# Méthodes Numériques Probabilistes

## Exercices sur la méthode de Monte Carlo et les chaînes de Markov

# 1 Monte Carlo methods

## 1.1 Antithetic variables

Let $f : [0,1] \to \mathbb{R}$ be such that

$$\int_{u=0}^{1} f(u)^2 \mathrm{d}u < +\infty.$$

We study a Monte Carlo method to approximate

$$\mathcal{I} := \int_{u=0}^{1} f(u) \mathrm{d}u.$$

1. Let $U \sim \mathcal{U}[0,1]$. Show that $\mathcal{I} = \frac{1}{2}\left(\mathbb{E}[f(U)] + \mathbb{E}[f(1-U)]\right)$.
2. Let $(U_n)_{n \geq 1}$ be a sequence of independent copies of $U$. Show that

   $$\widehat{\mathcal{I}}^{\mathrm{a}}_{2n} := \frac{1}{2n} \sum_{i=1}^{n} \left(f(U_i) + f(1-U_i)\right)$$

   converges almost surely to $\mathcal{I}$ and compute $\mathrm{Var}(\widehat{\mathcal{I}}^{\mathrm{a}}_{2n})$.
3. Let

   $$\widehat{\mathcal{I}}_{2n} := \frac{1}{2n} \sum_{i=1}^{2n} f(U_i)$$

   be the standard Monte Carlo estimator of $\mathcal{I}$ which requires the same number of evaluations of the function $f$ as $\widehat{\mathcal{I}}^{\mathrm{a}}_{2n}$ (but twice more random samples). Show that $\mathrm{Var}(\widehat{\mathcal{I}}^{\mathrm{a}}_{2n}) \leq \mathrm{Var}(\widehat{\mathcal{I}}_{2n})$ if and only if $\mathrm{Cov}(f(U), f(1-U)) \leq 0$.
4. Assume that $f$ is monotonic. Show that

   $$\mathbb{E}\left[(f(U_1) - f(U_2))(f(1-U_1) - f(1-U_2))\right] \leq 0.$$

   Deduce that in this case, $\mathrm{Cov}(f(U), f(1-U)) \leq 0$.
5. Conclude on the practical interest of the method.

## 1.2 Stratification

Let $X$ be a random variable in $\mathbb{R}^d$ with law $P$ and $f \in \mathbf{L}^2(P)$. Let

$$\mathcal{I} = \int_{x \in \mathbb{R}^d} f(x) \mathrm{d}P(x) = \mathbb{E}[f(X)].$$

We assume that there is a finite partition of $\mathbb{R}^d$ into $m$ measurable subsets $(A_k)_{1 \leq k \leq m}$, called *strates*, such that for any $k \in \{1, \ldots, m\}$:

- $p_k := P(A_k) = \mathbb{P}(X \in A_k)$ is known (and positive);
- we know how to draw random samples $(X_n^k)_{n\geq 1}$ under the law $P(\cdot|A_k) = \mathbb{P}(X \in \cdot|X \in A_k)$.

For integers $n_1, \dots, n_m \geq 1$ such that $n_1 + \dots + n_m = n$, we set

$$\widehat{\mathfrak{I}}_n^{\mathrm{s}} := \sum_{k=1}^{m} p_k \widehat{\mathfrak{I}}_{n_k}^k, \qquad \widehat{\mathfrak{I}}_{n_k}^k := \frac{1}{n_k}\sum_{i=1}^{n_k} f(X_i^k),$$

where the samples $(X_i^1)_{1\leq i\leq n_1}, \dots, (X_i^m)_{1\leq i\leq n_m}$ are independent from each other. Last, we define

$$\forall k \in \{1, \dots, m\}, \qquad \mu_k := \mathbb{E}[f(X_1^k)], \quad \sigma_k^2 := \mathrm{Var}(f(X_1^k)).$$

### 1.2.1 Generalities

1. Show that
$$\mathrm{Var}(f(X)) = \sum_{k=1}^{m} p_k\sigma_k^2 + \sum_{k=1}^{m} p_k \left(\mu_k - \sum_{\ell=1}^{m} p_\ell \mu_\ell\right)^2.$$
Give an interpretation of this formula.
2. Compute $\mathbb{E}[\widehat{\mathfrak{I}}_n^{\mathrm{s}}]$.
3. How does $\widehat{\mathfrak{I}}_n^{\mathrm{s}}$ behave when $\min(n_1, \dots, n_m) \to +\infty$?
4. Show that $\mathrm{Var}(\widehat{\mathfrak{I}}_n^{\mathrm{s}}) = \sum_{k=1}^{m} \frac{p_k^2\sigma_k^2}{n_k}$.

