DOI: 10.1214/ECP.v14-1447

Elect. Comm. in Probab. **14** (2009), 90–103

ELECTRONIC COMMUNICATIONS in PROBABILITY

ORTHOGONALITY AND PROBABILITY: BEYOND NEAREST NEIGHBOR TRANSITIONS

YEVGENIY KOVCHEGOV
Department of Mathematics, Oregon State University, Corvallis, OR 97331-4605, USA
email: kovchegy@math.oregonstate.edu

Submitted October 28, 2008, *accepted in final form* February 2, 2009

AMS 2000 Subject classification: 60G05, 05E35
Keywords: reversible Markov chains, orthogonal polynomials, Karlin-McGregor representation

Abstract
In this article, we will explore why Karlin-McGregor method of using orthogonal polynomials in the study of Markov processes was so successful for one dimensional nearest neighbor processes, but failed beyond nearest neighbor transitions. We will proceed by suggesting and testing possible fixtures.

1 Introduction

This paper was influenced by the approaches described in Deift [2] and questions considered in Grünbaum [6].
The Karlin-McGreogor diagonalization can be used to answer recurrence/transience questions, as well as those of probability harmonic functions, occupation times and hitting times, and a large number of other quantities obtained by solving various recurrence relations, in the study of Markov chains, see [8], [9], [10], [11], [7], [16], [15], [13]. However with some exceptions (see [12]) those were nearest neighbor Markov chains on half-line. Grünbaum [6] mentions two main drawbacks to the method as (a) “typically one cannot get either the polynomials or the measure explicitly", and (b) “the method is restricted to ‘nearest neighbour’ transition probability chains that give rise to tridiagonal matrices and thus to orthogonal polynomials". In this paper we attempt to give possible answers to the second question of Grünbaum [6] for general reversible Markov chains. In addition, we will consider possible applications of the newer methods in orthogonal polynomials such as using Riemann-Hilbert approach, see [2], [3] and [14], and their probabilistic interpretations.
In Section 2, we will give an overview of the Karlin-McGregor method from a naive college linear algebra perspective. In 2.3, we will give a Markov chain interpretation to the result of Fokas , Its and Kitaev, connecting orthogonal polynomials and Riemann-Hilbert problems. Section 3 deals with one dimensional random walks with jumps of size $\leq m$, the $2m+1$ diagonal operators. There we consider diagonalizing with orthogonal functions. In 3.2, as an example we consider a pentadiagonal operator and use Plemelj formula, and a two sided interval to obtain the respective diagonalization. In Section 4, we use the constructive approach of Deift [2] to produce the Karlin-

McGregor diagonalization for all irreducible reversible Markov chains. After that, we revisit the example from Section 3.

2 Eigenvectors of probability operators

Suppose P is a tridiagonal operator of a one-dimensional Markov chain on $\{0,1,\dots\}$ with forward probabilities p_k and backward probabilities q_k. Suppose λ is an eigenvalue of P and $\mathbf{q}^T(\lambda) = \begin{pmatrix} Q_0 \\ Q_1 \\ Q_2 \\ \vdots \end{pmatrix}$ is the corresponding right eigenvector such that $Q_0 = 1$. So $\lambda \mathbf{q}^T = P\mathbf{q}^T$ generates the recurrence relation for Q_j. Then each $Q_j(\lambda)$ is a polynomial of j-th degree. The Karlin-McGregor method derives the existence of a probability distribution ψ such that polynomials $Q_j(\lambda)$ are orthogonal with respect to ψ. In other words, if π is stationary with $\pi_0 = 1$ and $< \cdot,\cdot >_\psi$ is the inner product in $L^2(d\psi)$, then

$$< Q_i, Q_j >_\psi = \frac{\delta_{i,j}}{\pi_j}$$

Thus $\{\sqrt{\pi_j}Q_j(\lambda)\}_{j=0,1,\dots}$ are orthonormal polynomials, where $\pi_0 = 1$ and $\pi_j = \frac{p_0 \cdots p_{j-1}}{q_1 \cdots q_j}$ $(j = 1,2,\dots)$. Also observe from the recurrence relation that the leading coefficient of Q_j is $\frac{1}{p_0 \cdots p_{j-1}}$.

Now, $\lambda^t \mathbf{q}^T = P^t \mathbf{q}^T$ implies $\lambda^t q_i = (P^t \mathbf{q}^T)_i$ for each i, and

$$< \lambda^t Q_i, Q_j >_\psi = < (P^t \mathbf{q}^T)_i, Q_j >_\psi = \frac{p_t(i,j)}{\pi_j}$$

Therefore

$$p_t(i,j) = \pi_j < \lambda^t Q_i, Q_j >_\psi$$

Since the spectrum of P lies entirely inside $(-1,1]$ interval, then so is the support of ψ. Hence, for $|z| > 1$, the generating function

$$G_{i,j}(z) = \sum_{t=0}^{+\infty} z^{-t} p_t(i,j) = -z\pi_j < \frac{Q_i}{\lambda - z}, Q_j >_\psi = -z\pi_j \int \frac{Q_i(\lambda)Q_j(\lambda)}{\lambda - z} d\psi(\lambda)$$

