Almost k-wise vs. k-wise independent permutations, and uniformity for general group actions

Noga Alon *
Tel-Aviv University and the Institute for Advanced Study
nogaa@tau.ac.il

Shachar Lovett†
The Institute for Advanced Study
slovett@math.ias.edu

December 10, 2011

Abstract

A family of permutations in S_n is k-wise independent if a uniform permutation chosen from the family maps any distinct k elements to any distinct k elements equally likely. Efficient constructions of k-wise independent permutations are known for $k = 2$ and $k = 3$, but are unknown for $k \geq 4$. In fact, it is known that there are no nontrivial subgroups of S_n for $n \geq 25$ which are 4-wise independent. Faced with this adversity, research has turned towards constructing almost k-wise independent families, where small errors are allowed. Optimal constructions of almost k-wise independent families of permutations were achieved by several authors.

Our first result is that any such family with small enough error is statistically close to a distribution which is perfectly k-wise independent. This allows for a simplified analysis of algorithms: an algorithm which uses randomized permutations can be analyzed assuming perfect k-wise independence, and then applied to an almost k-wise independent family. In particular, it allows for an oblivious derandomization of two-sided randomized algorithms which work correctly given any k-wise independent distribution of permutations.

Another model is that of weighted families of permutations, or equivalently distributions of small support. We establish two results in this model. First, we show that a small random set of $n^{O(k)}$ permutations w.h.p supports a k-wise independent distribution. We then derandomize this by showing that any almost $2k$-wise independent family supports a k-wise independent distribution. This allows for oblivious derandomization of algorithms for search problems which work correctly given perfect k-wise independent distributions.

These results are all in fact special cases of a general framework where a group acts on a set. In the aforementioned case, the group of permutations acts on tuples of k elements. We prove all the above results in the general setting of the action of a finite group on a finite set.

*Supported in part by an ERC advanced grant and by NSF grant DMS-0835373.
†Supported by NSF grant DMS-0835373.

1 Introduction

Small probability spaces of limited independence are widely used in many applications. Specifically, if the analysis of a randomized algorithm depends only on the assumption that the entries are k-wise independent, one can replace the random tape by a tape selected from a k-wise independent distribution. One application of this is a derandomization of the algorithm by enumerating over all possible random strings. Another application is when the random string needs to be saved, for example in data structures, where using k-wise independence allows one to maintain a succinct data structure.

The case of k-wise independent distributions over $\{0,1\}^n$ has been widely studied, and there are optimal constructions of k-wise independent probability spaces of size $n^{O(k)}$ (see e.g. [ABI86]). Moreover, these constructions are *strongly explicit*: given an index of an element $i \in [n^{O(k)}]$ and an index of a bit $j \in [n]$, one can compute the j-th bit of the i-th string in time $O(k \log n)$. This is crucial for several applications, for example for streaming algorithms and cryptography, where operations need to be performed in poly-logarithmic time.

Another widely studied case is that of k-wise independent permutations of n elements. This problem is motivated by cryptographic applications, as k-wise independent permutations allow perfect secrecy even if one allows k oracle queries to the encryption. For more details on the role of k-wise independent permutations in cryptography, see, e.g., [RW06, Vau98, Vau00, Vau03].

Here, the situation is much less understood. For $k = 2$ the group of invertible affine transformations $x \mapsto ax + b$ over a finite field $\mathbb{F}$ yields a 2-wise independent family; and for $k = 3$ the group of Möbius transformations $x \mapsto (ax + b)/(cx + d)$ with $ad - bc = 1$ over the projective line $\mathbb{F} \cup \{\infty\}$ yields a 3-wise independent family. For $k \geq 4$ (and n large enough), however, no k-wise independent family is known, other than the full symmetric group S_n and the alternating group A_n. In fact, it is known (c.f.., e.g., [Cam95], Theorem 5.2) that for $n \geq 25$ and $k \geq 4$ there are no other subgroups of S_n which form a k-wise independent family[1]. This is a major obstacle, while as groups are by no means the only way to produce such families, algebraic techniques are among the most useful in combinatorics, and the lack of algebraic structure is a serious drawback.

Faced with this adversity, research has turned towards constructing families of permutations which are *almost k-wise independent*, allowing for small errors. There has been much research towards constructing explicit almost k-wise independent families of minimal size. This was achieved, up to polynomial factors, by Kaplan, Naor and Reingold [KNR05], who gave a construction of such a family of size $n^{O(k)}$. Alternatively, one can start with the constant size expanding set of S_n given by Kassabov [Kas07], and take a random walk on it of length $O(k \log n)$. Both of these constructions are also strongly explicit: given an index of a permutation $i \in [n^{O(k)}]$ and an element $j \in [n]$, one can compute the image of the i-th permutation on j in time $O(k \log n)$. Again, this is crucial for applications such as streaming algorithms or cryptography.

For many applications, almost k-wise independent families are just as good as perfect k-wise independent families. However, the analysis must take into account the error, which in some cases is not trivial. Our first result shows that by choosing the error small enough, one can analyze an algorithm using perfect k-wise independent permutations, and then apply almost k-wise independent permutations to achieve almost the same results.

Theorem 1.1. *Let μ be a distribution taking values in S_n which is almost k-wise independent with error $\varepsilon \cdot n^{-O(k)}$. Then there exists a distribution over permutations μ' which is k-wise independent, and such that the statistical distance between μ and μ' is at most ε.*

[1] In the language of group theory, these are k-transitive groups. The currently known proof of this fact is hard, as it requires the classification of finite simple groups.

A similar result for k-wise independent hash functions was obtained by Alon, Goldreich and Mansour [AGM03], and more generally over product spaces by Rubinfeld and Xie [RX10]. Our proof technique is similar in spirit, although technically more involved. This allows for an oblivious derandomization of two-sided algorithms which "work" given any k-wise independent distribution over permutations: let f be a boolean function, and let A be a randomized algorithm such that

$$\Pr_{\pi\sim\mu}[A(x,\pi) = f(x)] \geq 2/3$$

for any k-wise independent distribution over permutations μ. Then A can be derandomized by letting π be chosen uniformly from an almost k-wise independent distribution with error $n^{-O(k)}$. Since such distributions can be generated strongly explicitly, the overhead (in terms of the number of bits needed to sample from the distribution) is just $O(k \log n)$.

A relaxation of the problem of constructing small families of k-wise independent permutations is that of considering weighted families, or equivalently distributions of small support which are k-wise independent. Contrary to the case of unweighted families, it is simple to establish that there exist distributions of small support which are k-wise independent. First, note that given a family S of permutations, it is easy to decide if there exists a distribution μ supported on S which is k-wise independent, using linear programming: for a permutation π define the matrix $M_k(\pi)$ to be the permutation on distinct k-tuples induced by π. It is an $(n)_k \times (n)_k$ permutation matrix, where $(n)_k := \prod_{i=0}^{k-1}(n-i)$. Let U denote the uniform matrix all whose elements are $(n-k)!/n!$. Then there exists a k-wise independent distribution supported on S iff U belongs to the convex hull of $\{M_k(\pi) : \pi \in S\}$. The latter condition can be easily verified using linear programming. Now, starting with any set of permutations which support k-wise independent permutations (for example the set of all permutations), one can apply Carathéodory theorem, and deduce that U lies in the convex hull of at most n^{2k} permutations. That is, there exist k-wise independent distributions which are supported on at most n^{2k} permutations. Moreover, and somewhat surprisingly, one can algorithmically find a k-wise independent distribution with small support in a *weakly explicit* manner (i.e. in time $n^{O(k)}$) using the ideas of Karp and Papadimitriou [KP82] and Koller and Megiddo [KM94][2].

