These notes have not received the scrutiny of publication. They could be missing important references, etc.

Properties of Hermite Polynomials, Proving the Poincaré Inequality

1 Hermite Expansion

Recap.
Last lecture, we briefly defined the Hermite polynomials, a set of orthogonal polynomials which serve as an orthogonal basis for $L^2(\mu)$. Since polynomials are dense in $L^2(\mu)$, writing functions in terms of Hermite polynomials will allow us to easily prove the Poincaré inequality.

Recall that

$$\operatorname{He}_n(x) = (-1)^n e^{x^2/2}[\frac{d^n}{dx^n} e^{-x^2/2}], \ \ \operatorname{He}_0(x) = 1 \tag{1}$$

We'll need the generating function identity for hermite polynomials:

$$e^{tx - \frac{x^2}{2}} = \sum_{n=0}^{\infty} \frac{\operatorname{He}_n(x)}{n!} t^n \tag{2}$$

We could also generate them by induction or Equation (2), but starting with (3) will make proving what follows easier.

Basic Properties of Hermite Polynomials

We prove the following three properties of Hermite Polynomials:

1. Orthogonality: $\mathbb{E}_{Z\sim N(0,1)}[\operatorname{He}_n(z)\operatorname{He}_m(z)] = \begin{cases} 0, & \text{if } n \neq m. \\ n!, & \text{if } n = m. \end{cases}$
2. Inductive Formula: $\operatorname{He}_{n+1}(x) = x\operatorname{He}_n(x) - n\operatorname{He}_{n-1}(x)$
3. Derivative Formula: $\operatorname{He}_n'(x) = n\operatorname{He}_{n-1}(x)$

Proof of 1.

$$\mathbb{E}[e^{tz - \frac{t^2}{2}} e^{sz - \frac{s^2}{2}}] = \sum_{n,m} \frac{\mathbb{E}[\operatorname{He}_n(z)\operatorname{He}_m(z)] t^n s^m}{n!m!}$$

$$\mathbb{E}[e^{tz - \frac{t^2}{2}} e^{sz - \frac{s^2}{2}}] = e^{-\frac{t^2}{2} - \frac{s^2}{2}} \mathbb{E}[e^{(t+s)z}] = e^{-\frac{t^2}{2} - \frac{s^2}{2}} e^{\frac{(t+s)^2}{2}} = e^{ts} = \sum_n \frac{(ts)^n}{n!}$$

Where in the third step we used the fact that the moment generating function of a $N(0,1)$ distribution is $E[e^{tz}] = e^{\frac{t^2}{2}}$.

Matching powers of t and s, we see that for $n \neq m$, $E[\operatorname{He}_n(z)\operatorname{He}_m(z)]$ must be 0, while for $n = m$, $E[\operatorname{He}_n(z)\operatorname{He}_m(z)]$ must equal $n!$, which is the desired result.

$\square$

Proof of 2.
$$\frac{\partial}{\partial t}\sum_{n=0}^{\infty}\frac{\mathrm{He}_n(x)t^n}{n!}=\sum_{n=1}^{\infty}\frac{n\,\mathrm{He}_n(x)t^{n-1}}{n!}=\sum_{n=1}^{\infty}\frac{\mathrm{He}_n(x)t^{n-1}}{(n-1)!}=\sum_{n=0}^{\infty}\frac{\mathrm{He}_{n+1}(x)t^n}{(n)!}$$

$$\frac{\partial}{\partial t}\sum_{n=0}^{\infty}\frac{\mathrm{He}_n(x)t^n}{n!}=\frac{\partial}{\partial t}e^{tx-\frac{t^2}{2}}=(x-t)e^{tx-\frac{t^2}{2}}=(x-t)\sum_{n=0}^{\infty}\frac{\mathrm{He}_n(x)t^n}{n!}=\sum_{n=0}^{\infty}\frac{x\,\mathrm{He}_n(x)t^n}{n!}-\sum_{n=1}^{\infty}\frac{n\,\mathrm{He}_{n-1}(x)t^n}{(n)!}$$

$$\text{So }\sum_{n=0}^{\infty}\frac{\mathrm{He}_{n+1}(x)t^n}{(n)!}=\sum_{n=0}^{\infty}\frac{x\,\mathrm{He}_n(x)t^n}{n!}-\sum_{n=1}^{\infty}\frac{n\,\mathrm{He}_{n-1}(x)t^n}{(n)!}$$

Again matching coefficients, we see that $\begin{cases}\mathrm{He}_{n+1}(x)=x\,\mathrm{He}_n(x)-n\,\mathrm{He}_{n-1}(x), & n\geq 1\\ \mathrm{He}_{n+1}(x)=x\,\mathrm{He}_n(x), & n=0\end{cases}$, which is the desired result. □

Proof of 3.

