SÉMINAIRE DE PROBABILITÉS (STRASBOURG)

OLIVIER CATONI

Simulated annealing algorithms and Markov chains with rare transitions

Séminaire de probabilités (Strasbourg), tome 33 (1999), p. 69-119

<http://www.numdam.org/item?id=SPS_1999__33__69_0>

SIMULATED ANNEALING ALGORITHMS AND MARKOV CHAINS WITH RARE TRANSITIONS

OLIVIER CATONI

ABSTRACT. In these notes, written for a D.E.A. course at University Paris XI during the first term of 1995, we prove the essentials about stochastic optimisation algorithms based on Markov chains with rare transitions, under the weak assumption that the transition matrix obeys a large deviation principle. We present a new simplified line of proofs based on the Freidlin and Wentzell graphical approach. The case of Markov chains with a periodic behaviour at null temperature is considered. We have also included some pages about the spectral gap approach where we follow Diaconis and Stroock [13] and Ingrassia [23] in a more conventional way, except for the application to non reversible Metropolis algorithms (subsection 6.2.2) where we present an original result.

ALGORITHMES DE RECUIT SIMULÉ ET CHAÎNES DE MARKOV À TRANSITIONS RARES: Dans ces notes, tirées d'un cours de D.E.A. donné au premier trimestre 1995, nous établissons les bases de la théorie des algorithmes d'optimisation stochastiques fondés sur des chaînes de Markov à transitions rares, sous l'hypothèse faible selon laquelle la matrice des transitions vérifie un principe de grandes déviations. Nous présentons un nouvel ensemble de preuves originales fondées sur l'approche graphique de Freidlin et Wentzell. Le cas des chaînes présentant un comportement périodique à température nulle est traité. De plus nous avons aussi inclus quelques pages sur les méthodes de trou spectral, dans lesquelles nous suivons Diaconis et Stroock [13] et Ingrassia [23] d'une façon plus conventionnelle, si ce n'est pour l'application aux algorithmes de Metropolis non réversibles de la section 6.2.2, qui est originale.

INTRODUCTION

These lecture notes were written on the occasion of a course of lectures which took place from January to April 1995. We seized the opportunity of the present English translation to add some proofs which were left to the reader and to correct some misprints and omissions. Sections 4.1, 4.2 and 4.3 contain standard material from [13] and [23]. The rest is more freely inspired by the existing literature. The presentation of the cycle decomposition is new, as well as lemma 1. We chose to make weak large deviation assumptions on the transition matrix p_β at inverse temperature β, and to give results which are accordingly concerned

Date: May 1995, English translation January 1997, in revised form November 1998.

only with equivalents for the logarithm of the probability of some events of interest. In the study of simulated annealing, we considered piecewise constant temperature sequences, in order to avoid introducing specifically non-homogeneous techniques. Our aim was to give tools to study a wide variety of stochastic optimisation algorithms with discrete time and finite state space. For related results directed towards applications to statistical mechanics, we refer to [8].

1. Examples of homogeneous Markov Chains

We are going to study in this section homogeneous Markov chains related to stochastic optimisation algorithms.

1.1. The Metropolis Algorithm.

This algorithm can be applied to any finite state space E on which an energy function $U : E \to \mathbb{R}$ is defined (U can be any arbitrary real valued function). Its purpose can be either:

- to simulate the equilibrium distribution of a system from statistical mechanics with state space E and energy U interacting with a heat bath at temperature T,
- or to find a state $x \in E$ for which $U(x)$ is close to $\min_{y \in E} U(y)$.

We will mainly be interested in the second application in these notes.

Description of the algorithm

Let us consider a Markov matrix $q : E \times E \to [0,1]$ which is irreducible and reversible with respect to its invariant measure. In other words let us assume that

- $\sum_{y \in E} q(x,y) = 1, \quad x \in E,$
- $\sup_m q^m(x,y) > 0, \quad x, y \in E.$

(This last equation means that there is a path $x_0 = x, x_1, \ldots, x_m = y$ leading from x to y such that $q(x_i, x_{i+1}) > 0$, $i = 0, \ldots, l-1$.)

- the invariant probability distribution μ of q (which is unique under the preceding assumptions) is such that

$$\mu(x)q(x,y) = \mu(y)q(y,x).$$

Let us consider also an inverse temperature $\beta > 0$, $\beta \in \mathbb{R}$. To this temperature corresponds the Gibbs distribution $G(E, \mu, U, \beta)$, defined by

$$G(E, \mu, U, \beta)(x) = \frac{\mu(x)}{Z} \exp(-\beta U(x))$$

where Z (the "partition function") is

$$Z = \sum_{x \in E} \mu(x) \exp(-\beta U(x)).$$

The distribution $G(E, \mu, U, \beta)$ describes the thermal equilibrium of the thermodynamic system (E, μ, U, β). We then define the transition matrix at inverse temperature β. This is the Markov matrix $p_\beta : E \times E \to [0,1]$ defined by

$$p_\beta(x,y) \quad = \quad q(x,y) \exp -\beta(U(y) - U(x))^+, \quad x \neq y \in E,$$

where $r^+ = \max\{0, r\}$.

Proposition 1.1. *The matrix p_β is irreducible. It is aperiodic as soon as U is not constant, and therefore*

$$\forall \zeta, \nu \in \mathcal{M}_1^+(E), \quad \lim_{n \to +\infty} (\zeta - \nu) p_\beta^n = 0,$$

where $\mathcal{M}_1^+(E)$ is the set of probability measures on E. Moreover p_β is reversible with respect to $\mu_\beta = G(E, \mu, U, \beta)$.

Proof: It is irreducible because $p_\beta(x,y) > 0$ as soon as $q(x,y) > 0$. If U is not constant there are $x, y \in E$ such that $q(x,y) > 0$ and $U(x) < U(y)$, which implies that $p_\beta(x,x) > 0$ and therefore that p_β is aperiodic. Moreover

$$\begin{aligned} \mu_\beta(x) p_\beta(x,y) &= \frac{1}{Z} \mu(x) q(x,y) \exp\left(-\beta(U(x) \vee U(y))\right) \\ &= \mu_\beta(y) p_\beta(y,x), \quad x, y \in E, x \neq y. \end{aligned}$$

1.1.1. *Construction of the Metropolis algorithm.* On the canonical space $(E^{\mathbb{N}}, \mathcal{B})$ where $\mathcal{B}$ is the sigma field generated by the events depending on a finite number of coordinates, we consider the canonical process $(X_n)_{n \in \mathbb{N}}$ defined by

$$X_n(x) = x_n, \quad x \in E^{\mathbb{N}},$$

and the family of probability distributions $(P_\beta^x)_{x \in E}$ on $(E^{\mathbb{N}}, \mathcal{B})$ defined by

$$\begin{gathered} P_\beta^x \circ X_0^{-1} = \delta_x, \\ P_\beta^x(X_n = y \mid (X_0, \ldots, X_{n-1}) = (x_0, \ldots, x_{n-1})) = p_\beta(x_{n-1}, y). \end{gathered}$$

The homogeneous Markov chain $(E^{\mathbb{N}}, (X_n)_{n \in \mathbb{N}}, \mathcal{B}, (P_\beta^x)_{x \in E})$ is the canonical realization of the Metropolis algorithm with state space E, Markov matrix q, energy function U and inverse temperature β. We will use the notation $M(E, q, U, \beta)$.

1.1.2. *Computer implementation.* Assuming that $X_{n-1} = x \in E$, choose a state y according to the distribution $q(x,y)$, compute $U(y) - U(x)$, if $U(y) \leq U(x)$, put $X_n = y$, if $U(y) > U(x)$, put $X_n = y$ with probability $\exp -\beta(U(y) - U(x))$ and $X_n = x$ otherwise.

1.1.3. *Behaviour at temperature zero ($\beta = +\infty$).* Letting β tend to $+\infty$ in the definition of $M(E, q, U, \beta)$, we define the infinite inverse temperature algorithm $M(E, q, U, +\infty)$ by

$$P_{+\infty}(X_n = y \mid X_{n-1} = x) = q(x,y)\, \mathbf{1}(U(y) \leq U(x)), \quad x \neq y \in E.$$

This is a relaxation algorithm: $U(X_n)$ is almost surely non increasing. It is still homogeneous, but no more ergodic in general (if U is not constant on E, E has at least one transient component).

When β tends to infinity, $M(E, q, U, \beta)$ weakly tends to $M(E, q, U, +\infty)$, in the sense that for any function $f : E^{\mathbb{N}} \to \mathbb{R}$ depending on a finite number of coordinates we have

$$\lim_{\beta \to +\infty} \int_{E^{\mathbb{N}}} f(y) P_\beta(\mathrm{d}y) = \int_{E^{\mathbb{N}}} f(y) P_{+\infty}(\mathrm{d}y).$$

(Note that it implies that the same holds for any continuous function f, $E^{\mathbb{N}}$ being equipped with the product topology, because any such function is a uniform limit of functions depending on a finite number of coordinates.)

When it is observed during a fixed interval of time, $M(E, q, U, \beta)$ is a small perturbation of $M(E, q, U, +\infty)$ at low temperature.

We can see now that the Metropolis algorithm is suitable for the two purposes we announced at the beginning:

- *Simulation of the thermal equilibrium distribution* $G(E, \mu, U, \beta)$*:* As p_β is irreducible and aperiodic and as E is finite, $(P_\beta \circ X_0^{-1})p_\beta^n = P_\beta \circ X_n^{-1}$ tends to $G(E, \mu, U, \beta)$ when n tends to infinity (at exponential rate, as will be seen in the following).
- *Minimisation of* U*:* The Gibbs distributions $\mu_\beta = G(E, \mu, U, \beta)$ get concentrated around $\arg\min U$ when β tends to $+\infty$.

  Indeed, for any $\eta > 0$,

$$\begin{aligned} \mu_\beta(U(x) < \min U + \eta) &\geq 1 - \frac{1}{Z} \exp\left(-\beta(\eta + \min U)\right), \\ Z &\geq \mu(\arg\min U) \exp\left(-\beta \min U\right), \end{aligned}$$

  therefore we have the following rough estimate

$$\mu_\beta(U(x) < \min_E U + \eta) \geq 1 - \mu(\arg\min U)^{-1} e^{-\beta\eta}.$$

  Taking $\eta = \min\{U(y),\, y \in E \setminus \arg\min U\} - \min_E U$, we see that, as a consequence,

$$\lim_{\beta \to +\infty} G(E, \mu, U, \beta)(\arg\min U) = 1.$$

  Thus

  Proposition 1.2. *For any $\epsilon > 0$ there are $N \in \mathbb{N}$ and $\beta \in \mathbb{R}_+$ such that for any $n > N$*

$$P_\beta(U(X_n) = \min U) \geq 1 - \epsilon.$$

1.2. **The Gibbs sampler.** This algorithm is meant for a product state space $E = \prod_{i=1}^r F_i$, where the components F_i are finite sets. The purpose is the same as for the Metropolis algorithm (simulate the Gibbs distribution or minimise the energy).

Description: Let us consider

- An energy function $U : E \to \mathbb{R}$, which can in fact be any real valued function.
- An "infinite temperature" probability distribution $\mu \in \mathcal{M}_1^+$.
- An inverse temperature $\beta \in \mathbb{R}_+^*$.
- The Gibbs distribution

$$G(E, \mu, U, \beta)(x) = \frac{\mu(x)}{Z} \exp(-\beta U(x)).$$

- A permutation $\sigma \in \mathfrak{S}_r$ of $\{1, \dots, r\}$.

Let us define

- For any $i \in \{1, \dots, r\}$ the transition matrix $p^i_\beta : E \times E \to [0,1]$ at site i and inverse temperature β
$$p^i_\beta(x,y) = \mathbf{1}(\overline{y}^i = \overline{x}^i)\, G(E,\mu,U,\beta)(y \mid \overline{y}^i = \overline{x}^i), \quad x,y \in E,$$
where we have used the notations $x = (x^j)_{j=1}^r$, $x^j \in F_j$ and $\overline{x}^i = (x^j)_{j,j\neq i}$.
- The global transition matrix at temperature β
$$p_\beta = \prod_{i=1}^r p_\beta^{\sigma(i)} = p_\beta^{\sigma(1)} \cdots p_\beta^{\sigma(r)},$$
which corresponds to the scan of the sites defined by the permutation σ.

Properties of p_β:

- It is a full matrix, $(p_\beta(x,y) > 0, x,y \in E)$, thus it is irreducible and aperiodic.
- The Gibbs distribution G is p^i_β invariant for any $i \in \{1, \dots, r\}$, therefore G is also the (unique) invariant probability measure of p_β.

We consider then the Markov chain with canonical realization $(E^{\mathbb{N}}, (X_n)_{n\in\mathbb{N}}, \mathcal{B}, P_\beta)$ where P_β is the probability measure on $(E^{\mathbb{N}}, \mathcal{B})$ of the Markov chain defined by $P_\beta \circ X_0^{-1}$ and
$$P(X_n = y \mid X_{n-1} = x) = p_\beta(x,y), \quad x,y \in E.$$

The homogeneous Markov chain (X, P_β) is called a Gibbs sampler with state space E, energy function U, reference measure μ, scan function σ, inverse temperature β and initial distribution $P_\beta \circ X_0^{-1} = \mathcal{L}_0$. The notation $GS(E,\mu,\sigma,U,\beta,\mathcal{L}_0)$ will denote this process in the following. Let us describe its computer implementation with more details.

Computer implementation:

Each step of the chain corresponds to one scan of all the sites, in the order defined by σ. It includes thus r sub-steps.

To perform the ith sub-step, $i = 1, \dots, r$, if x is the starting configuration, we have to draw at random $f \in F_{\sigma(i)}$ according to the conditional thermal equilibrium distribution at site $\sigma(i)$ knowing that the configuration should coincide with x on the other sites.

This computation is easy if

- The number of elements of $F_{\sigma(i)}$ is small,
- The conditional distribution $G(X^{\sigma(i)} = f \mid X^j = x^j, j \neq \sigma(i))$ depends on few coordinates, as it is the case for a Markov random field. The new state at the end of the ith sub-step is $y \in E$, given by $y^{\sigma(i)} = f$ and $y^j = x^j$, $j \neq \sigma(i)$.

Behaviour at "zero temperature": Here again $\lim_{\beta\to+\infty} p^i_\beta$ exists, therefore $\lim_{\beta\to+\infty} p_\beta$ exists and defines a Markov chain at temperature zero. This zero temperature dynamic is a relaxation algorithm: the energy is almost surely non-increasing. It is not in general an ergodic process, and P_β converges weakly to $P_{+\infty}$, as in the case of the Metropolis dynamic. Moreover the purposes of simulation of the equilibrium distribution and of minimisation of the energy are fulfilled in the same way, and, as for the Metropolis algorithm, proposition 1.2 holds also for the Gibbs sampler.

2. Markov chains with rare transitions

2.1. Construction. We are going to put the two previous examples into a more general framework. Let us consider

- An arbitrary finite state space E,
- A rate function $V : E \times E \to \mathbb{R}_+ \cup \{+\infty\}$. Assume that V is irreducible in the sense that the matrix $\exp(-V(x,y))$ is irreducible.
- A family $\mathcal{F} = (E^{\mathbb{N}}, (X_n)_{n\in\mathbb{N}}, \mathcal{B}, P_\beta)_{\beta\in\mathbb{R}_+}$ of homogeneous Markov chains indexed by a real positive parameter β.

Definition 2.1. The family of homogeneous Markov chains $\mathcal{F}$ is said to have rare transitions with rate function V if for any $x, y \in E$

$$\lim_{\beta\to+\infty} \frac{-\log P_\beta(X_n = y \mid X_{n-1} = x)}{\beta} = V(x,y),$$

(with the convention that $\log 0 = -\infty$).

Remarks about this definition:

- This is a large deviation assumption with speed β and rate function V about the transition matrix. We will see that it implies large deviation estimates for the exit time and point from any subdomain of E.
- The two examples of algorithms given previously fit into this framework. Indeed the rate function of the Metropolis algorithm $M(E, q, U, \beta, \mathcal{L}_0)$ is

$$V(x,y) = \begin{cases} (U(y) - U(x))_+ & \text{if } p_\beta(x,y) > 0 \text{ for } \beta > 0 \\ +\infty & \text{otherwise.} \end{cases}$$

As for the Gibbs Sampler $GS(E, \mu, \sigma, U, \beta, \mathcal{L}_0)$ with $E = \prod_{i=1}^r F_i$, the rate function V is built in the following way:

For any $x, y \in E$, any $i \in \{1, \dots, r\}$, let us put

$$V^i(x,y) = \begin{cases} U(y) - \inf\{U(z) \mid \overline{z}^i = \overline{x}^i\}, & \text{if } \overline{x}^i = \overline{y}^i \\ +\infty & \text{otherwise,} \end{cases}$$

and let us consider the path $\gamma = (\gamma_k)_{k=0}^r$ defined by

$$\gamma_k^{\sigma(i)} = \begin{cases} y^{\sigma(i)} & \text{if } i \le k, \\ x^{\sigma(i)} & \text{otherwise.} \end{cases}$$

The rate function of the Gibbs sampler is

$$V(x,y) = \sum_{k=1}^r V^{\sigma(k)}(\gamma_{k-1}, \gamma_k).$$

2.2. Rate function induced by a potential.

Definition 2.2. We will say that the rate function $V : E \times E \to \mathbb{R}_+ \cup \{+\infty\}$ is induced by the potential $U : E \to \mathbb{R}$ if for all $x, y \in E$

$$U(x) + V(x,y) = U(y) + V(y,x),$$

with the convention that $+\infty + r = +\infty$ for any $r \in \mathbb{R}$.

Proposition 2.1. *The rate function of the Metropolis algorithm $M(E, \mu, U, \beta, \mathcal{L}_0)$ is induced by U.*

Proof:
As q is irreducible, $\mu(x) > 0$ for any $x \in E$. Indeed there is x_0 such that $\mu(x_0) > 0$ and there is n such that $q^n(x_0, x) > 0$, therefore $\mu(x) = \mu q^n(x) \geq \mu(x_0) q^n(x_0, x) > 0$. Thus $q(x,y) > 0$ if and only if $q(y,x) > 0$, from the μ reversibility of q. Therefore $V(x,y) = +\infty$ if and only if $V(y,x) = +\infty$. In the case when $q(x,y) > 0$, $x \neq y$,

$$V(x,y) - V(y,x) = (U(y) - U(x))^+ - (U(x) - U(y))^+ = U(y) - U(x). \qquad \square$$

3. Lemmas on irreducible Markov chains

Let E be a finite state space, $p : E \times E \to [0,1]$ an irreducible Markov matrix, $(E^{\mathbb{N}}, (X_n)_{n \in \mathbb{N}}, \mathcal{B}, P)$ an homogeneous Markov chain with transition matrix p, $W \subset E$ a given subset of E and $\overline{W} = E \setminus W$ its complement. For any oriented graph $g \subset E \times E$ and any $x \in E$, we write $g(x) = \{y \mid (x,y) \in g\}$ and more generally $g^n(x) = \bigcup_{y \in g^{n-1}(x)} g(y)$.

Definition 3.1. We let $G(W)$ be the set of oriented graphs $g \subset E \times E$ satisfying

1. For any $x \in E$, $|g(x)| = \mathbf{1}_{\overline{W}}$ (no arrow starts from W, exactly one arrow starts from each state outside W).
2. For any $x \in E$, $x \notin O_g(x)$, where $O_g(x) = \bigcup_{n=1}^{+\infty} g^n(x)$ is the orbit of x under g, (g is without loop).

Equivalently, the second condition can be replaced by: For any $x \in E \setminus W, O_g(x) \cap W \neq \emptyset$ (any point in $\overline{W}$ leads to W).