2.1 Converting to a Jacobi operator

Let $b_k = \sqrt{\frac{\pi_k}{\pi_{k+1}}} p_k$, then $b_k = \sqrt{\frac{\pi_{k+1}}{\pi_k}} q_{k+1}$ due to reversibility condition. Thus the recurrence relation for $\mathbf{q}$,

$$\lambda\sqrt{\pi_k}Q_k = q_k\sqrt{\pi_k}Q_{k-1} + (1 - q_k - p_k)\sqrt{\pi_k}Q_k + p_k\sqrt{\pi_k}Q_{k+1} \ ,$$

can be rewritten as

$$\lambda\sqrt{\pi_k}Q_k = b_{k-1}\sqrt{\pi_k}Q_{k-1} + a_k\sqrt{\pi_k}Q_k + b_k\sqrt{\pi_k}Q_{k+1},$$

where $a_k = 1 - q_k - p_k$. Therefore $\mathbf{q} = (\sqrt{\pi_0}Q_0, \sqrt{\pi_1}Q_1, \dots)$ solves $\widetilde{P}\mathbf{q} = \lambda\mathbf{q}$, where

$$\widetilde{P} = \begin{pmatrix} a_0 & b_0 & 0 & \dots \\ b_0 & a_1 & b_1 & \ddots \\ 0 & b_1 & a_2 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}$$

is a Jacobi (symmetric triangular with $b_k > 0$) operator. Observe that $\widetilde{P}$ is self-adjoint.
The above approach extends to all reversible Markov chains. Thus every reversible Markov operator is equivalent to a self-adjoint operator, and therefore has an all real spectrum.

2.2 Karlin-McGregor: a simple picture

It is a basic fact from linear algebra that if $\lambda_1, \dots, \lambda_n$ are distinct real eigenvalues of an $n \times n$ matrix A, and if $u_1, \dots, u_n$ and $v_1, \dots, v_n$ are the corresponding left and right eigenvectors. Then A diagonalizes as follows

$$A^t = \sum_j \frac{\lambda^t v_j^T u_j}{u_j v_j^T} = \int_{\sigma(A)} \lambda^t v^T(\lambda) u(\lambda) d\psi(\lambda) \,,$$

where $u(\lambda_j) = u_j$, $v(\lambda_j) = v_j$, spectrum $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$, and

$$\psi(\lambda) = \sum_j \frac{1}{u(\lambda)v^T(\lambda)} \delta_{\lambda_j}(\lambda) = \frac{n}{u(\lambda)v^T(\lambda)} U_{\sigma(A)}(\lambda)$$

Here $U_{\sigma(A)}(\lambda)$ is the uniform distribution over the spectrum $\sigma(A)$.
It is **important** to observe that the above integral representation is only possible if $u(\lambda)$ and $v(\lambda)$ are well defined - each eigenvalue has multiplicity one, i.e. all distinct real eigenvalues. As we will see later, this will become crucial for Karlin-McGregor diagonalization of reversible Markov chains. The operator for a reversible Markov chain is bounded and is equivalent to a self-adjoint operator, and as such has a real bounded spectrum. However the eigenvalue multiplicity will determine whether the operator's diagonalization can be expressed in a form of a spectral integral.
Since the spectrums $\sigma(P) = \sigma(P^*)$, we will extend the above diagonalization identity to the operator P in the separable Hilbert space $l^2(\mathbb{R})$. First, observe that $\mathbf{u}(\lambda) = (\pi_0 Q_0, \pi_1 Q_1, \dots)$ satisfies

$$\mathbf{u}P = \lambda P$$

due to reversibility. Hence, extending from a finite case to an infinite dimensional space $l^2(\mathbb{R})$, obtain

$$P^t = \int \lambda^t \mathbf{q}^T(\lambda)\mathbf{u}(\lambda) d\psi(\lambda) = \int \lambda^t \begin{pmatrix} \pi_0 Q_0 Q_0 & \pi_1 Q_0 Q_1 & \cdots \\ \pi_0 Q_1 Q_0 & \pi_1 Q_1 Q_1 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} d\psi(\lambda) \,,$$

where

$$\psi(\lambda) = \lim_{n \to +\infty} \psi_n(\lambda)$$

The above is the weak limit of

$$\psi_n(\lambda) = \frac{n}{\mathbf{u}(\lambda)\mathbf{q}^T(\lambda)} U_{\sigma(A_n)}(\lambda) ,$$

where A_n is the restriction of P to the first n coordinates, $< e_0, \dots, e_{n-1} >$

$$A_n = \begin{pmatrix} 1-p_0 & p_0 & 0 & \cdots & 0 \\ q_1 & 1-q_1-p_1 & p_1 & \ddots & \vdots \\ 0 & q_2 & 1-q_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & p_{n-2} \\ 0 & \cdots & 0 & q_{n-1} & 1-q_{n-1}-p_{n-1} \end{pmatrix}$$

Observe that if $Q_n(\lambda) = 0$ then $(Q_0(\lambda), \dots, Q_{n-1}(\lambda))^T$ is the corresponding right eigenvector of A_n. Thus the spectrum of $\sigma(A_n)$ is the roots of

$$Q_n(\lambda) = 0$$

So

$$\psi_n(\lambda) = \frac{n}{\mathbf{u}(\lambda)\mathbf{q}^T(\lambda)} U_{Q_n=0}(\lambda) = \frac{n}{\sum_{k=0}^{n-1} \pi_k Q_k^2(\lambda)} U_{Q_n=0}(\lambda) .$$

The orthogonality follows if we plug in $t = 0$. Since $\pi_0 Q_0 Q_0 = 1$, ψ should integrate to one.

