Markov Processes Relat. Fields **8**, 595–610 (2002)

Markov Processes and Related Fields

Log-Sobolev and Spectral Gap Inequalities for the Knapsack Markov Chain

P. Mathieu

CMI, Université de Provence, 39 Rue F. Joliot Curie, 13453 Marseille Cedex 13, France.
E-mail: pierre.mathieu@cmi.univ-mrs.fr

Received November 9, 2001, revised February 19, 2002

Abstract. We study a random walk moving on a discrete cube of large dimension and subject to a linear constraint. Under the assumption that more than the half of the points in the cube satisfy the constraint, we prove log-Sobolev and Poincaré inequalities with constants of the correct order.

KEYWORDS: knapsack problem, log-Sobolev inequalities, spectral gap, generalized Poincaré inequalities

AMS SUBJECT CLASSIFICATION: 60J10

1. Introduction

Research on Markov chains has recently focused on quantitative bounds, see [1] or [10]. Examples of Markov chains for which explicit computation can be carried over are very few; one of these is the simple random walk on a discrete cube $\{-1,+1\}^N$. In this paper, we investigate the effect of adding a linear constraint to the simple random walk on a discrete cube. Since the state space of the new Markov chain does not have any symmetry anymore, exact computation cannot be done and the best we can hope for are 'sharp' estimates. (Here 'sharp' will mean 'up to multiplicative constants'.)

There are different ways to estimate the law of a Markov chain; spectral gap bounds and coupling estimates are among the most popular. It turns out that, in our example, the most convenient approach is to prove a log-Sobolev inequality. (It even seems harder to get directly a spectral gap inequality than the log-Sobolev inequality.) As a consequence, not only do we bound the mixing time, but hypercontractivity is proved, i.e. some control of the law of the chain on scale N is obtained.

Let $a=(a_1,\dots,a_N)$ be a vector in $\mathbf{R}^N$ with positive entries. Let $\mathcal{S}=\{0,1\}^N$ be the discrete cube and define

$$\Delta=\Big\{x=(x_1,\dots,x_N)\in\mathcal{S}:\ \sum_{i=1}^N a_ix_i\le 1\Big\}.$$

Let η be the uniform probability measure on Δ. For any nonnegative integers $a\le b$, we use the notation $[[a,b]]$ for the set of integers between a and b.

Let $(X(t),t\in\mathbf{R})$ be the Markov chain on Δ defined as follows: starting at time $t=0$ from a given point $x\in\Delta$, we wait for an exponential time of parameter 1. Then choose one coordinate, say i, uniformly in $[[1,N]]$ and try to flip the ith coordinate of x. If the resulting configuration is still in Δ, we keep it. Otherwise we just do not change the configuration. One then repeats the same procedure again and again.

$(X(t),t\in\mathbf{R})$ is a continuous time Markov chain with state space Δ and generator

$$\mathcal{L}f(x)=\frac{1}{N}\sum_{i=1}^N \mathbf{1}\{T^ix\in\Delta\}\big(f(T^ix)-f(x)\big),$$

where T^ix is the point of $\mathcal{S}$ obtained by changing the value of the ith coordinate of x. Note that, since we chose positive values for a_j, if $x\in\Delta$ and $x_i=1$, then $T^ix\in\Delta$, i.e. Δ is a connected star-like subset of the cube $\mathcal{S}$ and therefore $X(t)$ is irreducible. Also note that the transition rates defining $\mathcal{L}$ satisfy the detailed balance condition with respect to the uniform probability; η is therefore the unique invariant probability for $X(t)$ and it is reversible.

One motivation for introducing a Markov chain on Δ is to solve the so-called knapsack problem, namely compute the cardinality of Δ. This question is known to be $\sharp P$ complete, which justifies the search for a stochastic algorithm that yields an approximate counting in polynomial time. Running the Markov chain X turns out to be a convenient way to simulate the uniform probability on Δ and thus construct the stochastic algorithm. We refer to [5] for further details. See also [4] for related questions.

It was recently proved, see [9], that the mixing time of X can be bounded by a polynomial in the size N. The bound does not depend on the choice of a but it is not sharp, as far as the exponent is concerned.

More precise estimates have been obtained under additional assumptions on the size of Δ: for $x\in\mathcal{S}$, let $|x|=\sum_{i=1}^N x_i$ be the number of 1's among the coordinates of x. For $\varepsilon\in[0,1]$, define $\Delta_\varepsilon=\{x\in\mathcal{S}:\ |x|\le\varepsilon N\}$. When $\varepsilon=1$, Δ_ε coincides with $\mathcal{S}$.

In [10], Theorem 4.1.2, Saloff-Coste uses comparison arguments to prove a sharp log-Sobolev inequality of the form (1.1) provided that, for all N, Δ contains $\Delta_{\varepsilon'}$ with $\varepsilon'=1-1/\sqrt{N}$. We extend this result under assumption:

(A) There exists $\varepsilon>1/2$ such that $\Delta_\varepsilon\subset\Delta$.

The technique we use mixes arguments from probability theory — coupling arguments — and estimates from semi-group theory — generalized Poincaré inequalities.

From the log-Sobolev inequality, one immediately gets estimates on the inverse spectral gap, see (1.2), hypercontractivity, see (1.3), and an upper bound for the mixing time, see (1.4).

In the very special case $\Delta = \mathcal{S}$, i.e. when X is the random walk on $\mathcal{S}$, very precise information is available, see [3] and [10] for instance. In particular it is known that $\mathcal{L}$ satisfies a log-Sobolev inequality with constant CN. The spectral gap can be computed by spectral analysis and is found to be C/N. The mixing time is of order $CN \log N$. (C is explicitly known.) Comparing with (1.1)–(1.4), we see that all our bounds are optimal up to multiplicative constants.

Let us further mention that Lowe and Meise also studied the knapsack problem under assumption (A), see [6]. They obtained an upper bound for the inverse spectral gap of the form $\lambda \geq C_\varepsilon/(N \log N)$, which implies that the mixing time satisfies $T \leq N^2 \log N$, but they were not able to get the correct order lower bound for λ nor to establish a log-Sobolev inequality.

Statement of the main results

Theorem 1.1 (log-Sobolev inequality). *Let $\varepsilon > 1/2$, and assume that assumption (A) is valid. There exists a constant C_ε that depends only on ε such that $\mathcal{L}$ satisfies the following log-Sobolev inequality:*

$$\eta(f^2 \log f^2) - \eta(f^2) \log \eta(f^2) \leq C_\varepsilon N \eta(-\mathcal{L} f \cdot f), \tag{1.1}$$

for any nonnegative function f on Δ.

