**Lecture 7: From junta degree to non-negative rank**
CSE 599S: Entropy optimality, Winter 2016
**Instructor:** James R. Lee **Last updated:** January 28, 2016

# 1 From junta degree to non-negative rank

Our goal now is to prove the following theorem; in the previous lecture, we saw how this implies that $\bar{\gamma}(\mathrm{CUT}_n)$ grows faster than any polynomial in $n$.

Recall that given $g : \{0,1\}^m \to \mathbb{R}$ and a number $n \geqslant m$, we define the matrix $M_n^g : \binom{[n]}{m} \times \{0,1\}^n \to \mathbb{R}$ by

$$M_n^g(S,x) = g_S(x)\,,$$

where $g_S(x) = g(x|_S)$, and $x|_S$ is the (ordered) restriction of $x \in \{0,1\}^n$ to the coordinates indexed by $S$. Here, $\binom{[n]}{m}$ denotes the collection of subsets $S \subseteq [n]$ with $|S| = m$.

**Theorem 1.1** (Chan-Lee-Raghavendra-Steurer 2013)**.** *For every $m \geqslant 1$ and $g : \{0,1\}^m \to \mathbb{R}_+$, there is a constant $C = C(g)$ such that for all $n \geqslant 2m$,*

$$\mathrm{rank}_+(M_n^g) \geqslant C\left(\frac{n}{\log n}\right)^{\deg_J(g)}\,. \tag{1.1}$$

## 1.1 Proof setup

We will work with the uniform measures on $\{0,1\}^n$ and $\binom{[n]}{m}$, and we will write $\mathbb{E}_x$ and $\mathbb{E}_S$ to denote expectation with respect to $x \in \{0,1\}^n$ or $S \in \binom{[n]}{m}$ chosen uniformly at random.

The proof will be easiest in the density setting (recall the last part of Lecture 4). $L^2(\{0,1\}^n)$ is the Euclidean space of real-valued functions $f : \{0,1\}^n \to R$ equipped with the inner product

$$\langle f, g\rangle = \mathop{\mathbb{E}}_x[f(x)g(x)]\,.$$

Let $d = \deg_J(g) - 1$. Recall that since $\deg_J(g) > d$, there must exist a $d$-locally positive functional $\varphi \in L^2(\{0,1\}^n)$ such that

$$\langle \varphi, g\rangle < 0\,. \tag{1.2}$$

By definition, such a functional satisfies $\langle \varphi, q\rangle > 0$ for every $d$-junta $q \in L^2(\{0,1\}^n)$ that is not identically 0.

Let us normalize this functional so that $\langle \varphi, \mathbf{1}\rangle = 1$, where $\mathbf{1} \in L^2(\{0,1\}^n)$ is the function that takes values 1 everywhere (observe that $\mathbf{1}$ is a 0-junta, so it must be that this inner product was positive before we normalized it). Under this normalization, $\varphi$ is often referred to as a *pseudo-density.*

**Exercise (1 point) 1.2** (Justify the name "pseudo-density."). Suppose that $\varphi : \{0,1\}^n \to \mathbb{R}$ is a $k$-locally positive functional satisfying $\langle \varphi, \mathbf{1}\rangle = 1$. Prove that you can associate to every $S \subseteq [n]$ with $|S| \leqslant k$ an *actual* density $f_S : \{0,1\}^S \to \mathbb{R}_+$ (with respect to the uniform measure on $\{0,1\}^S$) such that for any $S$-junta $q$,

$$\langle \varphi, q\rangle = \langle f_S, q|_S\rangle\,.$$

Here, an $S$-junta is a function $q : \{0,1\}^n \to \mathbb{R}$ such that $q(x)$ only depends on the coordinates $\{x_i : i \in S\}$. This shows that, restricted to any set of at most $k$ bits, the pseudo-density agrees with an actual density.