Example. *Simple random walk and Chebyshev polynomials.* The Chebyshev polynomials of the first kind are the ones characterizing a one dimensional simple random walk on half line, i.e. the ones with generator

$$P_{ch} = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & \ddots \\ 0 & 0 & \frac{1}{2} & 0 & \ddots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}$$

So, $T_0(\lambda) = 1$, $T_1(\lambda) = \lambda$ and $T_{k+1}(\lambda) = 2\lambda T_k(\lambda) - T_{k-1}(\lambda)$ for $k = 2, 3, \dots$. The Chebyshev polynomials satisfy the following trigonometric identity:

$$T_k(\lambda) = \cos(k \cos^{-1}(\lambda))$$

Now,

$$\psi_n(\lambda) = \frac{n}{\sum_{k=0}^{n-1} \pi_k T_k^2(\lambda)} U_{\{\cos(n\cos^{-1}(\lambda))=0\}}(\lambda) ,$$

where $\pi(0) = 1$ and $\pi(1) = \pi(2) = \cdots = 2$. Here

$$U_{\{\cos(n\cos^{-1}(\lambda))=0\}}(\lambda) = U_{\{\cos^{-1}(\lambda))=\frac{\pi}{2n}+\frac{\pi k}{n},\ k=0,1,\dots,n-1\}}(\lambda)$$

Thus if $X_n \sim U_{\{\cos(n\cos^{-1}(\lambda))=0\}}$, then $Y_n = \cos^{-1}(X_n) \sim U_{\{\frac{\pi}{2n}+\frac{\pi k}{n},\ k=0,1,\dots,n-1\}}$ and Y_n converges weakly to $Y \sim U_{[0,\pi]}$. Hence X_n converges weakly to

$$X = \cos(Y) \sim \frac{1}{\pi\sqrt{1-\lambda^2}}\chi_{[-1,1]}(\lambda)d\lambda\ ,$$

i.e.

$$U_{\{\cos(n\cos^{-1}(\lambda))=0\}}(\lambda) \to \frac{1}{\pi\sqrt{1-\lambda^2}}\chi_{[-1,1]}(\lambda)d\lambda$$

Also observe that if $x = \cos(\lambda)$, then

$$\sum_{k=0}^{n-1}\pi_k T_k^2(\lambda) = -1 + 2\sum_{k=0}^{n-1}\cos^2(kx) = n - \frac{1}{2} + \frac{\sin((2n-1)x)}{2\sin(x)}$$

Thus

$$d\psi_n(\lambda) \to d\psi(\lambda) = \frac{1}{\pi\sqrt{1-\lambda^2}}\chi_{[-1,1]}(\lambda)d\lambda$$

2.3 Riemann-Hilbert problem and a generating function of $p_t(i,j)$

Let us write $\sqrt{\pi_j}Q_j(\lambda) = k_j P_j(\lambda)$, where $k_j = \frac{1}{\sqrt{p_0\cdots p_{j-1}}\sqrt{q_1\cdots q_j}}$ is the leading coefficient of $\sqrt{\pi_j}Q_j(\lambda)$, and $P_j(\lambda)$ is therefore a *monic* polynomial.

In preparation for the next step, let $w(\lambda)$ be the probability density function associated with the spectral measure ψ: $d\psi(\lambda) = w(\lambda)d\lambda$ on the compact support, $supp(\psi) \subset [-1,1] = \Sigma$. Also let

$$C(f)(z) = \frac{1}{2\pi i}\int_\Sigma \frac{f(\lambda)}{\lambda - z}d\psi(\lambda)$$

denote the Cauchy transform w.r.t. measure ψ.

First let us quote the following theorem.

Theorem 1. [Fokas, Its and Kitaev, 1990] *Let*

$$v(z) = \begin{pmatrix} 1 & w(z) \\ 0 & 1 \end{pmatrix}$$

be the jump matrix. Then, for any $n \in \{0,1,2,\dots\}$,

$$m^{(n)}(z) = \begin{pmatrix} P_n(z) & C(P_n w)(z) \\ -2\pi i k_{n-1}^2 P_{n-1}(z) & -2\pi i k_{n-1}^2 C(P_{n-1}w)(z) \end{pmatrix},\ \ \textit{for all } z \in \mathbb{C}\setminus\Sigma,$$

is the unique solution to the Riemann-Hilbert problem with the above jump matrix $v(x)$ and Σ that satisfies the following condition

$$m^{(n)}(z)\begin{pmatrix} z^{-n} & 0 \\ 0 & z^n \end{pmatrix} \to I \ \ as \ \ z \to \infty\ . \tag{1}$$

