Spectral Graph Theory Lecture 23

Quadrature for the Finite Free Convolution

Daniel A. Spielman November 30, 2015

Disclaimer

These notes are not necessarily an accurate representation of what happened in class. The notes written before class say what I think I should say. I sometimes edit the notes after class to make them way what I wish I had said.

There may be small mistakes, so I recommend that you check any mathematically precise statement before using it in your own work.

These notes were last revised on December 1, 2015.

23.1 Overview

The material in today's lecture comes from [**?**] and [**?**]. My goal today is to prove simple analogs of the main quadrature results, and then give some indication of how the other quadrature statements are proved. I will also try to explain what led us to believe that these results should be true.

Recall that last lecture we considered the expected characteristic polynomial of a random matrix of the form $\boldsymbol{A} + \Pi \boldsymbol{B} \Pi^T$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric. We do not know a nice expression for this expected polynomial for general $\boldsymbol{A}$ and $\boldsymbol{B}$. However, we will see that there is a very nice expression when $\boldsymbol{A}$ and $\boldsymbol{B}$ are Laplacian matrices or the adjacency matrices of regular graphs.

23.2 The Finite Free Convolution

In Free Probability [**?**], one studies operations on matrices in a large dimensional limit. These matrices are determined by the moments of their spectrum, and thus the operations are independent of the eigenvectors of the matrices. We consider a finite dimensional analog.

For n-dimensional symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, we consider the expected characteristic polynomial

$$\mathbb{E}_{\boldsymbol{Q} \in \mathcal{O}(n)} \chi_x(\boldsymbol{A} + \boldsymbol{Q}\boldsymbol{B}\boldsymbol{Q}^T),$$

where $\mathcal{O}(n)$ is the group of n-by-n orthonormal matrices, and $\boldsymbol{Q}$ is a random orthonormal matrix chosen according to the Haar measure. In case you are not familiar with "Haar measure", I'll quickly explain the idea. It captures our most natural idea of a random orthnormal matrix. For example, if $\boldsymbol{A}$ is a Gaussian random symmetric matrix, and $\boldsymbol{V}$ is its matrix of eigenvectors, then $\boldsymbol{V}$

is a random orthonormal matrix chosen according to Haar measure. Formally, it is the measure that is invariant under group operations, which in this case are multiplication by orthnormal matrices. That is, the Haar measure is the measure under which for every $S \subseteq \mathcal{O}(n)$ and $\boldsymbol{P} \in \mathcal{O}(n)$, S has the same measure as $\{\boldsymbol{Q}\boldsymbol{P} : \boldsymbol{Q} \in S\}$.

This expected characteristic polynomial does not depend on the eigenvectors of $\boldsymbol{A}$ and $\boldsymbol{B}$, and thus can be written as a function of the characteristic polynomials of these matrices. To see this, write $\boldsymbol{A} = \boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^T$ and $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{C}\boldsymbol{U}^T$ where $\boldsymbol{U}$ and $\boldsymbol{V}$ are the orthnormal eigenvectors matrices and $\boldsymbol{C}$ and $\boldsymbol{D}$ are the diagonal matrices of eigenvalues. We have

$$\chi_x(\boldsymbol{V}\boldsymbol{D}\boldsymbol{V}^T + \boldsymbol{Q}\boldsymbol{U}\boldsymbol{C}\boldsymbol{U}^T\boldsymbol{Q}^T) = \chi_x(\boldsymbol{D} + \boldsymbol{V}^T\boldsymbol{Q}\boldsymbol{U}\boldsymbol{C}\boldsymbol{U}^T\boldsymbol{Q}^T\boldsymbol{V}) = \chi_x(\boldsymbol{D} + (\boldsymbol{V}^T\boldsymbol{Q}\boldsymbol{U})\boldsymbol{C}(\boldsymbol{V}^T\boldsymbol{Q}\boldsymbol{U})^T).$$

If $\boldsymbol{Q}$ is distributed according to the Haar measure on $\mathcal{O}(n)$, then so is $\boldsymbol{V}^T\boldsymbol{Q}\boldsymbol{U}$.

