Spectral Graph Theory

Lorenzo Orecchia, Erasmo Tani

March 8, 2019

Preface

Chapter 1

First chapter

Chapter 2

1

'Random Walks and Averaging Processes Last time we discussed two different processes related to the adjacency matrix of a graph. The first process is the **random walk process** described by the random walk operator $W = AD^{-1}$ which is the transition matrix of the natural random walk over the vertices of G. The second process is the **averaging process** by means of which the vertices of the graph, assigned an initial belief value, try to reach a global consensus by repeating a local step. This second process corresponds to the operator $\frac{1}{2}(I + W^T) = \frac{1}{2}(I + D^{-1}A)$

In the random walk process we work with probability vectors $p \in \mathbb{R}^V$ representing the probability of being at a particular vertex at a specific time. In the averaging process we work with x vectors that represent the amount of mass on the arcs (/ edges). Recall that these two representations are related by:

$$\boxed{p = Dx} \tag{2.1}$$

The random walk process converges to the standard random walk stationary distribution π where $\pi \in \Delta_n$ and for any vertex $i \in V$: $\pi_i \sim d_i$ where d_i denotes the degree of i in G. On the other hand, the second process converges to:

$$\bar{x} = \frac{\sum_i d_i x_i}{vol(G)} \vec{1}$$

One can check that the averaging process does, in fact, correspond to having each vertex compute the average of its neighbours:

$$\left(\frac{1}{2}(I + W^T)x\right)_i = \frac{1}{2}x_i + \frac{1}{d_i}\sum_{j \sim i} x_j.$$

In particular, the action of W^T corresponds to the averaging the value of every neighbour (Figure 2.1), and $\frac{1}{2}(I + W^T)$ simply averages the result with the current value.

Figure 2.1: The action of W^T on the vector $x = (0.1, 0.6, 0.2, 0.1)$.

2.1 Hilbert Space Interpretation

One can think of the x vectors as living in the Hilbert space with inner product $\langle \cdot, \cdot \rangle_D : \mathbb{R}^V \to \mathbb{R}$ given by:

$$\langle x, y \rangle_D = x^T D y$$

And similarly the p vectors live in the dual space with inner product: $\langle \cdot, \cdot \rangle_{D^{-1}} : \mathbb{R}^V \to \mathbb{R}$ given by:

$$\langle p, q \rangle_{D^{-1}} = p^T D^{-1} q$$

for any x the corresponding vector p will equal $\Phi(x)$ where Φ is the isomorphism guaranteed by the Riesz representation theorem*.

Figure 2.2: Vertex probability vectors p live in the dual Hilbert space to that of edge mass vectors x.

2.2 Measuring Convergence to the Fixed Point

An important choice to make when tracking the dynamics of these processes, is the choice of potential function: how do we measure distance from the fixed point?

In the **averaging process** it makes sense to track the distance of a mass assignment x from the optimum by:

$$\phi_{avg}(x) := \sum_{i \in V} \pi_i \left(x_i - (\sum_{j \in V} \pi_j x_j) \right)^2 = Var_{i \sim \pi}(x_i)$$

By applying the conversion in (2.1) we see that this corresponds to:

*Read more about it: https://en.wikipedia.org/wiki/Riesz_representation_theorem

$$\phi_{avg}(x) = ||x - \bar{x}||_\pi^2 = ||D^{-1}p - \bar{x}D^{-1}\vec{1}||_D^2 = ||p - \bar{x}\vec{1}||_{D^{-1}}^2 = ||p - \pi||_{D^{-1}}^2$$

which corresponds naturally to a way to measure distance from the optimum in the **random walk** process.