The Riemann-Hilbert problem, for an oriented smooth curve Σ, is the problem of finding $m(z)$, analytic in $\mathbb{C} \setminus \Sigma$ such that

$$m_+(z) = m_-(z)v(z), \quad \text{for all } z \in \Sigma,$$

where m_+ and m_- denote respectively the limit from the left and the limit from the right of function m as the increment approaches a point on Σ.
Suppose we are given the weight function $w(\lambda)$ for the Karlin-McGregor orthogonal polynomials $\mathbf{q}$. If $m^{(n)}(z)$ is the solution of the Riemann-Hilbert problem as in the above theorem, then for $|z| > 1$,

$$m^{(n)}(z) = \begin{pmatrix} \frac{1}{k_n\sqrt{\pi_n}}Q_n(z) & -\frac{1}{2\pi i k_n\sqrt{\pi_n}z^{n+1}}G_{0,n} \\ -2\pi i\frac{k_{n-1}}{\sqrt{\pi_{n-1}}}Q_{n-1}(z) & \frac{k_{n-1}}{\sqrt{\pi_{n-1}}z^n}G_{0,n-1}(z) \end{pmatrix}$$
$$= \begin{pmatrix} q_1\dots q_n Q_n(z) & -\frac{q_1\dots q_n}{2\pi i z^{n+1}}G_{0,n} \\ \frac{-2\pi i}{p_0\cdots p_{n-2}}Q_{n-1}(z) & \frac{1}{p_0\cdots p_{n-2}z^n}G_{0,n-1}(z) \end{pmatrix}$$

3 Beyond nearest neighbor transitions

Observe that the Chebyshev polynomials were used to diagonalize a simple one dimensional random walk reflecting at the origin. Let us consider a random walk where jumps of sizes one and two are equiprobable

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & \dots \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & \dots \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \dots \\ 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} & \ddots \\ 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \ddots \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & 0 & \ddots \\ \dots & \dots & \dots & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

The above random walk with the reflector at the origin is reversible with $\pi_0 = 1$ and $\pi_1 = \pi_2 = \dots = 2$. The Karlin-McGregor representation with orthogonal polynomials will not automatically extend to this case. However this does not rule out obtaining a Karlin-McGregor diagonalization with orthogonal functions.
In the case of the above pentadiagonal Chebyshev operator, some eigenvalues will be of geometric multiplicity two as

$$P = P_{ch}^2 + \frac{1}{2}P_{ch} - \frac{1}{2}I \ ,$$

where P_{ch} is the original tridiagonal Chebyshev operator.

3.1 $2m+1$ diagonal operators

Consider a $2m+1$ diagonal reversible probability operator P. Suppose it is Karlin-McGregor diagonalizable. Then for a given $\lambda \in \sigma(P)$, let $\mathbf{q}^T(\lambda) = \begin{pmatrix} Q_0 \\ Q_1 \\ Q_2 \\ \vdots \end{pmatrix}$ once again denote the corresponding right eigenvector such that $Q_0 = 1$. Since the operator is more than tridiagonal, we encounter the problem of finding the next $m-1$ functions, $Q_1(\lambda) = \mu_1(\lambda)$, $Q_2(\lambda) = \mu_2(\lambda), \ldots, Q_{m-1}(\lambda) = \mu_{m-1}(\lambda)$.

Observe that $\mathbf{q} = \mathbf{q_0} + \mathbf{q_1}\mu_1 + \cdots + \mathbf{q_{m-1}}\mu_{m-1}$, where each $\mathbf{q}_j^T(\lambda) = \begin{pmatrix} Q_{0,j} \\ Q_{1,j} \\ Q_{2,j} \\ \vdots \end{pmatrix}$ solves $P\mathbf{q}_j^T = \lambda \mathbf{q}_j^T$ recurrence relation with the initial conditions

$$Q_{0,j}(\lambda) = 0, \quad \ldots, \quad Q_{j-1,j}(\lambda) = 0, \; Q_{j,j}(\lambda) = 1, \; Q_{j+1,j}(\lambda) = 0, \quad \ldots, \quad Q_{m-1,j}(\lambda) = 0$$

In other words, $\mathbf{q}^T(\lambda) = \mathbf{Q}(\lambda)\mu^T$, where $\mathbf{Q}(\lambda) = \begin{bmatrix} | & | & & | \\ \mathbf{q}_0^T & \mathbf{q}_1^T & \cdots & \mathbf{q}_{m-1}^T \\ | & | & & | \end{bmatrix}$ and

$\mu^T = \begin{pmatrix} 1 \\ \mu_1(\lambda) \\ \vdots \\ \mu_{m-1}(\lambda) \end{pmatrix}$ is such that $\mathbf{q}(\lambda) \in l^2(\mathbb{R})$ for each $\lambda \in \sigma(P)$.

Let A_n denote the restriction of P to the first n coordinates, $< e_0, \ldots, e_{n-1} >$. Observe that if $Q_n(\lambda) = \cdots = Q_{n+m-1}(\lambda) = 0$ then $(Q_0(\lambda), \ldots, Q_{n-1}(\lambda))^T$ is the corresponding right eigenvector of A_n. Thus the spectrum of $\sigma(A_n)$ consists of the roots of

$$\det \begin{pmatrix} Q_{n,0}(\lambda) & Q_{n,1}(\lambda) & & Q_{n,m-1}(\lambda) \\ Q_{n+1,0}(\lambda) & Q_{n+1,1}(\lambda) & & Q_{n+1,m-1}(\lambda) \\ \vdots & \vdots & \cdots & \vdots \\ Q_{n+m-1,0}(\lambda) & Q_{n+m-1,1}(\lambda) & & Q_{n+m-1,m-1}(\lambda) \end{pmatrix} = 0$$

3.2 Chebyshev operators

Let us now return to the example generalizing the simple random walk reflecting at the origin. There one step and two step jumps were equally likely. The characteristic equation $z^4 + z^3 - 4\lambda z^3 + z^2 + z = 0$ for the recurrence relation

$$c_{n+2} + c_{n+1} - 4\lambda c_n + c_{n-1} + c_{n-2} = 0$$

can be easily solved by observing that if z is a solution then so are $\bar{z}$ and $\frac{1}{z}$. The solution in radicals is expressed as $z_{1,2} = r_1 \pm i\sqrt{1 - r_1^2}$ and $z_{3,4} = d_2 \pm i\sqrt{1 - r_2^2}$, where $r_1 = \frac{-1+\sqrt{9+16\lambda}}{4}$ and $r_2 = \frac{-1-\sqrt{9+16\lambda}}{4}$.