Definition 3.2. For any $x \in E$, $y \in W$, we will write

$$G_{x,y}(W) = \begin{cases} \{g \in G(W) \mid y \in O_g(x)\} & \text{if } x \in \overline{W} \\ G(W) & \text{if } x = y \\ \emptyset & \text{if } x \in W \setminus \{y\}. \end{cases}$$

Thus $G_{x,y}(W)$ is the set of graphs $g \in G(W)$ linking x to y. We will also write

$$G_{A,B}(W) = \{g \mid \forall x \in A, \exists y \in B \text{ such that } g \in G_{x,y}(W)\}.$$

We will give three formulas which express the equilibrium distribution of p, the probability distribution of the hitting point of W, and the expectation of the corresponding hitting time, as the ratio of two finite sums of positive terms.

They have been introduced in the large deviation theory of random dynamical systems by Freidlin and Wentzell [16]. The idea of using graphs to compute determinants has been known since the nineteenth century and presumably goes back to Kirchhoff [24]. The proofs which we propose are based on a preliminary lemma:

Lemma 3.1. *For any $W \subset E$, $W \neq \emptyset$, let $p_{|\overline{W}\times\overline{W}}$ be the matrix p restricted to $\overline{W}\times\overline{W}$:*

$$p_{|\overline{W}\times\overline{W}}(x,y) = p(x,y)\mathbf{1}(x \notin W)\mathbf{1}(y \notin W).$$

Let $\tau(W)$ be the first hitting time of W: $\tau(W) = \inf\{n \geq 0 | X_n \in W\}$. For any $x, y \in \overline{W}$ we have

$$\begin{aligned}
(\mathrm{id}_{|\overline{W}} - p_{|\overline{W}\times\overline{W}})^{-1}(x,y) &= \left(\sum_{n=0}^{+\infty} p^n_{|\overline{W}\times\overline{W}}\right)(x,y) \\
&= E_\beta\left(\sum_{n=0}^{\tau(W)} 1(X_n = y) \,|\, X_0 = x\right) \\
&= \left(\sum_{g\in G_{x,y}(W\cup\{y\})} p(g)\right)\left(\sum_{g\in G(W)} p(g)\right)^{-1},
\end{aligned}$$

where $p(g) = \prod_{(z,t)\in g} p(z,t)$.

Remark: The fact that $\mathrm{id}_{\overline{W}} - p_{\overline{W}\times\overline{W}}$ is non singular is a consequence of the fact that p is irreducible ($\lim_n p^n_{|\overline{W}\times\overline{W}} = 0$ and therefore all the eigenvalues of $p_{|\overline{W}\times\overline{W}}$ are of module lower than one).

Lemma 3.2. *The (unique) invariant probability distribution of p is given by*

$$\mu(x) = \left(\sum_{g\in G(\{x\})} p(g)\right)\left(\sum_{y\in E}\sum_{g\in G(\{y\})} p(g)\right)^{-1}, \; x \in E.$$

Lemma 3.3. *The distribution of the first hitting point can be expressed as*

$$P(X_{\tau(W)} = y \mid X_0 = x) = \left(\sum_{g\in G_{x,y}(W)} p(g)\right)\left(\sum_{g\in G(W)} p(g)\right)^{-1},$$

for any $W \neq \emptyset$, $x \in \overline{W}$, $y \in W$.

Lemma 3.4. *For any $W \neq \emptyset$, any $x \in \overline{W}$,*

$$E(\tau(W) \mid X_0 = x) = \left(\sum_{y\in\overline{W}}\sum_{g\in G_{x,y}(W\cup\{y\})} p(g)\right)\left(\sum_{g\in G(W)} p(g)\right)^{-1}.$$

Proof of lemma 3.1:

As p is irreducible, for any $W \neq \emptyset$, there is $g \in G(W)$ such that $p(g) > 0$ (the proof of this is left to the reader).

Let us write for any $x, y \in \overline{W}$

$$m(x,y) = \left(\sum_{g \in G_{x,y}(W \cup \{y\})} p(g)\right)\left(\sum_{g \in G(W)} p(g)\right)^{-1}.$$

We want to check that for any $x, y \in \overline{W}$

$$\text{(1)} \qquad \sum_{z \in \overline{W}} (\mathrm{id}(x,z) - p(x,z))\, m(z,y) = \mathrm{id}(x,y).$$

Using the equality

$$p(x,x) = 1 - \sum_{z \in E \setminus \{x\}} p(x,z),$$

we can equivalently check that

$$\text{(2)} \qquad \sum_{z \in \overline{\{x\}}} p(x,z) m(x,y) = \mathrm{id}(x,y) + \sum_{z \in \overline{W \cup \{x\}}} p(x,z) m(z,y).$$

The left hand side of this equation is equal to

$$\left(\sum_{(z,g) \in C_1} p(x,z)p(g)\right)\left(\sum_{g \in G(W)} p(g)\right)^{-1}.$$

where $C_1 = \{(z,g) \in \overline{\{x\}} \times G(W \cup \{y\}) \;:\; g \in G_{x,y}(W \cup \{y\})\}$, the right hand side is equal to

$$\mathrm{id}(x,y) + \left(\sum_{(z,g) \in C_2} p(x,z)p(g)\right)\left(\sum_{g \in G(W)} p(g)\right)^{-1},$$

where

$$C_2 = \{(z,g) \in \overline{W \cup \{x\}} \times G(W \cup \{y\}) \mid g \in G_{z,y}(W \cup \{y\})\}.$$

Let us consider first the case when $x \neq y$. Then we can define a one to one mapping $\varphi : C_1 \to C_2$ by

$$\varphi(z,g) = \begin{cases} (z,g) & \text{if } g \in G_{z,y}(W \cup \{y\}), \\ (g(x), (g \cup \{(x,z)\}) \setminus \{(x,g(x))\}) & \text{if } g \notin G_{z,y}(W \cup \{y\}). \end{cases}$$

The easiest way to check that φ is one to one is to check that

$$\varphi^{-1}(z,g) = \begin{cases} (z,g) & \text{if } g \in G_{x,y}(W \cup \{y\}), \\ (g(x), (g \cup \{(x,z)\}) \setminus \{(x,g(x))\}) & \text{if } g \notin G_{x,y}(W \cup \{y\}). \end{cases}$$

Let us write $\varphi = (\varphi_1, \varphi_2)$ to show the two components of φ. The following change of variable

$$\begin{aligned}
\sum_{(z,g)\in C_2} p(x,z)p(g) &= \sum_{(z,g)\in C_1} p(x,\varphi_1(z,g))p(\varphi_2(z,g)) \\
&= \sum_{(z,g)\in C_1} p(x,z)p(g)
\end{aligned}$$

shows that

$$\sum_{z\in\overline{\{x\}}} p(x,z)m(x,y) = \sum_{z\in\overline{W\cup\{x\}}} p(x,z)m(z,y).$$

We have now to check the case when $x = y$. In this case $C_2 \subset C_1$. Let us consider the one to one mapping $\varphi : C_1 \setminus C_2 \to G(W)$ defined by $\varphi(z,g) = g \cup \{(x,z)\}$, with inverse $\varphi^{-1}(g) = (g(x), g \setminus \{(x, g(x))\})$.

We have

$$\sum_{(z,g)\in C_1\setminus C_2} p(x,z)p(g) = \sum_{g\in G(W)} p(g),$$

and therefore

$$\begin{aligned}
&\left(\sum_{(z,g)\in C_1} p(x,z)p(g)\right)\left(\sum_{g\in G(W)} p(g)\right)^{-1} \\
&\qquad = 1 + \left(\sum_{(z,g)\in C_2} p(x,z)p(g)\right)\left(\sum_{g\in G(W)} p(g)\right)^{-1}. \qquad \square
\end{aligned}$$

Proof of lemma 3.3:

$$\begin{aligned}
P(X_{\tau(W)} = y \,|\, X_0 = x) &= \sum_{z\in\overline{W}}\sum_{n=0}^{+\infty} P(X_n = z, \tau(W) > n \,|\, X_0 = x)p(z,y) \\
&= \sum_{z\in\overline{W}} \left(\sum_{g\in G_{x,z}(W\cup\{z\})} p(g)p(z,y)\right)\left(\sum_{g\in G(W)} p(g)\right)^{-1} \\
&= \left(\sum_{g\in G_{x,y}(W)} p(g)\right)\left(\sum_{g\in G(W)} p(g)\right)^{-1}.
\end{aligned}$$

Proof of lemma 3.4 :

$$\begin{aligned} E(\tau(W)\,|\,X_0 = x) &= E\left(\sum_{n=0}^{\tau(W)-1} 1(X_n \in \overline{W})\,|\,X_0 = x\right) \\ &= E\left(\sum_{y\in\overline{W}}\sum_{n=0}^{+\infty} 1(X_n = y, \tau(W) > n)\,|\,X_0 = x\right) \\ &= \frac{\displaystyle\sum_{y\in\overline{W}}\ \sum_{g\in G_{x,y}(W\cup\{y\})} p(g)}{\displaystyle\sum_{g\in G(W)} p(g)}. \end{aligned}$$

Proof of lemma 3.2 :
Let $\nu(x) = \inf\{n \geq 1\,|\,X_n = x\}$.

$$\begin{aligned} \mu(x) &= E(\nu(x)\,|\,X_0 = x)^{-1} \\ &= \left(\sum_{y,y\neq x} p(x,y)E(\tau(\{x\})\,|\,X_0 = y) + 1\right)^{-1} \\ &= \left(\sum_{y,y\neq x}\ \sum_{z,z\neq x}\ \sum_{g\in G_{y,z}(\{x,z\})} p(x,y)p(g) + \sum_{g\in G(\{x\})} p(g)\right)^{-1}\left(\sum_{g\in G(\{x\})} p(g)\right) \\ &= \left(\sum_{g\in G(\{x\})} p(g)\right)\left(\sum_{z\in E}\ \sum_{g\in G(\{z\})} p(g)\right)^{-1}, \end{aligned}$$

because for any $z \neq x$ $\varphi_z : \{(y,g)\,|\,y \neq x, g \in G_{y,z}(\{x,z\})\} \to G(\{z\})$ defined by $\varphi_z(y,g) = g \cup \{(x,y)\}$ is one to one.

4. Cycle decomposition of a family of Markov chains with rare transitions

4.1. Behaviour of the invariant distribution, virtual energy.

Definition 4.1. The rate function $V : E \times E \longrightarrow \mathbb{R}_+ \cup \{+\infty\}$ is said to be irreducible when the matrix $(\exp -V(x,y))_{(x,y)\in E^2}$ is irreducible. This means namely that for any $x, y \in E$ there is a path $z_0 = x, \ldots, z_r = y$ such that

$$V(z_{i-1}, z_i) < +\infty, \quad i = 1, \cdots, r.$$

Proposition 4.1. *Let $\mathcal{F} = (E^{\mathbb{N}}, (X_n)_{n\in\mathbb{N}}, \mathcal{B}, P_\beta)_{\beta\in\mathbb{R}_+}$ be a family of homogeneous Markov chains with rare transitions, with irreducible rate function V. Then for β large enough (X, P_β) is irreducible and its invariant probability distribution μ_β is such that for any $x \in E$*

$$\lim_{\beta\to+\infty} -\beta^{-1}\log\mu_\beta(x) = \tilde{U}(x) \in \mathbb{R}_+.$$

The "virtual energy" function $\tilde{U} : E \to \mathbb{R}$ can be expressed as

$$\tilde{U}(x) = \min_{g \in G(\{x\})} V(g) - \min_{y \in E} \min_{g \in G(\{y\})} V(g),$$

where $V(g) = \sum_{(z,t) \in g} V(z,t)$. *In the case when V is induced by a potential function U, we have for any $x \in E$ that $\tilde{U}(x) = U(x) - \min_{y \in E} U(y)$.*

Corollary 4.1. *The family $\mathcal{F}$ describes an optimisation algorithm for the minimisation of the virtual energy $\tilde{U}$: For any $\epsilon > 0$, there are $N \in \mathbb{N}$ and $\beta \in \mathbb{R}_+$ such that, for any $n > N$,*

$$\min_{x \in E} P_\beta(\tilde{U}(X_n) = 0 \mid X_0 = x) \geq 1 - \epsilon.$$

This algorithm is called a "generalised Metropolis algorithm".

Proof: The first part of the proposition is a straightforward consequence of lemma 2. In the case when V is induced by U, consider the one to one mapping

$$\varphi : G(\{y\}) \longrightarrow G(\{x\}),$$

defined by

$$\varphi(g) = \{(z,t) \in g, t \notin O_g(x)\} \cup \{(t,z), (z,t) \in g, t \in O_g(x)\}.$$

It is obtained by reversing in $g \in G(\{y\})$ the path leading from x to y. We have

$$\begin{aligned}
\tilde{U}(y) + U(x) + \min_{z \in E} \min_{g \in G(\{z\})} V(g) &= \min_{g \in G(\{y\})} (V(g) + U(x)) \\
&= \min_{g \in G(\{y\})} \left(\sum_{\substack{z \notin O_g(x) \cup \{x\} \\ (z,t) \in g}} V(z,t) + U(x) + \sum_{\substack{z \in O_g(x) \cup \{x\} \\ (z,t) \in g}} V(z,t) \right) \\
&= \min_{g \in G(\{y\})} \left(\sum_{\substack{z \notin O_g(x) \cup \{x\} \\ (z,t) \in g}} V(z,t) + U(y) + \sum_{\substack{z \in O_g(x) \cup \{x\} \\ (z,t) \in g}} V(t,z) \right) \\
&= \min_{g \in G(\{y\})} (U(y) + V(\varphi(y))) \\
&= U(y) + \min_{g \in G(\{x\})} V(g) \\
&= U(y) + \tilde{U}(x) + \min_{z \in E} \min_{g \in G(\{z\})} V(g),
\end{aligned}$$

The proof of the corollary is the same as in the case of the classical Metropolis algorithm when the chain is aperiodic. When the chain has period d, then each chain $(X_{nd+k})_{n \in \mathbb{N}}$ is aperiodic for $k \in \{0, \ldots, d-1\}$, and the combination of the inequalities obtained for these d processes gives the result for $(X_n)_{n \in \mathbb{N}}$. $\square$

4.2. **Large deviation estimates for the exit time and exit point from a subdomain.** In this paragraph we will study the limiting behaviour of the law of the exit time and exit point from an arbitrary subdomain D of E. Let us recall some notations introduced in section 3:

$$\overline{D} = E \setminus D,$$
$$\tau(D) = \inf\{n \in \mathbb{N} : X_n \in D\}.$$

Proposition 4.2. *For any $D \subset E, D \neq \emptyset$, for any $x \in D$, under the same hypotheses as previously,*

$$\lim_{\beta \to +\infty} \frac{\log E_\beta(\tau(\overline{D}) \mid X_0 = x)}{\beta} = \min_{g \in G(\overline{D})} V(g) - \min_{y \in \overline{D}} \min_{g \in G_{x,y}(\overline{D} \cup \{y\})} V(g),$$

moreover, for any $y \in \overline{D}$

$$\lim_{\beta \to +\infty} -\frac{1}{\beta} P_\beta(X_{\tau(\overline{D})} = y \,|\, X_0 = x) = \min_{g \in G_{x,y}(\overline{D})} V(g) - \min_{g \in G(\overline{D})} V(g).$$

We will use the following notations for these new rate functions:

$$\lim_{\beta \to +\infty} -\beta^{-1} \log P_\beta(X_{\tau(\overline{D})} = y \,|\, X_0 = x) \stackrel{\text{def}}{=} V_D(x, y)$$
$$\lim_{\beta \to +\infty} \beta^{-1} \log E_\beta(\tau(\overline{D}) \mid X_0 = x) \stackrel{\text{def}}{=} H_D(x).$$

In the next paragraph, we will link the rate functions appearing in these two large deviation estimates with the virtual energy $\widetilde{U}$. For this purpose, we will introduce the decomposition of the state space into cycles due to Freidlin and Wentzell.

4.3. **Definition of cycles.**

Definition 4.2. Under the preceding hypotheses, a subdomain $C \subset E$ is said to be a cycle if it is a one point set or if for any $x, y \in C$, $x \neq y$, the probability, starting from x, to leave C without visiting y is exponentially small, by which we mean that

$$\lim_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(X_{\tau(\overline{C} \cup \{y\})} \neq y \,|\, X_0 = x) > 0.$$

As a consequence we have of course

$$\lim_{\beta \to +\infty} P_\beta(X_{\tau(\overline{C} \cup \{y\})} = y \,|\, X_0 = x) = 1.$$

4.4. **Some properties of cycles.**

Proposition 4.3. *The subdomain C of E is a cycle if and only if it is a one point set or for any $x, y \in C$, $x \neq y$ the number $N_C(x, y)$ of round trips including x and y performed by the chain starting from x before it leaves C satisfies*

$$\lim_{\beta \to +\infty} \frac{1}{\beta} \log E_\beta(N_C(x, y) \mid X_0 = x) > 0.$$

Remark: This property justifies the name "cycle".

Proof: Let us give a more formal mathematical definition of $N_C(x,y)$. For this, let us introduce the sequences of stopping times $(\mu_k(x,y), \nu_k(x,y))_{n\in\mathbb{N}}$ defined by the following induction

$$\begin{aligned}
\nu_{-1}(x,y) &= 0\\
\mu_k(x,y) &= \inf\{n > \nu_{k-1}(x,y) \,:\, X_n \in \{y\}\cup\overline{C}\}\\
\nu_k(x,y) &= \inf\{n > \mu_k(x,y) \,:\, X_n \in \{x\}\cup\overline{C}\},
\end{aligned}$$

then $N_C(x,y) = \inf\{k \,:\, X_{\mu_k(x,y)} \notin C \text{ or } X_{\nu_k(x,y)} \notin C\}$.

We have

$$\begin{aligned}
E_\beta(N_C(x,y)\,|\,X_0 = x) &= \sum_{n=0}^{+\infty} P_\beta(N_C(x,y) > n\,|\,X_0 = x)\\
&= \sum_{n=0}^{+\infty} \left(P_\beta(X_{\mu_0(x,y)} = y \text{ and } X_{\nu_0(x,y)} = x\,|\,X_0 = x)\right)^n\\
&= \left(1 - P_\beta(X_{\mu_0(x,y)} = y \text{ and } X_{\nu_0(x,y)} = x\,|\,X_0 = x)\right)^{-1}.
\end{aligned}$$

Moreover

$$\begin{aligned}
&P_\beta(X_{\mu_0(x,y)} = y \text{ and } X_{\nu_0(x,y)} = x\,|\,X_0 = x)\\
&\quad = P_\beta(X_{\tau(\overline{C}\cup\{y\})} = y\,|\,X_0 = x)P_\beta(X_{\tau(\overline{C}\cup\{x\})} = x\,|\,X_0 = y)\\
&\quad = \left(1 - P_\beta(X_{\tau(\overline{C}\cup\{y\})} \neq y\,|\,X_0 = x)\right)\left(1 - P_\beta(X_{\tau(\overline{C}\cup\{x\})} \neq x\,|\,X_0 = y)\right).
\end{aligned}$$

Therefore

$$\begin{aligned}
&\lim_{\beta\to+\infty} \frac{1}{\beta}\log E_\beta(N_C(x,y)\,|\,X_0 = x)\\
&\qquad\qquad = \inf\{V_{C\setminus\{z\}}(t,u) \,:\, (z,t) \in \{(x,y),(y,x)\} \text{ and } u \in \overline{C}\},
\end{aligned}$$

which proves that

$$\lim_{\beta\to+\infty} \frac{1}{\beta}\log E_\beta(N_C(x,y)\,|\,X_0 = x) > 0$$

for all $x, y \in C$, $x \neq y$, if and only if $V_{C\setminus\{y\}}(x,z) > 0$ for all $(x,y) \in C^2$, $z \in \overline{C}$.
□

Proposition 4.4. *Let $\mathcal{C}(E,V)$ be the set of cycles of (E,V). It has a tree structure for the inclusion relation, with root E and leaves the one point sets. This means that if C_1 and C_2 are cycles, either $C_1 \subset C_2$ or $C_2 \subset C_1$ or $C_1\cap C_2 = \emptyset$.*

Proof: If it were the case that $x \in C_1 \cap C_2$, $y \in C_1 \setminus C_2$ and $z \in C_2 \setminus C_1$, we would obtain a contradiction: we would have

$$\begin{aligned} 0 &= \lim_{\beta \to +\infty} \frac{1}{\beta} \log P_\beta(X_{\tau(\overline{C}_2 \cup \{z\})} = z \,|\, X_0 = x) \\ &\leq \lim_{\beta \to +\infty} \frac{1}{\beta} \log P_\beta(X_{\tau(\overline{C}_1 \cup \{y\})} \neq y \,|\, X_0 = x) \\ &< 0. \quad \square \end{aligned}$$

Proposition 4.5. *For any subdomain D of E, we define the principal boundary $B(D)$ of D by*

$$B(D) = \{y \notin D \,:\, V_D(x,y) = 0 \textit{ for some } x \in D\}$$

Then for any cycle $C \in \mathcal{C}(E,V)$, any subdomain $D \subset C$, $D \neq \emptyset$, $D \neq C$, $B(D) \subset C$.

