Simple Algorithms for Spectral Sum and Spectrum Approximation

Tyler Chen, Shashanka Ubaru, Thomas Trogdon

Introduction

The eigenvalues $\{\lambda_i\}_{i=1}^n$ of a $n \times n$ Hermitian matrix $\mathbf{A}$ are of extreme importance in a applications from nearly every area of science. However, computing the eigenvalues exactly is expensive.

In many settings, however, partial spectral information is still enough to compute a sufficiently accurate approximation to the quantity of interest. One task where this is frequently the case is computing a spectral sum

$$\operatorname{tr}(f[\mathbf{A}]) = \sum_{i=1}^n f(\lambda_i)$$

where $f : \mathbb{R} \to \mathbb{R}$ is a real valued function and $f[\mathbf{A}]$ is the induced matrix function.

Applications

A number of important quantities with applications in a wide range of disciplines can be written in terms of spectral sums. These quantities include the log-determinant, Shatten p-norms, the trace of the inverse matrix, the Estrada index, number of triangles in a graph and other graph motifs, etc.

A particularly important spectral sum corresponds to the indicator function $\mathbb{1}[\cdot \leq x]$ and induces the cumulative emperical spectral measure (CESM):

$$\Phi(x) := \sum_{i=1}^n \frac{1}{n} \mathbb{1}[\lambda_i \leq x].$$

The CESM contains information about the spectrum of $\mathbf{A}$.

The CESM is used in computational physics and chemistry to study spectral properties of various observables. In machine learning, approximate CESMs have found use in facilitating backpropagation through implicit likelihoods as well as for studying algorithms and models which are too large to inspect directly. This provides insight into differences between various training approaches and/or network architectures.

For a more complete overview of applications we refer readers to [UCS17, HMAS17, CTU21] and there references within.

Intuition

Define the weighted CESM $\Psi : \mathbb{R} \to [0,1]$ by

$$\Psi(x) := \mathbf{v}^\mathsf{T} \mathbb{1}[\mathbf{A} \leq x] \mathbf{v}, \qquad \mathbf{v} \sim \mathcal{U}(\mathbb{S}^{n-1}),$$

where $\mathcal{U}(\mathbb{S}^{n-1})$ is the uniform distribution on the unit hypersphere. Then $\mathbb{E}[\Psi(x)] = \Phi(x)$ and $\mathbb{E}[\int f \mathrm{d}\Psi] = \mathbb{E}[\mathbf{v}^\mathsf{T} f(\mathbf{A}) \mathbf{v}_\ell] = n^{-1} \operatorname{tr}(f[\mathbf{A}])$.

Let $\{\Psi_\ell\}_{\ell=1}^{n_\mathrm{v}}$ be identically distributed (iid) samples of the weighted CESM Ψ corresponding to vectors $\{\mathbf{v}_\ell\}_{\ell=1}^{n_\mathrm{v}}$ which are iid samples of $\mathbf{v}$. Then the average weighted CESM $\langle \Psi_\ell \rangle := \frac{1}{n_\mathrm{v}} \sum_{\ell=1}^{n_\mathrm{v}} \Psi_\ell$ is also an unbiased estimator of Φ and

$$\langle (\mathbf{v}_\ell)^\mathsf{T} f(\mathbf{A}) \mathbf{v} \rangle := \frac{1}{n_\mathrm{v}} \sum_{\ell=1}^{n_\mathrm{v}} (\mathbf{v}_\ell)^\mathsf{T} f(\mathbf{A}) \mathbf{v}_\ell = \int f \mathrm{d} \langle \Psi_\ell \rangle.$$

is an unbiased estimator for $n^{-1} \operatorname{tr}(f[\mathbf{A}])$.

In order to approximate a sample of Ψ (and therefore integrals against such samples) we use the information contained in the Krylov subspace

$$\mathcal{K}_k[\mathbf{A}, \mathbf{v}] := \operatorname{span}\{\mathbf{v}, \mathbf{A}\mathbf{v}, \ldots, \mathbf{A}^k \mathbf{v}\} = \{p(\mathbf{A})\mathbf{v} : \deg(p) \leq k\}.$$

A basis for the Krylov subspace $\mathcal{K}_k[\mathbf{A}, \mathbf{v}]$ can be computed using k matrix-vector products with $\mathbf{A}$ and it contains a wealth of information about the interaction of $\mathbf{A}$ with $\mathbf{v}$. In particular, such a basis contains the information necessary to compute the polynomial moments of Ψ through degree $2k$ (although it is sometimes more straightforward to obtain the moments through degree s, for some $s \leq 2k$).