Observe that r_1 and r_2 are the two roots of $s(x) = \lambda$, where $s(x) = x^2 + \frac{1}{2}x - \frac{1}{2}$ is the polynomial for which

$$P = s(P_{ch})$$

In general, the following is true for all operators P that represent symmetric random walks reflecting at the origin, and that allow jumps of up to m flights: there is a polynomial $s(x)$ such that

$P = s(P_{ch})$ and the roots z_j of the characteristic relation in $\lambda\mathbf{c} = P\mathbf{c}$ will lie on a unit circle with their real parts $Re(z_j)$ solving $s(x) = \lambda$. The reason for the latter is the symmetry of the corresponding characteristic equation of order $2m$, implying $\frac{1}{z_j} = \bar{z_j}$, and therefore the characteristic equation for $\lambda\mathbf{c} = P\mathbf{c}$ can be rewritten as

$$s\left(\frac{1}{2}\left[z+\frac{1}{z}\right]\right) = \lambda\ ,$$

where $\frac{1}{2}\left[z+\frac{1}{z}\right]$ is the Zhukovskiy function.
In our case, $s(x) = \left(x+\frac{1}{4}\right)^2 - \frac{9}{16}$, and for $\lambda \in \left(-\frac{9}{16}, 0\right]$, there will be two candidates for $\mu_1(\lambda)$,

$$\mu_+(\lambda) = r_1 = \frac{-1+\sqrt{9+16\lambda}}{4} \quad \text{and} \quad \mu_-(\lambda) = r_2 = \frac{-1-\sqrt{9+16\lambda}}{4}$$

Taking $0 \le \arg z < 2\pi$ branch of the logarithm $\log z$, and applying Plemelj formula, one would obtain

$$\mu_1(z) = -\frac{1}{4} + z^{\frac{1}{2}} \exp\left\{\frac{1}{2}\int_{-\frac{9}{16}}^{0} \frac{ds}{s-z}\right\}\ ,$$

where $\mu_+(\lambda) = \lim_{z\to\lambda,\ Im(z)>0} \mu_1(z)$ and $\mu_-(\lambda) = \lim_{z\to\lambda,\ Im(z)<0} \mu_1(z)$.

Now, as we defined $\mu_1(z)$, we can propose the limits of integration to be a contour in $\mathbb{C}$ consisting of the $[0,1]$ segment, $\left[-\frac{9}{16},0\right)_+ = \lim_{\varepsilon\downarrow 0}\left\{z = x+i\varepsilon\ :\ x \in \left[-\frac{9}{16},0\right)\right\}$, and $\left[-\frac{9}{16},0\right)_- = \lim_{\varepsilon\downarrow 0}\left\{z = x-i\varepsilon\ :\ x \in \left[-\frac{9}{16},0\right)\right\}$. Then

$$P^t = \int_{\left[-\frac{9}{16},0\right)_- \cup \left[-\frac{9}{16},0\right)_+ \cup [0,1]} \lambda^t \mathbf{q}^T(\lambda)\mathbf{u}(\lambda) d\psi(\lambda),$$

where $\mathbf{u}(\lambda)$ is defined as before, and

$$d\psi(\lambda) = \frac{1}{2\pi\sqrt{\lambda+\frac{9}{16}}}\left(\frac{\chi_{[-\frac{9}{16},0)_-}(\lambda)}{\sqrt{1-\left(\sqrt{\lambda+\frac{9}{16}}+\frac{1}{4}\right)^2}} + \frac{\chi_{[-\frac{9}{16},0)_+}(\lambda)+\chi_{[0,1]}(\lambda)}{\sqrt{1-\left(\sqrt{\lambda+\frac{9}{16}}-\frac{1}{4}\right)^2}}\right)d\lambda$$

Let us summarize this section as follows. If the structure of the spectrum does not allow Karlin-McGregor diagonalization with orthogonal functions over $(-1,1]$, say when there are two values of $\mu^T(\lambda)$ for some λ, then one may use Plemelj formula to obtain an integral diagonalization of P over the corresponding two sided interval.

4 Spectral Theorem and why orthogonal polynomials work

The constructive proofs in the second chapter of Deift [2] suggest the reason why Karlin-McGregor theory of diagonalizing with orthogonal polynomials works for all time reversible Markov chains. The approach goes back to the works of M.Krein and N.I.Akhiezer, see [1] and [5]. Using the same logical steps as in [2], we can construct a map $\mathscr{M}$ which assigns a probability measure $d\psi$ to a reversible transition operator P on a countable state space $\{0,1,2,\dots\}$. W.l.o.g. we can assume P is symmetric as one can instead consider

$$\begin{pmatrix} \sqrt{\pi_0} & 0 & \cdots \\ 0 & \sqrt{\pi_1} & \ddots \\ \vdots & \ddots & \ddots \end{pmatrix} P \begin{pmatrix} \frac{1}{\sqrt{\pi_0}} & 0 & \cdots \\ 0 & \frac{1}{\sqrt{\pi_1}} & \ddots \\ \vdots & \ddots & \ddots \end{pmatrix}$$