Proof:
If $y \in B(D) \setminus C$, $x \in D$, $z \in C \setminus D$, then

$$P_\beta(X_{\tau(\overline{D})} = y \mid X_0 = x) \leq P(X_{\tau(\overline{C} \cup \{z\})} = y \mid X_0 = x),$$

because in this case $y \in \overline{C} \cup \{z\}$ and $\overline{C} \cup \{z\} \subset \overline{D}$. This is in contradiction with the fact that

$$\lim_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(X_{\tau(\overline{D})} = y \,|\, X_0 = x) = 0$$

and

$$\lim_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(X_{\tau(\overline{C} \cup \{z\})} = y \,|\, X_0 = x) > 0.$$

Therefore $B(D) \subset C$. $\square$

An important property of a cycle is that, at low temperature, the exit time and exit point become independent from the starting point when it belongs to the cycle.

Proposition 4.6 (Independence from the starting point). *For any cycle $C \in \mathcal{C}(E,V)$, any $x \in C$, $y \in C$, $z \notin C$,*

$$V_C(x,z) = V_C(y,z) \stackrel{\text{def}}{=} V(C,z)$$

and

$$H_C(x) = H_C(y) \stackrel{\text{def}}{=} H(C).$$

The quantity $H(C)$ is called the depth of the cycle C.

Proof:

$$\begin{aligned} P_\beta(X_{\tau(\overline{C})} = z \,|\, X_0 = y) &= P_\beta(X_{\tau(\overline{C})} = z \,|\, X_0 = x) P_\beta(X_{\tau(\overline{C} \cup \{x\})} = x \,|\, X_0 = y) \\ &\quad + P_\beta(X_{\tau(\overline{C} \cup \{x\})} = z \,|\, X_0 = y) \\ &\geq P_\beta(X_{\tau(\overline{C})} = z \,|\, X_0 = x) P_\beta(X_{\tau(\overline{C} \cup \{x\})} = x \,|\, X_0 = y). \end{aligned}$$

Therefore $V_C(y,z) \leq V_C(x,z) + V_{C\setminus\{x\}}(y,x)$ by the definition of cycles $V_{C\setminus\{x\}}(y,x) = 0$, therefore $V_C(y,z) \leq V_C(x,z)$ and, exchanging x and y, $V_C(y,z) = V_C(x,z)$. Similarly we have

$$E_\beta(\tau(\overline{C}) \mid X_0 = x) = \sum_{u \in C} E_\beta\left(\sum_{n=0}^{\tau(\overline{C})} 1(X_n = u) \mid X_0 = x\right)$$

and

$$E_\beta\left(\sum_{n=0}^{\tau(\overline{C})} \mathbf{1}(X_n = u) \mid X_0 = x\right) = E_\beta\left(\sum_{n=0}^{\tau(\overline{C})} \mathbf{1}(X_n = u) \mid X_0 = u\right) \times P_\beta(X_{\tau(\overline{C}\cup\{u\})} = u \mid X_0 = x).$$

Therefore $H_C(x)$ is independent of $x \in C$. $\square$

Now we will give some properties of cycles linked with W-graph computations:

Proposition 4.7 (characterisation of cycles in terms of W-graphs). *A subset C of E is a cycle if and only if it is either a one point set or satisfies: for any $y \in C$, any $\hat{g} \in \arg \min_{g \in G(\overline{C}\cup\{y\})} V(g)$, $\hat{g}(C \setminus \{y\}) \subset C$.*

Proof: For any subset C of E, $|C| > 1$, any $y \in C$,

$$\begin{aligned}
&\min_{x \in C\setminus\{y\}} \lim_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(X_{\tau(\overline{C}\cup\{y\})} \neq y \mid X_0 = x) \\
&\quad = \min_{x \in C\setminus\{y\}} \min_{z \notin C} \min_{g \in G_{x,z}(\overline{C}\cup\{y\})} V(g) - \min_{g \in G(\overline{C}\cup\{y\})} V(g) \\
&\quad = \min_{g \in G(\overline{C}\cup\{y\}), g(C\setminus\{y\}) \not\subset C} V(g) - \min_{g \in G(\overline{C}\cup\{y\})} V(g),
\end{aligned}$$

therefore $\min_{x \in C\setminus\{y\}} \lim_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(X_{\tau(\overline{C}\cup\{y\})} \neq y \mid X_0 = x) > 0$ if and only if $\arg \min_{g \in G(\overline{C}\cup\{y\})} V(g) \subset \{g \in G(\overline{C} \cup \{y\}) \,:\, g(C \setminus \{y\}) \subset C\}$. $\square$

Proposition 4.8 (leading terms in a $\overline{C}$-graph). *For any cycle $C \in \mathcal{C}(V)$, any $x \in C$, any $y \notin C$ such that $V(C,y) < +\infty$, any graph $\hat{g} \in \arg\min_{g \in G_{x,y}(\overline{C})} V(g)$, $\hat{g}(C) \subset C \cup \{y\}$.*

Proof: Let us consider the state $z \in O_{\hat{g}}(x) \cup \{x\}$ such that $(z,y) \in \hat{g}$. Let $\tilde{g} \in \arg \min_{g \in G(\overline{C}\cup\{z\})} V(g)$, then according to the preceding proposition $\tilde{g}(C \setminus \{z\}) \subset C$, therefore $\tilde{g} \cup \{(z,y)\}$ belongs to $G_{x,y}(\overline{C})$. Thus

$$V(z,y) + V(\tilde{g}) = V(\tilde{g} \cup \{(z,y)\}) \geq V(\hat{g}) = V(\hat{g} \setminus \{(z,y)\}) + V(z,y).$$

This shows that $V(\hat{g} \setminus \{(z,y)\}) = \arg \min_{g \in G(\overline{C}\cup\{z\})} V(g)$, and therefore that $\hat{g}(C \setminus \{z\}) \subset C$. $\square$

Proposition 4.9 (local computation of the virtual energy). *For any cycle $C \in \mathcal{C}(V)$, any $x, y \in C$,*

$$\tilde{U}(x) - \tilde{U}(y) = \min_{g \in G(\overline{C} \cup \{x\})} V(g) - \min_{g \in G(\overline{C} \cup \{y\})} V(g).$$

This shows that the computation of the virtual energy within a cycle up to an additive constant depends only on the restriction of the rate function V to this cycle.

Proof: For any graph $g \in E \times E$, any subset $A \subset E$, let us put $g_{|A} = \{(u,v) \in g : u \in A\}$. With the notations of the proposition, let $g \in G(\{x\})$, then $g_{|C} \in G(\overline{C} \cup \{x\})$, $g_{|\overline{C}} \in G(C)$ and

$$V(g) = V(g_{|C}) + V(g_{|\overline{C}}).$$

Therefore

$$\min_{g \in G(\{x\})} V(g) \geq \min_{g \in G(\overline{C} \cup \{x\})} V(g) + \min_{g \in G(C)} V(g).$$

On the other hand, if $\hat{g} \in \arg\min_{g \in G(\overline{C} \cup \{x\})} V(g)$ and $g \in G(C)$, then $g \cup \hat{g}$ is without loop, because $\hat{g}(C) \subset C$, and thus $g \cup \hat{g} \in G(\{x\})$, and

$$\min_{g \in G(C)} V(g) + V(\hat{g}) = \min_{g \in G(C)} V(g \cup \hat{g}) \geq \min_{g \in G(\{x\})} V(g).$$

We have proved that $\min_{g \in G(C)} V(g) + \min_{g \in G(\overline{C} \cup \{x\})} V(g) = \min_{g \in G(\{x\})} V(g)$ and the proposition follows from the fact that

$$\tilde{U}(x) - \tilde{U}(y) = \min_{g \in G(\{x\})} V(g) - \min_{g \in G(\{y\})} V(g). \qquad \square$$

4.5. **Iterative construction of cycles and the virtual energy function.** For any subset $D \subset E$, let us put $\tilde{U}(D) = \min_{x \in D} \tilde{U}(x)$.

Proposition 4.10. *Let $E = \bigcup_{i \in I} C_i$ be a partition of E into disjoint cycles. Assume that it is not trivial, namely that $|I| \geq 2$. Let us consider on $\mathcal{C}_I = \{C_i \mid i \in I\}$ the graph s of the typical jumps, defined by*

$$s = \{(C_i, C_i) \mid i \in I\} \cup \{(C_i, C_j) \mid B(C_i) \cap C_j \neq \emptyset\}.$$

Let $\mathcal{C}_J = \{C_j \mid j \in J\}$ be an irreducible and stable component of s, that is a component of $\mathcal{C}_I$ for the equivalence relation

$$\mathcal{R}_s = \{(C_i, C_j) : i, j \in I, C_i \in O_s(C_j) \text{ and } C_j \in O_s(C_i)\} \cup \{(C_i, C_i) : i \in I\},$$

such that $s(\mathcal{C}_J) \subset \mathcal{C}_J$. There exists at least one such component, because s induces on $\mathcal{C}_J / \mathcal{R}_s$ a graph without loop, which has therefore at least one leaf (or terminal node). Moreover J is not reduced to one point, because this would mean that the principal boundary of the would be unique cycle in $\mathcal{C}_J$ would be empty, which is impossible.

Then $C = \bigcup_{j\in J} C_j$ *is a cycle, and* $C_j, j \in J$ *are the maximal strict subcycles of* C *for the inclusion relation. Moreover, for any* $i, j \in J$

$$\tilde{U}(C_i) + H(C_i) = \tilde{U}(C_j) + H(C_j),$$
$$H(C) = \min_{j\in J, y\notin C} V(C_j, y) + \max_{j\in J} H(C_j),$$

and for any $y \notin C$

$$V(C,y) = \min_{j\in J} V(C_j, y) - \min_{z\notin C, j\in J} V(C_j, z).$$

Remark: This proposition allows to build iteratively all the cycles, starting from the trivial partition of E into one point sets, computing in the same time the quantities $\tilde{U}(x) - \tilde{U}(C)$, $x \in C$, $H(C)$ and $V(C,y), y \notin C$.
Proof: Let $y \in C$. We will prove that for any $\hat{g} \in \arg\min_{g\in G(\overline{C}\cup\{y\})} V(g)$, $\hat{g}(C) \subset C$.

Let us assume that $y \in C_{j_0}$. As $\mathcal{C}_J$ is a component of $\mathcal{C}_I/\mathcal{R}_s$, it is possible to extract from $s/\mathcal{C}_J$ an oriented tree α with root C_{j_0} (we mean by this that a graph without loop connecting each point of $\mathcal{C}_J$ to C_{j_0}). Let $g^{j_0} \in \arg \min_{g\in G(\overline{C}_{j_0}\cup\{y\})} V(g)$ and for any $j \in J\setminus\{j_0\}$, let $g^j \in \arg\min_{g\in G(\overline{C_j})} V(g)$ be such that $g^j(C_j) \subset C_j \cup \alpha(C_j)$. Such a g^j exists, according to proposition 4.8, because $B(C_j)\cap\alpha(C_j) \neq \emptyset$. The graph $\bigcup_{j\in J} g^j$ is without loop, and therefore belongs to $G(\overline{C}\cup\{y\})$, thus

$$\sum_{j\in J} V(g^j) = V\left(\bigcup_{j\in J} g^j\right) \geq V(\hat{g}) = \sum_{j\in J} V(\hat{g}_{|C_j}).$$

This proves that

$$V(\hat{g}_{|C_j}) = \min_{g\in G(\overline{C}_j)} V(g), \qquad j \in J\setminus\{j_0\}$$

and

$$V(\hat{g}_{|C_{j_0}}) = \min_{g\in G(\overline{C}_{j_0}\cup\{y\})} V(g),$$

and therefore that

$$\hat{g}(C_{j_0}) = C_{j_0}$$

and

$$\hat{g}(C_j) \subset C_j \cup B(C_j), \qquad j \in J\setminus\{j_0\}.$$

Thus $\hat{g}(C) = C$.

This shows that C is a cycle. Let us prove now that $C_j \subset C$ are maximal among the subcycles of C distinct from C itself.

Assume that for some $j_0 \in J$ and some cycle $C' \in \mathcal{C}(V)$, $C_{j_0} \subset C' \subset C$, $C_{j_0} \neq C'$. As $\mathcal{C}(V)$ is a tree (proposition 4.4), there is $J' \subset J$ such that $C' = \bigcup_{j\in J'} C_j$ and $\{j_0\} \neq J'$. From a preceding proposition, for any $j \in J'$, $B(C_j) \subset C'$, since $C_j \neq C'$. Therefore, $s(C_j, j \in J') \subset \{C_j, j \in J'\}$, which implies that $J' = J$ and therefore that $C' = C$.

From the local computation of the virtual energy into cycles, we see that

$$\begin{aligned}\tilde{U}(C_i) - \tilde{U}(C_j) &= \min_{x \in C_i} \min\{V(g) \,:\, g \in G(\overline{C} \cup \{x\})\\ &\quad - \min_{y \in C_j} \min\{V(g) \,:\, g \in G(\overline{C} \cup \{y\})\}.\end{aligned}$$

From the preceding computation, for any $x \in C_i$

$$\begin{aligned}\min\{V(g) \,:\, g \in G(\overline{C} \cup \{x\})\} &= \sum_{k \in J \setminus \{i\}} \min\{V(g) \,:\, g \in G(\overline{C}_k)\}\\ &+ \min\{V(g) \,:\, g \in G(\overline{C}_i \cup \{x\})\}.\end{aligned}$$

Therefore

$$\begin{aligned}\tilde{U}(C_i) - \tilde{U}(C_j) &= \min_{x \in C_i} \min\{V(g) \,:\, g \in G(\overline{C}_i \cup \{x\})\}\\ &\quad - \min_{y \in C_j} \min\{V(g) \,:\, g \in G(\overline{C}_j \cup \{y\})\}\\ &+ \min\{V(g) \,:\, g \in G(\overline{C}_j)\} - \min\{V(g) \,:\, g \in G(\overline{C}_i)\}\\ &= H(C_j) - H(C_i).\end{aligned}$$

Similarly

$$\begin{aligned}H(C) &= \min\{V(g) \,:\, g \in G(\overline{C})\} - \min_{x} \min\{V(g) \,:\, g \in G(\overline{C} \cup \{x\})\}\\ &= \min_{z \in \overline{C}} \min_{j \in J} \Big\{ \min\{V(g) \,:\, g \in G_{C_j,z}(\overline{C}_j)\}\\ &\quad + \sum_{k \in J \setminus \{j\}} \min\{V(g) \,:\, g \in G(\overline{C}_k)\} \Big\}\\ &\quad - \min\{V(g) \,:\, x \in C, g \in G(\overline{C} \cup \{x\})\}\\ &= \min_{z \in \overline{C}} \min_{j \in J} V(C_j, z) + \sum_{k \in J} \min\{V(g) \,:\, g \in G(\overline{C}_k)\}\\ &- \min\{V(g) \,:\, g \in G(\overline{C} \cup \{x\}), x \in C\}\\ &= \min_{z \in \overline{C}} \min_{j \in J} V(C_j, z) - \max_{j \in J} H(C_j).\end{aligned}$$

We have also for any $x \in C$,

$$\begin{aligned}V(C,y) &= \min_{g \in G_{C,y}(\overline{C})} V(g) - \min_{g \in G(\overline{C})} V(g)\\ &= \min_{j} \min_{g \in G_{C_j,y}(\overline{C}_j)} V(g) + \sum_{k \in J \setminus \{j\}} \min_{g \in G(\overline{C}_j)} V(g)\\ &\quad - \min_{z} \min_{j} \left(\min_{g \in G_{C_j,z}(\overline{C}_j)} V(g) + \sum_{k \in J \setminus \{j\}} \min_{g \in \overline{C}_j} V(g) \right)\\ &= \min_{j \in J} V(C_j, y) - \min_{z} \min_{j \in J} V(C_j, z). \qquad \square\end{aligned}$$

4.6. **Maximal depth and maximal partition of a domain.** In this subsection we will compute the maximal depth $H(D) \stackrel{\text{def}}{=} \max_{x \in D} H_D(x)$, of a domain $D \subset E$ in terms of the maximal partition of D defined below:

Definition 4.3. For any domain $D \subset E$, we let $\mathcal{M}(D)$ be the set of maximal elements of $\{C \in \mathcal{C}(E)\,;\, C \subset D\}$ for the inclusion relation. Due to the tree structure of $\mathcal{C}(E)$, this is a partition of D. We call it the maximal partition of D.

From the graph point of view, the maximal partition has the important following property:

Lemma 4.1. *For any domain $D \subsetneq E$*

$$\min_{g \in G(\overline{D})} V(g) = \sum_{C \in \mathcal{M}(D)} \min_{g \in G(\overline{C})} V(g). \tag{3}$$

Proof. A first remark is that for any $g \in G(\overline{D})$

$$V(g) = \sum_{C \in \mathcal{M}(D)} V(g_{|C}),$$

This proves that the left hand side of equation (3) is not smaller than the right hand side. To prove the reverse inequality, consider the graph s on $\mathcal{M}(D) \cup \{\overline{D}\}$ defined by

$$(C_1, C_2) \in s \text{ iff } C_1 \in \mathcal{M}(D) \text{ and } C_2 \cap B(C_1) \neq \emptyset.$$

Then according to proposition 4.10, s is without any stable irreducible component and connects every cycle of $\mathcal{M}(D)$ to $\overline{D}$. Therefore it can be spanned by a disjoint union of oriented trees leading to $\overline{D}$, from which we can build as in the proof of proposition 4.10 a graph $\hat{g} \in G(\overline{D})$ such that $\hat{g}_{|C} = \min\{V(g)\,;\, g \in G(\overline{C})\}$ for any $C \in \mathcal{M}(D)$, proving that the right hand side is not smaller than the left hand side of equation (3). □

We are ready now to compute the maximal depth of a domain:

Proposition 4.11. *For any domain $D \subsetneq E$ let us define*

$$H(D) = \max_{x \in D} H_D(x).$$

Then

$$H(D) = \max\{H(C)\,;\, C \in \mathcal{M}(D)\}.$$

Proof. Let us put for short for any set of graphs G

$$V(G) = \min_{g \in G} V(g).$$

By definition we have

$$H(D) = V(G(\overline{D})) - \min_{y \in D} V(G(\overline{D} \cup \{y\})).$$

For any fixed $y \in D$, let $C \in \mathcal{M}(D)$ be such that $y \in C$. Remarking that

$$\mathcal{M}(D \setminus \{y\}) = \mathcal{M}(D \setminus C) \cup \mathcal{M}(C \setminus \{y\}),$$

and using the previous lemma we get that

$$\begin{aligned} H(D) &= \max_{C \in \mathcal{M}(D)} \max_{y \in C} V(G(\overline{C})) - V(G(\overline{C} \cup \{y\})) \\ &= \max_{C \in \mathcal{M}(D)} H(C). \end{aligned}$$

□

4.7. **Computing the cycles in term of path elevations.** In order to give a description of cycles and therefore of the behaviour of the trajectories which recalls what happens in the case when the rate function V derives from an energy function U, we will introduce a characterisation of the energy based on paths instead of graphs.

