OPERATOR FUNCTION BY CHEBYSHEV FIT

ROI BAER

Abstract. We explain the concept of orthogonal polynomials and specialize to Chebyshev polynomials. We then explain how to use this expansion for computing an operator function on a Hilbert-space function. As applications we discuss how to efficiently computer the free energy and the entropy of a particle in temperature T.

Jerusalem, January 13, 2020

1. Introduction

In many cases in Quantum Mechanics we need to apply a function $f\left(\hat{H}\right)$ of an operator $\hat{H}$ to a wavefunction ψ_0. For example, when solving the time-dependent Schrodinger, starting from a given initial state $\psi(0)=\psi_0$ at time $t=0$:

$$i\hbar\dot{\psi}(t)=\hat{H}\psi(t)\,, \tag{1.1}$$

where $\hbar$ is Planck's constant. The solution at time t is:

$$\psi(t)=e^{-i\hat{H}t/\hbar}\psi_0. \tag{1.2}$$

Here the operator function is based on the imaginary exponential function $f_t(x)=e^{-ixt/\hbar}$ and has time as a parameter t.

Another example is the partition function which is defined as a trace of an operator function:

$$Z=tr\left[e^{-\beta\hat{H}}\right] \tag{1.3}$$

where now the operator function is the real exponential function $f_\beta(x)=e^{-\beta x}$ and has the inverse temperature as a parameter β.

For non-interacting fermions the number of fermions as a function of temperature and chemical potential μ:

$$N(\beta,\mu)=tr\left[\frac{2}{1+e^{\beta\left(\hat{H}-\mu\right)}}\right] \tag{1.4}$$

Here we will describe how such an operator function can be computed using orthogonal polynomials.

2. Orthogonal Polynomials

Definition 1. A set of polynomials $p_n(x)$ $n = 0, 1, 2, ...$ is called an orthogonal set if $p_n(x)$ is of degree n and if

$$\int_a^b p_n(x)\, p_m(x)\, w(x)\, dx = c_n \delta_{nm} \tag{2.1}$$

We denote the coefficient of x^n (the leading coefficient) in $p_n(x)$ by A_n. $w(x) > 0$ is the weight function so clearly the norm c_n is positive.

Example 2. Examples of orthogonal polynomials are the Legendre polynomials, orthogonal in the interval $[-1, 1]$. These are obtained by a Gram Schmidt orthogonalization of the set $\{1, x, x^2, ...\}$. Remember that the Gram Schmidt procedure for orthogonalization of a set of vectors α_n is the set $\beta_0 = \alpha_0$ and

$$\beta_n = \alpha_n - \sum_{k=0}^{n-1} \frac{\beta_k \langle \beta_k | \alpha_n \rangle}{\langle \beta_k | \beta_k \rangle}. \tag{2.2}$$

In our case, take $p_0(x) = 1$. Then, since $\langle p_0 | x \rangle = 0$

$$p_1(x) = x. \tag{2.3}$$

Next,

$$p_2(x) = x^2 - p_1(x) \frac{\langle p_1 | x^2 \rangle}{\langle p_1 | p_1 \rangle} - p_0(x) \frac{\langle p_0 | x^2 \rangle}{\langle p_0 | p_0 \rangle} \tag{2.4}$$

$$= x^2 - \frac{2}{3} \tag{2.5}$$

and

$$p_3(x) = x^3 - p_1(x) \frac{\langle p_1 | x^3 \rangle}{\langle p_1 | p_1 \rangle} = x^3 - \frac{3}{5}x \tag{2.6}$$

Where we used: $\int_{-1}^{1} x^{2n} dx = \frac{2}{2n+1}$ and $\int_{-1}^{1} x^{2k} dx = 0$. Continuing this way we get $p_4(x) = x^4 - \frac{6}{7}x^2 + \frac{3}{35}$ etc.