which is symmetric, and its spectrum coinciding with spectrum $\sigma(P) \subset (-1,1]$.
Now, for $z \in \mathbb{C} \setminus \mathbb{R}$ let $G(z) = (e_0, (P - zI)^{-1} e_0)$. Then

$$ImG(z) = \frac{1}{2i}\left[(e_0,(P-zI)^{-1}e_0) - (e_0,(P-\bar{z}I)^{-1}e_0)\right] = (Im(z))|(P-zI)^{-1}e_0|^2$$

and therefore $G(z)$ is a **Herglotz function**, i.e. $G(z)$ is an analytic map from $\{Im(z) > 0\}$ into $\{Im(z) > 0\}$, and as all such functions, it can be represented as

$$G(z) = az + b + \int_{-\infty}^{+\infty} \left(\frac{1}{s-z} - \frac{s}{s^2+1}\right) d\psi(s), \quad Im(z) > 0$$

In the above representation $a \geq 0$ and b are real constants and $d\psi$ is a Borel measure such that

$$\int_{-\infty}^{+\infty} \frac{1}{s^2+1} d\psi(s) < \infty$$

Deift [2] uses $G(z) = (e_0, (P - zI)^{-1} e_0) = -\frac{1}{z} + O(z^{-2})$ to show $a = 0$ in our case, and

$$b = \int_{-\infty}^{+\infty} \frac{s}{s^2+1} d\psi(s)$$

as well as the uniqueness of $d\psi$. Hence

$$G(z) = \int \frac{d\psi(s)}{s-z}, \quad Im(z) > 0$$

The point of all these is to construct the spectral map

$\mathscr{M}$: {reversible Markov operators P} $\to$ {probability measures ψ on $[-1,1]$ with compact $supp(\psi)$}

The asymptotic evaluation of both sides in

$$(e_0, (P - zI)^{-1} e_0) = \int \frac{d\psi(s)}{s-z}, \quad Im(z) > 0$$

implies

$$(e_0, P^k e_0) = \int s^k d\psi(s)$$

Until now we were reapplying the logical steps in Deift [2] for the case of reversible Markov chains. However, in the original, the second chapter of Deift [2] gives a constructive proof of the following spectral theorem, that summarizes as

$$\mathscr{U}: \{\text{bounded Jacobi operators on } l^2\} \leftrightharpoons \{\text{probability measures } \psi \text{ on } \mathbb{R} \text{ with compact } supp(\psi)\},$$

where $\mathscr{U}$ is one-to-one onto.

Theorem 2. **[Spectral Theorem]** *For every bounded Jacobi operator $\mathscr{A}$ there exists a unique probability measure ψ with compact support such that*

$$G(z) = \left(e_0, (\mathscr{A} - zI)^{-1} e_0\right) = \int_{-\infty}^{+\infty} \frac{d\psi(x)}{x - z}$$

The spectral map $\mathscr{U} : \mathscr{A} \to d\psi$ is one-to-one onto, and for every $f \in L^2(d\psi)$,

$$(\mathscr{U} \mathscr{A} \mathscr{U}^{-1} f)(s) = s f(s)$$

in the following sense

$$(e_0, \mathscr{A} f(\mathscr{A}) e_0) = \int s f(s) d\psi(s)$$

So suppose P is a reversible Markov chain, then

$$\mathscr{M} : P \to d\psi \quad \text{and} \quad \mathscr{U}^{-1} : d\psi \to P_\triangle \,,$$

where $P_\triangle$ is a unique Jacobi operator such that

$$(e_0, P^k e_0) = \int s^k d\psi(s) = (e_0, P_\triangle^k e_0)$$

Now, if $Q_j(\lambda)$ are the orthogonal polynomials w.r.t. $d\psi$ associated with $P_\triangle$, then $Q_j(P_\triangle)e_0 = e_j$ and

$$\delta_{i,j} = (e_i, e_j) = (Q_i(P_\triangle)e_0, Q_j(P_\triangle)e_0) = (Q_i(P)e_0, Q_j(P)e_0)$$

Thus, if P is irreducible, then $f_j = Q_j(P)e_0$ is an orthonormal basis for Karlin-McGregor diagonalization. If we let $F = \begin{bmatrix} | & | & \\ f_0 & f_1 & \cdots \\ | & | & \end{bmatrix}$, then

$$P^t = \begin{pmatrix} & & \\ & (P^t e_i, e_j) & \\ & & \end{pmatrix} = F \begin{pmatrix} & & \\ & \int_{-1}^{1} s^t Q_i(s) Q_j(s) d\psi(s) & \\ & & \end{pmatrix} F^T,$$

where $F^T = F^{-1}$. Also Deift [2] provides a way for constructing

$$\mathscr{U}^{-1}\mathscr{M} : P \to P_\triangle$$

Since $P_\triangle$ is a Jacobi operator, it can be represented as

$$P_\triangle = \begin{pmatrix} a_0 & b_0 & 0 & \cdots \\ b_0 & a_1 & b_1 & \ddots \\ 0 & b_1 & a_2 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix} \qquad b_j > 0$$

Now,

$$(e_0, Pe_0) = (e_0, P_\triangle e_0) = a_0, \quad (e_0, P^2 e_0) = (e_0, P_\triangle^2 e_0) = a_0^2 + b_0^2$$

$$(e_0, P^3 e_0) = (e_0, P_\triangle^3 e_0) = (a_0^2 + b_0^2)a_0 + (a_0 + a_1)b_0^2$$

$$\text{and } (e_0, P^4 e_0) = (e_0, P_\triangle^4 e_0) = (a_0^2 + b_0^2)^2 + (a_0 + a_1)^2 b_0^2 + b_0^2 b_1^2$$

thus providing us with the coefficients of the Jacobi operator, a_0, b_0, $a_1, \ldots,$ and therefore the orthogonal polynomials Q_j.

