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 $\mu$. We use $\hat{G} = \mathbb{F}_2^\ast$ 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 orthonormal 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

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

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

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

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

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

$$d \leq 2 \frac{\text{Ent}_\mu(f)}{\delta^2}.$$
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 \( \epsilon = \delta \). Since \( \| u_\gamma \|_\infty = 1 \) for all \( \gamma \in \hat{G} \), we obtain a density \( \hat{f} \in \Delta_G \) such that
\[
\hat{f} = \frac{\exp \left( \sum_{i=1}^m c_i u_\gamma \right)}{\mathbf{E}_\mu \exp \left( \sum_{i=1}^m c_i u_\gamma \right)} ,
\]
for some real constants \( \{ c_i \} \) and \( \gamma_1, \ldots, \gamma_m \in \hat{G} \), and \( m \leq \frac{2}{\delta} \) \( \text{Ent}_\mu (f) \), and furthermore \( \text{Spec}_\delta (f) \subseteq \text{Spec}_0 (\hat{f}) \) because from the approximation property for every \( \gamma \in \text{Spec}_\delta (f) \), we have
\[
|\hat{f}(\gamma')| = |\langle u_\gamma, \hat{f} \rangle| \geq |\langle u_\gamma f \rangle| - \delta > 0.
\]
Thus we are left to prove that \( \text{Spec}_0 (\hat{f}) \) 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 \( \hat{f} \) 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 \( \text{Spec}_0 (\hat{f}) \subseteq \{ \sum_{i=1}^m a_i \gamma_i : a_i \in \{-1, 0, 1\} \} \), and we conclude that indeed \( \text{Spec}_0 (\hat{f}) \) 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 \( G = (\mathbb{Z}/p\mathbb{Z})^n \) 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 \( f : \mathbb{F}_2^n \rightarrow \mathbb{R}_+ \) 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 \( S \subseteq \text{Spec}_\delta (f) \) satisfying \( |S| \geq \delta |\text{Spec}_\delta (f)| \) and such that \( S \) is \( d \)-covered for
\[
d \leq O(1) \frac{\text{Ent}_\mu (f)}{\delta} + O \left( \frac{\log(1/\delta)}{\log \log(1/\delta)} \right) .
\]

Note that the second term in the bound (1.3) is only important when \( \text{Ent}_\mu (f) \ll 1 \) (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 \( F \subseteq L^2(X, \mu) \). Let \( f \in \Delta_X \) and \( \varepsilon > 0 \) be given. Then there exist non-negative constants \( \{ c_\varphi : \varphi \in F \} \) such that

\[
\sum_{\varphi \in F} c_\varphi \leq \frac{\max_{\varphi \in F} \| \varphi \|_\infty}{\varepsilon} \operatorname{Ent}_\mu(f),
\]

and the density

\[
\tilde{f} = \frac{\exp \left( \sum_{\varphi \in F} c_\varphi \varphi \right)}{\mathbb{E}_\mu \exp \left( \sum_{\varphi \in F} c_\varphi \varphi \right)}
\]

satisfies \( \langle \tilde{f}, \varphi \rangle \geq \langle f, \varphi \rangle - \varepsilon \) for all \( \varphi \in F \).

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 \( \max_{\varphi \in F} \| \varphi \|_\infty \leq 1 \). Then the number of non-zero coefficients \( c_\varphi \) is bounded by \( O \left( \frac{h}{\varepsilon^2} \right) \) where \( h = \operatorname{Ent}_\mu(f) \) because the decrease in the potential function for fixing an \( \varepsilon \)-violated constraint is proportional to \( \varepsilon^2 \), 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 \( \varepsilon \) in direction of the violated constraint. So each of the \( \approx \frac{h}{\varepsilon^2} \) phases only increases the sum of coefficients by \( \varepsilon \), leading to the bound of \( \approx \frac{h}{\varepsilon} \). A second method of proof simply computes the dual of a convex program.

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

\[
C(\delta) = \{ g \in L^2(X, \mu) : \langle g, \varphi \rangle \geq \langle f, \varphi \rangle - \delta \}.
\]

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

\[
\min_{g, \delta} \left\{ \operatorname{Ent}_\mu(g) + \frac{\operatorname{Ent}_\mu(f)}{\varepsilon} \delta : g \in C(\delta) \cap \Delta_X, \delta \geq 0 \right\}
\]

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

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

satisfies \( \sum_{\varphi \in F} c_\varphi \leq \frac{\operatorname{Ent}_\mu(f)}{\varepsilon} \).

[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 \( F_2^\mathbb{R} \) case.

