Covering the large spectrum and generalized Riesz products

James R. Lee*

Abstract

Chang's Lemma is a widely employed result in additive combinatorics. It gives optimal bounds on the dimension of the large spectrum of probability distributions on finite abelian groups. In this note, we show how Chang's Lemma and a powerful variant due to Bloom both follow easily from an approximation theorem for probability measures in terms of generalized Riesz products. The latter result involves no algebraic structure. The proofs are correspondingly elementary.

1 Introduction

Let G be a finite abelian group. Chang's Lemma [Cha02] asserts that, for every large subset $S \subseteq G$, the large Fourier coefficients of the indicator function $\mathbf{1}_S$ lie in a low-dimensional subspace. This has seen a large number of applications in additive combinatorics (in addition to Chang's original application to Freiman's theorem).

A theorem of Bloom [Blo14] shows that a large subset of the large spectrum can be contained in an even lower-dimensional subspace. We refer to Section 2.1 for the formal statements. Bloom employs his theorem as the key tool in obtaining the following quantitative version of Roth's theorem.

Theorem 1.1. *There exists a $c > 0$ such that for all sufficiently large N, the following holds: If $A \subseteq \{1, \ldots, N\}$ contains no non-trivial three-term arithmetic progression, then*

$$|A| \leqslant c \frac{(\log \log N)^4}{\log N} N \,.$$

This improves slightly over Sanders' [San11] breakthrough result that has $(\log \log N)^4$ replaced by $(\log \log N)^6$.

In this note, we state a general approximation theorem about probability measures on finite spaces equipped with no algebraic structure. From this theorem, the results of Chang and Bloom follow easily. Quite a few proofs of Chang's Lemma are known, and arguably the argument of [IMR14] is the simplest.

Indeed, that proof can be seen as closely related to the one presented here. See Section 2.3 for some remarks about the connection and the distinction that is drawn in light of Bloom's result. On the other hand, Bloom's proof uses the additive structure in a seemingly fundamental and intricate way; our argument is elementary and requires only a direct application of the fact that the characters of a finite abelian group are homomorphisms and bounded in ℓ_∞.

The statement and proof are inspired by the "entropy maximization" philosophy: Given a probability measure μ and a collection of linear observables $\mathcal{F}$, one can find a "simple" approximator

*University of Washington

$\tilde{\mu}$ (with respect to $\mathcal{F}$) by maximizing the entropy of $\tilde{\mu}$ over all probability measures having similar behavior on $\mathcal{F}$.

Our use of this philosophy borrows heavily from the work [LRS15] where it is employed in the setting of quantum states and von Neumann entropy. It is also related, at least in spirit, to the works [Gow10] and [RTTV08] on "dense model theorems," and to a long line of works employing an "entropy regularizer" in the setting of convex optimization.

In the next section, we state the approximation theorem in the context of finite probability spaces. We focus on the finite case for simplicity; there are no significant difficulties in extending the results at least to the setting of compact measure spaces (and compact abelian groups).

In Section 2.1, we show how the theorems of Chang and Bloom follow from the approximation theorem, and in Section 2.2, we prove the approximation result. Finally, in Section 2.3, we discuss how the statement and proof of the approximation theorem fall out naturally from the perspective of convex optimization.

2 An approximation theorem

Let X be a finite set equipped with a probability measure μ. We use $L^2(\mu)$ to denote the Hilbert space of real-valued functions on X equipped with inner product $\langle f, g\rangle = \sum_{x\in X} \mu(x) f(x) g(x)$.

For a function $h : X \to \mathbb{R}$, we will use the notation $\mathbb{E}_\mu h = \sum_{x\in X} \mu(x) h(x)$. We also denote by $\|h\|_p = (\mathbb{E}_\mu |h|^p)^{1/p}$ the $L^p(\mu)$ norm for $p \geqslant 1$.

