CSE 291: Analysis of Polynomial Threshold Functions Spring 2015

Lecture 4: Anticoncentration

Instructor: Daniel Kane **Date:** April 9

1 Overview

In this lecture, we wish to prove Carbery-Wright's Theorem on anticoncentration of polynomials with respect to an arbitrary log-concave distribution. We start with some basic properties of log-concave distributions. Then, we reduce the proof to 1-dimensional case, which will be looked in more details in the next lecture.

Theorem 1 (Carbery-Wright). *Let F be a log-concave probability distribution, p a degree-d polynomial. Then for $\epsilon > 0$,*

$$\mathbb{P}_{X\sim F}(|p(X)| \leq \epsilon \|p\|_{2,F}) = O(d\epsilon^{1/d})$$

When it is clear from the context, we write $\mathbb{P}$ instead of $\mathbb{P}_{X\sim F}$ and $\|p\|_2$ instead of $\|p\|_{2,F}$ for notational simplicity.

2 Properties of Log-concave Distributions

We start with some basic properties of log-concave distributions(and functions), which will be useful throughout the lecture.

Fact 1. *Marginals of log-concave distribution is log-concave. Consequently, any projection of log-concave distribution is log-concave.*

Fact 2. *Pointwise product of log-concave functions is log-concave.*

Fact 3. *If $\nu(\mathrm{d}x) = F(x)\,\mathrm{d}x$ is a log-concave probability distribution on $\mathbb{R}$, then for some μ, σ, $F(x) = O(\frac{1}{\sigma}e^{-\frac{|x-\mu|}{\sigma}})$.*

Proof of Fact 1. Consider a log-concave probability measure $\nu(\mathrm{d}x) = F(x)\,\mathrm{d}x$ defined on $\mathbb{R}^d$. We begin with the $d = 2$ case, that is, if $(X, Y) \sim \nu$, then the marginal distribution of X is log-concave. The marginal density of X is

$$F_X(x) = \int_{\mathbb{R}} F(x, y)\,\mathrm{d}y$$

To prove that $F_X(x)$ is log-concave, it suffices to show that for all $x_1, x_2 \in \mathbb{R}$,

$$F_X^2(\frac{x_1 + x_2}{2}) \geq F_X(x_1)F_X(x_2)$$

This is implied by Lemma 1 below by taking $f(\cdot) = F(x_1, \cdot)$, $g(\cdot) = F(x_2, \cdot)$, $h(\cdot) = F((x_1 + x_2)/2, \cdot)$, concluding the $d = 2$ case.

For general d, by similar argument, we get ν's every $(d-1)$-dimensional marginal density is log-concave. By applying the same logic repeatedly, we find out that the marginal distribution of any subset of variables is still log-concave.

□

Lemma 1 (Prekopa-Leindler). *Suppose functions $f, g, h : \mathbb{R} \to \mathbb{R}^+$ satisfies for all $x, y \in \mathbb{R}$,*

$$h^2(\frac{x+y}{2}) \geq f(x)g(y)$$

Then

$$\|h\|_1^2 \geq \|f\|_1 \|g\|_1$$

Proof. Note that if we multiply f, g and h by a, b and $\sqrt{ab}$ for some positive real numbers a and b, it does not affect the problem. Therefore, we may assume that $\sup_x f(x) = \sup_x g(x) = 1$. For any given function $\phi : \mathbb{R} \to \mathbb{R}$, $t > 0$, define ϕ's superlevel set as

$$\phi_t := \{x : \phi(x) \geq t\}$$

By assumption, if $x \in f_t$, $y \in g_t$, then $(x+y)/2 \in h_t$. Utilizing this fact, we claim that $m(h_t) \geq (m(f_t) + m(g_t))/2$. To see this, observe

$$h_t \subseteq \frac{\min(g_t) + f_t}{2} \cup \frac{g_t + \max(f_t)}{2}$$

where the intersection of two sets on the right hand side has only one point $\{\min(g_t) + \max(f_t)\}$ which is of zero Lebesgue measure.

Meanwhile, note that for any function $\phi : \mathbb{R} \to \mathbb{R}^+$,

$$\int_{\mathbb{R}} \phi(x)\,\mathrm{d}x = \int_{\mathbb{R}} \int_{\mathbb{R}} I(0 \leq t \leq \phi(x))\,\mathrm{d}t\,\mathrm{d}x = \int_{\mathbb{R}^+} m(\phi_t)\,\mathrm{d}t$$

Therefore,

$$\begin{aligned} \|h\|_1 &= \int_{\mathbb{R}^+} m(h_t)\,\mathrm{d}t \\ &\geq \int_{\mathbb{R}^+} \frac{m(f_t) + m(g_t)}{2}\,\mathrm{d}t \\ &= \frac{\|f\|_1 + \|g\|_1}{2} \\ &\geq \sqrt{\|f\|_1 \|g\|_1} \end{aligned}$$