We consider the problem of constructing small explicit sets which support k-wise independent distributions. First, we establish that most small sets support k-wise independent distributions.

Theorem 1.2. *Let S be a random subset of S_n of size n^{6k}. Then with high probability (w.h.p, for short) there exists a distribution μ supported on S which is k-wise independent.*

A similar result for k-wise independent hash functions was obtained by Austrin and Håstad [AH11]. Our result implies a somewhat surprising consequence for search algorithms which "work" given any k-wise independent distribution over permutations, which allows to transform weak guarantees to strong guarantees. Let f be a function and A an algorithm, such that for any k-wise independent distribution μ,

$$\Pr_{\pi\sim\mu}[A(x,\pi) = f(x)] > 0.$$

Then since almost all sets of size $n^{O(k)}$ support such a distribution, we must have that A has a noticeable fraction of witnesses in S_n,

$$\Pr_{\pi\in S_n}[A(x,\pi) = f(x)] \geq n^{-O(k)}.$$

[2]Essentially, the linear program for finding μ has $n!$ variables and $n^{O(k)}$ constraints. Its dual has $n^{O(k)}$ variables and $n!$ constraints. The dual problem can be solved efficiently using the ellipsoid method since it has an efficient separating-hyperplane oracle.

We also show that almost $2k$-wise independent permutations give an explicit construction of a set which supports k-wise independence, thus derandomizing Theorem 1.2.

Theorem 1.3. *Let S be a subset of S_n such that S is almost $2k$-wise independent with error $n^{-O(k)}$. Then there exists a distribution μ supported on S which is k-wise independent.*

We are not aware of a similar result, even in the case of k-wise independent hash functions. This allows for an oblivious derandomization of search algorithms which "work" given any k-wise independent distribution over permutations: let f be a function, and let A be a randomized algorithm such that

$$\Pr_{\pi\sim\mu}[A(x,\pi) = f(x)] > 0$$

for any k-wise independent distribution μ over permutations. Then taking S to be an almost $2k$-wise independent family of permutations with error $n^{-O(k)}$, we get that there exists $\pi \in S$ for which $A(x,\pi) = f(x)$, achieving an oblivious derandomization of A with overhead (measured in bits, as before) $O(k \log n)$.

Here is a toy example illustrating the way the last theorem and the discussion preceding it can be applied. Let $G = (V, E)$ be a graph on a set V of n vertices, and suppose that each vertex $v \in V$ has a real positive weight $w(v)$. Let $d(v)$ be the degree of v, and assume all degrees are bounded by k. We claim that G contains an independent set $U \subset V$ of total weight $W(U) = \sum_{u\in U} w(u)$ at least $\sum_{v\in V} \frac{w(v)}{d(v)+1}$. To prove it, let π be a random permutation of the set of vertices V, and let U consist of all vertices u so that $\pi(u)$ precedes $\pi(v)$ for every neighbor v of u. It is clear that U is an independent set, and for any vertex $u \in V$ the probability that $u \in U$ is exactly $\frac{1}{d(u)+1}$, as this is the probability that u precedes all its neighbors. By linearity of expectation, the expected value of the total weight of U is $\sum_{v\in V} \frac{w(v)}{d(v)+1}$ and hence there exists an independent set U of total weight at least as claimed.

The above proof clearly works even if π is only assumed to be $(k+1)$-wise independent (in fact, a weaker condition suffices, we only need π to be $(k+1)$-minwise independent). Therefore, the discussion preceding Theorem 1.3 implies that if π is chosen uniformly at random, then the probability it provides a set U satisfying $W(U) \geq \sum_{v\in V} \frac{w(v)}{d(v)+1}$, is at least $n^{-O(k)}$. The theorem itself shows that the support of any set of almost $(2k+2)$-wise independent permutations with sufficiently small error must contain a permutation π that provides an independent set U as above.

A similar reasoning can be applied to other arrangement problems. Given a k-uniform hypergraph with a weight for each permutation of the vertices in each of its edges, one may want to find a permutation maximizing the total weight of all orders induced on the sets of vertices in the edges. Problems of this type are called k-CSP-rank problems, (see, e.g., [AA07]), and include Betweenness and Feedback Arc Set. In most of these problems, finding the precise optimum is NP-hard, and the reasoning above provides some insight about algorithms for the (much easier) problem of finding a permutation in which the total weight is at least as large as the expected weight in a uniform random permutation.

1.1 Group action uniformity vs. almost uniformity

We actually prove all the aforementioned results in the general setting of *group actions*, of which k-wise independent permutations as well as k-wise independent random variables form specific instances. A group G acts on a set X if G acts as a group of permutations on X. That is, $g : X \to X$ is a permutation of X for all $g \in G$, and $(gh)(x) = g(h(x))$ for all $g, h \in G$ and $x \in X$. This gives a general framework: k-wise independent permutations correspond to the case of $G = S_n$

the group of permutations, and $X = [n]_k = \{i_1, \ldots, i_k \in [n] \text{ distinct}\}$ the set of (ordered) distinct k-tuples, where the action of G on X is straightforward. The case of k-wise independent distributions over $\{0,1\}^n$ corresponds to $G = \mathbb{F}_2^n$ and $X = [n]_k \times \mathbb{F}_2^k$, where the action of $g = (g_1, \ldots, g_n) \in \mathbb{F}_2^n$ on $x = ((i_1, \ldots, i_k), (b_1, \ldots, b_k)) \in [n]_k \times \mathbb{F}_2^k$ is given by $g(x) = ((i_1, \ldots, i_k), (b_1 + g_{i_1}, \ldots, b_k + g_{i_k}))$. Similarly, one can obtain in this way distributions supporting k-wise independent random variables, even when each variable is distributed over a different domain.

We now introduce some definitions. If G acts on X, a distribution μ over G is X-uniform if

$$\Pr_{g \sim \mu}[g(x) = y] = \Pr_{g \in G}[g(x) = y]$$

for all $x, y \in X$; and is almost X-uniform with error ε if

$$\left| \Pr_{g \sim \mu}[g(x) = y] - \Pr_{g \in G}[g(x) = y] \right| \leq \varepsilon$$

for all $x, y \in X$. These definitions coincide with k-wise independence and almost k-wise independence for permutations when $G = S_n$ and $X = [n]_k$. Theorem 1.1, Theorem 1.2 and Theorem 1.3 are immediate corollaries of the following general theorems, when applied to $G = S_n$ and $X = [n]_k$.