Example. *Pentadiagonal Chebyshev operator.* For the pentadiagonal P that represents the symmetric random walk with equiprobable jumps of sizes one and two,

$$(e_0, Pe_0) = 0, \;\; (e_0, P^2 e_0) = \frac{1}{4}, \;\; (e_0, P^3 e_0) = \frac{3}{32}, \;\; (e_0, P^4 e_0) = \frac{9}{64}, \;\; \ldots$$

Thus

$$a_0 = 0, \;\; b_0 = \frac{1}{2}, \;\; a_1 = \frac{3}{8}, \;\; b_1 = \frac{\sqrt{11}}{8}, \text{ etc.}$$

So

$$P_\triangle = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \cdots \\ \frac{1}{2} & \frac{3}{8} & \frac{\sqrt{11}}{8} & \ddots \\ 0 & \frac{\sqrt{11}}{8} & \ddots & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}$$

and

$$Q_0(\lambda) = 1, \quad Q_1(\lambda) = 2\lambda, \quad Q_2(\lambda) = \frac{32}{\sqrt{11}}\lambda^2 - \frac{6}{\sqrt{11}}\lambda - \frac{4}{\sqrt{11}}, \quad \ldots$$

Then applying classical Fourier analysis, one would obtain

$$\left(e_0, (P - zI)^{-1} e_0\right) = \frac{1}{2\pi} \int_0^{2\pi} \frac{d\theta}{\frac{1}{2}[\cos(\theta) + \cos(2\theta)] - z} = \int_{-\frac{9}{16}}^{1} \frac{d\psi(s)}{s - z} \,,$$

where

$$d\psi(s) = \frac{1}{2\pi\sqrt{s + \frac{9}{16}}} \left(\frac{\chi_{[-\frac{9}{16},1]}(s)}{\sqrt{1 - \left(\sqrt{s + \frac{9}{16}} - \frac{1}{4}\right)^2}} + \frac{\chi_{[-\frac{9}{16},0)}(s)}{\sqrt{1 - \left(\sqrt{s + \frac{9}{16}} + \frac{1}{4}\right)^2}} \right) ds$$

To obtain the above expression for $d\psi$ we used the fact that $\left(e_0, (P - zI)^{-1} e_0\right)$ would be the same if there were no reflector at zero.

4.1 Applications of Karlin-McGregor diagonalization

Let us list some of the possible applications of the diagonalization. First, the diagonalization can be used to extract the rate of convergence to a stationary probability distribution, if there is one. If we consider a nearest-neighbor Markov chain on a half-line with finite π, i.e.

$$\rho = \sum_k \pi_k < \infty,$$

then the spectral measure ψ will contain a point mass $\frac{\delta_1(\lambda)}{\sum_{k=0}^{\infty} \pi_k Q_k^2(\lambda)} = \frac{\delta_1(\lambda)}{\rho}$ at 1 (and naturally no point mass at -1).

In order to measure the rate of convergence to a stationary distribution, the following distance is used.

Definition 1. *If μ and ν are two probability distributions over a sample space Ω, then the total variation distance is*

$$\|\nu - \mu\|_{TV} = \frac{1}{2} \sum_{x \in \Omega} |\nu(x) - \mu(x)| = \sup_{A \subset \Omega} |\nu(A) - \mu(A)|$$

Observe that the total variation distance measures the coincidence between the distributions on a scale from zero to one.

In our case, $\nu = \frac{1}{\rho}\pi$ is the stationary probability distribution. Now, suppose the process commences at site $\mu_0 = i$, then the total variation distance between the distribution $\mu_t = \mu_0 P^t$ and ν is given by

$$\begin{aligned} \left\|\nu - \mu_t\right\|_{TV} &= \frac{1}{2} \sum_j \left| \frac{\pi_j}{\rho} - \pi_j \int_{(-1,1]} \lambda^t Q_i(\lambda) Q_j(\lambda) d\psi(\lambda) \right| \\ &= \frac{1}{2} \sum_j \pi_j \left| \int_{(-1,1)} \lambda^t Q_i(\lambda) Q_j(\lambda) d\psi(\lambda) \right| \end{aligned}$$

The rates of convergence are quantified via mixing times.