Denote the set of densities with respect to μ by $\Delta_X = \{f : X \to [0,\infty) : \|f\|_1 = 1\}$. For $f \in \Delta_X$, define the relative entropy

$$\mathrm{Ent}_\mu(f) = \mathbb{E}_\mu[f \log f]\,.$$

Generalized Riesz products. Suppose that $\mathcal{F} \subseteq L^2(\mu)$ is a collection satisfying $\sup_{\varphi\in\mathcal{F}} \|\varphi\|_\infty \leqslant 1$. Define the semi-norm $\|f\|_{\mathcal{F}} = \sup_{\varphi\in\mathcal{F}} |\langle \varphi, f\rangle|$.

Say that a function $R \in L^2(\mu)$ is a *degree-d Riesz $\mathcal{F}$-product* if

$$R(x) = \prod_{i=1}^{d} (1 + \varepsilon_i \varphi_i(x))$$

for some $d \geqslant 1$ and $\varphi_1, \ldots, \varphi_d \in \mathcal{F}$, $\varepsilon_1, \ldots, \varepsilon_d \in \{-1, 0, 1\}$. Observe that every such R is non-negative on X.

Theorem 2.1 (Approximation Theorem)**.** *For every $0 < \eta < \frac{1}{e^3}$ and $f \in \Delta_X$, there is a $g \in \Delta_X$ such that:*

1. $\|f - g\|_{\mathcal{F}} \leqslant \eta$.
2. *There is a subset $\mathcal{F}' \subseteq \mathcal{F}$, with*

$$|\mathcal{F}'| \leqslant 9\frac{\mathrm{Ent}_\mu(f)}{\eta^2}\,,$$

and such that g is a non-negative linear combination of degree-d Riesz $\mathcal{F}'$-products for

$$d \leqslant 18\frac{\mathrm{Ent}_\mu(f)}{\eta} + O\left(\frac{\log\frac{1}{\eta}}{\log\log\frac{1}{\eta}}\right).$$

2.1 The theorems of Chang and Bloom

Let G be a finite abelian group equipped with the uniform measure μ, and let $\hat{G}$ be the dual group. Let 0 denote the identity element in G and $\hat{G}$.

For $\gamma \in \hat{G}$, let $u_\gamma : G \to \mathbb{C}$ denote the corresponding character. One can write any $f : G \to \mathbb{C}$ as $f = \sum_{\gamma \in \hat{G}} \hat{f}(\gamma) u_\gamma$. We need only the properties that $u_\gamma u_{\gamma'} = u_{\gamma+\gamma'}$ for all $\gamma, \gamma' \in \hat{G}$ and $\max_{x \in G} |u_\gamma(x)| \leqslant 1$. One may consult [TV10, Ch. 4] for a treatment of discrete Fourier analysis tailored to applications in additive combinatorics.

For each value $\delta > 0$, we define the set

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

Say that a subset $S \subseteq \hat{G}$ is *d-covered* if there exists a subset $\Lambda \subseteq \hat{G}$ with $|\Lambda| \leqslant d$ such that

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

Let us define the family

$$\mathcal{F} = \left\{ \mathrm{Re}\, u_\gamma, \mathrm{Im}\, u_\gamma : \gamma \in \hat{G} \right\} \subseteq L^2(\mu).$$

Note that $\|\varphi\|_\infty \leqslant 1$ for every $\varphi \in \mathcal{F}$.

Lemma 2.2. *If R is a degree-d Riesz $\mathcal{F}$-product, then $\mathrm{Spec}_0(R) = \{\gamma \in \hat{G} : \hat{R}(\gamma) \neq 0\}$ is d-covered.*

Proof. Write $R = \prod_{i=1}^d (1 - \varepsilon_i \varphi_i)$ for $\{\varphi_i\} \subseteq \mathcal{F}$ and $\{\varepsilon_i\} \subseteq \{-1, 0, 1\}$. For each i, let $\gamma_i \in \hat{G}$ be such that $\varphi_i = \pm \mathrm{Re}\, u_{\gamma_i}$ or $\varphi_i = \pm \mathrm{Im}\, u_{\gamma_i}$. Since we can write $\mathrm{Re}\, u_\gamma = \frac{1}{2}(u_\gamma + u_{-\gamma})$ and $\mathrm{Im}\, u_\gamma = \frac{1}{2i}(u_\gamma - u_{-\gamma})$, upon expanding the product defining R, we see that every $\gamma \in \hat{G}$ with $\hat{R}(\gamma) \neq 0$ is a sum of elements from $\{\gamma_1, \ldots, \gamma_d, -\gamma_1, \ldots, -\gamma_d\} \subseteq \hat{G}$. (We are using the convention here that the empty sum is equal to the identity of $\hat{G}$ in order to handle $\hat{R}(0) \neq 0$.) □

The preceding proof yields easily the following fact as well.