Energy barrier between two points

For any two states $x, y \in E$, let $\Gamma_{x,y}$ be the set of paths joining x to y:

$$\Gamma_{x,y} = \{(x_0, \dots, x_r) \,:\, r > 0, x_0 = x, x_r = y\} \subset \bigcup_r E^r.$$

For any path $\gamma = (x_0, \dots, x_r)$, let

$$H(\gamma) = \max_{i=1,\dots,r} \tilde{U}(x_{i-1}) + V(x_{i-1}, x_i),$$

with the convention that when $r = 0$ we put

$$H((x_0)) = \tilde{U}(x_0).$$

The energy barrier between x and y is defined to be

$$H(x, y) = \min_{\gamma \in \Gamma_{x,y}} H(\gamma).$$

Proposition 4.12 (energy barrier of a cycle). *For any cycle $C \in \mathcal{C}(E, V)$, any $y \notin C$, we have*

$$\min_{x \in C} \tilde{U}(x) + V(x, y) = \tilde{U}(C) + H(C) + V(C, y).$$

Proof:

$$\begin{aligned} &\min_{x \in C} \tilde{U}(x) + V(x, y) - \tilde{U}(C) \\ &= \min_{x \in C} \left\{ \min_{g \in G(\overline{C} \cup \{x\})} V(g) + V(x, y) \right\} - \min_{z \in C} \min_{g \in G(\overline{C} \cup \{z\})} V(g) \\ &= \min_{g \in G_{C,y}(\overline{C})} V(g) - \min_{z \in C} \min_{g \in G(\overline{C} \cup \{z\})} V(g) \\ &= V(C, y) + H(C). \quad \square \end{aligned}$$

Proposition 4.13 (elevation of paths within a cycle). *For any cycle $C \in \mathcal{C}(V)$, any $x \in C$, any $y \notin C$, there is a path $\varphi \in \Gamma_{x,y}$, $\varphi = (\varphi_0, \dots, \varphi_s)$ such that $\varphi_i \in C$, $i = 0, \dots, s-1$ and $H(\varphi) = \tilde{U}(C) + H(C) + V(C, y)$. For any x, $y \in C$, there is a path $\varphi = (\varphi_0, \dots, \varphi_s) \in \Gamma_{x,y}$ such that $\varphi_i \in C$, $i = 0, \dots, s$ and $H(\varphi) \leq \tilde{U}(C) + \sup\{H(\tilde{C}) \,|\, \tilde{C} \in \mathcal{C}, \tilde{C} \subset C, \tilde{C} \neq C\}$ (with the convention that* $\sup \emptyset = 0$*).*

Proof. Let us proceed by induction on the size of cycles. For any $x, y \in C$, there are $C_0, \dots, C_k \in \mathcal{C}(V)$ such that $C_i \subset C$, $C_i \neq C$, C_i are maximal, $B(C_{i-1}) \cap C_i \neq \emptyset$, $i = 1, \dots, k$, $x \in C_0$, $y \in C_k$. This is a consequence of proposition 4.10 on the iterative construction of cycles. Let y_i, $i = 1, \dots, k$ be a point in $B(C_{i-1}) \cap C_i$ and let $y_0 = x$. According to our induction hypothesis that proposition 4.13 is true for the strict subcycles of C, we can find paths $\varphi^i \in \Gamma_{y_{i-1}, y_i}$, $i = 1, \dots, k$, such that $\varphi^i \subset C$ and $H(\varphi^i) = \tilde{U}(C_{i-1}) + H(C_{i-1})$. We can also find $\varphi^{k+1} \in \Gamma_{y_k, y}$ such that $\varphi^{k+1} \subset C_k \subset C$ and $H(\varphi^{k+1}) \leq \tilde{U}(C_k) + H(C_k)$. The concatenated path $\varphi_{x,y} = (\varphi^1, \dots, \varphi^{k+1}) \in \Gamma_{x,y}$ belongs to C and has an elevation lower than $\tilde{U}(C) + \max\{H(\tilde{C}) \; : \; \tilde{C} \in \mathcal{C}(V), \tilde{C} \subset C, \tilde{C} \neq C\}$. Let us now consider $x \in C$ and $z \notin C$, we can find according to proposition 4.12 a point y such that $\tilde{U}(y) + V(y, z) = \tilde{U}(C) + H(C) + V(C, z)$. Let $\varphi_{x,y}$ be constructed as above. The path $(\varphi_{x,y}, z)$ is included in C except its end point z and has an elevation equal to $\tilde{U}(C) + H(C) + V(C, z)$. Proposition 4.13 being easily seen to be true for one point cycles is therefore proved by induction. □

Proposition 4.14. *The elevation function is symmetric:*

$$H(x, y) = H(y, x), \quad x, y \in E.$$

Proof: Let $C_1 \in \mathcal{C}(V)$ be the largest cycle such that $x \in C_1$, $y \notin C_1$. Let $C_2 \in \mathcal{C}(V)$ be the largest cycle such that $x \notin C_2$, $y \in C_2$. Let $C_3 \in \mathcal{C}(V)$ be the smallest cycle such that $\{x, y\} \in C_3$. The cycles C_1 and C_2 are maximal strict subcycles of C_3, therefore $H(x, y) = H(C_1) + \tilde{U}(C_1) = H(C_2) + \tilde{U}(C_2) = H(y, x)$. □

Proposition 4.15. *For any cycle $C \in \mathcal{C}$,*

$$H(C) = \max_{x \in C} \min_{y \notin C} H(x, y) - \tilde{U}(x),$$

and more generally for any $D \subset E$, $D \neq E$, $D \neq \emptyset$,

$$H(D) = \max_{x \in D} \min_{y \notin D} H(x, y) - \tilde{U}(x).$$

Proof. The case of a cycle is a direct consequence of propositions 4.12 and 4.13 . In the case of a general domain D, one has to consider the maximal partition $\mathcal{M}(D)$ of D and apply proposition 4.11, to see that if C_0 is one of the deepest cycles in $\mathcal{M}(D)$ then $H(D) = H(C_0)$. Taking x in the bottom of C_0, and remarking that

$$\min_{y \notin D} H(x, y) \geq \min_{y \notin C_0} H(x, y),$$

we get that

$$H(D) = H(C_0) \leq \max_{x \in D} \min_{y \notin D} H(x, y) - \tilde{U}(x).$$

Now, for the converse, let x be any point in D and let C_0 be the maximal cycle of $\mathcal{M}(D)$ to which x belongs. As seen in the proof of equation (3), there is a sequence of cycles $C_0, \dots, C_r$ such that $B(C_i) \cap C_{i+1} \neq \emptyset$, $i = 0, \dots, r-1$ and $B(C_r) \cap \overline{D} \neq \emptyset$. Pemark that $\tilde{U}(C_i) + H(C_i)$ is decreasing: indeed, taking

$u \in C_i$ and $v \in B(C_i) \cap C_{i+1}$, we see that $\tilde{U}(C_i) + H(C_i) = H(u,v) = H(v,u) \geq \tilde{U}(C_{i+1}) + H(C_{i+1})$. With the help of proposition 4.13 we build a path γ, starting at x, going through this sequence of cycles and ending in $\overline{D}$ such that

$$\begin{aligned}
\min_{y \notin D} H(x,y) &\leq H(\gamma) \\
&\leq \max_{0 \leq i \leq r} \tilde{U}(C_i) + H(C_i) \\
&= \tilde{U}(C_0) + H(C_0) \\
&\leq \tilde{U}(x) + H(D).
\end{aligned}$$

□

Proposition 4.16 (Weak reversibility condition of Hajek and Trouvé). *Let $U : E \longrightarrow \mathbb{R}$ be an arbitrary real valued function defined on E. Let the elevation $H_U(\gamma)$ of a path $\gamma = (z_0, \dots, z_r) \in E^{r+1}$ with respect to U and V be defined by*

$$H_U(\gamma) = \max_{i=1,\dots,r} U(z_{i-1}) + V(z_{i-1}, z_i).$$

Let $H_U(x,y) = \min_{\gamma \in \Gamma_{x,y}} H_U(\gamma)$, $x, y \in E$, $x \neq y$.

Then $U(x) = \tilde{U}(x) + \min_{y \in E} U(y)$, $x \in E$ if and only if H_U is symmetric.

Proof: See question 10.3 of the appendix for some hints about the proof.

Proposition 4.17. *For any $x, y \in E$,*

$$\tilde{U}(y) \leq \tilde{U}(x) + V(x,y).$$

Consequently for any path $\gamma = (\gamma_0, \dots, \gamma_r)$

$$H(\gamma) \leq \tilde{U}(\gamma_0) + \sum_{k=1}^{r} V(\gamma_{k-1}, \gamma_k).$$

Proof:

$$\tilde{U}(y) \leq H(y,x) = H(x,y) \leq \tilde{U}(x) + V(x,y). \qquad \square$$

4.8. **Another construction of cycles.** For any $\lambda \in \mathbb{R}$, let us introduce the equivalence relation

$$\mathcal{R}_\lambda = \{(x,y) \in E^2 \mid x \neq y, H(x,y) < \lambda\} \cup \{(x,x) \mid x \in E\}.$$

Proposition 4.18. *The components of $E/\mathcal{R}_\lambda$ are the one point sets $\{x\}$ such that $\tilde{U}(x) \geq \lambda$ and the cycles $C \in \mathcal{C}(V)$ such that*

$$(4) \qquad \max\{\tilde{U}(\tilde{C}) + H(\tilde{C}) \mid \tilde{C} \in \mathcal{C}, \tilde{C} \subset C, \tilde{C} \neq C\} < \lambda \leq \tilde{U}(C) + H(C).$$

Thus $\mathcal{C}(V) = \bigcup_{\lambda \in \mathbb{R}_+} E/\mathcal{R}_\lambda$.

Proof: If $C \in \mathcal{C}(V)$ satisfies equation (4), then $C \in E/\mathcal{R}_\lambda$ according to previous propositions. On the other hand, let us compute for any $x \in E$ the component of x in $E/\mathcal{R}_\lambda$. Let us consider the maximal sequence of distinct cycles $C_0 = \{x\} \subset C_1 \subset C_2 \subset \cdots \subset E$ containing x (C_i is the smallest cycle strictly containing

$C_{i-1})$. If $H(C_0)+\tilde{U}(C_0) > \lambda$, then $\{x\} \in E/\mathcal{R}_\lambda$, otherwise let us consider $i_0 = \min\{i \,|\, H(C_i)+\tilde{U}(C_i) \geq \lambda\}$, then according to the first part of the proof $x \in C_{i_0} \in E/\mathcal{R}_\lambda$. $\square$

4.9. **Exit time from a subdomain.** From Freidlin and Wentzell's lemma we deduce that

Proposition 4.19. *For any subdomain $D \subset E$, $D \neq E$, any $x \in D$, any $\epsilon > 0$,*

$$\lim_{\beta\to+\infty} -\frac{1}{\beta}\log P_\beta(\tau(\overline{D}) > e^{\beta(H(D)+\epsilon)} \mid X_0 = x) = +\infty,$$

and

$$\liminf_{\beta\to+\infty} -\frac{1}{\beta}\log(\min_{y\in D} P_\beta(\tau(\overline{D}) < e^{\beta(H(D)-\epsilon)} \mid X_0 = y) \geq \epsilon,$$

where $H(D) = \max_{y\in D} H_D(y)$.

Proof: Applying the Markov property, we see that:

$$\begin{aligned}
&P(\tau(\overline{D}) > e^{\beta(H(D)+\epsilon)} \mid X_0 = x) \\
&\qquad \leq \left(\max_{y\in D} P(\tau(\overline{D}) > e^{\beta(H(D)+\epsilon/2)} \mid X_0 = y)\right)^{\lfloor e^{\beta\epsilon/2}\rfloor} \\
&\qquad \leq \left(\max_{y\in D} E(\tau(\overline{D}) \mid X_0 = y)e^{-\beta(H(D)+\epsilon/2)}\right)^{\lfloor e^{\beta\epsilon/2}\rfloor} \\
&\qquad\qquad\qquad \leq \exp -\beta\left(\frac{\epsilon}{4}\left(e^{\beta\epsilon/2}-1\right)\right),
\end{aligned}$$

To prove the second equation, let us notice that

$$\sum_{k=0}^{+\infty} P(\tau(\overline{D}) > ke^{\gamma\beta} \mid X_0 = x) \geq E(e^{-\gamma\beta}\tau(\overline{D}) \mid X_0 = x),$$

and that

$$\begin{aligned}
\sum_{k=0}^{+\infty} P(\tau(\overline{D}) > ke^{\gamma\beta} \mid X_0 = x) &\leq \sum_{k=0}^{+\infty}\left(\max_{y\in D} P(\tau(\overline{D}) > e^{\gamma\beta} \mid X_0 = y)\right)^k \\
&= \left(\min_{y\in D} P(\tau(\overline{D}) \leq e^{\gamma\beta} \mid X_0 = y)\right)^{-1},
\end{aligned}$$

thus

$$\min_{y\in D} P(\tau(\overline{D}) \leq e^{\gamma\beta} \mid X_0 = y) \leq e^{\gamma\beta}E(\tau(\overline{D}) \mid X_0 = x)^{-1}. \qquad \square$$

Proposition 4.20. *For any cycle $C \in \mathcal{C}(V)$, any sufficiently small $\epsilon > 0$, any $x \in C$,*

$$\liminf_{\beta\to+\infty} -\frac{1}{\beta}P_\beta(\tau(\overline{C}) < e^{\beta(H(C)-\epsilon)} \mid X_0 = x) \geq \epsilon.$$

Proof:

For any $x, y \in C$, $\gamma > 0$,

$$\begin{aligned} P(\tau(\overline{C}) < e^{\gamma\beta} \mid X_0 = x) &\leq P(\tau(\overline{C}) < e^{\beta\gamma} \mid X_0 = y) P(X_{\tau(\overline{C}\cup\{y\})} = y \mid X_0 = x) \\ &\quad + P(X_{\tau(\overline{C}\cup\{y\})} \neq y \mid X_0 = x). \end{aligned}$$

Let

$$\begin{aligned} \epsilon_0 &= \min_{x,y\in C, x\neq y} \lim_{\beta\to+\infty} -\frac{1}{\beta} \log P(X_{\tau(\overline{C}\cup\{y\})} \neq y \mid X_0 = x) \\ &= \min\{V_{C\setminus\{y\}}(x,z) \,:\, x,y \in C, x \neq y, z \in \overline{C}\} > 0, \end{aligned}$$

then for all $\epsilon < \epsilon_0$ and β large enough

$$P(\tau(\overline{C}) < e^{\beta\gamma} \mid X_0 = x) \leq \min_{y\in\overline{C}} P(\tau(\overline{C}) < e^{\beta\gamma} \mid X_0 = y) + e^{-\beta\epsilon}.$$

We end the proof by taking $\gamma = H(C) - \epsilon$ and applying the preceding proposition.
□

5. Convergence towards equilibrium

Proposition 5.1. *For any cycle $C \in \mathcal{C}(V)$, any $\gamma > 0$ such that $H(\{t \in C \mid \tilde{U}(t) > \tilde{U}(C)\}) < \gamma < H(C)$, any $x, y \in C$,*

$$\liminf_{\beta\to+\infty} -\frac{1}{\beta} \log P(X_{\lfloor e^{\beta\gamma}\rfloor} = y, \tau(\overline{C}) > e^{\gamma\beta} \mid X_0 = x) \geq \tilde{U}(y) - \tilde{U}(C).$$

Corollary 5.1.

$$\liminf_{\beta\to+\infty} -\frac{1}{\beta} \log P(\tilde{U}(X_{\lfloor e^{\beta\gamma}\rfloor}) \neq \tilde{U}(C) \mid X_0 = x) > 0$$

Proof: Let us put $N = \lfloor e^{\gamma\beta} \rfloor$. Let $A = \arg\min_{x\in C} \tilde{U}(x)$. For any $x, y \in C$,

$$\begin{aligned} P_\beta(X_n = y, \tau(\overline{C}) > e^{\gamma\beta} \mid X_0 = x) &\leq P_\beta(\tau(\overline{C} \cup A) > e^{\gamma\beta} \mid X_0 = x) \\ &\quad + \sup_{k\in\mathbb{N}, z\in A} P(X_k = y \mid X_0 = z). \end{aligned}$$

Let $f_k(x) = P_\beta(X_k = x \mid X_0 = z)\mu_\beta(x)^{-1}$. We have $\sum_{x\in E} f_k(x) p_\beta(x,y) \frac{\mu_\beta(x)}{\mu_\beta(y)} = f_{k+1}(y)$, and $\sum_{x\in E} p_\beta(x,y) \frac{\mu_\beta(x)}{\mu_\beta(y)} = 1$, therefore

$$\max_{y\in E} f_k(x) \leq \max_{y\in E} f_0(x) = \frac{1}{\mu_\beta(z)},$$

and $\sup_{k\in\mathbb{N}} P(X_k = y \mid X_0 = z) \leq \frac{\mu_\beta(y)}{\mu_\beta(z)}$. □

Proposition 5.2. *Let us assume that $C \in \mathcal{C}(V)$ is such that for some $z \in \arg\min_{x\in C} \tilde{U}(x)$, considering the graph s of the null cost jumps,*

$$s = \{(x,y) \in E^2 \,:\, V(x,y) = 0\},$$

the orbit $O_s(z)$ is aperiodic. Then for any $x, y \in C$, any γ such that $H(C \setminus \{z\}) < \gamma < H(C)$,

$$\lim_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = y, \tau(\overline{C}) > e^{\gamma\beta} \mid X_0 = x) = \tilde{U}(y) - \tilde{U}(C).$$

Proof:

Let us consider the Markov chain $(Y_n)_{n \in \mathbb{N}}$ on C with transitions

$$P_\beta(Y_n = y \mid Y_{n-1} = x) = \lim_{M \to +\infty} P_\beta(X_n = y \mid X_{n-1} = x, \tau(\overline{C}) > M).$$

The existence of this limit is a consequence of the Perron-Frobenius theorem applied to the (non stochastic) aperiodic irreducible non negative matrix $p_{\beta|C\times C}$. This theorem says that

$$p_{\beta|C\times C} = \rho\, \pi_1 + R \circ \pi_2,$$

where (π_1, π_2) forms a system of projectors (i.e. $\pi_1 \circ \pi_2 = \pi_2 \circ \pi_1 = 0$ and $\pi_1 + \pi_2 = \mathrm{Id}$), where π_1 is the projection on the one dimensional vector space generated by a positive eigenvector, where $\rho > 0$ is the spectral radius of $p_{\beta|C\times C}$ and where the spectral radius of R is strictly lower than ρ. This implies that

$$\lim_{M \to +\infty} \frac{\delta_z (p_{\beta|C\times C})^M \mathbf{1}}{\delta_y (p_{\beta|C\times C})^M \mathbf{1}} = \lim_{M \to +\infty} \frac{P_\beta(\tau(\overline{C}) > M \mid X_0 = z)}{P_\beta(\tau(\overline{C}) > M \mid X_0 = y)}$$

exists for any $y, z \in C$ and is equal to

$$\frac{\delta_z \pi_1 \mathbf{1}}{\delta_y \pi_1 \mathbf{1}}.$$

Therefore as soon as $p_\beta(x,y) > 0$,

$$\frac{P_\beta(X_n = z \mid X_{n-1} = x, \tau(\overline{C}) > M)}{P_\beta(X_n = y \mid X_{n-1} = x, \tau(\overline{C}) > M)} = \frac{p_\beta(x,z)}{p_\beta(x,y)} \frac{P_\beta(\tau(\overline{C}) > M - n \mid X_0 = z)}{P_\beta(\tau(\overline{C}) > M - n \mid X_0 = y)}$$

has a limit when M tends to infinity, which proves in turn the existence of the limit defining the transitions of Y at temperature β.