First, we show that distributions over G which are almost X-uniform with small enough error, are close in statistical distance to distributions which are X-uniform.

Theorem 2.1 (informal version). *Let μ be a distribution over G which is almost X-uniform with error $\varepsilon \cdot |X|^{-O(1)}$. Then there exists a distribution μ' on G which is X-uniform, and such that the statistical distance between μ and μ' is at most ε.*

Second, we show that a small random subset of G supports w.h.p a X-uniform distribution.

Theorem 3.1 (informal version). *Let $S \subset G$ be a random set of size $|X|^{O(1)}$. Then w.h.p there exists a distribution μ supported on S which is X-uniform.*

Finally, we derandomized Theorem 3.1. Recall that if G acts on X, then G also acts on $X \times X$ in the obvious manner, i.e. $g((x_1, x_2)) = (g(x_1), g(x_2))$. We show that if a distribution over G is almost $X \times X$ uniform with a small enough error, then it must support an X-uniform distribution.

Theorem 4.1 (informal version). *Let μ be a distribution supported on a set $S \subset G$ which is almost $(X \times X)$-uniform with error $|X|^{-O(1)}$. Then there exists a distribution μ' supported on S which is X-uniform.*

The proof of Theorem 3.1 is by a counting argument using the symmetry of the group action. The proofs of Theorem 2.1 and Theorem 4.1 rely on representation theory of finite groups. In the language of Fourier analysis literature, we prove results regarding quadrature rules for the representations appearing in the action of G on X. Technically, our arguments involving representation theory are quite basic, and as such are similar in spirit to several known results in the Fourier analysis literature. In particular, Theorem 3.1 is similar to theorems established in [KORT01, AR11]. However, our proof in arguably simpler, as it applies the Carathéodory theorem instead of a more involved second moment argument. Also, some technical parts used in the proof of Theorem 4.1 are related to known results in the Fourier analysis literature, e.g. in [Mas98, RS09].

Paper organization Theorem 2.1 is proved in Section 2, Theorem 3.1 in Section 3 and Theorem 4.1 in Section 4. We conclude with some open problems in Section 5. For lack of space, we present preliminary definitions in Appendix A, and defer some of the technical claims proofs to Appendix B. Note that throughout the paper we do not attempt to optimize constants.

2 Almost X-uniform distributions are statistically close to X-uniform distributions

We prove in this section Theorem 2.1, which states that almost X-uniform distributions with small enough error are statistically close to X-uniform distributions.

Theorem 2.1. *Let μ be a distribution on G which is almost X-uniform with error ε. Then there exists a distribution μ' on G which is X-uniform, and such that the statistical distance between μ and μ' is at most $\varepsilon \cdot 3|X|^4$.*

We first rephrase the conditions for a distribution to be X-uniform, or almost X uniform, in terms of representations. Let R_X be the representation of the action of G on X, i.e. $R_X(g)_{x,y} = 1_{g(x)=y}$. Let U_G denote the uniform distribution over G.

Claim 2.2. *Let μ be a distribution on G. Then*

1. *μ is X-uniform iff $R_X(\mu) = R_X(U_G)$.*
2. *μ is almost X-uniform with error ε iff $\|R_X(\mu) - R_X(U_G)\|_\infty \leq \varepsilon$.*

Proof. The claim is immediate from the definitions of X-uniform and almost X-uniform distributions, since $R_X(\mu)_{x,y} = \Pr_{g\sim\mu}[g(x) = y]$ and $R_X(U_G)_{x,y} = \Pr_{g\in G}[g(x) = y]$. □

The first step is to decompose R_X into its irreducible representations. Let $R_X \equiv e_0\mathbf{1} + e_1R_1 + \ldots + e_tR_t$, where $R_1, \ldots, R_t$ are unitary nonequivalent non-trivial irreducible representations, and e_i is the multiplicity of R_i in R_X. We next transform the conditions of Claim 2.2 to the basis of the irreducible representations.

Claim 2.3. *Let μ be a distribution on G. Then*

1. *μ is X-uniform iff $R_i(\mu) = 0$ for all $i \in [t]$.*
2. *If μ is almost X-uniform with error ε then $\|R_i(\mu)\|_\infty \leq \varepsilon|X|$ for all $i \in [t]$.*

The proof of Claim 2.3 is deferred to Appendix B. The main idea in the proof of Theorem 2.1 is to "correct" each element of $R_i(\mu)$ to be zero by making a small statistical change in μ, and without affecting the other elements of R_i or in any other $R_{i'}$. This is analogous to the proof idea of [AGM03] for almost k-wise independent bits (see also [AAK$^+$07]). Performing all these local changes sequentially over all elements of R_i, $i \in [t]$, will shift μ into an X-uniform distribution. Actually, as a first step we will get a general element in $\mathbb{C}[G]$, which we then rectify to be a distribution.

Let R_i be one of the irreducible representations, and let $d_i = \dim(R_i)$ be its dimension. For $j, k \in [d_i]$ we define $\Delta_{i,j,k} \in \mathbb{C}[G]$ as

$$\Delta_{i,j,k}(g) = \frac{d_i}{|G|}\overline{R_i(g)_{j,k}}.$$

We consider how shifting μ by a small multiple of $\Delta_{i,j,k}$ affects the entries of $R_1, \ldots, R_t$.

Claim 2.4. *Let $i \in [t], j, k \in [d_i]$ and $i' \in [t], j', k' \in [d_{i'}]$. For any $\delta \in \mathbb{R}$ we have*

$$R_{i'}(\mu + \delta\Delta_{i,j,k})_{j',k'} = R_{i'}(\mu)_{j',k'} + \delta \cdot 1_{(i,j,k)=(i',j',k')}.$$

Proof. First, note that by additivity

$$R_{i'}(\mu + \delta\Delta_{i,j,k})_{j',k'} = R_{i'}(\mu)_{j',k'} + \delta \cdot R_{i'}(\Delta_{i,j,k})_{j',k'}.$$

The claim follows from the orthogonality of the entries of the irreducible representations. By Schur's Lemma,

$$R_{i'}(\Delta_{i,j,k})_{j',k'} = \frac{d_i}{|G|} \sum_{g \in G} R_{i'}(g)_{j',k'} \overline{R_i(g)_{j,k}} = 1_{(i,j,k)=(i',j',k')}.$$

□

We will also need the following claim, which asserts that $\mathbf{1}(\Delta_{i,j,k}) = 0$ and that $\|\Delta_{i,j,k}\|_\infty$ is bounded.

Claim 2.5. *Let $i \in [t], j, k \in [d_i]$. Then*

1. $\mathbf{1}(\Delta_{i,j,k}) = 0$.
2. $\|\Delta_{i,j,k}\|_\infty \leq \frac{|X|}{|G|}$.