Lemma 2.3. *If $h : G \to \mathbb{R}$ is a linear combination of Riesz $\mathcal{F}'$-products for some $\mathcal{F}' \subseteq \mathcal{F}$, then $\mathrm{Spec}_0(h)$ is $|\mathcal{F}'|$-covered.*

Denote $\Delta_G = \{f : G \to [0, \infty) : \mathbb{E}_\mu f = 1\}$. Now we can prove the theorems of Chang [Cha02] and Bloom [Blo14].

Theorem 2.4 (Chang)**.** *For every $f \in \Delta_G$ and $\delta > 0$, the set $\mathrm{Spec}_\delta(f)$ is d-covered for*

$$d \leqslant 18 \frac{\mathrm{Ent}_\mu(f)}{\delta^2}.$$

Proof. Set $\eta = \delta/\sqrt{2}$. Then there exists a $g \in \Delta_G$ satisfying the conclusion of Theorem 2.1 applied to f. Note that $\|f - g\|_{\mathcal{F}} \leqslant \eta$ implies $\mathrm{Spec}_\delta(f) \subseteq \mathrm{Spec}_0(g)$. Combining Theorem 2.1(2) and Lemma 2.3 yields the desired conclusion. □

We remark that a tighter analysis using the same method is able to achieve a leading constant of $2\sqrt{2}$ in general, and 2 in the case $G = \mathbb{F}_2^n$ (which is tight; see [IMR14]).

Theorem 2.5 (Bloom)**.** *For every $f \in \Delta_G$ and $0 < \delta < \frac{1}{e^3}$, there exists a subset $S \subseteq \mathrm{Spec}_\delta(f)$ such that $|S| \geqslant \frac{\delta}{2}|\mathrm{Spec}_\delta(f)|$ and S is d-covered for*

$$d \leqslant 36\sqrt{2} \frac{\mathrm{Ent}_\mu(f)}{\delta} + O\left(\frac{\log \frac{1}{\delta}}{\log \log \frac{1}{\delta}} \right).$$

Proof. Setting $\eta = \delta/(2\sqrt{2})$ and applying Theorem 2.1, there exists a $g \in \Delta_G$ such that

$$g = \sum_{i=1}^{N} c_i R_i$$

with $N \geqslant 1$, $c_1, \ldots, c_N > 0$, and where $R_1, \ldots, R_N$ are degree-d Riesz $\mathcal{F}$-products for d as in Theorem 2.1(3), and furthermore $\|f - g\|_{\mathcal{F}} \leqslant \eta$.

Observe that since $g \in \Delta_G$, we have $\sum_{i=1}^{N} c_i \mathbb{E}_\mu R_i = \mathbb{E}_\mu g = 1$. Thus we can define a random variable $Z \in \{1, 2, \ldots, N\}$ so that

$$\mathbb{P}[Z = i] = c_i \mathbb{E}_\mu R_i \,.$$

Since $\|f - g\|_{\mathcal{F}} \leqslant \eta$, we deduce that if $\gamma \in \mathrm{Spec}_{2\sqrt{2}\eta}(f)$, then $\gamma \in \mathrm{Spec}_{\sqrt{2}\eta}(g)$. For such γ, we have

$$\mathop{\mathbb{E}}_{z}\left[\left|\langle u_\gamma, \tfrac{R_z}{\mathbb{E}_\mu R_z}\rangle\right|\right] = \sum_{i=1}^{N} c_i(\mathbb{E}_\mu R_i)\left|\left\langle u_\gamma, \tfrac{R_i}{\mathbb{E}_\mu R_i}\right\rangle\right| \geqslant |\langle u_\gamma, g\rangle| \geqslant \sqrt{2}\eta = \frac{\delta}{2}\,.$$

