Hermite Polynomial Features for Private Data Generation

Seongho Joo

SNU MILAB

Warm-up

Definition 1 (Gegenbauer polynomial)

The Gegenbauer polynomial of degree $\ell \geq 0$ in dimension $d \geq 2$ is given by

$$P_d^\ell(t) := \sum_{j=0}^{\lfloor \ell/2 \rfloor} c_j \cdot t^{\ell-2j} \cdot (1-t^2)^j$$

where $c_0 = 1$ and $c_{j+1} = -\frac{(\ell-2j)(\ell-2j-1)}{2(j+1)(d-1+2j)} c_j$ for $j = 0, 1, \ldots, \lfloor \ell/2 \rfloor - 1$.

Chebyshev polynomials $(d=2)$, Legendre polynomials $(d=3)$, monomials $(d=\infty)$.

1. (Orthogonality)

$$\int_{-1}^{-1} P_d^\ell(t) P_d^{\ell'}(t)(1-t^2)^{\frac{d-3}{2}} \,\mathrm{d}t = \frac{\left|\mathbb{S}^{d-1}\right| \mathbf{1}_{\{\ell=\ell'\}}}{\alpha_{\ell,d} \cdot \left|\mathbb{S}^{d-2}\right|}$$

2. (Reproducing property) For any $x, y, \in \mathbb{S}^{d-1}$,

$$P_d^\ell(\langle x, y\rangle) = \alpha_{\ell,d} \cdot \mathop{\mathbb{E}}_{w \sim \mathcal{U}(\mathbb{S}^{d-1})} \left[P_d^\ell(\langle x, w\rangle) P_d^\ell(\langle y, w\rangle) \right].$$

Warm-up

Definition 2 (Generalized zonal kernels)

For an integers $s \geq 1$ and a sequence of vector-valued functions $h_\ell : \mathbb{R} \to \mathbb{R}^s$ for $\ell = 0, l1, \dots,$ we define the generalized zonal kernel (GZK) of order s as

$$k(x, y) := \sum_{\ell=0}^{\infty} \langle h_\ell(\|x\|), h_\ell(\|y\|) \rangle \, P_d^\ell \left(\frac{\langle x, y \rangle}{\|x\| \, \|y\|} \right)$$

Examples: dot-product, Gaussian, and neural tangent kernels.

Spectral Approximation of GZK

Let ϕ_{x_j} be the feature map defined by

$$\phi_{x_j}(w) := \sum_{\ell=0}^{\infty} h_\ell(\|x_j\|) P_d^\ell\left(\frac{\langle x, w\rangle}{\|x_j\|}\right) \in \mathbb{R}^s.$$

Question: Can we define an operator $\boldsymbol{\Phi} : \mathbb{R}^n \to L^2(\mathbb{S}^{d-1}, \mathbb{R}^s)$ such that $\boldsymbol{\Phi}^* \boldsymbol{\Phi} = \boldsymbol{K}$?

Spectral Approximation of GZK

Let ϕ_{x_j} be the feature map defined by

$$\phi_{x_j}(w) := \sum_{\ell=0}^{\infty} h_\ell(\|x_j\|) P_d^\ell\left(\frac{\langle x, w\rangle}{\|x_j\|}\right) \in \mathbb{R}^s.$$

Question: Can we define an operator $\boldsymbol{\Phi} : \mathbb{R}^n \to L^2(\mathbb{S}^{d-1}, \mathbb{R}^s)$ such that $\boldsymbol{\Phi}^* \boldsymbol{\Phi} = \boldsymbol{K}$?

