Lecture 4: Covering the large spectrum via dual-sparse approximation
CSE 599S: Entropy optimality, Winter 2016
Instructor: James R. Lee **Last updated:** January 19, 2016

1 Discrete Fourier analysis

In this lecture, we use the dual-sparse approximation theorem from the last lecture to prove some results in discrete Fourier analysis. For simplicity, we restrict ourselves to the setting of $G = \mathbb{F}_2^n$, but the theorems hold (when suitably restated) for any finite abelian group G.

Fourier analysis over $\mathbb{F}_2^n$. We use $\mathbb{F}_2 = \{0, 1\}$ to denote the field on two elements. Let $G = \mathbb{F}_2^n$ be equipped with the uniform measure μ. We use $\hat{G} = \mathbb{F}_2^n$ to denote the dual group (though we use the notations G and $\hat{G}$ to distinguish primal and dual objects). We will use the definitions from Lecture 3 (Section 3).

For every $\gamma \in \hat{G}$, we define the corresponding character $u_\gamma : G \to \mathbb{R}$ by

$$u_\gamma(x) = (-1)^{\gamma_1 + \cdots + \gamma_n} .$$

The functions $\{u_\gamma : \gamma \in \hat{G}\}$ form an orthornormal basis for $L^2(G, \mu)$, and thus every $f \in L^2(G, \mu)$ can be written uniquely as

$$f = \sum_{\gamma \in \hat{G}} \hat{f}(\gamma) u_\gamma ,$$

where $\hat{f}(\gamma) = \langle f, u_\gamma \rangle$.

We will be interested in the "large spectrum" of a function $f \in L^2(G, \mu)$: For a parameter $\delta > 0$, define

$$\mathrm{Spec}_\delta(f) = \{\gamma \in \hat{G} : |\hat{f}(\gamma)| > \delta\} .$$

Say that a subset $S \subseteq \hat{G}$ is *d-covered* if

$$S \subseteq \left\{ \sum_{\lambda \in \Lambda} a_\lambda \lambda : a_\lambda \in \{-1, 0, 1\} \right\} \tag{1.1}$$

for some $\Lambda \subseteq \hat{G}$ with $|\Lambda| \leqslant d$. When $G = \mathbb{F}_2^n$, (1.1) is the same as saying that S is contained in the span of Λ (in the vector space $\mathbb{F}_2^n$).

1.1 Chang's Lemma

Recall that $\Delta_G = \{f : G \to \mathbb{R}_+ : \mathbb{E}_\mu f = 1\}$ is the set of densities on G (with respect to the uniform measure μ).

Lemma 1.1 (Chang)**.** *For any $f \in \Delta_G$ and $\delta > 0$, the set $\mathrm{Spec}_\delta(f)$ is d-covered for*

$$d \leqslant 2 \frac{\mathrm{Ent}_\mu(f)}{\delta^2} .$$

Proof. We prove this using Theorem 3.1 (the dual-sparse approximation theorem) from Lecture 3. Let $\mathcal{F} = \{\pm u_\gamma : \gamma \in \hat{G}\}$ and apply the approximation theorem with $\varepsilon = \delta$. Since $\|u_\gamma\|_\infty = 1$ for all $\gamma \in \hat{G}$, we obtain a density $\tilde{f} \in \Delta_G$ such that

$$\tilde{f} = \frac{\exp\left(\sum_{i=1}^m c_i u_{\gamma_i}\right)}{\mathbb{E}_\mu \exp\left(\sum_{i=1}^m c_i u_{\gamma_i}\right)}, \tag{1.2}$$

for some real constants $\{c_i\}$ and $\gamma_1, \ldots, \gamma_m \in \hat{G}$, and $m \leqslant \frac{2}{\delta^2}\mathrm{Ent}_\mu(f)$, and furthermore $\mathrm{Spec}_\delta(f) \subseteq \mathrm{Spec}_0(\tilde{f})$ because from the approximation property for every $\gamma \in \mathrm{Spec}_\delta(f)$, we have

$$|\hat{\tilde{f}}(\gamma)| = |\langle u_\gamma, \tilde{f}\rangle| \geqslant |\langle u_\gamma f\rangle| - \delta > 0\,.$$