Let $r = \mathrm{rank}_+(M_n^g)$. By definition, this means that for every $S \in \binom{[n]}{m}$, we can write

$$g_S(x) = M_n^g(x) = \sum_{i=1}^{r} A_i(S) B_i(x)$$

for some functions $A_i : \binom{[n]}{m} \to \mathbb{R}_+$ and $B_i : \{0,1\}^n \to \mathbb{R}_+$. (Here we are using a factorization $M_n^g = AB$ where $A_{S,i} = A_i(S)$ and $B_{x,i} = B_i(x)$.) Or, more succinctly:

$$g_S = \sum_{i=1}^{r} A_i(S) B_i\,. \tag{1.3}$$

Without loss of generality, we can assume that $\mathbb{E}\, B_i = 1$ for each $i = 1, 2, \dots, r$ because we can scale $B_i$ by some positive number to achieve this, and correspondingly scale $A_i$ to maintain equality in (1.3).

Define functions $\varphi_S : \{0,1\}^n \to R$ by $\varphi_S(x) = \varphi(x|_S)$ for every $S \in \binom{[n]}{m}$. We will try to prove that (1.3) requires $r$ to be large by averaging under $\varphi_S(x)$ on both sides. To illustrate, let's assume for the moment that each $B_i$ is actually a $d$-junta. In that case, we have

$$\mathop{\mathbb{E}}_{S} \mathop{\mathbb{E}}_{x} \varphi_S(x) g_S(x) = \mathop{\mathbb{E}}_{y \in \{0,1\}^m} \varphi(y) g(y) = \langle \varphi, g \rangle < 0\,. \tag{1.4}$$

On the other hand,

$$\sum_{i=1}^{r} \mathop{\mathbb{E}}_{S} \mathop{\mathbb{E}}_{x} A_i(S) B_i(x) = \sum_{i=1}^{r} \mathop{\mathbb{E}}_{S} A_i(S) \mathop{\mathbb{E}}_{x} \left[\varphi_S(x) B_i(x)\right] \geqslant 0 \tag{1.5}$$

because for each $i \in [m]$, it holds that $\mathbb{E}_x[\varphi_S(x) B_i(x)] = \mathbb{E}_{y \in \{0,1\}^m} \varphi(y) \mathbb{E}_x[B_i(x) \mid x|_S = y] \geqslant 0$, since $\varphi$ is $d$-locally positive and the function $y \mapsto \mathbb{E}_x[B_i(x) \mid x|_S = y]$ is a $d$-junta.

Combining (1.4) and (1.5) contradicts (1.3) in the case when each $B_i$ is a $d$-junta. To prove that (1.3) requires $r$ to be large in the general case, we will proceed in three steps:

1. **Truncation.** Most of the densities $\{B_i\}$ have small relative entropy to the uniform measure. More precisely, recalling the definition $\mathrm{Ent}(f) = \mathbb{E}[f \log f]$, we will show that most of contribution to (1.3) comes from $B_i$'s with $\mathrm{Ent}(B_i) \leqslant O(\log r)$. This is where we will use the assumption (for the sake of contradiction) that the rank $r$ is small.

2. **Junta approximation.** Every density $f$ can be approximated by $k$-juntas with respect to the tests $\{\varphi_S\}$, where $k \approx m \|\varphi\|_\infty^2 \,\mathrm{Ent}(f)$.

3. **Random restriction.** Our argument above required each $B_i$ to be a $d$-junta, and we only achieve $k$-juntas for some $k \approx m \log r$ and certainly $m \geqslant d$. Still, our junta approximators will be decent (we will still have $k$ much smaller than the trivial bound of $n$). The last step will use our random choice of $S \in \binom{[n]}{m}$. For a randomly chosen subset $S$, it will be very likely that only a few of the $k$ junta coordinates fall into $S$, so on most subsets we will get a $d$-junta, allowing the argument above to go through.

## 1.2 Truncation

This step doesn't rely on the specific structure of our problem, so we do it more generally. Consider finite sets $X$ and $Y$ and a matrix $M : X \times Y \to \mathbb{R}_+$. Suppose that $r = \text{rank}_+(M)$ and write

$$M(x,y) = \sum_{i=1}^{r} A_i(x) B_i(y)\,,$$

where $A_1, \ldots, A_r : X \to \mathbb{R}_+, B_1, \ldots, B_r : Y \to \mathbb{R}_+$.

The following discussion and definitions are probably overkill on a first reading, especially because the proofs themselves are rather simple. So we first state our goal. One might then skip to Section 1.3.

