My Title

Lev Stambler
levstamb@umd.edu
University of Maryland, College Park,
NeverLocal Ltd.

September 10, 2025

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri.

1 Core Idea: Datasets and Basic In-Distribution Testing

Ideally, we'd like to take the useful tools of Fourier Analysis (which assumes a product space) and generalize them to any distribution where the underlying super-space is a product space[1]. Though, we do not have a formal method of reasoning about this, we do not think that it is quite possible.

Rather, we will attempt to think about analysis as over a *dataset:* i.e. we will think of our distribution as being defined by a finite-sized dataset, $\mathcal{D}$, and the probability vector will be defined as

$$p(x) = \begin{cases} \frac{1}{|\mathcal{D}|} & \text{if } \ x \in \mathcal{D} \\ 0 \text{ otherwise} \end{cases}.$$

More formally, we will be over a space $\mathcal{T}^n$ (think of $\mathcal{T}$ as either $\mathbb{R}$ or the space of tokens etc.) and $\mathcal{D} \subset \mathcal{T}^n$. Then, we will be focusing on functions $f \in \mathcal{T}^n \to \mathbb{R}$. f can either be a trained model, ideal labeling function on the dataset, or some other efficiently computable function.

Then, we want to reason about Fourier coefficients as

$$\hat{f}(S) = \mathbb{E}_{x \sim \mathcal{D}}[f(x)\chi_S^{-1}(x)].$$

Further, we will abuse notation to denote $\mathcal{D} : \mathcal{T}^n \to \{0,1\}$ as the indicator function for the inclusion within the dataset (i.e. an in-distribution tester). We will write $\hat{f}_{\mathrm{OG}}(S)$ to denote the normal ("original") Fourier coefficient:

$$\hat{f}_{\mathrm{OG}}(S) = \mathbb{E}_{x \sim \mathcal{T}^n}[f(x)\chi_S^{-1}(x)].$$

Importantly, notice that

$$\hat{f}(S) = \frac{1}{|\mathcal{D}|} \sum_{x \in \mathcal{D}} f(x)\chi_S(x) = \frac{|\mathcal{T}^n|}{|\mathcal{D}|} \cdot \hat{f}_{\mathrm{OG}}(\mathcal{D} \circ f)$$

where $\circ$ is the element wise composition. TODO: I think we need to use χ^{-1}?

[1]This includes, but is not limited too, language datasets over tokens, images over pixels, and other similar data modalities.

$\frac{\mathcal{T}^n}{|\mathcal{D}|}$ is a normalizing constant which we will frequently arise. As such, we will denote $\frac{|\mathcal{T}|^n}{|\mathcal{D}|}$ as $C_{\mathcal{D}}$ and its inverse as $C_{\mathcal{D}}^{-1}$ for ease of notation.

We conveniently have our first lemma.

Lemma 1.1: For all $x \in \mathcal{D}$,

$$f(x) = C_{\mathcal{D}}^{-1} \sum_{S} \hat{f}(S)\chi_S(x)$$

Proof: First, note that $f(x) = \mathcal{D}(x) \cdot f(x)$ for $x \in \mathcal{D}$ and then $\widehat{\mathcal{D} \circ f}_{\mathrm{OG}}(S) = C_{\mathcal{D}}^{-1} \cdot \hat{f}(S)$ and as such, using the standard Fourier identity

$$\mathcal{D}(x) \cdot f(x) = \sum_{S} \widehat{\mathcal{D} \circ f}(S)\chi_S(x) = C_{\mathcal{D}}^{-1} \sum_{S} \hat{f}(S).$$

■

1.1 In Distribution Testing and Related Notions

Unfortunately, we were not able to find a good way to define property testing, *soley* over the distribution. Rather, we will need to test both in-distribution and out-of-distribution samples. Still, we will only need to test out-of-distribution samples in a very limited way: samples which are only hamming distance 1 away from in-distribution samples for most of our properties.

To denote the need for in-distribution testing, we will append $\mathcal{D}$ as a subscript to various operators.

Definition 1.1 *(Distribution-Testing Coordinate Averaging)*: Let $E^i_{\mathcal{D}}$ for $i \in n$ be the i-th in distribution operator for $x \in \mathcal{D}$:

$$E^i_{\mathcal{D}}[f](x) = \mathbb{E}_{a \in \mathcal{T}}[\mathcal{D}(x_i \mapsto a) \circ f(x^{x_i \mapsto a})].$$

Note that Definition 1.1 *zeros out* any coordinate setting which does not remain within the dataset. Intuitively, we can think about $E^i_{\mathcal{D}}$ as the generic coordinate averaging operator for a function which will test whether an input is in the dataset and output 0 if it is not.