Lemma 3. *The polynomial x^n is a linear combination of the orthogonal set of polynomials $p_m(x)$ where $m \le n$.*

Proof. Using induction. For $n = 0$ it is trivial. Now suppose it is true for $n > 0$ then for $n+1$ we have that $x^{n+1} - \frac{1}{A_{n+1}} p_{n+1}(x)$ is a polynomial of degree n and thus equal to $\sum_{m=0}^{n} b_m p_m(x)$. Thus, $x^{n+1} = \frac{1}{A_{n+1}} p_{n+1}(x) - \sum_{m=0}^{n} b_m p_m(x)$ i.e. it is a linear combination of $p_m(x)$'s. □

Corollary 4. *Within a set of orthogonal polynomials, p_n is orthogonal to all m-degree polynomials $q_m(x)$ where $m < n$.*

Proof. For $m < n$ $p_m(x)$ is a linear combination of $p_k(x)$ $k \le m$, all of which are orthogonal to $p_n(x)$ by definition. □

Corollary 5. *Any m-degree polynomial $q(x)$ can be written as $q(x) = \sum_{n=0}^{m} b_n p_n(x)$ where $b_n = c_n^{-1} \int q(x)\, p_n(x)\, w(x)\, dx$.*

Definition 6. A root of a polynomial $p(x)$is a point of sign change. Specifically, a point t for which $p(x)\, p(y) < 1$ for all $x < t < y$ in an infinitesimal interval around t. Naturally $p(t) = 0$.

Theorem 7. *All n roots of $p_n(x)$ are real, distinct and appear inside the interval $[a,b]$.*

Proof. Denote by $N <= n$ the number of sign changes of $p_n(x)$ in $[a,b]$. If $N = 0$ then define $q_0(x) = 1$ and if $N > 0$ define polynomial $q_N(x) = \Pi_{i=1}^{N}(x - t_i)$, where t_i are the N roots of $p_n(x)$ in the interval. $p_n(x)\, q_N(x)$ does not change sign even once in the interval and therefore $\int_a^b p_n(x)\, q_N(x)\, w(x)\, dx \neq 0$ i.e. p_n and q_N are non-orthogonal. This can only be true if $N = n$ since $p_n$is orthogonal to all polynomials of degree $N < n$. Hence $p_n(x)$ changes sign n times in the interval, which means that all its roots are distinct and real. □

3. Quadrature as approximations to integrals

Definition 8. A rule $\sum_{n=1}^{N} a_n f(x_n)$ is called a quadrature formula for $f(x)$ with N nodes x_n or a N-quadrature; a_n are called the weights. We use quadratures to approximate integrals $\int_a^b f(x)\, w(x)\, dx$. A $N-$quadrature is called "k-exact" if it is exact for all $f(x) = q_m(x)$ polynomials of degrees $m \leq k$.

Theorem 9. *A N-quadrature* cannot *be $2N$-exact.*

Proof. By counter example: $f(x) = \Pi_{n=1}^{N}(x - x_n)^2$ is a non-negative $2N$-degree polynomial. Its integral is positive while the quadrature formula yields zero. □

Definition 10. A N-quadrature which is $(2N-1)$-exact is called a **Gaussian N-quadrature**.

Theorem 11. *A N-quadrature which is $(N-1)$-exact is also $(2N-1)$-exact if and only if the nodes x_n are the roots of $p_N(x)$.*

Proof. Follow these:

(1) If the nodes x_n are the roots of $p_N(x)$ then the quadrature is $(2N-1)$exact: Any polynomial $s_k(x)$ of degree $k < 2N$ can be written as $s_k(x) = q(x)\, p_N(x) + r(x)$ where both $q(x)$ and $r(x)$ are polynomials of degree less than N. Now $\int s_k(x)\, w(x)\, dx = \int r(x)\, w(x)\, dx$ since $p_N(x)$ is orthogonal to $q(x)$. Finally, since $q(x_n)\, p_N(x_n) = 0$ and the quadrature is $(N-1)$-exact we have: $\int s_k(x)\, w(x)\, dx = \sum_{n=1}^{N} a_n r(x_n) = \sum_{n=1}^{N} a_n s_k(x_n)$, hence it is also $2N-1$exact.