If $p(x)$ and $q(x)$ are the characteristic polynomials of $\boldsymbol{A}$ and $\boldsymbol{B}$, then we define their *finite free convolution* to be the polynomial

$$p(x) \boxplus_n q(x) \stackrel{\text{def}}{=} \mathbb{E}_{\boldsymbol{Q}\in\mathcal{O}(n)}\chi_x(\boldsymbol{A} + \boldsymbol{Q}\boldsymbol{B}\boldsymbol{Q}^T).$$

In today's lecture, we will establish the following formula for the finite free convolution.

Theorem 23.2.1. *Let*

$$p(x) = \sum_{i=0}^{n} x^{n-i}(-1)^i a_i \quad \text{and} \quad q(x) = \sum_{i=0}^{n} x^{n-i}(-1)^i b_i.$$

Then,

$$p(x) \boxplus_n q(x) = \sum_{k=0}^{n} x^{n-k}(-1)^k \sum_{i+j=k} \frac{(n-i)!(n-j)!}{n!(n-i-j)!} a_i b_j. \tag{23.1}$$

This convolution was studied by Walsh [**?**], who proved that when p and q are real rooted, so is $p \boxplus_n q$.

Our interest in the finite free convolution comes from the following theorem, whose proof we will also sketch today.

Theorem 23.2.2. *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be symmetric matrices with constant row sums. If $\boldsymbol{A}\mathbf{1} = a\mathbf{1}$ and $\boldsymbol{B}\mathbf{1} = b\mathbf{1}$, we may write their characteristic polynomials as*

$$\chi_x(\boldsymbol{A}) = (x-a)p(x) \quad \text{and} \quad \chi_x(\boldsymbol{B}) = (x-b)q(x).$$

We then have

$$\mathbb{E}_{\Pi\in S_n}\chi_x(\boldsymbol{A} + \Pi\boldsymbol{B}\Pi^T) = (x-(a+b))(p(x) \boxplus_{n-1} q(x)).$$

We know that $\mathbf{1}$ is an eigenvector of eigenvalue $a+b$ of $\boldsymbol{A} + \Pi\boldsymbol{B}\Pi^T$ for every permutation matrix Π. Once we work orthogonal to this vector, we discover the finite free convolution.

We describe this theorem as a *quadrature* result, because it obtains an integral over a continuous space as a sum over a finite number of points.

Before going in to the proof of the theorem, I would like to explain why one might think something like this could be true. The first answer is that it was a lucky guess. We hoped that this expectation would have a nice formula. The nicest possible formula would be a bi-linear map: a function that is linear in p when q is held fixed, and vice versa. So, we computed some examples by holding $\boldsymbol{B}$ and q fixed and varying $\boldsymbol{A}$. We then observed that the coefficients of the resulting expected polynomial are in fact a linear functions of the coefficients of p. Once we knew this, it didn't take too much work to guess the formula.

I now describe the main quadrature result we will prove today. Let $\mathcal{B}(n)$ be the nth hyperoctahedral group. This is the group of symmetries of the generalized octahedron in n dimensions. It may be described as the set of matrices that can be written in the form $\boldsymbol{D}\Pi$, where $\boldsymbol{D}$ is a diagonal matrix of ± 1 entries and Π is a permutation. It looks like the family of permutation matrices, except that both 1 and -1 are allowed as nonzero entries. $\mathcal{B}(n)$ is a subgroup of $\mathcal{O}(n)$.

Theorem 23.2.3. *For all symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$,*

$$\mathbb{E}_{\boldsymbol{Q}\in\mathcal{O}(n)}\chi_x(\boldsymbol{A}+\boldsymbol{Q}\boldsymbol{B}\boldsymbol{Q}^T)=\mathbb{E}_{\boldsymbol{P}\in\mathcal{B}(n)}\chi_x(\boldsymbol{A}+\boldsymbol{P}\boldsymbol{B}\boldsymbol{P}^T).$$

We will use this result to prove Theorem 23.2.1. The proof of Theorem 23.2.2 is similar to the proof of Theorem 23.2.3. So, we will prove Theorem 23.2.3 and then explain the major differences.

