Exercise Sheet 3
Poincaré and *log*-Sobolev inequalities in the context of Mixing times

Most of material of this exercise sheet is taken from Ravi Montenegro and Prasad Tetali (2006), *Mathematical Aspects of Mixing Times in Markov Chains*, Foundations and Trends in Theoretical Computer Science, which is accessible https://tetali.math.gatech.edu/PUBLIS/survey.pdf. **We do not expect you to solve all the exercises in this sheet. You can use these exercises as useful references and perhaps to practice in the future.**

Let $(\Omega, \mathsf{P}, \pi)$ be a *Markov kernel* of a finite Markov chain on a finite space Ω with a unique invariant measure π. That is,

- $\mathsf{P}(x, y) \geq 0 \; \forall x, y \in \Omega$,
- $\sum_{y\in\Omega} \mathsf{P}(x, y) = 1 \; \forall x \in \Omega$,
- $\sum_{x\in\Omega} \pi(x)\mathsf{P}(x, y) = \pi(y) \; \forall x, y \in \Omega$.

Let us assume that P is irreducible and π has full support on Ω.
If $A, B \subset \Omega$ the *ergodic flow* is

$$\mathsf{Q}(A, B) = \sum_{x\in A, y\in B} \pi(x)\mathsf{P}(x, y)$$

Exercise 1

Let P be a Markov kernel. Define $\mathsf{Q} = \frac{\mathsf{P}+\mathsf{I}}{2}$. Show that Q is also a Markov kernel.
Q is said to be *lazy*.

The *minimal holding probability* $\alpha \in [0, 1]$ satisfies $\forall x \in \Omega \; \mathsf{P}(x, x) \geq \alpha$.
If P is aperiodic[1] then $\mathsf{P}^n(x, \cdot)$ approach π as $n \to \infty$. We only consider aperiodic Markov chains.
The *total variation* between two probability measures μ, ν is defined as

$$||\mu - \nu||_{\mathrm{TV}} = \max_{A\subset\Omega} |\mu(A) - \nu(A)| \,. \tag{1}$$

Exercise 2

Prove that

$$||\mu - \nu||_{\mathrm{TV}} = \frac{1}{2} \sum_{x\in\Omega} |\mu(x) - \nu(x)| \,. \tag{2}$$

The *mixing time with respect to total variation* is defined as

$$\tau(\varepsilon) := \min \left\{ n : \forall x \in \Omega \text{ such that } ||\mathsf{P}^n(x, \cdot) - \pi||_{\mathrm{TV}} \leq \varepsilon \right\} .$$

A Markov chain is *time-reversal* P^* if

$$\pi(x)\mathsf{P}^*(x, y) = \pi(y)\mathsf{P}(y, x) \; \forall x, y \in \Omega \,.$$

We define the standard inner product on $L^2(\pi)$, i.e.

$$\langle f, g\rangle_\pi = \sum_{x\in\Omega} \pi(x)f(x)g(x) \,.$$

[1] A Markov chain is said to be aperiodic if the chain does not exhibit any regularity in its behavior, meaning that it doesn't tend to revisit certain states at regular intervals.

Exercise 3
Prove

$$\langle f, \mathsf{P}g\rangle_\pi = \langle \mathsf{P}^* f, g\rangle_\pi \,.$$

Recall that the *Dirichlet form* is defined as

$$\mathcal{E}_{\mathsf{P}}(f,g) = \langle f, (\mathsf{I}-\mathsf{P})g\rangle_\pi \,.$$

and the *Poincaré inequality* has the form

$$\lambda \mathrm{Var}_\pi(f) \le \mathcal{E}(f,f) \;\forall f : \Omega \to \mathbb{R}\,, \tag{3}$$

where λ is optimal.

Exercise 4
Prove

$$\mathcal{E}_{\mathsf{P}}(f,f) = \mathcal{E}_{\mathsf{P}^*}(f,f) = \mathcal{E}_{\frac{\mathsf{P}+\mathsf{P}^*}{2}}(f,f)\,.$$

