S$.

b. For every t, ω fixed, $x \mapsto \varphi(t, \omega, x)$ is increasing.

In this case $\varphi(t, \theta_{-t}\overline{\omega}, \cdot)$ is constant if and only if $\varphi(t, \theta_{-t}\overline{\omega}, \underline{x}) = \varphi(t, \theta_{-t}\overline{\omega}, \overline{x})$, i.e. it is enough to keep track of only two trajectories.

When a Markov chain *admits* a graphical construction with this monotonicity property, we say it is *completely monotone*. A more usual but *strictly* weaker notion of monotonicity is when the semigroup S_t preserves increasing functions.

1.5 Reversible dynamics: examples

Recall: π is stationary if and only if $\pi L = o$, i.e.

$$\sum_y \pi(y) L_{yx} = 0 \quad \text{for every } x \in S.$$

A simpler sufficient condition for stationarity is: for every $(x, y) \in S_2$

$$\pi(x) L_{xy} = \pi(y) L_{yx}$$

that is called *Detailed Balance condition* (DB). A chain satisfying (DB) is said to be *reversible.* (DB) is equivalent to either one of the following.

- if $(X_t)_{t\in[0,T]}$ is a stationary Markov chain with generator L and $X_t \sim \pi$, then $(X_{T-t})_{t\in[0,T]}$ is also a Markov chain with generator L.
- L is a symmetric operator in $L^2(\pi)$.

Example 1

$H : S \to \mathbb{R}$ "energy".

$$\pi(x) = \frac{1}{Z} e^{-\beta H(x)} \quad \beta^{-1} = \text{"temperature"}$$

Several choices of generator reversible for π are possible, for example

$$L_{xy} = e^{\frac{\beta}{2}[H(x)-H(y)]} \quad \text{or} \quad L_{xy} = e^{\beta[H(x)-H(y)]^+}$$

One can modify the above examples by choosing a set E of *edges* (unordered pairs of elements of S) such that (S, E) is a connected graph. Then set L_{xy} as above if $\{x, y\}$ is an edge in E, $L_{xy} = 0$ otherwise.

Example 2: spin systems

Λ be a finite subset of $\mathbb{Z}^d$ ($\Lambda \subset\subset \mathbb{Z}^d$), $S = \{-1, 1\}^\Lambda$. For $\eta \in S$ and $\tau \in \{-1, 1\}^{\Lambda^c}$ define $\eta\tau \in \{-1, 1\}^{\mathbb{Z}^d}$ in the obvious way:

$$(\eta\tau)_i = \begin{cases} \eta_i & \text{for } i \in \Lambda \\ \tau_i & \text{otherwise.} \end{cases}$$

For $\eta \in \{-1, 1\}^{\mathbb{Z}^d}$ and $A \subset\subset \mathbb{Z}^d$, η_A denotes its restriction to A. For each such A let $\Phi_A : \{-1, 1\}^A \to \mathbb{R}$. The family $\{\Phi_A : A \subset\subset \mathbb{Z}^d\}$ is called a *summable potential* if

$$\sup_{x\in\mathbb{Z}^d} \sum_{A\ni x} \|\Phi_A\|_\infty < +\infty.$$

For a summable potential we can define, for $\eta \in S$ and $\tau \in \{-1, 1\}^{\Lambda^c}$

$$H_\Lambda^\tau(\eta) := \sum_{A: A\cap\Lambda\neq\emptyset} \Phi_A((\eta\tau)_A)$$

Thus we define the *finite volume Gibbs measure with potential* Φ*, inverse temperature* β*, boundary conditions* τ *on the volume* Λ: $\eta \in S$

$$\nu_\Lambda^\tau(\eta) = \frac{1}{Z_\Lambda^\tau} \exp\left[-\beta H_\Lambda^\tau(\eta)\right].$$

One key property of these families of measure is the fact that they can be interpreted as *conditional distribution*: $\nu_\Lambda^\tau(\eta)$ is the probability of observing the spins η in Λ *conditioned* on the spin τ in Λ^c. In particular, for $A \subseteq \Lambda$

$$\nu_\Lambda^\tau(\eta_A|\eta_{\Lambda\setminus A}) = \nu_A^{\tau\eta_{\Lambda\setminus A}}(\eta_A).$$

where we write $\nu_\Lambda^\tau(\eta_A|\eta_{\Lambda\setminus A})$ for $\nu_\Lambda^\tau\left(\{\xi \in S : \xi_A = \eta_A\}|\eta_{\Lambda\setminus A}\right)$

Elements of S will be called *configurations*. We now define edges (admissible transitions) between configurations. In the sampler we are going to define the only nonzero rates $L_{\eta\sigma}$ are when $\sigma = \eta^i$ for some $i \in \Lambda$, where

$$\eta_j^i = \begin{cases} \eta_j & \text{for } j \neq i \\ -\eta_i & \text{for } j = i \end{cases}$$

We call *spin system* any sample as in Example 1 for the energy H_Λ^τ. For example

$$L_\Lambda^\tau f(\eta) = \sum_{i\in\Lambda} c(i,\eta)\nabla_i f(\eta)$$

where $\nabla_i f(\eta) = f(\eta^i) - f(\eta)$ and $c(i,\eta) = \exp\left[-\frac{\beta}{2}\nabla_i H_\Lambda^\tau(\eta)\right]$.

It is instructive to define a "good" graphical construction of these spin systems. Let $M := \max_{i,\eta} c(i,\eta)$. To each $i \in \Lambda$ we associate a Poisson process N_i of intensity M, and all these processes are independent. Every $t \in N_i$ is a "possible" jump time for η_i. To such time we associate a uniform random variable U_t in $[0,1]$. For $\eta \in S$ we define the process $N_{\eta,i}$ by

$$t \in N_{\eta,i} \iff t \in N_i \text{ and } U_t \leq \frac{c(i,\eta)}{M}.$$

All these r.v. U_t are assumed to be independent. The resulting point processes $N_{\eta,i}$ are Poisson processes of intensity $c(i,\eta)$, that can be used for the graphical construction of the process (i.e. they satisfy properties i.-iii. above).

Example 3: birth and death processes

In this example $S = \mathbb{N}$, and the only admissible transitions are $n \mapsto n+1$ (birth of an individual) and $n \mapsto n-1$ (for $n > 0$ death of an individual). Therefore the generator has the form

$$Lf(n) = a(n)\nabla^+ f(n) + b(n)\nabla^- f(n)$$

with $\nabla^+ f(n) = f(n+1) - f(n)$, $\nabla^- f(n) = f(n-1) - f(n)$. The system is reversible for a probability π if

$$a(n)\pi(n) = b(n+1)\pi(n+1).$$

For example, for $\pi(n) := e^{-\lambda}\frac{\lambda^n}{n!}$, we may choose $a(n) = \lambda$, $b(n) = n$ (a new individual is born with rate λ, and each individual dies with rate 1 independently of others).

Example 4: interacting birth and death processes

Suppose at each site $i \in \Lambda \subset\subset \mathbb{Z}^d$ we have a birth and death process. The simplest case is when these processes are all independent, and each of them has $a(n) = \lambda$, $b(n) = n$. In this case, if η_i is the number of individuals at $i \in L$, and $\eta = (\eta_i)_{i\in\Lambda} \in S = \mathbb{N}^\Lambda$, the generator of the overall process would be

$$Lf(\eta) = \sum_{i\in\Lambda} \left[\lambda \nabla_i^+ f(\eta) + \eta_i \nabla_i^- f(\eta)\right]$$

where

$$\nabla f(\eta) = f(\eta^{i,+}) - f(\eta) \quad \text{with } \eta_j^{i,+} = \begin{cases} \eta_j & \text{for } j \neq i \\ \eta_i + 1 & \text{for } j = i \end{cases}$$

and similarly for $\nabla_i^- f(\eta)$.

The stationary reversible distribution for this independent case is of course a product of Poisson measures:

$$\overline{\pi}(\eta) = e^{-|\Lambda|\lambda} \prod_{i\in\Lambda} \frac{\lambda^{\eta_i}}{\eta_i!}.$$

Now, we may perturb this product measure by an interaction term as follows. Let $K : \mathbb{Z}^d \to \mathbb{R}$ be a *pair interaction*, and

$$\pi(\eta) = \frac{1}{Z}\overline{\pi}(\eta) \exp\left[-\beta \sum_{i,j\in\Lambda} K(i-j)\eta_i\eta_j\right]$$

For $K(z) \geq 0$ for every $z \in \mathbb{Z}^d$, we say the interaction is *repulsive*. A Markov chain that satisfies detailed balance for this π is

$$Lf(\eta) = \sum_{i\in\Lambda} \left[\lambda \exp\left[-\beta \sum_{j\in\Lambda} K(i-j)\eta_j\right] \nabla_i^+ f(\eta) + \eta_i \nabla_i^- f(\eta)\right].$$

Example 5. A system with a conservation law: interacting random walks

Consider the probability π on N^Λ in the previous example but, for a given $N > 0$, we restrict it to the set $S_N := \{\eta : \sum_{i\in\Lambda} \eta_i = N\}$ of configuration with exactly N particles. We obtain the following probability

$$\pi_N(\eta) = \frac{1}{Z_N} \prod_{i\in\Lambda} \frac{1}{\eta_i!} \exp\left[-\beta \sum_{i,j\in\Lambda} K(i-j)\eta_i\eta_j\right].$$

We are now going to define a Markov chain on S_N (that therefore conserves the total number of particles) which is reversible for π_N. The only allowed transitions will be motions of one particle from a site $i \in \Lambda$ to another site $j \in \Lambda$.

Suppose first particles are allowed to jump from $i \in \Lambda$ to *any* $j \in \Lambda$ (model on the complete graph). To define the Markov chain we just have to define the rates $c(i, j, \eta)$ at which a particle at the site i jumps at j when the current state is η. A choice for which detailed balance w.r.t. π_N holds is

$$c(i, j, \eta) = \frac{1}{|\Lambda|}\eta_i \exp\left[\beta \sum_{k\in\Lambda} K(i-k)\eta_k\right].$$

We thus obtained the *complete graph generator*

$$L_{\text{c.g.}} f(\eta) = \sum_{i,j\in\Lambda} c(i, j, \eta)\nabla_{ij} f(\eta)$$

where $\nabla_{ij} f(\eta) = f(\eta^{ij}) - f(\eta)$ with

$$\eta^{ij}_k = \begin{cases} \eta_k & \text{for } k \notin \{i, j\} \text{ or } k \in \{i, j\} \text{ and } \eta_i = 0 \\ \eta_i - 1 & \text{for } k = i \text{ and } \eta_i > 0 \\ \eta_j + 1 & \text{for } k = j \text{ and } \eta_i > 0. \end{cases}$$

A more popular model is that in which particles are only allowed to move to nearest neighbor sites. In this case the generator is (nearest neighbor model)

$$L_{\text{n.n.}} f(\eta) = \sum_{i\sim j} \eta_i \exp\left[\beta \sum_{k\in\Lambda} K(i-k)\eta_k\right] \nabla_{ij} f(\eta),$$

where $i \sim j$ means $|i - j| = 1$.

