Analysis of Boolean Functions	**(CMU 18-859S, Spring 2007)**

Lecture 22: Majority is the Stablest

Apr. 5, 2007

Lecturer: Ryan O'Donnell — *Scribe: Elaine Shi*

In this lecture, our goal is to prove the "Majority is Stablest" Theorem for balanced functions $f : \{-1,1\}^n \to [-1,1]$ with low influence. We shall first prove the "Invariance Principle" for multilinear polynomials with low influence and bounded degree. The Invariance Principle shows that to prove that "Majority is Stablest" Theorem on uniform probability distribution on $\{-1,1\}^n$, it suffices to prove it on product Gaussian distribution. Then we prove the "Majority is Stablest" Theorem on product Gaussian distribution.

1 Invanriance Principle

Theorem 1.1 *(Invariance Principle.) Let $Q(u_1, u_2, \ldots, u_n)$ be a multi-linear polynomial, with formal variables $u_1, u_2, \ldots, u_n$, $\deg(Q) \leq d$, i.e.,*

$$Q(u_1, u_2, \ldots, u_n) = \sum_{S \subseteq [n], |S| \leq d} \left(\alpha_S \cdot \prod_{i \in S} u_i \right)$$

where $\alpha_S \in \mathbb{R}$. Assume $\sum_{S \neq \emptyset} \alpha_S^2 = 1$. Write $\tau_i = \mathrm{Inf}_i(Q) = \sum_{S \ni i} \alpha_S^2$, and assume $\forall i = 1, 2, \ldots, n, \tau_i \leq \tau$. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ denote i.i.d. random ± 1 bits; let $\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n$ denote i.i.d. gaussians $N(0,1)$. Let $\mathbf{X} = Q(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, $\mathbf{Y} = Q(\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n)$. Let $\psi : \mathbb{R} \to \mathbb{R}$ denote a B-nice function, i.e., ψ is smooth and $\forall x \in \mathbb{R}, |\psi''''(x)| \leq B$. Then

$$|\mathbf{E}[\psi(\mathbf{X})] - \mathbf{E}[\psi(\mathbf{Y})]| \leq O(d \cdot 9^d \cdot B \cdot \tau)$$

Remark 1.2 *Note that when $u_i = \mathbf{g}_i$ or when $u_i = \mathbf{x}_i$, $\mathbf{Var}[Q(u_1, u_2, \ldots, u_n)] = \sum_{S \neq \emptyset} \alpha_S^2 = 1$. In addition, in case $\mathbf{Var}[Q(u_1, u_2, \ldots, u_n)] \neq 1$, to apply this theorem, we can scale the α_S's accordingly such that $\mathbf{Var}[Q(u_1, u_2, \ldots, u_n)] = 1$.*

Proof: The proof resembles that of the Berry-Esseen Theorem in the previous lecture.

As before, hybridize and write $\mathbf{z}_i = Q(\mathbf{g}_1, \ldots, \mathbf{g}_i, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n)$.

Claim 1.3

$$|\mathbf{E}[\psi(\mathbf{z}_{i-1})] - \mathbf{E}[\psi(\mathbf{z}_i)]| \leq O(B \cdot 9^d \cdot \tau_i^2).$$

Note that given the above claim, we can bound the overall error as below:

$$
\begin{aligned}
&|\mathbf{E}[\psi(\mathbf{X})] - \mathbf{E}[\psi(\mathbf{Y})]| = |\mathbf{E}[\psi(\mathsf{z}_0)] - \mathbf{E}[\psi(\mathsf{z}_n)]| \\
\leq &\sum_{i=1}^{n} |\mathbf{E}[\psi(\mathsf{z}_{i-1})] - \mathbf{E}[\psi(\mathsf{z}_i)]| = \sum_{i=1}^{n} O(B \cdot 9^d \cdot \tau_i^2) \\
= &O(B \cdot 9^d) \cdot \sum_{i=1}^{n} \tau_i^2 \leq O(B \cdot 9^d) \cdot \max_{i=1}^{n}\{\tau_i\} \cdot \sum_{i=1}^{n} \tau_i \\
\leq &O(B \cdot 9^d \cdot \tau) \cdot \sum_{i=1}^{n} \sum_{S \ni i} \alpha_S^2
\end{aligned}
$$

Note that

$$
\sum_{i=1}^{n} \sum_{S \ni i} \alpha_S^2 = \sum_{|S| \leq d} |S| \cdot \alpha_S^2 \leq d \cdot \sum_{S \neq \emptyset} \alpha_S^2 = d \cdot \mathbf{Var}[Q] = d
$$

Therefore, we conclude that $|\mathbf{E}[\psi(\mathbf{X})] - \mathbf{E}[\psi(\mathbf{Y})]| \leq O(d \cdot 9^d \cdot B \cdot \tau)$.