A log-Sobolev inequality like (1.1) immediately yields various estimates on $\mathcal{L}$. Let us now mention some of these consequences:

(i) *Spectral gap estimates*: let λ be the spectral gap of $\mathcal{L}$, i.e. λ is the smallest non zero eigenvalue of $-\mathcal{L}$. From Proposition 3.7 of [2] and Theorem 1.1, we deduce that there exists a constant C_ε that depends only on ε such that

$$\lambda \geq \frac{1}{C_\varepsilon N}. \tag{1.2}$$

(ii) *Hypercontractivity*: let $P_t = e^{t\mathcal{L}}$ be the semi-group generated by $\mathcal{L}$. Let $q > 2$. According to Proposition 3.4 in [2], (1.1) implies that, for some constant $A_{\varepsilon,q}$ that depends only on ε and q and for any function f,

$$\eta\big(|P_{A_{\varepsilon,q}N} f|^q\big)^{1/q} \leq \eta(f^2)^{1/2}. \tag{1.3}$$

(iii) *Mixing time*: from log-Sobolev inequalities one can deduce estimates on the time it takes for the Markov chain generated by $\mathcal{L}$ to equilibrate. If, for instance, we define the equilibrium time:

$$T = \inf\Big\{t : \ \forall x \in \Delta, \forall f, \ \ |P_t f(x) - \eta(f)| \leq \frac{1}{2}\eta(f^2)^{1/2}\Big\},$$

then Corollary 2.2.7 in [10] tells us that

$$T \leq C_\varepsilon N \log N. \tag{1.4}$$

Other similar estimates can be found in [10].

We observed that it seems easier to prove at the same time both the log-Sobolev inequality (1.1) and the spectral gap inequality (1.2) than to get lower bounds for λ only. Also note that we deduce the log-Sobolev inequality from hypercontractivity estimates on P_t, see Lemma 3.1. These two points are rather unusual in the literature on log-Sobolev inequalities.

Let Γ be a connected subset of $\mathcal{S}$. Replacing Δ by Γ in the definition of $\mathcal{L}$ defines a Markov generator and it makes sense to ask whether the conclusion of Theorem 1.1 still holds. Our proof applies provided that, on one hand, Γ satisfies condition (A) (i.e. Γ contains Δ_ε, for some $\varepsilon > 1/2$), and, on the other hand, Γ is a decreasing subset of $\mathcal{S}$. (By 'decreasing subset', we mean that if $x \in \Gamma$ and $y \in \mathcal{S}$ satisfies $y_i \le x_i$ for all i, then $y \in \Gamma$.)

2. Hitting time estimates

In this section we shall derive some estimates on hitting times for the usual random walk on the cube $\mathcal{S}$. ε is a fixed number in $]1/2, 1]$ and the constants, A_ε, $C_\varepsilon, \ldots$ are universal functions of ε that do not depend on N. Their value may change from line to line.

We use the notation $(X^0(t), t \in \mathbf{R})$ to denote the random walk on $\mathcal{S}$, $X^0(t)$ is thus the continuous time Markov process on $\mathcal{S}$ with generator

$$\mathcal{L}^0 f(x) = \frac{1}{N} \sum_{i=1}^{N} f(T^i x) - f(x).$$

The unique invariant probability measure of $X^0(t)$ is the uniform probability on $\mathcal{S}$, say η^0; η^0 is also reversible for $\mathcal{L}^0$.

Define

$$\tau_\varepsilon^0 = \inf\Big\{t : \; |X^0(t)| \le \Big(\frac{1}{2} + \frac{1}{3}\Big(\varepsilon - \frac{1}{2}\Big)\Big)N\Big\}$$

and

$$\sigma_\varepsilon^0 = \inf\Big\{t : \; |X^0(t)| > \Big(\frac{1}{2} + \frac{2}{3}\Big(\varepsilon - \frac{1}{2}\Big)\Big)N\Big\}.$$

Lemma 2.1. *There exist constants A_ε and C_ε such that:*

(i) *for any $x \in \mathcal{S}$ with*

$$|x| \le \Big(\frac{1}{2} + \frac{1}{3}\Big(\varepsilon - \frac{1}{2}\Big)\Big)N,$$

we have

$$\mathsf{P}\big[\sigma_\varepsilon^0 \ge A_\varepsilon N \mid X^0(0) = x\big] \ge 1 - \exp(-N/C_\varepsilon); \tag{2.1}$$

(ii) *for any $x \in \mathcal{S}$,*

$$\mathsf{P}\big[\tau_\varepsilon^0 \le A_\varepsilon N \mid X^0(0) = x\big] \ge 1 - \exp(-N/C_\varepsilon). \tag{2.2}$$

As will be clear from the proof, these two estimates are not very sharp and one could get bounds on the constants A_ε and C_ε. We only stated the lemma as it is because (i) and (ii) are sufficient for our purposes.

Proof of (i). Let $m^0(t) = |X^0(t)|$. The symmetry of the walk X^0 with respect to any permutation of the coordinates implies that $m^0(t)$ is a Markov process and thus

$$\mathsf{P}\big[\sigma_\varepsilon^0 \geq A_\varepsilon N \mid X^0(0) = x\big] = \mathsf{P}\big[\sigma_\varepsilon^0 \geq A_\varepsilon N \mid m^0(0) = |x|\big].$$

The state space of m^0 is $[[0, N]]$ and $m^0(t)$ can only perform jumps of size 1. m^0 has a unique invariant probability measure, say g^0, which is also reversible. For $m \in [[0, N]]$, $g^0(m) = \binom{N}{m} 2^{-N}$.

For convenience, set

$$N_\varepsilon = \mathit{Int}\Big[\Big(\frac{1}{2} + \frac{2}{3}\Big(\varepsilon - \frac{1}{2}\Big)\Big)N\Big] + 1,$$

where $\mathit{Int}(a)$ is the integer part of a. Note that $\sigma_\varepsilon^0 = \inf\{t : m^0(t) = N_\varepsilon\}$. From the reversibility of the process $m^0(t)$ with respect to g^0, it follows that, for any $x \in \mathcal{S}$ with

$$|x| \leq \Big(\frac{1}{2} + \frac{1}{3}\Big(\varepsilon - \frac{1}{2}\Big)\Big)N$$

and for any time t,

$$\mathsf{P}\big[\sigma_\varepsilon^0 \leq t \mid m^0(0) = |x|\big] \leq (t+1)\frac{g^0(N_\varepsilon)}{g^0(|x|)}. \tag{2.3}$$

For the reader's convenience, we indicate the proof of (2.3) although we believe this is a classical fact from symmetric Markov chains: let $(\widehat{X}^0(t), t \in \mathbf{N})$ be the usual random walk on $\mathcal{S}$ with discrete time. Quantities defined from $\widehat{X}^0(t)$ will have a hat $\widehat{\ }$. Let $n(t)$ be the Poisson process that counts the number of jumps of X^0 until time t. Note that $n(t)$ has parameter 1 and in particular $\mathsf{E}[n(t)] = t$. It is possible to construct the three processes X^0, n and $\widehat{X}^0$ on the same probability space in such a way that $X^0(t) = \widehat{X}^0(n(t))$ and n and $\widehat{X}^0$ are independent processes. Therefore, if $\widehat{\sigma}_\varepsilon^0$ denotes the hitting time of N_ε by $\widehat{X}^0$, we have $\widehat{\sigma}_\varepsilon^0 = n(\sigma_\varepsilon^0)$ and

$$\begin{aligned}
\mathsf{P}\big[\sigma_\varepsilon^0 \leq t \mid m^0(0) = |x|\big] &= \mathsf{P}\big[\widehat{\sigma}_\varepsilon^0 \leq n(t) \mid \widehat{m}^0(0) = |x|\big] \\
&= \sum_{k \geq 0} \mathsf{P}\big[\widehat{\sigma}_\varepsilon^0 = k;\ k \leq n(t) \mid \widehat{m}^0(0) = |x|\big] \\
&= \sum_{k \geq 0} \mathsf{P}\big[\widehat{\sigma}_\varepsilon^0 = k \mid \widehat{m}^0(0) = |x|\big]\, \mathsf{P}\big[k \leq n(t)\big].
\end{aligned} \tag{2.4}$$

(Of course, $\widehat{m}^0(t) = |\widehat{X}^0(t)|$.)