We can now define influence:

Definition 1.2 *(i-th Coordinate Distribution-Testing Influence Operator)*:

$$\mathrm{Inf}^i_{\mathcal{D}} f = \mathbb{E}_{x \in \mathcal{D}}\left[\left(f(x) - E^i_{\mathcal{D}} f(x)\right)^2\right].$$

Proposition 1.1: We now prove that basic identities still hold as in O'Donnell's Analysis of Boolean Functions [1]:

$$\mathrm{Inf}^i_{\mathcal{D}} f = C_{\mathcal{D}} \langle \mathcal{D} \circ f, \mathcal{D} \circ (f - E^i_D) \rangle,$$

$$E^{i_{\mathcal{D}}}_{D} f = C_{\mathcal{D}} \sum_{S, S_i = 0} \hat{f}(S)\chi_S(x)$$

$$\mathrm{Inf}^i_{\mathcal{D}} f = C_{\mathcal{D}} \sum_{S, S_i \neq 0} \hat{f}(S)^2.$$

Proof: We start with the second equality. TODO: check constants Note that,

$$\begin{aligned} E_{\mathcal{D}}^{i_{\mathcal{D}}^{i}}(x) &= \mathbb{E}_{a \in \mathcal{T}}[\mathcal{D}(x_i \mapsto a) \circ f(x^{x_i \mapsto a})] \\ &= \sum_{S, S_i = 0} \hat{f}_{\text{OG}}(f)(S)\chi_S(x) \\ &= C_{\mathcal{D}} \sum_{S, S_i = 0} \hat{f}(S)\chi_S(x). \end{aligned}$$

The first equality holds as we note that,

$$\begin{aligned} \text{Inf}^i_{\mathcal{D}} f &= C_{\mathcal{D}}^{-1} \langle f(x) - E^i_{\mathcal{D}}(x), f(x) - E^i_{\mathcal{D}}(x) \rangle \\ &= C_{\mathcal{D}}^{-1} \left(\mathbb{E}[f(x)^2] - 2E_x \left[f(x) E^i_{\mathcal{D}}(x) \right] + \mathbb{E}_{\mathcal{D}} \left[E^i_{\mathcal{D}}(x)^2 \right] \right). \end{aligned}$$

TODO: norm constant Then, note that by the second equality and Parseval's theorem, $E_{x \sim \mathcal{T}^n} \left[f(x) E^i_{\mathcal{D}}(x) \right] = \sum_{S, S_i = 0} \hat{f}_{\text{OG}}(S)^2$ and $\mathbb{E}_{\mathcal{D}} \left[E^i_{\mathcal{D}}(x)^2 \right] = \sum_{S, S_i = 0} \hat{f}(S)^2$ as desired.

Finally, we note that as $\text{Inf}^i_{\mathcal{D}} f = \mathbb{E}_{\mathcal{D}}[f(x)^2] - \mathbb{E}_{\mathcal{D}} \left[E^i_{\mathcal{D}}(x)^2 \right]$, we can see that

$$\mathbb{E}_{\mathcal{D}}[f(x)^2] - E_x \left[f(x) E^i_{\mathcal{D}}(x) \right] = \sum_S \hat{f}(S)^2 - \sum_{S, S_i = 0} \hat{f}(S)^2 = \sum_{S, S_i \neq 0} \hat{f}(S)^2.$$

■

Just as in TODO: cite , we can define total influence and get some convenient corollaries:

Definition 1.3 *(Total Influence, $\boldsymbol{I}_{\mathcal{D}}$)*:

$$\boldsymbol{I}_{\mathcal{D}}[f] = \sum_{i \in [n]} \text{Inf}^i_{\mathcal{D}}[f].$$

We immediately get:

Proposition 1.2:

$$\boldsymbol{I}_{\mathcal{D}}[f] = \sum_S \#S \ \hat{f}(S)^2$$

which, for finite groups, we get

$$\boldsymbol{I}_{\mathcal{D}}[f] = \sum_S k \cdot W^k[f].$$

as in page 213 of TODO: a where $\#S = |\text{supp }(S)| = \text{supp }(S) = \{i : S_i \neq 0\}$.

Finally, we introduce one more definition which will capture the closeness of two functions, f, g over $\mathcal{D}$.

