Free Convolution and Jacobi Matrices

Anuran Makur
EECS Department
Massachusetts Institute of Technology
Email: a_makur@mit.edu

I. Introduction

We consider the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and two probability measures $\mathbb{P}$ and $\mathbb{Q}$ on this space. We further assume the regularity condition that the moments of $\mathbb{P}$ and $\mathbb{Q}$ are all finite. It is well-known that there are several equivalent representations of probability measures. Some common examples include the moment generating function (or Laplace transform) when it exists in an open interval around zero, the characteristic function (or Fourier transform), the set of moments (where we essentially "sample" the moment generating function by expanding it using Taylor's theorem), the cumulant generating function (logarithm of the moment generating function) or its Taylor coefficients known as cumulants, the Cauchy transform,[1] and the R-transform or its Taylor coefficients the free cumulants. Naturally, there are several regularity conditions for each of these descriptions to completely and uniquely characterize a probability measure. For example, the moments of a distribution completely characterize a distribution if Carleman's condition holds; more generally, the "inverse problem" of determining distributions from moments is known as the moment problem whose variants include the Hamburger and Stieltjes moments problems. We will be particularly interested in the representation of probability measures using Jacobi matrices, which are symmetric tridiagonal matrices.

Recall that any probability measure $\mathbb{P}$ defines a real Hilbert space of $\mathbb{P}$-square integrable functions on $\mathbb{R}$, denoted $\mathcal{L}^2(\mathbb{R}, \mathbb{P})$, equipped with correlation as the inner product:

$$\forall f, g \in \mathcal{L}^2(\mathbb{R}, \mathbb{P}), \quad \langle f, g \rangle_{\mathbb{P}} \triangleq \int_{\mathbb{R}} f g \, d\mathbb{P}. \tag{1}$$

If $\mathbb{P}$ is positive definite over the space of polynomials on $\mathbb{R}$ (which is easily verified by checking the positive definiteness of an infinite Hankel matrix of moments using Sylvester's criterion), then there is a unique set of orthonormal polynomials up to arbitrary sign changes, which we will denote $\{\pi_n : \mathbb{R} \to \mathbb{R} \,|\, n \in \mathbb{N} = \{0, 1, 2, \dots\}\}$ where π_n is a polynomial with degree n, corresponding to $\mathbb{P}$ that (usually) form a complete orthonormal basis for the Hilbert space. This set of orthonormal polynomials also serves as a valid representation of $\mathbb{P}$. While orthonormal polynomials can be naïvely defined using the Gram-Schmidt algorithm, perhaps a more intriguing and fundamental property they satisfy is the following three-term recurrence relation:

$$\forall n \in \mathbb{N}, \, \forall x \in \mathbb{R}, \quad x\pi_n(x) = b_n \pi_{n-1}(x) + a_{n+1}\pi_n(x) + b_{n+1}\pi_{n+1}(x) \tag{2}$$

where the sequences $\{a_n : n \in \mathbb{N}, \, n \geq 1\}$ and $\{b_n : n \in \mathbb{N}, \, n \geq 1\}$ characterize the recurrence. We note that the recursion in (2) has the base case $n = 0$ where we define $\pi_{-1}(x) \triangleq 0$ and $\pi_0(x) = 1$ (regardless of $\mathbb{P}$). Moreover, Favard's theorem asserts that certain sequences $\{a_n : n \in \mathbb{N}, \, n \geq 1\}$ and $\{b_n : n \in \mathbb{N}, \, n \geq 1\}$ define unique families of orthonormal polynomials. So, the sequences of

[1] In signal processing and communications theory, this is known as the Hilbert transform. It is useful in phase retrieval problems.

coefficients $\{a_n : n \in \mathbb{N},\, n \geq 1\}$ and $\{b_n : n \in \mathbb{N},\, n \geq 1\}$ are yet another valid representation of $\mathbb{P}$, and we construct a tridiagonal matrix using these sequences as follows:

$$T_n(\mathbb{P}) \triangleq \begin{bmatrix} a_1 & b_1 & 0 & \cdots & 0 \\ b_1 & a_2 & b_2 & \cdots & 0 \\ 0 & b_2 & a_3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & b_{n-1} \\ 0 & 0 & \cdots & b_{n-1} & a_n \end{bmatrix}. \tag{3}$$

Such matrices are dubbed "Jacobi matrices," and the infinite Jacobi matrix $T_\infty(\mathbb{P})$ (as $n \to \infty$) fully characterizes $\mathbb{P}$.

