Electronic Colloquium on Computational Complexity, Report No. 88 (2023)

A High Dimensional Goldreich-Levin Theorem

Parker Newton* Silas Richelson† Chase Wilson‡

Abstract

In this work we prove a high dimensional analogue of the beloved Goldreich-Levin theorem (STOC 1989). We consider the following algorithmic problem: given oracle access to a function $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ such that $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big] \geq \varepsilon$ for some $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ and $\varepsilon > 0$, recover $\mathbf{A}$ (or a list of all such matrices). We focus on the case $\varepsilon \leq 1/q$ since when $\varepsilon \geq 1/q + \delta$, the problem is solved by the original Goldreich-Levin theorem. As stated, this problem cannot be efficiently solved, since when $\varepsilon \leq 1/q$ the list of $\mathbf{A}$ with good agreement with f might be exponentially large. Our main theorem gives an algorithm which efficiently recovers a list of linear maps of size $\mathcal{O}\big(1/\varepsilon\big)$ which have good agreement with f, and such that every linear map which has good agreement with f, also has good agreement with some map in our list. Our proof makes novel use of Fourier analysis.

1 Introduction

The celebrated Goldreich-Levin Theorem [GL89] is a cornerstone theorem in theoretical computer science. It yielded fundamental applications in cryptography [Blu83, HILL99], led to the development of new categories of error-correcting codes [Sud97, KT00], and was an early success in boolean learning theory [KM93] (to name but a few of the uses of this terrific theorem). The technical core of the Goldreich-Levin theorem is the "prediction implies inversion" lemma which states that a function $f : \{0,1\}^n \to \{0,1\}$ which predicts random inner products with a secret $\mathbf{y} \in \{0,1\}^n$, with *any advantage at all over guessing randomly*, must "know" $\mathbf{y}$, in the sense that $\mathbf{y}$ can be recovered efficiently given oracle access to f. This lemma has been generalized in many different ways. One line of follow up work proves prediction implies inversion lemmas for general group homomorphisms $f : G \to H$ [GKS06, DGKS08, BBW18]. Another work proves a degree 2 analogue using quadratic Fourier analysis [TW14]. With this work, we add to the Goldreich-Levin fandom by generalizing the original theorem to the case when f has high dimensional output.

1.1 Statement of the Problem and our Main Theorem

We consider a function $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ for integers $n, m, q \in \mathbb{N}$ with q prime which has the following linear agreement guarantee:

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big] \geq \varepsilon, \tag{1}$$

*University of California, Riverside. Email: pnewt001@ucr.edu. Work completed prior to joining Amazon.
†University of California, Riverside. Email: silas@cs.ucr.edu
‡University of California, San Diego. Email: c7wilson@ucsd.edu

ISSN 1433-8092

for some $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ and $\varepsilon > 0$. We ask whether, given oracle access to such a function, it is possible to efficiently recover $\mathbf{A}$. More precisely, and in the list decoding spirit of [GL89], we ask whether it is possible to efficiently output a short list of matrices $\mathsf{L} = \{\mathbf{A}_1, \ldots, \mathbf{A}_\ell\}$ such that any $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ which has good agreement with f is in L. When $\varepsilon \geq 1/q + \delta$, list decoding algorithms from prior work [GL89, DGKS08], indeed recover such a list. However, these algorithms fail when $\varepsilon \leq 1/q$. Actually, when $\varepsilon \leq 1/q$, the problem is not possible as stated, since the list might be exponentially large (and so cannot be efficiently recovered). For example, suppose that f always outputs $\mathbf{Ax}$ except for the first coordinate, which it chooses randomly. In other words,

$$f(\mathbf{x})_i = \begin{cases} \$ \sim \mathbb{Z}_q, & i = 1 \\ (\mathbf{Ax})_i, & i \geq 2 \end{cases}$$

where $(\mathbf{Ax})_i$ denotes the $i-$th coordinate of $\mathbf{Ax} \in \mathbb{Z}_q^n$. Clearly, in this case $f(\mathbf{x}) = \mathbf{Ax}$ occurs with probability $1/q$. However, $f(\mathbf{x}) = \mathbf{A}'\mathbf{x}$ also holds with probability $1/q$ for any $\mathbf{A}'$ whose final $n-1$ rows are the same as those in $\mathbf{A}$. Indeed, the function f possesses no information about the first row of $\mathbf{A}$, so any matrix which equals $\mathbf{A}$ outside of the first row (there are q^n such matrices) will have just as good agreement with f as $\mathbf{A}$ does.

So to summarize, the algorithmic question above is solved by prior work when $\varepsilon \geq 1/q + \delta$ and is impossible when $\varepsilon \leq 1/q$. This is unfortunate because the setup is very natural when $\varepsilon \leq 1/q$. When $n = 1$, the barrier of $1/q + \delta$ makes sense conceptually since a random function (from which no secret can be extracted) will have agreement $1/q$. So the original (one-dimensional) Goldreich-Levin theorem promises that a secret can be extracted from any function which has a prediction advantage over guessing randomly. In higher dimensions, a random function will agree with a linear function with probability $q^{-n} \ll 1/q$ and so one might hope that some information about $\mathbf{A}$ would be recoverable from a function with agreement probability $\varepsilon > q^{-n}$.

The problem is that we have asked the wrong question. Rather than aiming to recover every matrix with good agreement with f, we should try to recover a short list of matrices such that every $\mathbf{A}$ which has good agreement with f has good agreement with some matrix in the list. Indeed, in the previous example, all of the matrices which agree with $\mathbf{A}$ outside of the first row (a q^n-sized family) agree with f for the same reason (because they agree with $\mathbf{A}$ on the final $n-1$ rows). Thus, recovering multiple matrices from this family is repetitive and should not be our goal. Once the question has been modified, we are able to answer it positively.