It remains to prove Claim 1.3. Write

$$
Q(u_1, u_2, \ldots, u_n) = r(u_1, u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n) + u_i \cdot s(u_1, u_2, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n)
$$

where r and s are multi-linear polynomials of degree at most d.

Let $\mathbf{R} = r(\mathbf{g}_1, \ldots, \mathbf{g}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n)$, $\mathbf{S} = s(\mathbf{g}_1, \ldots, \mathbf{g}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_n)$. We then have $\mathbf{z}_{i-1} = \mathbf{R} + \mathbf{x}_i\mathbf{S}$, $\mathbf{z}_i = \mathbf{R} + \mathbf{g}_i\mathbf{S}$. Note that $\mathbf{R}$ and $\mathbf{S}$ are mutually independent of $\mathbf{x}_i$ and $\mathbf{g}_i$. Now

$$
\begin{aligned}
&|\mathbf{E}[\psi(\mathbf{z}_{i-1})] - \mathbf{E}[\psi(\mathbf{z}_i)]| = |\mathbf{E}[\psi(\mathbf{R} + \mathbf{x}_i\mathbf{S})] - \mathbf{E}[\psi(\mathbf{R} + \mathbf{g}_i\mathbf{S})]| \qquad \text{(Use Taylor Expansion)} \\
= &\Big|\mathbf{E}[\psi(\mathbf{R}) + \mathbf{x}_i\mathbf{S} \cdot \psi'(\mathbf{R}) + (\mathbf{x}_i\mathbf{S})^2\psi''(\mathbf{R})/2 + (\mathbf{x}_i\mathbf{S})^3\psi'''(\mathbf{R})/6 + \{\leq \frac{B}{24} \cdot (\mathbf{x}_i\mathbf{S})^4\}] - \\
&\mathbf{E}[\psi(\mathbf{R}) + \mathbf{g}_i\mathbf{S} \cdot \psi'(\mathbf{R}) + (\mathbf{g}_i\mathbf{S})^2\psi''(\mathbf{R})/2 + (\mathbf{g}_i\mathbf{S})^3\psi'''(\mathbf{R})/6 + \{\leq \frac{B}{24} \cdot (\mathbf{g}_i\mathbf{S})^4\}]\Big|
\end{aligned}
$$

It is not hard to see that all of the first 4 terms cancel out. For example, consider the term $\mathbf{E}[(\mathbf{x}_i\mathbf{S})^2\psi''(\mathbf{R})/2]$.

$$
\begin{aligned}
&\mathbf{E}[\mathbf{x}_i^2 \cdot \mathbf{S}^2 \cdot \psi''(\mathbf{R})/2] = \mathbf{E}[\mathbf{x}_i^2] \cdot \mathbf{E}[\mathbf{S}^2 \cdot \psi''(\mathbf{R})/2] \qquad \text{(independence)} \\
= &\mathbf{E}[\mathbf{g}_i^2] \cdot \mathbf{E}[\mathbf{S}^2 \cdot \psi''(\mathbf{R})/2]
\end{aligned}
$$

Hence,

$$|\mathbf{E}[\psi(\mathbf{z}_{i-1})]-\mathbf{E}[\psi(\mathbf{z}_i)]| \le \frac{B}{24}\cdot\mathbf{E}[\mathbf{x}_i^4\cdot\mathbf{S}^4]+\frac{B}{24}\cdot\mathbf{E}[\mathbf{g}_i^4\cdot\mathbf{S}^4]$$
$$=\frac{B}{24}\cdot\mathbf{E}[\mathbf{x}_i^4]\cdot\mathbf{E}[\mathbf{S}^4]+\frac{B}{24}\cdot\mathbf{E}[\mathbf{g}_i^4]\cdot\mathbf{E}[\mathbf{S}^4]$$
$$=\frac{B}{24}\cdot 1\cdot\mathbf{E}[\mathbf{S}^4]+\frac{B}{24}\cdot 3\cdot\mathbf{E}[\mathbf{S}^4]$$
$$\le O(B)\cdot\mathbf{E}[\mathbf{S}^4]$$
$$\le O(B)\cdot 9^d\cdot\mathbf{E}[\mathbf{S}^2]^2 \qquad \text{(Hyper-contractivity)}$$
$$=O(B)\cdot 9^d\cdot\mathbf{E}[(\sum_{T\subseteq[n],T\ni i}\alpha_T\prod_{j\in T,j\neq i}\mathbf{y}_j)^2]^2$$