(2) In a N-quadrature which is $(2N-1)$-exact the nodes x_n must be the roots of $p_N(x)$: Let $q_N(x) = \Pi_{n=1}^{N}(x - x_n)$ then for $k < N$:$\int q_N(x)\, p_k(x)\, w(x)\, dx = \sum_{n=1}^{N} a_n q_N(x_n)\, p_k(x_n) = 0$ since then the quadrature is exact. Thus $q_N(x)$ is a N degree polynomial which is orthogonal to all $p_k(x)$ i.e. it must be a constant times $p_N(x)$.

□

Corollary 12. *The weights are obtained from solving the N linear equations with N unknowns*

$$\sum_{n=1}^{N} a_n p_k(x_n) = \frac{c_0}{p_0(0)} \delta_{k0}, k = 0, \ldots, K \tag{3.1}$$

.

Theorem 13. *For a Gaussian quadrature* $\sum_{n=1}^{N} a_n p_k(x_n)\, p_j(x_n) = c_k \delta_{kj},\ k, j = 0, \ldots, K$

4. Polynomial fit approximation to functions

Given a function $f(x)$, an interval $[a,b]$ and a weight $w(x)$and a set of polynomials of degree $k=0,1,\ldots$ $p_k(x)$ we ask what is the best fit we can obtain for $f(x)$

$$f(x) \approx \sum_{n=0}^{N-1} \alpha_n p_n(x) \tag{4.1}$$

we can obtain for $f(x)$using these polynomials, so that the difference is minimal in the sense

$$J = \int_a^b \left(f(x) - \sum_{n=0}^{N-1} \alpha_n p_n(x) \right)^2 w(x)\, dx \tag{4.2}$$

We can solve this by taking derivatives with respect to α_k:

$$0 = \frac{\partial J}{\partial \alpha_k} = -2 \int_a^b \left(f(x) - \sum_{n=0}^{N-1} \alpha_n p_n(x) \right) p_k(x)\, w(x)\, dx \tag{4.3}$$

which can be rearranged to read:

$$\sum_{n=0}^{N-1} \alpha_n \int_a^b p_n(x)\, p_k(x)\, w(x)\, dx = \int_a^b f(x)\, p_k(x)\, w(x)\, dx. \tag{4.4}$$

we see that, if we choose the polynomials $p_n(x)$ as the orthogonal set with respect to $w(x)$ we obtain a simple expression :

$$\alpha_k = \frac{1}{c_k} \int_a^b f(x)\, p_k(x)\, w(x)\, dx. \tag{4.5}$$

Hence we have a systematic way of generating *optimal* approximations to functions. If we know how the perform the integrals of Eq. 4.5 we can use the obtained α_k for approximating the function of an operator having its eigenvalues within the interval $[a,b]$.

Now, if we cannot use exact integration, we can use Gauss integration:

$$\alpha_k = \frac{1}{c_k} \sum_{n=1}^{N'} f(t_n)\, p_k(t_n)\, w_n, \quad k = 0, \ldots, N-1 \tag{4.6}$$

One can choose $N' = N$ or even $N' = 2N$ for higher precision.

5. Chebyshev Polynomials

The Chebyshev polynomials $T_n(x)$ are a family of polynomials, defined through trigonometric functions. For $|x| \le 1$ we set $x = \cos\theta$ and define

$$T_n(x) = T_n(\cos\theta) = \cos n\theta \tag{5.1}$$

Is this really a polynomial?? First we note:

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= x \end{aligned}$$

Now, using the trigonometric relation:

$$\cos(n+1)\theta = 2\cos\theta\cos n\theta - \cos(n-1)\theta$$

It is straightforward to unveil the polynomial nature: $T_{n+1}(x)$ can be defined in terms of lower degree T's :

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x) \tag{5.2}$$

Thus the Chebyshev polynomials can be generated, for example

$$\begin{aligned} T_2(x) &= 2x^2 - 1 \\ T_3(x) &= 4x^3 - 3x \\ T_4(x) &= 8x^4 - 8x^2 + 1 \end{aligned}$$

The orthogonality of the Chebyshev polynomials over the interval $[-1, 1]$ derives from:

$$\frac{2}{\pi}\int_0^{\pi} \cos(n\theta)\cos(m\theta)\,d\theta = (\delta_{n0}+1)\,\delta_{nm} \tag{5.3}$$

and by making the change of variable $x = \cos\theta$, we obtain:

$$\frac{2}{\pi}\int_{-1}^{1} \frac{T_n(x)\,T_m(x)}{\sqrt{1-x^2}}dx = \delta_{nm}(\delta_{n0}+1) \tag{5.4}$$

Hence, the weight $w(x) = \frac{2}{\pi}\frac{1}{\sqrt{1-x^2}}$ emerges.