Define $\boldsymbol{\Phi}$ (a.k.a. quasi-matrix) as follows:

$$\boldsymbol{\Phi} \cdot v := \sum_{j=1}^{n} v_j \cdot \phi_{x_j}$$

Then adjoint of this operator $\boldsymbol{\Phi}^* : L^2(\mathbb{S}^{d-1}, \mathbb{R}^s) \to \mathbb{R}^n$ is the following for $f \in L^2(\mathbb{S}^{d-1}, \mathbb{R}^s)$ and $j \in [n]$

$$[\phi^* f]_j = \left\langle \phi_{x_j}, f \right\rangle_{L^2(\mathbb{S}^{d-1}, \mathbb{R}^s)}$$

With this definition, it follows that $\boldsymbol{\Phi}^* \boldsymbol{\Phi} \overset{(a)}{=} \boldsymbol{K}$.

(a): $\mathbb{E}_{w \sim \mathcal{U}}[\phi_{x_i}(w) \phi_{x_j}(w)] = K_{ij}$

Spectral Approximation of GZK

The approach for spectrally approximating $\boldsymbol{K}$ is sampling the rows of the quasi-matrix $\mathbf{\Phi}$ with probabilities proportional to their ridge leverage scores. The ridge leverage scores of $\mathbf{\Phi}$ are defined as follows:

Spectral Approximation of GZK

The approach for spectrally approximating $\boldsymbol{K}$ is sampling the rows of the quasi-matrix $\boldsymbol{\Phi}$ with probabilities proportional to their ridge leverage scores. The ridge leverage scores of $\boldsymbol{\Phi}$ are defined as follows:

Definition 3 (Ridge leverage scores Φ)

Let $\boldsymbol{\Phi} : \mathbb{R}^n \to L^2(\mathbb{S}^{d-1}, \mathbb{R}^s)$ be the operator. Also,for every $w \in \mathbb{S}^{d-1}$, define $\boldsymbol{\Phi}_w \in \mathbb{R}^{n \times s}$ as,

$$\Phi_w := [\phi_{x_1}(w), \phi_{x_2}(w), \dots \phi_{x_n}(w)]^\top. \tag{0.1}$$

For any $\lambda > 0$, the row leverage scores of $\boldsymbol{\Phi}$ are defines as,

$$\tau_{\lambda(w)} := \mathrm{Tr}(\Phi_w^\top \cdot (\boldsymbol{K} + \lambda \boldsymbol{I})^{-1} \Phi_w).$$

An important quantity for the spectral approximation to $\boldsymbol{K}$ is the average of the ridge leverage scores with respect to the uniform distribution on $\mathbb{S}^{d-1}$ which is equal to *statistical dimension:*

$$\mathop{\mathbb{E}}_{w \sim \mathcal{U}(\mathbb{S}^{d-1})} [\tau_\lambda(w)] = \mathrm{Tr}(\boldsymbol{K}(\boldsymbol{K} + \lambda I)^{-1}) = s_\lambda$$

Upper bound on leverage scores of GZK

For any dataset $\boldsymbol{X} = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{d\times n}$, let $\boldsymbol{\Phi}$ be the feature operator for the order s GZK on $\boldsymbol{X}$. For any $\lambda > 0$ and $w \in \mathbb{S}^{d-1}$, the ridge leverage scores of $\boldsymbol{\Phi}$ are uniformly upper bounded by

$$\tau_\lambda(w) \leq \sum_{\ell=0}^{\infty} \alpha_{\ell,d} \min \left\{ \frac{\pi^2(\ell+1)^2}{6\lambda} \sum_{j\in[n]} \|h_\ell(\|x_j\|)\|^2, s \right\}$$

Definition 4 (Random feature for Generalized Zonal Kernels)

For any GZK and dataset $\boldsymbol{X}^{d\times n}$, sample i.i.d point $w_1, \ldots, w_m \sim \mathcal{U}(S^{d-1})$ and let $\phi_{w_1}, \ldots, \phi_{w_m} \in \mathbb{R}^{n\times s}$ be defined as the previous. Then define the features matrix $\boldsymbol{Z} \in \mathbb{R}^{(m\times s)\times n}$:

$$\boldsymbol{Z} := \frac{1}{\sqrt{m}} \cdot [\phi_{w_1}, \ldots, \phi_{w_m}]^\top$$

These random features are *unbiased*, i.e., $\mathbb{E}[\boldsymbol{Z}^\top \boldsymbol{Z}] = \boldsymbol{K}$

s 는 kernel element-wise approximation을 위해서 m은 전체 approximation을 위해서?