Now that the definition of Y is justified, let us return to the main stream of our proof. We have

$$\begin{aligned} &P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = y, \tau(\overline{C}) > \lfloor e^{\gamma\beta} \rfloor \mid X_0 = x) P_\beta(\tau(\overline{C}) > M - \lfloor e^{\gamma\beta} \rfloor \mid X_0 = y) \\ &\quad = P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = y, \tau(\overline{C}) > M \mid X_0 = x), \end{aligned}$$

and therefore

$$\begin{aligned} &P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = y, \tau(\overline{C}) > e^{\gamma\beta} \mid X_0 = x) \\ &\quad = P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = y \mid \tau(\overline{C}) > M, X_0 = x) \frac{P_\beta(\tau(\overline{C}) > M \mid X_0 = x)}{P_\beta(\tau(\overline{C}) > M - \lfloor e^{\gamma\beta} \rfloor \mid X_0 = y)}. \end{aligned}$$

Moreover

$$\begin{aligned} &P_\beta(\tau(\overline{C}) > M \mid X_0 = x) = \\ &\quad \sum_{z \in C} P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = z, \tau(\overline{C}) > e^{\gamma\beta} \mid X_0 = x) P_\beta(\tau(\overline{C}) > M - \lfloor e^{\gamma\beta} \rfloor \mid X_0 = z). \end{aligned}$$

Let $K = M - \lfloor e^{\gamma\beta} \rfloor$,

$$P_\beta(\tau(\overline{C}) > K \mid X_0 = z) \geq P_\beta(X_{\tau(\overline{C} \cup \{y\})} = y \mid X_0 = z) P_\beta(\tau(\overline{C}) > K \mid X_0 = y),$$

therefore

$$\limsup_{\beta \to +\infty} \sup_{K \in \mathbb{N}} \left| \frac{P_\beta(\tau(\overline{C}) > K \mid X_0 = z)}{P_\beta(\tau(\overline{C}) > K \mid X_0 = y)} - 1 \right| = 0.$$

Thus

$$\begin{aligned} &\lim_{\beta \to +\infty} \sup_{M > e^{\gamma\beta}} \left| \frac{P_\beta(\tau(\overline{C}) > M \mid X_0 = x)}{P_\beta(\tau(\overline{C}) > M - \lfloor e^{\gamma\beta} \rfloor \mid X_0 = y)} - 1 \right| \\ &= \lim_{\beta \to +\infty} \left| \sum_{z \in C} P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = z, \tau(\overline{C}) > e^{\gamma\beta} \mid X_0 = x) - 1 \right| \\ &= 0, \end{aligned}$$

and, letting $M \to +\infty$,

$$\lim_{\beta \to +\infty} \frac{P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = y, \tau(\overline{C}) > e^{\gamma\beta} \mid X_0 = x)}{P(Y_{\lfloor e^{\gamma\beta} \rfloor} = y \mid Y_0 = x)} = 1.$$

In the same way, we can prove that for any $x, y \in C$,

$$\lim_{\beta \to +\infty} \frac{p_\beta(x,y)}{P_\beta(Y_1 = y \mid Y_0 = x)} = 1.$$

Therefore Y is a Markov chain with rare transitions and rate function $V_{C \times C}$. According to proposition 4.9, the virtual energy of Y is $(\tilde{U}(x) - \tilde{U}(C))_{x \in C}$. Therefore it is enough to prove the proposition in the special case when $C = E$. We will assume in the following of the proof that we are in this case. Let us consider the family of product Markov chains $\left((E \times E)^{\mathbb{N}}, (X^1, X^2)_{n \in \mathbb{N}}, \mathcal{B} \otimes \mathcal{B}, P_\beta^1 \otimes P_\beta^2\right)_{\beta \in \mathbb{R}_+}$, where P_β^1 and P_β^2 have the same transitions as P_β and have the following initial distributions:

$$\begin{aligned} P_\beta^1 \circ (X_0^1)^{-1} &= \delta_x, \\ P_\beta^2 \circ (X_0^2)^{-1} &= \mu_\beta, \end{aligned}$$

(here μ_β is as usual the invariant distribution at inverse temperature β). It is a family of Markov chains with rare transitions with rate function

$$V^2(x,y) = V(x^1, y^1) + V(x^2, y^2).$$

Moreover $H^2((E \times E) \setminus \{(z,z)\}) = H(E \setminus \{z\})$. Indeed there is n_0 such that for any $n \geq n_0$, there is a path $(\varphi_1, \dots, \varphi_n)$ such that $\varphi_1 = \varphi_n = z$ and $V(\varphi_{i-1}, \varphi_i) = 0$. For any $x \in E$ there is an infinite path $(\psi_n)_{n \in \mathbb{N}^*}$ such that $\psi_1 = x$ and $\tilde{U}(\psi_i) \leq \tilde{U}(x)$, (take a path such that $V(\psi_{i-1}, \psi_i) = 0$). Moreover, for any $i \in \mathbb{N}$, $H(\psi_i, z) \leq H(x, z)$, indeed

$$\begin{aligned} H(\psi_i, z) &\leq \max(H(\psi_i, x), H(x, z)) \\ &= \max(H(x, \psi_i), H(x, z) \\ &= \max(\tilde{U}(x), H(x, z)) \\ &= H(x, z). \end{aligned}$$

With these two types of paths, it is easy to build in $E \times E$ a path $\psi \in \Gamma_{(x,y),(z,z)}$ such that

$$H(\psi) \leq (\tilde{U}(x) + H(y, z)) \vee H(x, z)$$

(Let the first component follow ψ while the second component is led to z via a path of minimal elevation $H(y, z)$, then let the first component follow a path of minimum elevation, while the second component follows a path φ of suitable length.) This proves that $H^2(E \times E \setminus \{(z, z)\}) = H(E \setminus \{z\})$, because it cannot obviously be lower.

Now for any $y \in C$, putting $N = \lfloor e^{\gamma\beta} \rfloor$, applying the Markov property at time $\tau^2(\{(z, z)\})$, and remarking that X^1 and X^2 conditioned by the same initial condition have the same distribution, we have

$$\begin{aligned} P_\beta^1 \otimes P_\beta^2(X_N^1 = y) &\geq P_\beta^1 \otimes P_\beta^2((X_N^1 = y) \text{ and } \tau^2(\{(z, z)\}) \leq N) \\ &= P_\beta^1 \otimes P_\beta^2((X_N^2 = y) \text{ and } \tau^2(\{(z, z)\}) \leq N) \\ &\geq P_\beta^2(X_N^2 = y) - P_\beta^1 \otimes P_\beta^2(\tau^2(\{(z, z)\}) > N). \end{aligned}$$

(This argument is equivalent to considering a "coupled" Markov chain where X^1 and X^2 are glued together once they meet.) As

$$\lim_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta^1 \otimes P_\beta^2(\tau^2(\{(z, z)\}) > N) = +\infty,$$

we get the desired result. □

Theorem 5.1 (convergence rate). *Let us put*

$$\begin{aligned} H_1 &= H(E \setminus \arg\min \tilde{U}) \\ H_2 &= H(E \setminus \{z\}), \; z \in \arg\min \tilde{U}, \\ H_3 &= H^2((E \times E) \setminus \Delta), \end{aligned}$$

where the value of H_2 is independent from the choice of $z \in \arg\min_{x \in E} \tilde{U}(x)$ and where $\Delta = \{(x, x) : x \in E\}$. For any $\gamma > H_1$, any $x \in E$, $y \in E$,

$$\liminf_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = y \mid X_0 = x) \geq \tilde{U}(y).$$

For any $\gamma > H_2$

$$\lim_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(\tau(\{z\}) > e^{\gamma\beta} \mid X_0 = x) = +\infty, \; x \in E, z \in \arg\min_{x \in E} \tilde{U}(x).$$

For any $\gamma > H_3$, any x, $y \in E$

$$\lim_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = y \mid X_0 = x) = \tilde{U}(y).$$

In general the constants $H_1 \leq H_2 \leq H_3$ are distinct. However, when the null cost graph $s = \{(x,y) \in E^2 \mid V(x,y) = 0\} \cup \Delta$ has an aperiodic component in $\arg\min \tilde{U}$, *we have $H_2 = H_3$. Moreover if* $\arg\min \tilde{U}$ *is a one point set, then $H_1 = H_2 = H_3$.*

Eventually the following non-convergence results holds: for any $\gamma < H_1$, there is $x \in E$ such that

$$\liminf_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(\tilde{U}(X_{\lfloor e^{\gamma\beta} \rfloor}) = 0 \mid X_0 = x) > 0;$$

for any $\gamma < H_2$, any $z \in$ $\arg\min \tilde{U}$, *there is $x \in E$ such that*

$$\liminf_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(\tau(\{z\}) \leq e^{\gamma\beta} \mid X_0 = x) > 0,$$

for any $\gamma < H_3$, any $z \in$ $\arg\min \tilde{U}$, *there is $x \in E$ such that*

$$\limsup_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = z \mid X_0 = x) > 0.$$

Remark 5.1. The second and the third critical depths are distinct when the chain is "almost" periodic on the set on ground states, that is when it behaves as a periodic chain on a time scale larger than $e^{H_2\beta}$. The non convergence results show that H_1, H_2 and H_3 are sharp.

Proof. The first convergence result is a consequence of proposition 5.1, the second one is a consequence of proposition 4.19, and the third one is proved exactly as the end of the proof of proposition 5.2. The first and second non convergence results are easy corollaries of proposition 4.20. The third non convergence result is proved in the following way: take $(x,y) \in E^2$ in the bottom of the deepest cycle of $E \times E \setminus \{(z,z)\,;\, z \in E\}$. By definition, the depth of this cycle is the third critical depth H_3, therefore for any $\gamma < H_3$, any $z \in \arg\min \tilde{U}$,

$$\lim_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(\tau^2(\{(z,z)\}) \leq e^{\gamma\beta} \mid (X^1,X^2)_0 = (x,y)) > 0.$$

But

$$\begin{aligned}
\min\{P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = z \mid X_0 = x), P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = z \mid X_0 = y)\} \\
\leq \sqrt{P_\beta((X^1,X^2)_{\lfloor e^{\gamma\beta} \rfloor} = (z,z) \mid (X^1,X^2)_0 = (x,y))} \\
\leq \sqrt{P_\beta(\tau(\{(z,z)\}) \leq e^{\gamma\beta} \mid (X^1,X^2)_0 = (x,y))}.
\end{aligned}$$

This proves that either

$$\limsup_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = z \mid X_0 = x) > 0$$

or

$$\limsup_{\beta \to +\infty} -\frac{1}{\beta} \log P_\beta(X_{\lfloor e^{\gamma\beta} \rfloor} = z \mid X_0 = y) > 0$$

□

Corollary 5.2 (choice of β as a function of N). *For any $\eta > 0$ and any $\gamma > H(E \setminus \arg\min \tilde{U}) = H_1$, we have*

$$\liminf_{N\to+\infty} -\frac{1}{\log N} \log P_{(\log N)/\gamma}(\tilde{U}(X_N) \geq \eta \mid X_0 = x) \geq \frac{\eta}{\gamma}.$$

(The probability of failure of the algorithm with N steps has an upper bound of order $\left(\frac{1}{N}\right)^{\eta/H_1}$.) On the contrary for any $\gamma < H_1$, there is $x \in E$ such that

$$\liminf_{N\to+\infty} -\frac{1}{\log N} \log P_{(\log N)/\gamma}(\tilde{U}(X_N) < \eta \mid X_0 = x) > 0,$$

(the probability of failure consequently tends to one.)

Remarks:

- The inverse temperature parameter β has to be chosen as a function of the number of iterations N.
- To get an approximate solution y such that $\tilde{U}(y) < \eta$ with probability $1 - \epsilon$, the number of iterations needed is of order $\epsilon^{-H_1/\eta}$.
- To get an exact solution with probability $1 - \epsilon$, it is necessary to set in the previous estimate the value of the constant η to $\eta = \min\{\tilde{U}(z) \mid z \in E, \tilde{U}(z) > 0\}$, which may be very close to zero, in which case the number of iterations needed is very large. Therefore, in some situations, the Metropolis algorithm is very slow and speed-up methods are required.
- Another weakness of the Metropolis algorithm is that it is as a rule impossible to compute explicitly the value of H_1, whereas this value is needed to set the temperature parameter in an efficient way.

6. Geometric inequalities for eigenvalues of Markov chains

6.1. Reversible Markov chains.

6.1.1. *Spectral gap estimates.*

Theorem 6.1. *Let E be a finite set and $(E^{\mathbb{N}}, (X_n)_{n\in\mathbb{N}}, \mathcal{B}, P)$ be the canonical realization of a Markov chain with irreducible and reversible transition matrix p and invariant probability distribution π. Let us define the operator $p : L^2(\pi) \to L^2(\pi)$ by $pf(x) = \sum_{y\in E} p(x,y)f(y)$. This operator is self-adjoint, therefore it can be put in diagonal form and its eigenvalues $\lambda_0 \geq \cdots \geq \lambda_{m-1}$ (where $m = |E|$), counted with their multiplicities, satisfy:*

$$1 = \lambda_0 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{m-1} \geq -1.$$

For any probability distribution $\mu \in \mathcal{M}_1^+(E)$, any integer $n \in \mathbb{N}$,

$$\|\mu p^n - \pi\|_{2,\pi} \leq (\max(\lambda_1, -\lambda_{m-1}))^n \|\mu - \pi\|_{2,\pi}.$$

Moreover, for any subset D of E, any $n \in \mathbb{N}$,

$$|P(X_n \in D \mid X_0 = x) - \pi(D)| \\ \leq \left(\frac{1-\pi(x)}{\pi(x)}\right)^{1/2} \min\left(\pi(D)^{1/2}, \frac{1}{2}\right) (\max(\lambda_1, -\lambda_{m-1}))^n.$$

Proof: For any functions $f, g \in L^2(\pi)$,

$$\begin{aligned} (f, pg) &= \sum_{x\in E, y\in E} f(x)p(x,y)g(y)\pi(x) \\ (pf, g) &= \sum_{x\in E, y\in E} p(x,y)f(y)g(x)\pi(x). \end{aligned}$$

As $\pi(x)p(x,y) = \pi(y)p(y,x)$, we have that $(f,pg) = (pf,g)$. The strict inequality between $\beta_0 > \beta_1$ is part of the Perron-Frobenius theorem which we will not prove here. We have $\lambda_{m-1} = -1$ when p is 2-periodic.

The matrix p being irreducible, its invariant measure π is everywhere strictly positive. Therefore we can define a representation

$$i : \mathcal{M}_1^+(E) \longrightarrow L^2(\pi)$$

by $i(\mu) = \dfrac{\mathrm{d}\mu}{\mathrm{d}\pi}$ and put on $\mathcal{M}_1^+(E)$ the corresponding Euclidean norm $||\mu||_{2,\pi} = ||i(\mu)||_{L^2(\pi)}$. The adjoint operator $\tilde{p} = i^{-1} \circ p \circ i : \mathcal{M}_1^+(E) \longrightarrow \mathcal{M}_1^+(E)$ is nothing but the right action of p: $\tilde{p}(\mu)(y) = (\mu p)(y) = \sum_x \mu(x)p(x,y)$. Note that it is self adjoint for $||\ ||_{2,\pi}$, with the same spectrum as p.

Let $\rho = \max_{k\geq 1} |\lambda_k| = \max(\lambda_1, -\lambda_{m-1})$. We have $||\mu p^n - \pi||_{2,\pi} = ||(\mu - \pi)p^n||_{2,\pi}$. Moreover

$$(\mu - \pi, \pi)_{2,\pi} = \int \left(\frac{d\mu}{d\pi} - 1\right) d\pi = 0,$$

therefore $\mu - \pi$ is in the space generated by the eigenvectors of $\tilde{p}$ corresponding to the eigenvalues $\lambda_1, \ldots, \lambda_{m-1}$. Let $\nu_1, \ldots, \nu_{m-1}$ be some choice of these eigenvectors

$$\begin{aligned} (\mu - \pi) &= \sum_{k=1}^{m-1} \alpha_k \nu_k, \\ (\mu - \pi)p^n &= \sum_{k=1}^{m-1} \alpha_k \lambda_k^n \nu_k, \\ ||(\mu - \pi)p^n||_{2,\pi}^2 &= \sum_{k=1}^{m-1} |\alpha_k|^2 |\lambda_k|^{2n} ||\nu_k||_{2,\pi}^2 \\ &\leq \rho^{2n} \sum_{k=1}^{m-1} |\alpha_k|^2 ||\nu_k||_{2,\pi}^2 \\ &= \rho^{2n} ||\mu - \pi||_{2,\pi}^2. \end{aligned}$$

Moreover

$$|P(X_n \in D \mid X_0 = x) - \pi(D)| = \left| \int_D (f_n(y) - 1)\, d\pi(y) \right|$$

where $f_n(y) = \dfrac{P(X_n = y \mid X_0 = x)}{\pi(y)} = \dfrac{\delta_x p^n(y)}{\pi(y)}$. Applying the Cauchy-Schwartz inequality, we obtain that

$$\begin{aligned}
|P(X_n \in D \mid X_0 = x) - \pi(D)| &\leq \left(\int_D (f_n(y) - 1)^2 d\pi(y) \right)^{1/2} \pi(D)^{1/2} \\
&\leq ||\delta_x p^n - \pi||_{2,\pi} \pi(D)^{1/2} \\
&\leq \rho^n ||\delta_x - \pi||_{2,\pi} \pi(D)^{1/2}.
\end{aligned}$$

Moreover

$$||\delta_x - \pi||_{2,\pi} = \left(\frac{(1 - \pi(x))^2}{\pi(x)} + 1 - \pi(x) \right)^{1/2} = \left(\frac{1 - \pi(x)}{\pi(x)} \right)^{1/2}.$$

In the same way

$$\begin{aligned}
|P(X_n \in D \mid X_0 = x) - \pi(D)| &\leq \int_{y, f_n(y) > 1} (f_n(y) - 1) d\pi(y) \\
&\leq \frac{1}{2} \int_E |f_n(y) - 1| d\pi(y) \\
&\leq \frac{1}{2} ||\delta_x p^n - \pi||_{2,\pi} \\
&\leq \frac{1}{2} \rho^n ||\delta_x - \pi||_{2,\pi}. \qquad \square
\end{aligned}$$

6.1.2. *Poincaré inequalities.* Let us call a "routing function" any function $\gamma : E^2 \longrightarrow \bigcup_{n=2}^{+\infty} E^n$ such that $\gamma(x, y) = (z_0 = x, z_1, \ldots, z_{r(x,y)} = y)$ is a path (of arbitrary length $r(x, y)$) going from x to y, with the supplementary condition that $r(x, y)$ be odd when $x = y$. Let Γ be the set of all routing functions.