23.3 Quadrature

In general, quadrature formulas allow one to evaluate integrals of a family of functions over a fixed continuous domain by summing the values of those functions at a fixed number of points. There is an intimate connection between families of orthogonal polynomials and quadrature formulae that we unfortunately do not have time to discuss.

The best known quadrature formula allows us to evalue the integral of a polynomial around the unit circle in the complex plane. For a polynomial $p(x)$ of degree less than n,

$$\int_{\theta=0}^{2\pi}p(e^{i\theta})d\theta=\frac{1}{n}\sum_{k=0}^{n-1}p(\omega^k),$$

where $\omega=e^{2\pi i/n}$ is a primitive nth root of unity.

We may prove this result by establishing it separately for each monomial. For $p(x)=x^k$ with $k\neq 0$,

$$\int_{\theta=0}^{2\pi}p(e^{i\theta})d\theta=\int_{\theta=0}^{2\pi}e^{i\theta k}d\theta=0.$$

And, for $|k|<n$, the corresponding sum is the sum of nth roots of unity distributed symmetrically about the unit circle. So,

$$\sum_{j=0}^{n-1}\omega^{jk}=0.$$

We used this fact in the start of the semester when we computed the eigenvectors of the ring graph and observed that all but the dominant are orthogonal to the all-1s vector.

On the other hand, for $p(x) = 1$ both the integral and sum are 1.

We will use an alternative approach to quadrature on groups, encapsulted by the following lemma.

Lemma 23.3.1. *For every n and function $p(x) = \sum_{|k|<n} c_k x^k$, and every $\theta \in [0, 2\pi]$,*

$$\sum_{j=0}^{n} p(e^{i(2\pi j/n+\theta)}) = \sum_{j=0}^{n} p(e^{i(2\pi j/n)}).$$

This identity implies the quadrature formula above, and has the advantage that it can be experimentally confirmed by evaluating both sums for a random θ.

Proof. We again evaluate the sums monomial-by-monomial. For $p(x) = x^k$, with $|k| < n$, we have

$$\sum_{j=0}^{n} (e^{i(2\pi j/n+\theta)})^k = e^{i\theta k} \sum_{j=0}^{n} (e^{i(2\pi j/n)})^k.$$

For $k \neq 0$, the latter sum is zero. For $k = 0$, $e^{i\theta k} = 1$. $\square$

23.4 Quadrature by Invariance

For symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, define the function

$$f_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{Q}) = \det(\boldsymbol{A} + \boldsymbol{Q}\boldsymbol{B}\boldsymbol{Q}^T).$$

We will derive Theorem 23.2.3 from the following theorem.

Theorem 23.4.1. *For all $\boldsymbol{Q} \in \mathcal{O}(n)$,*

$$\mathbb{E}_{\boldsymbol{P}\in\mathcal{B}(n)} f(\boldsymbol{P}) = \mathbb{E}_{\boldsymbol{P}\in\mathcal{B}(n)} f(\boldsymbol{Q}\boldsymbol{P}).$$

Proof of Theorem 23.2.3. First, observe that it suffices to consider determinants. For every $\boldsymbol{P} \in \mathcal{B}(n)$, we have

$$\int_{\boldsymbol{Q}\in\mathcal{O}(n)} \det(\boldsymbol{A} + \boldsymbol{Q}\boldsymbol{B}\boldsymbol{Q}^T) = \int_{\boldsymbol{Q}\in\mathcal{O}(n)} f(\boldsymbol{Q}) = \int_{\boldsymbol{Q}\in\mathcal{O}(n)} f(\boldsymbol{Q}\boldsymbol{P}).$$

So,

$$\mathbb{E}_{\boldsymbol{P}\in\mathcal{B}(n)} \int_{\boldsymbol{Q}\in\mathcal{O}(n)} f(\boldsymbol{Q}\boldsymbol{P}) = \int_{\boldsymbol{Q}\in\mathcal{O}(n)} f(\boldsymbol{Q}).$$