About proof of Thm 9.

Suppose

$$\left\|\boldsymbol{Z}^\top\boldsymbol{Z}-\boldsymbol{K}\right\|_{\text{op}} \le \varepsilon \left\|\boldsymbol{K}+\lambda\boldsymbol{I}\right\|_{\text{op}}$$

Then RHS is $\varepsilon\Sigma^2$ by the assumption, then

$$\iff \left\|\Sigma^{-2}\right\|_{\text{op}}\left\|\boldsymbol{Z}^\top\boldsymbol{Z}-\boldsymbol{K}\right\|_{\text{op}} \le \varepsilon \tag{0.2}$$

$$\iff \left\|\Sigma^{-2}\right\|_{\text{op}}\left\|V(\boldsymbol{Z}^\top\boldsymbol{Z}-\boldsymbol{K})V^\top\right\|_{\text{op}} \le \varepsilon \tag{0.3}$$

$$\iff \left\|\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\cdot\boldsymbol{Z}^\top\boldsymbol{Z}\cdot\boldsymbol{V}^\top\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\cdot\boldsymbol{K}\cdot\boldsymbol{V}^\top\boldsymbol{\Sigma}^{-1}\right\|_{\text{op}} \le \varepsilon \tag{0.4}$$

부정확, 괜히 복잡하게 생각함

About proof of Thm 9.

The spectral approximation condition is equivalent to

$$-\varepsilon(\boldsymbol{K}+\lambda\boldsymbol{I}) \preceq \boldsymbol{Z}^\top\boldsymbol{Z}-\boldsymbol{K} \preceq \varepsilon(\boldsymbol{K}+\lambda\boldsymbol{I})$$

Left multiply by $\boldsymbol{\Sigma}^{-1}\boldsymbol{V}$ and right multiply by $\boldsymbol{V}^\top\boldsymbol{\Sigma}^{-1}$

$$-\varepsilon\boldsymbol{I} \preceq \boldsymbol{\Sigma}^{-1}V(\boldsymbol{Z}^\top\boldsymbol{Z}-\boldsymbol{K})\boldsymbol{V}\boldsymbol{\Sigma}^{-1} \preceq \varepsilon\boldsymbol{I}$$

$$\iff \left\|\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\cdot\boldsymbol{Z}^\top\boldsymbol{Z}\cdot\boldsymbol{V}^\top\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\boldsymbol{V}\cdot\boldsymbol{K}\cdot\boldsymbol{V}^\top\boldsymbol{\Sigma}^{-1}\right\|_{\mathsf{op}} \leq \varepsilon$$

About proof of Thm 9.

Lemma 1 (Tail bound for matrix approximation in operator norm)

Let $\boldsymbol{B}$ be a fixed $n \times n$ matrix. Construct an $n \times n$ matrix $\boldsymbol{R}$ that, almost surely, satisfies,

$$\mathbb{E}[\boldsymbol{R}] = \boldsymbol{B} \text{ and } \|\boldsymbol{R}\|_{\text{op}} \leq L.$$

Let $\boldsymbol{B}$ be a fixed $n \times n$ matrix. Construct an $n \times n$ matrix $\boldsymbol{R}$ that, almost surely, satisfies Let $\boldsymbol{M}_1$ and $\boldsymbol{M}_2$ be semi-definite upper bounds for the expected squares

$$\mathbb{E}[\boldsymbol{R}\boldsymbol{R}^*] \preceq \boldsymbol{M}_1, \quad \mathbb{E}[\boldsymbol{R}^*\boldsymbol{R}] \preceq \boldsymbol{M}_2$$