For any Markov matrix p, irreducible and reversible with respect to its invariant probability distribution π, we define the length of $(z_0, \ldots, z_r)$ with respect to p by

$$|(z_0, \ldots, z_r)|_p = \sum_{i=1}^{r} \left(\pi(z_{i-1}) p(z_{i-1}, z_i) \right)^{-1},$$

with the convention that $0^{-1} = +\infty$.

Let us introduce the constants

$$\kappa = \min_{\gamma \in \Gamma} \max_{(z,t) \in E^2 \setminus \Delta} \sum_{\substack{(x,y) \in E^2 \setminus \Delta, \\ (z,t) \in \gamma(x,y)}} |\gamma(x, y)|_p \pi(x) \pi(y)$$

$$\iota = \min_{\gamma \in \Gamma} \max_{(z,t) \in E^2} \sum_{\substack{x \in E, \\ (z,t) \in \gamma(x,x)}} |\gamma(x, x)|_p \pi(x).$$

Theorem 6.2. *With the previous notations, the spectrum $\lambda_0 = 1 > \lambda_1 \geq \cdots \geq \lambda_{m-1} \geq -1$ of p satisfies*

$$\lambda_1 \leq 1 - \frac{1}{\kappa}$$
$$-\lambda_{m-1} \leq 1 - \frac{2}{\iota}.$$

Proof:
Let us write λ_1 as

$$\lambda_1 = \sup_{\substack{\varphi \in L^2(\pi), \\ E_\pi(\varphi) = 0}} \frac{(\varphi, p\varphi)_\pi}{(\varphi, \varphi)_\pi},$$

this gives

$$1 - \lambda_1 = \inf_{\varphi, E_\pi(\varphi)=0} \frac{\mathcal{E}(\varphi, \varphi)}{(\varphi, \varphi)_\pi},$$

where

$$\begin{aligned}
\mathcal{E}(\varphi, \varphi) &= (\varphi, \varphi - p\varphi)_\pi \\
&= \sum_{x,y \in E} \varphi(x)(\varphi(x) - p(x,y)\varphi(y))\pi(x) \\
&= \sum_{x,y \in E} \varphi(x)(\varphi(x)p(x,y) - p(x,y)\varphi(y))\pi(x).
\end{aligned}$$

Let us put $\pi(x)p(x,y) = Q(x,y)$. We have

$$\begin{aligned}
\mathcal{E}(\varphi, \varphi) &= \sum_{x,y \in E} \varphi(x)(\varphi(x) - \varphi(y))Q(x,y) \\
&= \frac{1}{2} \sum_{x,y \in E} (\varphi(x) - \varphi(y))^2 Q(x,y)
\end{aligned}$$

(*Remark:* The quadratic form $\mathcal{E}$ is called the Dirichlet form of p.)

When $E_\pi(\varphi) = 0$, we have for any routing function $\gamma \in \Gamma$

$$\begin{aligned}
(\varphi, \varphi)_\pi &= \frac{1}{2} \sum_{x,y \in E^2 \setminus \Delta} (\varphi(x) - \varphi(y))^2 \pi(x)\pi(y) \\
&= \frac{1}{2} \sum_{x,y \in E^2 \setminus \Delta} \left(\sum_{(z,t) \in \gamma(x,y)} \varphi(t) - \varphi(z) \right)^2 \pi(x)\pi(y) \\
&\leq \frac{1}{2} \sum_{(x,y) \in E^2 \setminus \Delta} \pi(x)\pi(y) \left(\sum_{(z,t) \in \gamma(x,y)} \frac{1}{Q(z,t)} \right) \\
&\quad \times \sum_{(z,t) \in \gamma(x,y), z \neq t} Q(z,t) \left(\varphi(t) - \varphi(z)\right)^2 \\
&\leq \frac{1}{2} \sum_{(z,t) \in E^2 \setminus \Delta} Q(z,t) \left(\varphi(t) - \varphi(z)\right)^2 \\
&\quad \times \sum_{\substack{(x,y) \in E^2 \setminus \Delta, \\ (z,t) \in \gamma(x,y)}} \pi(x)\pi(y) |\gamma(x,y)|_p \\
&\leq \mathcal{E}(\varphi, \varphi) \max_{(z,t) \in E^2 \setminus \Delta} \sum_{\substack{(x,y) \in E^2 \setminus \Delta, \\ (z,t) \in \gamma(x,y)}} \pi(x)\pi(y) |\gamma(x,y)|_p .
\end{aligned}$$

This being true for any choice of $\gamma \in \Gamma$, we have

$$(\varphi, \varphi)_\pi \leq \kappa \mathcal{E}(\varphi, \varphi),$$

whence $1 - \lambda_1 \geq \dfrac{1}{\kappa}$.

Let us come now to the second inequality. We have

$$1 + \lambda_{m-1} = \inf_{\varphi \in L^2(\pi)} \frac{(\varphi, p\varphi + \varphi)_\pi}{(\varphi, \varphi)_\pi}.$$

Moreover

$$\begin{aligned}
(\varphi, p\varphi + \varphi)_\pi &= \sum_{(x,y) \in E} \varphi(x)(p(x,y)\varphi(y) + \varphi(x))\pi(x) \\
&= \sum_{(x,y) \in E} \varphi(x)(\varphi(x) + \varphi(y)) Q(x,y) \\
&= \frac{1}{2} \sum_{(x,y) \in E} (\varphi(x) + \varphi(y))^2 Q(x,y).
\end{aligned}$$

$$
\begin{aligned}
(\varphi,\varphi)_\pi &= \sum_{x\in E,\gamma(x,x)=(z_0,\dots,z_r)} \pi(x)\frac{1}{4}\left(\sum_{i=0}^{r-1}(-1)^i\left(\varphi(z_i)+\varphi(z_{i+1})\right)\right)^2\\
&\le \frac{1}{4}\sum_{x\in E,\gamma(x,x)=(z_0,\dots,z_r)} \pi(x)\left(\sum_{i=0}^{r-1}\left(\varphi(z_i)+\varphi(z_{i+1})\right)^2 Q(z_i,z_{i+1})\right)\\
&\quad\times\left(\sum_{i=0}^{r-1}\frac{1}{Q(z_i,z_{i+1})}\right)\\
&\le \frac{1}{4}\sum_{x\in E}\pi(x)|\gamma(x,x)|_p\sum_{(z,t)\in\gamma(x,x)}\left(\varphi(z)+\varphi(t)\right)^2 Q(z,t)\\
&\le (\varphi,p\varphi+\varphi)_\pi\frac{1}{2}\max_{(z,t)\in E^2}\sum_{x\in E,(z,t)\in\gamma(x,x)}\pi(x)|\gamma(x,x)|_p.
\end{aligned}
$$

Thus

$$
(\varphi,\varphi)_\pi \le \frac{\iota}{2}(\varphi,p\varphi+\varphi)_\pi,
$$

and $1+\lambda_{m-1}\ge\dfrac{2}{\iota}$. $\square$.

6.2. **Application to the generalised Metropolis algorithm.**

6.2.1. *Reversible case.*

Theorem 6.3. *Let us consider a family $(E^{\mathbb{N}},(X_n)_{n\in\mathbb{N}},\mathcal{B},P_\beta)_{\beta\in\mathbb{R}_+}$ of homogeneous Markov chains with rate function V and transition matrix p_β. Let us assume that V is irreducible and that p_β is reversible with respect to its invariant probability distribution μ_β. Let us assume moreover that for some strictly positive constants β_0,a,b,c,d, for any $\beta\ge\beta_0$, any x, $y\in E$,*

$$
\begin{aligned}
ae^{-\beta U(x)} &\le \mu_\beta(x)\le be^{-\beta U(x)}\\
ce^{-\beta V(x,y)} &\le p_\beta(x,y).
\end{aligned}
$$

Let us assume that $V(x,x)=0$, for any $x\in E$. Let $\gamma\in\Gamma$ be a routing function such that for any $(x,y)\in E^2\setminus\Delta$

$$
H(\gamma(x,y))=H(x,y),
$$

and let

$$
\begin{aligned}
L(\gamma) &= \max_{(x,y)\in E^2\setminus\Delta}|\gamma(x,y)|, \qquad \textit{(nb. of edges)}\\
D(\gamma) &= \max_{(z,t)\in E^2\setminus\Delta}\left|\{(x,y)\in E^2\setminus\Delta \mid (z,t)\in\gamma(x,y)\}\right|.
\end{aligned}
$$

Then the eigenvalues of p_β, $\lambda_0=1>\lambda_1\ge\cdots\ge\lambda_{m-1}\ge-1$, satisfy, for any $\beta\ge\beta_0$,

$$
\begin{aligned}
\lambda_1 &\le 1-\frac{a\ c}{b^2\ L(\gamma)D(\gamma)}e^{-\beta H_2}\\
-\lambda_{m-1} &\le 1-2\ c.
\end{aligned}
$$

Consequently for any $\beta \geq \beta_0$, any $D \subset E$, any $x \in E$,

$$|P_\beta(X_n \in D \mid X_0 = x) - \mu_\beta(D)| \leq a^{-1/2} e^{\beta U(x)/2} (\mu_\beta(D))^{1/2} \times \left(1 - \min\left(2c, \frac{ac \exp(-\beta H_2)}{b^2 L(\gamma) D(\gamma)}\right)\right)^n.$$

Proof: The upper bound for λ_1 is a consequence of the expression for κ. To get the lower bound for λ_{m-1}, consider the routing function $\gamma(x,x) = (x,x)$.

6.2.2. *The non-reversible case.* Let us consider a family $(E^{\mathbb{N}}, (X_n)_{n\in\mathbb{N}}, \mathcal{B}, P_\beta)_{\beta\in\mathbb{R}_+}$ of Markov chains with rare transitions with irreducible rate function V and transition matrix p_β.

Given some real number $\lambda \in)0,1($, let us consider the Markov matrices

$$\begin{aligned} q_\beta(x,y) &= \lambda\delta(x,y) + (1-\lambda)p_\beta(x,y) \\ \overline{q}_\beta(x,y) &= \sum_{z\in E} q_\beta(x,z) q_\beta(y,z) \frac{\mu_\beta(y)}{\mu_\beta(z)}. \end{aligned}$$

The matrices q_β and $\overline{q}_\beta$ are irreducible and μ_β is their common invariant distribution. Moreover $\overline{q}_\beta$ is reversible, it is a non negative self-adjoint operator in $L^2_{\mu_\beta}$, since it is the product of q_β and of its adjoint. Let ρ_β be the spectral gap of $\overline{q}_\beta$,

$$\begin{aligned} \rho_\beta &= 1 - \max\{|\xi| \mid \xi \in Sp(\overline{q}_\beta), \xi \neq 1\} \\ &= 1 - \max\{\xi \mid \xi \in Sp(\overline{q}_\beta), \xi \neq 1\}. \end{aligned}$$

Theorem 6.4. *We have*

$$\limsup_{\beta\to+\infty} -\frac{1}{\beta}\log \rho_\beta \leq H_2 = H(E \setminus \{z\})$$

for any $z \in \arg\min \tilde{U}$. Moreover for any $D \subset E$, any $x \in E$, any $n \in \mathbb{N}$,

$$P_\beta(\tau(D) \leq n \mid X_0 = x) \geq \mu_\beta(D) - \left(\frac{1-\mu_\beta(x)}{\mu_\beta(x)}\right)^{1/2} \mu_\beta(D)^{1/2}(1-\rho_\beta)^{n/2}.$$

Proof. Let $(E^{\mathbb{N}}, (X_n)_{n\in\mathbb{N}}, \mathcal{B}, Q_\beta)_{\beta\in\mathbb{R}_+}$ be the canonical realization of a family of Markov chains with transition matrix q_β (and some irrelevant arbitrary initial distribution). We have

$$\begin{aligned} Q_\beta(X_n \in D \mid X_0 = x) &= \sum_{k=0}^n \binom{n}{k} \lambda^{n-k}(1-\lambda)^k P_\beta(X_k \in D \mid X_0 = x) \\ &\leq \max_{k=0,\dots,n} P_\beta(X_k \in D \mid X_0 = x) \\ &\leq P_\beta(\tau(D) \leq n \mid X_0 = x). \end{aligned}$$

Moreover

$$|Q_\beta(X_n \in D \mid X_0 = x) - \mu_\beta(D)| \leq \mu_\beta(D)^{1/2} \|\delta_x q_\beta^n - \mu_\beta\|_{2,\mu_\beta}.$$

For any probability distribution ν

$$\begin{aligned}\left\|\nu\frac{q_\beta}{\mu_\beta}-1\right\|^2_{2,\mu_\beta} &= \left(\frac{(\nu-\mu_\beta)\overline{q}_\beta}{\mu_\beta},\frac{(\nu-\mu_\beta)}{\mu_\beta}\right)_{\mu_\beta}\\ &\leq \left\|\frac{\nu-\mu_\beta}{\mu_\beta}\right\|^2_{2,\mu_\beta}(1-\rho_\beta),\end{aligned}$$

therefore

$$P_\beta(\tau(D)\leq n \mid X_0=x)\geq \mu_\beta(D)-\left(\frac{1-\mu_\beta(x)}{\mu_\beta(x)}\right)^{1/2}\mu_\beta(D)^{1/2}(1-\rho_\beta)^{n/2}.$$

Let $\overline{V}$ be the rate function corresponding to $(\overline{q}_\beta)_{\beta\in\mathbb{R}_+}$. It is easy to check that for any x, $y\in E$,

$$\overline{V}(x,y)\leq V(x,y),$$

thus $H_2(\overline{V}) = \max_{x\in E} H_{\overline{V}}(x,y)-\tilde{U}(x)\leq \max_{x\in E} H_V(x,y)-\tilde{U}(x) = H_2$, where y is an arbitrary point in $\arg\min\tilde{U}$. Thus, considering a routing function satisfying $H_{\overline{V}}(\gamma(x,y)) = H_{\overline{V}}(x,y)$, we see that

$$\limsup_{\beta\to+\infty} -\frac{1}{\beta}\log\rho_\beta\leq H_2.$$

□

Remark 6.1. We have in fact more precisely that

$$\lim_{\beta\to+\infty} -\frac{\log\rho_\beta}{\beta} = H_2.$$

This can be seen from [29] where the reader will find an alternative approach to the non reversible case, or from the fact that a lower limit would contradict the optimal rate of convergence of q_β^n towards its invariant distribution proved in theorem 5.1. Another interesting reference for the "multiplicative reversiblization" method is [15].

Remark 6.2. Theorem 6.4 can be used as an alternative tool to prove convergence results, such as proposition 5.1 and 5.2. Indeed these propositions rely on upper bounds for the tail distribution of the exit times from domains. As an illustration, let us see how we can prove the first bound in proposition 4.19 from theorem 6.4.

Let $D\subsetneq E$ be a domain and $\tau(\overline{D})$ its exit time. The estimate we are looking for does not depend on the behaviour of the Markov chain outside from D, so we can modify the state space, creating a unique outer state Δ standing for $\overline{D}$. Let $E' = D\cup\{\Delta\}$ be the new state space. Let the modified transition matrix p'_β be

$$p'_\beta = \begin{cases} p_\beta(x,y) & \text{if } (x,y)\in D^2,\\ \displaystyle\sum_{z\notin D} p_\beta(x,z) & \text{if } x\in D, y=\Delta,\\ \dfrac{1}{|D|}\exp(-\beta H_\Delta) & \text{if } x=\Delta, y\in D.\end{cases}$$

By taking $H_\Delta > H(D)$, we get a modified virtual energy $\tilde{U}'$ such that $\tilde{U}'(\Delta) = 0$ and

$$H_\Delta - H(D) \leq \tilde{U}'(x) \leq H_\Delta, \quad x \in D.$$

Indeed

$$\begin{aligned} H_\Delta - \tilde{U}'(x) &= H'(\Delta, x) - \tilde{U}'(x) \\ &= H'(x, \Delta) - \tilde{U}'(x) \\ &\leq H(D), \end{aligned}$$

and

$$\tilde{U}'(x) \leq H'(x, \Delta) = H_\Delta.$$

Moreover the critical depths of the modified landscape are $H_1' = H_2' = H_3' = H(D)$. We see immediately that we have

$$\begin{aligned} \lim_{\beta \to +\infty} \mu_\beta'(\Delta) &= 1, \\ \lim_{\beta \to +\infty} \mu_\beta'(x) \exp\big((H_\Delta + \epsilon)\beta\big) &= +\infty, & x \in D, \epsilon > 0, \\ \lim_{\beta \to +\infty} \rho_\beta' \exp\big((H(D) + \epsilon)\beta\big) &= +\infty, & \epsilon > 0. \end{aligned}$$

Plugging this altogether into theorem 6.4 applied to $\{\Delta\}$ gives that for any $\epsilon > 0$

$$\lim_{\beta \to +\infty} \max_{x \in D} P_\beta(\tau(\overline{D}) > \exp\big(\beta(H(D) + \epsilon)\big) \mid X_0 = x) = 0,$$

(taking $H_\Delta = H(D) + \epsilon/2$). This can be immediately strengthened to

$$\lim_{\beta \to +\infty} -\frac{1}{\beta} \log \max_{x \in D} P_\beta(\tau(\overline{D}) > \exp\big(\beta(H(D) + \epsilon)\big) \mid X_0 = x) = +\infty$$

using the Markov property as in the beginning of the proof of proposition 4.19.

We have sketched the link between theorem 6.4 and proposition 4.19 to show that semigroup methods can be extended to the same generality as the Freidlin and Wentzell approach, however the reader should keep in mind that their main interest is to provide more explicit bounds and constants when stronger assumptions are made on the transition matrix p_β than what is assumed in these notes.

7. Simulated annealing algorithms

7.1. **Description.** Let us consider a finite state space E and a family $(p_\beta)_{\beta \in \mathbb{R}_+}$ of Markov matrices with rare transitions and irreducible rate function V.

For any increasing inverse temperature sequence $(\beta_n)_{n \in \mathbb{N}^*}$ (of real positive numbers), we can construct a non-homogeneous Markov chain $(E^{\mathbb{N}}, (X_n)_{n \in \mathbb{N}}, \mathcal{B}, P_{(\beta_n)_{n \in \mathbb{N}^*}})$ with transitions

$$P_{(\beta_.)}(X_n = y \mid X_{n-1} = x) = p_{\beta_n}(x, y), \quad x, y \in E.$$

This chain describes the generalised simulated annealing algorithm. It is used to minimise the virtual energy $\tilde{U}$ corresponding to (E, V).