This Markov chain is reversible for the same probability π_N as for the complete graph model.

Note that for $\beta = 0$ the dynamics reduce to that of N independent, simple symmetric random walks.

Bibliographic remarks.
A good reference on various aspects of discrete-time Markov Chains is the book of E. Behrends [2]. For both discrete and continuous time chains see the book by P. Bremaud [4], which also put emphasis on simulation aspects. For Markov processes in more general state spaces see [14]. Spin systems and many other systems motivated by statistical mechanics can be found in [17]. Perfect sampling was first proposed by J. G. Propp and D. B. Wilson in [18], which is still a good reference, together with the paper by J. Fill [15].

2 Functional inequalities

2.1 Distances from equilibrium

The aim of this section is to provide some basic tools for giving quantitative estimates of the rate of convergence to equilibrium of an irreducible Markov chain.

Let π be the stationary distribution of a Markov chain with semigroup $S_t = e^{tL}$. The Ergodic Theorem establishes that, for each "observable" $f : S \to \mathbb{R}$ and $x \in S$,

$$\lim_{t \to +\infty} S_t f(x) = \sum_y f(y)\pi(y) =: \pi[f].$$

One way to view this convergence is in the space $L^2(\pi)$, where observable are provided with the norm $\|f\|_2 := \sqrt{\pi\left[f^2\right]}$. This suggests to measure distance from equilibrium by

$$\|S_t - \pi\|_{2\to 2} := \sup\{\|S_t f - \pi[f]\|_2 : \|f\|_2 = 1\}.$$

It is interesting to compare this mode of convergence with a more natural one. Given a probability μ on S consider the *total variation distance*

$$\|\mu - \pi\|_{TV} := \sum_{x \in E} |\mu(x) - \pi(x)|$$

It can be shown that, letting $\pi^* := \min_x \pi(x)$,

$$\max_{x \in S} \|\delta_x S_t - \pi\|_{TV} \leq \frac{1}{\pi^*}\|S_t - \pi\|_{2\to 2}.$$

Thus if we have good estimates on how fast $\|S_t - \pi\|_{2\to 2}$ converges to zero, we have estimates on the rate of convergence to equilibrium in total variation. The problem is that, for large S, π^* may be very small.

Another popular notion to measure distance from equilibrium is that of *relative entropy*, defined, for a probability μ on S, by

$$h(\mu|\pi) := \sum_x \pi(x)\left(\frac{\mu(x)}{\pi(x)} \log \frac{\mu(x)}{\pi(x)}\right) = \pi\left[\frac{\mu}{\pi}\log\frac{\mu}{\pi}\right].$$

By Jensen's inequality it is easily shown that $h(\mu|\pi) \geq 0$ and $h(\mu|\pi) = 0$ if and only if $\mu = \pi$.

The so-called Czisar's inequality holds:

$$\|\mu - \pi\|_{TV}^2 \leq h(\mu|\pi).$$

For later use we also introduce the notion of entropy for nonnegative functions $f \geq 0$:

$$Ent_\pi(f) := \pi[f \log f] - \pi[f]\log \pi[f],$$

so that

$$h(\mu|\pi) = Ent_\pi\left(\frac{\mu}{\pi}\right)$$

2.2 Functional inequalities

We are therefore interested in the rate of convergence to zero of quantities as

$$\|S_t - \pi\|_{2\to 2} \quad \text{and} \quad h(\delta_x S_t|\pi).$$

A key notion is that of *Dirichlet form*:

$$\mathcal{E}(f,g) := -\pi\,[fLg]\,.$$

A simple computation shows that

$$\mathcal{E}(f,f) = \frac{1}{2}\sum_{y\in S}\pi\left[L_{xy}(f(y)-f(x))^2\right].$$

Moreover, if the chain is reversible

$$\mathcal{E}(f,g) = \frac{1}{2}\sum_{y\in S}\pi\,[L_{xy}(f(y)-f(x))(g(y)-g(x))]\,.$$

Consider an observable f; an elementary computation yields

$$\frac{d}{dt}\|S_t f - \pi[f]\|_2^2 = -2\mathcal{E}(S_t f, S_t f).$$

The idea is to close the above identity as a differential inequality for $\|S_t f - \pi[f]\|_2^2$. Suppose there is a constant $\gamma > 0$ such that the following inequality holds:

$$(\mathbf{PI}) \qquad Var_\pi[f] := \pi\left[f^2\right] - \pi^2[f] \le \frac{1}{\gamma}\mathcal{E}(f,f)$$

for *every* observable f, which is called the *Poincaré inequality*. Under (**PI**), observing that $\|S_t f - \pi[f]\|_2^2 = Var_\pi(S_t f)$,

$$\frac{d}{dt}\|S_t f - \pi[f]\|_2^2 \le -2\gamma\|S_t f - \pi[f]\|_2^2 \;\Rightarrow\; \|S_t f - \pi[f]\|_2^2 \le e^{-2\gamma t}Var_\pi(f)$$

In particular this shows that

$$\|S_t - \pi\|_{2\to 2} \le e^{-\gamma t}.$$

With a bit of spectral theory one can show:

- if L is reversible, the largest constant γ for which (**PI**) holds is the opposite of the second largest eigenvalue of L (the largest is 0). For this reason this γ is called the *spectral gap*. We use the same terminology in the non-reversible case too: in general $-\gamma$ is the second largest eigenvalue of $\frac{L+L^*}{2}$, where $L^*_{xy} := \frac{L_{yx}\pi(y)}{\pi(x)}$ is the adjoint of L in $L^2(\pi)$.

- If γ is the spectral gap, we have

$$\|S_t - \pi\|_{2\to 2} = e^{-\gamma t}.$$

We also have
$$\max_{x\in S}\|\delta_x S_t - \pi\|_{TV} \le \frac{1}{\pi^*}e^{-\gamma t}.$$

For the decay of the relative entropy a similar argument is used. We first notice that, for any probability μ on S

$$h(\mu S_t|\pi) = Ent_\pi(S_t^* f)$$

where $S_t^* = e^{tL^*}$ and $f = \frac{\mu}{\pi}$. For simplicity, assume that the system is reversible, so $S_t = S_t^*$. A simple computation yields

$$\frac{d}{dt}Ent_\pi(S_t f) = -\mathcal{E}(S_t f, \log S_t f).$$

The inequality that allows to close the former identity is

$$(\mathbf{MLSI}) \qquad Ent_\pi(f) \le \frac{1}{\alpha}\mathcal{E}(f, \log f),$$

with $\alpha > 0$, for every $f > 0$, which implies

$$Ent_\pi(S_t f) \le e^{-\alpha t}Ent_\pi(f) \text{ i.e. } h(\mu S_t|\pi) \le e^{-\alpha t}h(\mu|\pi).$$

The inequality above is called the *modified logarithmic Sobolev inequality.*

Using the simple fact that $h(\delta_x|\pi) = \log\left(\frac{1}{\pi(x)}\right)$, and using Czisar's inequality, we get
$$\max_{x\in S}\|\delta_x S_t - \pi\|_{TV} \le \log\left(\frac{1}{\pi^*}\right)e^{-\alpha t/2}$$
to be compared with the estimate obtained above

$$\max_{x\in S}\|\delta_x S_t - \pi\|_{TV} \le \frac{1}{\pi^*}e^{-\gamma t}.$$

We will show in next section that $\alpha \le 2\gamma$. Thus the estimate obtained with the (**MLSI**) could be worse in the exponential rate, but for moderate times could be much better since, for large S, $\log\left(\frac{1}{\pi^*}\right) \ll \frac{1}{\pi^*}$.

In these lectures we will deal also with a third functional inequality, the *logarithmic Sobolev inequality*:

$$(\mathbf{LSI}) \qquad Ent_\pi(f) \le \frac{1}{s}\mathcal{E}(\sqrt{f}, \sqrt{f})$$

for all $f \ge 0$. The deepest meaning of this inequality will not be dealt with in these lectures. Its interpretation in terms of rate of convergence to equilibrium is not as straightforward as for (**PI**) and (**MLSI**). We only mention the nontrivial fact that one can derive from (**LSI**):

$$\max_{x\in S}\|\delta_x S_t - \pi\|_{TV} \le e\left(\log\frac{1}{\pi^*}\right)^{\frac{\gamma}{2s}}e^{-\gamma t}.$$

This certainly improves

$$\max_{x\in S} \|\delta_x S_t - \pi\|_{TV} \le \frac{1}{\pi^*} e^{-\gamma t}.$$

and it is often better than

$$\max_{x\in S} \|\delta_x S_t - \pi\|_{TV} \le \log\left(\frac{1}{\pi^*}\right) e^{-\alpha t/2}$$

even though we will prove that $2s \le \gamma$.

2.3 Hierarchy of functional inequalities

We show now that the three functional inequalities introduced are, in a suitable sense, hierarchically ordered. We begin with

Lemma 8. *For every $f > 0$*

$$\mathcal{E}(f, \log f) \ge 2\mathcal{E}(\sqrt{f}, \sqrt{f})$$

that, if the chain is reversible, can be improved to

$$\mathcal{E}(f, \log f) \ge 4\mathcal{E}(\sqrt{f}, \sqrt{f})$$

Proof. Using the inequality, for $a, b > 0$, $b(\log a - \log b) \le 2\sqrt{b}\left(\sqrt{a} - \sqrt{b}\right)$, we have

$$\begin{aligned}\mathcal{E}(f, \log f) &= -\sum_x \pi(x) f(x) \sum_y L_{xy} \left[\log f(y) - \log f(x)\right] \\ &\ge -2\sum_x \pi(x)\sqrt{f(x)} \sum_y L_{xy}\left[\sqrt{f(y)} - \sqrt{f(x)}\right] = 2\mathcal{E}(\sqrt{f}, \sqrt{f}).\end{aligned}$$

In the reversible case the proof is similar, but uses the nicer expression for the Dirichet form off the diagonal:

$$\mathcal{E}(f, \log f) = \sum_x \pi(x) \sum_y L_{xy} \left[f(y) - f(x)\right]\left[\log f(y) - \log f(x)\right],$$

and that one uses the inequality

$$(\log a - \log b)(a - b) \ge 4\left(\sqrt{a} - \sqrt{b}\right)^2.$$

Theorem 9. *Consider an irreducible, finite state Markov chain, and let γ, α, s denote the largest constants in (**PI**), (**MLSI**) and (**LSI**) respectively. Then*

$$s > 0 \quad \alpha \ge 2s \quad 2\gamma \ge \alpha \quad \gamma \ge 2s.$$

For reversible chains these improve to $2\gamma \ge \alpha \ge 4s$.