**Lemma 1.3** (Smooth non-negative factorizations)**.** *Let $r = \text{rank}_+(M)$. For any $\delta > 0$, there exists a matrix $\tilde{M} \in \mathbb{R}_+^{X \times Y}$ such that*

$$\mathop{\mathbb{E}}_{x \in X, y \in Y} |M(x,y) - \tilde{M}(x,y)| \leqslant \delta\,,$$

*and $\tilde{M} = UV$ where $U \in \mathbb{R}_+^{X \times k}$, $V \in \mathbb{R}_+^{k \times Y}$. If $\{U^{(i)}\}$ are the columns of $U$ and $\{V_i\}$ are the rows of $V$, then the factorization further satisfies*

$$\mathop{\mathbb{E}}_{y \in Y} V_i(y) = 1 \qquad i = 1, 2, \ldots, k$$

$$\max\{\|V_1\|_\infty, \ldots, \|V_k\|_\infty\} \leqslant r\|M\|_\infty\,,$$

*and*

$$\sum_{i=1}^{k} \|U^{(i)}\|_\infty \leqslant \frac{r}{\delta}\|M\|_\infty\,.$$

**Lower bounds via separating hyperplanes.** Recall that our goal is to show that $\text{rank}_+(M)$ is large. It would be nice if we could argue that $M$ cannot be too correlated with any map $(x,y) \mapsto A_i(x)B_i(y)$ and therefore $r$ must be large. This would avoid having to argue about a subtle relationship between $\{A_i\}$ and $\{B_i\}$ for different values of $i$. For instance, we could try to find a functional $F : X \times Y \to \mathbb{R}$ such that $\mathbb{E}_{x,y} F(x,y)M(x,y) < 0$ while $\mathbb{E}_{x,y} F(x,y)A_i(x)B_i(y) \geqslant 0$ for all $i = 1, \ldots, r$.

In other words, we would like to define a convex set of "low non-negative rank" matrices and show that $M$ is not in this set (by convex duality, this separation would always be accomplished with such a linear functional $F$). Note that matrices of the form $(x,y) \mapsto A_i(x)B_i(y)$ are exactly those of non-negative rank 1. But the convex hull of $\{N \in \mathbb{R}_+^{X \times Y} : \text{rank}_+(N) = 1\}$ is precisely the set of all non-negative matrices (which certainly contains $M$!).

Instead, let us proceed analytically. For simplicity, let us equip both $X$ and $Y$ with the uniform measure. Let $\mathcal{Q} = \{b : Y \to \mathbb{R}_+ \mid \|b\|_1 = 1\}$ denote the set of probability densities on $Y$, where we define $\|b\|_1 = \frac{1}{|Y|}\sum_{y \in Y} |b(y)|$.

Now define

$$\alpha_+(N) = \min\left\{ \max_{i \in [k]} \|B_i\|_\infty \cdot \sum_{i=1}^{k} \|A^{(i)}\|_\infty : A \in \mathbb{R}_+^{X \times k},\, B \in \mathbb{R}_+^{k \times Y} \text{ with } N = AB, \{B_1, \ldots, B_k\} \subseteq \mathcal{Q} \right\}$$

Here $\{A^{(i)}\}$ are the columns of $A$ and $\{B_i\}$ are the rows of $B$. Note that now $k$ is unconstrained.

Observe that the function $\alpha_+$ is convex (unlike the non-negative rank!). To see this, consider a pair $N = AB$ and $N' = A'B'$. Define $\gamma = \max_i \|B_i\|_\infty$ and $\gamma' = \max_i \|B'\|_\infty$, and write

$$\frac{N+N'}{2} = \begin{pmatrix} \frac{1}{2\gamma'}A & \frac{1}{2\gamma}A' \end{pmatrix} \begin{pmatrix} \gamma' B \\ \gamma B' \end{pmatrix},$$

witnessing the fact that

$$\alpha_+(\tfrac{1}{2}(N+N')) \leqslant \gamma\gamma' \left( \frac{1}{2\gamma'} \frac{\alpha(N)}{\gamma} + \frac{1}{2\gamma} \frac{\alpha(N')}{\gamma'} \right) = \frac{\alpha(N) + \alpha(N')}{2}.$$

### 1.2.1 Relating $\alpha_+$ and $\operatorname{rank}_+$

We will see now that low non-negative rank matrices are close to matrices with $\alpha_+$ small. In standard communication complexity/discrepancy arguments, this corresponds to discarding "small rectangles."