Proof. The first item follows because $\sum_{g \in G} R_i(g)_{j,k} = 0$ by Schur's lemma, since R_i is a nontrivial irreducible representation. The second item follows because $d_i \leq |X|$ and because $|R_i(g)_{j,k}| \leq 1$ since $R_i(g)$ is a unitary matrix. □

Applying Claim 2.4 and Claim 2.5 iteratively over all elements of $R_1, \ldots, R_t$, we obtain the following corollary.

Corollary 2.6. *Let μ be a distribution over G which is almost X-uniform with error ε. Define $\Delta \in \mathbb{C}[G]$ by*

$$\Delta(g) = -\sum_{i \in [t]} \sum_{j,k \in [d_i]} R_i(\mu)_{j,k} \cdot \Delta_{i,j,k}(g).$$

Then

1. $R_X(\mu + \Delta) = R_X(U_G)$.
2. $\|\Delta\|_\infty \leq \frac{\varepsilon |X|^4}{|G|}$.

Proof. The first item holds since $R_i(\mu + \Delta)_{j,k} = R_i(U_G)_{j,k}$ for all $i \in [t]$ and $j, k \in d_i$ by Claim 2.4, and since $\mathbf{1}(\mu + \Delta) = \mathbf{1}(U_G) = 1$ by the first item in Claim 2.5. The second item holds since $\sum d_i^2 \leq |X|^2$ as $\dim(R_X) = |X|$, $|R_i(\mu)_{j,k}| \leq \varepsilon |X|$ by Claim 2.3, and $|\Delta_{i,j,k}(g)| \leq |X|/|G|$ by the second item in Claim 2.5. □

We are nearly done. The only problem is that $\mu + \Delta$ may not be a distribution: it may be complex, or have negative values. This can be fixed, without increasing the statistical distance too much. We are now ready to prove Theorem 2.1.

Proof of Theorem 2.1. Let $\lambda = |G| \cdot \|\Delta\|_\infty \leq \varepsilon |X|^4$. Define

$$\mu' = (1 - \lambda) \left(\mu + \frac{\Delta + \overline{\Delta}}{2} \right) + \lambda U_G.$$

We claim that μ' is a distribution which is X-uniform. First let us show that $R_X(\mu') = R_X(U_G)$. We already know by Corollary 2.6 that $R_X(\mu + \Delta) = R_X(U_G)$. Conjugating this equality, since

R_X is a real representation (i.e. all elements in $R_X(g)$ are real), and since $\mu, U_G \in \mathbb{R}[G]$ are also real, we obtain that also

$$R_X(\mu + \overline{\Delta}) = R_X(U_G).$$

Thus $R_X(\mu') = R_X(U_G)$ since $R_X(\mu')$ is a convex combination of $R_X(\mu + \Delta), R_X(\mu + \overline{\Delta})$ and $R_X(U_G)$.

To conclude we need to show that μ' in a distribution, i.e. it is real, nonnegative and sums to one. By definition of μ' it is real, and since $R_X(\mu') = R_X(U_G)$ we have $\sum_{g\in G} \mu'(g) = \mathbf{1}(\mu') = \mathbf{1}(U_G) = 1$. The bound $\mu'(g) \geq 0$ for all $g \in G$ follows by elementary calculations from $\mu(g) \geq 0$, $|\Delta(g)| \leq \lambda/|G|$ and $U_G(g) = 1/|G|$. □

3 Random sets support X-uniform distributions

We establish Theorem 3.1 in this section, which states that w.h.p a random set of size $|X|^{O(1)}$ supports an X-uniform distribution.

Theorem 3.1. *Let $S \subset G$ be a random set of size $O(|X|^6)$. Then with probability 0.99 over the choice of S, there exists a distribution μ supported on S which is X-uniform.*

Recall that a distribution μ is X-uniform if $\Pr_{g\sim\mu}[g(x) = y] = \Pr_{g\in G}[g(x) = y]$ for all $x, y \in X$. We say a set S *supports X-uniformity* if there exists a distribution supported on S which is X-uniform. We first establish that this a purely geometric property of S.

Let R_X be the representation of the action of G on X, that is, $R_X(g)_{x,y} = 1_{g(x)=y}$. Let $U = R_X(U_G) = \mathbb{E}_{g\in G}[R_X(g)]$ denote the matrix which corresponds to the action on X of the uniform distribution over G. We consider these matrices as points in $\mathbb{R}^d$ for $d = |X|^2$.

Claim 3.2. *A set $S \subset G$ supports X-uniformity iff the convex hull of the matrices $\{R_X(g) : g \in S\}$ contains the matrix U.*

Proof. A point in the convex hull is given by $M = \sum_{g\in S} \mu(g) \cdot R_X(g)$ where $\mu(g) \geq 0$ and $\sum_{g\in S} \mu(g) = 1$. Thus, each point in the convex hull corresponds to a distribution μ over S, and vice versa. Note that $M_{x,y} = \Pr_{g\sim\mu}[g(x) = y]$, hence an X-uniform distribution corresponds to the matrix U. □

Let $S \subset G$ be a random set. By Claim 3.2 it is enough to show that the matrix U lies in the convex hull of $\{R_X(g) : g \in S\}$. Suppose this is not the case; then there must exist a hyperplane H in $\mathbb{R}^d$ which passes through U and such that all matrices $\{R_X(g) : g \in S\}$ lie on one side of H. We first show that any hyperplane which passes through U has a noticeable fraction of the matrices $\{R_X(g) : g \in G\}$ on both sides.

Claim 3.3. *Let H be a hyperplane which passes through U. The number of matrices $\{R_X(g) : g \in G\}$ on any side of H is at least $|G|/(|X|^2 + 1)$.*

Proof. Let H^+ denote a halfspace defined by H, and let $G^+ = \{g \in G : R_X(g) \in H^+\}$ denote the set of permutations whose corresponding matrices lie in H^+. The matrix U can be written by Carathéodory theorem as the convex combination of $d+1$ matrices $R_X(g_0), \ldots, R_X(g_d)$. We claim that for any $h \in G$, the matrix U also belongs to the convex hull of $R_X(g_0h), \ldots, R_X(g_dh)$. This follows since $R_X(g_ih) = R_X(g_i)R_X(h)$ and $UR_X(h) = U$. Thus, at least one of $g_0h, \ldots, g_dh$ must lie in G^+, for any choice of $h \in G$. This concludes the proof since for a randomly chosen h,

$$1 = \Pr_{h\in G}[\exists i, g_ih \in G^+] \leq \sum_{i=0}^{d} \Pr_{h\in G}[g_ih \in G^+] = (d+1) \cdot \frac{|G^+|}{|G|}.$$

□

We now establish Theorem 3.1.