On the other hand, as $\mathcal{B}(n)$ is discrete we can reverse the order of integration to obtain

$$\int_{\boldsymbol{Q}\in\mathcal{O}(n)} f(\boldsymbol{Q}) = \int_{\boldsymbol{Q}\in\mathcal{O}(n)} \mathbb{E}_{\boldsymbol{P}\in\mathcal{B}(n)} f(\boldsymbol{Q}\boldsymbol{P}) = \int_{\boldsymbol{Q}\in\mathcal{O}(n)} \mathbb{E}_{\boldsymbol{P}\in\mathcal{B}(n)} f(\boldsymbol{P}) = \mathbb{E}_{\boldsymbol{P}\in\mathcal{B}(n)} f(\boldsymbol{P}),$$

where the second equality follows from Theorem 23.4.1. $\square$

23.5 Structure of the Orthogonal Group

To prove Theorem 23.4.1, we need to know a little more about the orthogonal group. We divide the orthonormal matrices into two types, those of determinant 1 and those of determinant -1. The orthonormal matrices of determinant 1 form the *special orthogonal group*, $\mathcal{SO}(n)$, and every matrix in $\mathcal{O}(n)$ may be written in the form $\boldsymbol{D}\boldsymbol{Q}$ where $\boldsymbol{Q} \in \mathcal{SO}(n)$ and $\boldsymbol{D}$ is a diagonal matrix in which the first entry is ± 1 and all others are 1. Every matrix in $\mathcal{SO}(n)$ may be expressed as a product of 2-by-2 rotation matrices. That is, for every $\boldsymbol{Q} \in \mathcal{SO}(n)$ there are matrices $\boldsymbol{Q}_{i,j}$ for $1 \leq i < j \leq n$ so that $\boldsymbol{Q}_{i,j}$ is a rotation in the span of $\boldsymbol{\delta}_i$ and $\boldsymbol{\delta}_j$ and so that

$$\boldsymbol{Q} = \boldsymbol{Q}_{1,2}\boldsymbol{Q}_{1,3}\cdots\boldsymbol{Q}_{1,n}\boldsymbol{Q}_{2,3}\cdots\boldsymbol{Q}_{2,n}\cdots\boldsymbol{Q}_{n-1,n}.$$

If you learned the QR-factorization of a matrix, then you learned an algorithm for computing this decomposition.

These facts about the structure of $\mathcal{O}(n)$ tell us that it suffices to prove Theorem 23.4.1 for the special cases in which $\boldsymbol{Q} = \mathsf{diag}(-1, 1, 1, \ldots, 1)$ and when $\boldsymbol{Q}$ is rotation of the plane spanned by $\boldsymbol{\delta}_i$ and $\boldsymbol{\delta}_j$. As the diagonal matrix is contained in $\mathcal{B}(n)$, the result is immediate in that case.

For simplicity, consider the case $i = 1$ and $j = 2$, and let $\boldsymbol{R}_\theta$ denote the rotation by angle θ in the first two coordinates:

$$\boldsymbol{R}_\theta \stackrel{\text{def}}{=} \begin{bmatrix} \cos\theta & \sin\theta & \boldsymbol{0} \\ -\sin\theta & \cos\theta & \boldsymbol{0} \\ 0 & 0 & \boldsymbol{I}_{n-2} \end{bmatrix}.$$

The hyperoctahedral group $\mathcal{B}(n)$ contains the matrices $\boldsymbol{R}_\theta$ for $\theta \in \{0, \pi/2, \pi, 3\pi/2\}$. As $\mathcal{B}(n)$ is a group, for these θ we know

$$\mathbb{E}_{\boldsymbol{P}\in\mathcal{B}(n)} f(\boldsymbol{P}) = \mathbb{E}_{\boldsymbol{P}\in\mathcal{B}(n)} f(\boldsymbol{R}_\theta \boldsymbol{P}),$$

as the set of matrices in the expectations are identical. This identity implies

$$\frac{1}{4}\sum_{j=0}^{3} \mathbb{E}_{\boldsymbol{P}\in\mathcal{B}(n)} f_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{R}_{2\pi j/4}\boldsymbol{P}) = \mathbb{E}_{\boldsymbol{P}\in\mathcal{B}(n)} f(\boldsymbol{P}).$$

We will prove the following lemma, and then show it implies Theorem 23.4.1.