where

$$\mathbf{y}_j=\begin{cases}\mathbf{g}_j & \text{if } j<i,\\ \mathbf{x}_j & \text{if } j>i.\end{cases}$$

Notice that $\mathbf{E}[\mathbf{g}_i\mathbf{x}_j]=0$, $\mathbf{E}[\mathbf{g}_i^2]=\mathbf{E}[\mathbf{x}_i^2]=1$. Therefore,

$$|\mathbf{E}[\psi(\mathbf{z}_{i-1})]-\mathbf{E}[\psi(\mathbf{z}_i)]| \le O(B)\cdot 9^d\cdot(\sum_{T\subseteq[n],T\ni i}\alpha_T^2)^2$$
$$=O(B)\cdot 9^d\cdot\tau_i^2$$

□

2 Majority is the Stablest, Proof Sketch

Recall that we have been trying to prove the theorem that "Majority is the stablest", namely, fix $0\le\rho\le 1$, for all balanced functions $f:\{-1,1\}^n\to[-1,1]$ with "small" influnce $\mathrm{Inf}_i(f)$ for all i, then $\mathbb{S}_\rho(f)\le\frac{2}{\pi}\mathrm{arcsin}\rho+$ "small", where the term $\frac{2}{\pi}\mathrm{arcsin}\rho$ is the noise stability of Maj_n as n goes to infinity. We now prove the theorem for Gaussian random variables. To do this, we need the following definition.

Definition 2.1 *$\vec{\mathbf{g}}=(\mathbf{g}_1,\mathbf{g}_2,\ldots,\mathbf{g}_n)$ is an n-dimensional random Gaussian if $\mathbf{g}_1,\mathbf{g}_2,\ldots,\mathbf{g}_n$ are i.i.d. Gaussians $N(0,1)$.*

Definition 2.2 *$\vec{\mathbf{g}}$ and $\vec{\mathbf{h}}$ are ρ-correlated n-dimensional Gaussians $(0\le\rho\le 1)$, if $\vec{\mathbf{g}}$ is a random Gaussian, and $\vec{\mathbf{h}}$ is formed by*

$$\mathbf{h}_i=\rho\cdot\mathbf{g}_i+\sqrt{1-\rho^2}\cdot N(0,1)$$

independently across i's. Denote $\vec{\mathbf{h}}=\rho\cdot\vec{\mathbf{g}}+\sqrt{1-\rho^2}\cdot\vec{\mathbf{g}}'$ where $\vec{\mathbf{g}}'$ is an n-dimensional random Gaussian independent of $\vec{\mathbf{g}}$.

Definition 2.3 *Let $f : \mathbb{R}^n \to \mathbb{R}$, s.t. $\mathbf{E}[f(\vec{\mathbf{g}})^2] < \infty$. Let $0 \le \rho \le 1$. The Ornstein-Uhlenbeck operator U_ρ is defined by*

$$U_\rho f(\vec{\mathbf{g}}) = \mathbf{E}_{\vec{\mathbf{h}} \sim_\rho \vec{\mathbf{g}}}[f(\vec{\mathbf{h}})]$$

where $\vec{\mathbf{h}} \sim_\rho \vec{\mathbf{g}}$ denotes that $\vec{\mathbf{h}}$ is an n-dimensional random Gaussian ρ-correlated with $\vec{\mathbf{g}}$.

Remark 2.4 *U_ρ is defined analogous to the noise operator T_ρ for functions $f : \mathbb{R}^n \to \mathbb{R}$. Recall the definition of T_ρ on functions $f : \{-1,1\}^n \to \mathbb{R}$:*

$$T_\rho f(\mathbf{x}) = \mathbf{E}_{\mathbf{y} \sim_\rho \mathbf{x}}[f(\mathbf{y})]$$

For $\mathbf{x} \in \{-1,1\}^n$, $\mathbf{y} \sim_\rho \mathbf{x}$ denotes that we pick $\mathbf{y}$ ρ-correlated with $\mathbf{x}$ as follows: each coordinate $\mathbf{y}_i$ is set to be $\mathbf{x}_i$ with probability $1/2 + \rho/2$, and $-\mathbf{x}_i$ with probability $1/2 - \rho/2$.