$$\frac{\partial}{\partial x}\sum_{n=0}^{\infty}\frac{\mathrm{He}_n(x)t^n}{n!}=\sum_{n=0}^{\infty}\frac{\mathrm{He}'_n(x)t^n}{n!}=\frac{\partial}{\partial x}e^{tx-\frac{t^2}{2}}=te^{tx-\frac{t^2}{2}}=\sum_{n=0}^{\infty}\frac{\mathrm{He}_n(x)t^{n+1}}{n!}=\sum_{n=1}^{\infty}\frac{\mathrm{He}_{n-1}(x)t^n}{(n-1)!}$$

Matching coefficients one last time, we see that $\mathrm{He}'_n(x)=n\,\mathrm{He}_{n-1}(x)\ \forall n\geq 1$ which is the desired result. □

Three Brief Notes

1. We also can define normalized Hermite Polynomials $h_n(x)=\frac{\mathrm{He}_n(x)}{\sqrt{n}}$. These are, unsurprisingly, orthonormal rather than just orthogonal.

2. The fact that we have such a convenient formula for getting $\mathrm{He}_{n-1}(x)$ from $\mathrm{He}_n(x)$ in the form of the derivative formula inspires the question of whether there is a convenient formula for getting $\mathrm{He}_n(x)$ from $\mathrm{He}_{n-1}(x)$. Of course there is, as we can plug the derivative formula into the induction formula to get $\mathrm{He}_{n+1}(x)=(x-\frac{d}{dx})\,\mathrm{He}_n(x)$.

3. $\deg(\mathrm{He}_n(x))=n$. This can clearly be seen from the induction formula and the derivative formula, along with the definition $\mathrm{He}_0(x)=1$. By the derivative formula, when we go from n to $n-1$ we take a derivative which reduces the degree by 1. Meanwhile, by the induction formula when we go from $n-1$ to n we multiply by x which increases the degree by 1. The induction formula and definition of $\mathrm{He}_0(x)$ are sufficient: Since we know that $\deg(\mathrm{He}_0(x))=0$ and that $\deg(\mathrm{He}_n(x))=\deg(\mathrm{He}_{n-1}(x))+1$, it is clear by induction that $\deg(\mathrm{He}_n(x))=n$.

Creation and Annihilation Operators

The fact that we have formulas which allow us to go up and down the "ladder" of Hermite Polynomials inspires the use of Creation and Annihilation Operators. These are used a lot in Quantum Mechanics and QFT. They are defined as follows:

$$A=\frac{d}{dx},\ A^{\dagger}=x-\frac{d}{dx}$$

as shown above, they obey the following formulas:

$$Ah_n=\sqrt{n}h_{n-1},\ A^{\dagger}h_n=\sqrt{n+1}h_{n+1}$$

Multivariate Hermite Polynomials

One can easily define multivariate Hermite Polynomials, which take on values in $\mathbb{R}^d$ rather than $\mathbb{R}$, as follows:

$$\mathrm{H}_\alpha(x) = \prod_{i=1}^{d} h_{\alpha_i}(x_i),\ \ \alpha \in \mathbb{N}^d$$

So for example, $\mathrm{H}_{(2,1)}(x_1, x_2) = h_2(x_1)h_1(x_2)$

Fact: H_α is an orthonormal basis for polynomials in x1,...,xd. Polynomials are dense in gaussian spaces, so orthonormal basis for all functions is not hard to check.

We are ready to prove the Poincaré inequality!

2 Proof of the Poincaré inequality

First we recall the Poincaré inequality.

Theorem 1 (Poincaré inequality)**.** *Let f be differentiable, $Z \sim N(0, \sigma^2 I_n)$. Then*

$$\mathrm{Var}(f(Z)) \leq \sigma^2 \mathbb{E}[|\nabla f|_2^2$$

Example 1. Let $f(z) = z$. Then $\mathrm{Var}(f(z)) = \mathrm{Var}(z) = \sigma^2$, while $\sigma^2\mathbb{E}[|\nabla f|_2^2] = \sigma^2$. The inequality is tight in this case. In particular, it is clear that the inequality is tight when f is any linear function of z.

We now need one more piece of background:

Since hermite polynomials are a basis for polynomials, we can write f as $f(x) = \sum_{\alpha\in\mathbb{N}^d} \hat{f}(\alpha)H_\alpha(x)$ (in L^2). We use the hat because this is like a fourier transform, for which the convention is to use hats. The $\hat{f}(\alpha)$ are referred to as fourier coefficients or hermite coefficients.