Proof. The inequality $\alpha \geq 2s$ ($\alpha \geq 4s$ in the reversible case) follows immediately from last Lemma. To prove $2\gamma \geq \alpha$ and $\gamma \geq 2s$, consider a function g, and define, for $\varepsilon > 0$, $f_\varepsilon := (1 + \varepsilon g)^2$. By elementary calculus one gets

$$\begin{aligned} Ent_\pi(f_\varepsilon) &= 2\varepsilon^2 Var_\pi(g) + o(\varepsilon^2) \\ \mathcal{E}(\sqrt{f_\varepsilon}, \sqrt{f_\varepsilon}) &= \varepsilon^2 \mathcal{E}(g, g) \\ \mathcal{E}(f_\varepsilon, \log f_\varepsilon) &= 4\varepsilon^2 \mathcal{E}(g, g) + o(\varepsilon^2) \end{aligned}$$

Thus

$$\frac{1}{\varepsilon^2} Ent_\pi(f_\varepsilon) \leq \frac{1}{\alpha\varepsilon^2} \mathcal{E}(f_\varepsilon, \log f_\varepsilon) \ \forall \varepsilon > 0 \ \Rightarrow \ 2Var_\pi(g) \leq \frac{4}{\alpha} \mathcal{E}(g, g) \\ \Rightarrow \ 2\gamma \geq \alpha$$

The inequality $\gamma \geq 2s$ is proved similarly.

So we are left to prove that, for a finite, irreducible Markov chain, $s > 0$, which by the inequalities above implies $\gamma, \alpha > 0$. By definition of (**LSI**)

$$s = \inf \left\{ \frac{\mathcal{E}(\sqrt{f}, \sqrt{f})}{Ent_\pi(f)} : f \geq 0, \ Ent_\pi(f) > 0 \right\}$$

Since both $\mathcal{E}(\sqrt{f}, \sqrt{f})$ and $Ent_\pi(f)$ are continuous and homogeneous of degree 1,

$$s = \inf \left\{ \frac{\mathcal{E}(\sqrt{f}, \sqrt{f})}{Ent_\pi(f)} : f \geq 0, \ Ent_\pi(f) = 1 \right\}$$

By irreducibility: $Ent_\pi(f) = 1 \Rightarrow f \neq$ const. $\Rightarrow \mathcal{E}(\sqrt{f}, \sqrt{f}) > 0$. Moreover, since S is finite, observables belong to a finite dimensional space, and it is easy to show that $\{f : Ent_\pi(f) = 1\}$ is compact. Thus s is the infimum over a compact set of a continuous function that is strictly positive on that set. Thus the infimum is a minimum, and therefore strictly positive.

Remark 2.1. *For Markov chains with infinite state space the inequalities in the above Lemma still hold, but the constants may be zero.*

2.4 Tensor property of functional inequalities

The aim of this section is to explain the behavior of functional inequalities in the case our Markov chain is a finite family of independent Markov chains. So assume $\pi = \mu^{\otimes N}$ be a product probability on S^N. A function $f : S^N \to \mathbb{R}$ can be viewed as a function of N variables $x_1, x_2, \ldots, x_N$. When we think of f as a function of x_i with all other variables "frozen", we write f_i. So, for instance,

$$\mu[f_i] = \sum_{x \in S} f(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_N)\mu(x).$$

Accordingly, we define $Var_\mu(f_i)$ and $Ent_\mu(f_i)$.

Proposition 2.1.

$$\begin{aligned} Var_\pi(f) &\leq \sum_{i=1}^N \pi\left[Var_\mu(f_i)\right] \\ Ent_\pi(f) &\leq \sum_{i=1}^N \pi\left[Ent_\mu(f_i)\right] \end{aligned}$$

for every f ($f \geq 0$ for the second inequality).

We will only prove the second inequality. The first can be shown along the same lines, and it is slightly simpler. We will obtain the inequality from a stronger result, that will be useful again later.

Lemma 10. *Let $(\Omega, \mathcal{F}, \pi)$ be a probability space, $\mathcal{F}_1, \mathcal{F}_2$ be two sub-σ-fields of $\mathcal{F}$, $f \geq 0$ be a measurable function with $\pi[f] = 1$ and $f \log f \in L^1(\pi)$. Then*

$$Ent_\pi(f) \leq \pi\left[Ent_\pi(f|\mathcal{F}_1)\right] + \pi\left[Ent_\pi(f|\mathcal{F}_2)\right] + \log \pi\left[\pi[f|\mathcal{F}_1]\pi[f|\mathcal{F}_2]\right]$$

where $Ent_\pi(f|\mathcal{F}_i) := \pi\left[f \log f|\mathcal{F}_i\right] - \pi[f|\mathcal{F}_i] \log \pi[f|\mathcal{F}_i]$.

We first note that from the Lemma it follows that, for $N = 2$

$$Ent_\pi(f) \leq \pi\left[Ent_\pi(f_1)\right] + \pi\left[Ent_\pi(f_2)\right],$$

since it is enough to take $\mathcal{F}_i$ to be the σ-field generated by the projection on the i-th component. In this case the independence of $\mathcal{F}_1$ and $\mathcal{F}_2$ easily implies that $\log \pi\left[\pi[f|\mathcal{F}_1)\pi[f|\mathcal{F}_2)\right] = 0$. For $N > 2$ one just proceeds by induction.

Proof of the Lemma. It is enough (!) to observe that

$$\begin{aligned} &\pi\left[Ent_\pi(f|\mathcal{F}_1)\right] + \pi\left[Ent_\pi(f|\mathcal{F}_2)\right] + \log \pi\left[\pi[f|\mathcal{F}_1]\pi[f|\mathcal{F}_2]\right] \\ &\qquad\qquad\qquad\qquad = Ent_\pi(f) + h(\nu_1|\nu_2) \end{aligned}$$

where

$$\begin{aligned} d\nu_1 &:= f d\pi \\ d\nu_2 &:= \frac{\pi[f|\mathcal{F}_1]\pi[f|\mathcal{F}_2]}{\pi\left[\pi[f|\mathcal{F}_1]\pi[f|\mathcal{F}_2]\right]} d\pi \end{aligned}$$

□

Now, let L be the generator of an irreducible Markov chain on S, and consider N independent copies of this chain (possibly with different starting points). It is easy to show that this process on S^N is also an irreducible Markov chain, and that its generator L_N is given by

$$L_N f = \sum_{i=1}^N L f_i$$

If μ is the stationary distribution for L, then $\pi := \mu^{\otimes N}$ is the stationary distribution for L_N.

Thus, if we denote by $\mathcal{E}_N$ the Dirichlet form of the product process, and $\mathcal{E}$ the one of the single component, we have

$$\mathcal{E}_N(f,g) = \sum_{i=1}^{N} \pi\left[\mathcal{E}(f_i, g_i)\right].$$

Theorem 11. *Let $K(L)$ (resp $K(L_N)$) be one of the three best constants in (**PI**), (**MLSI**) or (**LSI**) for L (resp. L_N). Then*

$$K(L_N) = K(L)$$

Proof. We give it for (**LSI**), the others are the same.

$$Ent_\pi(f) \leq \sum_{i=1}^{N} \pi\left[Ent_\mu(f_i)\right] \leq \sum_{i=1}^{N} \pi\left[\frac{1}{s(L)}\mathcal{E}(\sqrt{f_i}, \sqrt{f_i})\right]$$
$$= \frac{1}{s(L)}\mathcal{E}_N(\sqrt{f}, \sqrt{f})$$

which implies $s(L_N) \geq s(L)$. To get equality it is enough to take functions of only one variable.

Bibliographic remarks.
What contained in this section, and much more, con be found in [13]. The proof of Lemma 10 is in [11].

3 Applications

3.1 (LSI) for spin systems at high temperature: the method of bisection

We begin by recalling the model: $S = \{-1,1\}^\Lambda$,

$$H_\Lambda^\tau(\eta) := \sum_{A:A\cap\Lambda\neq\emptyset} \Phi_A((\eta\tau)_A) \qquad \pi(\eta) = \nu_\Lambda^\tau(\eta) = \frac{1}{Z_\Lambda^\tau}\exp\left[-\beta H_\Lambda^\tau(\eta)\right].$$

$$L_\Lambda^\tau f(\eta) = \sum_{i\in\Lambda} c(i,\eta)\nabla_i f(\eta)$$

where $\nabla_i f(\eta) = f(\eta^i) - f(\eta)$ and $c(i,\eta) = \exp\left[-\frac{\beta}{2}\nabla_i H_\Lambda^\tau(\eta)\right]$. For $\beta = 0$ we just have a family of independent Markov Chains, one for each site of Λ. In this case, by the tensor property, $s(L_\Lambda^\tau)$ is independent of Λ (and of τ, of course). Is there a way to partially extend this property (i.e. proving $\inf_{\Lambda,\tau} s(L_\Lambda^\tau) > 0$) to $\beta > 0$ small enough?

In what follows we assume the potential to be *finite range* and *translation invariant*:

$$\begin{aligned}\Phi_A &\equiv 0 \ \text{ if } \operatorname{diam}(A) \geq R \\ \Phi_{A+i}(\eta_{A+i}) &= \Phi_A(\eta_{A+i}).\end{aligned}$$

By the finite range property, for Λ fixed, the families $\{\nu_\Lambda^\tau : \tau \in \{-1,1\}^{\Lambda^c}\}$ and $\{L_\Lambda^\tau : \tau \in \{-1,1\}^{\Lambda^c}\}$ are finite. In particular

$$s_\Lambda := \inf\{s(L_\Lambda^\tau) : \tau \in \{-1,1\}^{\Lambda^c}\} > 0.$$

Suppose, for simplicity, $d = 2$. Let R_{L_1,L_2} be any rectangle in $\mathbb{Z}^2$ with side lengths L_1 and L_2. By translation invariance $s_{R_{L_1,L_2}}$ is invariant by translation of R_{L_1,L_2}, so it is not necessary to specify the position in space of R_{L_1,L_2}.

Let

$$s_{L_1,L_2} := \inf\{s_\Lambda : \Lambda \subseteq R_{L_1,L_2}\}.$$

We will show that

$$\inf_{L_1,L_2} s_{L_1,L_2} > 0$$

by setting up an induction in L_1, L_2.

Consider $\Lambda := R_{\frac{3}{2}L_1,L_2}$ and write it as the union of two overlapping rectangles A, B with first side length $\leq L_1$, second side length $= L_2$ and $A \cap B$ is a rectangle with sides $\sqrt{L_1}, L_2$ (we should take the integer part of $\sqrt{L_1}$, but we ignore this trivial complication).