But

$$\begin{aligned}
&\mathsf{P}\big[\widehat{\sigma}_\varepsilon^0 = k \mid \widehat{m}^0(0) = |x|\big] \\
&\quad = \sum_{m_1, \ldots, m_{k-1}} \mathsf{P}\big[\widehat{m}^0(0) = m_0, \ldots, \widehat{m}^0(k) = m_k \mid \widehat{m}^0(0) = |x|\big],
\end{aligned}$$

where we sum over all paths $(m_0,\dots,m_k)$ starting from $m_0=|x|$, ending at $m_k=N_\varepsilon$ and such that $m_i\neq N_\varepsilon$ for $i<k$.

Since $\widehat{X}^0$ is reversible with respect to g^0, we have

$$\begin{aligned}&\mathsf{P}\big[\widehat{m}^0(0)=m_0,\dots,\widehat{m}^0(k)=m_k\mid\widehat{m}^0(0)=|x|\big]\\ &=\ \frac{g^0(m_k)}{g^0(m_0)}\,\mathsf{P}\big[\widehat{m}^0(0)=m_k,\dots,\widehat{m}^0(k)=m_0\mid\widehat{m}^0(0)=m_k\big].\end{aligned}$$

Replacing m_k and m_0 by their values, we therefore get

$$\begin{aligned}&\mathsf{P}\big[\widehat{\sigma}^0_\varepsilon=k\mid\widehat{m}^0(0)=|x|\big]\\ &=\ \sum_{m_1,\dots,m_{k-1}}\frac{g^0(N_\varepsilon)}{g^0(|x|)}\,\mathsf{P}\big[\widehat{m}^0(0)=N_\varepsilon,\dots,\widehat{m}^0(k)=|x|\,\big|\,\widehat{m}^0(0)=N_\varepsilon\big]\\ &\le\ \frac{g^0(N_\varepsilon)}{g^0(|x|)},\end{aligned}$$

since $\sum_{m_1,\dots,m_{k-1}}\mathsf{P}\big[\widehat{m}^0(0)=N_\varepsilon,\dots,\widehat{m}^0(k)=|x|\,\big|\,\widehat{m}^0(0)=N_\varepsilon\big]\le 1$. Inserting this last inequality in (2.4) gives

$$\begin{aligned}\mathsf{P}\big[\sigma^0_\varepsilon\le t\mid\widehat{m}^0(0)=|x|\big]\ &\le\ \frac{g^0(N_\varepsilon)}{g^0(|x|)}\sum_{k\ge 0}\mathsf{P}\big[k\le n(t)\big]\\ &=\ \frac{g^0(N_\varepsilon)}{g^0(|x|)}\,(\mathsf{E}[n(t)]+1)\ =\ (t+1)\frac{g^0(N_\varepsilon)}{g^0(|x|)}.\end{aligned}$$

Obvious coupling arguments show that it is sufficient to prove (2.1) when

$$|x|=Int\Big[\Big(\frac{1}{2}+\frac{1}{3}\Big(\varepsilon-\frac{1}{2}\Big)\Big)N\Big].$$

From (2.3) we then get that

$$\begin{aligned}&\mathsf{P}\big[\sigma^0_\varepsilon\le A_\varepsilon N\mid m^0(0)=|x|\big]\\ &\le\ (A_\varepsilon N+1)\,\frac{g^0\Big(Int\Big[\Big(\frac{1}{2}+\frac{2}{3}\Big(\varepsilon-\frac{1}{2}\Big)\Big)N\Big]+1\Big)}{g^0\Big(Int\Big[\Big(\frac{1}{2}+\frac{1}{3}\Big(\varepsilon-\frac{1}{2}\Big)\Big)N\Big]\Big)}.\end{aligned}$$

One easily checks, with the help of Stirling's formula, that

$$\frac{g^0\Big(Int\Big[\Big(\frac{1}{2}+\frac{2}{3}\Big(\varepsilon-\frac{1}{2}\Big)\Big)N\Big]+1\Big)}{g^0\Big(Int\Big[\Big(\frac{1}{2}+\frac{1}{3}\Big(\varepsilon-\frac{1}{2}\Big)\Big)N\Big]\Big)}\le\exp(-N/C_\varepsilon).$$

Therefore $\mathsf{P}\big[\sigma^0_\varepsilon\le A_\varepsilon N\mid X^0(0)=x\big]\le\exp(-N/C_\varepsilon)$.

Proof of (ii). We use the notation 1 to denote the point of $\mathcal{S}$ with all coordinates equal to 1. It clearly suffices to prove (2.2) when $X^0(0) = 1$.

Consider the following stochastic algorithm, $(Z(t) = (X^1(t), X^2(t)), t \in \mathbf{R}_+)$ defined on $\mathcal{S} \times \mathcal{S}$: we start with all the coordinates of $X^1(0)$ equal to 0 and all the coordinates of $X^2(0)$ equal to 1. Wait for an exponential time of mean $1/2$, then choose an integer i uniformly in $[[1, N]]$. Next, set the ith coordinates of X^1 and X^2 to 0 with probability $1/2$ or to 1 with probability $1/2$. This way the ith coordinates of X^1 and X^2 coincide. Then iterate this procedure.

$Z(t)$ is a Markov process. The law of $(X^1(t), t \in \mathbf{R}_+)$ (resp. $(X^2(t), t \in \mathbf{R}_+)$) is the same as the law of $(X^0(t), t \in \mathbf{R}_+)$ started with all coordinates equal to 0 (resp. equal to 1). Let $m^1(t) = |X^1(t)|$ and $m^2(t) = |X^2(t)|$. For any t one has $m^1(t) \leq m^2(t)$.