2.3 Why Laplacians?

The above analysis can provide motivation for the use of Laplacian matrices to study graph algorithms. One can in fact prove the following claim:

Claim 2.3.1. *The potential function defined above is the quadratic form of the Laplacian of a suitably defined graph. I.e there exists some graph H such that, for all $x \in \mathbb{R}^V$: $\phi_{avg}(x) = x^T L(H)x$.*

Proof. One can define a graph K_G as follows: $V(K_G) = V(G) = V$ and $E(K_G) = \binom{V}{2}$, so that the graph is complete, with weigths $w_{K_G}(ij) = \frac{d_i d_j}{vol(G)}$ *

□

We also observe that the value of the quadratic form $x^T L(G)x$ defined by the Laplacian matrix of G measures the rate of convergence towards the stationary distribution at x. In fact, consider the continuous version of the averaging process given by:

$$\frac{dx(t)}{dt} = -(I - W^T)x(t)$$

We have:

$$\begin{aligned}
\frac{d}{dt}\left(||x - \bar{x}||_D^2\right) &= \left(\frac{d}{dt}x(t)\right)^T D(x(t) - \bar{x}) \\
&= \left(-(I - W^T)x(t)\right)^T D(x(t) - \bar{x}) \\
&= -x(t)(D - A)(x - \bar{x}) = -x^T(D - A)x
\end{aligned}$$

Where the last step above uses the fact that:

$$L\vec{1} = 0.$$

*Recall that for any $S \subset V$: $vol(S) := \sum_{i \in S} d_i$ and $vol(G) := vol(V)$.

Chapter 3

Positive Semidefinite Matrices

In this chapter we will discuss positive semidefinite (PSD) matrices, a class of operators that plays and essential role both in spectral graph theory and in optimization as a whole.

Definition 3.0.1 (PD/PSD Matrix)**.** A symmetric $n \times n$ matrix M is said to be *positive semidefinite* (PSD) if, for all vectors $x \in \mathbb{R}^n$ we have:

$$x^T M x \geq 0$$

we denote this by $M \succeq 0$. Similarly, a symmetric $n \times n$ matrix M is said to be *positive definite* (PD) if, for all $x \in \mathbb{R}^n \setminus \{0\}$:

$$x^T M x > 0$$

we denote this fact as $M \succ 0$.

From the spectral theorem for symmetric matrices we get the following useful characterization of PSDness:

Theorem 3.0.2. *A symmetric $n \times n$ matrix M is PSD if and only if all of its eigenvalues are non-negative, and PD if they are strictly positive.*

Proof. Every eigenvalue of M arises as the value of $x^T M x$ for some $x \in \mathbb{R}^n$ (namely the corresponding normalized eigenvector) and hence if M is PSD(/PD) then all eigenvalues must be non-negative(/strictly positive). On the other hand suppose that all the eigenvalues of M are non-negative ($0 \leq \lambda_1 \leq ... \leq \lambda_n$) then we can write:

$$M = \sum_{i=1}^{n} \lambda_n v_i v_i^T$$

where $v_1, ..., v_n$ is an orthonormal basis of eigenvectors of M. Hence, for every $x \in \mathbb{R}^n$:

$$x^T M x = \sum_{i=1}^{n} \lambda_n \langle v_i, x \rangle^2 \geq 0$$

(similarly for PD matrices we get > 0). $\square$

We also state the following useful result without a proof:

Theorem 3.0.3 (Sylvester's Criterion)**.** *A matrix is PD if and only if all its leading principal minors are positive, and PSD if and only if all it's principal minors are non-negative.*

We now prove a characterization of PSDness for matrices * :

Proposition 3.0.4 (Characterization of PSD Matrices)**.** *A square matrix $M \in \mathbb{R}^{n \times n}$ is PSD (and hence symmetric) if and only if there exists vectors $x^{(1)}, ..., x^{(n)} \in \mathbb{R}^n$ such that $M_{i,j} = \langle x^{(i)}, x^{(j)} \rangle$.*