Exercise 5
If P is reversible, prove

$$\mathcal{E}_{\mathsf{P}}(f,g) = \mathcal{E}_{\mathsf{P}}(g,f)\,.$$

Let $k_n^x(y) = \frac{\mathsf{P}^n(x,y)}{\pi(y)}$ be the density with respect to π at time $n \ge 0$. It holds for $p \in [1,\infty)$

$$||k_n - 1||_{p,\pi}^p \to 0 \text{ as } n \to \infty\,,$$

where

$$||k_n - 1||_{p,\pi}^p = \sum_{y\in\Omega} |k_n(y) - 1|^p \pi(y)\,. \tag{4}$$

Exercise 6
Prove

$$||\mu - \pi||_{\mathrm{TV}} = \frac{1}{2}\left|\left|\frac{\mu}{\pi} - 1\right|\right|_{1,\pi}\,, \tag{5}$$

$$\mathrm{Var}_\pi\left(\frac{\mu}{\pi}\right) = \left|\left|\frac{\mu}{\pi} - 1\right|\right|_{2,\pi}^2\,. \tag{6}$$

The L^2*-mixing time* is defined as

$$\tau_2(\varepsilon) = \min\left\{n : \forall x \in \Omega \text{ such that} ||k_n^x - 1||_{2,\pi} \le \varepsilon\right\}.$$

The *informal divergence* is

$$\mathtt{D}\big(\mathsf{P}^n(x,\cdot)||\pi\big) = \mathrm{Ent}_\pi(k_n^x) = \sum_{y\in\Omega} \mathsf{P}^n(x,y)\log\frac{\mathsf{P}^n(x,y)}{\pi(y)}\,, \tag{7}$$

and the *mixing time wrt relative entropy* is defined as

$$\tau_{\mathtt{D}}(\varepsilon) = \min\left\{n : \forall x \in \Omega \text{ such that } \mathtt{D}\big(\mathsf{P}^n(x,\cdot)||\pi\big) \le \varepsilon\right\}.$$

Exercise 7
Prove that the informal divergence are convex with respect μ, ν.

Exercise 8
Let d be defined as (1) or (4) or (7). Use the fact that they are convex in order to prove

$$d(\sigma \mathsf{P}^n, \pi) \leq \max_{x\in\Omega} d(\mathsf{P}^n(x,\cdot), \pi)\,.$$

Assume ν is absolutely continuous with respect μ. *Pinsker inequality* bounds the total variation via the informal divergence, i.e.

$$||\nu - \mu||_{\mathrm{TV}}^2 \leq \frac{1}{2}\mathtt{D}(\nu||\mu)\,.$$

We define the *heat kernel*[2] as

$$H_t = e^{t\mathcal{L}}, \qquad \text{where} \qquad \mathcal{L} = -(\mathsf{I} - \mathsf{P})\,.$$

Let $h_t^x(y)$ denote its density, i.e.

$$h_t^x = \frac{H_t(x,y)}{\pi(y)}\ \forall y \in \Omega$$

and also let

$$H_t^* = e^{t\mathcal{L}^*}$$

with semigroup associated to the dual $-\mathcal{L}^* = \mathsf{I} - \mathsf{P}^*$.

Exercise 9
For all h_0 and for all $t \geq 0$ we set $h_t = H_t^* h_0$. Prove

$$\frac{\mathrm{d}}{\mathrm{d}t} h_t(x) = \mathcal{L}^* h_t(x)\,.$$

Exercise 10
For all h_0 and for all $t \geq 0$ we set $h_t = H_t^* h_0$. Prove

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Var}(h_t) = -2\mathcal{E}(h_t, h_t)\,, \tag{8}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Ent}(h_t) = -\mathcal{E}(h_t, \log h_t)\,. \tag{9}$$

Exercise 11
Let $\pi_* = \min_{x\in\Omega}\pi(x)$. Use (5) and (8) and the fact that $\mathrm{Var}(h_0) \leq \frac{1-\pi_*}{\pi_*}$ to prove that in continuous time

$$\tau_2(\varepsilon) \leq \frac{1}{\lambda}\left(\frac{1}{2}\log\frac{1-\pi_*}{\pi_*} + \log\frac{1}{\varepsilon}\right),$$

where λ is the optimal constant of (3).

Let us recall the log-*Sobolev inequality* in the context of Markov kernels. ρ_0 is the optimal constant that satisfies

$$\rho_0 \mathrm{Ent}_\pi(f) \leq \mathcal{E}(f, \log f)\ \forall f : \Omega \to \mathbb{R}_+\,. \tag{10}$$

[2] We are a bit sloppy. For continuous case we should write

$$H_t = e^{-t}\sum_{n\geq 0}\frac{t^n \mathsf{P}^n}{n!}\ \forall t \geq 0.$$

Exercise 12
Let $\pi_* = \min_{x\in\Omega} \pi(x)$. Use the fact that $\mathrm{Ent}(h_0) \le \log \frac{1}{\pi_*}$ to prove that in continuous time

$$\tau_{\mathrm{D}}(\varepsilon) \le \frac{1}{\rho_0}\left(\log\log\frac{1}{\pi_*} + \log\frac{1}{\varepsilon}\right).$$