### 1.2.2 Optimal allocation

We now fix $n$ and look for the *optimal* allocation of $(n_1, \dots, n_m)$.

1. Show that, for any $n_1, \dots, n_m$,
$$\left(\sum_{k=1}^{m} p_k\sigma_k\right)^2 \leq n\sum_{k=1}^{m}\frac{p_k^2\sigma_k^2}{n_k}.$$
2. Deduce the optimal allocation $(n_1^*, \dots, n_m^*)$ in terms of variance (without taking into account the constraint that $n_k$ must be an integer).
3. What do you think of the practical use of this optimal allocation?

### 1.2.3 Proportional allocation

We finally study the *proportional* allocation $n_k = np_k$, assuming for simplicity that $np_k$ is an integer.

1. Show that in this case $n\,\mathrm{Var}(\widehat{\mathfrak{I}}_n^{\mathrm{s}}) \leq \mathrm{Var}(f(X))$. Interpret this result.
2. State and prove a Central Limit Theorem for $\widehat{\mathfrak{I}}_n^{\mathrm{s}}$.
3. How to choose the strates to reduce the statistical error?

### 1.3 Splitting for rare events

Let $X$ be a random variable in $\mathbb{R}^d$ with law $P$. Let $V : \mathbb{R}^d \to (0, +\infty)$ and $a > 0$. We are interested in the estimation of the probability

$$p := \mathbb{P}(V(X) > a),$$

which we assume to be very small. In practice, the function $V$ typically measures a risk, $a$ a threshold and $p$ a probability of failure.

*Preliminary question.* For $n \geq 1$, let

$$\widehat{p}_n := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{V(X_i)>a\}}$$

be the standard Monte Carlo estimator of $p$, where $X_1, \ldots, X_n$ are iid under $P$. For a fixed value of $n$, recall the asymptotic behaviour of the *relative error* $\sqrt{\mathrm{Var}(\widehat{p}_n)}/p$ when $p \to 0$.

#### 1.3.1 Splitting with given levels

1. The splitting method is defined as follows. Let $0 = a_0 < a_1 < \cdots < a_m = a$ be subdivision of the interval $[0, a]$. Show that

$$p = \prod_{k=1}^{m} p_k, \qquad p_k := \mathbb{P}\left(V(X) > a_k | V(X) > a_{k-1}\right).$$

2. For $k \in \{1, \ldots, m\}$, we assume that we known how to sample random variables $(X_n^k)_{n \geq 1}$ under the law $P(\cdot | V > a_{k-1})$ (that is to say, the conditional measure $P$ on the set $\{x \in \mathbb{R}^d : V(x) > a_{k-1}\}$). We assume that the sequences $(X_n^k)_{n \geq 1}$, $k \in \{1, \ldots, m\}$ are independent from each other. We then consider the estimator

$$\widehat{p}_n^{\mathrm{s}} := \prod_{k=1}^{m} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{V(X_i^k)>a_k\}}.$$

   Show that $\mathbb{E}[\widehat{p}_n^{\mathrm{s}}] = p$ and that, almost surely, $\widehat{p}_n^{\mathrm{s}} \to p$.

3. Show that

$$\lim_{n \to +\infty} n \,\mathrm{Var}\left(\widehat{p}_n^{\mathrm{s}}\right) = mp^2 \left( -1 + \frac{1}{m} \sum_{k=1}^{m} \frac{1}{p_k} \right).$$

4. Show that, for any $p_1, \ldots, p_m > 0$ such that $\prod_{k=1}^{m} p_k = p$,

$$\frac{1}{m} \sum_{k=1}^{m} \frac{1}{p_k} \geq \frac{1}{p^{1/m}},$$

   and compute the vector $(p_1^*, \ldots, p_m^*)$ for which this lower bound is reached.

5. From now on we set $\alpha = p^{1/m}$. How to choose the levels $a_1, \ldots, a_{m-1}$ to minimise the variance of $\widehat{p}_n^{\mathrm{s}}$ in the $n \to +\infty$ limit?

6. Express the value of $\lim_{n \to +\infty} n \,\mathrm{Var}(\widehat{p}_n^{\mathrm{s}})$ in terms of $\alpha$ for the optimal choice of $a_1, \ldots, a_{m-1}$. For which value of $\alpha$, and therefore of $m$, is this quantity minimal?