Theorem 1. *Let $m, n, q \in \mathbb{N}$ be integers with q prime, m sufficiently large, let $\varepsilon > 0$ be a parameter with $\varepsilon > 12 \cdot \max\{q^{-m/9}, q^{-n/3}\}$, and let $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ be a function. There exists a randomized oracle algorithm $\mathcal{A}$ which has the following syntax, runtime and correctness guarantees.*

- **Syntax:** *$\mathcal{A}$ takes no input, gets oracle access to f, and outputs $\mathsf{L} \subset \mathbb{Z}_q^{n \times m}$ of size $|\mathsf{L}| = \mathcal{O}\big(1/\varepsilon\big)$.*
- **Runtime:** *$\mathcal{A}$ runs in expected time $\mathsf{poly}\big(n, \log q, m^{\log_q(1/\varepsilon)}, \varepsilon^{-\log_q(1/\varepsilon)}\big)$.*
- **Correctness:** *With probability at least $1 - 2^{-\Omega(m)}$ over the randomness of $\mathsf{L} \sim \mathcal{A}^{(\varepsilon, f)}$, the following both hold:*
  - *for all $\mathbf{A} \in \mathsf{L}$, $\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{Ax}\big] = \Omega(\varepsilon)$;*
  - *for all $\mathbf{A}' \in \mathbb{Z}_q^{n \times m}$, either*
    - *$\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}'\mathbf{x}\big] < \varepsilon$, or*

– $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[\mathbf{A}'\mathbf{x} = \mathbf{A}\mathbf{x}\right] = \Omega(\varepsilon^2)$ *for some* $\mathbf{A} \in \mathsf{L}$.

All uses of $\mathcal{O}(\cdot)$ *and* $\Omega(\cdot)$ *above hide absolute constants which are independent of all other parameters.*

Remarks. We make two brief remarks about the parameters of our result.

1. The running time of $\mathcal{A}$ is exponential in $\log_q(1/\varepsilon)$, and so our result promises a polynomial time matrix recovery algorithm only in parameter regimes where this quantity is constant. While $\log_q(1/\varepsilon)$ is constant in some parameter regimes of interest (*e.g.*, $q = \mathsf{poly}(m)$, $\varepsilon = \mathsf{poly}(1/m)$), there are others where it is super-constant (*e.g.*, $q = 2$ and $\varepsilon = 1/m$). Removing the exponential dependence on $\log_q(1/\varepsilon)$ is a nice open question.

2. The requirement that $\varepsilon \geq 12 \cdot \max\{q^{-m/9}, q^{-n/3}\}$ is due to our use of the Chernoff-Hoeffding inequality in a few places. In the body of the paper (Section 4.1), we design another algorithm which does not require $\varepsilon > 12q^{-n/3}$, instead requiring just that $\varepsilon \geq q^{-n} + \delta$ for $\delta > 0$, but paying with a running time of $\mathsf{poly}(\delta^{-n})$. This alternative algorithm is best when n is small, which incidentally is the case when the requirement that $\varepsilon \geq 12q^{-n/3}$ is most intrusive.

1.2 Related Work

Approximate List Decodable Codes. Theorem 1 says that it is possible to efficiently recover from f a short list of matrices such that any $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ with good agreement with f also has good agreement with some matrix in the list. This type of relaxed list decoding guarantee is called *approximate list decoding*. Approximate list decodable codes are used in coding theory to build list decodable codes via code concatenation [Tre03, DHK+19]. They were also used in earlier work [IJK09] to prove hardness amplification theorems.

Effective Property Testing. The BLR test [BLR93] says that if $f : \{0,1\}^n \to \{0,1\}$ is such that $f(\mathbf{x}+\mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y})$ holds with good probability over $\mathbf{x}, \mathbf{y} \sim \{0,1\}^n$ then there exists a linear map $\varphi : \{0,1\}^n \to \{0,1\}$ such that f has good agreement with φ. In the high dimensional and large modulus setting, such linearity tests are much harder to prove. Samorodnitsky [Sam07] showed using methods from additive combinatorics that if $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ passes this linearity test with probability ε then it is $\varepsilon'-$close to a linear function where ε' depends exponentially on ε and n. A breakthrough result of Sanders [San12] obtains a better (quasipolynomial) relationship between ε' and ε. The holy grail of this area would be a proof that ε' depends polynomially on ε. This is known to follow from the polynomial Frieman-Ruzsa conjecture in additive combinatorics (and is, in fact, equivalent to a special case of PRF). Our work makes any high dimensional linearity testing theorem effective by offering an algorithm which would recover the linear map which is close to f (whose existence would be ensured by the linearity testing theorem).

Other Relevant Prior Work. A recent work of Asadi, Golovnev, Gur and Shinkar [AGGS22] proves effective property testing theorems en route to giving worst-case to average-case reductions for matrix multiplication and related problems. The effective property testing theorems are proved by converting Sanders' result [San12] into an efficient algorithm. This result is general and is used to give worst-case to average-case reductions for several problems simultaneously. All of their results inherit the quasi-polynomial dependence on the test-passing probability from [San12]. Our result is less general as it

focuses only on the problem of matrix recovery, however we obtain better parameters as we do not inherit the quasi-polynomial dependence of Sanders' theorem.

1.3 Technical Overview

Suppose $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ and $\varepsilon > 0$ are such that $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big] \geq \varepsilon$ holds for some $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$. We give a high level overview of how to algorithmically recover $\mathbf{A}$. This is the crux of our algorithm; the full list is recovered by repeating this procedure. In this discussion we isolate and over-explain what we feel are the key ideas, in order to highlight them as much as possible. Because of this, there are several important (but in our opinion secondary) points which are swept under the rug, and so the algorithm presented in this section is a significant oversimplification of our work. Full details can be found in the body of the paper.

Our first observation is that for any $\mathbf{z} \in \mathbb{Z}_q^n$, we can get a map $f_{\mathbf{z}} : \mathbb{Z}_q^m \to \mathbb{Z}_q$ via $f_{\mathbf{z}}(\mathbf{x}) = \langle \mathbf{z}, f(\mathbf{x})\rangle$. The key point here is that

$$\mathbb{E}_{\mathbf{z}\sim\mathbb{Z}_q^n}\Big[\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f_{\mathbf{z}}(\mathbf{x}) = \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{x}\rangle\big]\Big] = 1/q + \big(1 - 1/q\big)\cdot\varepsilon,$$

from which it follows that if $\mathbf{z} \sim \mathbb{Z}_q^n$ is chosen uniformly, then $f_{\mathbf{z}}$ is likely to have non-trivial agreement with the linear function $\mathbf{x} \mapsto \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{x}\rangle$. In this case, the vector $\mathbf{A}^{\mathsf{t}}\mathbf{z} \in \mathbb{Z}_q^m$ can be recovered with non-negligible probability by running the (one-dimensional) Goldreich-Levin algorithm.