Proof of Lemma 1.3. We will apply Theorem 1.4 with \( F = \{ \pm u_\gamma : \gamma \in \hat{G} \} \) 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 \( F \)), we have

\[
\operatorname{Spec}_\delta(f) \subseteq \operatorname{Spec}_{2\delta/3}(\tilde{f}).
\]

(1.5)
By scaling the numerator and denominator by the same constant, we can write
\[ \tilde{f} = \frac{\exp \left( \sum_{\gamma \in \hat{G}} c_\gamma (1 + \varphi_\gamma) \right)}{E_\mu \exp \left( \sum_{\gamma \in \hat{G}} c_\gamma (1 + \varphi_\gamma) \right)}, \]
where \( \varphi_\gamma \in \{-u_\gamma, u_\gamma\} \) and \( \sum_{\gamma \in \hat{G}} c_\gamma \leq \frac{\text{Ent}_\mu(f)}{\varepsilon} \). In particular, since \( |\varphi_\gamma| \leq 1 \), every term in the sum is non-negative everywhere.

Note also that
\[ \left\| \sum_{\gamma \in \hat{G}} c_\gamma (1 + \varphi_\gamma) \right\|_\infty \leq 2 \frac{\text{Ent}_\mu(f)}{\varepsilon}. \]

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
\[ \sup_{x \in [0, 1]} \left| \frac{e^x - p_m(x)}{e^x} \right| \leq B^{m+1}. \]

In particular, we can choose \( m \leq 3B + O \left( \frac{\log(1/\delta)}{\log \log(1/\delta)} \right) \) with \( B = 2 \frac{\text{Ent}_\mu(f)}{\varepsilon} \) so that
\[ g = \frac{p_m \left( \sum_{\gamma \in \hat{G}} c_\gamma (1 + \varphi_\gamma) \right)}{E_\mu p_m \left( \sum_{\gamma \in \hat{G}} c_\gamma (1 + \varphi_\gamma) \right)} \in \Delta_G, \]
then \( \| \tilde{f} - g \|_1 \leq \delta/3 \). Observe that for any \( \gamma \in \hat{G}, \)
\[ |\hat{f}(\gamma) - \hat{g}(\gamma)| = |\langle \tilde{f} - g, u_\gamma \rangle| \leq \| \tilde{f} - g \|_1 \cdot \| u_\gamma \|_\infty \leq \delta/3, \]

hence \( \text{Spec}_{\delta/3}(\tilde{f}) \subseteq \text{Spec}_{\delta/3}(g) \). Combined with (1.5), this yields \( \text{Spec}_0(f) \subseteq \text{Spec}_{\delta/3}(g) \). 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 \( \alpha \) with at most \( m \) coordinates).

Define a probability distribution on terms in this sum (indexed by \( \alpha \))
\[ p_\alpha = c_\alpha E_\mu R_\alpha. \]

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

Observe that for any \( \psi \in L^2(X, \mu) \), we have
\[ \sum_{\alpha} p_\alpha |\langle \psi, \tilde{R}_\alpha \rangle| \geq |\langle \psi, g \rangle| \quad (1.6) \]

Consider \( \psi = u_\gamma \) for some \( \gamma \in \text{Spec}_{\delta/3}(g) \). If we choose \( \alpha \) randomly according to the distribution \( \{p_\alpha\} \), then (1.6) yields \( E[|\langle u_\gamma, \tilde{R}_\alpha \rangle|] \geq \delta/3 \). On the other hand, \( |\langle u_\gamma, \tilde{R}_\alpha \rangle| \leq 1 \) holds with probability one, hence
\[ \mathbb{P}[\gamma \in \text{Spec}_0(\tilde{R}_\alpha)] \geq \delta/3. \]
In particular, there must exist some \( \alpha \) such that \( |\text{Spec}_0(\bar{R}_\alpha) \cap \text{Spec}_\delta(f)| \geq \frac{\delta}{3}|\text{Spec}_\delta(f)| \), recalling that \( \text{Spec}_\delta(f) \subseteq \text{Spec}_{\delta/3}(g) \).

Finally, observe that since \( |\alpha| \leq m \), it follows that \( \text{Spec}_\delta(\bar{R}_\alpha) \) is \( m \)-covered since the non-zero Fourier coefficients of \( R_\alpha \) 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 \( \mathbb{F}_2^n \). \( \square \)

## 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 \( \mathcal{V} = V_1 \times \cdots \times V_n \subseteq [0, 1]^n \) 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:

\[
\text{maximize } \sum_{v \in \mathcal{V}} \pi(v)p(v)
\]

subject to the basic constraints:

\[
x_i(v) \in [0, 1] \quad i \in \{1, 2, \ldots, n\}, v \in \mathcal{V} \\
p(v) \geq 0 \quad v \in \mathcal{V}.
\]

There is also a set of truthfulness constraints:

\[
\sum_{i=1}^n v_i x_i(v) - p(v) \geq \sum_{i=1}^n v_i x_i(w) - p(w) \quad \text{for all } v, w \in \mathcal{V}.
\] (2.1)

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

\[
p(v) \leq \sum_{i=1}^n v_i x_i(v) \quad \text{for all } v \in \mathcal{V}.
\]
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.)