**Lemma 1.4.** *For every non-negative $M \in \mathbb{R}_+^{X\times Y}$ with $\operatorname{rank}_+(M) \leqslant r$ and every $\delta \in (0,1)$, there is a matrix $\tilde{M} \in \mathbb{R}_+^{X\times Y}$ such that*

$$\|M - \tilde{M}\|_1 \leqslant \delta$$

*and*

$$\alpha_+(\tilde{M}) \leqslant \frac{r^2\|M\|_\infty^2}{\delta}.$$

*Proof.* Suppose that $M = AB$ with $A \in \mathbb{R}_+^{X\times r}$, $B \in \mathbb{R}_+^{r\times Y}$, and let us interpret this factorization in the form

$$M(x,y) = \sum_{i=1}^{r} A_i(x) B_i(y) \tag{1.6}$$

(where $\{A_i\}$ are the columns of $A$ and $\{B_i\}$ are the rows of $B$). By rescaling the columns of $A$ and the rows of $B$, respectively, we may assume that $\mathbb{E}[B_i] = 1$ for every $i \in [r]$.

Let $\Lambda = \{i : \|B_i\|_\infty > \tau\}$ denote the "bad set" of indices (we will choose $\tau$ momentarily). Observe that if $i \in \Lambda$, then

$$\|A_i\|_\infty \leqslant \frac{\|M\|_\infty}{\tau},$$

from the representation (1.6) and the fact that all summands are positive.

Define the matrix $\tilde{M}(x,y) = \sum_{i\notin\Lambda} A_i(x)B_i(y)$. It follows that

$$\|M - \tilde{M}\|_1 = \mathop{\mathbb{E}}_{x,y}\left[|M(x,y) - \tilde{M}(x,y)|\right] = \sum_{i\in\Lambda} \mathop{\mathbb{E}}_{x,y}[A_i(x)B_i(y)].$$

Each of the latter terms is at most $\|A_i\|_\infty\|B_i\|_1 \leqslant \frac{\|M\|_\infty}{\tau}$ and $|\Lambda| \leqslant r$, thus

$$\|M - \tilde{M}\|_1 \leqslant r\frac{\|M\|_\infty}{\tau}.$$

Next, observe that

$$\mathop{\mathbb{E}}_{y}[M(x,y)] = \sum_{i=1}^{r} A_i(x)\|B_i\|_1 = \sum_{i=1}^{r} A_i(x),$$

implying that $\|A_i\|_\infty \leqslant \|M\|_\infty$ and thus $\sum_{i=1}^{r} \|A_i\|_\infty \leqslant r\|M\|_\infty$.

Setting $\tau = r\|M\|_\infty/\delta$ yields the statement of the lemma. ◻

### 1.3 Approximation by juntas

Let $B : \{0,1\}^n \to \mathbb{R}_+$ be a density and suppose we wish to approximate it by a "simple" density $\tilde{B}$ with respect to the family of tests $\{\varphi_S\}$. This fits exactly into the dual-sparse approximation framework of Lecture 3. For concreteness, we write down the corresponding minimum relative entropy optimization; the variables are the values $b(x)$ for $x \in \{0,1\}^n$, and $\varepsilon$ is a constant we will choose later:

$$\begin{aligned} &\text{minimize} && \mathrm{Ent}(b) && (1.7)\\ &\text{subject to} && \mathbb{E}[b] = 1 && (1.8)\\ & && b(x) \geqslant 0 \quad \forall x \in \{0,1\}^n && (1.9)\\ & && \langle \varphi_S, b\rangle \leqslant \langle \varphi_S, b\rangle + \varepsilon \quad \forall S \in \binom{[n]}{m}. && (1.10)\end{aligned}$$

From the dual-sparse approximation theorem, we know there exists a density $\tilde{B}$ of the form

$$\tilde{B} = \frac{\exp\left(\sum_S c_S \varphi_S\right)}{\mathbb{E}\exp\left(\sum_S c_S \varphi_S\right)},$$

where

$$\#\left\{S \in \binom{[n]}{m} : c_S \neq 0\right\} \leqslant 2\frac{\|\varphi\|_\infty^2}{\varepsilon^2}\mathrm{Ent}(B),$$

and such that $\langle \varphi_S, \tilde{B}\rangle \leqslant \langle \varphi_S, B\rangle + \varepsilon$ for all $S \in \binom{[n]}{m}$.