Thus we are left to prove that $\mathrm{Spec}_0(\tilde{f})$ can be m-covered. To this end, use the Taylor expansion $e^x = \sum_{k=0}^\infty \frac{x^k}{k!}$ to see that the non-zero Fourier coefficients of $\tilde{f}$ must be products of the form

$$\prod_{i\in\alpha} u_{\gamma_i} = u_{\sum_{i\in\alpha}\gamma_i}$$

for some subset $\alpha \subseteq [m]$. Therefore $\mathrm{Spec}_0(\tilde{f}) \subseteq \left\{\sum_{i=1}^m a_i\gamma_i : a_i \in \{-1, 0, 1\}\right\}$, and we conclude that indeed $\mathrm{Spec}_0(\tilde{f})$ is m-covered, completing the proof. □

Remark 1.2. The essential use of $G = \mathbb{F}_2^n$ in the preceding argument came in the last step, where we argued that the sum $\sum_{i\in\alpha}\gamma_i$ can be written as a linear combination with only $\{-1, 0, 1\}$ coefficients (indeed, only with $\{0, 1\}$ coefficients). This relies on the fact that we are working over $\mathbb{F}_2$ so that $2\gamma = \gamma + \gamma = 0$ for all $\gamma \in \mathbb{F}_2^n$. Doing the same argument over $G = \left(\mathbb{Z}/p\mathbb{Z}\right)^n$ would lose a factor of p in the bound on d. While this might be fine for p small and n large, it becomes uninteresting in the case $n = 1$, say.

Exercise 1.1. Prove that the bound in Lemma 1.1 is tight by considering, for n odd, the density $f : \mathbb{F}_2^n \to \mathbb{R}_+$ given by

$$f(x) = \begin{cases} 2 & \sum_{i=1}^n x_i > n/2 \\ 0 & \sum_{i=1}^n x_i < n/2\,. \end{cases}$$

You may need to consult the O'Donnell book to understand the Fourier spectrum of f.

1.2 Bloom's Lemma

In [Bloom, 2014], the following variant of Chang's lemma is proved.

Lemma 1.3 (Bloom). *For any $f \in \Delta_G$ and $\delta > 0$, there is a subset $S \subseteq \mathrm{Spec}_\delta(f)$ satisfying $|S| \geqslant \delta|\mathrm{Spec}_\delta(f)|$ and such that S is d-covered for*

$$d \leqslant O(1)\frac{\mathrm{Ent}_\mu(f)}{\delta} + O\left(\frac{\log(1/\delta)}{\log\log(1/\delta)}\right). \tag{1.3}$$

Note that the second term in the bound (1.3) is only important when $\mathrm{Ent}_\mu(f) \ll 1$ (which is not a particularly interesting regime).

To prove this, we need a variant of the dual-sparse approximation theorem.

Theorem 1.4. *Consider some $\mathcal{F} \subseteq L^2(X,\mu)$. Let $f \in \Delta_X$ and $\varepsilon > 0$ be given. Then there exist non-negative constants $\{c_\varphi : \varphi \in \mathcal{F}\}$ such that*

$$\sum_{\varphi \in \mathcal{F}} c_\varphi \leqslant \frac{\max_{\varphi \in \mathcal{F}} \|\varphi\|_\infty}{\varepsilon} \mathrm{Ent}_\mu(f),$$

and the density

$$\tilde{f} = \frac{\exp\left(\sum_{\varphi \in \mathcal{F}} c_\varphi \varphi\right)}{\mathbb{E}_\mu \exp\left(\sum_{\varphi \in \mathcal{F}} c_\varphi \varphi\right)} \tag{1.4}$$

satisfies $\langle \tilde{f}, \varphi \rangle \geqslant \langle f, \varphi \rangle - \varepsilon$ for all $\varphi \in \mathcal{F}$.