Because $\left|\left\langle u_\gamma, \tfrac{R_i}{\mathbb{E}_\mu R_i}\right\rangle\right| \leqslant 1$, we conclude that

$$\mathop{\mathbb{P}}_{z}\left(\hat{R}_z(\gamma) \neq 0\right) = \mathop{\mathbb{P}}_{z}\left(|\langle u_\gamma, R_z\rangle| > 0\right) \geqslant \frac{\delta}{2}\,.$$

By linearity, $\mathbb{E}_z\,|\mathrm{Spec}_0(R_z)| \geqslant \frac{\delta}{2}\,|\mathrm{Spec}_\delta(f)|$. Moreover, by Lemma 2.3, every set $\mathrm{Spec}_0(R_i)$ is d-covered. Thus there exists at least one such set that completes the proof of the theorem. □

2.2 Proof of the Approximation Theorem

We now prove Theorem 2.1. We assume that $\eta > 0$ and $f \in \Delta_X$ are given as in the theorem.

The gradient descent. We need the notion of the *relative entropy* between two densities $h, h' \in \Delta_X$:

$$\mathbb{D}_\mu(h \,\|\, h') = \mathop{\mathbb{E}}_{\mu}\left[h \log \frac{h}{h'}\right].$$

This definition makes sense whenever $\mathrm{supp}(h) \subseteq \mathrm{supp}(h')$. Otherwise, we take the value to be $+\infty$.

For some value $T > 0$, define a family $\{g_t : t \in [0, T]\} \subseteq \Delta_X$ by

$$g_t = \frac{\exp\left(\int_0^t \varphi_s\, ds\right)}{\mathbb{E}_\mu \exp\left(\int_0^t \varphi_s\, ds\right)}\,, \tag{2.1}$$

where $s \mapsto \varphi_s \in L^2(\mu)$ is a measurable function to be specified shortly. Observe that $g_0 = \mathbf{1}$ is the constant 1 function.

A simple calculation yields: For $t \in [0, T)$,

$$\frac{d}{dt}\mathbb{D}_\mu(f \,\|\, g_t) = \langle \varphi_t, g_t - f\rangle\,. \tag{2.2}$$

We define the maps $s \mapsto \varphi_s$ to be piecewise constant on a finite sequence of intervals. Given the definition on intervals $[0, t_1), [t_1, t_2), \ldots, [t_{i-1}, t_i)$ with $0 < t_1 < t_2 < \cdots < t_i$, we define it on an interval $[t_i, t_{i+1})$ as follows.

If there exists a functional $\varphi \in \mathcal{F}$ such that

$$|\langle g_{t_i}, \varphi\rangle - \langle f, \varphi\rangle| > \frac{2\eta}{3},$$

then we put

$$\varphi_s = \operatorname{sign}\big(\langle f - g_{t_i}, \varphi\rangle\big) \cdot \varphi \tag{2.3}$$

for $s \in [t_i, t_{i+1})$ where $t_{i+1} = \inf\{t \geqslant t_i : |\langle g_t, \varphi\rangle - \langle f, \varphi\rangle| \leqslant \eta/3\}$. We will see momentarily why such a t_{i+1} must exist.

If there is no such functional φ at time t_i, then we set $T = t_i$ and $i_{\max} = i$. By construction, we have the property that $\|f - g_T\|_{\mathcal{F}} \leqslant \frac{2}{3}\eta$.

Lemma 2.6. $T \leqslant 3\frac{\mathrm{Ent}_\mu(f)}{\eta}$.

Proof. Simply observe that for $t \in [0, T)$, the calculation (2.2) combined with the definition of the sequence $\{t_i\}$ and the choice (2.3) yields

$$\frac{d}{dt}\mathbb{D}_\mu(f \,\|\, g_t) \leqslant -\frac{\eta}{3}.$$

On the other hand, $\mathbb{D}_\mu(f \,\|\, g_0) = \mathrm{Ent}_\mu(f)$ and $D_\mu(f \,\|\, g_t) \geqslant 0$ is always true. This yields the claim. □

Lemma 2.7. *It holds that* $i_{\max} \leqslant 9\frac{\mathrm{Ent}_\mu(f)}{\eta^2}$.