Note each $\varphi_S$ is an $m$-junta (since it depends only on the coordinates in $S$), hence the approximator $\tilde{B}$ is a $2m\frac{\|\varphi\|_\infty^2}{\varepsilon^2}\mathrm{Ent}(B)$-junta.

The troublesome "width" parameter $\|\varphi\|_\infty$ does not play such an important role for us presently, because it is bounded by some constant depending only on the function $g$. But to get improved bounds (or extend these techniques to other settings), the dependence on $\|\varphi\|_\infty$ is important.

### 1.4 Random restriction: Putting everything together

Recall (1.3) and (1.4). We now do the corresponding analysis for the right-hand side. Our goal is to prove a lower bound on the quantity

$$\mathop{\mathbb{E}}_{S,x} \varphi_S(x) M_n^g(S,x)$$

that depends on $r$. For $r$ small enough, we will contradict (1.4).

First, let us apply Lemma 1.3 with a parameter $\delta$ to obtain a matrix $\tilde{M}_n^g$ satisfying $\|M_n^g - \tilde{M}_n^g\|_1 \leqslant \delta$ and such that

$$\tilde{M}_n^g(S,x) = \sum_{i=1}^k A_i(x) B_i(y),$$

where each $B_i$ satisfies $\mathbb{E}\, B_i = 1$ and $\|B_i\|_\infty \leqslant r$ and where

$$\sum_{i=1}^k \|A_i\|_\infty \leqslant \frac{r}{\delta}\|g\|_\infty^2. \tag{1.11}$$

Observe that

$$\mathop{\mathbb{E}}_{S,x} \varphi_S(x) M_n^g(S,x) \geqslant -\delta \|\varphi\|_\infty + \mathop{\mathbb{E}}_{S,x} \varphi_S(x) \tilde{M}_n^g(S,x)\,, \tag{1.12}$$

so now we can focus on the latter term.

Note that

$$\|B_i\|_\infty \leqslant r \implies \mathrm{Ent}(B_i) = \mathbb{E}[B_i \log B_i] \leqslant \|\log B_i\|_\infty = \log r\,.$$

Apply the results of Section 1.3 with a parameter $\varepsilon > 0$ to obtain densities $\{\tilde{B}_i\}$ such that $\tilde{B}_i$ is an $h$-junta for

$$h \leqslant 2m \frac{\|\varphi\|_\infty^2}{\varepsilon^2} \log r\,, \tag{1.13}$$

and $\langle \varphi_S, \tilde{B}_i \rangle \leqslant \langle \varphi_S, B_i \rangle + \varepsilon$ for all $|S| = m$.

Then:

$$\begin{aligned}
\mathop{\mathbb{E}}_{S,x} \varphi_S(x) \tilde{M}_n^g(S,x) &= \sum_{i=1}^r \mathop{\mathbb{E}}_S A_i(S) \langle \varphi_S, B_i \rangle \\
&\geqslant \sum_{i=1}^r \mathop{\mathbb{E}}_S A_i(S) \left( \langle \varphi_S, \tilde{B}_i \rangle - \varepsilon \right) \\
&\geqslant -\varepsilon \|g\|_1 - \delta + \sum_{i=1}^r \mathop{\mathbb{E}}_S A_i(S) \mathop{\mathbb{E}}_x \varphi_S(x) \tilde{B}_i(x)\,,
\end{aligned}$$

where in the last line we used

$$\sum_{i=1}^r \mathop{\mathbb{E}}_S A_i(S) = \mathop{\mathbb{E}}_{S,x} M_n^g(S,x) \geqslant -\delta + \mathop{\mathbb{E}}_{S,x} M_n(S,x) = -\delta + \mathop{\mathbb{E}}_{S,x} g_S(x) = -\delta + \|g\|_1\,.$$