7.2. **Convergence results.** These results make use of two important constants of (E, V). We have already used the first, it is the first critical depth

$$H_1 = \max\{H(C) \mid C \in \mathcal{C}(V), \tilde{U}(C) > 0\}$$

The second will be called the difficulty of (E, V), and is defined to be

$$D = \max\{\frac{H(C)}{\tilde{U}(C)} \mid C \in \mathcal{C}(V), \tilde{U}(C) > 0\}.$$

Theorem 7.1. *With the preceding hypotheses and notations, for any bounds $\overline{H}$ and $\underline{D}$ such that $\overline{H} > H_1$, and $0 < \underline{D} < D$, for any η such that $0 < \eta < \overline{H}/\underline{D}$, any integer $r > 0$, the triangular sequence of inverse temperatures*

$$\beta_n^N = \frac{1}{\overline{H}} \log\left(\frac{N}{r}\right) \left(\frac{\overline{H}}{\underline{D}\eta}\right)^{\frac{1}{r}\left\lfloor\frac{(n-1)r}{N}\right\rfloor}, \qquad 1 \leq n \leq N,$$

satisfies for any $x \in E$

$$\liminf_{N \to +\infty} -\frac{1}{\log N} \log P_{(\beta^N)}(\tilde{U}(X_N) \geq \eta \mid X_0 = x) \geq \frac{1}{D}\left(\frac{\underline{D}\eta}{\overline{H}}\right)^{1/r}.$$

Remarks:

- For r large, the order of magnitude of the upper bound for $P_{(\beta^N)}(\tilde{U}(X_N) \geq \eta \mid X_0 = x)$ is close to $N^{-1/D}$. More precisely, for any $\epsilon > 0$,

  any $r \geq \dfrac{\log\left(\overline{H}/(\underline{D}\eta)\right)}{\log(1+\epsilon)}$, any $x \in E$,

  $$\liminf_{N \to +\infty} -\frac{1}{\log N} \log P_{(\beta^N)}(\tilde{U}(X_N) \geq \eta \mid X_0 = x) \geq \frac{1}{(1+\epsilon)D}.$$

  The number of iterations needed to bring down the probability of failure to a given order of magnitude is therefore independent of the precision $\eta > 0$. The upper bound for the probability of failure is at best of order $N^{-1/D}$. One can show that this is the best one can achieve using non-decreasing inverse temperature sequences (see [6, 31, 33]).
- The choice of parameters is robust: it is not necessary to know the exact value of D to choose the values of the parameters. We get a probability of failure of order $\left(\dfrac{1}{N}\right)^{\frac{1}{D}\left(\frac{\eta\underline{D}}{\overline{H}}\right)^{1/r}}$ uniformly for any rate function V such that $\overline{H} > H_1(E, V)$ and $\underline{D} < D(E, V)$. This is not the case with the Metropolis algorithm in which the choice of β requires a precise knowledge of $H_1(V)$ (namely the proved exponent of convergence of simulated annealing $\dfrac{1}{D}\left(\dfrac{\eta\underline{D}}{\overline{H}}\right)^{1/r}$ is uniformly close to the optimal exponent $\dfrac{1}{D}$ when $\left(\dfrac{\eta\underline{D}}{\overline{H}}\right)^{1/r}$ is close to one, which can be obtained by taking a large value for

r, even when the gaps $\overline{H} - H_1$ and $D - \underline{D}$ are large, whereas the exponent of convergence of the Metropolis algorithm $\frac{\eta}{\gamma}$ is close to optimal only when $\frac{\gamma - H_1}{H_1}$ is small.)

- Triangular sequences of inverse temperatures are absolutely needed: one can show that for any infinite (unique) non-decreasing temperature sequence (β_n)

$$\min_{x\in E} \limsup_{N\to+\infty} -\frac{1}{\log N} \log P_{(\beta_\cdot)}(\tilde{U}(X_N) \geq \eta \mid X_0 = x) \leq \frac{\eta}{H_1(E,V)}.$$

(See [5].) When a non triangular sequence is used, the convergence speed is in first approximation of the same order as for the Metropolis algorithm. This means that triangular sequences are crucial to get a significative speed-up with respect to the Metropolis algorithm.

Proof:

Let us put, to simplify notations,

$$\gamma_k^N = \beta_n^N = \frac{1}{\overline{H}} \log \frac{N}{r} \left(\frac{\overline{H}}{\underline{D}\eta}\right)^{k/r} \quad , k\frac{N}{r} < n \leq (k+1)\frac{N}{r}.$$

Let us also put $P_{(\beta_\cdot^N)} = P_N$ and assume that $N/r \in \mathbb{N}$ (the modifications needed to handle the general case are left to the reader).

Let $\xi > 0$ be fixed and let

$$\eta_k = \frac{\overline{H}}{(1+\xi)D} \left(\frac{\overline{H}}{\underline{D}\eta}\right)^{-\frac{k+1}{r}}, \quad k = 0, \dots, r-1.$$

$$\begin{cases} \lambda_0 &= +\infty \\ \lambda_k &= \frac{(1+\frac{1}{D})\overline{H}}{(1+\xi)} \left(\frac{\overline{H}}{\underline{D}\eta}\right)^{-k/r}, \quad k = 1, \dots, r-1. \end{cases}$$

Let us consider the events

$$\begin{aligned} B_k &= \{\tilde{U}(X_n) + V(X_n, X_{n+1}) \leq \lambda_k, k\frac{N}{r} \leq n < (k+1)\frac{N}{r}\}, \\ A_k &= B_k \cap \{\tilde{U}(X_{(k+1)N/r}) < \eta_k\}. \end{aligned}$$

We have

$$\begin{aligned} \exp \overline{H}\gamma_0^N &= \frac{N}{r}, \\ \exp\left((1+\xi)\left(1+\frac{1}{D}\right)^{-1} \lambda_k \gamma_k^N\right) &= \frac{N}{r}, \quad k > 0, \end{aligned}$$

and

$$\eta_{r-1} = \frac{1}{\xi+1}\frac{\underline{D}}{D}\eta \leq \eta.$$

Therefore

$$\begin{aligned}
P_N(\tilde{U}(X_N) \geq \eta \mid X_0 = x) &\leq P_N(\tilde{U}(X_N) \geq \eta_{r-1} \mid X_0 = x) \\
&\leq 1 - P_N(\bigcap_{k=0}^{r-1} A_k \mid X_0 = x) \\
&\leq \sum_{k=0}^{r-1} P_N(\overline{A}_k \cap \bigcap_{l=0}^{k-1} A_l \mid X_0 = x).
\end{aligned}$$

Moreover

$$\begin{aligned}
P_N(\overline{A}_k \cap \bigcap_{l=0}^{k-1} A_l \,|\, X_0 = x) \leq P_N(\overline{B}_k \cap \bigcap_{l=0}^{k-1} A_l \,|\, X_0 = x) \\
+ P_N((\tilde{U}(X_{(k+1)N/r}) \geq \eta_k) \cap \bigcap_{l=0}^{k-1} A_l \cap B_k \,|\, X_0 = x).
\end{aligned}$$

Let us remark first that for any cycle C such that $\tilde{U}(C) > 0$ and $H(C)+\tilde{U}(C) \leq \lambda_k$ we have

$$H(C) \leq \left(1 + \frac{1}{D}\right)^{-1} \lambda_k.$$

For any $z \in E$ such that $\tilde{U}(z) < \eta_{k-1}$, let us consider the smallest cycle $C_z \in \mathcal{C}(V)$ containing z such that $\tilde{U}(C_z)+H(C_z) > \lambda_k$. We have $\tilde{U}(C_z) = 0$. Indeed, if we had $\tilde{U}(C_z) > 0$, we would have also $\tilde{U}(C_z) + H(C_z) \leq \tilde{U}(z)(1 + D) \leq \eta_{k-1}(1 + D) = \lambda_k$, which is a contradiction.

From the preceding remarks, we deduce that

$$H_1(C_z, V_{|C_z \times C_z}) \leq \left(1 + \frac{1}{D}\right)^{-1} \lambda_k.$$

Indeed any cycle $C \subset C_z$ such that $C \neq C_z$ and $\tilde{U}(C) > 0$ satisfies $H(C) \leq \left(1 + \frac{1}{C}\right)^{-1} \lambda_k$, and

$$\begin{aligned}
H_1(C_z, V_{|C_z \times C_z}) &= \max\{H(C) : C \subset C_z, C \in \mathcal{C}(V), \tilde{U}(C) > 0\} \\
&= H(C_z \setminus \arg\min_{y \in C_z} \tilde{U}(y)).
\end{aligned}$$

Let us remark now that

$$\begin{aligned}
&P(X_{(k+1)N/r} = y, B_k \,|\, X_{kN/r} = z) \\
&\leq P(X_{(k+1)N/r} = y, X_n \in C_z, k\frac{N}{r} < n \leq (k+1)\frac{N}{r} \,|\, X_{kN/r} = z).
\end{aligned}$$

Let us consider on C_z the Markov chain $(Y_n)_{n\in\mathbb{N}}$ with transitions $P(Y_n = y \,|\, Y_{n-1} = x) = q(x, y)$ defined by

$$q(x,y) = \begin{cases} p_{\gamma_k^N}(x,y) & \text{if } x \neq y \in C_z, \\ 1 - \displaystyle\sum_{w\in(C_z\setminus\{x\})} q(x,w) & \text{otherwise.} \end{cases}$$

(We obtain this new chain by reflecting $(X_n)_{n\in\mathbb{N}}$ on the boundary of C_z.)

As for any $x, y \in C_z$, $p_{\gamma_k^N}(x, y) \leq q(x, y)$, we have

$$P_N(X_{(k+1)N/r} = y, X_n \in C_z, k\frac{N}{r} < n \leq (k+1)\frac{N}{r} \,|\, X_{kN/r} = z) \leq P(Y_{N/r} = y \,|\, Y_0 = z).$$

Applying to Y the theorem on the convergence speed of the Metropolis algorithm, we see that for any $\epsilon > 0$, there is N_0 such that for any $N \geq N_0$,

$$P_N((X_{(k+1)N/r} = y) \cap \bigcap_{l=0}^{k-1} A_l \cap B_k \,|\, X_0 = x) \leq \exp\left(-\gamma_k^N(\tilde{U}(y) - \epsilon)\right).$$

We have now to find an upper bound for

$$P_N(\overline{B}_k \cap \bigcap_{l=0}^{k-1} A_l \,|\, X_0 = x).$$

For any $z \in E$ such that $\tilde{U}(z) < \eta_{k-1}$, we have

$$\begin{aligned} P_N(\overline{B}_k \,|\, X_{kN/r} = z) &\leq \sum_{n=kN/r+1}^{(k+1)N/r} P_N(\tilde{U}(X_{n-1}) + V(X_{n-1}, X_n) > \lambda_k \,|\, X_{kN/r} = z) \\ &= \sum_{\substack{kN/r<n\leq(k+1)N/r,\\ (u,v)\in E^2,\\ \tilde{U}(u)+V(u,v)>\lambda_k}} P_N(X_{n-1} = u \mid X_{kN/r} = z) p_{\gamma_k^N}(u, v) \\ &\leq \sum_{\substack{kN/r<n\leq(k+1)N/r,\\ (u,v)\in E^2,\\ \tilde{U}(u)+V(u,v)>\lambda_k}} \frac{\mu_{\gamma_k^N}(u)}{\mu_{\gamma_k^N}(z)} p_{\gamma_k^N}(u, v) \\ &\leq \frac{N}{r} \exp\left(-\gamma_k^N(\lambda_k - \tilde{U}(z) - \epsilon)\right), \end{aligned}$$

for any $\epsilon > 0$ and N large enough. Thus for any $\epsilon > 0$, for N large enough,

$$\begin{aligned}
P_N(\overline{B}_k \cap \bigcap_{l=0}^{k-1} A_l \,|\, X_0 = x) &= \sum_{z, \tilde{U}(z) < \eta_{k-1}} P_N(\overline{B}_k \,|\, X_{kN/r} = z) \\
&\quad \times P_N((X_{kN/r} = z) \cap \bigcap_{l=0}^{k-2} A_l \cap B_{k-1} \,|\, X_0 = x) \\
&\leq \sum_{z, \tilde{U}(z) < \eta_{k-1}} \frac{N}{r} \exp\left(-\gamma_k^N(\lambda_k - \tilde{U}(z) - \epsilon)\right) \\
&\quad \times \exp\left(-\gamma_{k-1}^N(\tilde{U}(z) - \epsilon)\right) \\
&\leq \frac{N}{r} \exp\left(-\gamma_k^N(\lambda_k - \eta_{k-1} - \epsilon)\right) \\
&\quad \times \exp\left(-\gamma_{k-1}^N(\eta_{k-1} - 2\epsilon)\right).
\end{aligned}$$

Therefore, for any $\epsilon > 0$, there is N_0 such that, for any $N \geq N_0$,

$$\begin{aligned}
&P(\overline{A}_k \cap \bigcap_{l=0}^{k-1} A_l \,|\, X_0 = x) \\
&\quad \leq \frac{N}{r} \exp\left(-(\lambda_k - \eta_{k-1})\gamma_k^N - \eta_{k-1}\gamma_{k-1}^N + \epsilon(\gamma_k^N + \gamma_{k-1}^N)\right) \\
&\qquad + \exp\left(-(\eta_k - \epsilon)\gamma_k^N\right).
\end{aligned}$$

Coming back to the definitions, we see that

$$\begin{aligned}
\eta_k \gamma_k^N &= \log\left(\frac{N}{r}\right) \frac{1}{(1+\xi)} \frac{1}{D} \left(\frac{D\eta}{\overline{\overline{H}}}\right)^{1/r}, \\
\lambda_k \gamma_k^N + \eta_{k-1}(\gamma_{k-1}^N - \gamma_k^N) &= \left(1 + \frac{1}{D}\right)(1+\xi)^{-1} \log\frac{N}{r} \\
&\quad + \frac{1}{(1+\xi)D}\left(\left(\frac{\eta\underline{D}}{\overline{\overline{H}}}\right)^{1/r} - 1\right) \log\frac{N}{r} \\
&= \log\frac{N}{r}\left((1+\xi)^{-1}\left(1 + \frac{1}{D}\left(\frac{\eta\underline{D}}{\overline{\overline{H}}}\right)^{1/r}\right)\right), \\
\epsilon\gamma_k^N &\leq \frac{\epsilon}{\eta\underline{D}} \log\frac{N}{r}.
\end{aligned}$$

Thus

$$\begin{aligned}
P(\overline{A}_k \cap \bigcap_{l=0}^{k-1} A_l \,|\, X_0 = x) &\leq \left(\frac{N}{r}\right)^{-(1+\xi)^{-1}\left(\frac{1}{D}\left(\frac{\eta\underline{D}}{\overline{\overline{H}}}\right)^{1/r} - \xi\right) + \frac{2\epsilon}{\eta\underline{D}}} \\
&\quad + \left(\frac{N}{r}\right)^{-(1+\xi)^{-1}\frac{1}{D}\left(\frac{D\eta}{\overline{\overline{H}}}\right)^{1/r} + \frac{\epsilon}{\eta\underline{D}}}.
\end{aligned}$$

Letting ξ and ϵ tend to zero, we get eventually that

$$\liminf_{N\to+\infty} -\frac{1}{\log N} \log P_N(\tilde{U}(X_N) \geq \eta \mid X_0 = x) \geq \frac{1}{D}\left(\frac{D\eta}{\overline{\overline{H}}}\right)^{1/r}.$$

For more precise computations under the stronger hypothesis that for some constant $a > 0$

$$a^{-1}e^{-\beta V(x,y)} \leq p_\beta(x,y) \leq ae^{-\beta V(x,y)},$$

we refer to [11].

8. The energy transformation algorithm

8.1. The energy transformation method. The purpose of this algorithm is to minimise a function $U : E \longrightarrow \mathbb{R}$ defined on a finite set E, using to explore the states of E an irreducible Markov matrix $q : E \times E \longrightarrow [0,1]$ with a fixed symmetric support. The method is to use a rate function of the form

$$V(x,y) = (F \circ U(y) - F \circ U(x))^+, \quad q(x,y) > 0,$$

where $F : \mathbb{R} \longrightarrow \mathbb{R} \cup \{-\infty\}$ is a suitable increasing function.

8.2. Convergence result for a single transformation.

Proposition 8.1. *Let $q : E \times E \longrightarrow [0,1]$ be an irreducible Markov matrix with symmetric support. Let $(E^{\mathbb{N}}, (X_n)_{n\in\mathbb{N}}, \mathcal{B}, P_{\beta,\eta})_{\beta\in\mathbb{R}_+,\eta\in\mathbb{R}_+^*}$ be the canonical realization of a family of Markov chains with transitions*

$$p_{\beta,\eta}(x,y) = q(x,y)\exp\left(-\beta\left(F_\eta \circ U(y) - F_\eta \circ U(x)\right)^+\right), \quad x \neq y,$$

where $F_\eta(u) = \log(u+\eta)$, where $\eta + U_{\min} > 0$.

Let us introduce the two rate functions:

$$V(x,y) = \begin{cases} (U(y) - U(x))^+, & p_{\beta,\eta}(x,y) > 0, \\ +\infty & \textit{otherwise} \end{cases}$$

$$W_\eta(x,y) = \begin{cases} (F_\eta \circ U(y) - F_\eta \circ U(x))^+, & p_{\beta,\eta}(x,y) > 0 \\ +\infty & \textit{otherwise.} \end{cases}$$

Then W_η is the rate function describing the rare transitions of the sub-family $(P_{\beta,\eta})_{\beta\in\mathbb{R}_+}$, and for any $\eta > -U_{\min}$, $\rho > 0$, $\epsilon > 0$ and any $x \in E$,

$$\liminf_{N\to+\infty} -\frac{1}{\log N} \log P_{\beta_{N,\eta},\eta}(U(X_N) - U_{\min} \geq \rho(\eta + U_{\min}) \mid X_0 = x) \geq \frac{\log(1+\rho)}{\log(1 + D_{(\eta+U_{\min})})}(1+\epsilon)^{-1},$$

with

$$\beta_{N,\eta} = \frac{\log N}{(1+\epsilon)\log(1 + D_{(\eta+U_{\min})})},$$

$$D_\alpha = \max\{\frac{H_V(C)}{U(C) - U_{\min} + \alpha} \mid C \in \mathcal{C}(V), U(C) > U_{\min}\},$$

where $H_V(C)$ is the depth of C with respect to the rate function V, induced by U.

Remark: If it is known in advance that $a < U_{\min} < b$, it is possible to take $F(u) = \log(u-a)$. This ensures a probability of failure bounded by $\left(\frac{1}{N}\right)^{\frac{\log(1+\rho)}{\log(1+D_{(U_{\min}-a)})}(1+\epsilon)^{-1}}$ when failure means $U(X_N) \geq U_{\min} + \delta$ with $\delta = \rho(b-a)$. The interesting thing is that the exponent $\alpha = (1+\epsilon)^{-1}\dfrac{\log(1+\rho)}{\log(1+D_{(U_{\min}-a)})} = (1+\epsilon)^{-1}\dfrac{\log\left(1+\frac{\delta}{b-a}\right)}{\log\left(1+D_{(U_{\min}-a)}\right)}$ describing the convergence speed depends on the precision $(b-a)$ with which $U_{\min}$ is known in advance, and that, for a fixed value of δ, α tends to $+\infty$ when the precision $b-a$ tends to 0.

Proof:

As F_η is increasing, it is easy to see that $\mathcal{C}(V) = \mathcal{C}(W_\eta)$. In the case when $H_1(V) = H_1(W_\eta) = 0$, there are no local minimum, $D_{(\eta+U_{\min})} = 0$, and the proposition is true with the convention that $1/0 = +\infty$, since the convergence of the probability of error to zero is in this case easily seen to be exponential and not polynomial in N. Therefore we will assume in this proof that $H_1(W_\eta) > 0$. For any cycle $C \in \mathcal{C}(V) = \mathcal{C}(W_\eta)$ such that $U(C) > U_{\min}$,

$$\begin{aligned} H_{W_\eta}(C) &= F_\eta(U(C) + H_V(C)) - F_\eta(U(C)) \\ &= \log\left(1 + \frac{H_V(C)}{U(C)+\eta}\right) \\ &\leq \log(1 + D_{(\eta+U_{\min})}), \end{aligned}$$

Moreover

$$\begin{aligned} F_\eta(\rho(\eta+U_{\min}) + U_{\min}) - F_\eta(U_{\min}) &= \log(1+\rho), \\ \text{and} \quad \exp\left(\beta_{N,\eta} H_1(W_\eta)(1+\epsilon)\right) &\leq N, \end{aligned}$$

therefore

$$\liminf_{N\to+\infty} -\frac{1}{\beta_{N,\eta}} \log P_{\beta_{N,\eta},\eta}(U(X_N) - U_{\min} \geq \rho(\eta+U_{\min}) \,|\, X_0 = x) \geq \log(1+\rho). \qquad \square$$

In the following paragraph, we will use the energy transformation method repeatedly to improve a rough initial lower bound for $U_{\min}$.