Let $K(t)$ be the number of coordinates that coincide in $X^1(t)$ and $X^2(t)$. Then $K(t)$ increases. For all t, one has

$$m^2(t) \leq m^1(t) + (N - K(t)) \tag{2.5}$$

and

$$\begin{aligned}
&\mathsf{P}\big[\tau_\varepsilon^0 > A_\varepsilon N \mid X^0(0) = 1\big] \\
&\quad = \mathsf{P}\Big[\forall t \leq A_\varepsilon N,\ m^2(t) > \Big(\frac{1}{2} + \frac{1}{3}\Big(\varepsilon - \frac{1}{2}\Big)\Big)N\Big] \\
&\quad \leq \mathsf{P}\Big[m^2(A_\varepsilon N) > \Big(\frac{1}{2} + \frac{1}{3}\Big(\varepsilon - \frac{1}{2}\Big)\Big)N\Big] \\
&\quad \leq \mathsf{P}\Big[m^1(A_\varepsilon N) + \big(N - K(A_\varepsilon N)\big) > \Big(\frac{1}{2} + \frac{1}{3}\Big(\varepsilon - \frac{1}{2}\Big)\Big)N\Big] \\
&\quad \leq \mathsf{P}\Big[m^1(A_\varepsilon N) > \Big(\frac{1}{2} + \frac{1}{6}\Big(\varepsilon - \frac{1}{2}\Big)\Big)N\Big] \\
&\qquad + \mathsf{P}\Big[N - K(A_\varepsilon N) > \frac{1}{6}\Big(\varepsilon - \frac{1}{2}\Big)N\Big].
\end{aligned} \tag{2.6}$$

In the right part, the first probability is bounded by $\exp(-N/C_\varepsilon)$ for any choice of A_ε, as the proof of part (i) of the Lemma shows. We claim that

$$\mathsf{P}\Big[N - K(A_\varepsilon N) > \frac{1}{6}\Big(\varepsilon - \frac{1}{2}\Big)N\Big] \leq \exp(-N/C_\varepsilon),$$

if we properly choose A_ε.

Indeed, let $n(t)$ be the number of marks in the interval $[0, t]$ of the Poisson clock that defines Z; $n(t)$ is thus of order $2t$ for large t. For any subset $B \subset [[1, N]]$, the probability that the coordinates of $X^1(t)$ and $X^2(t)$ differ for all indices in B is $(1 - \sharp B/N)^{n(t)}$. When $t = A_\varepsilon N$, and for a set B of cardinality of order

$$\frac{1}{6}\Big(\varepsilon - \frac{1}{2}\Big)N,$$

the probability that the coordinates of $X^1(A_\varepsilon N)$ and $X^2(A_\varepsilon N)$ with indices in B differ is

$$\Big(1 - \frac{1}{6}\Big(\varepsilon - \frac{1}{2}\Big)\Big)^{n(A_\varepsilon t)}.$$

Since the number of possibilities for B is bounded by 2^N, we thus get that

$$\begin{aligned} &\mathsf{P}\Big[N - K(A_\varepsilon N) > \frac{1}{6}\Big(\varepsilon - \frac{1}{2}\Big)N\Big] \\ &\qquad \leq \; 2^N\Big(1 - \frac{1}{6}\Big(\varepsilon - \frac{1}{2}\Big)\Big)^{A_\varepsilon N} + \mathsf{P}\big[n(A_\varepsilon N) \leq A_\varepsilon N\big]. \end{aligned}$$

We can choose A_ε sufficiently large so that these two terms are bounded by $\exp\{-N/C_\varepsilon\}$. □

3. Coupling and log-Sobolev inequality

Recall that X^0 denotes the random walk on $\mathcal{S}$. We shall use a coupling argument to compare the Markov chain $X(t)$ with the walk $X^0(t)$ and derive a log-Sobolev inequality for $\mathcal{L}$. Before giving the details of the proof, let us briefly sketch our approach: consider the semi-group $P_t f(x) = \mathsf{E}[f(X(t)) \mid X(0) = x]$. As a bounded operator on $L_2(\Delta, \eta)$, P_t could also be defined by the formula: $P_t = \exp(t\mathcal{L})$. Similarly, let P_t^0 be the semi-group of the random walk on $\mathcal{S}$.

As pointed out in the introduction and as is well known (see [2]), proving a log-Sobolev inequality is equivalent to proving the ultracontractivity of the associated semi-group. For instance the log-Sobolev inequality of Lemma 3.1 is equivalent to saying that, for some or any $q > 2$ (that does not depend on N), there exist constants, $A_{\varepsilon,q}$ and $C_{\varepsilon,q}$, such that, for all functions f, one has:

$$\eta\big(|P_{A_{\varepsilon,q}N} f|^q\big)^{1/q} \leq C_{\varepsilon,q}\eta\big(|f|^2\big)^{1/2} \tag{3.1}$$

and the conclusion of Theorem 1.1 is equivalent to saying that one can take $C_{\varepsilon,q} = 1$ in (3.1). (See [2], Propositions 3.4 and 3.6.)

We already know that a similar inequality holds for P_t^0:

$$\eta^0\big(|P_{AN}^0 f|^q\big)^{1/q} \leq \eta^0\big(|f|^2\big)^{1/2}, \tag{3.2}$$

where A is some constant that depends on q (see [10, Chapter 2.2]).

We shall deduce (3.1) from (3.2) using a coupling argument. This is a rather unusual approach since in most examples one rather uses log-Sobolev inequalities to control the semi-group. We do not know how to get directly (3.1) from (3.2) with the correct constant $C_{\varepsilon,q} = 1$. This is mainly due to the fact that, on the event where the coupling is not successful, we completely loose the control of the semi-group.

Lemma 3.1. *There exists a constant C_ε such that the following log-Sobolev inequality is satisfied:*

$$\eta(f^2 \log f^2) - \eta(f^2)\log \eta(f^2) \leq C_\varepsilon N \eta(-\mathcal{L}f \cdot f) + C_\varepsilon \eta(f^2), \tag{3.3}$$

for any nonnegative function f on Δ.

Proof. As we said, it is sufficient to prove (3.1) for some $q > 2$ (q is not allowed to depend on N). Let

$$\tau_\varepsilon = \inf\Big\{t:\ |X(t)| \le \Big(\frac{1}{2}+\frac{1}{3}\Big(\varepsilon-\frac{1}{2}\Big)\Big)N\Big\}.$$

We claim that, for an appropriate constant A_ε, we have, for any $x \in \Delta$,

$$\mathsf{P}\big[\tau_\varepsilon \le A_\varepsilon N \mid X(0)=x\big] \ge 1-\exp(-C_\varepsilon/N). \tag{3.4}$$

Inequality (3.4) directly follows from Lemma 2.1. Indeed, for a given $x \in \Delta$, define a stochastic process $(Z(t) = (X^1(t), X^2(t)), t \in \mathbf{R}_+)$ in $\mathcal{S} \times \Delta$ as follows: set $X^1(0) = X^2(0)$. Wait for an exponential time of mean $1/2$ and then choose i uniformly at random in $[[1, N]]$. Let α take the value 0 or 1 with probability $1/2$. Then replace the ith coordinate of X^1 by α. If the point obtained by replacing the ith coordinate of X^2 by α is still in Δ, then replace it. Otherwise X^2 remains constant. Iterate this procedure. The law of $(X^1(t), t \in \mathbf{R}_+)$ is the same as the law of $(X^0(t), t \in \mathbf{R}_+)$. The law of $(X^2(t), t \in \mathbf{R}_+)$ is the same as the law of $(X(t), t \in \mathbf{R}_+)$. Besides, at any time t and for any coordinate i, one has $X_i^1(t) \ge X_i^2(t)$. Therefore

$$\mathsf{P}\big[\tau_\varepsilon \le A_\varepsilon N \mid X(0)=x\big] \ge \mathsf{P}\big[\tau_\varepsilon^0 \le A_\varepsilon N \mid X^0(0)=x\big]$$

and (3.4) follows from Lemma 2.1 (ii).