II. Properties of Jacobi Matrices

Here, we review how to translate between $\mathbb{P}$ and its Jacobi matrix $T_\infty(\mathbb{P})$.

A. From Jacobi Matrix to Probability Measure

Given $T_n(\mathbb{P})$ for some large $n \in \mathbb{N}$, we first compute its spectral decomposition:

$$T_n(\mathbb{P}) = Q\Lambda Q^T \tag{4}$$

where Q is an orthogonal matrix of eigenvectors (in the columns), and Λ is a diagonal matrix of eigenvalues $\{\lambda_i = \Lambda_{i,i} : 1 \leq i \leq n\}$. Let $q(T_n(\mathbb{P})) \triangleq Q^T e_1$ be the first row of Q, where $e_1 = [1\ \ 0 \ldots 0]^T$ is a standard basis vector. Then, we define the *spectral measure* of $T_n(\mathbb{P})$ as:

$$\mu_n(\mathbb{P}) \triangleq \sum_{i=1}^{n} q(T_n(\mathbb{P}))_i^2\, \delta_{\lambda_i} \tag{5}$$

where δ_x denotes a Dirac delta measure at $x \in \mathcal{X}$. The spectral measure satisfies the following Gaussian quadrature property for any polynomial $\tau : \mathbb{R} \to \mathbb{R}$ with degree at most $2n-1$:

$$\forall n \in \mathbb{N},\quad \int_{\mathbb{R}} \tau\, d\mathbb{P} = \int_{\mathbb{R}} \tau\, d\mu_n(\mathbb{P}) = \sum_{i=1}^{n} q(T_n(\mathbb{P}))_i^2\, \tau(\lambda_i) \tag{6}$$

where the integrals are abstract Lebesgue integrals. Naturally, we may extend (6) to get $\mu_n(\mathbb{P}) \to \mathbb{P}$ weakly (in distribution) as $n \to \infty$ using the Portmanteau theorem.

Suppose we now ask for n independent samples of the probability measure $\mathbb{P}$. We may obtain this sampling from the Jacobi matrix $T_n(\mathbb{P})$ by exploiting the weak convergence of the spectral measure to $\mathbb{P}$. In particular, we need to "re-sample" the eigenvalues in Λ so that they resemble samples drawn from the spectral measure in (5). This "re-sampling" can be performed by constructing a vector where each λ_i appears $\left\lfloor nq(T_n(\mathbb{P}))_i^2 \right\rceil$ times, where $\lfloor\cdot\rceil$ denotes the rounding operation. We will refer to this "re-sampling" operation as $\mathsf{Resample} : \mathbb{R}^n \times \{x \in \mathbb{R}^n : \|x\|_2^2 = 1\} \to \mathbb{R}^n$, where $\mathsf{Resample}\left([\lambda_1\ \ \lambda_2 \ldots \lambda_n]^T, q(T_n(\mathbb{P}))\right)$ produces the "re-sampled" vector we delineated earlier.