The roots of the Chebyshev polynomial $T_N(x)$ are derived from the zeros of the cosine $\cos N\theta = 0$ which are

$$\theta_n^{(N)} = \frac{\pi}{N}\left(n - \frac{1}{2}\right) \tag{5.5}$$

where n is any integer. Hence, the Gaussian-Chebyshev sampling points, based on the $T_N(x)$ the are

$$x_n^{(N)} = \cos\theta_n^{(N)},\ \text{n=1,}\ldots N \tag{5.6}$$

The N quadrature weights weights a_n are determined from Eq. 3.1 one gets: $\sum_{n=1}^{N} a_n T_k\left(x_n^{(N)}\right) = 2\delta_{k0}$ or

$$\sum_{n=1}^{N} a_n^{(N)} \cos\left(k\theta_n^{(N)}\right) = 2\delta_{k0},\ k = 0,\ldots,N-1 \tag{5.7}$$

Due to the fact that This is a set of N equations and N unknowns. We note that for $k \neq 0$

$$\sum_{n=1}^{N} \cos\left(k\theta_n^{(N)}\right) = Re\sum_{n=0}^{N-1} e^{ik\frac{1}{N}\left(\frac{\pi}{2}+n\pi\right)} \tag{5.8}$$

$$= Re\left[e^{ik\frac{\pi}{2N}}\frac{(-)^k - 1}{2i\sin\frac{k\pi}{2N}}\right] \tag{5.9}$$

$$= 0 \tag{5.10}$$

this gives:

$$a_n^{(N)} = \frac{2}{N} \tag{5.11}$$

The quadrature is then :

$$\frac{2}{\pi}\int_{-1}^{1}\frac{f(x)}{\sqrt{1-x^2}}dx \approx \frac{2}{N}\sum_{n=1}^{N} f\left(\cos\theta_n^{(N)}\right) \tag{5.12}$$

6. Chebyshev approximation

Further reading: [1]

Suppose we have a set of functions $f(x;p)$ where x is a variable in the interval $[-1,1]$ and p is some set of parameters. We can use Chebyshev polynomials and write,

$$f(x;p) \approx \sum_{k=0}^{N-2} b_k(p)\,T_k(x) \tag{6.1}$$

where due to Eq. 5.4:

$$b_k(p) = \frac{1}{1+\delta_{k0}}\frac{2}{\pi}\int_{-1}^{1}\frac{T_k(x)\,f(x;p)}{\sqrt{1-x^2}}dx. \tag{6.2}$$

Then, using the Gaussian quadrature of with N terms in Eq. (5.12) we estimate b_k, $k=0,\ldots,N-2$ as:

$$b_k(p) = \frac{1}{1+\delta_{n0}}\frac{2}{N}\sum_{n=1}^{N}\cos\left(k\theta_n^{(N)}\right) f\left(\cos\theta_n^{(N)};p\right) \tag{6.3}$$

Note: The constant $b_k(p)$ can be calculated efficiently using a FFT. Denoting $f_n(p) = f\left(\cos\theta_n^{(N)};p\right)$, for real functions $f(x)$ We write Eq. (6.3) as:

$$b_k(p) = \frac{1}{1+\delta_{n0}}\frac{2}{N}Re\left\{e^{i\frac{\pi}{2N}k}\sum_{n=0}^{2N-1} e^{i\frac{2\pi}{2N}nk}\phi_n(p)\right\},\quad k=0,\ldots,N-2 \tag{6.4}$$

Where the sum can be done using a FFT, by setting

$$\phi_n = \begin{cases} f_{n+1} & 0\le n<N \\ 0 & N\le n<2N \end{cases} \tag{6.5}$$

so the sum in the last term is a FFT of $f_n(p)$ (for each value of parameters p an FFT needs to be performed) and all the b's for this set of parameters are obtained.