There are two ways to prove this. One is to revisit the proof of Theorem 3.1 from Lecture 3. Let us assume (by scaling) that $\max_{\varphi \in \mathcal{F}} \|\varphi\|_\infty \leqslant 1$. Then the number of non-zero coefficients c_φ is bounded by $O(h/\varepsilon^2)$ where $h = \mathrm{Ent}_\mu(f)$ because the decrease in the potential function for fixing an ε-violated constraint is proportional to ε^2, and the potential can only change by h over the course of the algorithm. On the other hand, to achieve this potential decrease, we only "move" (exponentially) by ε in direction of the violated constraint. So each of the $\approx h/\varepsilon^2$ phases only increases the sum of coefficients by ε, leading to the bound of $\approx h/\varepsilon$. A second method of proof simply computes the dual of a convex program.

Exercise (2 points) 1.1. Let $\mathcal{F} \subseteq L^2(X,\mu)$ be a family satisfying $\|\varphi\|_\infty \leqslant 1$ for $\varphi \in \mathcal{F}$. Let $C(\delta) \subseteq L^2(X,\mu)$ be the polytope described by the linear inequality constraints:

$$C(\delta) = \left\{ g \in L^2(X,\mu) : \langle g, \varphi \rangle \geqslant \langle f, \varphi \rangle - \delta \right\}.$$

Given f and $\varepsilon > 0$, consider the optimization:

$$\underset{g,\delta}{\text{minimize}} \quad \left\{ \mathrm{Ent}_\mu(g) + \frac{\mathrm{Ent}_\mu(f)}{\varepsilon}\delta : g \in C(\delta) \cap \Delta_X, \delta \geqslant 0 \right\}$$

Show that (i) the optimal solution (g^*, δ^*) is unique, (ii) it satisfies $\delta^* \leqslant \varepsilon$, and (iii) that

$$g^* = \frac{\exp\left(\sum_{\varphi \in \mathcal{F}} c_\varphi \varphi\right)}{\mathbb{E}_\mu \exp\left(\sum_{\varphi \in \mathcal{F}} c_\varphi \varphi\right)}$$

satisfies $\sum_{\varphi \in \mathcal{F}} c_\varphi \leqslant \frac{\mathrm{Ent}_\mu(f)}{\varepsilon}$.
[Hint: This can be done by understanding Chapter 5 (Duality) of the Boyd-Vandenberghe book. For convex programs of this form, the dual can be calculated explicitly.]

Now we prove Bloom's lemma in the $\mathbb{F}_2^n$ case.

Proof of Lemma 1.3. We will apply Theorem 1.4 with $\mathcal{F} = \{\pm u_\gamma : \gamma \in \hat{G}\}$ and $\varepsilon = \delta/3$. Let $\tilde{f}$ be the resulting approximator from (1.4). Observe that from the approximation property (with respect to the functionals in $\mathcal{F}$), we have

$$\mathrm{Spec}_\delta(f) \subseteq \mathrm{Spec}_{2\delta/3}(\tilde{f}). \tag{1.5}$$

By scaling the numerator and denominator by the same constant, we can write

$$\tilde{f} = \frac{\exp\left(\sum_{\gamma\in\hat{G}} c_\gamma(1+\varphi_\gamma)\right)}{\mathbb{E}_\mu \exp\left(\sum_{\gamma\in\hat{G}} c_\gamma(1+\varphi_\gamma)\right)},$$

where $\varphi_\gamma \in \{-u_\gamma, u_\gamma\}$ and $\sum_{\gamma\in\hat{G}} c_\gamma \leqslant \frac{\mathrm{Ent}_\mu(f)}{\varepsilon}$. In particular, since $|\varphi_\gamma| \leqslant 1$, every term in the sum is non-negative everywhere.