Proof of the Poincaré inequality.

$$\mathrm{Var}(f) = \mathrm{Cov}(\sum_\alpha \hat{f}(\alpha)\,\mathrm{H}_\alpha, \sum_\beta \hat{f}(\beta)\,\mathrm{H}_\beta)$$

Since $\mathrm{H}_{(0,0,\ldots,0)}$ is constant in z, we can drop the $\alpha = (0,0,...,0)$ and $\beta = (0,0,...,0)$ terms from the sums. We define $|\alpha| = \sum_i \alpha_i$.

$$\mathrm{Var}(f) = \mathrm{Cov}(\sum_\alpha \hat{f}(\alpha)\,\mathrm{H}_\alpha, \sum_\beta \hat{f}(\beta)\,\mathrm{H}_\beta) = \mathrm{Cov}(\sum_{|\alpha|\geq 1} \hat{f}(\alpha)\,\mathrm{H}_\alpha, \sum_{|\beta|\geq 1} \hat{f}(\beta)\,\mathrm{H}_\beta)$$

$$= \sum_{|\alpha|\geq 1}\sum_{|\beta|\geq 1} \hat{f}(\alpha)\hat{f}(\beta)\mathrm{Cov}(\mathrm{H}_\alpha, \mathrm{H}_\beta) = \sum_{|\alpha|\geq 1}\sum_{|\beta|\geq 1} \hat{f}(\alpha)\hat{f}(\beta)\mathbb{E}(\mathrm{H}_\alpha\,\mathrm{H}_\beta) = \sum_{|\alpha|\geq 1} |\hat{f}(\alpha)|^2$$

Where in the second to last equality we used that $\mathbb{E}[\mathrm{H}_\alpha] = 0$ unless $\alpha = (0,0,...,0)$. This can be shown easily: $\mathbb{E}[\mathrm{H}_\alpha] = \mathbb{E}[\mathrm{H}_\alpha\,\mathrm{H}_{(0,0,\ldots,0)}] = 0$ if $\alpha \neq 0$ by the orthogonality of Hermite Polynomials.

This equality $\mathrm{Var}(f) = \sum_{|\alpha|\geq 1} |\hat{f}(\alpha)|^2$ is known as the Plancherel Theorem/Parseval's Identity, and applies to fourier series as well.

To find ∇f, we first find $\frac{\partial}{\partial x_i} f$.

$$\frac{\partial}{\partial x_i} f = \frac{\partial}{\partial x_i} \sum_{\alpha\in\mathbb{N}^d} \hat{f}(\alpha) H_\alpha(x) = \sum_{\alpha\in\mathbb{N}^d s.t. \alpha_i \geq 1} \hat{f}(\alpha) \frac{\partial\,\mathrm{H}_\alpha(x)}{\partial x_i}$$

$$\frac{\partial\, \mathrm{H}_\alpha(x)}{\partial x_i} = (\prod_{i\neq j} h_{\alpha_j})\frac{\partial h_{\alpha_i}(x_i)}{\partial x_i} = (\prod_{i\neq j} h_{\alpha_j})\sqrt{\alpha_i} h_{\alpha_i - 1} = \sqrt{\alpha_i}\, \mathrm{H}_{\alpha - e_i}$$

where e_i is a vector full of 0's with a 1 in the ith coordinate.

$$\frac{\partial}{\partial x_i} f = \sum_{\alpha\in\mathbb{N}^d s.t.\alpha_i\geq 1} \hat{f}(\alpha)\frac{\partial\, \mathrm{H}_\alpha(x)}{\partial x_i} = \sum_{\alpha\in\mathbb{N}^d s.t.\alpha_i\geq 1} \hat{f}(\alpha)\sqrt{\alpha_i}\, \mathrm{H}_{\alpha-e_i}$$

$$\mathbb{E}[|\nabla f|_2^2] = \mathbb{E}[\sum_i |\frac{\partial f}{\partial x_i}|^2] = \sum_i \mathbb{E}[(\sum_\alpha \hat{f}(\alpha)\sqrt{\alpha_i}\, \mathrm{H}_{\alpha-e_i})^2] = \sum_i\sum_\alpha \alpha_i |\hat{f}(\alpha)|^2$$

$$= \sum_\alpha |\hat{f}(\alpha)|^2 \sum_i \alpha_i = \sum_\alpha |\hat{f}(\alpha)|^2|\alpha| = \sum_{\alpha s.t.|\alpha|\geq 0} |\hat{f}(\alpha)|^2|\alpha|$$