Now for the random restriction. For each $i = 1, \ldots, k$, let $J_i$ denote the set of variables on which $\tilde{B}_i$ depends. Recall that $|J_i| \leqslant h$. We have

$$\mathop{\mathbb{E}}_x \varphi_S(x) \tilde{B}_i(x) = \mathop{\mathbb{E}}_{y \in \{0,1\}^S} \varphi(y) \mathop{\mathbb{E}}_{x \in \{0,1\}^n} \left[ \tilde{B}_i(x) \mid x|_S = y \right]$$

Note that the map $y \mapsto \mathbb{E}_{x \in \{0,1\}^n} \left[ \tilde{B}_i(x) \,\middle|\, x|_S = y \right]$ is a junta on $J_i \cap S$. Thus if $|J_i \cap S| \leqslant d$, then the contribution from this term is non-negative since $\varphi$ is $d$-locally positive.

But if we think of $|S| = m$ as fixed and $n$ as growing, thus $|J_i \cap S| > d$ is quite rare. Formally,

$$\mathop{\mathbb{E}}_{S,x} \left[ \varphi_S(x) A_i(S) \tilde{B}_i(x) \right] \geqslant -\|A_i\|_\infty \mathbb{P}_S[|J_i \cap S| > d] \geqslant -\|A_i\|_\infty \frac{h^{d+1} (2m)^{d+1}}{n^{d+1}}\,.$$

In the last estimate, we have used a simple union bound and $n \geqslant 2m$.

Now if we recall (1.11), this yields

$$\sum_{i=1}^k \mathop{\mathbb{E}}_{S,x} \left[ \varphi_S(x) A_i(S) \tilde{B}_i(x) \right] \geqslant -\frac{r}{\delta} \|g\|_\infty^2 \frac{h^{d+1} (2m)^{d+1}}{n^{d+1}}\,.$$

Note that by choosing $n$ only moderately large, we will make this error term very small.

To choose parameters correctly, let's unwind our whole sequence of inequalities to obtain:

$$\langle \varphi, g\rangle = \mathop{\mathbb{E}}_{S,x} \varphi_S(x) M_n^g(S,x) \geqslant -\delta(1+\|\varphi\|_\infty) - \varepsilon\|g\|_1 - \frac{r}{\delta}\|g\|_\infty^2 \frac{h^{d+1}(2m)^{d+1}}{n^{d+1}} \tag{1.14}$$

Recall that we are seeking to obtain a contradiction to (1.4). Let $\beta = |\langle \varphi, g\rangle|$ and observe that $\beta, \|\varphi\|_\infty, \|g\|_1, \|g\|_\infty, m$ are all quantities that depend only on $g$. Since we are thinking of $g$ as fixed, the bound (1.13) yields $h = O(\log r)$.

By taking $\varepsilon, \delta$ sufficiently small (dependning only on $g$), we conclude that

$$r \geqslant c\left(\frac{n}{2m\log r}\right)^{d+1}$$

for some $c = c(g)$ sufficiently large . Now either $r \geqslant n^{d+1}$ and we are done, or else $m \log r = O(\log n)$, and we have prove the bound of Theorem 1.1.

# 2 Exercises

## 2.1 Stronger lower bounds

So far, we only saw how to prove a lower bound of $\bar{\gamma}(\mathrm{CUT}_n) \geqslant n^{\omega(1)}$. To obtain stronger quantitative lower bounds, one has to analyze carefully the parts of the argument that read "some constant depending only on $g$." To do this properly, it turns out that one needs an appropriate definition of "approximate" junta degree. Basically, the "proof" that a function has large junta degree (the locally positive functional) has to be robust.