Definition 1.4 *(In-Distribution Closeness, $\epsilon_{\mathcal{D}}$-closeness)*: We say that a function g is $\epsilon_{\mathcal{D}}$ close to f if:

$$\mathop{\mathbb{E}}_{x \in \mathcal{D}}[(f(x) - g(x))^2] \leq \epsilon$$

or equivalently,

$$C_{\mathcal{D}}^{-1} \mathop{\mathbb{E}}_{x \in \mathcal{T}^{\otimes n}} \left[(\mathcal{D} \circ f(x) - \mathcal{D} \circ g(x))^2 \right] \leq \epsilon.$$

1.2 Immediate Consequences

Learning from Random Examples

We can already adopt learning low-degree functions from random examples to our setting. Specifically, if the in-distribution Fourier mass is concentrated on low-degree terms, we can learn the function from random examples drawn from $\mathcal{D}$.

TODO: this is auto-gened, check over and make right! CLAUDE CHECK

Theorem 1.1: Let $f : \mathcal{T}^n \to \mathbb{R}$ be a function such that $\sum_{S:\ |S|\ > d} \hat{f}(S)^2 \leq \frac{\epsilon^2}{4}$. Then, there exists an algorithm which, given $O\left(\frac{n^d}{\epsilon^2} \log(n)\right)$ random examples from $\mathcal{D}$, outputs a function g which is $\epsilon_{\mathcal{D}}$ close to f with probability at least $\frac{2}{3}$.

One-Way Property Testing for Computable via Decision Tree

1.3 In Distribution Testing Definitions

Surprisingly, if want to characterize the complexity (or sensitivity or one of many Fourier properties) of our function f over distribution $\mathcal{D}$ and in-distribution testing, we can more or less use standard analysis.

In more detail, we will model our modified function $g : |\mathcal{T}|^n \to {\mathbb{R}^+}^m \cup \{0\}$ as

$$g(x) = \begin{cases} f(x) & \text{if } \ x \in \mathcal{D} \\ 0 & \text{otherwise} \end{cases}$$

And, taking the standard inner-product (TODO: we are over sphere/ real numbers, first take the inner-product then the other thing!)

Note that g is required to differentiate between in and out-of distribution.

Influence and Related Operators

We would like coordinate wise influence to be defined in the standard way, but adopted to our setting:

$$\mathrm{Inf}^i_{\mathcal{D}}[g] = \mathop{\mathbb{E}}_{x \in \mathcal{D}}\left[\left(g - E^i_{\mathcal{D}} g\right)^2\right]$$

where $E^i_{\mathcal{D}}$ is the i-th coordinate expectation operator:

$$\left[E^i_{\mathcal{D}} g\right](x) = \mathop{\mathbb{E}}_{x_i \in \mathcal{T}}\left[g\left(x^{i \mapsto x_i}\right)\right].$$

In words, the i-th expectation operator "averages out" the i-th coordinate over the dataset while the Influence measures the difference.

TODO: define distribution coeffs! TODO: This prob dist thing doesn't work??? Just have a subset thingy for now!!! Then, we have the following proposition:

Proposition 1.3:

$$E^i_{\mathcal{D}} g = \sum_{s : s_i \neq 0} \widehat{p_{\text{scale}} \cdot f}(s) \hat{f}_{\mathcal{D}}$$

Lemma 1.2: We can re-write the influence as an inner-product

$$\mathrm{Inf}^i_{\mathcal{D}}[f] = \left\langle p \cdot \left(g - E^i_{\mathcal{D}} g\right), g\right\rangle = \left\langle p \cdot \left(g - E^i_{\mathcal{D}} g\right), g - E^i_{\mathcal{D}} g\right\rangle$$

Proof: The second equality follows directly from the definition. The first is because

$$\left\langle p \cdot \left(g - E^i_{\mathcal{D}} g\right), g - E^i_{\mathcal{D}} g\right\rangle = \mathbb{E}_x\left[g^2\right] - 2\mathbb{E}_x\left[g \cdot E^i_{\mathcal{D}} f\right] + \mathbb{E}_x\left[\left(E^i_{\mathcal{D}} f\right)^2\right]$$

■

TODO: What happens if you take $f = \mathcal{D}$? TODO: Is there a monotonic question of degree over dataset vs degree over full space? (you can then know if a model class can learn it or not) (!!!!

This is a good idea. People would care a lot more !!!!) Look into: Statistical Learning theory: *distribution agnostic learning theory*

2 Gaussian Space Analysis

We reformulate the Fourier-analytic framework over Gaussian space, following O'Donnell [1] Chapter 11. The key insight is that Hermite polynomials serve as the orthonormal basis for $L^2(\mathbb{R}^n, \gamma_n)$, analogous to characters χ_S on the Boolean cube.