Note also that

$$\left\| \sum_{\gamma\in\hat{G}} c_\gamma(1+\varphi_\gamma) \right\|_\infty \leqslant 2\frac{\mathrm{Ent}_\mu(f)}{\varepsilon}.$$

Let $p_m(x) = \sum_{k=0}^m \frac{x^k}{k!}$ be the degree-m truncation of the Taylor series for e^x. We can use Taylor's theorem to write

$$\sup_{x\in[0,B]} \frac{|e^x - p_m(x)|}{e^x} \leqslant \frac{B^{m+1}}{m!}.$$

In particular, we can choose $m \leqslant 3B + O\left(\frac{\log(1/\delta)}{\log\log(1/\delta)}\right)$ with $B = 2\frac{\mathrm{Ent}_\mu(f)}{\varepsilon}$ so that if

$$g = \frac{p_m\left(\sum_{\gamma\in\hat{G}} c_\gamma(1+\varphi_\gamma)\right)}{\mathbb{E}_\mu\, p_m\left(\sum_{\gamma\in\hat{G}} c_\gamma(1+\varphi_\gamma)\right)} \in \Delta_G,$$

then $\|\tilde{f} - g\|_1 \leqslant \delta/3$. Observe that for any $\gamma \in \hat{G}$,

$$|\hat{\tilde{f}}(\gamma) - \hat{g}(\gamma)| = |\langle \tilde{f} - g, u_\gamma\rangle| \leqslant \|\tilde{f} - g\|_1 \cdot \|u_\gamma\|_\infty \leqslant \delta/3,$$

hence $\mathrm{Spec}_{2\delta/3}(\tilde{f}) \subseteq \mathrm{Spec}_{\delta/3}(g)$. Combined with (1.5), this yields $\mathrm{Spec}_\delta(f) \subseteq \mathrm{Spec}_{\delta/3}(g)$. Thus we now focus on g.

By expanding out p_m, we can write

$$g = \sum_{k=0}^m \sum_{\alpha\in\hat{G}^k} c_\alpha \prod_{i=1}^k (1+\varphi_{\alpha_i})$$

for some non-negative constants $\{c_\alpha\}$. Let us write $g = \sum_\alpha c_\alpha R_\alpha$ (and recall that every summand involves a vector α with at most m coordinates).

Define a probability distribution on terms in this sum (indexed by α):

$$p_\alpha = c_\alpha \mathop{\mathbb{E}}_\mu R_\alpha.$$

The fact that $\sum_\alpha p_\alpha = 1$ follows from $\mathbb{E}_\mu g = 1$. So we have $g = \sum_\alpha p_\alpha \bar{R}_\alpha$ where $\bar{R}_\alpha = R_\alpha / \mathbb{E}_\mu(R_\alpha)$.

Observe that for any $\psi \in L^2(X,\mu)$, we have

$$\sum_\alpha p_\alpha |\langle \psi, \bar{R}_\alpha\rangle| \geqslant |\langle \psi, g\rangle| \tag{1.6}$$

Consider $\psi = u_\gamma$ for some $\gamma \in \mathrm{Spec}_{\delta/3}(g)$. If we choose α randomly according to the distribution $\{p_\alpha\}$, then (1.6) yields $\mathbb{E}[|\langle u_\gamma, \bar{R}_\alpha\rangle|] \geqslant \delta/3$. On the other hand, $|\langle u_\gamma, \bar{R}_\alpha\rangle| \leqslant 1$ holds with probability one, hence

$$\mathbb{P}[\gamma \in \mathrm{Spec}_0(\bar{R}_\alpha)] \geqslant \delta/3.$$

In particular, there must exist some α such that $|\mathrm{Spec}_0(\bar{R}_\alpha) \cap \mathrm{Spec}_\delta(f)| \geqslant \frac{\delta}{3}|\mathrm{Spec}_\delta(f)|$, recalling that $\mathrm{Spec}_\delta(f) \subseteq \mathrm{Spec}_{\delta/3}(g)$.