### 1.3.2 Adaptive splitting

We now study an algorithm which allows to generate the levels $a_k$ so that, approximately,

$$p_1 = \ldots = p_m = \alpha = 1 - \frac{1}{n},$$

and therefore the number $m$ of levels will be such that, approximately, $(1-1/n)^m = p$. More precisely, we consider the *adaptive splitting* algorithm, which makes evolve a set of $n$ random variables as follows:

- *Initialisation.* At iteration $k = 0$, sample iid variables $X_1^0, \ldots, X_n^0$ with law $P$, and set

  $$a_0 := \min_{1 \leq i \leq n} V(X_i^0), \qquad i_0 := \arg\min_{1 \leq i \leq n} V(X_i^0).$$

  We assume that $V(X)$ has a density on $(0, +\infty)$, so that $i_0$ is almost surely uniquely defined.

- *Iterations.* At iteration $k \geq 1$, the variables $X_1^k, \ldots, X_n^k$ are obtained from $X_1^{k-1}, \ldots, X_n^{k-1}$ as follows:

  - For $i = i_{k-1}$, the variable $X_i^k$ is freshly drawn according to the measure $P(\cdot | V > a_{k-1})$.
  - For $i \neq i_{k-1}$, we set $X_i^k = X_i^{k-1}$.

  We next set

  $$a_k := \min_{1 \leq i \leq n} V(X_i^k), \qquad i_k := \arg\min_{1 \leq i \leq n} V(X_i^k).$$

- *Stopping criterion.* The algorithm stops when $a_k > a$, and we set

  $$m = \min\{k \geq 1 : a_k > a\}, \qquad \widehat{p}_n^{\mathrm{as}} = \left(1 - \frac{1}{n}\right)^m.$$

Notice that the numbers of levels $m$ and their values $a_0, \ldots, a_m$ are random.

1. Let $F$ be the CDF of $V(X)$, and $\Lambda(y) := -\log(1 - F(y))$. Check that $\Lambda$ is increasing and compute $\Lambda(0)$ and $\Lambda(a)$.

2. Compute the law of $\Lambda(V(X))$.

3. Let $b \geq 0$ and $Y$ a random variable with law $P(\cdot | V > b)$. Show that, for any $z \in \mathbb{R}$,

   $$\mathbb{P}\left(\Lambda(V(Y)) > z\right) = \exp\left(\Lambda(b) - \max(z, \Lambda(b))\right).$$

4. Show that if $X$ and $Y$ are independent random variables, then for any $f$ such that $f(X, Y) \in \mathbf{L}^1(\mathbb{P})$,

   $$\mathbb{E}[f(X, Y)] = \mathbb{E}[g(X)], \qquad g(x) := \mathbb{E}[f(x, Y)].$$

5. Show that the random variables $(\Lambda(V(X_i^1)) - \Lambda(a_0))_{1 \leq i \leq n}$ are independent and exponentially distributed with parameter 1, and that they are independent from the random variable $\Lambda(a_0)$, which is exponentially distributed with parameter $n$. *Hint: you may for instance compute, for any $z, z_1, \ldots, z_n$, $\mathbb{P}(\Lambda(V(X_1^1)) - \Lambda(a_0) > z_1, \ldots, \Lambda(V(X_n^1)) - \Lambda(a_0) > z_n, \Lambda(a_0) > z)$.*

6. Show that for any $k \geq 1$, the random variables $(\Lambda(V(X_i^k)) - \Lambda(a_{k-1}))_{1 \leq i \leq n}$ are independent and exponentially distributed with parameter 1, and that they are independent from the random variable $(\Lambda(a_{\ell-1}) - \Lambda(a_{\ell-2})_{1 \leq \ell \leq k}$, which are exponentially distributed with parameter $n$. *Hint: you may argue by induction.*

7. Deduce the law of $m$ and show that $\mathbb{E}[\widehat{p}_n^{\mathrm{as}}] = p$.

8. Compute $\mathrm{Var}(\widehat{p}_n^{\mathrm{as}})$. How does the relative error $\sqrt{\mathrm{Var}(\widehat{p}_n^{\mathrm{as}})}/p$ behave when $p \to 0$?