Optional

Exercise 13
In discrete time, let

$$\begin{aligned} \mathsf{P}^*(x,y) &= \frac{\pi(y)}{\pi(x)}\mathsf{P}(y,x)\,, \\ \pi_* &= \min_{x\in\Omega}\pi(x), \\ ||\mathsf{P}^*|| &= \sup_{f:\Omega\to\mathbb{R},\ \mathbb{E}[f]=0} \frac{||\mathsf{P}^* f||_2}{||f||_2}\,. \end{aligned}$$

Prove

$$\tau_2(\varepsilon) \le \left\lceil \frac{1}{1-||\mathsf{P}^*||}\log\frac{1}{\varepsilon\sqrt{\pi^*}} \right\rceil\,.$$

Exercise 14

1. Given P and a function $f:\Omega\to\mathbb{R}$, prove that
$$\mathrm{Var}(\mathsf{P}^* f) - \mathrm{Var}(f) \le -\mathrm{Var}(f)\lambda_{\mathsf{PP}^*}\,.$$
2. Suppose that P is a discrete time Markov chain. Prove
$$\tau_2(\varepsilon) \le \left\lceil \frac{2}{\lambda_{\mathsf{PP}^*}}\log\frac{1}{\varepsilon\sqrt{\pi^*}} \right\rceil\,.$$

It is preferable to work with P instead of PP^*. Several simplifications make this possible.

Exercise 15
In discrete time, prove that a Markov chain with holding probability α satisfies

$$\tau_2(\varepsilon) \le \left\lceil \frac{1}{\alpha\lambda}\log\frac{1}{\varepsilon\sqrt{\pi^*}} \right\rceil\,.$$

Advanced

Suppose it exist $G:\mathbb{R}_+\to\mathbb{R}_+$ and $f:\Omega\to\mathbb{R}_+$ with $\mathbb{E}[f]=1$ such that

$$\mathcal{E}(f,f) \ge G(\mathrm{Var}(f)).$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Var}(h_t) = -2\mathcal{E}(h_t,h_t) \le -2G(\mathrm{Var}(h_t)).$$

We change variables to $I = \mathrm{Var}(h_t)$, then

$$\frac{\mathrm{d}I}{\mathrm{d}t} \le -2G(I)$$

and

$$\tau_2(\varepsilon) = \int_0^{\tau_2(\varepsilon)} 1 \mathrm{d}t \leq \int_{\mathrm{Var}(h_0)}^{\varepsilon^2} \frac{\mathrm{d}I}{-2G(I)} \,. \tag{11}$$

Exercise 16

Set $G(r) = \lambda r$. Convince yourself that (11) is equivalent to the bounds we derived before.

Exercise 17

If f is non-negative, prove

$$\mathrm{Ent}(f^2) \geq \mathbb{E}f^2 \log \frac{\mathbb{E}f^2}{(\mathbb{E}f)^2} \,.$$

Derive a Poincaré inequality in the case $\mathbb{E}[f] = 1$.

Exercise 18

Given a spectral gap λ and the log-Sobolev constant ρ_0, prove

$$\tau_2(\varepsilon) \leq \frac{1}{2\rho} \log\log \frac{1}{\pi_*} + \frac{1}{\lambda}\left(\frac{1}{4} + \log \frac{1}{\varepsilon}\right) \,.$$

Exercise 19

Given a Nash inequality of the form

$$||f||_2^{2+\frac{1}{D}} \leq C \left(\mathcal{E}(f,f) + \frac{1}{T}||f||_2^2\right) ||f||_1^{\frac{1}{D}} \,, \tag{12}$$

which holds $\forall f : \Omega \to \mathbb{R}$ and some constants $C, D, T \in \mathbb{R}_+$. Prove for all $f \geq 0$ and $\mathbb{E}[f] = 1$

$$\mathcal{E}(f,f) \geq (1 + \mathrm{Var}(f)) \left(\frac{(1 + \mathrm{Var}(f))^{\frac{1}{D}}}{C} - \frac{1}{T}\right) .$$

Exercise 20

Given a Nash inequality of the form (12) with $DC \geq T$ and $D \geq 2$ and a spectral gap λ. Prove

$$\tau_2(\varepsilon) \leq T + \frac{1}{\lambda}\left(\frac{D}{2} \log \frac{DC}{T} + \log \frac{1}{\varepsilon}\right) .$$