Define the quantities $M = \max\left\{\|M_1\|_{\text{op}}, \|M_2\|_{\text{op}}\right\}$. Form the matrix sampling estimator

$$\bar{\boldsymbol{R}} = \frac{1}{m}\sum_{j=1}^{m} \boldsymbol{R}_j$$

where each $\boldsymbol{R}_j$ is an independent copy of $\boldsymbol{R}$. Then,

$$\Pr[\|\bar{\boldsymbol{R}} - \boldsymbol{B}\|_{\text{op}} \geq \varepsilon] \geq 4 \cdot \frac{\operatorname{Tr}(\boldsymbol{M}_1 + \boldsymbol{M}_2)}{\boldsymbol{M}} \cdot \exp\left(\frac{-m\varepsilon^2/2}{M + 2L\varepsilon/3}\right).$$

Matrix의 second momentt에 대한 bound가 있어야 함

Ridge leverage score

Lemma 2 (Minimization characterization of ridge leverage scores)

For any $\lambda > 0$, let $\mathbf{\Phi}$ be the operator with leverage score $\tau_\lambda(\cdot)$. Let Φ_w^i denote the i^{th} column of the matrix $\Phi_w \in \mathbb{R}^{n \times s}$ for any $i \in [s]$, the following holds,

$$\tau_\lambda(w) = \sum_{i \in [s]} \left(\min_{g_i \in L^2(\mathbb{S}^{d-1}, \mathbb{R}^s)} \|g_i\|_{L^2(\mathbb{S}^{d-1}, \mathbb{R}^s)} + \lambda^{-1} \left\| \mathbf{\Phi}^* g_i - \Phi_w^i \right\|_2^2 \right) \quad \text{for } w \in \mathbb{S}^{d-1}$$

Lemma 7의 증명을 위해서는 $g_w^i(\sigma) := \left(\sum_{\ell=0}^{\infty} \alpha_{\ell,d} \mathbf{1}_{\{R_\ell \geq \mu\}} \cdot P_d^\ell(\langle \sigma, w \rangle)\right) \cdot e_i$ 로 setting, where

$$R_\ell := \frac{(\ell+1)^2}{n} \cdot \sum_{j \in [n]} \|h_\ell(\|x_j\|)\|^2, \quad \text{for } \ell = 0, 1, 2, \ldots$$

$$\mu := \frac{6\lambda s}{\pi^2 n}$$

About the proof of dot-product kernel

Why

$$(1 - 8\varepsilon/10) \cdot (\widetilde{\boldsymbol{K}} + \lambda \boldsymbol{I}) \preceq \boldsymbol{Z}^\top \boldsymbol{Z} + \lambda \boldsymbol{I} \preceq (1 + 8\varepsilon/10) \cdot (\widetilde{\boldsymbol{K}} + \lambda \boldsymbol{I})$$

with $\left\| \widetilde{\boldsymbol{K}} - \boldsymbol{K} \right\|_F \le \frac{\varepsilon \lambda}{10}$ implies (ε, δ)-spectral approximation?

$$(1 - \varepsilon)(\boldsymbol{K} + \lambda \boldsymbol{I}) \preceq \boldsymbol{Z}^\top \boldsymbol{Z} + \lambda \boldsymbol{I} \preceq (1 + \varepsilon)(\boldsymbol{K} + \lambda \boldsymbol{I})$$

About the proof of dot-product kernel

Why

$$(1 - 8\varepsilon/10) \cdot (\widetilde{\boldsymbol{K}} + \lambda \boldsymbol{I}) \preceq \boldsymbol{Z}^\top \boldsymbol{Z} + \lambda \boldsymbol{I} \preceq (1 + 8\varepsilon/10) \cdot (\widetilde{\boldsymbol{K}} + \lambda \boldsymbol{I})$$

with $\left\| \widetilde{\boldsymbol{K}} - \boldsymbol{K} \right\|_F \leq \frac{\varepsilon\lambda}{10}$ implies (ε, δ)-spectral approximation?