Let $\mathcal{F}_1$ be the σ-field generated by projection onto $\{-1,1\}^{A\setminus B}$, and $\mathcal{F}_2$ be the σ-field generated by projection onto $\{-1,1\}^{B\setminus A}$. Since Gibbs measures are conditional measures,

$$\nu_\Lambda^\tau(\cdot|\mathcal{F}_1) = \nu_B^{\tau\eta_{A\setminus B}}(\cdot)$$

Also, remember the inequality, for $\nu := \nu_\Lambda^\tau$

$$\begin{aligned}Ent_\nu(f) &\leq \nu\left[Ent_\nu(f|\mathcal{F}_1) + Ent_\nu(f|\mathcal{F}_2)\right] + \log\nu\left[\nu(f|\mathcal{F}_1)\nu(f|\mathcal{F}_2)\right] \\ &= \nu\left[Ent_{\nu_B^\cdot}(f) + Ent_{\nu_A^\cdot}(f)\right] + \log\nu\left[\nu_A^\cdot(f)\nu_B^\cdot(f)\right]\end{aligned}$$

But

$$Ent_{\nu_B^\cdot}(f) \leq s_{L_1,L_2}^{-1}\nu_B^\cdot\left[\sum_{i\in B} c(i,\eta)\left(\nabla_i\sqrt{f(\eta)}\right)^2\right]$$

since $B \subseteq R_{L_1,L_2}$ this last expression is the Dirichlet form in the volume B evaluated at $(\sqrt{f}, \sqrt{f})$.

Thus, we can proceed from

$$Ent_\nu(f) \leq \nu\left[Ent_{\nu_B^\cdot}(f) + Ent_{\nu_A^\cdot}(f)\right] + \log\nu\left[\nu_A^\cdot(f)\nu_B^\cdot(f)\right]$$

by noticing that

$$
\begin{aligned}
&\nu\left[Ent_{\nu_B^\cdot}(f)+Ent_{\nu_A^\cdot}(f)\right] \\
&\qquad\qquad \frac{1}{2}\le s_{L_1,L_2}^{-1}\nu\left[\nu_B^\cdot\left[\sum_{i\in B}c(i,\eta)\left(\nabla_i\sqrt{f(\eta)}\right)^2\right]\right. \\
&\qquad\qquad\qquad \left.+\,\nu_A^\cdot\left[\sum_{i\in A}c(i,\eta)\left(\nabla_i\sqrt{f(\eta)}\right)^2\right]\right] \\
&\qquad\qquad = s_{L_1,L_2}^{-1}\mathcal{E}_\Lambda^\tau(\sqrt{f},\sqrt{f})+s_{L_1,L_2}^{-1}\nu\left[\sum_{i\in A\cap B}c(i,\eta)\left(\nabla_i\sqrt{f(\eta)}\right)^2\right]
\end{aligned}
$$

Let temporarily forget the term $\nu\left[\sum_{i\in A\cap B}c(i,\eta)\left(\nabla_i\sqrt{f(\eta)}\right)^2\right]$, and concentrate on the term

$$\log\nu\left[\nu_A^\cdot(f)\nu_B^\cdot(f)\right]$$

The idea is that, for β small, $\nu_A^\cdot(f)$ and $\nu_B^\cdot(f)$ are almost independent under ν. Remember that $\mathcal{F}_1$ is the σ-field generated by projection onto $\{-1,1\}^{A\setminus B}$, and $\mathcal{F}_2$ is the σ-field generated by projection onto $\{-1,1\}^{B\setminus A}$.

Lemma 12. *Suppose there is a measure $\overline{\nu}$ on $\{-1,1\}^\Lambda$ such that $\overline{\nu}_{|\mathcal{F}_i}=\nu_{|\mathcal{F}_i}$ for $i=1,2$, and $\mathcal{F}_1,\mathcal{F}_2$ are independent under $\overline{\nu}$. Set $h:=\frac{\overline{\nu}}{\nu}$. Then, for each $f\ge 0$ with $\nu[f]=1$*

$$\log\nu\left[\nu_A^\cdot(f)\nu_B^\cdot(f)\right]\le 4\|h-1\|_\infty Ent_\nu(f).$$

Proof. First notice that

$$
\begin{aligned}
\log\nu\left[\nu[f|\mathcal{F}_1]\nu[f|\mathcal{F}_2]\right] &= \log\overline{\nu}\left[h\nu[f|\mathcal{F}_1]\nu[f|\mathcal{F}_2]\right] \\
&= \log\left\{\overline{\nu}\left[(h-1)\nu[f|\mathcal{F}_1]\nu[f|\mathcal{F}_2]\right]+\overline{\nu}\left[\nu[f|\mathcal{F}_1]\nu[f|\mathcal{F}_2]\right]\right\} \\
&= \log\left\{\overline{\nu}\left[(h-1)\nu[f|\mathcal{F}_1]\nu[f|\mathcal{F}_2]\right]+1\right\} \\
&\le \overline{\nu}\left[(h-1)\nu[f|\mathcal{F}_1]\nu[f|\mathcal{F}_2]\right],
\end{aligned}
$$

where we used the inequality $\log(1+x)\le x$. Therefore

$$
\begin{aligned}
&\log\nu\left[\nu[f|\mathcal{F}_1]\nu[f|\mathcal{F}_2]\right]\le\overline{\nu}\left[(h-1)\nu[f|\mathcal{F}_1]\nu[f|\mathcal{F}_2]\right] \\
&\quad=\overline{\nu}\left[(h-1)\left[\nu[f|\mathcal{F}_1]-\nu\left(\sqrt{\nu[f|\mathcal{F}_1]}\right)^2\right]\left[\nu[f|\mathcal{F}_2]-\nu\left(\sqrt{\nu[f|\mathcal{F}_2]}\right)^2\right]\right] \\
&\quad\le\|h-1\|_\infty\overline{\nu}\left[\left|\nu[f|\mathcal{F}_1]-\nu\left(\sqrt{\nu[f|\mathcal{F}_1]}\right)^2\right|\left|\nu[f|\mathcal{F}_2]-\nu\left(\sqrt{\nu[f|\mathcal{F}_2]}\right)^2\right|\right] \\
&\quad=\|h-1\|_\infty\nu\left[\left|\nu[f|\mathcal{F}_1]-\nu\left(\sqrt{\nu[f|\mathcal{F}_1]}\right)^2\right|\right]\nu\left[\left|\nu[f|\mathcal{F}_2]-\nu\left(\sqrt{\nu[f|\mathcal{F}_1]}\right)^2\right|\right]
\end{aligned}
$$

Now notice that, for $k = 1, 2$,

$$\nu\left[\left|\nu[f|\mathcal{F}_k] - \nu\left(\sqrt{\nu[f|\mathcal{F}_k]}\right)^2\right|\right] \leq 2\sqrt{Var_\nu(\sqrt{f})}.$$

In fact

$$\begin{aligned}
&\nu\left[\left|\nu[f|\mathcal{F}_k] - \nu\left(\sqrt{\nu[f|\mathcal{F}_k]}\right)^2\right|\right] \\
&\quad = \nu\left[\left|\sqrt{\nu(f|\mathcal{F}_k)} - \nu\left[\sqrt{\nu(f|\mathcal{F}_k)}\right]\right|\left|\sqrt{\nu(f|\mathcal{F}_k)} + \nu\left[\sqrt{\nu(f|\mathcal{F}_k)}\right]\right|\right] \\
&\quad\quad \leq \sqrt{Var_\nu\left(\sqrt{\nu(f|\mathcal{F}_k)}\right)}\sqrt{2\left[1 + \nu\left[\sqrt{\nu(f|\mathcal{F}_k)}\right]^2\right]} \\
&\quad\quad\quad \leq 2\sqrt{Var_\nu\left(\sqrt{\nu(f|\mathcal{F}_k)}\right)} \leq 2\sqrt{Var_\nu(\sqrt{f})}.
\end{aligned}$$

Summing all up we get

$$\log \nu\left[\nu[f|\mathcal{F}_1]\nu[f|\mathcal{F}_2]\right] \leq 4\|h-1\|_\infty Var_\nu\left(\sqrt{f}\right),$$

and the proof is completed by the inequality (see [10])

$$Var_\nu\left(\sqrt{f}\right) \leq \nu\left[f \log f\right]$$

□

There is a "canonical" choice of $\overline{\nu}$:

$$\overline{\nu}(\eta) := \nu(\eta_{B\setminus A})\nu(\eta_{A\setminus B})\nu_{A\cap B}^{\tau\eta_{\Lambda\setminus(A\cap B)}}(\eta_{A\cap B})$$

where by $\nu(\eta_{B\setminus A})$ we mean the probability of $\eta_{B\setminus A}$ under the restriction of ν to $\{-1,1\}^{B\setminus A}$.

The property given in the following (highly nontrivial) result, is known as *strong mixing* for Gibbs measures.

Proposition 3.1. *For $\beta > 0$ sufficiently small, there are constants $C, D > 0$ independent of A, B, Λ, τ such that*

$$\|h-1\|_\infty \leq Ce^{-D\,dist(A\setminus B, B\setminus A)} = Ce^{-D\sqrt{L_1}}.$$

Summarizing:

$$\begin{aligned}
Ent_\nu(f) &\leq s_{L_1,L_2}^{-1}\mathcal{E}_\Lambda^\tau(\sqrt{f},\sqrt{f}) \\
&\quad + s_{L_1,L_2}^{-1}\nu\left[\sum_{i\in A\cap B} c(i,\eta)\left(\nabla_i\sqrt{f(\eta)}\right)^2\right] + Ce^{-D\sqrt{L_1}}Ent_\nu(f)
\end{aligned}$$

We still have to deal with the term

$$\nu \left[\sum_{i \in A \cap B} c(i,\eta) \left(\nabla_i \sqrt{f(\eta)} \right)^2 \right].$$

We cannot bound it by $\mathcal{E}^\tau_\Lambda(\sqrt{f}, \sqrt{f})$, as otherwise we would get a horrible factor 2 in the recursion. Note that there is a remarkable freedom in the choice of A and B. In particular, we can choose A_n and B_n in $\sqrt{L_1}/3$ different ways in such a way that the sets $A_n \cap B_n$ are all disjoint. Averaging over n:

$$\frac{3}{\sqrt{L_1}} \nu \left[\sum_n \sum_{i \in A_n \cap B_n} c(i,\eta) \left(\nabla_i \sqrt{f(\eta)} \right)^2 \right] \leq \frac{3}{\sqrt{L_1}} \mathcal{E}^\tau_\Lambda(\sqrt{f}, \sqrt{f}).$$

Thus

$$\left(1 - Ce^{-D\sqrt{L_1}}\right) Ent_\nu(f) \leq s^{-1}_{L_1,L_2} \left(1 + \frac{3}{\sqrt{L_1}}\right) \mathcal{E}^\tau_\Lambda(\sqrt{f}, \sqrt{f})$$

which implies, for some constant $C > 0$ and L_1 large enough

$$Ent_\nu(f) \leq s^{-1}_{L_1,L_2} \left(1 + \frac{C}{\sqrt{L_1}}\right) \mathcal{E}^\tau_\Lambda(\sqrt{f}, \sqrt{f})$$