For some constant A_ε to be chosen later, we now define an auxiliary process $((X(t), \widetilde{X}(t)),\, t \in \mathbf{R}_+) \in \Delta \times \mathcal{S}$. ($X$ will behave like the Markov chain on Δ and therefore our notation should not be confusing.) Let $X(0) = \widetilde{X}(0) \in \Delta$. Let $n(t)$ be a Poisson clock of parameter 1. When the clock n rings for a time $t \le A_\varepsilon N$, we choose, uniformly at random, $i \in [[1, N]]$, and flip the ith coordinates of X and $\widetilde{X}$ if the resulting point is still in Δ. Otherwise, we do not change the current point. Thus $X(t)$ and $\widetilde{X}(t)$ just behave like the Markov chain on Δ. Besides, for all $t \le A_\varepsilon N$, we have $\widetilde{X}(t) = X(t)$. Next, when the clock n rings after time $A_\varepsilon N$, we choose, uniformly at random, $i \in [[1, N]]$, and

(i) flip the ith coordinates of X if the resulting point is still in Δ,

(ii) flip the ith coordinates of $\widetilde{X}$ whatever is the resulting point.

Thus the full process $(X(t), t \in \mathbf{R}_+)$ is the Markov chain on Δ, whereas, after time $A_\varepsilon N$, $\widetilde{X}$ starts behaving like the random walk on $\mathcal{S}$, i.e. does not take into account the constraint anymore. We have:

$$\mathsf{E}[f(\widetilde{X}(t)) \mid X(0) = \widetilde{X}(0) = x] = \mathsf{E}[P^0_{t-A_\varepsilon N} f(X(A_\varepsilon N)) \mid X(0) = x] \tag{3.5}$$

for all $t \ge A_\varepsilon N$. Also note that, until time

$$\inf\Big\{t \ge A_\varepsilon N \ \text{ such that }\ |X(t)| \ge \Big(\frac{1}{2}+\frac{2}{3}\Big(\varepsilon-\frac{1}{2}\Big)\Big)N\Big\},$$

the processes X and $\widetilde{X}$ coincide.

Let f be a nonnegative function defined on $\mathcal{S}$ and vanishing outside Δ. Then

$$\begin{aligned}
\eta\big(&\mathsf{E}[f(\widetilde{X}(2A_\varepsilon N)) \mid \widetilde{X}(0)=\cdot\,]^q\big)^{1/q} \\
&= \eta\big(\mathsf{E}[P^0_{A_\varepsilon N} f(X(A_\varepsilon N)) \mid X(0)=\cdot\,]^q\big)^{1/q} \\
&\le \eta\big(\mathsf{E}[(P^0_{A_\varepsilon N} f)^q(X(A_\varepsilon N)) \mid X(0)=\cdot\,]\big)^{1/q} \\
&= \eta\big((P^0_{A_\varepsilon N} f)^q\big)^{1/q} \\
&\le \Big(\frac{2^N}{\sharp\Delta}\Big)^{1/q} \eta^0\big[(P^0_{A_\varepsilon N} f)^q\big]^{1/q} \;\le\; 2^{1/q}\eta^0\big[(P^0_{A_\varepsilon N} f)^q\big]^{1/q} \\
&\le 2^{1/q}\eta^0[f^2]^{1/2} \;\le\; 2^{1/q}\eta(f^2)^{1/2},
\end{aligned} \tag{3.6}$$

where in the fourth line we used the fact that η is invariant for X and in the fifth line we used that from assumption (A), we have $2\sharp\Delta \ge \sharp\mathcal{S}$. The last inequality follows by (3.2) (we have assumed that A_ε is larger than the constant A in (3.2)). Also, by Holder's inequality

$$\begin{aligned}
\eta\big(&\mathsf{E}[f(X(2A_\varepsilon N)); \widetilde{X}(2A_\varepsilon N)\neq X(2A_\varepsilon N) \mid X(0)=\cdot\,]^q\big)^{1/q} \\
&\le \eta\big(\mathsf{E}[f^q(X(2A_\varepsilon N)) \mid X(0)=\cdot\,] \\
&\quad \times \mathsf{P}\big[\widetilde{X}(2A_\varepsilon N)\neq X(2A_\varepsilon N) \mid X(0)=\cdot\,\big]^{q-1}\big)^{1/q} \\
&= \eta\big(P_{2A_\varepsilon N}(f^q)\,\mathsf{P}\big[\widetilde{X}(2A_\varepsilon N)\neq X(2A_\varepsilon N) \mid X(0)=\cdot\,\big]^{q-1}\big)^{1/q} \\
&\le \eta\big[P_{2A_\varepsilon N}(f^q)\big]^{1/q} \sup_{x\in\Delta} \mathsf{P}\big[\widetilde{X}(2A_\varepsilon N)\neq X(2A_\varepsilon N) \mid X(0)=x\big]^{1-1/q} \\
&= \eta[f^q]^{1/q} \sup_{x\in\Delta} \mathsf{P}\big[\widetilde{X}(2A_\varepsilon N)\neq X(2A_\varepsilon N) \mid X(0)=x\big]^{1-1/q}.
\end{aligned} \tag{3.7}$$

Note that Lemma 2.1 implies

$$\sup_{x\in\Delta} \mathsf{P}[\widetilde{X}(2A_\varepsilon N)\neq X(2A_\varepsilon N) \mid X(0)=x] \le \exp(-N/C_\varepsilon). \tag{3.8}$$

Indeed, neglecting events of probability less than $\exp(-N/C_\varepsilon)$, we may assume that $\tau_\varepsilon \le A_\varepsilon N$. In particular we then have $\widetilde{X}(\tau_\varepsilon) = X(\tau_\varepsilon)$ and

$$|\widetilde{X}(\tau_\varepsilon)| \le \Big(\frac{1}{2}+\frac{1}{3}\Big(\varepsilon-\frac{1}{2}\Big)\Big)N.$$

Also note that, by assumption (A), as long as the process $X(t)$ satisfies

$$|X(t)| \le \Big(\frac{1}{2}+\frac{2}{3}\Big(\varepsilon-\frac{1}{2}\Big)\Big)N,$$

the constraint is not felt and $X(t) = \widetilde{X}(t)$. Since, due to Lemma 2.1 (i), this happens with probability larger than $1-\exp(-N/C_\varepsilon)$ for times $t \in [\tau_\varepsilon, 2A_\varepsilon N]$, we get (3.8).

Finally note that

$$\begin{aligned}
&\mathsf{E}[f(X(2A_\varepsilon N)) \mid X(0)=x] \\
&\quad \le \quad \mathsf{E}[f(\widetilde{X}(2A_\varepsilon N)) \mid X(0)=x] \\
&\qquad\quad + \mathsf{E}[f(X(2A_\varepsilon N)); \widetilde{X}(2A_\varepsilon N) \neq X(2A_\varepsilon N) \mid X(0)=x],
\end{aligned}$$

and therefore

$$\begin{aligned}
&\eta\big[(P_{2A_\varepsilon N} f)^q\big]^{1/q} \\
&\quad = \quad \eta\big(\mathsf{E}[f(X(2A_\varepsilon N)) \mid X(0)=\cdot\,]^q\big)^{1/q} \\
&\quad \le \quad \eta\big(\mathsf{E}[f(\widetilde{X}(2A_\varepsilon N)) \mid X(0)=\cdot\,]^q\big)^{1/q} \\
&\qquad\quad + \eta\big(\mathsf{E}[f(X(2A_\varepsilon N)); \widetilde{X}(2A_\varepsilon N) \neq X(2A_\varepsilon N) \mid X(0)=\cdot\,]^q\big)^{1/q} \\
&\quad \le \quad 2^{1/q}\eta(f^2)^{1/2} + \exp\Big(-\frac{N}{C_\varepsilon}\Big(1-\frac{1}{q}\Big)\Big)\eta(f^q)^{1/q}.
\end{aligned}$$

We used (3.6), (3.7) and (3.8).