2.1 Gaussian Space Foundations

Definition 2.1 *(Standard Gaussian Measure)*: The *standard Gaussian measure* gam on $\mathbb{R}$ has density

$$d\gamma(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx. \tag{1}$$

The n-dimensional Gaussian measure γ_n on $\mathbb{R}^n$ is the product measure $\gamma^{\otimes n}$.

Definition 2.2 *(Gaussian Inner Product)*: For $f, g \in L^2(\mathbb{R}^n, \gamma_n)$, the inner product is

$$\langle f, g \rangle_{\gamma_n} = \mathbb{E}_{x \sim \gamma_n}[f(x) g(x)]. \tag{2}$$

Definition 2.3 *(Dataset in Gaussian Space)*: Let $\mathcal{D} \subset \mathbb{R}^n$ be a finite dataset. We define the uniform distribution over $\mathcal{D}$ by

$$p(x) = \begin{cases} \frac{1}{|\mathcal{D}|} & \text{if } x \in \mathcal{D} \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

We abuse notation and write $\mathcal{D} : \mathbb{R}^n \to \{0, 1\}$ for the indicator function of $\mathcal{D}$.

2.2 Hermite Polynomials

The Hermite polynomials form a complete orthonormal basis for $L^2(\mathbb{R}, \gamma)$.

Definition 2.4 *(Univariate Hermite Polynomials)*: The *probabilist's Hermite polynomials* $h_j : \mathbb{R} \to \mathbb{R}$ are the orthonormalized polynomials with respect to gam. The first few are:

$$h_0(x) = 1, \quad h_1(x) = x, \quad h_2(x) = \frac{x^2 - 1}{\sqrt{2}}, \quad h_3(x) = \frac{x^3 - 3x}{\sqrt{6}}. \tag{4}$$

Proposition 2.1: The Hermite polynomials satisfy $\langle h_j, h_k \rangle_\gamma = \delta_{jk}$.

Definition 2.5 *(Multivariate Hermite Polynomials)*: For a multi-index $\alpha \in \mathbb{N}^n$, define

$$H_\alpha(x) = \prod_{i=1}^{n} h_{\alpha_i}(x_i). \tag{5}$$

The *degree* of H_α is $|\alpha| = \sum_{i=1}^{n} \alpha_i$.

Theorem 2.1 (Hermite Expansion): Every $f \in L^2(\mathbb{R}^n, \gamma_n)$ has a unique expansion

$$f = \sum_{\alpha \in \mathbb{N}^n} \hat{f}(\alpha) H_\alpha \tag{6}$$

where the *Hermite coefficients* are $\hat{f}(\alpha) = \langle f, H_\alpha \rangle_{\gamma_n} = \mathbb{E}_{x \sim \gamma_n}[f(x) H_\alpha(x)]$.

2.3 Noise Operator

Definition 2.6 *(ρ-Correlated Gaussians)*: For $\rho \in [-1, 1]$, we say (x, y) are *ρ-correlated Gaussians* if $x \sim \gamma_n$ and

$$y = \rho x + \sqrt{1 - \rho^2} z \tag{7}$$

where $z \sim \gamma_n$ is independent of x. We write $y \sim N_\rho(x)$ for the conditional distribution of y given x.

Definition 2.7 *(Ornstein-Uhlenbeck Operator)*: The *noise operator U_ρ* is defined by

$$U_\rho f(x) = \mathbb{E}_{y \sim N_\rho(x)}[f(y)]. \tag{8}$$

Proposition 2.2 *(Hermite Eigenfunction Property)*: The Hermite polynomials are eigenfunctions of U_ρ:

$$U_\rho H_\alpha = \rho^{|\alpha|} H_\alpha. \tag{9}$$

Proof: By independence of coordinates and linearity, it suffices to check the univariate case. For $y = \rho x + \sqrt{1 - \rho^2} z$ with $z \sim \gamma$ independent, we have $\mathbb{E}_z\left[h_j\left(\rho x + \sqrt{1 - \rho^2} z\right)\right] = \rho^j h_j(x)$ by the generating function of Hermite polynomials. ■

2.4 Influence in Gaussian Space

Definition 2.8 *(Gaussian Influence)*: The *influence of coordinate i* on $f \in L^2(\mathbb{R}^n, \gamma_n)$ is

$$\mathrm{Inf}_i[f] = \mathbb{E}_{x \sim \gamma_n}\left[(\partial_i f(x))^2\right] \tag{10}$$

where $\partial_i f = \frac{\partial f}{\partial x_i}$ is the partial derivative.