Proof. Fix an interval $[t_{i-1}, t_i)$ with $i \leqslant i_{\max}$. Let $\varphi = \varphi_{t_{i-1}}$. We calculate

$$\frac{d}{dt}\langle \varphi, g_t\rangle = -\langle \varphi, g_t(\varphi - \langle \varphi, g_t\rangle)\rangle = -\langle \varphi^2, g_t\rangle + \langle \varphi, g_t\rangle^2.$$

Notice that the latter quantity is at least $-\|\varphi\|_\infty^2\|g_t\|_1 \geqslant -1$. Therefore $t_i - t_{i-1} \geqslant \frac{\eta}{3}$. We conclude that $i_{\max} \leqslant 3T/\eta$ and combine this with Lemma 2.6. □

Truncating the exponential. Note now that

$$g_T = \frac{\exp\left(\int_0^T \varphi_s\, ds\right)}{\mathbb{E}_\mu \exp\left(\int_0^T \varphi_s\, ds\right)} = \frac{\exp\left(\int_0^T (1+\varphi_s)\, ds\right)}{\mathbb{E}_\mu \exp\left(\int_0^T (1+\varphi_s)\, ds\right)} \tag{2.4}$$

with $\|f - g_T\|_{\mathcal{F}} \leqslant 2\eta/3$. We are left to truncate the Taylor expansion of e^x to approximate g_T by a non-negative combination of Riesz $\mathcal{F}$-products.

Toward this end, let $p_m(x) = \sum_{j \leqslant m} \frac{x^j}{j!}$ and recall that for $B \geqslant 0$,

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

Let us choose $m \leqslant 3B + O\left(\frac{\log\frac{1}{\eta}}{\log\log\frac{1}{\eta}}\right)$ so as to make this quantity less than $\frac{\eta}{4}$.

Now put $B = 2T$ and $\psi = \int_0^T (1+\varphi_s)\, ds$. Note that $\|\psi\|_\infty \leqslant 2T$. Applying Taylor approximation to the numerator in (2.4) yields

$$\left\|e^\psi - p_m(\psi)\right\|_1 \leqslant \frac{\eta}{4}\mathbb{E}_\mu(e^\psi).$$

Finally, we define:

$$g = \frac{p_m(\psi)}{\mathbb{E}_\mu\, p_m(\psi)}\,.$$

Observe that $g \in \Delta_X$ and $\|g_T - g\|_1 \leqslant \frac{\eta}{4}\frac{\mathbb{E}_\mu\, e^\psi}{\mathbb{E}_\mu\, p_m(\psi)} \leqslant \frac{\eta}{4(1-\frac{\eta}{4})} \leqslant \frac{\eta}{3}$. This implies $\|g_T - g\|_{\mathcal{F}} \leqslant \frac{\eta}{3}$ since $\|\varphi\|_\infty \leqslant 1$ for $\varphi \in \mathcal{F}$. Therefore $\|f - g\|_{\mathcal{F}} \leqslant \|f - g_T\|_{\mathcal{F}} + \|g_T - g\|_{\mathcal{F}} \leqslant \eta$.

Let $\mathcal{F}' = \{\varphi_s : s \in [0, T)\} \subseteq \mathcal{F}$ be the set of functionals encountered during the argument. Note that $|\mathcal{F}'| \leqslant i_{\max}$. Moreover, we have $\psi = \sum_{\varphi\in\mathcal{F}'} c_\varphi(1 + \varphi)$ for some non-negative constants $\{c_\varphi\}$. Thus $p_m(\psi)$ expands into a non-negative linear combination of degree-m Riesz $\mathcal{F}'$-products, hence the same is true for g, completing the proof.