Remember $\nu = \nu^\tau_{R_{\frac{3}{2}L_1,L_2}}$. However, with no modifications, the argument can be extended to ν^τ_Λ for any $\Lambda \subseteq R_{\frac{3}{2}L_1,L_2}$. Therefore

$$s^{-1}_{\frac{3}{2}L_1,L_2} \leq s^{-1}_{L_1,L_2} \left(1 + \frac{C}{\sqrt{L_1}}\right)$$

Iterating twice on both sides, we get for a possibly different constant C

$$s^{-1}_{2L,2L} \leq s^{-1}_{L,L} \left(1 + \frac{C}{\sqrt{L}}\right)$$

which implies (exercise) $\inf_L s_{L,L} > 0$. We have therefore proved

Theorem 13. *There exists $\overline{\beta} > 0$ such that for every $\beta < \overline{\beta}$*

$$\inf_\Lambda s_\Lambda > 0,$$

3.2 The Bochner-Bakry-Emery approach to PI and MLSI

In this section we assume reversibility of L ($\iff$ the Dirichlet form $\mathcal{E}(f,g)$ is symmetric). We recall that the proof that the rate of L^2 convergence to equilibrium is the best constant in (**PI**) is based on

$$\frac{d}{dt} Var_\pi(S_t f) = -2\mathcal{E}(S_t f, S_t f).$$

Taking one more derivative we get

$$\frac{d^2}{dt^2}Var_\pi(S_t f) = -2\frac{d}{dt}\mathcal{E}(S_t f, S_t f) = 4\pi\left[(LS_t f)^2\right].$$

Suppose now $\exists k > 0$ such that for every f

$$\mathcal{E}(f,f) \le \frac{1}{k}\pi\left[(Lf)^2\right].$$

We get

$$\frac{d}{dt}\mathcal{E}(S_t f, S_t f) \le -2k\mathcal{E}(S_t f, S_t f)$$

In particular $\mathcal{E}(S_t f, S_t f) \to 0$ as $t \to +\infty$. Rewriting the last inequality as

$$\frac{d}{dt}\mathcal{E}(S_t f, S_t f) \le k\frac{d}{dt}Var_\pi(S_t f)$$

and integrating from t to $+\infty$ we get

$$\mathcal{E}(f,f) \ge kVar_\pi(f) \;\Rightarrow\; k \le \gamma\ !$$

By a bit of spectral Theory it can be shown that the best constant k in

$$(\mathbf{PI}') \qquad \mathcal{E}(f,f) \le \frac{1}{k}\pi\left[(Lf)^2\right]$$

is *equal* to the spectral gap γ.

The same argument can be implemented with he entropy replacing the variance. We obtain, for $f > 0$

$$\frac{d^2}{dt^2}Ent_\pi(S_t f) = -\frac{d}{dt}\mathcal{E}(S_t f, \log S_t f) = \pi\left[LS_t f L\log S_t f\right] + \pi\left[\frac{(LS_t f)^2}{S_t f}\right]$$

We therefore have that the inequality

$$(\mathbf{MLSI}') \qquad k\mathcal{E}(f,\log f) \le \pi\left[LfL\log f\right] + \pi\left[\frac{(Lf)^2}{f}\right]$$

for every $f > 0$, implies the (**MLSI**) $kEnt_\pi(f) \le \mathcal{E}(f,\log f)$.

This time the converse is not necessarily true: the entropy may decay exponentially fast, but not necessarily in a convex way.

In order to understand how useful the above inequalities are, we write generators of Markov chains in the following form:

$$Lf(x) = \sum_{\gamma\in G} c(x,\gamma)[f(\gamma(x)) - f(x)] =: \sum_{\gamma\in G} c(x,\gamma)\nabla_\gamma f(x)$$

where G is some set of functions from S to S (*allowed movements*). It is clear that every Markov chain can be written in this way: for $(x,y) \in S_2$, define

$$\gamma_{xy}(z) := \begin{cases} z & \text{for } z \neq x \\ y & \text{for } z = x \end{cases}$$

$$c(z,\gamma_{xy}) := \begin{cases} 0 & \text{for } z \neq x \\ L_{xy} & \text{for } z = x \end{cases}$$

and $G := \{\gamma_{xy} : (x,y) \in S_2\}$. This, however, is not necessarily the most convenient representation, as shown in examples below.

Now we give a condition which implies the Markov chain with generator $Lf(x) = \sum_{\gamma\in G} c(x,\gamma)\nabla_\gamma f(x)$ is reversible with respect to a probability π.

(**Rev**) For every $\gamma \in G$ there exists $\gamma^{-1} \in G$ such that $\gamma^{-1}\gamma(x) = x$ for every $x \in S$ such that $c(x,\gamma) > 0$. Moreover

$$\pi(x)c(x,\gamma) = \pi(\gamma(x))c(\gamma(x),\gamma^{-1})$$

Under (**Rev**) it is easy to see that

$$\begin{aligned} \mathcal{E}(f,g) &= \frac{1}{2}\pi\left[\sum_{\gamma\in G} c(x,\gamma)\nabla_\gamma f(x)\nabla_\gamma g(x)\right] \\ \pi\left[(Lf)^2\right] &= \pi\left[\sum_{\gamma,\delta\in G} c(x,\gamma)c(x,\delta)\nabla_\gamma f(x)\nabla_\delta f(x)\right] \end{aligned}$$

The idea is that we could establish (**PI**) if we could make a pointwise (i.e. for x fixed) comparison between

$$\sum_{\gamma\in G} c(x,\gamma)\nabla_\gamma f(x)\nabla_\gamma f(x) \quad \text{and} \quad \sum_{\gamma,\delta\in G} c(x,\gamma)c(x,\delta)\nabla_\gamma f(x)\nabla_\delta f(x)$$

as quadratic forms in $\nabla_\gamma f$. This is not possible in this terms, but it surprisingly works after a suitable "reshuffling".

Theorem 14. *Let $R : S \times G \times G \to [0,+\infty)$ be such that for each x,γ,δ with $R(x,\gamma,\delta) > 0$ we have*

$$\begin{aligned} &\mathbf{P1}: \quad R(x,\gamma,\delta) = R(x,\delta,\gamma) \\ &\mathbf{P2}: \quad \pi(x)R(x,\gamma,\delta) = \pi(x)R(\gamma(x),\gamma^{-1},\delta) \\ &\mathbf{P3}: \quad \gamma\delta(x) = \delta\gamma(x) \end{aligned}$$

Then, for every f,g the following Bochner*-type identity holds*

$$\pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\nabla_\gamma f(x)\nabla_\delta f(x)\right] = \frac{1}{4}\pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\nabla_\gamma\nabla_\delta f(x)\nabla_\gamma\nabla_\delta f(x)\right] \geq 0$$

In particular

$$\begin{aligned}
\pi\left[(Lf)^2\right] &= \pi\left[\sum_{\gamma,\delta\in G} c(x,\gamma)c(x,\delta)\nabla_\gamma f(x)\nabla_\delta f(x)\right] \\
&\geq \pi\left[\sum_{\gamma,\delta\in G} \left(c(x,\gamma)c(x,\delta) - R(x,\gamma,\delta)\right)\nabla_\gamma f(x)\nabla_\delta f(x)\right] \\
&=: \pi\left[\sum_{\gamma,\delta\in G} \Gamma_R(x,\gamma,d)\nabla_\gamma f(x)\nabla_\delta f(x)\right]
\end{aligned}$$

Proof. First, by (**P3**), $\nabla_\gamma\nabla_\delta f(x)\nabla_\gamma\nabla_\delta f(x) = \nabla_\gamma\nabla_\delta f(x)\nabla_\delta\nabla_\delta f(x)$. Then write

$$\begin{aligned}
\nabla_\gamma\nabla_\delta f(x)\nabla_\delta\nabla_\gamma f(x) = \nabla_\delta f(\gamma(x))\nabla_\gamma f(\delta(x)) - \nabla_\delta f(\gamma(x))\nabla_\gamma f(x) \\
- \nabla_\delta f(x)\nabla_\gamma f(\delta(x)) + \nabla_\delta f(x)\nabla_\gamma f(x)
\end{aligned}$$

We show that each one of the four summands in the r.h.s. of this last formula, when multiplied by $R(x,\gamma,\delta)$, summed over γ,δ and averaged over π gives

$$\pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\nabla_\delta f(x)\nabla_\gamma f(x)\right].$$

For the fourth summand there is nothing to prove. Moreover, by (**P2**),

$$\begin{aligned}
\pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\nabla_\delta f(x)\nabla_\gamma f(x)\right] &= \pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\nabla_\delta f(\gamma(x))\nabla_{\gamma^{-1}} f(\gamma(x))\right] \\
&= -\pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\nabla_\delta f(\gamma(x))\nabla_\gamma f(x)\right]
\end{aligned}$$

which takes care of the second and, by symmetry, of the third summand. For the first summand we use first (**P2**), then (**P2**), (**P2**) again and (**P3**):

$$\begin{aligned}
&\pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\nabla_\delta f(x)\nabla_\gamma f(x)\right] = \pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\nabla_\delta f(\gamma(x))\nabla_{\gamma^{-1}} f(\gamma(x))\right] \\
&= -\pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\nabla_\gamma f(\delta(x))\nabla_\delta(x)\right] = -\pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\nabla_{\gamma^{-1}} f(\delta\gamma(x))\nabla_\delta(\gamma(x))\right] \\
&= \pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\nabla_\gamma f(\delta(x))\nabla_\delta(\gamma(x))\right]
\end{aligned}$$

□

The point is that one can often choose R such that, uniformly in $x \in X$, for every $u \in \mathbb{R}^g$

$$\sum_{\gamma,\delta} \Gamma_R(x,\gamma,d) u_\gamma u_\delta \geq k \sum_{\gamma} c(x,\gamma) u_\gamma^2.$$

for some $k > 0$. If so it is immediately seen that (**PI'**) holds with the constant k, thus $\gamma \geq k$.

We do not really have a general method for computing a “good” R. The following result applies, however, to many cases.

Proposition 3.2. *Let us write G in the form $G = J \cup J^{-1}$, where $J \subseteq G$, and $J^{-1} := \{\gamma : \gamma^{-1} \in J\}$. J and J^{-1} are not necessarily disjoint. Define $R(x,\gamma,\delta)$ as follows:*

$$R(x,\gamma,\delta) = \begin{cases} \frac{c(x,\gamma)[c(x,\delta)+c(\gamma(x),\delta)]}{2} & \text{if } \gamma\circ\delta = \delta\circ\gamma,\ \gamma,\delta \in J \cap J^{-1} \\ c(x,\gamma)c(\gamma(x),\delta) & \text{if } \gamma\circ\delta = \delta\circ\gamma,\ \gamma,\delta \in J\setminus J^{-1} \text{ or } \gamma,\delta \in J^{-1}\setminus J \\ c(x,\gamma)c(x,\delta) & \text{if } \gamma\circ\delta = \delta\circ\gamma,\ \left\{ \begin{array}{l} \gamma \in J\setminus J^{-1},\ \delta \in J^{-1}\setminus J \\ \text{or } \gamma \in J^{-1}\setminus J,\ \delta \in J\setminus J^{-1} \end{array} \right. \\ 0 & \text{otherwise.} \end{cases}$$

*Then properties (**P2**-**P3**) hold.*

Proof. The proof consists in a simple “checking by hands”, using the reversibility condition (**Rev**), and it is left as an exercise. □

Condition (**P1**) must be checked separately, since it may depend on the special choice of the rates. We shall refer to the R in this Proposition as the **canonical** R.