Proof of Theorem 3.1. Let $S \subset G$ be a random set of N elements, chosen with repetitions. Let $K \lhd G$ be the normal subgroup of G which acts trivially on X, i.e. $K = \{g \in G : g(x) = x \; \forall x \in X\}$. Observe that the quotient group G/K also acts on X, and that $\{R_X(g) : g \in G\} = \{R_X(g) : g \in G/K\}$. Thus the number of distinct matrices $R_X(g)$ is bounded by $|G/K| \leq |X|!$, and by Fact A.2 the number of ways to partition this set of matrices by any hyperplane, and in particular one which passes through U, is bounded by $(|X|!)^d$. Fix such a partition. The number of matrices $\{R_x(g) : g \in G\}$ which lies on each side of the partition is at least $|G|/(d+1)$ by Claim 3.3. Hence, the probability that S is contained in one side of the partition is bounded by $2(1 - 1/(d+1))^N$. Thus, by the union bound, the probability that there exists a hyperplane passing through U, such that S is contained in one side of it, is at most

$$|G/K|^d \cdot 2\left(1 - \frac{1}{d+1}\right)^N \leq 2\exp(-N/(d+1) + d\log(|X|!)),$$

which is at most 0.01 for $N = O(d^2 \log(|X|!)) \leq O(|X|^6)$. □

4 Almost X-uniform distributions support X-uniform distributions

We prove in this section Theorem 4.1, which states that if μ is an almost $X \times X$-uniform distribution with small enough error, then there exists an X-uniform distribution μ' supported on the support of μ.

Theorem 4.1. *Let μ be a distribution supported on a set $S \subset G$ which is almost $(X \times X)$-uniform with error $\varepsilon < 0.5|X|^{-7}$. Then there exists a distribution μ' supported on S which is X-uniform.*

Fix such a distribution μ, and let S denote its support, $S = \{g : \mu(g) > 0\}$. Let R_X be the representation of G acting on X. By Claim 3.2, S supports an X-uniform distribution iff $R_X(U_G) = \mathbb{E}_{g \in G}[R_X(g)]$ lies in the convex hull of $\{R_X(g) : g \in S\}$. Assume this is not the case; then there exists an hyperplane H which passes through $R_X(U_G)$ and such that all $\{R_X(g) : g \in S\}$ lie on one side of H.

We first project H into an hyperplane with a simpler representation. Let $R_X \equiv e_0\mathbf{1} + e_1R_1 + \ldots + e_tR_t$ denote the decomposition of R_X into unitary nonequivalent irreducible representation, and let $d_i = \dim(R_i)$ denote the dimension of each irreducible representation. Essentially, we will project H to "use" only one copy from each nontrivial irreducible representation. That is, we will show that H can be projected to a hyperplane separating 0 from $\{R_1(g) \times \ldots \times R_t(g) : g \in S\}$.

Claim 4.2. *There exists a map $L : G \to \mathbb{R}$ given by*

$$L(g) := \sum_{i \in [t]} \sum_{j,k \in [d_i]} \lambda_{i,j,k} \cdot R_i(g)_{j,k}$$

for some coefficients $\{\lambda_{i,j,k} \in \mathbb{C} : i \in [t], j, k \in [d_i]\}$ such that

1. *$\mathbb{E}_{g \in G}[L(g)] = 0$.*
2. *For all $g \in S$, $L(g) > 0$.*

The proof of Claim 4.2 is deferred to Appendix B. We may assume w.l.o.g that $\mathbb{E}_{g\in G}[L^2(g)] = 1$ by multiplying all coefficients $\lambda_{i,j,k}$ by an appropriate factor. The main idea is to show that if μ is almost $X \times X$ uniform, then $\mathbb{E}_{g\sim\mu}[L^2(g)] \approx \mathbb{E}_{g\in G}[L^2(g)] = 1$ while $\mathbb{E}_{g\sim\mu}[L(g)] \approx \mathbb{E}_{g\in G}[L(g)] = 0$. Combining this with a bound on $\|L\|_\infty$ a simple calculation shows that it cannot be the case that $L(g) > 0$ for all g in the support of μ.

The first step is to show that the coefficients $\lambda_{i,j,k}$ cannot be very large.

Claim 4.3. *We have*

$$\sum_{i\in[t]} \sum_{j,k\in[d_i]} \frac{|\lambda_{i,j,k}|^2}{d_i} = 1.$$

In particular, $|\lambda_{i,j,k}| \le |X|^{1/2}$ *for all* i, j, k.

The proof of Claim 4.3 is deferred to Appendix B. An immediate corollary is that $L(g)$ can never be very large.

Corollary 4.4. $|L(g)| \le |X|^{2.5}$ *for all* $g \in G$.

Proof. We have $|R_i(g)_{j,k}| \le 1$ since R_i is unitary, hence $|L(g)| \le \sum_{i\in[t]} \sum_{j,k\in[d_i]} |\lambda_{i,j,k}| \le |X|^{2.5}$ since $\sum d_i^2 \le |X|^2$. □

The bound on $|\lambda_{i,j,k}|$ together with the assumption that μ is almost $X \times X$-uniform, implies that the first and second moment of L are approximately the same under μ and under the uniform distribution over G.

Claim 4.5. *Let* μ *be a distribution which is almost* $X \times X$*-uniform with error* ε*. Then*

1. $|\mathbb{E}_{g\sim\mu}[L(g)]| \le \varepsilon|X|^{4.5}$.
2. $|\mathbb{E}_{g\sim\mu}[L^2(g)] - 1| \le \varepsilon|X|^7$.

The proof of Claim 4.5 is deferred to Appendix B. We are now ready to prove Theorem 4.1.

Proof of Theorem 4.1. Let μ be almost $X \times X$ uniform with error $\varepsilon \le 0.5|X|^{-7}$. Summarizing Corollary 4.4 and Claim 4.5, we have

1. $\|L\|_\infty \le |X|^{2.5}$.
2. $E_{g\sim\mu}[L(g)] \le \varepsilon|X|^{4.5}$.
3. $E_{g\sim\mu}[L(g)^2] \ge 1 - \varepsilon|X|^7$.

However, since we assumed by contradiction that $L(g) > 0$ for all g in the support of μ, we have

$$\mathbb{E}_{g\sim\mu}[L(g)^2] \le \|L(g)\|_\infty \cdot \mathbb{E}_{g\sim\mu}[L(g)] \le |X|^{2.5} \cdot \varepsilon|X|^{4.5},$$

i.e. we have

$$1 - \varepsilon|X|^7 \le \varepsilon|X|^7,$$

which is false whenever $\varepsilon < 0.5|X|^{-7}$. □

5 Summary and open problems

We showed that almost X-uniform (or $X \times X$-uniform) distributions are close to perfect X uniform distributions in two ways: they are statistically close to some X-uniform distribution μ', and they support a X-uniform distribution μ''. It may be possible that both can be realized by the same X-uniform distribution, i.e. that $\mu' = \mu''$. We leave this as an open problem.

Another interesting combinatorial problem is to construct small sets which are perfectly uniform. This is unknown even in the special case of k-wise independent permutations, not to mention the general setting of group actions. Recently, Greg Kuperberg, Ron Peled and the second author gave a non-explicit proof for the existence of small families of k-wise independent permutations.

Acknowledgements We thank Avi Wigderson for helpful discussions and reference to the work of Karp and Papadimitriou [KP82].

References

[AA07] N. Ailon and N. Alon. Hardness of fully dense problems. *Inform. and Comput.*, 205(8):1117–1129, 2007.