The lemma follows. □

Proof of Fact 3. We only consider the case where $\ln F(x)$ is differentiable.(when $\ln F(x)$ is not differentiable, we can replace gradient with subgradient in Equation (1).) Let μ be the maximizer of $F(\cdot)$. Without loss of generality, assume $\mu = 0$. Let σ_+ be the minimum value of $\sigma > 0$ such that

$$-\frac{\mathrm{d}}{\mathrm{d}x}(\ln F(x))\bigg|_{x=\sigma} \geq \frac{1}{\sigma} \tag{1}$$

Define σ_- similarly on the left hand side. Let $\sigma = \max(\sigma_+, \sigma_-)$. Without loss of generality, suppose $\sigma = \sigma_+$. Now, by concavity of $\ln F(x)$, we know that for all x,

$$\ln F(x) \leq \ln F(\sigma_+) - \frac{1}{\sigma_+}(x - \sigma_+)$$

In the meantime, since $F(x)$ is monotonically decreasing in $[0, +\infty)$, $F(\sigma)$ is a probability density function, we have $\int_0^{\sigma_+} F(x)\,\mathrm{d}x \leq 1$, hence $F(\sigma_+) \leq \frac{1}{\sigma_+}$. Thus, $F(\sigma_+) \leq F(0) \leq F(\sigma_+)e \leq \frac{e}{\sigma_+}$. We conclude that for $x > 0$,

$$F(x) \leq F(\sigma_+)e^{-(x-\sigma_+)/\sigma_+} \leq eF(0)e^{-|x|/\sigma_+}$$

Similarly, by definition of σ_-, for $x < 0$

$$F(x) \le eF(0)e^{-|x|/\sigma_-} \le eF(0)e^{-|x|/\sigma_+}$$

Hence, for all $x \in \mathbb{R}$,

$$F(x) \le eF(0)e^{-|x|/\sigma_+} \le \frac{e^2}{\sigma_+} e^{-|x|/\sigma_+} = O(\frac{1}{\sigma} e^{-|x-\mu|/\sigma})$$

□

3 Reducing the Problem to One Dimensional Case

We observe that Theorem 1 is homogeneous in the polynomial p, that is, scaling p by a factor of $c > 0$ does not change the its statement. It therefore suffices to show that for an arbitrary polynomial p, for all $\epsilon > 0$,

$$\text{either } \|p\|_2 \le 1 \text{ or } \mathbb{P}(|p(X)| < \epsilon) \le O(d\epsilon^{1/d}) \tag{2}$$

To see why, note that consider an arbitrary polynomial p, if $\|p\|_2 = 0$, then the theorem is vacuously true. Otherwise consider normalized polynomial $\bar{p} = 2\frac{p}{\|p\|_2}$, $\|\bar{p}\|_2 = 2 > 1$. Thus, by Equation (2), for all $\epsilon > 0$,

$$\mathbb{P}(|\bar{p}(X)| < \epsilon) \le O(d\epsilon^{1/d})$$

Therefore for the original polynomial p,

$$\mathbb{P}(|p(X)| < \epsilon\|p\|_2) \le O(d\epsilon^{1/d})$$

Our problem now reduces to showing that there exists a numerical constant $C > 0$, such that for any log-concave probability measure ν, for all $\epsilon > 0$ and all degree-d polynomial p, the following two inequalities cannot both hold:

$$\int (p^2(x) - 1)\nu(\mathrm{d}x) > 0 \text{ and } \int (I(|p(x)| < \epsilon) - Cd\epsilon^{1/d})\nu(\mathrm{d}x) > 0 \tag{3}$$

It is difficult to reason with high dimensional integrals in general. Thankfully, there is a generic tool that can reduce such problem to its 1-dimensional counterpart. First we need some notations.

Definition 1. *A needle in $\mathbb{R}^n$ is a pair (l, ν), where $l \subseteq \mathbb{R}^n$ is a line segment, ν is a log-concave measure on l.*

Theorem 2 (Localization, (Lovasz-Simonovits))**.** *Suppose $f, g : \mathbb{R}^n \to \mathbb{R}$ are continuous functions, μ is a log-concave measure on $\mathbb{R}^n$, then if*

$$\int_{\mathbb{R}} f(x)\mu(\mathrm{d}x) > 0 \textit{ and } \int_{\mathbb{R}} g(x)\mu(\mathrm{d}x) > 0 \tag{4}$$

then there exists a needle (l, ν) such that

$$\int_l f(x)\nu(\mathrm{d}x) > 0 \textit{ and } \int_l g(x)\nu(\mathrm{d}x) > 0 \tag{5}$$

We remark that the measure ν on the needle (l, ν) we find does not have to be equal to the marginal of μ on l. In fact, as we will see in the proof, their densities typically differ by a factor of a log-concave function.