While this is a good start, this procedure only obtains what amounts to a single row of $\mathbf{A}$ with non-negligible probability. Thus, we have burned a considerable amount of probabilistic "good will" in order to make one step of progress. We cannot hope to recover every row of $\mathbf{A}$ in this fashion because the probability of running the procedure successfully n times in a row is exponentially small. For this reason, once we recover $\mathbf{A}^{\mathsf{t}}\mathbf{z} \in \mathbb{Z}_q^m$, we *double down*: we modify the function f to disregard $f(\mathbf{x})$ whenever $\langle \mathbf{z}, f(\mathbf{x})\rangle \neq \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{x}\rangle$. The new function stands to have better (conditional) agreement with $\mathbf{A}$ since

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\Big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\Big|\langle \mathbf{z}, f(\mathbf{x})\rangle = \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{x}\rangle\Big] \geq \varepsilon\big/\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[\langle \mathbf{z}, f(\mathbf{x})\rangle = \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{x}\rangle\big].$$

In other words, we hope to increase the chance that f agrees with $\mathbf{A}$ by conditioning on f agreeing with the rows of $\mathbf{A}$ which have already recovered. By looking at the above equation, we understand that whenever $\Pr_{\mathbf{x}}\big[\langle \mathbf{z}, f(\mathbf{x})\rangle = \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{x}\rangle\big]$ is bounded away from 1, we have made progress.

The above discussion suggests the following algorithm: repeat a small number of times the procedure 1) choose $\mathbf{z} \sim \mathbb{Z}_q^n$, 2) extract $\mathbf{A}^{\mathsf{t}}\mathbf{z}$ using Goldreich-Levin on $f_{\mathbf{z}}$, 3) update f to demand agreement with $\mathbf{x} \mapsto \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{x}\rangle$. Repeating only a small number of times ensures that we don't burn too much probabilistic good will. Our hope is that in only a few steps, the conditional agreement probability will grow so that it is close to 1. If this occurs, we will be able to easily recover the matrix.

This turns out not to fully work. However, before explaining the problems and how we fix them, let us start by looking at an example where it does work. The ideas from this simple algorithm make up an important special case of our main algorithm, so they are worthwhile to understand. Let $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ be a matrix, $T \subset \mathbb{Z}_q^m$ a subset of density $|T|/q^m = q^{-2}$, and let $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ be the function

$$f(\mathbf{x}) = \begin{cases} \mathbf{A}\mathbf{x}, & \mathbf{x} \in T \\ \$ \sim \mathbb{Z}_q^n, & \mathbf{x} \notin T \end{cases}$$

Suppose the above procedure is run three times with $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3 \in \mathbb{Z}_q^n$ and $\mathbf{A}^{\mathsf{t}}\mathbf{z}_1, \mathbf{A}^{\mathsf{t}}\mathbf{z}_2, \mathbf{A}^{\mathsf{t}}\mathbf{z}_3 \in \mathbb{Z}_q^m$ are recovered using the Goldreich-Levin theorem. Let $S \subset \mathbb{Z}_q^m$ be shorthand for the set of $\mathbf{x} \in \mathbb{Z}_q^m$ such that

$f_{\mathbf{z}_i}(\mathbf{x}) = \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}_i, \mathbf{x}\rangle$ holds for $i = 1, 2, 3$. The key point is that, conditioned on $\mathbf{x} \in S$, the chance that $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ is very good (about $1 - 1/q$). This means that membership in S (which can be efficiently tested) can be used as a proxy for f agreeing with $\mathbf{A}$, and this is good enough to allow recovering $\mathbf{A}$.

The problem with this plan, and the reason it does not fully work, is that in order to make progress we require the probability that $f_{\mathbf{z}}$ agrees with the linear function $\mathbf{x} \mapsto \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{x}\rangle$ to be both larger than $\frac{1}{q} + \delta$ (so that $\mathbf{A}^{\mathsf{t}}\mathbf{z}$ can be recovered using Goldreich-Levin) and also bounded away from 1 (so that sufficient progress is made). In general, it might be that for almost all $\mathbf{z}$, one but not both of these conditions hold. Indeed, consider the example function f which outputs $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ except with the first coordinate replaced with a random value. The functions $f_{\mathbf{z}}$ will have either perfect agreement with $\mathbf{x} \mapsto \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{x}\rangle$ (if $z_1 = 0$), or else $f_{\mathbf{z}}$ will be a random function (if $z_1 \neq 0$). In this case, our plan outlined above will not work since any $\mathbf{z}$ for which $f_{\mathbf{z}}$ has good agreement, enabling Goldreich-Levin recovery of $\mathbf{A}^{\mathsf{t}}\mathbf{z}$, will have perfect agreement, so the conditional probability does not increase.

Our main conceptual contribution is characterizing the conditions under which the above plan fails, and handling them. We prove using Fourier analysis that the only way the plan fails is if there exists a vector $\mathbf{w} \in \mathbb{Z}_q^n$ such that $f_{\mathbf{z}}$ having good agreement with $\mathbf{x} \mapsto \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{x}\rangle$ is heavily correlated with the event "$\mathbf{z} \perp \mathbf{w}$". Equivalently, the only way the plan fails is if there exists $\mathbf{w} \in \mathbb{Z}_q^n$ such that $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) \in \mathbf{A}\mathbf{x} + \mathrm{Span}(\mathbf{w})\right]$ is much larger than $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\right]$. In the above example where $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$ except with a random first coordinate, $f_{\mathbf{z}}$ has good agreement with $\mathbf{x} \mapsto \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{x}\rangle$ exactly when $\mathbf{z} \perp \mathbf{e}_1$ ($\mathbf{e}_1$ the first unit vector). Also, in this example $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\right] = 1/q$, while $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) \in \mathbf{A}\mathbf{x} + \mathrm{Span}(\mathbf{e}_1)\right] = 1$.