# 2 Markov chains

## 2.1 Coupling from the past

Given a probability measure $\pi$ on a discrete space $E$, we study an algorithm which returns a random variable $Y$ with *exact* law $\pi$, based on the construction of a Markov chain $(X_n)_{n\geq 0}$ which admits $\pi$ as stationary distribution. We first describe the algorithm in a general setting, and then its application to the Metropolis–Hastings algorithm for the simulation of the Ising model.

### 2.1.1 General description

We let $(X_n)_{n\geq 0}$ be a Markov chain in $E$ given under the form of a random dynamical system

$$X_{n+1} = f(X_n, Z_{n+1}),$$

where $(Z_n)_{n\geq 1}$ is a sequence of iid random variables in some measurable space $\mathcal{Z}$ and $f : E \times \mathcal{Z} \to E$ is measurable.

1. Recall the expression of the transition matrix $P$ of $(X_n)_{n\geq 0}$ in terms of $f$.

For any $z \in \mathcal{Z}$, we denote by $f_z : E \to E$ the function defined by $f_z(x) = f(x, z)$. We next define the random mappings $D_n, G_n : E \to E$ by

$$D_0 = G_0 = \mathrm{Id}, \qquad D_n = f_{Z_1} \circ f_{Z_2} \circ \cdots \circ f_{Z_n}, \quad G_n = f_{Z_n} \circ f_{Z_{n-1}} \circ \cdots \circ f_{Z_1}.$$

From now on, we assume that $E$ is finite, and that $P$ is irreducible and aperiodic, with unique stationary distribution $\pi$.

2. For any $x, y \in E$, express $\lim_{n\to+\infty} \mathbb{P}(G_n(x) = y)$ in terms of $\pi$.

We now define $N := \inf\{n \geq 0 : D_n \text{ is a constant function}\}$, with the convention that $\inf \emptyset = +\infty$.

3. Show that if $n \geq N$ then $D_n$ remains a constant function.
4. Assume that $N < +\infty$, almost surely, and let $Y \in E$ be the value of the constant function $D_N$. Show that $Y$ has law $\pi$. *Hint: you may compare the laws of $D_n(x)$ and $G_n(x)$ for fixed $n \geq 1$ and $x \in E$.*
5. In this question, we describe a sufficient condition for assumption $N < +\infty$ to hold. We assume that for any subset $A \subset E$ such that $|A| > 1$,

   $$\mathbb{P}(|f_Z(A)| < |A|) > 0.$$

   Let $\mathsf{m} = |E|$. Show that there exists $\kappa > 0$ such that $\mathbb{P}(N \leq \mathsf{m} - 1) \geq \kappa$, and deduce that $\mathbb{P}(N < +\infty) = 1$.
6. Which numerical difficulty does the simulation of the random variable $Y$ imply?

In the sequel, we assume that $E$ is endowed with an order relation $\leq$ and that there exist two elements $x_{\min}, x_{\max} \in E$ such that

$$\forall x \in E, \qquad x_{\min} \leq x \leq x_{\max}.$$

We moreover assume that $f$ is monotonic in the sense that

$$\forall x, y \in E, \quad \forall z \in \mathcal{Z}, \qquad x \leq y \Rightarrow f(x, z) \leq f(y, z).$$

In this case, the Markov chain $(X_n)_{n\geq 0}$ is called *monotonic*.

7. How to compute $N$ in practice if the Markov chain $(X_n)_{n\geq 0}$ is monotonic?

8. Let $M := \inf\{n \geq 0 : G_n(x_{\min}) = x_{\max}\}$. Show that for any $n \geq 0$, $\mathbb{P}(N > n) \leq \mathbb{P}(M > n)$, and deduce that $N < +\infty$, almost surely.

### 2.1.2 Application to the Ising model

We consider the Ising model $\pi_\beta$ on a finite graph $(\mathcal{V}, \mathcal{E})$ and inverse temperature $\beta$. We next let $(X_n)_{n\geq n}$ be the Markov chain defined by the Metropolis algorithm with:

- proposal matrix $Q$ which at each step picks a vertex uniformly in $\mathcal{V}$ and a new value for the spin uniformly in $\{-1, 1\}$;
- accept the move according to the Metropolis–Hastings rule for the target measure $\pi_\beta$.

1. Show that this chain takes the form $X_{n+1} = f(X_n, Z_{n+1})$ and describe precisely $f$ as well as $(Z_n)_{n\geq 1}$.
2. Justify the fact that $(X_n)_{n\geq 0}$ is irreducible and aperiodic, and has $\pi_\beta$ as stationary distribution.
3. Show that the chain $(X_n)_{n\geq 0}$ is monotonic.
4. Deduce a complete algorithm which returns a configuration exactly distributed under $\pi_\beta$.