Finally, observe that since $|\alpha| \leqslant m$, it follows that $\mathrm{Spec}_0(\bar{R}_\alpha)$ is m-covered since the non-zero Fourier coefficients of R_α correspond to those generated by sums of the characters $\alpha_1, \ldots, \alpha_m$ (and hence by $\{0,1\}$ sums of such characters). As in the proof of Lemma 1.1 (see Remark 1.2), this latter fact is only true over $\mathbb{F}_2^n$. □

2 Some open problems

These exercises are a bit open-ended.

Exercise (3+ points) 2.1. The proof of Lemma 1.3 proceeds by expanding the truncated power series for e^x and then sampling its terms at random. This is a bit mysterious. It seems plausible that one could prove it instead using a stochastic variant of the online mirror descent algorithm (see, e.g., [Bubeck, 2014]) or perhaps simply by writing the correct convex program as in Exercise 1.1.

Exercise (3+ points) 2.2. Here is a sparse approximation problem in auction design (that I learned from Matt Weinberg). There is one seller who is selling n items to one bidder. It's only one example of an array of similar questions.

Let $V_1, V_2, \ldots, V_n$ be independent random variables taking values in $[0,1]$. The value of a set of items $S \subseteq [n]$ to the bidder is $\sum_{i \in S} V_i$. The seller's goal is to maximize the (expected) revenue. It is known that, without loss, we can assume that a bidder acting in their own self interest is truthful (i.e., always reports their true valuation). Thus our goal is to design a revenue-maximizing truthful auction.

Denote by $\mathcal{V} = \mathcal{V}_1 \times \cdots \times \mathcal{V}_n \subseteq [0,1]^n$ the space of possible value vectors. For every $v \in \mathcal{V}$, the linear program has variables $\{x_i(v) : i = 1, 2, \ldots, n\}$ representing the probability that the bidder receives item i in the auction, and $p(v_1, \ldots, v_n)$ representing the price the bidder is charged (and thus pays).

For $i = 1, 2, \ldots, n$ our input consists of the probability mass functions $\pi_i : \mathcal{V}_i \to [0,1]$ for each V_i. Let us denote $\pi(v) = \pi_1(v_1)\pi_2(v_2)\cdots\pi_n(v_n)$.

Now the goal is to maximize (expected) revenue:

$$\text{maximize} \qquad \sum_{v \in \mathcal{V}} \pi(v) p(v)$$

subject to the basic constraints:

$$\begin{aligned} x_i(v) &\in [0,1] \qquad && i \in \{1, 2, \ldots, n\}, v \in \mathcal{V} \\ p(v) &\geqslant 0 \qquad && v \in \mathcal{V}\,. \end{aligned}$$

There is also a set of truthfulness constraints:

$$\sum_{i=1}^n v_i x_i(v) - p(v) \geqslant \sum_{i=1}^n v_i x_i(w) - p(w) \qquad \text{for all } v, w \in \mathcal{V}\,. \tag{2.1}$$

Let us assume that $(0, 0, \ldots, 0) \in \mathcal{V}$. Otherwise, we should add the rationally constraints:

$$p(v) \leqslant \sum_{i=1}^n v_i x_i(v) \qquad \text{for all } v \in \mathcal{V}\,.$$

The solution to this (infinite) linear program provides an optimal mechanism; the question is about whether there is a near-optimal mechanism with much smaller "menu complexity." In other words, we would like an auction that achieves expected revenue $R^* - \varepsilon n$ where R^* is the maximal expected revenue, but where the description of the auctioneer is simple. Can one construct a "simple" auction here using a dual-sparse approximation?

Note: It is acceptable to also relax the constraints (2.1) by subtracting $-\sqrt{\varepsilon} n$ from the right-hand side. (There are ways to convert such an auction to a truthful one losing only $\approx -\varepsilon n$ in the revenue.)