2.3 Remarks on convex programming and duality

The form of (2.1) is not accidental. Indeed, given $f \in \Delta_X$ and $\eta > 0$, one can consider the problem of minimizing $\mathrm{Ent}_\mu(g)$ over all $g \in \Delta_X$ satisfying $\|f - g\|_{\mathcal{F}} \leqslant \eta$. This is the problem of minimizing a convex function over a polytope.

The optimization is clearly feasible since f itself is a valid solution. Moreover, because of the allowed error $\eta > 0$, the program satisfies Slater's condition and thus strong duality holds (see [BV04, Ch. 5]). Setting the gradient of the Lagrangian equal to zero, one arrives at the form of the optimal solution

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

where $\{\lambda^*_\varphi \in \mathbb{R} : \varphi \in \mathcal{F}\}$ are the dual variables corresponding to the linear constraints implicit in $\|f - g\|_{\mathcal{F}} \leqslant \eta$.

Moreover, the algorithm and analysis of Section 2.2 are based on the "mirror descent" framework, analyzed using a Bregman divergence (in this case, the relative entropy). See, for instance, the monograph [Bub14]. The sparsity of the solution (captured by the "junta" property) is closely related to sparsity properties of the Frank-Wolfe algorithm [FW56].

Dependence in the constraints. As mentioned in the introduction, the work [IMR14] gives a proof of Chang's Lemma related to the one presented here. In particular, the authors also employ an entropy functional. A fundamental difference is that they first apply a change of basis so that the "approximation problem" becomes completely elementary—the "algorithm" can work independently in every coordinate.

A basic feature of the mirror descent approach is that one does not need the constraints to be disjoint: There is a global potential that controls convergence even as some constraints might become more or less violated. This difference is highlighted in a comparison between the theorems of Chang and Bloom.

Say that a subset $\Lambda \subseteq \hat{G}$ is *disassociated* if Λ cannot be d-covered for $d < |\Lambda|$. In proving Chang's theorem, we could have chosen a maximal disassociated subset Λ of the large spectrum and only required that our approximating density g have its Fourier coefficients in Λ be similar to those of f (this corresponds to setting $\mathcal{F} = \{\mathrm{Re}\, u_\gamma, \mathrm{Im}\, u_\gamma : \gamma \in \Lambda\}$).

This choice makes the constraints of the optimization problem independent and its analysis straightforward, merely using sub-additivity of entropy. On the other hand, for Bloom's theorem, we do not know ahead of time which disassociated set Λ will be important, and so the power of the Bregman-style convergence analysis seems to be more essential.

Acknowledgements

We thank Thomas Bloom, Prasad Raghavendra, and Julia Wolf for enlightening discussions.

References

[Blo14] T. Bloom. A quantitative improvement for Roth's theorem on arithmetic progressions. arXiv:1405.5800, 2014. 1, 3

[Bub14] S. Bubeck. Theory of convex optimization for machine learning. arXiv:1405.4980, 2014. 6

[BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004. 6

[Cha02] Mei-Chu Chang. A polynomial bound in Freiman's theorem. *Duke Math. J.*, 113(3):399–419, 2002. 1, 3

[FW56] Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval Res. Logist. Quart.*, 3:95–110, 1956. 6

[Gow10] W. T. Gowers. Decompositions, approximate structure, transference, and the Hahn-Banach theorem. *Bull. Lond. Math. Soc.*, 42(4):573–606, 2010. 2

[IMR14] Russell Impagliazzo, Cristopher Moore, and Alexander Russell. An entropic proof of Chang's inequality. *SIAM J. Discrete Math.*, 28(1):173–176, 2014. 1, 3, 6

[LRS15] J. R. Lee, P. Raghavendra, and D. Steurer. Lower bounds on the size of semidefinite programming relaxations. arXiv:1411.6317, 2015. 2

[RTTV08] Omer Reingold, Luca Trevisan, Madhur Tulsiani, and Salil P. Vadhan. Dense subsets of pseudorandom sets. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 76–85, 2008. 2

[San11] Tom Sanders. On Roth's theorem on progressions. *Ann. of Math. (2)*, 174(1):619–636, 2011. 1

[TV10] Terence Tao and Van H. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. 3