Note that $\sharp\Delta \le 2^N$. Therefore, for any nonnegative function f, we have $\eta(f^q)^{1/q} \le 2^{N(1/2-1/q)}\eta(f^2)^{1/2}$. It remains to choose q close enough to 2 such that $(1/2-1/q)\log 2 \le C_\varepsilon(1-1/q)$, and we obtain the inequality

$$\eta\big[(P_{2A_\varepsilon N} f)^q\big]^{1/q} \le (2^{1/q}+1)\,\eta\big(f^2\big)^{1/2}. \tag{3.9}$$

□

4. Poincaré inequalities

If we knew that the spectral gap of $-\mathcal{L}$ was of order C_ε/N, we would apply Proposition 3.9 of [2] and conclude that (3.3) implies (1.1). But we did not prove any lower bound on λ so far.

It is also known that one way to deduce a log-Sobolev inequality of type (1.1) from one of type (3.3) is to prove a so-called generalized Poincaré inequality (see [7]). Our strategy to conclude the proof of Theorem 1.1 is therefore as follows: use geometric bounds to prove a generalized Poincaré inequality — this is the content of the next lemma — then combine it with the hypercontractivity estimate (3.1) in order to estimate the spectral gap λ and conclude using Proposition 3.9 in [2]. Most of the arguments to follow are not new: geometric tools for bounding generalized Poincaré inequalities are used in [8], and the link between log-Sobolev and generalized Poincaré inequalities was studied in [7]. Still, for the reader's convenience, we tried to give a self-contained proof.

Lemma 4.1. *Let $q > 2$. For any $\varepsilon > 1/2$, there exist $p = p(\varepsilon, q) \in]0,1[$ and a constant $C_{\varepsilon,q}$ such that any function f on Δ such that $\eta(f) = 0$, satisfies the inequality*

$$\eta\big(f^2\big)^{2/p}] \le C_{\varepsilon,q} N\,\eta(-\mathcal{L}f\cdot f)\,\eta\big(|f|^q\big)^{(4-2p)/(pq)}. \tag{4.1}$$

Proof. C will denote a numerical constant whose value changes from line to line. We shall use repeatedly the fact that, due to assumption (A), $C\sharp\Delta_\varepsilon \geq \sharp\mathcal{S}$ and also $C\sharp\Delta \geq \sharp\mathcal{S}$ and $C\sharp\Delta_\varepsilon \geq \sharp\Delta$. C_ε depends only on ε and, in particular, not on q; $C_{\varepsilon,q}$ depends on ε and q. These constants do not depend on N.

We prove Lemma 4.1 by comparing the chain $X(t)$ with the Markov chain on Δ_ε. Recall that $\Delta_\varepsilon = \{x \in \mathcal{S} : |x| \leq \varepsilon N\}$. Let $\mathcal{L}_\varepsilon$ be the Markov generator on Δ_ε defined by

$$\mathcal{L}_\varepsilon f(x) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\{T^i x \in \Delta_\varepsilon\}\big(f(T^i x) - f(x)\big)$$

for $x \in \Delta_\varepsilon$ and f defined on Δ_ε. We also use the notation η_ε to denote the uniform probability measure on Δ_ε.

Let λ_ε be the spectral gap of $-\mathcal{L}_\varepsilon$ in $L_2(\Delta_\varepsilon, \eta_\varepsilon)$; λ_ε is solution of the variational problem:

$$\lambda_\varepsilon = \inf_{f:\, \eta_\varepsilon(f)=0} \frac{-\eta_\varepsilon(\mathcal{L}_\varepsilon f \cdot f)}{\eta_\varepsilon(f^2)}$$

where the infimum is taken over all f such that $\eta_\varepsilon(f) = 0$. We shall later prove that

$$\lambda_\varepsilon \geq \frac{1}{CN}. \tag{4.2}$$

Direct inspection shows that

$$\eta(-\mathcal{L}f \cdot f) = \frac{1}{2(\sharp\Delta)N} \sum_{x \sim y \in \Delta} \big(f(x) - f(y)\big)^2 \tag{4.3}$$

where we sum over points $x, y \in \Delta$ that are neighbours in $\mathcal{S}$. Also

$$\eta_\varepsilon(-\mathcal{L}_\varepsilon f \cdot f) = \frac{1}{2(\sharp\Delta_\varepsilon)N} \sum_{x \sim y \in \Delta_\varepsilon} \big(f(x) - f(y)\big)^2 \tag{4.4}$$

where we sum over points $x, y \in \Delta_\varepsilon$ that are neighbours in $\mathcal{S}$. Therefore, for any function f on Δ, we have $\eta_\varepsilon(-\mathcal{L}_\varepsilon f \cdot f) \leq C\eta(-\mathcal{L}f \cdot f)$.

Since for a function f with $\eta_\varepsilon(f) = 0$ one has

$$\eta_\varepsilon(f^2) = \frac{1}{2(\sharp\Delta_\varepsilon)^2} \sum_{x, y \in \Delta_\varepsilon} \big(f(x) - f(y)\big)^2,$$

we have the inequality

$$\frac{1}{2(\sharp\Delta_\varepsilon)^2} \sum_{x, y \in \Delta_\varepsilon} \big(f(x) - f(y)\big)^2 \leq \frac{1}{\lambda_\varepsilon}\eta_\varepsilon(-\mathcal{L}_\varepsilon f \cdot f) \leq C\frac{1}{\lambda_\varepsilon}\eta(-\mathcal{L}f \cdot f)$$

which, due to (4.2), implies that

$$\frac{1}{2(\sharp\Delta_\varepsilon)^2} \sum_{x, y \in \Delta_\varepsilon} \big(f(x) - f(y)\big)^2 \leq C_\varepsilon N \eta(-\mathcal{L}f \cdot f). \tag{4.5}$$

Let f be defined on Δ and such that $\eta(f)=0$ and start with the equality

$$\begin{aligned}\eta(f^2) &= \frac{1}{2(\sharp\Delta)^2}\sum_{x,y\in\Delta}\big(f(x)-f(y)\big)^2\\ &= \frac{1}{2(\sharp\Delta)^2}\sum_{x,y\in\Delta_\varepsilon}\big(f(x)-f(y)\big)^2\\ &\quad+\frac{1}{(\sharp\Delta)^2}\sum_{x\in\Delta-\Delta_\varepsilon,y\in\Delta}\big(f(x)-f(y)\big)^2 \qquad (4.6)\\ &\quad+\frac{1}{2(\sharp\Delta)^2}\sum_{x,y\in\Delta-\Delta_\varepsilon}\big(f(x)-f(y)\big)^2.\end{aligned}$$