If $f(x) = f^r(x) + if^i(x)$, is complex then you don't have to separate to real and imaginary parts. Just do the following FFT (note the :

$$b_k(p) = \frac{1}{1+\delta_{k0}}\frac{1}{N}\left\{e^{i\frac{\pi}{2N}k}\sum_{n=0}^{2N-1} e^{i\frac{2\pi}{2N}nk}f_{n+1}(p)\right\},\quad k=0,\ldots,N-1. \tag{6.6}$$

This works well for both the real and the complex cases and for real functions gives exactly the same coefficients as Eq. 6.4.

7. Chebyshev approximation for a operator function

Based on ref.[2].

We now show how to apply $F_p\left(\hat{H}\right)$ on a given function ψ. Given the operator $\hat{H}$ we first need to estimate the interval $[E_{min}, E_{max}]$ which contains all of $\hat{H}$'s eigenvalues. Then we define a function:$b - a = E_{min}$, $b + a = E_{max}$

$$f_p(x) = F_p\left(x\Delta E + \bar{E}\right) \tag{7.1}$$

where:

$$\bar{E} = \frac{E_{max} + E_{min}}{2}, \ \Delta E = \frac{E_{max} - E_{min}}{2} \tag{7.2}$$

Now find the Chebyshev approximation for $f(x)$:

$$f_p(x) = \sum_{k=0}^{K} b_k(p) T_k(x) \tag{7.3}$$

Then, here's how to apply $F_p\left(\hat{H}\right)$ on any given ψ :

$$F_p\left(\hat{H}\right)\psi = \sum_{k=0}^{K} b_k(p)\psi_k \tag{7.4}$$

where:

$$\psi_0 = \psi, \quad \psi_1 = \hat{H}_N\psi_0, \ \psi_{k+1} = 2\hat{H}_N\psi_k + \psi_{k-1} \tag{7.5}$$

and

$$\hat{H}_N = \frac{\hat{H} - \bar{E}}{\Delta E} \tag{7.6}$$

Note that the same set of ψ's can be used for many parameter s p, and that you only need to remember 4 functions.

8. Stochastic wave functions and traces using Chebyshev moments

The methods here are based on refs. [3, 4, 5][4][5].

If we want to calculate the trace of an operator function, (see for examples Eq. (1.3) or Eq. (1.4)):

$$A_p = tr\left[F_p\left(\hat{H}\right)\right] \tag{8.1}$$

We first employ the stochastic trace formula, where we take a random wave function at grid-point $\mathbf{r}_g$,

$$\psi(\mathbf{r}_g) = \frac{e^{i\theta_g}}{\sqrt{\Delta V}} \tag{8.2}$$

where ΔV is the grid spacing/area/volume. It can be shown that the average of the projection operator of such wave functions give the unit operator:

$$\langle|\psi\rangle\langle\psi|\rangle = \mathbf{I} \tag{8.3}$$

where the angular brackets denote averaging with respect to θ. Combining this with Eq. (8.1) we obtain:

$$A_p = tr\left[\langle|\psi\rangle\langle\psi|\rangle F_p\left(\hat{H}\right)\right] = \left\langle\left\langle\psi\left|F_p\left(\hat{H}\right)\right|\psi\right\rangle\right\rangle \tag{8.4}$$

Using the Chebyshev approximation of the function in Eq. (7.4) we find:

$$A_p = \sum_{k=0}^{K-1} b_k(p) M_k \tag{8.5}$$

where:

$$M_k = tr\left[T_k\left(\hat{H}_N\right)\right] = \langle\langle\psi|\psi_k\rangle\rangle \tag{8.6}$$

is the k Chebyshev moment. Note that the main numeric effort goes to calculation of the K moments, which are just numbers. One can save half of the Hamiltonian applications by noticing that from the trigonometric relations $\cos(n+m)\theta + \cos(n-m)\theta = 2\cos n\theta\cos m\theta$, we have, for $-1 \le x \le 1$:

