Probability and Computation: Problem sheet 5 Solutions

You are encouraged to submit your solutions by emailing them to luca.zanetti@cl.cam.ac.uk by Wednesday 27th of February.

Question 1. *Recall that $\tau(\epsilon)$ refers to the mixing time in total variation distance, while $\tau_2(\epsilon)$ refers to the mixing time in ℓ_2-norm. Prove that,*

$$\tau(\epsilon) \leq \tau_2(2\epsilon).$$

To prove it you might need Jensen's inequality: $\psi(\mathbf{E}[\,X\,]) \leq \mathbf{E}[\,\psi(X)\,]$ for any convex function ψ.

We have also mentioned in class that $\tau_2(2\epsilon) = O(\tau(\epsilon)\log(1/\pi_*))$, where $\pi_* = \min_x \pi(x)$. Can you find an example where this logarithmic factor is needed?

(NOTE: There was originally a typo regarding the constant 2 in front of ϵ in the relation between τ and τ_2)

Solution:

Recall that $\tau(\epsilon) = \min_{t\geq 0} \max_{x\in\Omega} \frac{1}{2}\left\|\frac{P_x^t}{\pi} - 1\right\|_{1,\pi} \leq \epsilon$ and $\tau_2(\epsilon) = \min_{t\geq 0} \max_{x\in\Omega} \frac{1}{2}\left\|\frac{P_x^t}{\pi} - 1\right\|_{2,\pi} \leq \epsilon$, where π is the stationary distribution. We can treat $\frac{P_x^t}{\pi} - 1$ as a random variable whose value is equal to $\frac{P^t(x,y)}{\pi(y)} - 1$ with probability $\pi(y)$. Then, by Jensen's inequality,

$$\left\|P_x^t - \pi\right\|_{1,\pi}^2 = \left(\sum_{y\in\Omega}\left|\frac{P^t(x,y)}{\pi(y)} - 1\right|\pi(y)\right)^2 = \left(\mathbf{E}_{y\sim\pi(y)}\left|\frac{P^t(x,y)}{\pi(y)} - 1\right|\right)^2 \leq \mathbf{E}_{y\sim\pi(y)}\left|\frac{P^t(x,y)}{\pi(y)} - 1\right|^2 = \left\|P_x^t - \pi\right\|_{2,\pi}^2$$

The relation between the two definitions of mixing time easily follows.

An example where the logarithmic factor is needed is the lazy random walk on K_n, the complete graph on n vertices. Consider a lazy random walk starting at vertex 1. After $t \geq 1$ steps, $P^t(x,x) = \frac{1}{2^t} + O(\frac{1}{n})$, while $P^t(x,y) = \Theta\left(\frac{1-(1/2)^t}{n}\right)$ for any $y \in V\setminus\{x\}$. Let's compute the ℓ_1 (total variation) and ℓ_2 distances to the stationary distribution π, which in this case is just the uniform distribution.

$$\left\|P_x^t - \pi\right\|_{TV} = \frac{1}{2}\left\|\frac{P_x^t}{\pi} - 1\right\|_{1,\pi} \leq \frac{1}{2}\left(\frac{n}{2^t} + (n-1)\cdot\frac{1}{2^t}\right)\frac{1}{n} \leq \frac{1}{2^t}$$

$$\left\|\frac{P_x^t}{\pi} - 1\right\|_{2,\pi} \approx \sqrt{\left(\left(\frac{n}{2^t}\right)^2 + (n-1)\left(\frac{1}{2^t}\right)^2\right)\frac{1}{n}} \geq \sqrt{\frac{n}{2^t}}.$$

Therefore, to make the total variation distance less than a small constant, say $1/4$, we just need t to be constant ($t \geq 2$ will suffice). However, for the ℓ_2-norm to be small, we need $t = \Omega(\log n)$.

In general, the total variation distance is "more forgiving" of points with high probability mass, as long as the rest of the distribution is not too far from the stationary distribution. On the contrary, the ℓ_2 distance gives a lot of weight to points that have too high probability mass. Indeed, a lazy random walk on an undirected graph will always have at list $\Omega(\log(n))$ ℓ_2-mixing time, while the ℓ_1-mixing time might be smaller (as in the case of the complete graphs).