For a function $f : \{0,1\}^n \to \mathbb{R}_+$ and $\varepsilon > 0$, define

$$\deg_{\mathrm{J}}^{\varepsilon}(f) = \min\left\{d : \forall\ d\text{-locally positive functionals } \varphi : \{0,1\}^n \to \mathbb{R},\ \mathop{\mathbb{E}}_x \varphi(x) f(x) > -\varepsilon\|\varphi\|_\infty \mathop{\mathbb{E}}_x f(x)\right\}.$$

**Exercise (1 point) 2.1.** Give an equivalent characterization of $\deg_{\mathrm{J}}^{\varepsilon}$ in terms of approximating $f$ by a non-negative sum of juntas, where the approximation is in the $L^1$-norm.

One can prove the following (see [Lee-Raghavendra-Steurer 2015]).

**Theorem 2.2.** *For any $g : \{0,1\}^m \to \mathbb{R}_+$ and $\varepsilon > 0$, the following holds. For all $n \geqslant 2m$,*

$$\mathrm{rank}_+(M_n^g) \geqslant \left(\frac{c\varepsilon^2 n}{m^2(m\log n + \log(\|g\|_\infty/\|g\|_1))}\right)^{\deg_{\mathrm{J}}^{\varepsilon}(g)},$$

*where $c > 0$ is a* universal constant.

The point of this result is that there is no hidden constant. This allows one to prove much stronger bounds.

**Exercise (1 point) 2.3.** In the definition of approximate junta degree, one notices the appearance of the uniform measure. More generally, Theorem 2.2 holds under the $p$-biased product measure $\mu_p^n$ for $p \in [0,1]$. Define:

$$\deg_{\mathrm{J}}^{\varepsilon,p}(f) = \min\Big\{d : \forall\ d\text{-locally positive functionals } \varphi : \{0,1\}^n \to \mathbb{R},$$

$$\underset{x\sim\mu_p^n}{\mathbb{E}} \varphi(x) f(x) > -\varepsilon \|\varphi\|_\infty \underset{x\sim\mu_p^n}{\mathbb{E}} f(x) \Bigg\} .$$

Prove that there is a constant $\varepsilon > 0$ such that for all $n \geqslant 3$, we have

$$\deg_{\mathsf{J}}^{\varepsilon,1/n}(f) \geqslant \frac{n}{2} + 1 ,$$

where $f(x) = (x_1 + \cdots + x_n - 1)^2$. Combined with Theorem 2.2, what lower bound does this yield for $\bar{\gamma}(\mathrm{CUT}_n)$ ?

## 2.2 Approximation

We have been concerned so far with *exact* characterization of polytopes (and, mainly, $\mathrm{CUT}_n$). But in this model, one can also talk about approximate lifts. For instance, consider the MAX-CUT problem: Given a non-negative weight $w : \binom{[n]}{2} \to \mathbb{R}_+$ on the edges of the (undirected) complete graph, the goal is to compute the *maximum-cut value*

$$\mathrm{opt}(w) \stackrel{\mathrm{def}}{=} \max_{z\in\mathrm{CUT}_n} \frac{\langle w, z\rangle}{\|w\|_1} .$$

The objective is the (normalized) weight of edges cut. (Strictly speaking, MAX-CUT involves finding the optimizer, not just its value.)

Fix the number of vertices $n$. For some constants $1 \geqslant c > s \geqslant 0$, let us consider the matrix

$$M^{c,s}(w,z) = c - \frac{\langle z, w\rangle}{\|w\|_1} ,$$

where $w$ ranges over all weighted graphs $w$ with $\mathrm{opt}(w) \leqslant s$ and $z$ ranges over all cuts (the extreme points of $\mathrm{CUT}_n$).

**Exercise (1 point) 2.4.** Argue that if $\mathrm{rank}_+(M^{c,s}) \geqslant r$, then the following holds. For any polytope $P$ defined by at most $r$ inequalities, if $P$ linearly projects to a polytope $\hat{P} \subseteq \mathbb{R}^{\binom{n}{2}}$ such that $\hat{P} \supseteq \mathrm{CUT}_n$, then there exists a weighted graph $w$ such that $\mathrm{opt}(w) \leqslant s$, but

$$\max_{z\in\hat{P}} \frac{\langle z, w\rangle}{\|w\|_1} \geqslant c .$$

In other words, $\hat{P}$ does a poor job of capturing even approximate MAX-CUT optimization over $\mathrm{CUT}_n$.