B. From Probability Measure to Jacobi Matrix

Given a probability measure $\mathbb{P}$, it is well-known that the QR decomposition of a certain Krylov matrix can be used to produce $T_n(\mathbb{P})$. Indeed, we consider the diagonal kernel:

$$\forall x, y \in \mathbb{R},\quad d(x,y) = x\delta(x-y) \tag{7}$$

where $\delta(\cdot)$ denotes the Dirac delta function. This defines the integral operator:

$$\forall x \in \mathbb{R},\quad (Df)(x) \triangleq \int_{\mathbb{R}} d(x,y) f(y)\, dy = xf(x) \tag{8}$$

for every function $f : \mathbb{R} \to \mathbb{R}$. Consider the Krylov space spanned by $\left\{D^k \mathbf{1} : k \in \mathbb{N}\right\}$, which is the subspace of monomials $\left\{\mathbf{1}, x, x^2, \dots\right\}$, where $\mathbf{1} : \mathbb{R} \to \mathbb{R}$ is the everywhere unity function. We perform a Gram-Schmidt orthonormalization of this set of vectors (where the inner product is defined in (1)) to obtain the orthonormal polynomials $\{\pi_n : \mathbb{R} \to \mathbb{R} | n \in \mathbb{N}\}$. The Jacobi matrix corresponding to $\mathbb{P}$ is then given by:

$$\forall i, j \geq 1, \;\; \left[T_\infty\left(\mathbb{P}\right)\right]_{i,j} = \int_{\mathbb{R}} x \pi_i(x) \pi_j(x) \, d\mathbb{P}(x). \tag{9}$$

To perform these operations numerically, assume we have m independent samples drawn from $\mathbb{P}$ and placed in a vector $X = \left[X_1 \;\; X_2 \dots X_m\right]^T \in \mathbb{R}^m$. We then construct the Krylov matrix:

$$K_{m,n} = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & X_1 & X_1^2 & \cdots & X_1^{n-1} \\ 1 & X_2 & X_2^2 & \cdots & X_2^{n-1} \\ 1 & X_3 & X_3^2 & \cdots & X_3^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_m & X_m^2 & \cdots & X_m^{n-1} \end{bmatrix} \tag{10}$$

for some large $n \in \mathbb{N}$. Taking a QR decomposition of this Krylov matrix: $K_{m,n} = Q^T R$, where Q lies in the Stiefel manifold and R is upper-triangular, we approximately obtain the sampled orthonormal polynomials:

$$Q^T \approx \begin{bmatrix} \pi_0\left(X_1\right) & \pi_1\left(X_1\right) & \pi_2\left(X_1\right) & \cdots & \pi_{n-1}\left(X_1\right) \\ \pi_0\left(X_2\right) & \pi_1\left(X_2\right) & \pi_2\left(X_2\right) & \cdots & \pi_{n-1}\left(X_2\right) \\ \pi_0\left(X_3\right) & \pi_1\left(X_3\right) & \pi_2\left(X_3\right) & \cdots & \pi_{n-1}\left(X_3\right) \\ \vdots & \vdots & \vdots & & \vdots \\ \pi_0\left(X_1\right) & \pi_1\left(X_m\right) & \pi_2\left(X_m\right) & \cdots & \pi_{n-1}\left(X_m\right) \end{bmatrix}. \tag{11}$$

This is because the QR decomposition effectively performs the Gram-Schmidt algorithm, and the Euclidean inner product between the jth and kth columns of the Krylov matrix tends to $\left\langle x^j, x^k \right\rangle_{\mathbb{P}}$:

$$\sum_{i=1}^{m} X_i^j X_i^k \longrightarrow \int_{\mathbb{R}} x^j x^k \, d\mathbb{P}(x) \;\; a.s. \tag{12}$$

as $m \to \infty$, by the strong law of large numbers. Finally, we may compute the Jacobi matrix:

$$T_n\left(\mathbb{P}\right) = Q \, \mathsf{diag}(X) \, Q^T \tag{13}$$

which is tridiagonal. Alternatively, we could have used the more numerically stable Lanczos algorithm (with initial vector $n^{-1/2}\left[1 \;\; 1 \cdots 1\right]^T$) to calculate this Jacobi matrix. We will refer to this process of obtaining $T_n\left(\mathbb{P}\right)$ from X as $\mathsf{Lanczos} : \mathbb{R}^m \to R^{n \times n}$, where $\mathsf{Lanczos}(X) = T_n\left(\mathbb{P}\right)$.