Proposition 2.3 *(Hermite Representation of Influence)*:

$$\mathrm{Inf}_i[f] = \sum_{\alpha : \alpha_i \geq 1} \alpha_i \cdot \hat{f}(\alpha)^2. \tag{11}$$

Proof: We have $\partial_i H_\alpha = \sqrt{\alpha_i} H_{\alpha - e_i}$ where e_i is the i-th standard basis vector (and the term vanishes if $\alpha_i = 0$). By Parseval,

$$\mathrm{Inf}_i[f] = \mathbb{E}\left[(\partial_i f)^2\right] = \sum_{\alpha : \alpha_i \geq 1} \alpha_i \hat{f}(\alpha)^2. \tag{12}$$

■

Definition 2.9 *(Total Influence)*: The *total influence* of f is

$$\boldsymbol{I}[f] = \sum_{i=1}^{n} \mathrm{Inf}_i[f] = \sum_{\alpha \in \mathbb{N}^n} |\alpha| \cdot \hat{f}(\alpha)^2. \tag{13}$$

2.5 Dataset-Specific Analysis

We now adapt the above to the setting where we have a finite dataset $\mathcal{D} \subset \mathbb{R}^n$. The key insight is relating dataset coefficients to "original" Gaussian coefficients via the indicator function.

Definition 2.10 *(Original Hermite Coefficient)*: For $f : \mathbb{R}^n \to \mathbb{R}$, the *original Hermite coefficient* is

$$\hat{f}_{\mathrm{orig}}(\alpha) = \mathbb{E}_{x \sim \gamma_n}[f(x) H_\alpha(x)]. \tag{14}$$

Definition 2.11 *(Dataset Hermite Coefficient)*: The *dataset Hermite coefficient* is

$$\hat{f}_{\mathcal{D}}(\alpha) = \mathbb{E}_{x\sim\mathcal{D}}[f(x)H_\alpha(x)] = \frac{1}{|\mathcal{D}|}\sum_{x\in\mathcal{D}} f(x)H_\alpha(x). \tag{15}$$

Definition 2.12 *(Normalizing Constant)*: Let $\gamma_n(\mathcal{D}) = \mathbb{E}_{x\sim\gamma_n}[\mathcal{D}(x)]$ be the Gaussian measure of the dataset. The normalizing constant is

$$C_{\mathcal{D}} = \frac{1}{\gamma_n(\mathcal{D})\cdot|\mathcal{D}|}. \tag{16}$$

Importantly, we have the relation

$$\hat{f}_{\mathcal{D}}(\alpha) = C_{\mathcal{D}} \cdot \widehat{\mathcal{D}\circ f}_{\mathrm{orig}}(\alpha) \tag{17}$$

where $(\mathcal{D}\circ f)(x) = \mathcal{D}(x)\cdot f(x)$.

Lemma 2.1 *(Reconstruction)*: For all $x \in \mathcal{D}$,

$$f(x) = C_{\mathcal{D}}^{-1}\sum_{\alpha\in\mathbb{N}^n} \hat{f}_{\mathcal{D}}(\alpha)H_\alpha(x). \tag{18}$$

Proof: For $x \in \mathcal{D}$, we have $f(x) = \mathcal{D}(x)\cdot f(x)$. Applying the standard Hermite expansion to $\mathcal{D}\circ f$ over $(\mathbb{R}^n, \gamma_n)$:

$$\mathcal{D}(x)\cdot f(x) = \sum_\alpha \widehat{\mathcal{D}\circ f}_{\mathrm{orig}}(\alpha)H_\alpha(x) = C_{\mathcal{D}}^{-1}\sum_\alpha \hat{f}_{\mathcal{D}}(\alpha)H_\alpha(x). \tag{19}$$

■

Definition 2.13 *(Dataset Coordinate Averaging)*: For $x \in \mathcal{D}$, define

$$E_{\mathcal{D}}^i[f](x) = \mathbb{E}_{a\sim\gamma}[\mathcal{D}(x^{i\mapsto a})\cdot f(x^{i\mapsto a})] \tag{20}$$

where $x^{i\mapsto a}$ denotes x with the i-th coordinate replaced by a. This operator *zeros out* any coordinate setting that leaves the dataset.