With the major ideas in place, we can now describe our algorithm and its analysis. Our algorithm begins by instantiating a set $S \subset \mathbb{Z}_q^m$ to $S = \mathbb{Z}_q^m$ and our analysis instantiates a subspace $\mathbf{W} \subset \mathbb{Z}_q^n$ to $\mathbf{W} = \{\mathbf{0}\}$. The analysis keeps track of the probability potential $\mathsf{P} := \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) \in \mathbf{A}\mathbf{x}+\mathbf{W}\middle|\mathbf{x} \in S\right]$. Note at the start of the algorithm $\mathsf{P} \geq \varepsilon$ holds. The algorithm and analysis now proceed in stages. In each stage P will be increased either by updating S or $\mathbf{W}$. Specifically, if conditions are such that progress is likely to be made by running the Goldreich-Levin procedure, then the algorithm draws $\mathbf{z} \sim \mathbb{Z}_q^n$, recovers $\mathbf{A}^{\mathsf{t}}\mathbf{z}$ using Goldreich-Levin, and sets $S = S \cap \left\{\mathbf{x} \in \mathbb{Z}_q^m : f_{\mathbf{z}}(\mathbf{x}) = \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{x}\rangle\right\}$. With non-negligible probability, P will increase. If, on the other hand, conditions are *not* such that progress is likely via the Goldreich-Levin procedure, then by the Fourier analysis argument, significant progress can be made by adding some $\mathbf{w} \in \mathbb{Z}_q^n$ to $\mathbf{W}$, thereby increasing $\mathbf{W}$ by one dimension. After about $\log_q(1/\varepsilon)$ stages, P will be close enough to 1 that $\mathbf{A}$ can be recovered with good probability.

2 Preliminaries

Basic Notation. For a prime $q \in \mathbb{N}$, we denote by $\mathbb{Z}_q$ the field of integers modulo q. We will denote scalars, vectors and matrices with lowercase italic, lowercase bold, and uppercase bold respectively (*e.g.*, $z \in \mathbb{Z}_q$, $\mathbf{z} \in \mathbb{Z}_q^n$ and $\mathbf{Z} \in \mathbb{Z}_q^{n\times m}$). Given a matrix $\mathbf{Z} \in \mathbb{Z}_q^{n\times m}$, we denote by $\mathbf{Z}^{\mathsf{t}} \in \mathbb{Z}_q^{m\times n}$ its transpose. For vectors $\mathbf{z}, \mathbf{w} \in \mathbb{Z}_q^n$, we write $\langle \mathbf{z}, \mathbf{w}\rangle$ for their dot product: $\langle \mathbf{z}, \mathbf{w}\rangle = z_1 w_1 + \cdots + z_n w_n$. For a distribution $\mathcal{D}$ (resp. set D), we write $r \sim \mathcal{D}$ (resp. $r \sim D$) to indicate that the random variable r is drawn according to $\mathcal{D}$ (resp. the uniform distribution on D). For an event $\mathbf{E}$, we denote by $\mathbb{1}_{\mathbf{E}}$ the indicator random variable corresponding to $\mathbf{E}$. Namely, $\mathbb{1}_{\mathbf{E}} = 1$ (resp. $\mathbb{1}_{\mathbf{E}} = 0$) when $\mathbf{E}$ occurs (resp. does not occur).

Linear Algebra. We assume familiarity with basic concepts from linear algebra. For example, if $\mathbf{W} \subset \mathbb{Z}_q^n$ is a subspace, then we denote by $\mathbf{W}^{\perp} \subset \mathbb{Z}_q^n$ the set $\{\mathbf{z} \in \mathbb{Z}_q^n : \langle \mathbf{z}, \mathbf{w}\rangle = 0 \,\forall\, \mathbf{w} \in \mathbf{W}\}$. It is

known that $\mathbf{W}^{\perp} \subset \mathbb{Z}_q^n$ is a subspace of dimension $n-d$, where $d = \dim(\mathbf{W})$. For a matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ and vectors $\mathbf{x} \in \mathbb{Z}_q^m$, $\mathbf{z} \in \mathbb{Z}_q^n$ we will use the identity $\langle \mathbf{z}, \mathbf{A}\mathbf{x} \rangle = \langle \mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{x} \rangle$.

Fourier Analysis. The set of functions $\{f : \mathbb{Z}_q^n \to \mathbb{C}\}$ is a complex vector space of dimension q^n. The Fourier basis is $\{\chi_{\mathbf{w}} : \mathbf{w} \in \mathbb{Z}_q^n\}$ where $\chi_{\mathbf{w}}(\mathbf{z}) = \omega^{\langle \mathbf{z}, \mathbf{w} \rangle}$ where $\omega = e^{2\pi i/q}$ is a primitive $q-$th root of unity on the complex unit circle. When a function $f : \mathbb{Z}_q^n \to \mathbb{C}$ is represented with respect to the Fourier basis, we denote by $\hat{f}(\mathbf{w})$ the complex coefficient in front of $\chi_{\mathbf{w}}$, so $f = \sum_{\mathbf{w} \in \mathbb{Z}_q^n} \hat{f}(\mathbf{w}) \chi_{\mathbf{w}}$. We will use the following well known facts.