III. Free Convolution

For the two probability measures defined earlier, $\mathbb{P}$ and $\mathbb{Q}$, we may define their free convolution as the inverse R-transform of the sum of their R-transforms. We will use the notation $\mathbb{P} \star_{\mathsf{free}} \mathbb{Q}$ for the free convolution. In our scenario, it is more pertinent to consider the matrix view of free convolution. Indeed, if $X \in \mathbb{R}^n$ and $Y \in \mathbb{R}^n$ are independent random samples drawn from $\mathbb{P}$ and $\mathbb{Q}$ respectively, then the vector of eigenvalues:

$$Z = \mathsf{eig}\left(\mathsf{diag}\left(X\right) + Q \, \mathsf{diag}\left(Y\right) Q^T\right) \tag{14}$$

where Q is an uniform Haar distributed orthogonal matrix, can be construed as a random independent sampling of the free convolution $\mathbb{P} \star_{\mathsf{free}} \mathbb{Q}$.

IV. Performing Free Convolutions using Jacobi Matrices

Given the Jacobi matrices $T_n(\mathbb{P})$ and $T_n(\mathbb{Q})$ corresponding to $\mathbb{P}$ and $\mathbb{Q}$ respectively, we may construct the Jacobi matrix of the free convolution $\mathbb{P} \star_{\mathsf{free}} \mathbb{Q}$ using the following three-step algorithm:

1) **Resample Step.** Construct the vectors:

$$X = \mathsf{Resample}\left(\mathsf{eig}\left(T_n(\mathbb{P})\right), q\left(T_n(\mathbb{P})\right)\right),$$
$$Y = \mathsf{Resample}\left(\mathsf{eig}\left(T_n(\mathbb{Q})\right), q\left(T_n(\mathbb{Q})\right)\right).$$

2) **Free Sum Step.** Then, as shown in (14), compute:

$$Z = \mathsf{eig}\left(\mathsf{diag}(X) + Q\,\mathsf{diag}(Y)\,Q^T\right)$$

where Q is uniform Haar distributed.

3) **Lanczos Step.** Finally, calculate:

$$T_n(\mathbb{P} \star_{\mathsf{free}} \mathbb{Q}) \approx \mathsf{Lanczos}(Z).$$

One obviously needs to rigorously argue that as $n \to \infty$, the final step of the above algorithm indeed converges (in some sense) to the Jacobi matrix $T_n(\mathbb{P} \star_{\mathsf{free}} \mathbb{Q})$. However, this algorithm suffices for simulation purposes.

V. Simulations for the Semicircle Distribution

We test this algorithm using simulations for the *Wigner semicircle distribution*. The semicircle pdf with radius $r > 0$ is defined as:

$$f_X^{(r)}(x) = \begin{cases} \frac{2}{\pi r^2}\sqrt{r^2 - x^2} & , \ -r \le x \le r \\ 0 & , \ \text{otherwise} \end{cases} \tag{15}$$

and it has mean and variance:

$$\mathbb{E}_{f_X^{(r)}}[X] = 0, \tag{16}$$

$$\mathbb{VAR}_{f_X^{(r)}}(X) = \frac{r^2}{4}. \tag{17}$$

The orthonormal polynomials with respect to the semicircle distribution are the *Chebyshev polynomials of the second kind*, $\{U_n^{(r)} : [-r, r] \to \mathbb{R} \,|\, n \in \mathbb{N}\}$, where $U_n^{(r)}$ has degree n. They satisfy the orthogonality relation:

$$\forall n, m \in \mathbb{N}, \quad \int_{-r}^{r} U_n^{(r)}(x) U_m^{(r)}(x) f_X^{(r)}(x)\, dx = \delta_{nm} \tag{18}$$

where δ_{nm} denotes a Kronecker delta function, and the three-term recurrence relation:

$$xU_n^{(r)}(x) = \frac{r}{2}U_{n-1}^{(r)}(x) + \frac{r}{2}U_{n+1}^{(r)}(x) + 0U_n^{(r)}(x) \tag{19}$$

