Lecture 5 - L^p distances and the Spectral Decomposition

Wednesday, August 25

In general it is difficult to determine the spectral decomposition, and even when possible it is non-trivial to write the t-step transition probabilities and turn this into a bound on variation distance. In this lecture we will focus first on relating the spectral decomposition to a quantitative bound on mixing time, and next on showing that the second largest (in magnitude) eigenvalue suffices.

First, consider a heuristic. Suppose that P can be diagonalized with eigenvalues λ_i and eigenbases $\overrightarrow{v_i}$ for $i = 1 \ldots n$, where $n = |\Omega|$. Then

$$\mathsf{p}^{(0)} = \sum_{k=1}^{n} \frac{< \mathsf{p}^{(0)}, \overrightarrow{v_i} >}{< \overrightarrow{v_i}, \overrightarrow{v_i} >} \overrightarrow{v_i}$$

for the standard scalar product $< \overrightarrow{a}, \overrightarrow{b} >= a \cdot b$. Then

$$\begin{aligned} \mathsf{p}^{(t)} &= \sum_{k=1}^{n} \frac{< \mathsf{p}^{(0)}, \overrightarrow{v_i} >}{< \overrightarrow{v_i}, \overrightarrow{v_i} >} \lambda^t \, \overrightarrow{v_i} \\ &\xrightarrow{t\to\infty} \frac{< \mathsf{p}^{(0)}, \overrightarrow{v_1} >}{< \overrightarrow{v_1}, \overrightarrow{v_1} >} \lambda_1^t \overrightarrow{v_1} + \frac{< \mathsf{p}^{(0)}, \overrightarrow{v_2} >}{< \overrightarrow{v_2}, \overrightarrow{v_2} >} \lambda_2^t \, \overrightarrow{v_2} \\ &= \pi + \frac{< \mathsf{p}^{(0)}, \overrightarrow{v_2} >}{< \overrightarrow{v_2}, \overrightarrow{v_2} >} \lambda_2^t \, \overrightarrow{v_2} \end{aligned}$$

if λ_2 is the second largest (in magnitude) eigenvalue. When $|\lambda_n| > \lambda_2$ then a similar bound holds with $\overrightarrow{v_2}$ replaced by $\overrightarrow{v_n}$ instead. In short, the second largest (in magnitude) eigenvalue will govern the mixing time.

In order to prove this we need a few preliminary steps.

Theorem 5.1. *If* P *is reversible and ergodic then it has a spectral decomposition with eigenvalues* $1 = \lambda_1 > \lambda_2 \geq \cdots \geq \lambda_n \geq -1$ *and moreover*

$$\left(\mathsf{P}^t\right)_{ij} = \sqrt{\frac{\pi(j)}{\pi(i)}} \sum_{m=1}^{n} \lambda_m^t \, (\overrightarrow{u_m})_i \, (\overrightarrow{u_m})_j$$

where u_m *is the* m*-th left eigenvector of the matrix* $S = D\,\mathsf{P}\,D^{-1}$ *with* $D = diag(\sqrt{\pi(1)}, \sqrt{\pi(2)}, \ldots, \sqrt{\pi(n)})$.

Proof. Our argument is a simplification of one found in Chapter 3, Section 4 of Aldous and Fill's book [1].

Observe from the definition that $S_{ij} = \pi(i)^{1/2}\,\mathsf{P}_{ij}\,\pi(j)^{-1/2}$. It is easily verified that S is symmetric if P is reversible (check this yourself). Then S is a symmetric real matrix and so the Spectral theorem says that S ha a spectral decomposition

$$S = U\,\Lambda\,U^T$$

where the columns of U are the orthonormal left eigenvectors, and $\Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ where λ_i are the eigenvalues of S.

Then the λ_i are eigenvalues of P as well, with left eigenvectors $(\overrightarrow{v_m})_i = \sqrt{\pi(i)}\,(\overrightarrow{u_m})_i$, i.e. $\sum_{i=1}^n (\overrightarrow{v_m})_i\,\mathsf{P}_{ij} = \lambda_m\,(\overrightarrow{v_m})_j$. (check this yourself)

Therefore,

$$\mathsf{P}^t = D^{-1}\,S^t\,D = D^{-1}\,U\,\Lambda^t\,U^T\,D$$

and the theorem follows by computing the elements $\left(\mathsf{P}^t\right)_{ij}$ in this expansion. $\square$

Exercise 1. Verify that this works on the uniform two-point example worked out earlier.

The upper bound on total variation distance will be done by bounding a different distance, which in turn upper bounds total variation distance. The L_p-norm of a function f is given by

$$\|f\|_{p,\pi} = \left[\sum_{x\in\Omega} |f(x)|^p \,\pi(x)\right]^{1/p} .$$

The total variation distance is then

$$\|\mu - \pi\|_{TV} = \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \pi(x)| = \frac{1}{2}\sum_{x\in\Omega} \left|1 - \frac{\mu(x)}{\pi(x)}\right| \pi(x) = \frac{1}{2}\left\|1 - \frac{\mu}{\pi}\right\|_{1,\pi} .$$

Cauchy-Schwartz shows that $\|f\|_{p,\pi} \leq \|f\|_{q,\pi}$ if $q \geq p$, so in particular

$$\boxed{\|\mu - \pi\|_{TV} \leq \frac{1}{2}\left\|1 - \frac{\mu}{\pi}\right\|_{2,\pi}}$$

Finally, we can bound the variation distance.