[AAK+07] N. Alon, A. Andoni, T. Kaufman, K. Matulef, R. Rubinfeld, and N. Xie. Testing k-wise and almost k-wise independence. In *STOC 07*, pages 496–505. ACM, New York, 2007.

[ABI86] N. Alon, L. Babai, and A. Itai. A fast and simple randomized algorithm for the maximal independent set problem. *Journal of algorithms*, 7:567–583, 1986.

[AGM03] N. Alon, O. Goldreich, and Y. Mansour. Almost k-wise independence versus k-wise independence. *Inf. Process. Lett.*, 88:107–110, November 2003.

[AH11] Per Austrin and Johan Håstad. Randomly supported independence and resistance. *SIAM J. Comput.*, 40(1):1–27, 2011.

[AR11] Gorjan Alagic and Alexander Russell. Spectral Concentration of Positive Functions on Compact Groups. *Journal of Fourier Analysis and Applications*, pages 1–19, February 2011.

[Cam95] P. J. Cameron. Permutation groups. In *Handbook of combinatorics, Vol. 1, 2*, pages 611–645. Elsevier, Amsterdam, 1995.

[FH91] W. Fulton and J. Harris. *Representation theory: a first course*, volume 129 of *Graduate texts in Mathematics*. Springer, 1 edition, 1991.

[Har67] E. F. Harding. The number of partitions of a set of N points in k dimensions induced by hyperplanes. *Proc. Edinburgh Math. Soc. (2)*, 15:285–289, 1966/1967.

[Kas07] M. Kassabov. Symmetric groups and expanders. *Inventiones Mathematicae*, 170(2):327–354, 2007.

[KM94] D. Koller and N. Megiddo. Constructing small sample spaces satisfying given constraints. *SIAM Journal on Discrete Mathematics*, 7:260274, 1994.

[KNR05] E. Kaplan, M. Naor, and O. Reingold. Derandomized constructions of k-wise (almost) independent permutations. In C. Chekuri, K. Jansen, J. D. P. Rolim, and L. Trevisan, editors, *Approximation, Randomization and Combinatorial Optimization*, volume 3624 of *Lecture Notes in Computer Science*, pages 354–365. Springer Berlin / Heidelberg, 2005.

[KORT01] K. Kueh, T. Olson, D. Rockmore, and K. Tan. Nonlinear approximation theory on compact groups. *Journal of Fourier Analysis and Applications*, 7:257–281, 2001.

[KP82] R.M. Karp and C.H. Papadimitriou. On linear characterizations of combinatorial optimization problems. *SIAM Journal on Computing*, 11(4):620–632, 1982.

[Mas98] David Maslen. Efficient computation of fourier transforms on compact groups. *Journal of Fourier Analysis and Applications*, 4:19–52, 1998.

[RS09] Aidan Roy and A. J. Scott. Unitary designs and codes. *Des. Codes Cryptography*, 53:13–31, October 2009.

[RW06] A. Russell and Hong Wang. How to fool an unbounded adversary with a short key. *IEEE Transactions on Information Theory*, 52(3):1130 –1140, march 2006.

[RX10] Ronitt Rubinfeld and Ning Xie. Testing non-uniform -wise independent distributions over product spaces. In *ICALP (1)*, pages 565–581, 2010.

[Sau72] N. Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13(1):145 – 147, 1972.

[She72] S. Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41:247–261, 1972.

[Vau98] Serge Vaudenay. Provable security for block ciphers by decorrelation. In Michel Morvan, Christoph Meinel, and Daniel Krob, editors, *STACS 98*, volume 1373 of *Lecture Notes in Computer Science*, pages 249–275. Springer Berlin / Heidelberg, 1998.

[Vau00] Serge Vaudenay. Adaptive-attack norm for decorrelation and super-pseudorandomness. In *Proceedings of the 6th Annual International Workshop on Selected Areas in Cryptography*, SAC '99, pages 49–61, London, UK, 2000. Springer-Verlag.

[Vau03] Serge Vaudenay. Decorrelation: a theory for block cipher security. *Journal of Cryptology*, 16(4):249–286, 2003.

[VC71] V. N. Vapnik and A. Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory Probab. Appl.*, 16:264–280, 1971.

A Preliminaries

Group action and uniformity A group G acts on a set X if there is a homomorphism from G to the permutation group on X. That is, each $g \in G$ is a permutation on X, and $(gh)(x) = g(h(x))$ for all $g, h \in G, x \in X$. We denote by U_G the uniform distribution over G. We recall some definitions from the introduction: a distribution μ over G is X-uniform if

$$\Pr_{g\sim\mu}[g(x) = y] = \Pr_{g\in G}[g(x) = y]$$

for all $x, y \in X$; and a distribution μ is almost X-uniform with error ε if

$$\left| \Pr_{g\sim\mu}[g(x) = y] - \Pr_{g\in G}[g(x) = y] \right| \leq \varepsilon$$

for all $x, y \in X$. A set $S \subset G$ is X-uniform (or almost X-uniform) if the uniform distribution over S is such. We will need some basic facts in linear algebra, geometry and representation theory, which are presented below.

Linear algebra Let $A = (a_{i,j})$ be a complex matrix. Recall some basic definitions. The L_∞ norm of A is $\|A\|_\infty = \max |a_{i,j}|$. The Frobenius norm of A is $\|A\|_{\mathrm{Fr}} = \sqrt{\sum |a_{i,j}|^2}$. Note that always $\|A\|_\infty \leq \|A\|_{\mathrm{Fr}}$. A matrix A is unitary if $A\overline{A^t} = I$. Note that the Frobenius norm of a matrix is invariant under any unitary basis change. That is, if U is unitary then $\|U^{-1}AU\|_{\mathrm{Fr}} = \|A\|_{\mathrm{Fr}}$. The tensor product of a $d_1 \times d_1$ matrix A_1 and a $d_2 \times d_2$ matrix A_2, denoted $A_1 \otimes A_2$, is a $(d_1 d_2) \times (d_1 d_2)$ matrix, whose entries are given by $(A_1 \otimes A_2)_{(i,i'),(j,j')} = (A_1)_{i,j}(A_2)_{i',j'}$. Note that the tensor product of unitary matrices is also unitary.

Geometry Let $X = \{x_1, \ldots, x_N\}$ be a set of points in $\mathbb{R}^d$. The *convex hull* of X is the set of points contained in the minimal convex set containing X; equivalently, it is the set of all points $\{\sum \lambda_i x_i : \lambda_1, \ldots, \lambda_N \geq 0, \sum \lambda_i = 1\}$.

Fact A.1 (Carathéodory theorem)**.** *Let X be a finite set of points in $\mathbb{R}^d$, and let y be a point in the convex hull of X. Then there exists a subset $Y \subset X$ of size $|Y| \leq d + 1$ such that y is in the convex hull of Y.*

Any hyperplane H partitions a set X of points into two sets: if $H = \{x : \langle a, x\rangle = b\}$ then the sets are $\{x \in X : \langle a, x\rangle \geq b\}$ and $\{x \in X : \langle a, x\rangle < b\}$. We need the following bound on the maximal number of ways a set can be partitioned by hyperplanes [3].