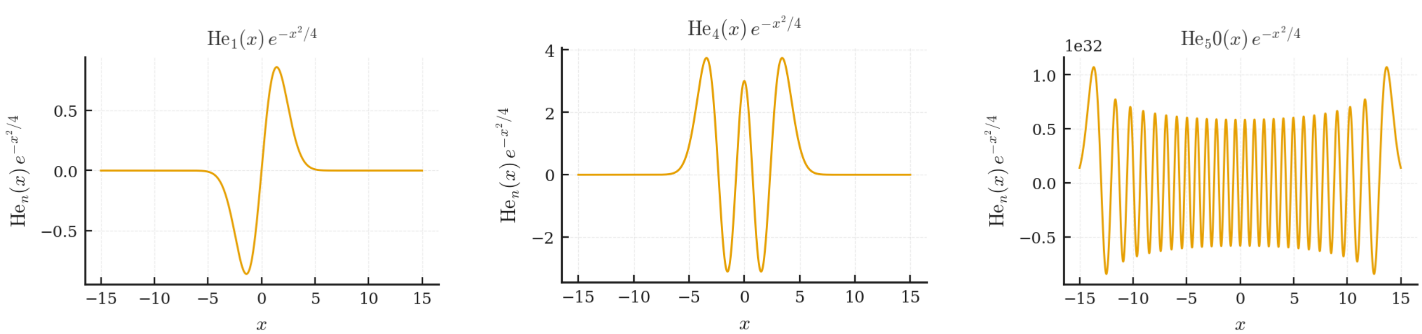
Since the sum is over α $s.t.|\alpha| \geq 0$, $|\alpha| \geq 1$ for every term. Thus, because of that and the fact that $|\hat{f}(\alpha)|^2 \geq 0$ $|\hat{f}(\alpha)|^2 \geq 0$, $\mathbb{E}[|\nabla f|_2^2] = \sum_{\alpha s.t.|\alpha|\geq 0} |\hat{f}(\alpha)|^2|\alpha| \geq \sum_{\alpha s.t.|\alpha|\geq 0} |\hat{f}(\alpha)|^2 = \mathrm{Var}(f)$.

□

While we only proved this for differentiable functions, it can be extended to Lipschitz/more general functions with some additional effort. As stated before, the two sides become equal for any linear function of z, since then $|\alpha| = 1$, which follows since otherwise the degree of f will be greater than 1.

Remark 1. Suppose $\mathrm{Var}(f) = 1, \mathbb{E}[f] = 0$. Then $\sum_\alpha |\hat{f}(\alpha)|_2^2 = 1$, and $\mathbb{E}[|\nabla f|_2^2] = \sum_{|\alpha|\geq 1} |\hat{f}(\alpha)|_2^2|\alpha|$ can be interpreted as providing a probability $|\hat{f}(\alpha)|_2^2$ of finding the function in each α, or in other words as the probability of finding the function with some level of smoothness. This can be made more precise, with the sum being $\mathbb{E}[\alpha]$ under some probability distribution.

One can understand the connection between α and smoothness by looking at the graphs of the "Hermite functions", defined by multiplying by $e^{-\frac{x^2}{4}}$ (Up to scaling), which allow us to see the behavior of the Hermite polynomials while ignoring their tail behavior.



(a) $n = 1$ (b) $n = 4$ (c) $n = 50$

Figure 1: Probabilists' $\mathrm{He}_n(x)\, e^{-x^2/4}$ for $n = 1, 4, 50$.

Informal Statement:
Functions supported only on small α (For example, $|\alpha| \leq D$ are as "smooth" as degree D polynomials.

3 Side Remarks and Complexity Theory

Question. A key question in modern average-case complexity theory: for some random variable such as $Z \sim N(0, I_n)$ or $\{\pm 1\}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$, when is $f(Z)$ "typically fast to calculate"?

Example 2 (Pure p-spin model / Random p-degree optimization)**.**

$$\text{compute } (f_1(x), ..., f_n(x)) \approx \arg\max_{x\in\{\pm 1\}^n} \sum_{\binom{[n]}{p}} g_\alpha x^\alpha, g_\alpha \sim N(0,1)$$

Here, the sum is over all subsets of indices of size p. The equality is not exact because we are interested in understanding the ability to compute approximate rather than exact minimizers/maximiers (note that the problems of minimizing/maximizing are equivalent).

When can we compute $x \subset \{\pm 1\}^n$ s.t. $p(x) = (1 + o(1)) \max_{y\in\{\pm 1\}^n} \sum_\alpha g_\alpha y^\alpha$ in poly(n) time?

It is believed (Gamarnik, Jagannath, Wein, 2020) that

- For p even if $p = 2$, it's possible (Already shown by Andrea Montanari before their paper)
- For p even if $p \geq 4$, it's not possible

What is the "evidence"? They present the heuristic argument that this is true for low degree polynomials. The proof is based on a type of physics-inspired landscape analysis technique called "overlap gap".

We saw a version of this property in low-temp (large β) random energy models. In these models $\max_{x\in\{\pm 1\}^n} E(x)$ is a random variable itself (here E(x) is the energy of state x, not its expected value). The number of approximate maximizers of E(x) is Poisson with some parameter determined by the model. The approximate maximizers will in general be far away from each other, since their locations are given by uniform sampling without replacement on the cube.