Theorem 5.2. *Given a reversible ergodic Markov chain then*

$$4\,\|\mathsf{P}^t(x,\cdot) - \pi\|_{TV}^2 \leq \left\|1 - \frac{\mathsf{P}^t(x,\cdot)}{\pi(x)}\right\|_{2,\pi}^2 = \frac{\mathsf{P}^{2t}(x,x)}{\pi(x)} - 1 = \frac{\sum_{m=1}^n \lambda_m^{2t}\,(\overrightarrow{u_m})_x^2}{\pi(x)} - 1 \leq \frac{1-\pi(x)}{\pi(x)}\,\lambda_{max}^{2t}$$

where $\lambda_{max} = \max\{\lambda_2,\, |\lambda_n|\}$.

Proof. From the definition it follows that

$$\left\|1 - \frac{\mathsf{P}^t(x,\cdot)}{\pi}\right\|_{2,\pi}^2 = \sum_{y\in\Omega} \pi(y) - 2\,\mathsf{P}^t(x,\cdot) + \frac{\mathsf{P}^t(x,y)^2}{\pi(y)} = \sum_{y\in\Omega} \frac{\mathsf{P}^t(x,y)}{\pi(y)} - 1\,.$$

To simplify further observe that

$$\frac{\mathsf{P}^t(x,y)^2}{\pi(y)} = \frac{\mathsf{P}^t(x,y)\,\pi(x)\mathsf{P}^t(x,y)}{\pi(x)\pi(y)} = \frac{\mathsf{P}^t(x,y)\,\pi(y)\mathsf{P}^t(x,y)}{\pi(x)\pi(y)} = \frac{\mathsf{P}^t(x,y)\mathsf{P}^t(y,x)}{\pi(x)}$$

where the second equality followed from $\pi(x)\,\mathsf{P}^t(x,y) = \pi(y)\,\mathsf{P}^t(y,x)$ (use reversibility and induction).

Therefore,

$$\left\|1 - \frac{\mathsf{P}^t(x,\cdot)}{\pi}\right\|_{2,\pi}^2 = \sum_{y\in\Omega} \frac{\mathsf{P}^t(x,y)\mathsf{P}^t(y,x)}{\pi(x)} - 1 = \frac{\mathsf{P}^{2t}(x,x)}{\pi(x)} - 1$$

completing the first equality.

For the second equality apply Theorem 5.1.

For the final inequality we simplify the eigenvalue bounds.

$$\sum_{m=1}^n \lambda_m^{2t}\,(\overrightarrow{u_m})_x^2 = (\overrightarrow{u_1})_x^2 + \sum_{m=2}^n \lambda_m^{2t}\,(\overrightarrow{u_m})_x^2 \leq (\overrightarrow{u_1})_x^2 + \lambda^{2t}\left(-(\overrightarrow{u_1})_x^2 + \sum_{m=1}^n (\overrightarrow{u_m})_x^2\right)$$

To finish, recall that $(\overrightarrow{u_m}) \propto \frac{(\overrightarrow{v_m})}{\sqrt{\pi}}$. It follows that $(\overrightarrow{u_1}) \propto \frac{\pi}{\sqrt{\pi}} = \sqrt{\pi}$. In fact, $\sqrt{\pi}$ is a unit vector, so $(\overrightarrow{u_1}) = \sqrt{\pi}$ and in particular $(\overrightarrow{u_1})_x^2 = \pi(x)$. For the term in the sum, recall that $U = ((\overrightarrow{u_1}), (\overrightarrow{u_2}), \ldots, (\overrightarrow{u_n}))$ is an orthonormal matrix. It follows that $U^{-1} = U^T$ and so $U\,U^T = I$. The matrix $U_{ij} = (\overrightarrow{u_j})_i$ and so it follows that the dot product of row i with row j is 1 if $i = j$ and 0 otherwise, in particular $\sum_{m=1}^{n} (\overrightarrow{u_m})_x^2 = 1$. Substitution into the equation above finishes the proof. □

Remark 5.3. *A similar argument shows that*

$$\left(\mathsf{P}^t\right)_{ij} = \sqrt{\frac{\pi_j}{\pi_i}} \sum_{m=1}^{n} \lambda_m^t \, (\overrightarrow{u_m})_i \, (\overrightarrow{u_m})_j \xrightarrow{t\to\infty} \sqrt{\frac{\pi_j}{\pi_i}} \left(\sqrt{\pi_i \pi_j} + \lambda_{max}^t\right) = \pi_j + \sqrt{\frac{\pi_j}{\pi_i}} \lambda_{max}^t$$

Therefore λ also can be used to find the rate at which the t-step distribution approaches π.

Exercise 2. Compare the total variation bound for the uniform two-point space to the L^2 bound and the bound derived from λ_{max}.

References

[1] D. Aldous and J. Fill. Reversible markov chains and random walks on graphs (book to appear). *URL for draft http://www.stat.Berkeley.edu/users/aldous.*