Claim 1 (Basic Fourier Analysis). *Let $q, n \in \mathbb{N}$ be integers with q prime.*

- **(a) The Fourier Basis is Orthonormal:** *For any subspace $\mathbf{W} \subset \mathbb{Z}_q^n$, $\mathbb{E}_{\mathbf{z} \sim \mathbf{W}}\left[\omega^{\langle \mathbf{z}, \mathbf{w} \rangle}\right] = \mathbb{1}_{\mathbf{w} \in \mathbf{W}^{\perp}}$.*
- **(b) Explicit Formula for the Fourier Coefficients:** *For any function $f : \mathbb{Z}_q^n \to \mathbb{C}$, the Fourier coefficients are given by: $\hat{f}(\mathbf{w}) = \mathbb{E}_{\mathbf{z} \sim \mathbb{Z}_q^n}\left[f(\mathbf{z}) \cdot \omega^{-\langle \mathbf{z}, \mathbf{w} \rangle}\right]$.*
- **(c) Parseval's Identity:** *For any $f : \mathbb{Z}_q^n \to \mathbb{C}$, we have $\mathbb{E}_{\mathbf{z} \sim \mathbb{Z}_q^n}\left[|f(\mathbf{z})|^2\right] = \sum_{\mathbf{w} \in \mathbb{Z}_q^n} |\hat{f}(\mathbf{w})|^2$.*

The One-Dimensional Goldreich-Levin Theorem. The following generalization of the Goldreich-Levin theorem [GL89] to large fields was proved in [AGS03].

Claim 2. *Let $q, m \in \mathbb{N}$ be integers such that q is prime. There is a randomized oracle algorithm $\mathcal{A}_{\mathsf{GL}}$ which has the following syntax, runtime and correctness guarantees.*

- **Syntax:** *$\mathcal{A}_{\mathsf{GL}}$ is parametrized by $\varepsilon > 0$, takes no input, gets oracle access to $f : \mathbb{Z}_q^m \to \mathbb{Z}_q$ and outputs $\mathbf{u} \in \mathbb{Z}_q^m$.*
- **Running Time:** *$\mathcal{A}_{\mathsf{GL}}$ runs in time $\mathsf{poly}\big(m, \log q, \log(1/\varepsilon)\big)$.*
- **Correctness:** *If f is such that $\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\left[f(\mathbf{x}) = \langle \mathbf{u}, \mathbf{x} \rangle\right] \geq 1/q + \varepsilon$, then $\mathcal{A}_{\mathsf{GL}}$ outputs $\mathbf{u}$ with probability at least ε^2 (probability over the random coins of $\mathcal{A}_{\mathsf{GL}}$).*

The Chernoff-Hoeffding Inequality. Let $n \in \mathbb{N}$ be an integer, $\mu_1, \ldots, \mu_n \in [0,1]$ and $\delta > 0$. Suppose $X_1, \ldots, X_n$ are n independent copies of the $0/1$ random variables whose expectations are $\mathbb{E}[X_i] = \mu_i$. Let $X = \frac{1}{n} \cdot \big(X_1 + \cdots + X_n\big)$ be their mean, so $\mathbb{E}[X] = \frac{1}{n} \cdot \big(\mu_1 + \cdots + \mu_n\big) =: \mu$. The Chernoff-Hoeffding inequality says:

$$\Pr\left[|X - \mu| > \delta\right] \leq 2e^{-\frac{\delta^2}{3\mu}n}.$$

The following form will be useful to us.

Claim 3. *Let $m, n, q \in \mathbb{N}$ be integers with $m \geq 18$, $n \geq 2$ and q prime, let $\delta \geq q^{-m/3}$ be a parameter, $T \subset \mathbb{Z}_q^m$ a subset of density $\tau := |T|/q^m$, and let $\mathcal{G}$ be the set of all functions mapping T into $\mathbb{Z}_q^n$, so $\mathcal{G} := \big\{g : T \to \mathbb{Z}_q^n\big\}$. Then*

$$\Pr_{g \sim \mathcal{G}}\Big[\exists\, \mathbf{A} \in \mathbb{Z}_q^{n \times m} \text{ st } \big|\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\left[g(\mathbf{x}) = \mathbf{A}\mathbf{x} \;\&\; \mathbf{x} \in T\right] - \tau q^{-n}\big| > \delta\Big] \leq e^{-m}.$$

Proof. For $g \in \mathcal{G}$ and $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$, let $\mathsf{P}(g,\mathbf{A}) := \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[g(\mathbf{x}) = \mathbf{A}\mathbf{x} \,\&\, \mathbf{x} \in T\big]$ be shorthand. Note that a random function $g \sim \mathcal{G}$ is specified by independently choosing random outputs in $\mathbb{Z}_q^n$ for every input $\mathbf{x} \in T$. For a matrix $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$, define the $0/1$ random variables $\{X_{\mathbf{x}}^{\mathbf{A}}\}_{\mathbf{x}\in\mathbb{Z}_q^m}$ via $X_{\mathbf{x}}^{\mathbf{A}} = \mathbb{1}_{\mathbf{x}\in T}\cdot\mathbb{1}_{g(\mathbf{x})=\mathbf{A}\mathbf{x}}$ (randomness over $g\sim\mathcal{G}$). Note that for all $\mathbf{A}\in\mathbb{Z}_q^{n\times m}$, the $X_{\mathbf{x}}^{\mathbf{A}}$ are independent copies of the $0/1$ random variables whose expectations are

$$\mathbb{E}\big[X_{\mathbf{x}}^{\mathbf{A}}\big] = \begin{cases} q^{-n}, & \mathbf{x}\in T \\ 0, & \mathbf{x}\notin T \end{cases},$$

and furthermore, the quantity $\mathsf{P}(g,\mathbf{A})$ is equal to $q^{-m}\cdot\sum_{\mathbf{x}\in\mathbb{Z}_q^m} X_{\mathbf{x}}^{\mathbf{A}}$. Thus, the union bound and the Chernoff-Hoeffding inequality combine to give

$$\begin{aligned}\Pr_{g\sim\mathcal{G}}\Big[\exists\,\mathbf{A}\in\mathbb{Z}_q^{n\times m}\text{ s.t. }\big|\mathsf{P}(g,\mathbf{A})-\tau q^{-n}\big|>\delta\Big] &\le q^{mn}\cdot\max_{\mathbf{A}\in\mathbb{Z}_q^{n\times m}}\Big\{\Pr_{g\sim\mathcal{G}}\Big[\big|\mathsf{P}(g,\mathbf{A})-\tau q^{-n}\big|>\delta\Big]\Big\}\\ &\le 2q^{mn}e^{-\frac{\delta^2 q^{m+n}}{3\tau}} \le 2e^{-(\frac{1}{3}\cdot q^{m/3+n}-mn\ln q)},\end{aligned}$$