$$(1 - \varepsilon)(\boldsymbol{K} + \lambda \boldsymbol{I}) \preceq \boldsymbol{Z}^\top \boldsymbol{Z} + \lambda \boldsymbol{I} \preceq (1 + \varepsilon)(\boldsymbol{K} + \lambda \boldsymbol{I})$$

Not hard to show: using that $\|\boldsymbol{A}\|_{\mathsf{op}} \leq \|\boldsymbol{A}\|_F$ for matrix $\boldsymbol{A}$, and start from the

$$(1 - 8\varepsilon/10)(K + \lambda \boldsymbol{I}) + (1 - 8\varepsilon/10)(\tilde{\boldsymbol{K}} - \boldsymbol{K}) \preceq (1 - \varepsilon)(K + \lambda \boldsymbol{I})$$

About the proof of dot-product kernel

$$
\begin{aligned}
\left|\sum_{\ell>q}\left(\sum_{i=0}^{\infty}\tilde{h}_{\ell,i}(\|x\|)\tilde{h}_{\ell,i}(\|y\|)\right)\cdot P_d^{\ell}\left(\frac{\langle x,y\rangle}{\|x\|\|y\|}\right)\right| &\le \frac{C_\kappa\cdot\Gamma\left(\frac{d}{2}\right)\cdot e^{-r^2\beta_\kappa}}{4\cdot(d-1)!}\cdot\sum_{\ell>q}\frac{(\ell+d-1)!}{2^\ell\cdot\ell!}\cdot\frac{(r^2\beta_\kappa)^\ell}{\Gamma\left(\ell+\frac{d}{2}\right)} \\
&\le \frac{C_\kappa\cdot\Gamma\left(\frac{d}{2}\right)\cdot e^{-r^2\beta_\kappa}}{4\cdot(d-1)!}\cdot\sum_{\ell>q}\frac{1}{\ell^{\ell-d/2}}\cdot\left(\frac{e\cdot r^2\beta_\kappa}{2}\right)^{\ell}\cdot\left(1+\frac{d-1}{\ell}\right)^{d/2} \\
&\le \frac{C_\kappa\cdot\Gamma\left(\frac{d}{2}\right)\cdot 2^{\frac{d}{2}}\cdot e^{r^2\beta_\kappa}}{5\cdot(d-1)!}\cdot\sum_{\ell>q}\frac{1}{\ell^{\ell-d/2}}\cdot\left(\frac{e\cdot r^2\beta_\kappa}{2}\right)^{\ell} \\
&\le \frac{C_\kappa\cdot e^{r^2\beta_\kappa}}{20(d/2)d/2}\cdot\sum_{\ell>q}\frac{1}{\ell^{\ell-d/2}}\cdot\left(\frac{e\cdot r^2\beta_\kappa}{2\ell}\right)^{\ell} \\
&\le \frac{C_\kappa\cdot e^{r^2\beta_\kappa}}{20}\cdot\left(\frac{e\cdot r^2\beta_\kappa}{d}\right)^{d/2}\cdot\sum_{\ell>q}\left(\frac{e\cdot r^2\beta_\kappa}{2\ell}\right)^{\ell-d/2} \\
&\le \frac{\varepsilon\lambda}{20n}.
\end{aligned}
$$

아마도 $\ell >> d$ 가정 들어간듯, $n! = \Theta(\sqrt{n}(n/e)^n)$

About the proof of dot-product kernel

Gautschi inequality.