Given access to the moments of Ψ through degree s, a natural approach to obtaining an approximation $[\Psi]_s^{\circ\mathrm{q}}$ to Ψ is through the definition

$$\int f \mathrm{d}[\Psi]_s^{\circ\mathrm{q}} := \int [f]_s^{\circ\mathrm{p}} \mathrm{d}\Psi$$

where $[\,\cdot\,]_s^{\circ\mathrm{p}}$ is an operator such that $[f]_s^{\circ\mathrm{p}}$ is a *polynomial* of degree at most s which approximates a continuous function f in a suitable way. indeed, integrals of polynomials against Ψ can be computed using just the moments of Ψ.

Algorithms falling into this framework include Stochastic Lanczos Quadrature and the Kernel Polynomial Method. This approach provides a *unified* perspective.

Algorithm

The prototypical algorithm is a few lines of pseudocode:

for $\ell = 1, 2, \ldots, n_\mathrm{v}$ **do**
  sample $\Psi_\ell \overset{\text{iid}}{\sim} \Psi$ by sampling $\mathbf{v}_\ell \overset{\text{iid}}{\sim} \mathcal{U}(\mathbb{S}^{n-1})$
  compute moments of Ψ_ℓ through deg. s by constructing $\mathcal{K}_k[\mathbf{A}, \mathbf{v}_\ell]$
  approximate Ψ_ℓ by $[\Psi_\ell]_s^{\circ\mathrm{q}}$ induced by a polynomial operator $[\,\cdot\,]_s^{\circ\mathrm{p}}$
return $\langle [\Psi_\ell]_s^{\circ\mathrm{q}} \rangle = \frac{1}{n_\mathrm{v}} \sum_{\ell=1}^{n_\mathrm{v}} [\Psi_\ell]_s^{\circ\mathrm{q}}$

Analysis

To analyze the algorithm, simply apply the triangle inequality several times:

$$\begin{aligned}\left| \int f \mathrm{d}\big(\Phi - \langle [\Psi_\ell]_s^{\circ\mathrm{q}} \rangle\big) \right| &\leq \left| \int f \mathrm{d}\big(\Phi - \langle \Psi_\ell \rangle\big) \right| + \left| \int f \mathrm{d}\big(\langle \Psi_\ell \rangle - \langle [\Psi_\ell]_s^{\circ\mathrm{q}} \rangle\big) \right| \\ &\leq \left| \int f \mathrm{d}\big(\Phi - \langle \Psi_\ell \rangle\big) \right| + \left\langle \left| \int f \mathrm{d}\big(\Psi_\ell - [\Psi_\ell]_s^{\circ\mathrm{q}}\big) \right| \right\rangle.\end{aligned}$$

We have now decoupled the analysis into understanding:
– convergence of $\langle \Psi_\ell \rangle$ to $\Phi[\mathbf{A}]$ in terms of n_v
– for each ℓ, convergence of $[\Psi_\ell]_s^{\circ}$ to Ψ_ℓ in terms of s.

The first term can be characterized exactly [CTU21].

The second term can be bounded using standard tools from approximation theory.

Theorem. Suppose $f \in \mathsf{Lip}(L)$. Set $n_\mathrm{v} \geq \left(\frac{4I[\mathbf{A}]^2 L^2}{n+2}\right) \varepsilon^{-2} \ln(2n\eta^{-1})$ and

$$s \geq \begin{cases} \frac{1}{\ln(\rho)} \ln\left(\frac{4c_2 M}{\rho - 1} \cdot \varepsilon^{-1}\right) & f \in \mathsf{Anl}(\rho, M) \\ d + \left(\frac{2c_2c_3V}{\pi d}\right)^{1/d} \cdot \varepsilon^{-1/d} & f \in \mathsf{BV}(d, V) \\ \frac{\pi^2 c_1 c_3 L}{2} \cdot \varepsilon^{-1} & f \in \mathsf{Lip}(L) \end{cases}$$

Then the output $\langle [\Psi_\ell]_s^{\circ\mathrm{q}} \rangle$ satisfies

$$\mathbb{P}\left[\left| n^{-1} \operatorname{tr}(f[\mathbf{A}]) - \int f \mathrm{d} \langle [\Psi_\ell]_s^{\circ\mathrm{q}} \rangle \right| > \varepsilon \right] \leq \eta.$$

Here c_1, c_2, c_3 are small constants which differ slightly for various choices of $[\,\cdot\,]_s^{\circ\mathrm{p}}$.