For example, suppose $f(\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n) = \sum_{S \subseteq [n]} \left(\alpha_S \prod_{i \in S} \mathbf{g}_i\right)$. Then

$$\begin{aligned} & U_\rho f(\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n) = \mathbf{E}_{(\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n) \sim_\rho (\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n)}[f(\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n)] \\ =& \mathbf{E}_{\vec{\mathbf{h}}}[\sum_{S \subseteq [n]} \alpha_S \prod_{i \in S} \mathbf{h}_i] = \sum_{S \subseteq [n]} \alpha_S \prod_{i \in S} \mathbf{E}[\mathbf{h}_i] \\ =& \sum_{S \subseteq [n]} \alpha_S \prod_{i \in S} \rho \cdot \mathbf{g}_i \\ =& \sum_{S \subseteq [n]} \alpha_S \cdot \rho^{|S|} \cdot \prod_{i \in S} \mathbf{g}_i \end{aligned}$$

Now suppose $n = 1$, and $f(\mathbf{g}) = \mathbf{g}^2$. Then

$$U_\rho f(\mathbf{g}) = \mathbf{E}_{\mathbf{h} \sim_\rho \mathbf{g}}[\mathbf{h}^2] = \mathbf{E}_{\mathbf{h} \sim_\rho \mathbf{g}}[(\rho \cdot \mathbf{g} + \sqrt{1-\rho^2} N(0,1))^2] = \rho^2 \mathbf{g}^2 + 1 - \rho^2$$

Definition 2.5 *Let $f : \mathbb{R}^n \to \mathbb{R}$,*

$$\mathbb{S}_\rho(f) = \mathbf{E}_{\vec{\mathbf{g}}}[f(\vec{\mathbf{g}}) U_\rho f(\vec{\mathbf{g}})] = \mathbf{E}_{\vec{\mathbf{g}}}[f(\vec{\mathbf{g}}) \mathbf{E}_{\vec{\mathbf{h}} \sim_\rho \vec{\mathbf{g}}}[f(\vec{\mathbf{h}})]] = \mathbf{E}_{\substack{(\vec{\mathbf{g}}, \vec{\mathbf{h}}) \\ \rho\text{-correlated}}}[f(\vec{\mathbf{g}}) f(\vec{\mathbf{h}})]$$

Exercise. If $f = \sum_S \alpha_S \prod_{i \in S} \mathbf{g}_i$, then $\mathbb{S}_\rho(f) = \sum_S \alpha_S^2 \rho^{|S|}$.

Let $\vec{\mathbf{g}}$ be a random n-dimensional Gaussian, $\|\vec{\mathbf{g}}\|_2^2 = \sum_{i=1}^n \mathbf{g}_i^2$.

Observation 2.6 *Observe that $\|\vec{\mathbf{g}}\|_2^2$ is the sum of n independent random variables each having mean 1 and variance 2. Due to the Central Limit Theorem, with very high probability $\|\vec{\mathbf{g}}\|_2^2 = n \pm O(\sqrt{n})$; hence, with high probability, $\|\vec{\mathbf{g}}\|_2 = \sqrt{n}(1 \pm \frac{O(1)}{\sqrt{n}})$.*

Observation 2.7 *$\vec{\mathbf{g}}$ is distributed spherically symmetrically. This is because the p.d.f. of $\vec{\mathbf{g}}$ at $\vec{u} = (u_1, u_2, \ldots, u_n)$ is equal to*

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-u_i^2/2} = (\frac{1}{\sqrt{2\pi}})^n e^{-(u_1^2 + u_2^2 + \ldots + u_n^2)/2} = (\frac{1}{\sqrt{2\pi}})^n e^{-\|\vec{u}\|_2^2/2}$$

Clearly, the p.d.f. only depends on $\|\vec{u}\|_2^2$.

We conclude that $\vec{\mathbf{g}}$ is very much like a random point on the surface of a sphere having radius $\sqrt{n}$.