$$x^{1-s} < \frac{\Gamma(x+1)}{\Gamma(x+s)} < (x+1)^{1-s}, \quad x > 0, 0 < s < 1$$

$$\text{Tail term} \le \frac{C_k \cdot e^{r^2\beta_k}}{20} \cdot \left(\frac{e \cdot r^2\beta_k}{d}\right)^{d/2} \sum_{\ell > q} \left(\frac{e \cdot r^2\beta_k}{2\ell}\right)^{\ell - \frac{d}{2}}$$

$$\le \frac{C_k \cdot e^{r^2\beta_k}}{20} \cdot \left(\frac{e \cdot r^2\beta_k}{d}\right)^{d/2} \sum_{\ell > q} \left(\frac{e \cdot r^2\beta_k}{2\ell}\right)^{\ell} \cdot \left(\frac{1}{\ell}\right)^{-\frac{d}{2}} \cdot \left(\frac{1}{2}\right)^{-\frac{d}{2}}$$

$$\ell \leftarrow 3.7r^2\beta_k \quad \ell \leftarrow r^2\beta_k + \frac{d}{2}\log\frac{3r^2\beta_k}{d} + \log\frac{C_\kappa n}{\varepsilon\lambda}$$

정확히는 ℓ을 대입하고 나서 summation 안에 term이 $e^{-\ell}$ 이 되니까 실제 q보다 작아서 upper-bound를 잡게 됨.

About the proof of dot-product kernel

Frullani integral

$$\int_0^{\infty} \frac{f(ax) - f(bx)}{x} \, \mathrm{d}x = [f(\infty) - f(0)] \log \frac{a}{b}$$

적분놀이

$$
\begin{aligned}
\int_0^\infty \frac{1-e^{-x}(1+x)}{x(e^x-1)(e^x+e^{-x})}\,\mathrm{d}x &= \int_0^\infty \frac{e^x-(1+x)}{x(e^x-1)(e^{2x}+1)}\,\mathrm{d}x \\
&= \int_0^\infty \sum_{k=2}^\infty \frac{x^{k-1}}{k!} \sum_{j=1}^\infty \left(\frac{1}{k(4j-1)^k} + \frac{1}{k(4j)^k} \right) \\
&\overset{a}{=} \sum_{j=1}^\infty \sum_{k=2}^\infty \left(\frac{1}{k(4j-1)^k} + \frac{1}{k(4j)^k} \right) \\
&\overset{b}{=} \sum_{j=1}^\infty \left[\log\left(\frac{4j-1}{4j-2} \right) - \frac{1}{4j-1} \right] + \sum_{j=1}^\infty \left[\log\left(\frac{4j}{4j-1} \right) - \frac{1}{4j} \right]
\end{aligned}
$$

(a) : perform the integral using the Gamma integral

(b) : $\log(1-x) = -\sum_{r=1}^\infty \frac{x^r}{r}$ for $-1 \le x < 1$

(c) : $\prod_{k=1}^{n-1}(k+x) = \frac{\Gamma(n+x)}{\Gamma(1+x)}$, gautschi's inequality

Stochastic Chebyshev Gradient for Spectral Optimization

Lemma. Then, for any $\boldsymbol{v} \in \mathbb{R}^d$, it holds that

$$\nabla_\theta \boldsymbol{v}^\top p_n(\boldsymbol{A})\boldsymbol{v} = 2\sum_{i=0}^{n-1}(2 - \mathbf{1}_{i=0})\boldsymbol{w}_i \left(\sum_{j=1}^{n-1} b_{j+1}\boldsymbol{y}_{j-i}\right)^\top \theta,$$

where $\boldsymbol{w}_{j+1} = 2\boldsymbol{A}\boldsymbol{w}_j - \boldsymbol{w}_{j-1}, \boldsymbol{w}_1 = \boldsymbol{A}\boldsymbol{v}, \boldsymbol{w}_0 = \boldsymbol{v}$ and $\boldsymbol{y}_{j+1} = 2\boldsymbol{w}_{j+1} + \boldsymbol{y}_{j-1}, \boldsymbol{y}_1 = 2\boldsymbol{A}\boldsymbol{v}, \boldsymbol{y}_0 = \boldsymbol{v}$

$\sum_{j=0}^{\infty}(2 - \mathbf{1}_{k=0}\boldsymbol{w}_k\boldsymbol{y}_{j-k}^\top) = \sum_{k=0}^{j}(2 - \mathbf{1}_{k=0}\boldsymbol{y}_{j-k}\boldsymbol{w}_k^\top) \leftarrow$ 이거 진작에 $2w_1 = y_1, w_0 = y_0$ 대입하니까 쉽네