The argument starting from Bochner identity for (**MLSI'**) follows exactly the same lines, leading to the following. One first observes that

$$\pi[LfL\log f] + \pi\left[\frac{(Lf)^2}{f}\right] = \pi\left[\sum_{\gamma,\delta} c(x,\gamma)c(x,\delta)\nabla_\gamma f(x)\nabla_\delta \log f(x)\right] + \pi\left[\sum_{\gamma,\delta} c(x,\gamma)c(x,\delta)\frac{\nabla_\gamma f(x)\nabla_\delta f(x)}{f(x)}\right]$$

If we use Bochner's identity in the first term

$$\pi\left[\sum_{\gamma,\delta} c(x,\gamma)c(x,\delta)\nabla_\gamma f(x)\nabla_\delta \log f(x)\right]$$

we can rewrite it as

$$\pi\left[\sum_{\gamma,\delta} \Gamma(x,\gamma,\delta)\nabla_\gamma f(x)\nabla_\delta \log f(x)\right] + \frac{1}{4}\pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\nabla_\gamma\nabla_\delta f(x)\nabla_\gamma\nabla_\delta \log f(x)\right]$$

The problem is that the last summand is not necessarily nonnegative. However:

Lemma 15. *For every $f > 0$*

$$\frac{1}{4}\pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\nabla_\gamma\nabla_\delta f(x)\nabla_\gamma\nabla_\delta \log f(x)\right] + \pi\left[\sum_{\gamma,\delta} c(x,\gamma)c(x,\delta)\frac{\nabla_\gamma f(x)\nabla_\delta f(x)}{f(x)}\right] \geq \pi\left[\sum_{\gamma,\delta} \Gamma(x,\gamma,\delta)\frac{\nabla_\gamma f(x)\nabla_\delta f(x)}{f(x)}\right]$$

Proof. Using the same argument as in Theorem 14 we obtain after some computations:

$$\begin{aligned}
&\pi\left[\sum_{\gamma,\delta} c(x,\gamma)c(x,\delta)\frac{\nabla_\gamma f(x)\nabla_\delta f(x)}{f(x)}\right] \\
&= \pi\left[\sum_{\gamma,\delta} \Gamma(c,\gamma,\delta)\frac{\nabla_\gamma f(x)\nabla_\delta f(x)}{f(x)}\right] + \pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\frac{\nabla_\gamma f(x)\nabla_\delta f(x)}{f(x)}\right] \\
&= \pi\left[\sum_{\gamma,\delta} \Gamma(c,\gamma,\delta)\frac{\nabla_\gamma f(x)\nabla_\delta f(x)}{f(x)}\right] \\
&+\frac{1}{4}\pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\left\{\nabla_\gamma\left(\frac{\nabla_\delta f(x)}{f(\delta(x))}\right)\nabla_\gamma\nabla_\delta f(x) - \nabla_\gamma\left(\frac{(\nabla_\delta f(x))^2}{f(x)f(\delta(x))}\right)\nabla_\gamma f(x)\right\}\right]
\end{aligned}$$

Thus we are left to show that

$$\pi\left[\sum_{\gamma,\delta} R(x,\gamma,\delta)\left\{\nabla_\gamma\nabla_\delta f(x)\nabla_\gamma\nabla_\delta \log f(x) + \nabla_\gamma\left(\frac{\nabla_\delta f(x)}{f(\delta(x))}\right)\nabla_\gamma\nabla_\delta f(x) - \nabla_\gamma\left(\frac{(\nabla_\delta f(x))^2}{f(x)f(\delta(x))}\right)\nabla_\gamma f(x)\right\}\right] \geq 0$$

Setting $a := f(x), b := f(\delta(x)), c := f(g(x)), d := f(\delta\gamma(x))$, one checks that $\{\cdots\}$ equals the sum of the following 4 expressions

$$\begin{gathered}
d\log d - d\log(bc/a) + (bc/a) - d \\
c\log c - c\log(da/b) + (da/b) - c \\
b\log b - b\log(da/c) + (da/c) - b \\
a\log a - a\log(bc/d) + (bc/d) - a
\end{gathered}$$

which are all nonnegative, since $\alpha \log \alpha - \alpha \log \beta + \beta - \alpha \geq 0$ for every $\alpha, \beta > 0$. $\square$

Putting all together we have

$$\begin{aligned}\pi[LfL\log f] + \pi\left[\frac{(Lf)^2}{f}\right] \\ \geq \pi\left[\sum_{\gamma,\delta} \Gamma(x,\gamma,\delta)\left(\nabla_\gamma f(x)\nabla_\delta \log f(x) + \frac{\nabla_\gamma f(x)\nabla_\delta f(x)}{f(x)}\right)\right]\end{aligned}$$

It follows that if we can find $k > 0$ such that

$$\begin{aligned}\sum_{\gamma,\delta} \Gamma(x,\gamma,\delta)\left(\nabla_\gamma f(x)\nabla_\delta \log f(x) + \frac{\nabla_\gamma f(x)\nabla_\delta f(x)}{f(x)}\right) \\ \geq k\sum_\gamma c(x,\gamma)\nabla_\gamma f(x)\nabla_\gamma \log f(x)\end{aligned}$$

for every $x \in S$ and $f > 0$, then $\alpha \geq k$.

This pointwise comparison is typically much harder than for (**PI'**).

Before proceeding to examples, we state a simple result that will be useful in the following two sections.

Lemma 16. *Let us write G in the form $G = J \cup J^{-1}$, where $J \subseteq G$, and $J^{-1} := \{\gamma : \gamma^{-1} \in J\}$, and assume $J \cap J^{-1} = \emptyset$ (we agree the the identity map is not in G). Then*

$$\begin{aligned}\mathcal{E}(f,g) = \frac{1}{2}\pi\left[\sum_{\gamma\in G} c(x,\gamma)\nabla_\gamma f(x)\nabla_\gamma g(x)\right] \\ = \pi\left[\sum_{\gamma\in J} c(x,\gamma)\nabla_\gamma f(x)\nabla_\gamma g(x)\right] \\ = \pi\left[\sum_{\gamma\in J^{-1}} c(x,\gamma)\nabla_\gamma f(x)\nabla_\gamma g(x)\right]\end{aligned}$$

3.3 MLSI for birth and death processes

Recall the general generator of a birth and death process

$$Lf(n) = a(n)\nabla^+ f(n) + b(n)\nabla^- f(n)$$

while the detailed balance equation is

$$a(n)\pi(n) = b(n+1)\pi(n+1).$$

The set G of "movements" consists of two elements $\{+, -\}$, where $+(n) = n+1$ and $-(n) = (n-1)\mathbf{1}_{n>0}$, and one is the inverse of the other, so we can write $J := \{+\}$, $J^{-1} := \{-\}$.

In the language of the previous section, we have $c(n,+) = a(n)$, $c(n,-) = b(n)$.

Moreover, the canonical R reads

$$\begin{aligned} R(n,+,+) &:= a(n)a(n+1) \\ R(n,-,-) &:= b(n)b(n-1) \\ R(n,+,-) = R(n,-,+) &:= a(n)b(n). \end{aligned}$$

It is a simple exercise to show that also Condition (**P1**) of the previous section is satisfied, while (**P2-P3**) are guaranteed by the "canonical" choice. In particular, letting as before $\Gamma(n,\delta,\gamma) = c(n,\gamma)c(n,\delta) - R(n,\gamma,\delta)$, we have

$$\begin{aligned} &\pi[LfL\log f] + \pi\left[\frac{(Lf)^2}{f}\right] \\ &\quad\geq \pi\left[\sum_{\gamma,\delta\in G}\Gamma((n,\delta,\gamma)\left(\nabla_\gamma f(n)\nabla_\delta \log f(n) + \frac{\nabla_g f(n)\nabla_\delta f(n)}{f(n)}\right)\right] \\ &\quad\geq \pi\left[\sum_{\gamma,\delta\in G}\Gamma((n,\delta,\gamma)\nabla_\gamma f(n)\nabla_\delta \log f(n)\right] \\ &\quad= \pi\left[a(n)[a(n)-a(n+1)]\nabla^+ f(n)\nabla^+ \log f(n)\right. \\ &\qquad\left. + b(n)[b(n)-b(n-1)]\nabla^- f(n)\nabla^- \log f(n)\right] \end{aligned}$$

Thus, the MLSI holds with the constant α if this last expression is greater or equal to $\alpha\mathcal{E}(f,\log f)$ for every $f > 0$. Moreover

$$\mathcal{E}(f,g) = \pi\left[a(n)\nabla^+ f(n)\nabla^+ g(n)\right] = \pi\left[b(n)\nabla^- f(n)\nabla^- g(n)\right]$$

The comparison becomes simple under, for example, the following assumptions:

(**A**) there exists $c > 0$ such that for every $n \geq 0$, $a(n) \geq a(n+1)$ and $b(n+1) - b(n) \geq c$.

These assumptions are satisfied with $c = 1$ in the case π is a Poisson measure, $a(n) = \lambda$ and $b(n) = n$.

Under the above assumptions, noting that $\nabla^\pm f(n)\nabla^\pm \log f(n) \geq 0$,

$$\begin{aligned} &\pi\left[a(n)[a(n)-a(n+1)]\nabla^+ f(n)\nabla^+ \log f(n)\right. \\ &\qquad\left. + b(n)[b(n)-b(n-1)]\nabla^- f(n)\nabla^- \log f(n)\right] \\ &\qquad\qquad \geq c\pi\left[b(n)\nabla^- f(n)\nabla^- \log f(n)\right] = c\mathcal{E}(f,\log f) \end{aligned}$$

We have therefore proved the following result.

Theorem 17. *Under assumption A, the MLSI holds with constant c.*

Despite of its simplicity, this result is sharper that it may appear, as the following facts illustrate.

Proposition 3.3. *Under assumption A, the LSI fails, i.e. there is no $s > 0$ such that*

$$sEnt_\pi(f) \le \mathcal{E}(\sqrt{f}, \sqrt{f})$$

for every $f \ge 0$.

Proposition 3.4. *In the Poisson case ($a(n) = \lambda$ and $b(n) = n$) the constant $c = 1$ is the optimal constant in the MLSI.*

The proof of both Propositions above are elementary, and consist in using suitable test functions.