for $n \in \mathbb{N}$ with the base cases $U_{-1}^{(r)}(x) \triangleq 0$ and $U_0^{(r)}(x) = 1$. The sequences of coefficients for this recursion are $a_n = 0$ and $b_n = \frac{r}{2}$ for every $n \ge 1$. Hence, the $n \times n$ Jacobi matrix corresponding to the semicircle distribution $f_X^{(r)}$ is:

$$T_n\left(f_X^{(r)}\right) \triangleq \begin{bmatrix} 0 & \frac{r}{2} & 0 & \cdots & 0 \\ \frac{r}{2} & 0 & \frac{r}{2} & \cdots & 0 \\ 0 & \frac{r}{2} & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \frac{r}{2} \\ 0 & 0 & \cdots & \frac{r}{2} & 0 \end{bmatrix}. \tag{20}$$

The semicircle distribution has the interesting property that it is closed under free convolutions. For example, we have:

$$f_X^{(2)} \star_{\text{free}} f_X^{(2)} = f_X^{(2\sqrt{2})} \tag{21}$$

where we represent the probability measure corresponding to $f_X^{(r)}$ using $f_X^{(r)}$ by abusing notation. Observe that the variances on the left hand side of (21) add to produce the variance on the right hand side of (21). Therefore, if we apply the algorithm from Section IV to the inputs $T_n(f_X^{(2)})$ and $T_n(f_X^{(2)})$ for some large value of $n \in \mathbb{N}$, we expect to see an output resembling $T_n(f_X^{(2\sqrt{2})})$. The next subsection presents Julia code for this algorithm and plots illustrating the results.

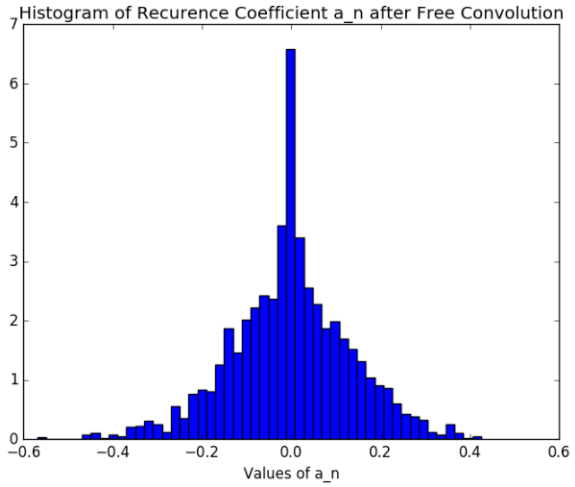
A. Julia Code and Associated Plots

```
using PyPlot

function reSample(T,n,k) # inputs = Jacobi matrix, its dimension n,
                         # and scale factor k, output = iid samples
                         # from spectral measure
    lambda,Q = eig(T)
    q = Q[1,:]
    samplingRate = round(Int,n*k*(q.*q))
    resample = zeros(1,1)
    for i = 1:n
        if samplingRate[i] > 0
            resample = [resample; lambda[i]*ones(samplingRate[i],1)]
        end
    end
    resample = resample[2:end]
end

function freeSum(sample1,sample2,m) # inputs = iid samples from two
                                    # distributions and their length,
                                    # output = iid samples from
                                    # free convolution
    Q = qrfact(randn(m,m) + im*randn(m,m))[:Q]
    U = full(Q*diagm(sign(randn(m,1) + im*randn(m,1))[:,1])) # Haar
    outputSamples = eigvals(diagm(sample1) + U*(diagm(sample2)*U'))
end

function Lanczos(sample,m) # inputs = iid samples from a distribution
                           # and its length, output = Jacobi matrix
    vprev = zeros(m)
    v = ones(m)/sqrt(m)
    a = zeros(m)
    b = zeros(m)
    for i = 1:m-1
        w = sample.*v
        a[i] = (w' * v)[1]
        wproj = w - a[i]*v - b[i]*vprev
```