Proposition 2.4 *(Hermite Expansion of Coordinate Average)*:

$$E_{\mathcal{D}}^i[f](x) = C_{\mathcal{D}}^{-1}\sum_{\alpha:\alpha_i=0} \hat{f}_{\mathcal{D}}(\alpha)H_\alpha(x). \tag{21}$$

Proof: By the reconstruction lemma, $f(y) = C_{\mathcal{D}}^{-1}\sum_\alpha \hat{f}_{\mathcal{D}}(\alpha)H_\alpha(y)$ for $y \in \mathcal{D}$. The multivariate Hermite polynomial factors as $H_\alpha(x^{i\mapsto a}) = H_{\alpha_{-i}}(x_{-i})\cdot h_{\alpha_i}(a)$. Applying $E_{\mathcal{D}}^i$:

$$\begin{aligned} E_{\mathcal{D}}^i[f](x) &= \mathbb{E}_{a\sim\gamma}[\mathcal{D}(x^{i\mapsto a})\cdot f(x^{i\mapsto a})] \\ &= C_{\mathcal{D}}^{-1}\sum_\alpha \hat{f}_{\mathcal{D}}(\alpha)H_{\alpha_{-i}}(x_{-i})\cdot \mathbb{E}_{a\sim\gamma}\left[\mathcal{D}(x^{i\mapsto a})h_{\alpha_i}(a)\right]. \end{aligned} \tag{22}$$

When $\alpha_i \geq 1$, the integral $\mathbb{E}_{a\sim\gamma}\left[h_{\alpha_i}(a)\cdot(\cdots)\right]$ vanishes since $\mathbb{E}_{a\sim\gamma}[h_k(a)] = 0$ for $k \geq 1$. Thus only $\alpha_i = 0$ terms survive. ■

Definition 2.14 *(Dataset Influence)*:

$$\mathrm{Inf}_{\mathcal{D}}^i[f] = \mathbb{E}_{x\sim\mathcal{D}}\left[\left(f(x) - E_{\mathcal{D}}^i[f](x)\right)^2\right]. \tag{23}$$

Proposition 2.5 *(Influence via Hermite Coefficients)*:

$$\mathrm{Inf}_{\mathcal{D}}^i[f] = \sum_{\alpha:\alpha_i\geq 1} \hat{f}_{\mathcal{D}}(\alpha)^2. \tag{24}$$

Proof: We first show that $\mathbb{E}_{\mathcal{D}}\left[f\cdot E_{\mathcal{D}}^i f\right] = \mathbb{E}_{\mathcal{D}}\left[\left(E_{\mathcal{D}}^i f\right)^2\right]$. This follows because $E_{\mathcal{D}}^i$ is idempotent: applying it twice gives the same result as applying it once, since averaging over $x_i \sim \gamma$ on a function already independent of x_i has no effect.

Now expand the definition:

$$\begin{aligned}\mathrm{Inf}^i_{\mathcal{D}}[f] &= \mathbb{E}_{\mathcal{D}}\left[\left(f - E^i_{\mathcal{D}} f\right)^2\right] \\ &= \mathbb{E}_{\mathcal{D}}[f^2] - 2\mathbb{E}_{\mathcal{D}}\left[f \cdot E^i_{\mathcal{D}} f\right] + \mathbb{E}_{\mathcal{D}}\left[\left(E^i_{\mathcal{D}} f\right)^2\right] \\ &= \mathbb{E}_{\mathcal{D}}[f^2] - \mathbb{E}_{\mathcal{D}}\left[\left(E^i_{\mathcal{D}} f\right)^2\right].\end{aligned} \tag{25}$$

By Parseval's identity over $\mathcal{D}$, we have $\mathbb{E}_{\mathcal{D}}[f^2] = \sum_\alpha \hat{f}_{\mathcal{D}}(\alpha)^2$. By the Hermite expansion of $E^i_{\mathcal{D}} f$ (which only includes terms with $\alpha_i = 0$), we get $\mathbb{E}_{\mathcal{D}}\left[\left(E^i_{\mathcal{D}} f\right)^2\right] = \sum_{\alpha : \alpha_i = 0} \hat{f}_{\mathcal{D}}(\alpha)^2$.