To show that LSI fails, one first observe that, by reversibility and Assumption A,

$$\frac{\pi(n+1)}{\pi(n)} = \frac{a(n)}{b(n+1)} \le \frac{a(0)}{c(n+1)}$$

from which it follows that $\pi(n) \le \frac{B}{n!}$ for some constant $B > 0$.

Using this estimate, and considering the sequence of test functions

$$f_k(n) := \mathbf{1}_{(k,+\infty)}(n),$$

one shows by direct computation that

$$\lim_{k\to+\infty} \frac{\mathcal{E}(\sqrt{f_k}, \sqrt{f_k})}{Ent_\pi(f_k)} = 0.$$

Similarly, to show that $\alpha = 1$ is the optimal constant for the MLSI in the Poisson case, one takes the test functions

$$f_k(n) := e^{-n/k},$$

and it is easily shown that

$$\lim_{k\to+\infty} \frac{\mathcal{E}(f_k, \log f_k)}{Ent_\pi(f_k)} = 1.$$

3.4 Spectral gap for interacting birth and death processes

We shall be considering Markov chains on $\mathbb{N}^\Lambda$, with $\Lambda \subset\subset \mathbb{Z}^d$, whose generator is given by

$$Lf(\eta) = \sum_{i\in\Lambda} \left[\lambda \exp\left[-\beta \sum_{j\in\Lambda} K(i-j)\eta_j\right] \nabla_i^+ f(\eta) + \eta_i \nabla_i^- f(\eta)\right].$$

and stationary distribution

$$\pi(\eta) = \frac{1}{Z}\overline{\pi}(\eta) \exp\left[-\beta \sum_{i,j\in\Lambda} K(i-j)\eta_i\eta_j\right]$$

For $\beta = 0$ this is a family of independent birth and death processes with Poisson invariant measure. Thus, by tensorization, the MLSI holds with best constant $\alpha = 1$.

Argument similar to those of the previous section show that the spectral gap γ of a single birth and death process as well as that of the whole independent systems is $\gamma = 1$.

By using the Bochner-Bakry-Emery method we show how to obtain an explicit lower bound for the spectral gap for $\beta > 0$ sufficiently small, under the assumption that the interaction is repulsive ($K \geq 0$).

We will also point out why the same method fails when one tries to prove the MLSI for $\beta > 0$; as far as we know, this is still unproved.

It is simple to put a system of interacting birth and death processes in the general framework of the Bochner-Bakry-Emery method. Let

$$G := \{+_i, -_i : i \in \Lambda\}$$

where $+_i$ ($-_i$ resp.) is the function on $\mathbb{N}^\Lambda$ that add a particle at i (resp. removes a particle at i if there is at least one).

We have therefore

$$c(\eta, +_i) = \lambda \exp\left[-\beta \sum_{j\in\Lambda} K(i-j)\eta_j\right] \quad c(\eta, -_i) = \eta_i$$

The Dirichlet form is given by, taking $J := G := \{+_i : i \in \Lambda\}$,

$$\mathcal{E}(f,g) = \pi\left[\sum_{i\in\Lambda} \eta_i \nabla_i^- f(\eta) \nabla_i^- g(\eta)\right] = \pi\left[\sum_{i\in\Lambda} c(\eta, +_i) \nabla_i^+ f(\eta) \nabla_i^+ g(\eta)\right].$$

We try again with the canonical R:

$$\begin{array}{rcl} R(\eta, +_i, +, j) & = & c(\eta, +_i) c(\eta^{i,+}, +_j) \\ R(\eta, -_i, -_j) & = & \left\{ \begin{array}{ll} \eta_i(\eta_i - 1) & \text{for } i = j \\ \eta_i \eta_j & \text{for } i \neq j \end{array} \right. \\ R(\eta, +_i, -_j) & = & c(\eta, +_i)\eta_j \end{array}$$

which is easily shown to satisfy also (**P1**).

We obtain

$$\begin{aligned} \pi\left[(Lf)^2\right] \geq \sum_{\gamma,\delta\in G} \pi\left[\left(c(\eta,\gamma)c(\eta,\delta) - R(\eta,\gamma,\delta)\right)\nabla_\gamma f(\eta) \nabla_\delta f(\eta)\right] \\ = \sum_{i\in\Lambda} \pi\left[\eta_i \left(\nabla_i^- f(\eta)\right)^2\right] \\ + \sum_{i,j\in\Lambda} \pi\left[c(\eta,+_i)c(\eta,+_j)\left(1 - e^{-\beta K(i-j)}\right)\nabla_i^+ f(\eta) \nabla_j^+ f(\eta)\right] \end{aligned}$$

The first term
$$\sum_{i\in\Lambda}\pi\left[\eta_i\left(\nabla_i^- f(\eta)\right)^2\right]=\mathcal{E}(f,f).$$
For the second term, we first observe that $K\geq 0$ implies $c(\eta,+_j)\leq 1$. Using the inequality
$$\left|\nabla_i^+ f(\eta)\nabla_j^+ f(\eta)\right|\leq\frac{1}{2}\left[\left(\nabla_i^+ f(\eta)\right)^2+\left(\nabla_j^+ f(\eta)\right)^2\right]$$
we obtain the following simple bound
$$\begin{aligned}\sum_{i,j\in\Lambda}\pi&\left[c(\eta,+_i)c(\eta,+_j)\left(1-e^{-\beta K(i-j)}\right)\nabla_i^+ f(\eta)\nabla_j^+ f(\eta)\right]\\&\geq-\sum_{i,j\in\Lambda}\pi\left[c(\eta,+_i)\left|1-e^{-\beta K(i-j)}\right|\left(\nabla_i^+ f(\eta)\right)^2\right]\\&=-\mathcal{E}(f,f)\sum_{l\in\mathbb{Z}^d}\left(1-e^{-\beta K(l)}\right)\end{aligned}$$
We have therefore obtained a bound uniform over $\Lambda\subset\subset\mathbb{Z}^d$:
$$\pi\left[(Lf)^2\right]\geq[1-\varepsilon(\beta)]\mathcal{E}(f,f).$$
with
$$\varepsilon(\beta):=\sum_{l\in\mathbb{Z}^d}\left(1-e^{-\beta K(l)}\right).$$
This implies that the spectral gap γ satisfies the inequality
$$\gamma\geq 1-\varepsilon(\beta)$$
which is a nice bound when $\varepsilon(\beta)<1$. This is true for β sufficiently small if $\sum_{l\in\mathbb{Z}^d}K(l)<+\infty$.

Note that the above argument is based essentially on the inequality
$$\left|\nabla_i^+ f(\eta)\nabla_j^+ f(\eta)\right|\leq\frac{1}{2}\left[\left(\nabla_i^+ f(\eta)\right)^2+\left(\nabla_j^+ f(\eta)\right)^2\right]$$
When one goes along the same lines for MLSI, one gets terms of the form
$$\nabla_i^+ f(\eta)\nabla_j^+\log f(\eta).$$
By analogy, one would try a bound of the type
$$\begin{aligned}&\|\nabla_i^+ f(\eta)\nabla_j^+\log f(\eta)\|\\&\qquad\leq C\left[\nabla_i^+ f(\eta)\nabla_i^+\log f(\eta)+\nabla_j^+ f(\eta)\nabla_j^+\log f(\eta)\right]\end{aligned}$$
for some constant $C>0$ independent of $f>0$. This simply does **not** hold true.

3.5 Spectral Gap for interacting Random walks

We begin by considering a system of interacting random walks on the complete graph:

$$S = S_N := \left\{ \eta : \sum_{i\in\Lambda} \eta_i = N \right\}$$

$$L_{\text{c.g.}} f(\eta) = \sum_{i,j\in\Lambda} c(i,j,\eta) \nabla_{ij} f(\eta)$$

with

$$c(i,j,\eta) = \frac{1}{|\Lambda|} \eta_i \exp\left[\beta \sum_{k\in\Lambda} K(i-k)\eta_k \right]$$

and stationary distribution

$$\pi_N(\eta) = \frac{1}{Z_N} \prod_{i\in\Lambda} \frac{1}{\eta_i!} \exp\left[-\beta \sum_{i,j\in\Lambda} K(i-j)\eta_i\eta_j \right].$$

Again, we put it in the framework of the Bochner-Bakry-Emery method.

$$G = \{\gamma_{ij}, (i,j) \in \Lambda^2\}$$

where γ_{ij} is the map on S_N that moves a particle from site i to site j, if there is a particle at i, and it does nothing otherwise.

$$c(\eta, \gamma_{ij}) = c(i,j,\eta).$$

Here the canonical R does not work, since (**P1**) fails. We rather set

$$R(\eta, \gamma_{ij}, \gamma_{hl}) := c(\eta, \gamma_{ij}) c(\eta^{i,-}, \gamma_{h,l}).$$

In more explicit terms

$$R(\eta, \gamma_{ij}, \gamma_{hl}) = \begin{cases} c(\eta, \gamma_{ij}) c(\eta, \gamma_{il}) \frac{\eta_i - 1}{\eta_i} e^{-\beta K(0)} & \text{for } h = i \\ c(\eta, \gamma_{ij} c(\eta, \gamma_{hl}) e^{-\beta K(i-h)} & \text{for } h \neq i \end{cases}$$

One checks by hands that (**P1-P3**) are satisfied.

We thus obtain

$$
\begin{aligned}
&\pi_N \left[\left(L_{\text{c.g.}} f(\eta) \right)^2 \right] \\
&\qquad \geq \sum_{\gamma,\delta \in G} \pi_N \left[\left(c(\eta,\gamma) c(\eta,\delta) - R(\eta,\gamma,\delta) \right) \nabla_\gamma f(\eta) \nabla_\delta f(\eta) \right] \\
&\quad = \sum_{i,j,l} \pi_N \left[c(\eta,\gamma_{ij}) c(\eta,\gamma_{il}) \left(1 - \frac{\eta_i - 1}{\eta_i} e^{-\beta K(0)} \right) \nabla_{ij} f(\eta) \nabla_{il} f(\eta) \right] \\
&\quad + \sum_{i,j,h,l : i \neq h} \pi_N \left[c(\eta,\gamma_{ij}) c(\eta,\gamma_{hl}) \left(1 - e^{-\beta K(i-h)} \right) \nabla_{ij} f(\eta) \nabla_{hl} f(\eta) \right] \\
&\qquad\qquad = \sum_i \pi_N \left[\left(1 - \frac{\eta_i - 1}{\eta_i} e^{-\beta K(0)} \right) u_i^2(\eta) \right] \\
&\qquad\qquad\qquad + \sum_{i \neq h} \left(1 - e^{-\beta K(i-h)} \right) \pi_N \left[u_i(\eta) u_h(\eta) \right]
\end{aligned}
$$

where we have set $u_i(\eta) := \sum_j c(\eta,\gamma_{ij}) \nabla_{ij} f(\eta)$.