8.3. The Iterated Energy Transformation algorithm.

Theorem 8.1. *Let $\gamma < U_{\min}$ be a lower bound for $U_{\min}$ which is assumed to be known beforehand. Let $\eta_0 \geq 0$ be a non negative parameter, and let us consider the (non-Markovian) stochastic process $(E^{\mathbb{N}}, (X_n)_{n=1,\dots,N}, \mathcal{B}, P_N)$ with transitions*

$$P_N(X_n = y \,|\, (X_0,\dots,X_{n-1}) = (x_0,\dots,x_{n-1})) = p_{\beta_{N,\tau_k}}(x_{n-1}, y), \quad k\frac{N}{r} < n \leq (k+1)\frac{N}{r},$$

where

$$\begin{aligned}
\beta_N &= \frac{\log(N/r)}{(1+\epsilon)\log(1+D_{\eta_0})} \\
\tau_k &= \tau_{k-1} - \frac{1}{(1+\rho)}(\tau_{k-1} + U(x_{kN/r})) + \eta_0, \\
\tau_0 &= \eta_0 - \gamma,
\end{aligned}$$

and where r is the number of steps of the algorithm. Then for any $x \in E$

$$\liminf_{N\to+\infty} -\frac{1}{\log N}$$

$$\log P_N\left(U(X_N) - U_{\min} \geq \rho\left(\frac{\rho}{1+\rho}\right)^{r-1}(U_{\min} - \gamma + \eta_0) + \eta_0\rho(1+\rho) \,|\, X_0 = x\right)$$

$$\geq \frac{\log(1+\rho)}{(1+\epsilon)\log(1+D_{\eta_0})}.$$

Remarks:

- The probability of failure can be reduced to order N^{ξ} with ξ arbitrarily large by increasing ρ and r and decreasing η_0. A more precise study of the algorithm (see [7]) would allow us to choose r and ρ as functions of N and to get a convergence speed better than polynomial.
- The I.E.T. algorithm is well suited when D is large and $|E|$ is moderate. In order to fight against the number of states in E, it is possible to use an energy transform of the form $\alpha u + \beta \log(u + \eta)$.
- The energy transformation method can also be used for the simulated annealing algorithm: any concave increasing energy transformation will decrease the difficulty (see [2]).

Proof: Let us introduce the events

$$A_k = \{U(X_{(k+1)N/r}) - U_{\min} < (\tau_k + U_{\min})\rho\}.$$

We have

$$\begin{aligned}
P_N(U(X_N) \geq U_{\min} + \rho(U_{\min} + \tau_{r-1}) \,|\, X_0 = x) &= P_N(\overline{A}_{r-1} \,|\, X_0) \\
&\leq P_N\left(\overline{\bigcap_{k=0}^{r-1} A_k} \,|\, X_0 = x\right) \\
&\leq \sum_{k=0}^{r-1} P_N(\overline{A}_k \cap \bigcap_{l=0}^{k-1} A_l \,|\, X_0 = x) \\
&\leq \sum_{k=0}^{r-1} P_N(\overline{A}_k \,|\, X_0 = x, \bigcap_{l=0}^{k-1} A_l).
\end{aligned}$$

When $\bigcap_{l=0}^{k-1} A_l$ holds,

$$U(X_{kN/r}) + \tau_{k-1} < (U_{\min} + \tau_{k-1})(\rho + 1),$$

therefore $U_{\min} + \tau_k > \eta_0$ and, applying the previous proposition,

$$\liminf_{N \to +\infty} -\frac{1}{\log N} \log P_N(\overline{A}_k \,|\, X_0 = x, \bigcap_{l=0}^{k-1} A_l) \geq \frac{\log(1+\rho)}{(1+\epsilon)\log(1+D_{\eta_0})},$$

thus

$$\liminf_{N \to +\infty} -\frac{1}{N} \log P_N(U(X_N) \geq U_{\min} + \rho(U_{\min} + \tau_{r-1}) \,|\, X_0 = x) \\ \geq \frac{\log(1+\rho)}{(1+\epsilon)\log(1+D_{\eta_0})}.$$

Moreover

$$\begin{aligned} \tau_k + U_{\min} &\leq (\tau_{k-1} + U_{\min})\frac{\rho}{1+\rho} + \eta_0 \\ &\leq \eta_0 \sum_{l=0}^{k-1} \left(\frac{\rho}{1+\rho}\right)^l + \left(\frac{\rho}{1+\rho}\right)^k (\tau_0 + U_{\min}), \end{aligned}$$

whence

$$(\tau_{r-1} + U_{\min}) \leq \eta_0(1+\rho) + \left(\frac{\rho}{1+\rho}\right)^{r-1} (U_{\min} + \eta_0 - \gamma). \qquad \square$$

9. A general remark about the interest of repeated optimisation schemes

All the algorithms we have encountered in these notes have a probability of failure bounded by $\epsilon(N)$, where N is their number of iterations and where $\lim_{N\to+\infty} N^{-1}\log\epsilon(N) = 0$. Due to this slow convergence speed, these algorithms should be used repeatedly. Indeed performing $N/\hat{M}$ repetitions of the algorithm with $\hat{M}$ iterations, where $\hat{M} \in \arg\min_{M\in\mathbb{N}} M^{-1}\log\epsilon(M)$, and keeping in the end the best solution among the $N/\hat{M}$ computed solutions, gives a probability of failure bounded from above by ξ^N with $\xi = \epsilon(\hat{M})^{1/\hat{M}}$ (when $N/\hat{M} \in \mathbb{N}$). The fact that $\lim_{M\to+\infty} M^{-1}\log\epsilon(M) = 0$ ensures that $\arg\min_{M\in\mathbb{N}} M^{-1}\log\epsilon(M)$ is not void and is bounded. See [2] and [7, 10] for more details.

10. Problem

The different questions are independent. The integer part of r is noted $\lfloor r \rfloor = \max\{n \in \mathbb{Z} \mid n \leq r\}$.

10.1. **Question 1.** Let us consider the state space $E = \{1,2,3,4,5\}$ and the rate function $V : E \times E \to \mathbb{R}_+ \cup \{+\infty\}$ defined by the following matrix

$$\begin{pmatrix} 0 & 1 & 3 & 0 & 2 \\ 8 & 0 & 2 & 2 & 3 \\ 9 & 5 & 0 & 7 & 4 \\ 0 & 2 & +\infty & 0 & +\infty \\ 8 & 5 & 4 & 11 & 0 \end{pmatrix}$$

(for instance $V(3,4)=7$.)

1.1. Compute the virtual energy of each state and construct all the cycles by induction.

1.2. Compute $H_1(V)$, $H_2(V)$ et $H_3(V)$.

1.3. Let us consider a family $(E^{\mathbb{N}}, (X_n)_{n\in\mathbb{N}}, \mathcal{B}, P_\beta)_{\beta\in\mathbb{R}_+}$ of homogeneous Markov chains with rare transitions with rate function V. For any subset D of E, we put $\tau(D) = \inf\{n \in \mathbb{N} \mid X_n \in D\}$. Compute

$$\lim_{\beta\to+\infty} -\frac{1}{\beta} \log P_\beta(X_{\tau(\{2,3,5\})} = 3 \mid X_0 = 4),$$

and

$$\lim_{\beta\to+\infty} \frac{1}{\beta} \log E_\beta(\tau(\{2,5\}) \mid X_0 = 3).$$

10.2. **Question 2.** Let us consider a family $(E^{\mathbb{N}}, (X_n)_{n\in\mathbb{N}}, \mathcal{B}, P_\beta)_{\beta\in\mathbb{R}_+}$ of homogeneous Markov chains with rare transitions defined on a finite state space E, with an irreducible rate function $V : E \times E \to \mathbb{R}_+ \cup \{+\infty\}$. Let $\tilde{U}$ be its virtual energy.

Let us assume that for some real positive constants a and b and for any $(x,y) \in E^2$

$$a \exp(-\beta V(x,y)) \leq p_\beta(x,y) \leq b \exp(-\beta V(x,y)),$$

where $p_\beta : E \times E \to [0,1]$ is the transition matrix of the chain P_β: For any $n \in \mathbb{N}$, $n > 0$,

$$p_\beta(x,y) = P_\beta(X_n = y \mid X_{n-1} = x).$$

2.1. Show that there is a positive real constant c such that for any subset D of E, $D \neq E$, $D \neq \emptyset$, any $x \in E \setminus D$, any $n \in \mathbb{N}$, any $\beta \in \mathbb{R}_+$,

$$P_\beta(\tau(D) > n \mid X_0 = x) \leq \exp\left(-\left\lfloor cne^{-\beta H(E\setminus D)}\right\rfloor\right),$$

where $\tau(D)$ is the first hitting time of D:

$$\tau(D) = \inf\{n \in \mathbb{N} \mid X_n \in D\}.$$

2.2. Deduce from this that there is a positive real constant d such that for any real positive $\eta \in \mathbb{R}_+$, any $x \in E$, any $\beta \in \mathbb{R}_+$,

$$P_\beta(\tilde{U}(X_n) \geq \eta \mid X_0 = x) \leq \exp\left(-\left\lfloor \frac{n}{d} e^{-\beta H_1(V)}\right\rfloor\right) + de^{-\eta\beta}.$$

2.3. Using the preceding inequalities, state a convergence theorem concerning $P_\beta(\tilde{U}(X_{N(\beta)} \geq \eta \mid X_0 = x)$ for a suitable function $N(\beta)$.

10.3. **Question 3: Weak reversibility condition of Hajek and Trouvé.** On a finite state space E, let us consider an irreducible rate function $V : E \times E \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ and a real valued function $U : E \rightarrow \mathbb{R}$. Let us define the elevation $H_U(\gamma)$ of a path $\gamma = (z_0, \cdots, z_r) \in E^{r+1}$ with respect to U by the formula

$$H_U(\gamma) = \max_{i=1,\cdots,r} U(z_{i-1}) + V(z_{i-1}, z_i).$$

For any $(x, y) \in E^2$, let $\Gamma_{x,y}$ be the set of paths joining x to y:

$$\Gamma_{x,y} = \bigcup_{r=1}^{+\infty} \{(z_0, \cdots, z_r) \in E^{r+1} \mid z_0 = x, z_r = y\}.$$

Let us define the minimum elevation between two states $x \in E$ and $y \in E$ by

$$H_U(x, y) = \min\{H_U(\gamma) \mid \gamma \in \Gamma_{x,y}\}.$$

3.1. Let us assume that the function $H_U(x, y)$ is symmetric. Namely, let us assume that for any $(x, y) \in E^2$

$$H_U(x, y) = H_U(y, x).$$

(This is a "weak reversibility condition", due to Hajek in the case when $p_\beta(x, y) = q(x, y) \exp(-\beta(U(y) - U(x))_+)$ with a non reversible kernel q and to Trouvé in the general case). Let $\tilde{U}$ be the virtual energy corresponding to (E, V). For any cycle $C \in \mathcal{C}(V)$, consider the following property $\mathcal{H}(C)$:

$$\forall (x, y) \in C^2, \quad U(x) - \tilde{U}(x) = U(y) - \tilde{U}(y).$$

Show by induction on $|C|$ that $\mathcal{H}(C)$ is true for any cycle $C \in \mathcal{C}(E, V)$.
Hints:

- Consider the partition $(C_i)_{i \in I}$ of C in strict maximal subcycles. Introduce the constants $c_i \in \mathbb{R}$, $i \in I$, defined by
$$U(x) = \tilde{U}(x) + c_i, \quad x \in C_i.$$
- Show that if $B(C_i) \cap C_j \neq \emptyset$, (where $B(C_i)$ is the principal boundary of C_i), then $c_i \geq c_j$. (For $x \in C_i$ and $y \in B(C_i) \cap C_j$ compare $H_{\tilde{U}}(x, y)$, $H_U(x, y)$, $H_U(y, x)$ and $H_{\tilde{U}}(y, x)$.)
- Draw from this the conclusion that $c_i = c_j$ for any $(i, j) \in I^2$.

This shows that $(H_U(x, y))_{(x,y) \in E^2}$ is symmetric if and only if for any $x \in E$

$$U(x) = \min_{y \in E} U(y) + \tilde{U}(x).$$

10.4. **Question 4.**

4.1. Give an example of a finite state space E and of an irreducible rate function $V : E \times E \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ such that

$$\begin{aligned} H_1(E, V) &= 1, \\ H_2(E, V) &= 2, \\ H_3(E, V) &= 3. \end{aligned}$$

4.2. Could-you give such an example in which $|E| = 4$?

4.3. Could-you give such an example in which $|E| = 5$?

Acknowledgement: I would like to thank Alain Trouvé and Cécile Cot for their useful remarks on a first French draft of these lecture notes. My thanks also go to Robert Azencott, who encouraged my interest in stochastic optimisation since the early times of my Phd dissertation. I am grateful to the referee for his careful reading and numerous suggestions that were very precious to help me improve the presentation of many results.

REFERENCES

[1] Azencott Robert (1988) Simulated Annealing, *Séminaire Bourbaki 40ième année, 1987-1988* **697**.

[2] Azencott Robert (1992) Sequential Simulated Annealing: Speed of Convergence and Acceleration Techniques, in *Simulated Annealing: Parallelization Techniques*, R. Azencott Ed., Wiley Interscience.

[3] Azencott Robert (1992) A Common Large Deviations Mathematical Framework for Sequential Annealing and Parallel Annealing, in *Simulated Annealing: Parallelization Techniques*, R. Azencott Ed., Wiley Interscience.

[4] Azencott Robert and Graffigne Christine (1992) Parallel Annealing by Periodically Interacting Multiple Searches: Acceleration Rates, in *Simulated Annealing: Parallelization Techniques*, R. Azencott Ed., Wiley Interscience.

[5] Catoni Olivier (1991) Exponential Triangular Cooling Schedules for Simulated Annealing Algorithms: a case study, *Applied Stochastic Analysis, Proceedings of a US-French Workshop, Rutgers University, April 29 - May 2, 1991,* Karatzas I. and Ocone D. eds., Lecture Notes in Control and Information Sciences No 177, Springer Verlag, 1992.

[6] Catoni Olivier (1992) Rough Large Deviation Estimates for Simulated Annealing: Application to Exponential Schedules, *The Annals of Probability*, Vol. 20, nb. 3, pp. 1109 - 1146.

[7] Catoni Olivier, (1998) The Energy Transformation Method for the Metropolis Algorithm Compared with Simulated Annealing. *Probab. Theory Related Fields 110 (1998), no. 1.*, pages 69–89.

[8] Catoni Olivier and Cerf Raphael (1997) The Exit Path of a Markov Chain with Rare Transitions, *ESAIM:P&S*, vol 1, pp. 95-144, http://www.emath.fr/Maths/Ps/ps.html.

[9] Catoni Olivier (1998) Solving Scheduling Problems by Simulated Annealing. *SIAM J. Control Optim. 36, no. 5, (electronic)*, pages 1539–1575.

[10] Catoni Olivier (1996) Metropolis, Simulated Annealing and I.E.T. Algorithms: Theory and Experiments. *Journal of Complexity 12, special issue on the conference Foundation of Computational Mathematics, January 5-12 1997, Rio de Janeiro*, pages 595–623, December 1996.

[11] Cot Cécile and Catoni Olivier (1998) Piecewise constant triangular cooling schedules for generalized simulated annealing algorithms. *Ann. Appl. Probab. 8, no. 2,*, pages 375–396.

[12] Deuschel J.D. and Mazza C. (1994) L^2 convergence of time nonhomogeneous Markov processes: I. Spectral Estimates, *The annals of Applied Probability*, vol. 4, no. 4, 1012-1056.

[13] Diaconis Persi and Stroock Daniel (1991) Geometric Bounds for Eigenvalues of Markov Chains, *The Annals of Applied Probability*, Vol. 1, No 1, 36 - 61.

[14] Duflo M. (1996) *Algorithmes Stochastiques*, Mathématiques & Applications (Paris), Springer Verlag.

[15] Fill J. A. (1991) Eigenvalue bounds on the convergence to stationarity for nonreversible Markov chains, with an application to the exclusion process, *Ann. Applied Probab.*, **1**.

[16] Freidlin, M. I. and Wentzell, A. D. (1984). *Random Perturbations of Dynamical Systems.* Springer, New York.

[17] Geman S., Geman D., *Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images*, I.E.E.E. Transactions on Pattern Analysis and Machine Intelligence, 6, 721- 741, 1984.

[18] Götze F. (1991) Rate of Convergence of Simulated Annealing Processes, *preprint.*

[19] Graffigne Christine (1992) Parallel Annealing by Periodically Interacting Multiple Searches: An Experimental Study, in *Simulated Annealing: Parallelization Techniques*, R. Azencott Ed., Wiley Interscience.

[20] Holley R. and Stroock D. (1988) Annealing via Sobolev inequalities, *Comm. Math. Phys.*, 115:553-559.

[21] Holley, R. A., Kusuoka, S. and Stroock, D. W. (1989), Asymptotics of the spectral gap with applications to the theory of simulated annealing, *Journal of functional analysis*, **83**, 333-347.

[22] Hwang, C. R. and Sheu, S. J. (1992) Singular perturbed Markov chains and exact behaviour of simulated annealing processes. *J. Theoret. Prob.*, **5**, 2, 223-249.

[23] Ingrassia S. (1994) On the rate of convergence of the Metropolis algorithm and Gibbs sampler by geometric bounds, *Ann. Appl. Probab.* **4**, no.2, 347-389.

[24] Kirchhoff G. (1847) Über die Auflösung der Gleichungen, auf welche man beider Untersuchung der linearen Verteilung galvanischer Ströme gefuhrt wird, *Ann. Phys. Chem.*, 72, pp. 497-508.(English transl. IRE Trans. Circuit Theory CT-5 (1958) 4-7).

[25] Kirkpatrick S., Gelatt C. D. and Vecchi M. P., (1983) Optimization by simulated annealing, *Science*, 220, 621-680, 1983.

[26] Miclo Laurent (1991) Evolution de l'énergie libre. Application à l'étude de la convergence des algorithmes du recuit simulé. *Doctoral Dissertation*, Université d'Orsay, February 1991.

[27] Miclo Laurent (1996) Sur les problèmes de sortie discrets inhomogènes *Ann. Appl. Probab.* **6**, no 4, 1112-1156.

[28] Miclo Laurent (1995) Sur les temps d'occupations des processus de Markov finis inhomogènes à basse température, *submitted to Stochastics and Stochastics Reports.*

[29] Miclo Laurent (1997) Remarques sur l'hypercontractivité et l'évolution de l'entropie pour des chaînes de Markov finies, *Séminaire de Probabilités XXXI*, Lecture Notes in Mathematics 1655, Springer.

[30] Saloff-Coste, Laurent (1997) Lectures on finite Markov chains *Lectures on probability theory and statistics (Saint-Flour, 1996)*, 301–413, Lecture Notes in Math., 1665, Springer, Berlin.

[31] Trouvé Alain (1993) Parallélisation massive du recuit simulé, *Doctoral Dissertation*, Université Paris 11, January 5 1993.

[32] Trouvé Alain (1994) Cycle Decomposition and Simulated Annealing, *S.I.A.M. J. Control Optim., 34(3)*, 1996.

[33] Trouvé Alain (1995) Rough Large Deviation Estimates for the Optimal Convergence Speed Exponent of Generalized Simulated Annealing Algorithms, *Ann. Inst. H. Poincaré, Probab. Statist.*, 32(2), 1996.