Proof. Suppose that $x^{(1)}, ..., x^{(n)} \in \mathbb{R}^n$ exist as above, then clearly the matrix M is symmetric and furthermore:

$$y^T M y = \sum_{i,j} y_i M_{ij} y_j = \left\langle \sum_i y_i x^{(i)}, \sum_j y_j x^{(j)} \right\rangle = \left\| \sum_i y_i x^{(i)} \right\|^2 \geq 0$$

so M is PSD. Conversely we can use the spectral decomposition of M to see that:

$$M = \sum_{k=1}^{n} \lambda_k v^{(k)} v^{(k)^T}$$

giving that:

$$M_{ij} = \sum_{k=1}^{n} \lambda_k v_i^{(k)} v_j^{(k)}$$

so that we can set:

$$x_k^{(i)} = \sqrt{\lambda_k} v_i^{(k)}$$

and this satisfies the desired properties. $\square$

*from Luca Trevisan's notes

Note the the above proposition goes hand in hand with the fact that any PSD matrix can be decomposed as $M = B^T B$. One example of such factorization is the Cholesky Decomposition / Cholesky Factorization, where M is written as LL^* for some lower-triangular matrix L. This exists for every hermitian PSD matrix, and when the matrix M only has real entries, the matrix L can also be chosen to only have real entry and the factorization becomes $M = LL^T$. If M is PD then L the Cholesky Factorization is unique, this is not necessarily true in the PSD case.

Proposition 3.0.5 (PSD Matrices Form a Cone)**.** *Let A, B be $n \times n$ PSD matrices, we then have:*

1. $A + B \succeq 0$,
2. *if* $\alpha \geq 0$ *then* $\alpha A \succeq 0$,
3. *if* $\alpha, \beta \geq 0$ *then* $\alpha A + \beta B \succeq 0$.

In particular, this means that the set of PSD matrices forms a cone.

Proof. We prove each of the above individually:

1. For any $x \in \mathbb{R}^n$:
$$x^T(A + B)x = x^T Ax + x^T Bx \geq 0.$$
2. For any $x \in \mathbb{R}^n$:
$$x^T(\alpha A)x = \alpha x^T Ax \geq 0.$$
3. Follows from the previous two parts.

$\square$

We now introduce the matrix inner product:

Definition 3.0.6 (Matrix Inner Product)**.** Given any two $n \times n$ matrices A, B their (dot) inner product is the value:

$$A \bullet B := tr(A^T B) = \sum_{i,j} A_{i,j} B_{i,j}.$$

we now go on to prove that the one define above is, indeed, an inner product:

Proposition 3.0.7. *The function $\bullet : \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n} \to \mathbb{R}$ defined above is a real inner product, i.e. satisfies:*

Symmetry *For any A, B we have $A \bullet B = B \bullet A$,*

Linearity *For any A, B, C and $\alpha, \beta \in \mathbb{R}^n$ we have $(\alpha A + \beta B) \bullet C = \alpha(A \bullet C) + \beta(B \bullet C)$,*

Positive Definiteness *$A \bullet A \geq 0$ and $A \bullet A = 0 \Longleftrightarrow A = 0$.*

Proof. The inner product is clearly symmetric ($A^T B = (B^T A)^T$ so $A^T B$ and $B^T A$ have the same trace). For linearity one checks:

$$(\alpha A + \beta B) \bullet C = \sum_{i,j} (\alpha A_{i,j} + \beta B_{i,j}) C_{i,j} = \alpha \sum_{i,j} A_{i,j} C_{i,j} + \beta \sum_{i,j} B_{i,j} C_{i,j} \tag{3.1}$$

$$= \alpha(A \bullet C) + \beta(B \bullet C) \tag{3.2}$$

Positive definitiness is also clear since $A \bullet A$ is a sum of squares. □

Proposition 3.0.8 (Properties of the Matrix Inner Product)**.** *Let A and B be symmetric matrices such that $A, B \succeq 0$ then $A \bullet B \geq 0$.*

Proof. Using the spectral expansion of A:

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^T$$

we have:

$$A \bullet B = tr\left(\sum_{i=1}^{n} \lambda_i v_i v_i^T B\right) = \sum_{i=1}^{n} \lambda_i tr\left(v_i v_i^T B\right) = \sum_{i=1}^{n} \lambda_i tr\left(v_i v_i^T B\right) = \sum_{i=1}^{n} \lambda_i tr\left(v_i^T B v_i\right) \geq 0$$

□

Appendix A

First Appendix

Last note