Therefore:

$$\mathrm{Inf}^i_{\mathcal{D}}[f] = \sum_\alpha \hat{f}_{\mathcal{D}}(\alpha)^2 - \sum_{\alpha : \alpha_i = 0} \hat{f}_{\mathcal{D}}(\alpha)^2 = \sum_{\alpha : \alpha_i \geq 1} \hat{f}_{\mathcal{D}}(\alpha)^2. \tag{26}$$

■

Definition 2.15 *(Total Dataset Influence)*:

$$\boldsymbol{I}_{\mathcal{D}}[f] = \sum_{i=1}^{n} \mathrm{Inf}^i_{\mathcal{D}}[f] = \sum_{\alpha \in \mathbb{N}^n} |\alpha| \cdot \hat{f}_{\mathcal{D}}(\alpha)^2. \tag{27}$$

Definition 2.16 *(Dataset Closeness)*: We say g is *ϵ-close* to f over $\mathcal{D}$ if

$$\mathbb{E}_{x \sim \mathcal{D}}\left[(f(x) - g(x))^2\right] \leq \epsilon. \tag{28}$$

2.6 Hypercontractivity

Theorem 2.2 ((2,q)-Hypercontractivity): For $q \geq 2$ and $\rho \leq \frac{1}{\sqrt{q-1}}$,

$$\left\|U_\rho f\right\|_q \leq \|f\|_2. \tag{29}$$

Corollary 2.1 *(Level-k Inequality)*: If f has Hermite expansion supported on degree at most k, then for $q \geq 2$:

$$\|f\|_q \leq (q-1)^{\frac{k}{2}} \|f\|_2. \tag{30}$$

Proof: Apply hypercontractivity with $\rho = \frac{1}{\sqrt{q-1}}$; since f is degree-k, we have $U_\rho f = \rho^{\leq k} f$ where the lowest eigenvalue is ρ^k. ■

2.7 Invariance Principle

The Invariance Principle connects Boolean analysis to Gaussian analysis.

Theorem 2.3 (Gaussian Invariance Principle (Informal)): Let $f : \{-1, 1\}^n \to \mathbb{R}$ be a multilinear polynomial with small influences. Let $g : \mathbb{R}^n \to \mathbb{R}$ be its *Gaussian version*: the function with the same multilinear coefficients, viewed as Hermite coefficients. Then $f(x)$ for $x \in \{-1, 1\}^n$ uniform and $g(z)$ for $z \sim \gamma_n$ have approximately the same distribution.

This principle allows us to transfer results between Boolean and Gaussian settings, and is crucial for applications like the Majority Is Stablest theorem.

2.8 Learning Low-Degree Functions

Theorem 2.4 (Learning Low-Degree Hermite Functions): Let $f : \mathbb{R}^n \to \mathbb{R}$ be bounded with $|f(x)| \leq 1$ and satisfy $\sum_{|\alpha| > d} \hat{f}_{\mathcal{D}}(\alpha)^2 \leq \frac{\epsilon^2}{4}$. Given $m = O\left(\frac{n^d}{\epsilon^2} \cdot \log n\right)$ random samples from $\mathcal{D}$, there exists an algorithm that outputs g with

$$\mathbb{E}_{x \sim \mathcal{D}}\left[(f(x) - g(x))^2\right] \leq \epsilon \tag{31}$$

with probability at least $\frac{2}{3}$.

Proof: **Algorithm:** Draw m i.i.d. samples $x^{(1)}, ..., x^{(m)} \sim \mathcal{D}$. For each multi-index α with $|\alpha| \leq d$, compute the empirical estimate

$$\tilde{f}(\alpha) = \frac{1}{m} \sum_{j=1}^{m} f\left(x^{(j)}\right) H_\alpha\left(x^{(j)}\right). \tag{32}$$

Output $g = \sum_{|\alpha| \leq d} \tilde{f}(\alpha) H_\alpha$.

Analysis: Decompose the error as

$$\begin{aligned} \mathbb{E}_{\mathcal{D}}\left[(f-g)^2\right] &= \mathbb{E}_{\mathcal{D}}\left[\left(f - f_{\leq d} + f_{\leq d} - g\right)^2\right] \\ &\leq 2\mathbb{E}_{\mathcal{D}}\left[\left(f - f_{\leq d}\right)^2\right] + 2\mathbb{E}_{\mathcal{D}}\left[\left(f_{\leq d} - g\right)^2\right] \end{aligned} \tag{33}$$

where $f_{\leq d} = \sum_{|\alpha| \leq d} \hat{f}_{\mathcal{D}}(\alpha) H_\alpha$ is the degree-d truncation.