using $\delta \ge q^{-m/3}$ and $\tau \le 1$. Finally, note that

$$\frac{1}{3}q^{m/3+n} - mn\ln q \ge \frac{1}{3}q^{m/3+2} - 2m\ln q \ge \frac{1}{3}2^{m/3+2} - 2m\ln 2 \ge 2^{m/3} - 2m \ge m+1,$$

completing the proof. The first inequality above holds because the function $\phi(x) = \frac{1}{3}q^{m/3+x} - (m\ln q)x$ has positive derivative for all $x \ge 2$ when $q \ge 2$ and $m \ge 8$; the second inequality holds because $\psi(x) = \frac{1}{3}x^{m/3+2} - 2m\ln x$ has positive derivative for all $x \ge 2$ when $m \ge 9$; the final inequality holds because $\chi(x) = 2^{x/3} - 3x - 1$ is positive for all $x \ge 18$. □

3 Proving the Main Theorem

We restate Theorem 1 below in a more quantitative form.

Theorem 1 (Restated). *Let $m,n,q\in\mathbb{N}$ be integers with $m\ge 18$ and q prime, let $\varepsilon > 0$ be such that $\varepsilon \ge 12\cdot\max\{q^{-m/9}, q^{-n/3}\}$, and let $f:\mathbb{Z}_q^m\to\mathbb{Z}_q^n$ be a function. There exists an algorithm $\mathcal{A}_{\mathsf{List.Dec}}$ with the following syntax, runtime and correctness guarantees.*

- **Syntax:** *$\mathcal{A}_{\mathsf{List.Dec}}$ takes no input, gets oracle access to f, and outputs a set $\mathsf{L}\subset\mathbb{Z}_q^{n\times m}$ of size at most $|\mathsf{L}|\le 9/\varepsilon$.*
- **Runtime:** *$\mathcal{A}_{\mathsf{List.Dec}}$ runs in expected time $\mathsf{poly}\big(n,\log q, m^{\log_q(1/\varepsilon)}, \varepsilon^{-\log_q(1/\varepsilon)}\big)$.*
- **Correctness:** *With probability at least $1-2^{-\Omega(m)}$ over the randomness of $\mathsf{L}\sim\mathcal{A}_{\mathsf{List.Dec}}^{(\varepsilon,f)}$, the following both hold:*
  - *for all $\mathbf{A}\in\mathsf{L}$, $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x})=\mathbf{A}\mathbf{x}\big]\ge\varepsilon/10$;*
  - *for all $\mathbf{A}'\in\mathbb{Z}_q^{n\times m}$, either $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x})=\mathbf{A}'\mathbf{x}\big]<\varepsilon$, or $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[\mathbf{A}'\mathbf{x}=\mathbf{A}\mathbf{x}\big]\ge\varepsilon^2/36$ for some $\mathbf{A}\in\mathsf{L}$.*

3.1 Agreement Decoding Implies List Decoding

Proving Theorem 1 requires designing an algorithm which outputs a list of matrices which "explains" all of f's linear agreement. In this section, we show that it suffices to design an algorithm which reconstructs a single matrix which has good agreement with f. The list decoding algorithm works by simply calling this algorithm repeatedly.

Theorem 2. *Let $m, n, q \in \mathbb{N}$ be integers with $m \geq 18$, $n \geq 2$ and q prime, let $\varepsilon \geq 12q^{-m/3}$, and let $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ be a function. There exists a randomized algorithm $\mathcal{A}_{\mathsf{Matrix.Rec}}$ which has the following syntax, running time and correctness guarantees.*

- **Syntax:** *$\mathcal{A}_{\mathsf{Matrix.Rec}}$ takes no input, gets oracle access to f, and outputs a matrix $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n \times m}$.*
- **Running Time:** *$\mathcal{A}_{\mathsf{Matrix.Rec}}$ runs in expected $\mathsf{poly}\big(m, n, \log q, 1/\varepsilon\big)$ time.*
- **Correctness:** *With probability at least $\mathsf{poly}\big(1/n, \varepsilon^{\log_q(1/\varepsilon)}, m^{-\log_q(1/\varepsilon)}\big)$,*

$$\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_{\mathsf{out}}\mathbf{x}\big] \geq \max_{\mathbf{A} \in \mathbb{Z}_q^{n \times m}} \Big\{\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big]\Big\} - \varepsilon.$$

Proof of Theorem 1 Using Theorem 2. Let $\varepsilon > 0$ be the parameter from Theorem 1, and assume that $n \geq 2$ since when $n = 1$, Theorem 1 follows easily from Claim 2 (the one-dimensional Goldreich-Levin theorem). Our algorithm $\mathcal{A}_{\mathsf{List.Dec}}$ will call the algorithm $\mathcal{A}_{\mathsf{Matrix.Rec}}$ from Theorem 2 several times with parameter $\varepsilon/3$, and our proof will invoke Claim 3 several times with parameter $\delta = \frac{1}{2} \cdot \big(\frac{\varepsilon}{6}\big)^3$, so that $\delta \geq q^{-m/3}$ is satisfied. Note $q^{-n} \leq \frac{1}{2} \cdot \big(\frac{\varepsilon}{6}\big)^3$ ensures that $\big(q^{-n} + \delta\big) \leq \big(\frac{\varepsilon}{6}\big)^3$. Our algorithm $\mathcal{A}_{\mathsf{List.Dec}}$ works by initializing a list $\mathsf{L} \subset \mathbb{Z}_q^{n \times m}$ to empty $\mathsf{L} = \{\,\}$, a function $g : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ to $g = f$, and doing the following ℓ times, where $\ell = 9/\varepsilon$ is an upper bound on the size of the list we will output:

- call $\mathcal{A}^{(\varepsilon/3, g)}_{\mathsf{Matrix.Rec}}$ $N = m/p$ times, where $p = \mathsf{poly}\big(1/n, \varepsilon^{\log_q(1/\varepsilon)}, m^{-\log_q(1/\varepsilon)}\big)$ is the success probability of $\mathcal{A}_{\mathsf{Matrix.Rec}}$. Let $\mathbf{A}_1, \ldots, \mathbf{A}_N \in \mathbb{Z}_q^{n \times m}$ be the outputs.
- For each $j = 1, \ldots, N$, approximate $\mathsf{P}_j := \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[g(\mathbf{x}) = \mathbf{A}_j\mathbf{x}\big]$ using m/ε samples.
- Let $j \in \{1, \ldots, N\}$ be such that P_j is maximal. If $\mathsf{P}_j \geq 2\varepsilon/9$, add $\mathbf{A}_j$ to L and update g by changing $g(\mathbf{x})$ to random for all $\mathbf{x} \in \mathbb{Z}_q^m$ such that $g(\mathbf{x}) = \mathbf{A}_j\mathbf{x}$.