Question 2. *The d-dimensional hypercube is an undirected graph $G = (V, E)$ such that V can be represented as the set of binary strings of length d, i.e., $V = \{0,1\}^d$, and $\{x, y\} \in E$ if and only if x and y differ by exactly one bit.*

1. *Compute $|V|$ and $|E|$.*
2. *Prove that G has conductance $O(1/\log(|V|))$.*

Solution:

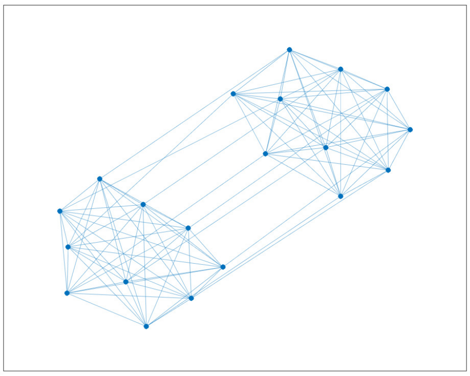
1. $V = 2^d$ and each vertex u has exactly d neighbours. Hence, $|E| = d|V|/2 = d2^{d-1}$.
2. Take $S = \{x \in V : x_1 = 0\}$. Clearly, $|S| = 2^{d-1} = |V|/2$, and $\mathrm{vol}(S) = d2^{d-1}$. Any $u \in S$ has exactly one neighbour in $V \setminus S$: the string of bits which is equal to u in every coordinate apart from the first one. Therefore, $w(S, V \setminus S) = |S| = 2^{d-1}$ and

$$\phi(G) \leq \phi(S) = \frac{w(S, V \setminus S)}{\mathrm{vol}(S)} = \frac{2^{d-1}}{d2^{d-1}} = \frac{1}{d} = \frac{1}{\log |V|}.$$

Question 3. *Let G be a graph of $2n$ vertices that consists in two complete graphs of n vertices, connected by a perfect matching (see picture below). Notice this is a regular graph of degree n and its edge-set can be described as follows:*

$$E = \{\{u,v\} : u, v \in \{1, \dots, n\} \text{ s.t. } u \neq v\} \bigcup \{\{u,v\} : u, v \in \{n+1, \dots, 2n\} \text{ s.t. } u \neq v\}$$
$$\bigcup \{\{u,v\} : v = u + n, \text{ for } u = 1, \dots, n\}.$$

1. *What is the conductance of G?*
2. *Prove that $1 - \lambda_2 = O(1/n)$, which implies that the mixing time of G is at least $\Omega(n)$.*



Solution:

1. Clearly the conductance is minimised on the set $S = \{1, \dots, n\}$:

$$\phi(G) = \phi(S) = \frac{w(S, V \setminus S)}{\mathrm{vol}(S)} = \frac{n}{n^2} = \frac{1}{n}$$

2. By the Cheeger inequality, $1 - \lambda_2 = O(\phi(G)) = O(1/n)$. Alternatively, we can prove this via the variational characterisation of $1-\lambda_2$. Let P the transition matrix of a lay random walk in G. Since G is regular, the stationary distribution π is uniform. Therefore,

$$1 - \lambda_2 = \min_{0 \neq f \perp 1} \sum_{\{u,v\}} \frac{(f(u) - f(v))^2}{2n \sum_{z \in V} f(z)^2}.$$

To obtain an upper bound on $1 - \lambda_2$ (which implies a *lower* bound on mixing), we don't necessarily need to minimise this expression: we need to find a nonzero $f \perp 1$ so that this expression is small enough. A good choice is a function f constant on each of the two complete subgraph, but that takes opposite sign on the two, e.g.,

$$f(u) = \begin{cases} 1 & \text{if } u \in \{1, \ldots, n\} \\ -1 & \text{if } u \in \{n+1, \ldots, 2n\} \end{cases}$$

First, check that f is perpendicular to 1:

$$\langle f, 1\rangle_\pi = \sum_{u\in\{1,\ldots,n\}} f(u) - \sum_{v\in\{n+1,\ldots,2n\}} f(v) = n - n = 0.$$

Then,

$$1 - \lambda_2 \le \sum_{\{u,v\}} \frac{(f(u) - f(v))^2}{2n\sum_{z\in V} f(z)^2} = \frac{4n}{4n^2} = \frac{1}{n},$$

which shows that $1 - \lambda_2 = O(1/n)$. (It turns out this is actually the optimal choice of f)

Question 4. *Use the variational characterisation to show that a cycle on n vertices has spectral gap $1 - \lambda = O(n^{-2})$. To obtain such a result, we suggest you look at the proof for the lower bound of the spectral gap in regular graphs and understand why such a proof gives us an asymptotically tight result for the cycle.*

Solution: Since the n-cycle has degree 2, the variational characterisation can be restated as

$$1 - \lambda_2 = \min_{0\ne f\perp 1} \sum_{\{u,v\}} \frac{(f(u) - f(v))^2}{4\sum_{z\in V} f(z)^2}.$$