## 2.2 Asymptotic variance for nonreversible Markov chains

### 2.2.1 Asymptotic variance and Poisson equation

Let us consider a stochastic matrix $P$ on $E$, which is irreducible, aperiodic, and with stationary distribution $\pi$. Let us recall that for any function $f \in \mathbb{R}^E$,

$$\lim_{n\to\infty} \sqrt{n} \left( \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) - \pi f \right) = \mathcal{N}\left(0, \sigma^2(f)\right), \qquad \text{in distribution,}$$

where $(X_n)_{n\geq 0}$ is a Markov chain with transition matrix $P$ and

$$\sigma^2(f) = \mathrm{Var}_\pi(f(X_0)) + 2\sum_{n=1}^{\infty} \mathrm{Cov}_\pi(f(X_0), f(X_n)).$$

In the sequel it will be convenient to denote by $\sigma^2(f, P)$ the asymptotic variance, to keep track of the transition matrix of the underlying chain $(X_n)_{n\geq 0}$.

We define $\mathbb{R}^E_0 = \{g \in \mathbb{R}^E : \pi g = 0\}$ and for $f \in \mathbb{R}^E$ we set $\widetilde{f} = f - \pi f \in \mathbb{R}^E_0$.

1. Show that if $g \in \mathbb{R}^E_0$ then $Pg \in \mathbb{R}^E_0$. We denote by $P_0$ the restriction of $P$ to $\mathbb{R}^E_0$.
2. Show that for any $f \in \mathbb{R}^E$, there is a unique $g \in \mathbb{R}^E_0$ such that $(I - P)g = \widetilde{f}$.
3. Show that $g(x) = \sum_{n=0}^{+\infty} \mathbb{E}_x[\widetilde{f}(X_n)]$.
4. Deduce that $\sigma^2(f, P) = 2\langle \widetilde{f}, g\rangle_\pi - \|\widetilde{f}\|^2_\pi$.

### 2.2.2 Reversible and nonreversible chains

Let $S$ be a stochastic matrix on $E$ which is irreducible, aperiodic and reversible with respect to $\pi$. Let $A$ be a $E \times E$ matrix which satisfies:

$$\forall (x,y) \in E \times E, \quad \pi(x)A(x,y) = -\pi(y)A(y,x)\,, \tag{1}$$

$$\forall (x,y) \in E \times E, \quad \begin{cases} A(x,y) = 0 \text{ if } S(x,y) = 0, \\ |A(x,y)| < S(x,y) \text{ if } S(x,y) \neq 0, \end{cases} \tag{2}$$

$$\forall x \in E, \quad \sum_{y \in E} A(x,y) = 0. \tag{3}$$

Let us finally define $P = S + A$.

1. Show that for all $(x,y) \in E \times E$, $S(x,y) > 0$ if and only if $P(x,y) > 0$.
2. Show that $P$ is a stochastic and aperiodic matrix which admits $\pi$ as a unique stationary distribution.

Let us recall a version of the Spectral Theorem adapted to our framework, and which will be useful in the following. Let us first extend the scalar product $\langle \cdot, \cdot \rangle_\pi$ to complex valued functions: for any functions $u : E \to \mathbb{C}$ and $v : E \to \mathbb{C}$,

$$\langle u, v \rangle_\pi = \sum_{x \in E} \overline{u(x)} v(x) \pi(x),$$

where $\overline{z}$ denotes the complex conjugate of $z \in \mathbb{C}$. For a complex valued matrix $M \in \mathbb{C}^{E \times E}$, let us denote its adjoint by $M^* \in \mathbb{C}^{E \times E}$, defined by: for any functions $u : E \to \mathbb{C}$ and $v : E \to \mathbb{C}$,

$$\langle M^* u, v \rangle_\pi = \langle u, Mv \rangle_\pi.$$

From the Spectral Theorem, if $M$ is Hermitian (i.e. $M^* = M$), then $M$ is diagonalisable in an orthonormal (for the Hermitian inner product $\langle \cdot, \cdot \rangle_\pi$) basis of $\mathbb{C}^{\mathsf{m}}$, with real eigenvalues.