Trace estimator

Lemma.

$$\mathrm{Var}_{\boldsymbol{v}}[\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v}] = 2\left(\|\boldsymbol{A}\|_F^2 - \sum_{i=1}^{d} \boldsymbol{A}_{ii}^2\right) \leq 2\,\|\boldsymbol{A}\|_F^2$$

for Rademacher random variable $\boldsymbol{v} \in [-1,1]^d$ and $\boldsymbol{A} \in \mathcal{S}^{d\times d}$.

For Frobenius norm

$$\sum_{i=1}^{d'} \left\| \frac{\partial \boldsymbol{A}}{\partial \theta_i} \right\|_F^2 = \left\| \frac{\partial \boldsymbol{A}}{\partial \theta} \right\|_F^2$$

right?

How to deal with infinite-dimensional programming?

문제: KKT theorem을 infinite-dimensional problem에 적용할 수 없음.

1. 먼저 Finite version의 solution를 구한다. Solution의 limit q^* 를 구한다.
2. Objective function이 continuous 한 것을 보이고, feasible set이 non-decreasing set인 것을 보인다.
3. Berge의 maximum theorem에 의해서 finite version의 minimum이 infinite version으로 converge 한다.
4. Step 1에서 구한 q^* 가 minimizer라는 것을 보인다.

Lemma

Lemma 3 (Weighted Regularity Bounds for Modified Chebyshev Coefficients)

Suppose that q_n^* is the optimal degree distribution and b_j is the Chebyshev coefficient of the analytic function f. Define the weighted coefficient $\hat{b}_j$ as $\hat{b}_j = b_j/(1 - \sum_{i=0}^{j-1} q_i^*)$ for $j \geq 0$ (with convention $q_{-1}^* = 0$). Then, there exists constants $D_1', D_2' > 0$ independent of M, N such that

$$\sum_{n=1}^{\infty} q_n \left(\sum_{j=1}^{\infty} |\hat{b}_j| j^4| \right)^2 \leq D_1' + \frac{D_2' N^8}{\rho^{2N}}$$

어디다 써먹음?

$$\mathbb{E}_{n,\boldsymbol{v}}[\|\psi - \psi'\|_2^2] \leq D_0 \left(\frac{L_A^4 + \beta_A^2}{M} + L_A^4 \right) \|\Delta\theta\|_2^2 \, \mathbb{E}_n \left[\left(\sum_{j=1}^{n} | \left| \hat{b}_j^2 \right| j^4 \right) \right]$$

비슷한 Lemma는 여기다 적용 $N \leftarrow \sqrt{N}$

$$\mathbb{E}_{n,v}[\psi^2] \leq \frac{4}{(b-a)^2} \left(\frac{2L_A^2}{M} + d' L_{nuc}^2 \right) \mathbb{E}_n \left[\left(\sum_{j=1}^{n} \left| \hat{b}_j \right| j^2 \right)^2 \right]$$

About Chebyshev Polynomial

Lemma 4 (Chebyshev Stability Bounds)

Suppose that $A, A+E \in \mathbb{R}^{d\times d}$ are symmetric matrices and they have eigenvalues in $[-1,1]$. Let T_i and U_i be the first and the second kind of Chebyshev basis polynomial with degree $i \geq 0$, respectively. Then, is holds that

$$\|T_i(A+E) - T_i(A)\| \leq i^2 \|E\|, \quad \|U_i(A+E) - U_i(A)\| \leq \frac{i(i+1)(i+2)}{3} \|E\|$$

where $\|\cdot\|$ can be $\|\cdot\|_2$ (spectral norm) or $\|\cdot\|_2$ (Frobenius norm)

T_i 는 i^2-Lipschitz 이고 U_i 는 $\frac{i(i+1)(i+2)}{3}$-Lipschitz 이다.