As in the solution above, we don't need to find the exact minimum, but just construct a nonzero f that is perpendicular to the all-one function and that gives us a good bound. Let's start labelling the vertices as $V = \{0, \ldots, n-1\}$ where there's an edge between u and v if and only if $u \pm 1 \equiv v \mod n$. To come up with a good guess for f, recall the upper bound on $1 - \lambda_2$ for regular graphs: the main "loss" comes from the inequality

$$(f(u_0) - f(u_1) + f(u_1) - f(u_2) + \cdots + f(u_{\ell-1}) - f(u_\ell))^2 \le \ell \sum_{i=0}^{\ell-1} (f(u_i) - f(u_{i+1}))^2 .$$

This inequality is tight when $(f(u_i) - f(u_{i+1}))^2$ is the same for any i. Then, a good strategy to come up with our function f is to make $|f(i) - f(i+1)|$ constant . For example, let $f(i) = \min\{i, n-i\}$. Assume first that n is odd. We check that $f \perp 1$: $\langle f, 1\rangle_\pi = \frac{1}{n}\sum_{i=0}^{n-1} f(i) = \sum_{i=0}^{(n-1)/2} i + \sum_{i=(n+1)/2}^{n-1}(n-i) = 0$. Moreover,

$$1 - \lambda_2 \le \sum_{\{u,v\}} \frac{(f(u) - f(v))^2}{4\sum_{z\in V} f(z)^2} \le \frac{\sum_{i=0}^{n-1} 1}{\sum_{i=0}^{(n-1)/2} i^2} = O(1/n^2).$$

In the case n is even, f is not exactly perpendicular to 1. We can make it so by considering instead $g = f - \langle f, 1\rangle_\pi 1$ (we are just shifting every coordinate of f by a small amount).

Question 5. *Prove the claim made in Lecture 10: for a regular undirected graph with degree d and diameter δ,*

$$d \cdot \delta = O(n).$$

Solution: Let u, v be two vertices at distance δ from each other. We claim that for any $1 \le i \le \delta - 1$, there must be at least d nodes at distance $i-1, i$, or $i+1$ from u. Consider a shortest path from u to v: $u = u_0, u_1, \ldots, u_\delta = v$. For any $1 \le i \le \delta - 1$, u_i can be connected only with nodes at distance $i-1, i$, or $i+1$ from u (otherwise we could create a "shortcut", contradicting that this is a shortest path). This implies that $\delta \le 3n/d$.

Question 6. *(Bonus) In this exercise you are asked to prove the variational characterisation of the spectral gap. Let P be a self-adjoint transition matrix with stationary distribution π and eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$.*

1. *The first step consists mainly in a few algebraic manipulations. Prove that, for any nonzero $f : \Omega \to \mathbb{R}$,*
$$\langle f, (I-P)f \rangle_\pi = \frac{1}{2} \sum_{x,y\in\Omega} (f(x) - f(y))^2 P(x,y)\pi(x)$$

2. *Now use the spectral theorem to prove that, for any f such that $f \perp 1$,*
$$\langle f, (I-P)f \rangle_\pi \geq (1-\lambda_2)\|f\|_{2,\pi}^2$$

3. *Finally, prove that for an eigenvector f_2 with eigenvalue λ_2 and unit norm,*
$$\langle f_2, (I-P)f_2 \rangle_\pi = 1 - \lambda_2$$
and argue that this proves the variational characterisation.

4. *Modify the proof above to show a variational characterisation for λ_n:*
$$1 - \lambda_n = \max_{f\neq 0} \frac{\sum_{x,y\in\Omega}(f(x) - f(y))^2 P(x,y)\pi(x)}{2\|f\|_{2,\pi}^2}$$

Question 7. *Use the variational characterisations proved in the previous exercise to solve Question 5 of Problem Sheet 4.*

Hints

Q2.2: Look at the "dimension cut" $S = \{x \in V : x_1 = 0\}$.

Q3.2: Construct a function f which is constant on each complete subgraph, but has opposite sign on the two. Use this f to obtain an upper bond on the spectral gap using the variational characterisation.

Q6.2: First deduce from the spectral theorem (more precisely from the corollaries seen in the lecture), that for any $f \perp 1$, $\|f\|_{2,\pi}^2 = \sum_{i=2}^{n} \langle f, f_i \rangle_\pi^2$ (there was a typo here), where $\{1, f_2, \ldots, f_n\}$ is an orthonormal set of eigenvectors for P. You also need to use the linearity of the inner product.