Let $q>2$ and choose $p>0$. We first bound the first term in (4.6): using Holder's inequality, we get that

$$\begin{aligned}&\frac{1}{2(\sharp\Delta)^2}\sum_{x,y\in\Delta_\varepsilon}\big(f(x)-f(y)\big)^2\\ &\le \frac{1}{2}\left(\frac{\sum_{x,y\in\Delta_\varepsilon}|f(x)-f(y)|^q}{(\sharp\Delta)^2}\right)^{(2-p)/q}\left(\frac{\sum_{x,y\in\Delta_\varepsilon}|f(x)-f(y)|^2}{(\sharp\Delta)^2}\right)^{p/2}.\end{aligned}$$

We use (4.5) to bound the second term in the right-hand side and get

$$\begin{aligned}&\frac{1}{2(\sharp\Delta)^2}\sum_{x,y\in\Delta_\varepsilon}(f(x)-f(y))^2 \qquad (4.7)\\ &\le C_{\varepsilon,q,p}\big(\eta(|f|^q)\big)^{(2-p)/q}\big(N\eta(-\mathcal{L}f\cdot f)\big)^{p/2}.\end{aligned}$$

Consider now the second term in the right-hand side of (4.6). We assume that $q+p-pq-2\ge 0$ and use Holder's inequality to get

$$\begin{aligned}&\frac{1}{(\sharp\Delta)^2}\sum_{\substack{x\in\Delta-\Delta_\varepsilon,\\ y\in\Delta}}\big(f(x)-f(y)\big)^2\\ &\le \left(\frac{\sum_{x\in\Delta-\Delta_\varepsilon,y\in\Delta}|f(x)-f(y)|^q}{(\sharp\Delta)^2}\right)^{(2-p)/q}\cdots\\ &\quad\cdots\left(\frac{\sum_{x\in\Delta-\Delta_\varepsilon,y\in\Delta}|f(x)-f(y)|}{(\sharp\Delta)^2}\right)^{p}\big(\eta(\Delta-\Delta_\varepsilon)\big)^{(p+q-pq-2)/q}.\end{aligned}$$

On one hand, we have $\eta(\Delta-\Delta_\varepsilon)\le\exp(-N/C_\varepsilon)$. On the other hand, since Δ is connected, for any choice of $x\in\Delta-\Delta_\varepsilon, y\in\Delta$, there is a connected path, say $\gamma=(x_0,x_1,\ldots,x_{2N})$, of length at most $2N$ and linking x and y in Δ (i.e. $x_0=x$, $x_{2N}=y$ and x_i and x_{i+1} are neighbours). Writing $|f(x)-f(y)|=|\sum_{i=1}^{2N}f(x_{i+1})-f(x_i)|$ and using rough bounds on $\sharp\Delta$, it is not difficult to prove that

$$|f(x)-f(y)|^2\le e^{CN}\frac{1}{2(\sharp\Delta)N}\sum_{x\sim y\in\Delta}\big(f(x)-f(y)\big)^2=e^{CN}\eta(-\mathcal{L}f\cdot f)$$

for some numerical constant C. We thus obtain the bound

$$\begin{aligned}\frac{1}{(\sharp\Delta)^2}&\sum_{x\in\Delta-\Delta_\varepsilon, y\in\Delta}\big(f(x)-f(y)\big)^2\\ &\le\ C_{p,q}\big(\eta(|f|^q)\big)^{(2-p)/q}\exp\Big(\frac{CpN}{2}\Big)\\ &\qquad\times\big(\eta(-\mathcal{L}f\cdot f)\big)^{p/2}\exp\Big(-\frac{p+q-pq-2}{q}\frac{N}{C_\varepsilon}\Big).\end{aligned}$$

We can choose p small enough and get to:

$$\begin{aligned}\frac{1}{(\sharp\Delta)^2}&\sum_{x\in\Delta-\Delta_\varepsilon,\ y\in\Delta}\big(f(x)-f(y)\big)^2\\ &\le\ C_{\varepsilon,q,p}\big(\eta(|f|^q)\big)^{(2-p)/q}\big(\eta(-\mathcal{L}f\cdot f)\big)^{p/2} \qquad (4.8)\\ &\le\ C_{\varepsilon,q,p}\big(\eta(|f|^q)\big)^{(2-p)/q}\big(N\eta(-\mathcal{L}f\cdot f)\big)^{p/2}.\end{aligned}$$

The third term in the right-hand side of (4.6) can be bounded in the same way:

$$\frac{1}{(\sharp\Delta)^2}\sum_{x,y\in\Delta-\Delta_\varepsilon}\big(f(x)-f(y)\big)^2\le C_{\varepsilon,q,p}\big(\eta(|f|^q)\big)^{(2-p)/q}\big(N\eta(-\mathcal{L}f\cdot f)\big)^{p/2}. \qquad (4.9)$$

From (4.6)–(4.9) we deduce that

$$\eta(f^2)\le C_{\varepsilon,q,p}\big(\eta(|f|^q)\big)^{(2-p)/q}\big(N\eta(-\mathcal{L}f\cdot f)\big)^{p/2},$$

which is the desired result.

It remains to justify (4.2), i.e. to compute the spectral gap of $-\mathcal{L}_\varepsilon$. We denote by $(X^\varepsilon(t), t\in\mathbf{R}_+)$ the Markov chain in Δ_ε whose generator is $\mathcal{L}_\varepsilon$.

Let $\mathcal{H}$ be the set of functions over Δ_ε that depend on x only through $|x|$. Let $\mathcal{K}$ be the orthogonal complement of $\mathcal{H}$ in $L_2(\Delta_\varepsilon,\eta_\varepsilon)$. Note that $f\in\mathcal{K}$ iff $\sum_{|x|=m}f(x)=0$ for all $m\in[[0,\varepsilon N]]$, where the summation is over all x such that $|x|=m$. Since Δ_ε is invariant under permutations of the coordinates and due to the symmetry in $\mathcal{L}_\varepsilon$, $\mathcal{H}$ and $\mathcal{K}$ are left invariant by $\mathcal{L}_\varepsilon$. Therefore

$$\lambda_\varepsilon=\inf\Bigg(\inf_{\substack{f\in\mathcal{H};\\ \eta_\varepsilon(f)=0}}\frac{\eta_\varepsilon(-\mathcal{L}_\varepsilon f\cdot f)}{\eta_\varepsilon(f^2)};\ \inf_{f\in\mathcal{K}}\frac{\eta_\varepsilon(-\mathcal{L}_\varepsilon f\cdot f)}{\eta_\varepsilon(f^2)}\Bigg). \qquad (4.10)$$