Before getting into the proof of Theorem 2, we see how it relates to our problem. Fix $\epsilon > 0$. Note that there is a small technicality that prevents us from directly applying this result to Equation (3) – the function $I(|p(x)| < \epsilon)$ is not continuous. Nevertheless, we can define a new function $h_\epsilon(\cdot)$ such that h_ϵ is continuous and $I(|y| < \epsilon/2) \le h_\epsilon(y) \le I(|y| < \epsilon)$. Now, it suffices to prove that for any log-concave probability measure μ, the following two inequalities cannot both hold:

$$\int (p^2(x) - 1)\mu(\mathrm{d}x) > 0 \text{ and } \int (h_\epsilon(p(x)) - Cd\epsilon^{1/d})\mu(\mathrm{d}x) > 0 \tag{6}$$

Invoking Theorem 2, we only need to show for any needle (l, ν), the following two inequalities cannot both hold:

$$\int_l (p^2(x) - 1)\nu(\mathrm{d}x) > 0 \text{ and } \int_l (h_\epsilon(p(x)) - Cd\epsilon^{1/d})\nu(\mathrm{d}x) > 0 \tag{7}$$

Since $h_\epsilon(y) \le I(|y| < \epsilon/2)$, it is enough to show that there exists a numerical constant $C > 0$, such that for any needle (l, ν), for all $\epsilon > 0$ and all degree-d polynomial p, the following two inequalities cannot both hold:

$$\int_l (p^2(x) - 1)\nu(\mathrm{d}x) > 0 \text{ and } \int_l (I(|p(x)| < \epsilon) - Cd\epsilon^{1/d})\nu(\mathrm{d}x) > 0 \tag{8}$$

Since a needle is essentially a log-concave measure on an interval in $\mathbb{R}$ (up to linear transformation), it is thus equivalent to the 1-dimensional case of Equation (3). The final statement follows from 1-dimensional Carbery-Wright Theorem, letting us conclude that the "real work" lies in the 1-dimensional proof, which will be discussed in the next lecture.

4 Proof Idea of Theorem 2

For simplicity, we consider the case of $n = 2$. Our plan is to progressively build nested convex sets $\{K_i\}_{i=1}^\infty$ of rapidly decreasing volume, such that the sequence finally converges to a needle, and maintains the invariant

$$\int_{K_i} f(x)\mu(\mathrm{d}x) > 0 \text{ and } \int_{K_i} g(x)\mu(\mathrm{d}x) > 0 \tag{9}$$

We start with defining K_0 to be a large enough ball such that $\int_{K_0} f(x)d\mu(x) > 0$ and $\int_{K_0} g(x)d\mu(x) > 0$. (the existence is guaranteed by Dominated Convergence Theorem.) Taking

$$\delta = \frac{1}{2}\min\left(\int_{K_0} f(x)d\mu(x), \int_{K_0} g(x)d\mu(x)\right) > 0,$$

define $f_\delta(x) := f(x) - \delta$, $g_\delta(x) := g(x) - \delta$. One has $\int_{K_0} f_\delta(x)\mu(\mathrm{d}x) > 0$ and $\int_{K_0} g_\delta(x)\mu(\mathrm{d}x) > 0$.

To build K_i from K_{i+1}, we search for a hyperplane H_i (in $n = 2$, it is simply a line) that cuts K_i into two convex bodies A_i and B_i, such that

$$\begin{aligned} \int_{A_i} g_\delta(x)d\mu(x) &= \int_{B_i} g_\delta(x)d\mu(x) = \frac{1}{2}\int_{K_i} g_\delta(x)d\mu(x) > 0 \\ \int_{A_i} \mathrm{d}x &= \int_{B_i} \mathrm{d}x = \frac{1}{2}\int_{K_i} \mathrm{d}x \end{aligned} \tag{10}$$

(The existence is justified by Lemma 2, as we will see below.) Hence,

$$\int_{A_i} f(x)\mu(\mathrm{d}x) + \int_{B_i} f_\delta(x)d\mu(x) = \int_{K_i} f_\delta(x)d\mu(x) > 0$$

Therefore, f has positive integration value on at least one of A_i and B_i. Now, we can pick K_{i+1} to be the one on which f has positive integral. This ensures that not only Equation (9) is true for all i, but also $\mathrm{Vol}(K_i) \downarrow 0$.