3. Prove that $A$ is a diagonalisable matrix, and that its eigenvalues are purely imaginary with modules strictly smaller than 1.
4. What can be said of the results of the three previous questions if Assumption (2) is replaced by the weaker hypothesis: $\forall (x,y) \in E \times E$, $|A(x,y)| \leq S(x,y)$?

The objective of the following questions is to study why introducing the matrix $A$ may improve the efficiency of a Markov Chain Monte Carlo sampling algorithm, by comparing the asymptotic variance associated with the transition matrix $P = S + A$ with the asymptotic variance associated with the transition matrix $S$.

Let us recall that the matrices $S$, $A$, and $P$ leave $\mathbb{R}_0^E$ invariant. These matrices are thus seen in the following as endomorphisms of the linear $(\mathsf{m}-1)$-dimensional space $\mathbb{R}_0^E$, with $\mathsf{m}$ the cardinality of $E$.

5. The linear map $S : \mathbb{R}_0^E \to \mathbb{R}_0^E$ is symmetric for the scalar product $\langle \cdot, \cdot \rangle_\pi$, and is thus diagonalisable: let us denote by $(d_i)_{1 \leq i \leq \mathsf{m}-1} \in \mathbb{R}^{\mathsf{m}-1}$ its eigenvalues and by $(f_i : E \to \mathbb{R})_{1 \leq i \leq \mathsf{m}-1}$ the associated eigenvectors, which form an orthonormal basis of $\mathbb{R}_0^E$. For any $\alpha \in \mathbb{R}$, let us define the operator $(I - S)^\alpha : \mathbb{R}_0^E \to \mathbb{R}_0^E$ by:

   $$\forall i \in \{1, \dots, \mathsf{m}-1\}, \qquad (I - S)^\alpha f_i = (1 - d_i)^\alpha f_i.$$

   Check that for all $\alpha \in \mathbb{R}$, $(I - S)^\alpha$ is a well-defined symmetric operator (for the scalar product $\langle \cdot, \cdot \rangle_\pi$), and that $(I - S)^\alpha (I - S)^\beta = (I - S)^{\alpha + \beta}$, for all $\alpha, \beta \in \mathbb{R}$.

6. Let $Q = (I - S)^{-1/2} A (I - S)^{-1/2}$. Check that $I - Q$ is invertible. Show that for all $g \in \mathbb{R}_0^E$,

$$\langle (I - Q)^{-1} g, g \rangle_\pi \leq \langle g, g \rangle_\pi .$$

   When does equality hold?

7. Prove that for all $f : E \to \mathbb{R}$,

$$\langle (I - P)^{-1} \widetilde{f}, \widetilde{f} \rangle_\pi = \langle (I - Q)^{-1} (I - S)^{-1/2} \widetilde{f}, (I - S)^{-1/2} \widetilde{f} \rangle_\pi .$$

8. Deduce that for any function $f : E \to \mathbb{R}$, $\sigma^2(f, P) \leq \sigma^2(f, S)$. When does equality hold?

### 2.2.3 Construction of $A$

In these last two questions, we would like to build, for a given matrix $S$, a matrix $A$ which satisfies the three properties (1)-(2)-(3). In order to do so, let us consider a cycle for $S$, namely a path in $E$ with strictly positive probability for $S$, starting from a state and coming back to the same state, while visiting at least one other state.

1. For any $x \in E$, check that there exists a cycle for $S$ starting from $x$ and coming back to $x$.

Up to a renumbering of the states, we thus have:

$$\text{for some } \ell \geq 2, \; S(1,2) S(2,3) \ldots S(\ell, 1) > 0.$$

Let us then consider

$$\alpha = \min \left( \pi(1) S(1,2), \; \pi(2) S(2,3), \ldots, \pi(\ell - 1) S(\ell - 1, \ell), \; \pi(\ell) S(\ell, 1) \right).$$

For $t \in (-\alpha, \alpha)$, let us define $A_t$ by:

$$A_t(x,y) = \begin{cases} \frac{t}{\pi(x)} & \text{if } (x,y) = (\ell, 1) \text{ or } (x,y) = (i, i+1) \text{ for some } i \in \{1, \ldots, \ell - 1\}, \\ -\frac{t}{\pi(x)} & \text{if } (x,y) = (1, \ell) \text{ or } (x,y) = (i, i-1) \text{ for some } i \in \{2, \ldots, \ell\}, \\ 0 & \text{otherwise.} \end{cases}$$

2. Show that the matrix $A_t$ satisfies (1)-(2)-(3).