After the loop completes, $\mathcal{A}_{\mathsf{List.Dec}}$ outputs L. Note that $\mathcal{A}_{\mathsf{List.Dec}}$ has the required syntax and runtime, and also note that $|\mathsf{L}| \leq \ell$ clearly holds since the loop is executed ℓ times, with at most one matrix being added to L each time. Additionally, each of the matrices $\mathbf{A}_j \in \mathbb{Z}_q^{n \times m}$ which gets added to L has $\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[g(\mathbf{x}) = \mathbf{A}_j\mathbf{x}\big] \geq \varepsilon/9$ with probability at least $1 - 2^{-\Omega(m)}$. This is because $\mathsf{P}_j \geq 2\varepsilon/9$ and P_j is an approximation of $\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[g(\mathbf{x}) = \mathbf{A}_j\mathbf{x}\big]$ which is accurate to within $\varepsilon/9$ with probability $1 - 2^{-\Omega(m)}$ by the Chernoff-Hoeffding inequality. Furthermore, at all times during the algorithm, the function g is the same as f except that some of the inputs have had their output overwritten with a random value. So if we let $T \subset \mathbb{Z}_q^m$ be the set of inputs whose output has been overwritten with a random value at the time when $\mathbf{A}_j \in \mathbb{Z}_q^{n \times m}$ is added to L, we see that with probability $1 - 2^{-\Omega(m)}$ over the randomness used to overwrite the outputs,

$$\begin{aligned}\frac{\varepsilon}{9} &\leq \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[g(\mathbf{x}) = \mathbf{A}_j\mathbf{x}\big] = \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_j\mathbf{x} \;\&\; \mathbf{x} \notin T\big] + \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[g(\mathbf{x}) = \mathbf{A}_j\mathbf{x} \;\&\; \mathbf{x} \in T\big] \\ &\leq \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_j\mathbf{x}\big] + \big(q^{-n} + \delta\big) \leq \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_j\mathbf{x}\big] + \frac{\varepsilon}{90},\end{aligned}$$

holds by Claim 3, using $q^{-n} + \delta \leq \varepsilon/90$. It follows that $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_j\mathbf{x}\big] \geq \varepsilon/10$ holds for all $\mathbf{A}_j \in \mathsf{L}$. Therefore, to finish the proof, it remains to show that when $\mathcal{A}_{\mathsf{List.Dec}}$ terminates, at least one of the following holds for every $\mathbf{A}' \in \mathbb{Z}_q^{n\times m}$:

(1) $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}'\mathbf{x}\big] < \varepsilon$;

(2) there exists some $\mathbf{A} \in \mathsf{L}$ such that $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[\mathbf{A}\mathbf{x} = \mathbf{A}'\mathbf{x}\big] \geq \frac{\varepsilon}{4\ell}$.

For this purpose, we set some notation. For $i = 1, \ldots, \ell$, let $g_i : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ denote the function g during the $i-$th execution of the above loop; and let $\Phi_i := \max_{\mathbf{A}'\in\mathbb{Z}_q^{n\times m}}\big\{\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[g_i(\mathbf{x}) = \mathbf{A}'\mathbf{x}\big]\big\}$. Note that if $\Phi_i \geq 2\varepsilon/3$, then with probability at least $1 - 2^{-\Omega(m)}$, a matrix will be added to L during the $i-$th loop execution. Indeed, by Theorem 2, with probability $1 - 2^{-\Omega(m)}$, at least one of the N matrices $\{\mathbf{A}_j\}_{j=1,\ldots,N}$ computed by $\mathcal{A}_{\mathsf{Matrix.Rec}}$ is such that

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[g_i(\mathbf{x}) = \mathbf{A}_j\mathbf{x}\big] \geq \max_{\mathbf{A}'\in\mathbb{Z}_q^{n\times m}}\Big\{\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[g_i(\mathbf{x}) = \mathbf{A}'\mathbf{x}\big]\Big\} - \varepsilon/3 \geq \varepsilon/3,$$

in which case $\mathsf{P}_j \geq 2\varepsilon/9$ holds (and thus $\mathbf{A}_j$ gets added to L) with probability $1 - 2^{-\Omega(m)}$ by the Chernoff-Hoeffding inequality.

Now, let $\Phi_{\ell+1} := \max_{\mathbf{A}'\in\mathbb{Z}_q^{n\times m}}\big\{\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[g_{\ell+1}(\mathbf{x}) = \mathbf{A}'\mathbf{x}\big]\big\}$, where $g_{\ell+1}$ denotes the function g after the $\ell-$th loop execution has finished and $\mathcal{A}_{\mathsf{List.Dec}}$ has terminated. We complete the proof by proving two things: (i) that $\Phi_{\ell+1}$ is small; (ii) small $\Phi_{\ell+1}$ implies that for every matrix, at least one of the two conditions specified above holds. Specifically, we show:

- with probability at least $1 - 2^{-\Omega(m)}$, $\Phi_{\ell+1} < 3\varepsilon/4$ holds;
- if $\Phi_{\ell+1} < 3\varepsilon/4$, then either (1) or (2) holds for every $\mathbf{A}' \in \mathbb{Z}_q^{n\times m}$.