Fact A.2 (Harding [Har67])**.** *Let X be a set of N points in $\mathbb{R}^d$. The number of different ways to partition X into two sets by a hyperplane is at most $\sum_{i=0}^{d} \binom{N-1}{i} \leq N^d$.*

Representation theory Let G be a finite group. A representation of G (over $\mathbb{C}$) is a homomorphism $R : G \to \mathrm{GL}(d, \mathbb{C})$. That is, $R(g)$ for $g \in G$ is a $d \times d$ nonsingular complex matrix, and $R(gh) = R(g)R(h)$ for every $g, h \in G$. The dimension of the representation R is d. Two representations R, R' of G of dimension d are *equivalent* if there exists an invertible matrix A such that $R'(g) = A^{-1}R(g)A$ for all $g \in G$. This is denoted as $R \equiv R'$.

A representation R is *unitary* if $R(g)$ is unitary for all $g \in G$.

Fact A.3. *Any representation of G is equivalent to a unitary representation.*

We will restrict our attention only to unitary representations in this paper. We note that if R, R' are unitary and equivalent, then there exists a unitary matrix A such that $R'(g) = A^{-1}R(g)A$.

Let G be a group which acts on a set X, that is, $g : X \to X$ is a permutation of X for all $g \in G$, and $(gh)(x) = g(h(x))$ for all $g, h \in G$ and $x \in X$. The associated representation R_X maps each $g \in G$ to the permutation matrix it induces on the set X. That is, $R_X(g)$ is an $|X| \times |X|$ matrix, indexed by $x, y \in X$, defined as $(R_X(g))_{x,y} = 1$ if $g(x) = y$ and $(R_X(g))_{x,y} = 0$ otherwise. Note that R_X is always a unitary representation.

[3] A quick way to prove a slightly weaker estimate is as follows: the VC-dimension [VC71] of halfspaces is $d + 1$. Hence by the VC-dimension theorem [VC71, Sau72, She72] the number of partitions is at most $\sum_{i=0}^{d+1} \binom{N}{i} \leq N^{d+1}$.

The sum of two representations $R_1 : G \to \mathrm{GL}(d_1, \mathbb{C})$ and $R_2 : G \to \mathrm{GL}(d_2, \mathbb{C})$ is a representation $R : G \to \mathrm{GL}(d_1 + d_2, \mathbb{C})$ where $R(g)$ is defined as a block diagonal matrix with two blocks, given by $R_1(g)$ and $R_2(g)$. For $e \geq 1$ let $eR := R + \ldots + R$ where the sum is over e copies of R. A representation R is *reducible* if it is equivalent to the sum of two representations. Otherwise, the representation R is *irreducible*. We summarize a few basic properties of representations below. For details we refer the reader to any standard book on representation theory, e.g. [FH91].

Fact A.4 (Maschke's theorem)**.** *Any representation R of G is equivalent to a sum of irreducible representations $R \equiv e_1 R_1 + \ldots + e_t R_t$, where $R_1, \ldots, R_t$ are nonequivalent irreducible representations, and $e_i \geq 1$ is the multiplicity of R_i. We have $\sum e_i \dim(R_i) = \dim(R)$.*

Fact A.5 (Schur's lemma)**.** *Let R be an unitary irreducible representation of G of dimension d. Then for any $i, j, k, \ell \in [d]$,*

$$\frac{1}{|G|} \sum_{g \in G} R(g)_{i,j} \overline{R(g)_{k,\ell}} = \frac{1}{d} \delta_{i,k} \delta_{j,\ell}.$$

Let R', R'' be two non-equivalent unitary irreducible representations of G of dimensions d', d''. Then for any $i, j \in [d']$ and $k, \ell \in [d'']$,

$$\frac{1}{|G|} \sum_{g \in G} R'(g)_{i,j} \overline{R''(g)_{k,\ell}} = 0.$$

The trivial representation is given by $\mathbf{1}(g) = 1$ for all $g \in G$. An immediate corollary of Schur's lemma is that for every representation R which is irreducible and nontrivial, we have $\sum_{g \in G} R(g) = 0$.

The group algebra $\mathbb{C}[G]$ is the linear space of functions $\mu : G \to \mathbb{C}$. It is often written as $\mu = \sum_{g \in G} \mu(g) \cdot g$. Note that the distributions over G form a subset of $\mathbb{C}[G]$. For $\mu \in \mathbb{C}[G]$ and a representation R of G, let $R(\mu) := \sum_{g \in G} \mu(g) R(g)$. If μ is a distribution, this is equivalent to $R(\mu) = \mathbb{E}_{g \sim \mu}[R(g)]$.

B Deferred proofs

Proof of Claim 2.3. Note that as μ is a distribution, then $\mathbf{1}(\mu) = \sum_{g \in G} \mu(g) = 1$, and the same holds for U_G. Hence always $\mathbf{1}(\mu) = \mathbf{1}(U_G)$. Thus, $R_X(\mu) = R_X(U_G)$ iff $R_i(\mu) = R_i(U_G)$ for all $i \in [t]$. The first item follows since $R_i(U_G) = 0$ for all $i \in [t]$. To see that, note that by Schur's lemma

$$R_i(U_G)_{j,k} = \frac{1}{|G|} \sum_{g \in G} R_i(g)_{j,k} = \frac{1}{|G|} \sum_{g \in G} R_i(g)_{j,k} \overline{\mathbf{1}(g)} = 0$$

since R_i and $\mathbf{1}$ are nonequivalent unitary irreducible representations. To prove the second item, let μ be an almost X-uniform distribution with error ε. By Claim 2.2 this is equivalent to $\|R_X(\mu) - R_X(U_G)\|_\infty \leq \varepsilon$. The L_∞ norm is not convenient for the basis change required to switch to the basis of the irreducible representations. We thus switch to the Frobenius norm, which is trivially bounded by

$$\|R_X(\mu) - R_X(U_G)\|_{\mathrm{Fr}} \leq \varepsilon |X|.$$

Note that the Frobenius norm is invariant under unitary change of basis, and as both R_X and $R_1, \ldots, R_t$ are unitary, the basis change can also be assumed to be unitary. We thus have

$$\sqrt{\sum_{i=1}^{t} e_i \|R_i(\mu) - R_i(U_G)\|_{\mathrm{Fr}}^2} = \|R_X(\mu) - R_X(U_G)\|_{\mathrm{Fr}} \leq \varepsilon |X|,$$

which combined with the fact that $R_i(U_G) = 0$ implies that $\|R_i(\mu)\|_\infty \leq \varepsilon|X|$. □