Suppose that $(\vec{\mathbf{g}}, \vec{\mathbf{h}})$ are ρ-correlated random n-dimensional Gaussians; meanwhile, $\vec{\mathbf{g}}$ is very much like a random point on the surface of a sphere having radius $\sqrt{n}$. Recall that $\vec{\mathbf{h}} = \rho \cdot \vec{\mathbf{g}} + \sqrt{1-\rho^2} \cdot \vec{\mathbf{g}}'$ where $\vec{\mathbf{g}}'$ is an n-dimensional random Gaussian independent of $\vec{\mathbf{g}}$.

$$\mathbf{E}[\vec{\mathbf{g}} \cdot \vec{\mathbf{h}}] = \sum_{i=1}^{n} \mathbf{E}[\mathbf{g}_i \mathbf{h}_i] = \rho n \pm O(\sqrt{n})$$

So the angle between $\vec{\mathbf{g}}$ and $\vec{\mathbf{h}}$ is very close to $\mathrm{arccos}\rho \pm o(1)$. So we see that picking $(\vec{\mathbf{g}}, \vec{\mathbf{h}})$ to be ρ-correlated n-dimensional Gaussians is very much like picking two random vectors fixed at angle $\mathrm{arccos}\rho \pm o(1)$, on the surface of a radius $\sqrt{n}$ sphere. Note that since $0 \le \rho \le 1$, this angle is acute.

To prove that "Majority is the stablest", we consider the following question: Among all functions $f : \mathbb{R}^n \to \{-1, 1\}$ that are balanced (i.e., $\mathbf{E}[f(\vec{\mathbf{g}})] = 0$), what is the maximum value of $\mathbb{S}_\rho(f)$, where

$$\mathbb{S}_\rho(f) = \mathop{\mathbf{E}}_{\substack{(\vec{\mathbf{g}},\vec{\mathbf{h}}) \\ \rho\text{-correlated}}} [f(\vec{\mathbf{g}})f(\vec{\mathbf{h}})] = 2\mathop{\mathbf{Pr}}_{\substack{(\vec{\mathbf{g}},\vec{\mathbf{h}}) \\ \rho\text{-correlated}}} [f(\vec{\mathbf{g}}) = f(\vec{\mathbf{h}})] - 1$$

We can also think of this as picking a subset on the radius-$\sqrt{n}$ sphere, whose size is $1/2$ of the sphere, such that when we pick two random vectors on the surface of the sphere fixed at some acute angle, this set will maximize the probability that both vectors land inside or outside the set.

Intuitively, the hemisphere (in fact, any hemisphere) seems like the best set for our purpose. A theorem due to Borell in 1985 shows that this intuition is indeed true.

Theorem 2.8 *(Borell'85.) Pick $\vec{\mathbf{g}}$ and $\vec{\mathbf{h}}$ to be two random vectors fixed at some acute angle, on the surface of a sphere. Let S denote a subset of half of the sphere, and consider the probability that both $\vec{\mathbf{g}}$ and $\vec{\mathbf{h}}$ both land inside or outside the set S. This probability is maximized when S is any hemisphere.*

Proof:(idea.) The theorem can be proved by a symmetrization argument. Informally, let P denote an arbitrary plane across the center of the sphere, one can show that symmetrizing S across the plane P improves the probability that $\vec{\mathbf{g}}$ and $\vec{\mathbf{h}}$ both land inside or outside the set S. Sufficiently many symmetrizations bring it close to a hemisphere. □

Now we know that the maximum half sphere that maximizes $\mathbb{S}_\rho(f)$ is the hemisphere, we can compute the maximum value for $\mathbb{S}_\rho(f)$ by letting $f = \mathrm{sgn}(\mathbf{g}_1)$:

$$\mathop{\mathbf{E}}_{\substack{(\vec{\mathbf{g}},\vec{\mathbf{h}}) \\ \rho\text{-correlated}}} [\mathrm{sgn}(\mathbf{g}_1)\mathrm{sgn}(\mathbf{h}_1)] = \frac{2}{\pi}\mathrm{arcsin}\rho.$$

Remark 2.9 *Note that in the above, we only considered functions $f : \mathbb{R} \to \{-1, 1\}$. But in order to prove the "Majority is Stablest" Theorem, we need to prove for $f : \mathbb{R} \to [-1, 1]$. In fact, it is possible to show a similar result for functions $f : \mathbb{R} \to [-1, 1]$.*