Now assume $K(\cdot) \geq 0$, so that

$$
\begin{aligned}
\pi_N \left[\left(L_{\text{c.g.}} f(\eta) \right)^2 \right] \geq \left(1 - e^{-\beta K(0)} \right) \sum_i \pi_N \left[u_i^2(\eta) \right] \\
+ \sum_{i \neq h} \left(1 - e^{-\beta K(i-h)} \right) \pi_N \left[u_i(\eta) u_h(\eta) \right]
\end{aligned}
$$

Thus, by the inequality $2|u_i u_h| \leq u_i^2 + u_h^2$, we get the estimate

$$
\begin{aligned}
&\pi_N \left[\left(L_{\text{c.g.}} f(\eta) \right)^2 \right] \\
&\qquad \geq \left(1 - e^{-\beta K(0)} - \sum_{z \in \mathbb{Z}^d \setminus \{0\}} \left| 1 - e^{-\beta K(z)} \right| \right) \sum_i \pi_N \left[u_i^2(\eta) \right].
\end{aligned}
$$

Finally

$$
\begin{aligned}
\sum_i \pi_N \left[u_i^2(\eta) \right] &= \sum_{i,j,l} \pi_N \left[c(\eta,\gamma_{ij}) c(\eta,\gamma_{il}) \nabla_{ij} f(\eta) \nabla_{il} f(\eta) \right] \\
&\geq \sum_{i,j,l} \frac{1}{|\Lambda|} \pi_N \left[c(\eta,\gamma_{ij}) \nabla_{ij} f(\eta) \nabla_{il} f(\eta) \right] = \mathcal{E}(f,f)
\end{aligned}
$$

where first we have used the fact that $c(\eta,\gamma_{il}) \geq 1/|\Lambda|$ for $\eta_i \geq 1$ (due to $K(\cdot) \geq 0$) and it is independent of l, and then we have used the identity

$$
\sum_{i,j,l} \pi_N \left[c(\eta,\gamma_{ij}) \nabla_{ij} f(\eta) \nabla_{il} f(\eta) \right] = \frac{1}{2} |\Lambda| \sum_{i,j} \left[c(\eta,\gamma_{ij}) \left(\nabla_{ij} f(\eta) \right)^2 \right]
$$

which is shown next in Lemma 19.

Summing all up we have proved that

$$\pi_N\left[\left(L_{\text{c.g.}}f(\eta)\right)^2\right] \geq \left(1-e^{-\beta K(0)}-\sum_{z\in\mathbb{Z}^d\setminus\{0\}}\left|1-e^{-\beta K(z)}\right|\right)\mathcal{E}(f,f).$$

In other words:

Theorem 18. *Under the assumption*

$$1-e^{-\beta K(0)}-\sum_{z\in\mathbb{Z}^d\setminus\{0\}}\left|1-e^{-\beta K(z)}\right|>0$$

the system has a spectral gap bounded away from zero uniformly in the volume $|\Lambda|$ and in the number of particles N.

We still have to prove

Lemma 19.

$$\sum_{i,j,l}\pi_N\left[c(\eta,\gamma_{ij})\nabla_{ij}f(\eta)\nabla_{il}f(\eta)\right]=\frac{1}{2}|\Lambda|\sum_{i,j}\left[c(\eta,\gamma_{ij})\left(\nabla_{ij}f(\eta)\right)^2\right]$$

Proof. We rely on the fact that $c(\eta,\gamma_{ij})$ is independent of j, so we rather write $c(\eta,i)$. By reversibility

$$\pi_N\left[c(\eta,i)\nabla_{ij}f(\eta)\nabla_{il}f(\eta)\right]=-\pi_N\left[c(\eta,l)\nabla_{ij}f(\gamma_{li}\eta)\nabla_{li}f(\eta)\right].$$

Since $\nabla_{ij}f(\gamma_{li}\eta)=\nabla_{lj}f(\eta)-\nabla_{li}f(\eta)$, we have

$$\begin{aligned}&\pi_N\left[c(\eta,i)\nabla_{ij}f(\eta)\nabla_{il}f(\eta)\right]\\&\qquad=-\pi_N\left[c(\eta,l)\nabla_{lj}f(\eta)\nabla_{li}f(\eta)\right]+\pi_N\left[c(\eta,l)\nabla_{li}f(\eta)\nabla_{li}f(\eta)\right].\end{aligned}$$

Summing over i,j,l the conclusion follows easily. □

We now consider the system in which particles can only jump to one of the nearest site:

$$L_{\text{n.n.}}f(\eta)=\sum_{i\sim j}\eta_i\exp\left[\beta\sum_{k\in\Lambda}K(i-k)\eta_k\right]\nabla_{ij}f(\eta),$$

where $i\sim j$ means $|i-j|=1$. The stationary distribution is the same as in the complete graph model.

This model cannot be studied directly with the Bochner-Bakry-Emery method. However it can be easily reduced to the model in the complete graph, via the following Lemma.

Lemma 20. *Denote by $\mathcal{E}_{c.g.}$ and $\mathcal{E}_{n.n.}$ the Dirichlet forms of the complete graph and of the nearest neighbor model respectively. Then there exists a universal constant C, only depending on the dimension d, such that*

$$\mathcal{E}_{c.g.}(f,f) \leq C\,diam^2(\Lambda)\mathcal{E}_{n.n.}(f,f).$$

Proof. Let $i,j \in \Lambda$, and choose a "path" $i = i_0, i_1, \ldots, i_{n-1}, i_n = j$, i.e. a sequence of nearest neighbors with $n \leq D := \mathrm{diam}(\Lambda)$. Then

$$\begin{aligned}
&(\nabla_{ij} f(\eta))^2 = (\gamma_{ij} f(\eta) - f(\eta))^2 \\
&\quad = \left(\sum_{k=1}^{n} [\gamma_{ii_k} f(\eta) - \gamma_{ii_{k-1}} f(\eta) \right)^2 \leq D \sum_{k=1}^{n} \left(\gamma_{ii_{k-1}} [\gamma_{i_{k-1}i_k} f(\eta) - f(\eta)] \right).
\end{aligned}$$

By reversibility, using the notation $c(\eta, i) = c(\eta, \gamma_{ij})$,

$$\begin{aligned}
\pi_N \left[c(\eta,i)\,(\nabla_{ij} f(\eta))^2 \right] &\leq n \pi_N \left[c(\eta,i) \sum_{k=1}^{n} \left(\gamma_{ii_{k-1}} [\gamma_{i_{k-1}i_k} f(\eta) - f(\eta)]^2 \right) \right] \\
&= n \sum_{k=1}^{n} \pi_N \left[c(\eta, i_{k-1}) [\gamma_{i_{k-1}i_k} f(\eta) - f(\eta)]^2 \right].
\end{aligned}$$

This reduces the problem to nearest neighbor exchanges. Summing over i,j, by some simple symmetry argument we get

$$\sum_{i,j} \pi_N \left[c(\eta,i)\,(\nabla_{ij} f(\eta))^2 \right] \leq n^2 |\Lambda| \sum_{i \sim j} \left[c(\eta,i)\,(\nabla_{ij} f(\eta))^2 \right],$$

from which the conclusion follows, taking into accounts that for the n.n. model rates do not have the factor $\frac{1}{|\Lambda|}$ in front. $\square$

An immediate consequence of this Lemma (known as Yau's Lemma) is the following

Corollary 21. *If γ is the spectral gap of the system in the complete graph, then the spectral gap of the system with only nearest neighbor jumps is bounded from below by $\frac{\gamma}{C\,diam^2(\Lambda)}$.*

Bibliographic remarks

The duplication method is a variation of the martingale method in [19], developed in [5, 7, 8, 11]. The Bochner-Bakry-Emery method, inspired by [6] was developed for diffusion processes in [1] and [16], and extended to the discrete case in [3] and in [9].

References

[1] Bakry, D. and Emery, M., *Diffusions hypercontractives*. Séminaire de Probabilités XIX. Lecture Notes in Math. 1123, 177-206, Springer-Verlag, 1985.

[2] E. Behrends Introduction to Markov Chains (with Special Emphasis on Rapid Mixing) Vieweg Verlag 1999.

[3] A.S. Boudou, P. Caputo, P. Dai Pra, G. Posta, *Spectral gap estimates for interacting particle systems via a Bochner-type identity*, J. Func. Analysis 232, 222-258, 2006.

[4] P. Bremaud, *Markov Chains, Gibbs Fields, Monte-Carlo simulations and queues*, Springer, 2001.

[5] Bertini, L., Cancrini, N. and Cesi, F., *The spectral gap for a Glauber-type dynamics in a continuous gas*, Ann. I. H. Poincaré - PR 38, 1 (2002), 91-108.

[6] S. Bochner, *Vector fields and Ricci curvature*, Bull. Am. Math. Soc. 52, 776-797.

[7] Cancrini, N. and Martinelli, F., *On the spectral gap of Kawasaki dynamics under a mixing condition revisited*, J. Math. Phys. 41 (2000), n. 3, 1391-1423.

[8] Cancrini, N., Martinelli, F. and Roberto, C., *The logarithmic Sobolev constant of Kawasaki dynamics under a mixing condition revisited*, Ann. Inst. H. Poincaré Probab. Statist. 38 (2002), no. 4, 385–436.

[9] Caputo, P., Posta, G., *Entropy dissipation estimates in a zero range dynamics*, preprint `math.PR/0405455`, to appear of Prob.Th. Rel. Fields.

[10] F. Cesi, *Quasi-factorization of the entropy and the logarithmic Sobolev inequality for Gibbs random fields*, Probab.Th.Rel. Fields 120, 569-584, 2001.

[11] P. Dai Pra, A. Paganoni. G. Posta, *Entropy inequalities for unbounded spin systems*, Ann. Probab. 30 (4), 1959-1976, 2004.

[12] P. Diaconis and S. Holmes (eds.), *Stein's Method: Expository Lectures and Applications*, IMS, 2004.

[13] P. Diaconis, L. Saloff-Coste, *Logarithmic Sobolev inequalities for finite Markov chains*, Ann. Appl. Probab. 6 (3), 695-750, 1996.

[14] S.N. Ethier, T.G. Kurtz, *Markov Processes, Characterization and convergence*, John Wiley and Sons, 1986.

[15] J.A. Fill, *An interruptible algorithm for perfect sampling via Markov chains*, Ann. Appl. Probab. 8 (1), 131-162, 1998.

[16] Ledoux, M., *Logarithmic Sobolev inequalities for unbounded spin systems revisited*, Séminaire de Probabilités XXXV. Lecture Notes in Math. 1755, 167-194, Springer, Berlin, 2001.

[17] T. Ligget, *Interacting Particle Systems*, Springer Verlag, Berlin, 1995.

[18] J. G. Propp and D. B. Wilson, *Exact sampling with coupled Markov chains and applications to statistical mechanics*, Random Structures and Algorithms, 9(1&2):223–252, 1996.

[19] Yau, H-T., *Logarithmic Sobolev Inequality for Lattice Gases with Mixing Conditions*, Comm. Math. Phys. 181 (1996), 367-408.