$$T_{n+m}(x) + T_{|n-m|}(x) = 2T_n(x)T_m(x). \tag{8.7}$$

Using this relation with $n = m$ (we omit (x)):

$$T_{2n} = 2T_n^2 - 1 \tag{8.8}$$

and with $n = m + 1$ we find:

$$T_{2n+1} = 2T_nT_{n+1} - T_1. \tag{8.9}$$

Hence, applying this for the operator $\hat{H}_N$ instead of x, for even k, i.e. $k = 2n$, we find:

$$M_{2n} = 2\langle\langle\psi_n|\psi_n\rangle\rangle - M_0 \tag{8.10}$$

and for odd k, i.e. $k = 2n+1$, we find:

$$M_{2n+1} = 2\langle\langle\psi_n|\psi_{n+1}\rangle\rangle - M_1 \tag{8.11}$$

Clearly we only have to compute ψ_n's for $n = 0, \ldots, K/2$.

9. Algorithm for calculating the trace of $F\left(\hat{H}; p\right)$

The following steps are necessary:

1. Note that p is a set of parameters, like β, μ for the electron number function N_e
2. Prepare a grid representation of wave functions and the Hamiltonian operator
3. Estimate E_{min} and E_{max} (see: section 7).
4. Define $\hat{H}_N$ (Eq. 7.6).
5. For a given set of parameters p prepare the Chebyshev coefficients $b_k(p)$ $k = 0, \ldots, K-1$. K (see Eq. (6.6)) has to be even and large enough so that all $b_k(p)$ with $k > K$ are negligible (e.g. less than 10^{-8}).
6. Generate the moments M_k $k = 0, \ldots, K-1$:
   (a) Set
      (i) $M_0 = N_g$
      (ii) for all $k = 1, \ldots, K-1$ $M_k = 0$
   (b) Generate a stochastic wave function ψ_0 on the grid (Eq. 8.2). Note it must be zero on the grid boundaries.
   (c) Set

(i) $\psi_1 = \hat{H}_N \psi_0$
(ii) $M_1 = \langle \psi_0 | \psi_1 \rangle$,
(iii) $M_2 = 2 \langle \langle \psi_1 | \psi_1 \rangle \rangle - M_0$;

(d) For $n = 2, 3, ...$ to $n = K/2$ (remember K is even):
(i) $\psi_2 = 2\hat{H}_N \psi_1 - \psi_0$
(ii) $M_{2n-1} = M_{2n-1} + (2 \langle \psi_{n-1} | \psi_n \rangle - M_1)$
(iii) $M_{2n} = M_{2n} + (2 \langle \psi_n | \psi_n \rangle - M_0)$,
(iv) $\psi_0 = \psi_1$, $\psi_1 = \psi_2$

(e) Average M_k over I stochastic functions (see Eq. 8.2 on page 7). The statistical error should decrease in proportion to $I^{-1/2}$.

(7) Calculate the trace as: $tr\left[F\left(\hat{H}; p\right)\right] = \sum_{k=0}^{K-1} b_k(p) M_k$.

References

[1] T. J. Rivlin. *Chebyshev Polynomials: From approximation Theory to Algebra and Numbers Theory.* Wiley, New-York, 1990.

[2] R. Kosloff. Time-dependent quantum-mechanical methods for molecular- dynamics. *J. Phys. Chem.*, 92(8):2087–2100, 1988.

[3] D. A. Drabold and O. F. Sankey. Maximum-entropy approach for linear scaling in the electronic- structure problem. *Phys. Rev. Lett.*, 70(23):3631–3634, 1993.

[4] RN Silver and H Röder. Calculation of densities of states and spectral functions by chebyshev recursion and maximum entropy. *Phys. Rev. E*, 56(4):4822, 1997.

[5] Roi Baer, Daniel Neuhauser, and Eran Rabani. Self-averaging stochastic kohn-sham density-functional theory. *Phys. Rev. Lett.*, 111:106402, Sep 2013.

Fritz Haber Research Center for Molecular Dynamics, Institute of Chemistry, The Hebrew University of Jerusalem