Now, Let $N = \bigcap_{i=1}^\infty K_i$, then $\mathrm{Vol}(N) = 0$, and N is still a bounded convex set. The only possibilities are: (1) N is a single point x, (2) N is a line segment l. In the first case, by continuity, $f_\delta(x) \ge 0$ and

$g_\delta(x) \geq 0$, thus $f(x) \geq \delta > 0$, $g(x) \geq \delta > 0$. Taking a sufficiently small line segment along any direction, in the neighborhood of x (e.g. picking sufficiently small ϵ such that $l = \{(t_1, t_2) : x_1 - \epsilon \leq t_1 \leq x_1 + \epsilon, t_2 = x_2\}$) and ν to be the uniform distribution on l guarantees Equation (5) holds.

In the second case, we construct a measure ν on l so that (l, ν) has the desired property. For each i, define projection $\pi_i : K_i \to N$, such that $\pi_i(x)$ maps every x to its nearest neighbor in N. Pick i large enough such that the projection onto N is roughly equal to the perpendicular foot of the line that is perpendicular to I and passes through x. Also, i is sufficiently large that for all $x \in K_i$, $f(\pi_i(x)) \approx f(x)$. This implies that

$$\int_{K_i} f(\pi_i(x))d\mu(x) > 0 \text{ and } \int_{K_i} g(\pi_i(x))d\mu(x) > 0$$

Roughly, this implies

$$\int_I f(x)\mathrm{Vol}_{n-1}(\pi_i^{-1}(\{x\}))F(x)\,\mathrm{d}x > 0 \text{ and } \int_I g(x)\mathrm{Vol}_{n-1}(\pi_i^{-1}(\{x\}))F(x)\,\mathrm{d}x > 0$$

Note that the term $\mathrm{Vol}_{n-1}(\pi_i^{-1}(\{x\}))$ can be thought of as the density of projection of uniform distribution over K_i onto I, which is log-concave by Fact 1. Also, by Fact 2 we see that $\mathrm{Vol}_{n-1}(\pi_i^{-1}(\{x\}))F(x)$ is log-concave, thus we can use it as ν's density function. This construction guarantees Equation (5) to hold.

For the general n-dimensional case, we construct the K_i's a bit more carefully. Instead of reducing its volume in each step, we aim to reduce the volume of each of its axis-parallel 2 dimensional projections. We cycle through all pairs of coordinates $(j_1, j_2), 1 \leq j_1 < j_2 \leq d$ – in each time step i we only consider finding hyperplanes whose normal vector lives in the subspace of corresponding pair and maintains the invariant (9), which again is guaranteed to exist by Lemma 2. This procedure guarantees that the final N is of dimension at most 1. (Otherwise there must exist a pair of coordinates such that N's projection on it has nonzero area, which contradicts our construction.) The construction of ν is essentially the same as the $n = 2$ case.

We now come back to justify the existence of a hyperplane that simultaneously bisects two functions' integrals with respect to a given set.

Lemma 2 (Ham-Sandwich Cut). *Given a bounded set $K \subset \mathbb{R}^2$ and two continuous functions $f, g : K \to \mathbb{R}$ with $f(x) \geq 0$ for all x, there exists a hyperplane H that cuts K into two parts A and B, such that*

$$\int_A f\,\mathrm{d}x = \int_B f\,\mathrm{d}x \text{ and } \int_A g\,\mathrm{d}x = \int_B g\,\mathrm{d}x$$

Proof. For simplicity, we assume that K is a ball and that f is strictly positive on K (in general you can write f as a limit of such functions and use the limiting hyperplane as your H).

Consider any angle $\theta \in (0, 2\pi]$, By intermediate value theorem, there exists a unique offset $c(\theta)$ such that the line $x\cos\theta + y\sin\theta + c(\theta) = 0$ bisects the integration of f on K. Note that we can adjust c so that $c(\theta + \pi) = -c(\theta)$ and c is continuous. Formally, define $A(\theta) = K \cap \{x\cos\theta + y\sin\theta + c(\theta) \geq 0\}$ and $B(\theta) = K \cap \{x\cos\theta + y\sin\theta + c(\theta) \leq 0\}$, then we have $A(\theta + \pi) = B(\theta)$ and $B(\theta + \pi) = A(\theta)$, and

$$\int_{A(\theta)} f\,\mathrm{d}x = \int_{B(\theta)} f\,\mathrm{d}x$$

Now we use the extra degree of freedom in θ to balance the integral of g. Define

$$D(\theta) = \int_{A(\theta)} g\,\mathrm{d}x - \int_{B(\theta)} g\,\mathrm{d}x$$

which is a continuous function by the continuity of c. Also, $D(\theta + \pi) = -D(\theta)$. Again, by simple application of intermediate value theorem, there exists some θ_0 such that $D(\theta_0) = 0$, completing the proof. □