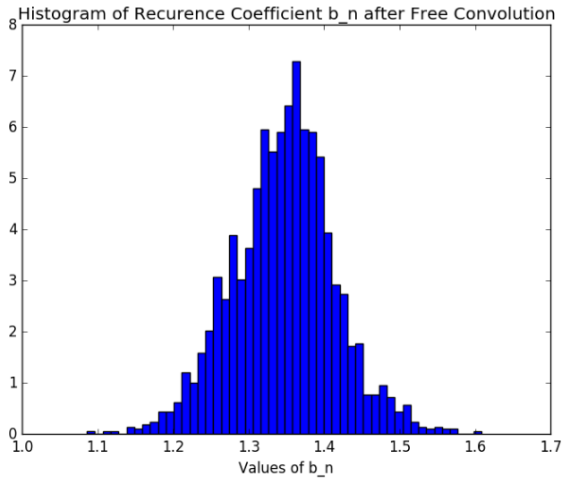
```
        b[i+1] = sqrt(wproj' * wproj)[1]
        vprev = v
        v = wproj/b[i+1]
    end
    a[m] = (v' * (sample.*v))[1]
    outputJacobi = SymTridiagonal(a,b[2:end])
end

function freeConvolveJacobi(T1,T2,n,k) # inputs = Jacobi matrices,
                                       # dimension n, scale factor k,
                                       # output = Jacobi matrix of
                                       # free convolution
    sample1 = reSample(T1,n,k)
    sample2 = reSample(T2,n,k)
    m = length(sample1) # length of both samples must be the same
    freeSumSample = freeSum(sample1,sample2,m)
    Lanczos(real(freeSumSample),m)
end

n = 100
T = SymTridiagonal(zeros(n),ones(n-1)) # semicircle pdf
freeConvT = freeConvolveJacobi(T,T,n,20) # expect to see zeros
                                         # in diagonal and sqrt(2)
                                         # in upper/lower diagonal

m1,m2 = size(freeConvT)
a = [freeConvT[k,k] for k = 1:m1]
b = [freeConvT[k+1,k] for k = 1:m1-1]

# Histogram of three-term recurrence coefficient a_n
bins = 50 # number of bins
plt[:hist](a,bins,normed="True")
xlabel("Values of a_n")
title("Histogram of Recurence Coefficient a_n
       after Free Convolution")

# Histogram of three-term recurrence coefficient b_n
bins = 50 # number of bins
plt[:hist](b,bins,normed="True")
xlabel("Values of b_n")
title("Histogram of Recurence Coefficient b_n
       after Free Convolution")
```

VI. Conclusion

In the semicircle distribution simulation, we expect the output Jacobi matrix to have zeros on the diagonal and $\sqrt{2}$ on the upper and lower diagonals (as one would see in $T_n(f_X^{(2\sqrt{2})})$). Figure 1 suggests that the values on the diagonal (which are values of the sequence $\{a_n : n \in \mathbb{N}\}$) indeed



(a) diagonal entries



(b) upper (or lower) diagonal entries

Fig. 1: Histograms of the diagonal and upper (or lower) diagonal entries of the output Jacobi matrix produced by the algorithm.

converge quickly to zero as we would expect. However, the values on the upper or lower diagonal (which are values of the sequence $\{b_n : n \in \mathbb{N}\}$) seem to require much larger values of n or a larger "scale factor" in the re-sampling phase (see Julia code in Subsection V-A) to converge. One could also try and improve the convergence of the upper or lower diagonal entries by performing many independent iterations in the free sum phase, but this would not correspond to a "nice" matrix operation theoretically. Finally, we remark that all material in this report was based on [Edelman, 2016] and [Horn and Johnson, 2013].

REFERENCES

[Edelman, 2016] Edelman, A. (2016). Random matrix theory. Lectures Notes for MIT Course 18.338.

[Horn and Johnson, 2013] Horn, R. A. and Johnson, C. R. (2013). *Matrix Analysis*. Cambridge University Press, New York, second edition.