Term 1 (Truncation error): By Parseval and the assumption,

$$\mathbb{E}_{\mathcal{D}}\left[\left(f - f_{\leq d}\right)^2\right] = \sum_{|\alpha| > d} \hat{f}_{\mathcal{D}}(\alpha)^2 \leq \frac{\epsilon^2}{4}. \tag{34}$$

Term 2 (Estimation error): We have

$$\mathbb{E}_{\mathcal{D}}\left[\left(f_{\leq d} - g\right)^2\right] = \sum_{|\alpha| \leq d} \left(\hat{f}_{\mathcal{D}}(\alpha) - \tilde{f}(\alpha)\right)^2. \tag{35}$$

For each α, the estimate $\tilde{f}(\alpha)$ is an average of m i.i.d. bounded random variables (since $|f \cdot H_\alpha|$ is bounded). By Hoeffding's inequality, for any $\delta > 0$:

$$\Pr\left[\left|\tilde{f}(\alpha) - \hat{f}_{\mathcal{D}}(\alpha)\right| > \delta\right] \leq 2\exp(-\Omega(m\delta^2)). \tag{36}$$

The number of multi-indices with $|\alpha| \leq d$ is at most $\binom{n+d}{d} \leq (n+d)^d \leq (2n)^d$. Setting $\delta = \frac{\epsilon}{2\sqrt{(2n)^d}}$ and taking a union bound, with $m = O\left(\frac{n^d}{\epsilon^2} \cdot \log n\right)$ samples we get

$$\sum_{|\alpha| \leq d} \left(\hat{f}_{\mathcal{D}}(\alpha) - \tilde{f}(\alpha)\right)^2 \leq \frac{\epsilon^2}{4} \tag{37}$$

with probability at least $\frac{2}{3}$.

Conclusion: Combining both terms, $\mathbb{E}_{\mathcal{D}}\left[(f-g)^2\right] \leq 2 \cdot \frac{\epsilon^2}{4} + 2 \cdot \frac{\epsilon^2}{4} = \epsilon^2 \leq \epsilon$. ∎

3 Conclusion and Future Directions

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet, ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum defuturum, quas natura non depravata desiderat. Et quem ad me accedis, saluto: 'chaere,' inquam, 'Tite!' lictores, turma omnis chorusque: 'chaere, Tite!' hinc hostis mi Albucius, hinc inimicus. Sed iure Mucius.

Acknowledgments

AI usage: the author would like to acknowledge extensive AI usage in the drafting, writing, and reviewing of this document. Most ideas are the author’s own, but AI tools increased the efficiency of the writing process dramatically, especially considering the author’s limited time budget for this work.

Bibliography

[1] Ryan O'Donnell. 2021. Analysis of boolean functions. *arXiv preprint arXiv:2105.10386* (2021).

A Appendix

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim aeque doleamus animo, cum corpore dolemus, fieri tamen permagna accessio potest, si aliquod aeternum et infinitum impendere malum nobis opinemur. Quod idem licet transferre in voluptatem, ut postea variari voluptas distinguique possit, augeri amplificarique non possit. At etiam Athenis, ut e patre audiebam facete et urbane Stoicos irridente, statua est in quo a nobis philosophia defensa et collaudata est, cum id, quod maxime placeat, facere possimus, omnis voluptas assumenda est, omnis dolor repellendus. Temporibus autem quibusdam et aut officiis debitis aut rerum necessitatibus saepe eveniet, ut et voluptates repudiandae sint et molestiae non recusandae. Itaque earum rerum defuturum, quas natura non depravata desiderat. Et quem ad me accedis, saluto: 'chaere,' inquam, 'Tite!' lictores, turma omnis chorusque: 'chaere, Tite!' hinc hostis mi Albucius, hinc inimicus. Sed iure Mucius. Ego autem mirari satis non queo unde hoc sit tam insolens domesticarum rerum fastidium. Non est omnino hic docendi locus; sed ita prorsus existimo, neque eum Torquatum, qui hoc primus cognomen invenerit, aut torquem illum hosti detraxisse, ut aliquam ex eo est consecutus? – Laudem et caritatem, quae sunt vitae sine metu degendae praesidia firmissima. – Filium morte multavit. – Si sine causa, nollem me ab eo delectari, quod ista Platonis, Aristoteli, Theophrasti orationis ornamenta neglexerit. Nam illud quidem physici, credere aliquid esse minimum, quod profecto numquam putavisset, si a Polyaeno, familiari suo, geometrica discere maluisset quam illum etiam ipsum dedocere. Sol Democrito magnus videtur, quippe homini erudito in geometriaque perfecto, huic pedalis fortasse; tantum enim esse omnino in nostris poetis aut inertissimae segnitiae est aut fastidii delicatissimi. Mihi quidem videtur, inermis ac nudus est. Tollit definitiones, nihil de dividendo ac partiendo docet, non quo ignorare vos arbitrer, sed ut.