Lemma 23.5.1. *For every symmetric $\boldsymbol{A}$ and $\boldsymbol{B}$, and every θ*

$$\frac{1}{4}\sum_{j=0}^{3} f_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{R}_{\theta+2\pi j/4}) = \frac{1}{4}\sum_{j=0}^{3} f_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{R}_{2\pi j/4}).$$

This lemma implies that for every $\boldsymbol{Q}_{1,2}$,

$$\mathbb{E}_{\boldsymbol{P}\in\mathcal{B}(n)} f(\boldsymbol{P}) = \mathbb{E}_{\boldsymbol{P}\in\mathcal{B}(n)} f(\boldsymbol{Q}_{1,2}\boldsymbol{P}).$$

This, in turn, implies Theorem 23.4.1 and thus Theorem 23.2.3.

We can use Lemma 23.3.1 to derive Lemma 23.5.1 follows from the following.

Lemma 23.5.2. *For every symmetric $\boldsymbol{A}$ and $\boldsymbol{B}$, there exist $c_{-2}, c_{-1}, c_0, c_1, c_2$ so that*

$$f_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{R}_\theta) = \sum_{k=-2}^{2} c_k (e^{i\theta})^k.$$

Proof. We need to express $f(\boldsymbol{R}_\theta)$ as a function of $e^{i\theta}$. To this end, recall that

$$\cos\theta = \frac{1}{2}(e^{i\theta} + e^{-i\theta}) \quad \text{and} \quad \sin\theta = \frac{-i}{2}(e^{i\theta} - e^{-i\theta}).$$

From these identities, we see that all two-by-two rotation matrices can be simultaneously diagonalized by writing

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \boldsymbol{U} \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} \boldsymbol{U}^*,$$

where

$$\boldsymbol{U} = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix},$$

and we recall that $\boldsymbol{U}^*$ is the conjugate transpose:

$$\boldsymbol{U}^* = \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}.$$

Let $\boldsymbol{D}_\theta$ be the digaonal matrix having its first two entries $e^{i\theta}$ and $e^{-i\theta}$, and the rest 1, and let $\boldsymbol{U}_n$ be the matrix with $\boldsymbol{U}$ in its upper 2-by-2 block and 1s on the diagonal beneath. So,

$$\boldsymbol{R}_\theta = \boldsymbol{U}_n \boldsymbol{D}_\theta \boldsymbol{U}_n^*.$$

Now, examine

$$\begin{aligned} f_{\boldsymbol{A},\boldsymbol{B}}(\boldsymbol{R}_\theta) &= \det(\boldsymbol{A} + \boldsymbol{R}_\theta \boldsymbol{B} \boldsymbol{R}_\theta^*) \\ &= \det(\boldsymbol{A} + \boldsymbol{U}_n \boldsymbol{D}_\theta \boldsymbol{U}_n^* \boldsymbol{B} \boldsymbol{U}_n \boldsymbol{D}_\theta^* \boldsymbol{U}_n^*) \\ &= \det(\boldsymbol{U}_n^* \boldsymbol{A} \boldsymbol{U}_n + \boldsymbol{D}_\theta \boldsymbol{U}_n^* \boldsymbol{B} \boldsymbol{U}_n \boldsymbol{D}_\theta^*) \\ &= \det(\boldsymbol{U}_n^* \boldsymbol{A} \boldsymbol{U}_n \boldsymbol{D}_\theta + \boldsymbol{D}_\theta \boldsymbol{U}_n^* \boldsymbol{B} \boldsymbol{U}_n). \end{aligned}$$