Proof of Claim 4.2. The hyperplane H separating $R_X(U_G)$ from $\{R_X(g) : g \in S\}$ implies there exists a map $L' : G \to \mathbb{R}$ defined as $L'(g) = \sum_{x,y\in X} \alpha_{x,y} R_X(g)_{x,y}$ for some real coefficients $\{\alpha_{x,y} \in \mathbb{R} : x, y \in X\}$ such that

$$L'(g) > \mathbb{E}_{h\in G}[L'(h)]$$

for all $g \in S$. Applying the linear transformation mapping R_X into the basis of irreducible representations, we get that $L'(g)$ can be expressed as

$$L'(g) = \sum_{\ell\in[e_0]} \beta_{0,\ell}\mathbf{1}(g) + \sum_{i\in[t]} \sum_{j,k\in[d_i]} \sum_{\ell\in[e_i]} \beta_{i,j,k,\ell} R_i(g)_{j,k},$$

where $\beta_{0,\ell}, \beta_{i,j,k,\ell} \in \mathbb{C}$ are obtained by a linear transformation (over $\mathbb{C}$) of $\alpha_{x,y}$. Observe that $\mathbb{E}_{g\in G}[L'(g)] = \sum_{\ell\in[e_0]} \beta_{0,\ell}$ by Schur's lemma, and define $L(g) := L'(g) - \mathbb{E}[L'(g)]$. Note that $L : G \to \mathbb{R}$ is real since $L' : G \to \mathbb{R}$ was real, that $\mathbb{E}[L] = 0$ and that $L(g) > 0$ for all $g \in S$. The coefficient $\lambda_{i,j,k}$ is given by the sum of all $\beta_{i,j,k,\ell}$ over $\ell \in [e_i]$. □

Proof of Claim 4.3. We assumed $\mathbb{E}[L^2] = 1$, which, since L is real, implies $\mathbb{E}[|L|^2] = 1$. Using Schur's lemma we get

$$\begin{aligned}
1 &= \mathbb{E}_{g\in G}[L(g) \cdot \overline{L(g)}] \\
&= \sum_{i,i'\in[t]} \sum_{j,k\in[d_i]} \sum_{j',k'\in[d_{i'}]} \lambda_{i,j,k}\overline{\lambda_{i',j',k'}} \mathbb{E}_{g\in G}[R_i(g)_{j,k}\overline{R_{i'}(g)}_{j',k'}] \\
&= \sum_{i\in[t]} \sum_{j,k\in[d_i]} \frac{|\lambda_{i,j,k}|^2}{d_i},
\end{aligned}$$

and in particular $|\lambda_{i,j,k}|^2 \leq d_i \leq |X|$. □

Proof of Claim 4.5. We have

$$|\mathbb{E}_{g\sim\mu}[L(g)]| \leq \sum_{i\in[t]} \sum_{j,k\in[d_i]} |\lambda_{i,j,k}||\mathbb{E}_{g\sim\mu}[R_i(g)_{j,k}]|.$$

The bound on the first moment follows since $\sum d_i^2 \leq |X|^2$; since $|\lambda_{i,j,k}| \leq |X|^{1/2}$ by Claim 4.3; and since μ is in particular X-uniform with error $\varepsilon|X|$, we have by Claim 2.3 that $|\mathbb{E}_{g\sim\mu}[R_i(g)_{j,k}]| \leq \varepsilon|X|^2$. The bound on the second moment is proved in a similar way. We have

$$\begin{aligned}
\mathbb{E}_{g\sim\mu}[|L(g)|^2] = &\sum_{i,j,k} |\lambda_{i,j,k}|^2 \cdot \mathbb{E}_{g\sim\mu}[|R_i(g)_{j,k}|^2] \\
&+ \sum_{(i,j,k)\neq(i',j',k')} \lambda_{i,j,k}\overline{\lambda_{i',j',k'}} \cdot \mathbb{E}_{g\sim\mu}[R_i(g)_{j,k}\overline{R_{i'}(g)_{j',k'}}].
\end{aligned}$$

To conclude the proof we need to show that $\mathbb{E}_{g\sim\mu}[|R_i(g)_{j,k}|^2] \approx 1/d_i$ and that $\mathbb{E}_{g\sim\mu}[R_i(g)_{j,k}R_{i'}(g)_{j',k'}] \approx 0$, and combine this with the identity $\sum |\lambda_{i,j,k}|^2/d_i = 1$ which we showed in Claim 4.3.

The condition that μ is almost $X \times X$-uniform with error ε is equivalent to

$$\|R_{X\times X}(\mu) - R_{X\times X}(U_G)\|_\infty \leq \varepsilon.$$

Switching to the Frobenius norm, this implies

$$\|R_{X\times X}(\mu) - R_{X\times X}(U_G)\|_{\mathrm{Fr}} \leq \varepsilon |X|^2.$$

We now decompose $R_{X\times X}$ to simpler representations, coming from the irreducible representations of R_X. We have that $R_{X\times X} = R_X \otimes R_X$, which since R_X is real, also gives $R_{X\times X} = R_X \otimes \overline{R_X}$. Now, if $R_X \equiv e_0\mathbf{1} + \sum_{i=1}^{t} e_i R_i$ is the decomposition of R_X into irreducible unitary nonequivalent representations, we have

$$R_{X\times X} \equiv e_0^2\mathbf{1} + \sum_{i=1}^{t} e_i(R_i + \overline{R_i}) + \sum_{i,i'=1}^{t} e_i e_{i'}(R_i \otimes \overline{R_{i'}}).$$

Note that this is not the decomposition of $R_{X\times X}$ into irreducible representations, since $R_i \otimes \overline{R_{i'}}$ is reducible! Nevertheless, we observe that as the basis change for R_X was unitary, so is the basis change for $R_{X\times X}$ (since the tensor product of two unitary matrices is again unitary). In particular, we get that $\|(R_i \otimes \overline{R_{i'}})(\mu) - (R_i \otimes \overline{R_{i'}})(U_G)\|_{\mathrm{Fr}} \leq \varepsilon|X|^2$, which implies that

$$\|(R_i \otimes \overline{R_{i'}})(\mu) - (R_i \otimes \overline{R_{i'}})(U_G)\|_{\infty} \leq \varepsilon|X|^2.$$

The matrix $R_i \otimes \overline{R_{i'}}$ is indexed by $((j,j'),(k,k'))$, where $(R_i \otimes \overline{R_{i'}})(g)_{(j,j'),(k,k')} = R_i(g)_{j,k}\overline{R_{i'}(g)_{j',k'}}$. We thus get that for any i,j,k,i',j',k' we have

$$\left|\mathbb{E}_{g\sim\mu}[R_i(g)_{j,k}\overline{R_{i'}(g)_{j',k'}}] - \mathbb{E}_{g\in G}[R_i(g)_{j,k}\overline{R_{i'}(g)_{j',k'}}]\right| \leq \varepsilon|X|^2.$$

To conclude, note that by Schur's lemma,

$$\mathbb{E}_{g\in G}[R_i(g)_{j,k}\overline{R_{i'}(g)_{j',k'}}] = \frac{1}{d_i}1_{(i,j,k)=(i',j',k')}.$$

The bound for the second moment now follows by elementary calculations analog to the ones for the first moment. □