Definition 2. *Suppose P is an irreducible and aperiodic Markov chain with stationary probability distribution ν. Given an $\epsilon > 0$, the mixing time $t_{mix}(\epsilon)$ is defined as*

$$t_{mix}(\epsilon) = \min\left\{t \ : \ \|\nu - \mu_t\|_{TV} \le \epsilon\right\}$$

Thus, in the case of a nearest-neighbor process commencing at $\mu_0 = i$, the corresponding mixing time has the following simple expression

$$t_{mix}(\epsilon) = \min\left\{ t \ : \ \sum_j \pi_j \left| \int_{(-1,1)} \lambda^t Q_i(\lambda) Q_j(\lambda) d\psi(\lambda) \right| \le 2\epsilon \right\}$$

Observe that the above expression is simplified when $\mu_0 = 0$. In general, if P is an irreducible aperiodic reversible Markov chain with a stationary distribution ν, that commences at $\mu_0 = 0$, then its mixing time is given by

$$t_{mix}(\epsilon) = \min\left\{ t \ : \ \left\| e_i F \left(\quad \int_{(-1,1)} s^t Q_j(s) Q_k(s) d\psi(s) \quad \right) F^T \right\|_{\ell_1} \leq 2\epsilon \right\},$$

where Q_j, F and ψ are as described earlier in the section.
Now, if $\mu_0 = 0$, then $e_0 F = ((e_0, f_0), (e_0, f_1), \dots) = e_0$ by construction, and therefore

$$\begin{aligned} t_{mix}(\epsilon) &= \min\left\{ t \ : \ \left\| \left(\int_{(-1,1)} s^t Q_0(s) d\psi(s), \int_{(-1,1)} s^t Q_1(s) d\psi(s), \dots \right) F^T \right\|_{\ell_1} \leq 2\epsilon \right\} \\ &= \min\left\{ t \ : \ \left\| e_0 \left(\sum_j \left[\int_{(-1,1)} s^t Q_j(s) d\psi(s) \right] \ Q_j(P) \right) \right\|_{\ell_1} \leq 2\epsilon \right\} \end{aligned}$$

in this case.
For example, if we estimate the rate of convergence of $\sum_j \left| \int_{(-1,1)} s^t Q_j(s) d\psi(s) \right| \left\| e_0 Q_j(P) \right\|_{\ell_1}$, we would obtain an upper bound on the mixing time. The estimate can be simplified if we use the ℓ^2 distance instead of the total variation norm in the definition of the mixing time.

We conclude by listing some other important applications of the method.

- The generator

$$G(z) = \left(\quad G_{i,j}(z) \quad \right) = F \left(\quad -z \int_{-1}^{1} \frac{Q_i(\lambda) Q_j(\lambda)}{\lambda - z} d\psi(\lambda) \quad \right) F^T$$

- One can use the Fokas, Its and Kitaev results, and benefit from the connection between orthogonal polynomials and Riemann-Hilbert problems.
- One can interpret random walks in random environment as a random spectral measure.

References

[1] N.I.Aheizer and M.Krein, *Some Questions in the Theory of Moments.* Translations of Math Monographs, Vol.**2**, Amer. Math. Soc., Providance, RI, (1962) (translation of the 1938 Russian edition) MR0167806

[2] P.Deift, *Orthogonal Polynomials and Random Matrices: A Riemann-Hilbert Approach.* Amer. Math. Soc., Providance, RI, (2000) MR1677884

[3] P.Deift, *Riemann-Hilbert Methods in the Theory of Orthogonal Polynomials* Spectral Theory and Mathematical Physics, Vol. **76**, Amer. Math. Soc., Providance, RI, (2006) pp.715-740 MR2307753

[4] H.Dym and H.P.McKean, *Gaussian processes, function theory, and the inverse spectral problem* Probability and Mathematical Statistics, **31**, Academic, New York - London (1976) MR0448523

[5] M.L.Gorbachuk and V.I.Gorbachuk, *M.G.Krein's Lectures on Entire Operators* Birkhäuser Verlag (1997)

[6] F.A. Grünbaum, *Random walks and orthogonal polynomials: some challenges* Probability, Geometry and Integrable Systems - MSRI Publications, Vol. **55**, (2007), pp.241-260. MR2407600

[7] S.Karlin, *Total Positivity* Stanford University Press, Stanford, CA (1968) MR0230102

[8] S.Karlin and J.L.McGregor, *The differential equations of birth and death processes, and the Stieltjes moment problem* Transactions of AMS, **85**, (1957), pp.489-546. MR0091566

[9] S.Karlin and J.L.McGregor, *The classification of birth and death processes* Transactions of AMS, **86**, (1957), pp.366-400. MR0094854

[10] S.Karlin and J.L.McGregor, *Random Walks* Illinois Journal of Math., **3**, No. 1, (1959), pp.417-431. MR0100927

[11] S.Karlin and J.L.McGregor, *Occupation time laws for birth and death processes* Proc. 4th Berkeley Symp. Math. Statist. Prob., **2**, (1962), pp.249-272. MR0137180

[12] S.Karlin and J.L.McGregor, *Linear Growth Models with Many Types and Multidimensional Hahn Polynomials* In: R.A. Askey, Editor, Theory and Applications of Special Functions, Academic Press, New York (1975), pp. 261-288. MR0406574

[13] Y.Kovchegov, N.Meredith and E.Nir *Occupation times via Bessel functions* preprint

[14] A.B.J.Kuijlaarsr, *Riemann-Hilbert Analysis for Orthogonal Polynomials* Orthogonal Polynomials and Special Functions (Springer-Verlag), Vol. **1817**, (2003) MR2022855

[15] W.Schoutens, *Stochastic Processes and Orthogonal Polynomials.* Lecture notes in statistics (Springer-Verlag), Vol. **146**, (2000) MR1761401

[16] G.Szegö, *Orthogonal Polynomials.* Fourth edition. AMS Colloquium Publications, Vol. **23**, (1975) MR0372517