In (4.10), the first term is the spectral gap for the Markov process $|X^\varepsilon(t)|$. Note that $|X^\varepsilon|$ is obtained by restricting the process $|X^0|$ to the set $[[0,\varepsilon N]]$. $|X^0|$ can only perform nearest neighbours jumps. Restricting a one-dimensional, nearest neighbours dynamics to a sub-interval can only increase the spectral gap. The spectral gap of the dynamics $|X^0|$ in $[[0,N]]$ is larger than the spectral gap of X^0 in $\mathcal{S}$, which is known to be of order C/N (see [10], Example 2.1.2). Therefore we may conclude that

$$\inf_{\substack{f\in\mathcal{H};\\ \eta_\varepsilon(f)=0}}\frac{\eta_\varepsilon(-\mathcal{L}_\varepsilon f\cdot f)}{\eta_\varepsilon(f^2)}\ge\frac{1}{CN}. \qquad (4.11)$$

We now turn to the second term in (4.10). Let $f \in \mathcal{K}$. Let $\mathcal{A}$ be the operator

$$\mathcal{A}f(x) = \frac{1}{N} \sum_{i,j=1} f(T^{i,j}x) - f(x),$$

where $T^{i,j}x$ is the point of $\mathcal{S}$ obtained by exchanging the ith and jth coordinates of x. $\mathcal{A}$ is thus the generator of the Bernoulli–Laplace Markov chain, see [3]. Note that the level sets $|x| = m$ are invariant. Besides, for all $m \in [[0, N]]$, since $\sum_{x\in\mathcal{S};|x|=m} f(x) = 0$, the following Poincaré inequality is satisfied:

$$\sum_{\substack{x\in\mathcal{S};\\ |x|=m}} f^2(x) \leq \frac{1}{4N} \sum_{\substack{x\in\mathcal{S};\\ |x|=m}} \sum_{i,j=1}^{N} \big(f(T^{i,j}x) - f(x)\big)^2,$$

i.e. the spectral gap of the Bernoulli–Laplace model is bounded by $4/N$ (see [3]).

Let $x \in \Delta_\varepsilon$. Note that if $T^{i,j}x \neq x$, then $T^{i,j}x = T^iT^jx$. Interchanging the roles of i and j if necessary, we may assume that $T^jx \in \Delta_\varepsilon$. Besides $T^{i,j}x \in \Delta_\varepsilon$. Thus

$$\begin{aligned}
&\sum_{\substack{x\in\mathcal{S};\\ |x|=m}} \sum_{i,j=1}^{N} \big(f(T^{i,j}x) - f(x)\big)^2 \\
&\leq \; 4 \sum_{\substack{x\in\mathcal{S};\\ |x|=m}} \sum_{i,j=1}^{N} \mathbf{1}\{T^jx, T^iT^jx \in \Delta_\varepsilon\}\big(f(T^iT^jx) - f(T^jx)\big)^2 \\
&\quad + 4 \sum_{\substack{x\in\mathcal{S};\\ |x|=m}} \sum_{i,j=1}^{N} \mathbf{1}\{x, T^jx \in \Delta_\varepsilon\}\big(f(T^jx) - f(x)\big)^2 \\
&\leq \; 8N \sum_{\substack{x\in\mathcal{S};\\ |x|\in\{m,m+1/N,m-1/N\}}} \sum_{i=1}^{N} \mathbf{1}\{x, T^ix \in \Delta_\varepsilon\}\big(f(T^ix) - f(x)\big)^2.
\end{aligned}$$

We can now sum the last inequality over the different values of $m \in [[0, \varepsilon N]]$, to get that for all functions $f \in \mathcal{K}$

$$\eta_\varepsilon(f^2) \leq CN\eta_\varepsilon(-\mathcal{L}_\varepsilon f \cdot f). \tag{4.12}$$

Inserting (4.11) and (4.12) in (4.10) leads to the desired lower bound on λ_ε. The proof of Lemma 4.1 is now complete. □

End of the proof of Theorem 1.1. Recall that λ is the spectral gap of $\mathcal{L}$. Let ψ be the corresponding eigenfunction. We assume that $\eta(\psi^2) = 1$.

Let $q > 2$. We start with Lemma 3.1: from (3.3) and Proposition 3.4 in [2] we deduce that

$$\eta\big(|P_{A_{\varepsilon,q}N} f|^q\big)^{1/q} \leq C_{\varepsilon,q}\eta(f^2)^{1/2}.$$

Applying this inequality to ψ yields

$$\exp(-\lambda A_{\varepsilon,q} N)\, \eta\big(|\psi|^q\big)^{1/q} \le C_{\varepsilon,q}. \tag{4.13}$$

We now apply inequality (4.1) to the function ψ: taking into account that $-\mathcal{L}\psi = \lambda\psi$, we get

$$1 \le C_{\varepsilon,q} \lambda N\, \eta\big(|\psi|^q\big)^{(4-2p)/(pq)}, \tag{4.14}$$

where p depends on q and ε but not on N.

From (4.13) and (4.14) we deduce that

$$\exp(-\lambda A_{\varepsilon,q} N) \le C_{\varepsilon,q} (\lambda N)^{p/(4-2p)},$$

hence

$$\lambda \ge \frac{1}{C_\varepsilon N}. \tag{4.15}$$

We now combine (4.15) with the log-Sobolev inequality (3.3) and Proposition 3.9 of [2] to get the *tight* inequality

$$\eta(f^2 \log f^2) - \eta(f^2) \log \eta(f^2) \le C_\varepsilon N \eta(-\mathcal{L} f \cdot f).$$

□

References

[1] D. Aldous and J. Fill Reversible Markov chains and random walks on graphs. In preparation. This article is available from http://www.stat.berkeley.edu/users/aldous/

[2] D. Bakry (1992) L'hypercontractivité et son utilisation en théorie des semi-groupes. In: *Ecole d'Été de Probabilités de Saint-Flour*, Lect. Notes Math. **1581**, Springer-Verlag, Berlin.

[3] P. Diaconis and M. Shahshahani (1987) Time to reach stationarity in the Bernoulli – Laplace diffusion model. *SIAM J. Math. Analysis* **18**, 208–218.

[4] M. Dyer, A. Frieze, R. Kannan, A. Kapoor, L. Perkovic and U. Vazirani (1993) A mildly exponential time algorithm for approximating the number of solutions to a multidimensional knapsack problem. *Combin. Probab. Comput.* **2**, 271–284.

[5] M. Jerrum and A. Sinclair (1996) The Markov chain Monte Carlo method: an approach to approximate counting and integration. In: *Approximation Algorithms for NP-hard Problems*, D.S. Hochbaum (ed.), PWS Publishing, Boston.

[6] M. Lowe and C. Meise (2001) Note on the knapsack Markov chain. *Stoch. Process. Appl.* **94**, 155–170.

[7] P. Mathieu (1997) Hitting times and spectral gap inequalities. *Ann. Inst. H. Poincaré, Probabilités et Statistiques*, **33** (4), 437–465.

[8] P. Mathieu and P. Picco (2000) Convergence to equilibrium for finite Markov processes, with application to the random energy model. Preprint.

[9] B. Morris and A. Sinclair (1999) Random walks on truncated cubes and sampling $0-1$ knapsack solutions. *Proc.* 40*th IEEE FOCS*.

[10] L. Saloff-Coste (1997) Lectures on finite Markov chains. In: *Ecole d'Été de Probabilités de Saint-Flour*, Lect. Notes Math. **1665**, 301–413.