The term $e^{i\theta}$ only appears in the first row and column of this matrix, and the term $e^{-i\theta}$ only appears in the second row and column. As a determinant can be expressed as a sum of products of matrix entries with one in each row and column, it is immediate that this determinant can be expressed in terms of $e^{ki\theta}$ for $|k| \leq 4$. As each such product can have at most 2 terms of the form $e^{i\theta}$ and at most two of the form $e^{-i\theta}$, we have $|k| \leq 2$. □

The difference between Theorem 23.2.3 and Theorem 23.2.2 is that the first involves a sum over the isometries of hyperoctahedron, while the second involves a sum over the symmetries of the regular n-simplex in $n-1$ dimensions. The proof of the appropriate quadrature theorem for the symmetries of the regular simplex is very similar to the proof we just saw, except that rotations of the plane through $\boldsymbol{\delta}_i$ and $\boldsymbol{\delta}_j$ are replaced by rotations of the plane parallel to the affine subspace spanned by triples of vertices of the simplex.

23.6 The Formula

To establish the formula in Theorem 23.2.1, we observe that it suffices to compute the formula for diagonal matrices, and that Theorem 23.2.3 makes this simple. Every matrix in $\mathcal{B}(n)$ can be written as a product $\Pi \boldsymbol{D}$ where $\boldsymbol{D}$ is a ± 1 diagonal matrix. If $\boldsymbol{B}$ is the diagonal matrix with entries $\mu_1, \dots, \mu_n$, then $\Pi \boldsymbol{D} \boldsymbol{B} \boldsymbol{D} \Pi^T = \Pi \boldsymbol{B} \Pi^T$, which is the diagonal matrix with entries $\mu_{\pi(1)}, \dots, \mu_{\pi(n)}$, where π is the permutation corresponding to Π.

Let $\boldsymbol{A}$ be diagonal with entries $\lambda_1, \dots, \lambda_n$. For a subset S of $\{1, \dots, n\}$, define

$$\lambda^S = \prod_{i \in S} \lambda_i.$$

We then have

$$a_i = \sum_{|S|=i} \lambda^S.$$

Let

$$p \boxplus_n q = \sum_{k=0}^{n} x^{n-k} (-1)^k c_k.$$

We first compute the expected determinant, c_n.

$$c_n = \frac{1}{n!} \sum_{\pi} \prod_{h} (\lambda_h + \mu_{\pi(h)}) = \frac{1}{n!} \sum_{\pi} \sum_{|S|=i} \lambda^S \prod_{h : \pi(h) \notin S} \mu_h.$$

As opposed to expanding this out, let's just figure out how often the product $\lambda^S \mu^T$ appears. We must have $|T| = n - |S|$, and then this term appears for each permutation such that $\pi(T) \cap S = \emptyset$. This happens $1/\binom{n}{i}$ fraction of the time, giving the formula

$$c_n = \sum_{i=0}^{n} \frac{1}{\binom{n}{i}} \sum_{|S|=i} \lambda^S \sum_{|T|=n-i} \mu^T = \sum_{i=0}^{n} \frac{1}{\binom{n}{i}} a_i b_{n-i} = \sum_{i=0}^{n} \frac{i!(n-i)!}{n!} a_i b_{n-i}.$$

For general c_k and $i + j = k$, we see that λ^S and μ^T appear whenever $\mu(T)$ is disjoint from S. The probability of this happening is

$$\frac{\binom{n-i}{j}}{\binom{n}{j}} = \frac{(n-i)!(n-j)!j!}{n!(n-i-j)!j!} = \frac{(n-i)!(n-j)!}{n!(n-i-j)!},$$

and so

$$c_k = \sum_{i+j=k} a_i b_j \frac{(n-i)!(n-j)!}{n!(n-i-j)!}.$$

23.7 Question

For which discrete subgroups of $\mathcal{O}(n)$ does a result like Theorem 23.2.3 hold? Can it hold for a substantially smaller subgroup than the symmetries of the simplex (which has size $(n+1)!$ in n dimensions).