We begin with the second point since it is easier. Suppose $\Phi_{\ell+1} < 3\varepsilon/4$, and let $\mathbf{A}' \in \mathbb{Z}_q^{n\times m}$ be such that $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}'\mathbf{x}\big] \geq \varepsilon$, *i.e.*, so that (1) does not hold for $\mathbf{A}'$. Let $T \subset \mathbb{Z}_q^m$ be the set of $\mathbf{x} \in \mathbb{Z}_q^m$ who have had their outputs overwritten with a random output in the final function $g_{\ell+1}$. So in other words, $T = \big\{\mathbf{x} \in \mathbb{Z}_q^m : \exists \mathbf{A} \in \mathsf{L} \text{ s.t. } f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big\}$. Since $f = g_{\ell+1}$ away from T, we have

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}'\mathbf{x} \,\&\, \mathbf{x} \notin T\big] \leq \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[g_{k+1}(\mathbf{x}) = \mathbf{A}'\mathbf{x}\big] < 3\varepsilon/4,$$

and so

$$\begin{aligned}\varepsilon \leq \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}'\mathbf{x}\big] &< \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}'\mathbf{x} \,\&\, \mathbf{x} \in T\big] + 3\varepsilon/4 \\ &\leq |\mathsf{L}| \cdot \max_{\mathbf{A}\in\mathsf{L}}\Big\{\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}'\mathbf{x} \,\&\, f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big]\Big\} + 3\varepsilon/4,\end{aligned}$$

by the union bound. Rearranging and using $|\mathsf{L}| \leq \ell$ implies that condition (2) holds for $\mathbf{A}'$.

To prove the first point, note that if there exists some $k \in \{1, \ldots, \ell\}$ such that $\Phi_k < 2\varepsilon/3$, then $\Phi_{\ell+1} < 3\varepsilon/4$ holds with probability $1 - 2^{-\Omega(m)}$. This is because $g_{\ell+1}$ is the same function as g_k except that some of the outputs have been changed to random values in $\mathbb{Z}_q^n$. Specifically, if we let $T \subset \mathbb{Z}_q^m$ be the inputs whose outputs under g_k and $g_{\ell+1}$ differ, then because $g_{\ell+1}$ is random on T, we have by Claim 3 that with probability at least $1 - 2^{-\Omega(m)}$, for all $\mathbf{A}' \in \mathbb{Z}_q^{n\times m}$:

$$\begin{aligned}\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[g_{\ell+1}(\mathbf{x}) = \mathbf{A}'\mathbf{x}\big] &= \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[g_{\ell+1}(\mathbf{x}) = \mathbf{A}'\mathbf{x} \,\&\, \mathbf{x} \notin T\big] + \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[g_{\ell+1}(\mathbf{x}) \,\&\, \mathbf{x} \in T\big] \\ &\leq \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[g_k(\mathbf{x}) = \mathbf{A}'\mathbf{x}\big] + \big(q^{-n} + \delta\big) \leq \Phi_k + \big(q^{-n} + \delta\big) < 3\varepsilon/4,\end{aligned}$$

using $q^{-n} + \delta < \varepsilon/12$. Therefore, assume that $\Phi_1, \ldots, \Phi_\ell \geq 2\varepsilon/3$ and let us show in this case that $\Phi_{\ell+1} < 3\varepsilon/4$ holds. This will complete the proof.

By the discussion above, since $\Phi_i \geq 2\varepsilon/3$ holds for all $i = 1, \ldots, \ell$, a matrix will be added to L in every loop execution with probability $1 - 2^{-\Omega(m)}$. Let $\mathbf{A}_1, \ldots, \mathbf{A}_\ell \in \mathbb{Z}_q^{n \times m}$ denote these matrices and let $T_i \subset \mathbb{Z}_q^m$ be the set of inputs whose outputs are overwritten by a random value because of agreement of g_i with $\mathbf{A}_i$; namely $T_i = \{\mathbf{x} \in \mathbb{Z}_q^m : g_i(\mathbf{x}) = \mathbf{A}_i\mathbf{x}\}$. By the discussion above, $|T_i| \geq (\varepsilon/9)q^m$ holds for all i. By Claim 3, with probability $1 - 2^{-\Omega(m)}$, we have

$$\left|T_i \cap T_{i'}\right| = q^m \cdot \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\left[g_{i'}(\mathbf{x}) = \mathbf{A}_{i'}\mathbf{x} \;\&\; \mathbf{x} \in T_i\right] \leq q^m \cdot \left(q^{-n} + \delta\right),$$

for all $i < i'$, since $g_{i'}$ is random on T_i. By inclusion-exclusion, we have

$$\left|\bigcup_{i=1}^{\ell} T_i\right| \geq \sum_{i=1}^{\ell} |T_i| - \sum_{i<i'} \left|T_i \cap T_{i'}\right| \geq q^m \cdot \left(\frac{\ell\varepsilon}{9} - \frac{\ell^2}{2}\left(q^{-n} + \delta\right)\right) = q^m \cdot \left(1 - \frac{\ell^2}{2} \cdot \left(q^{-n} + \delta\right)\right).$$

Using Claim 3 one more time gives that for all $\mathbf{A}' \in \mathbb{Z}_q^{n \times m}$,

$$\begin{aligned}\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\left[g_{\ell+1}(\mathbf{x}) = \mathbf{A}'\mathbf{x}\right] &\leq \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\left[g_{\ell+1}(\mathbf{x}) = \mathbf{A}'\mathbf{x} \;\&\; \mathbf{x} \in \bigcup_{i=1}^{\ell} T_i\right] + \frac{\ell^2}{2} \cdot \left(q^{-n} + \delta\right) \\ &\leq \left(\frac{\ell^2}{2} + 1\right) \cdot \left(q^{-n} + \delta\right),\end{aligned}$$

since $g_{\ell+1}$ is random on $\bigcup_i T_i$. Thus $\Phi_{\ell+1} \leq \left(\frac{\ell^2}{2} + 1\right) \cdot \left(q^{-n} + \delta\right) \leq 3\varepsilon/4$ (follows from $q^{-n} + \delta \leq \left(\frac{\varepsilon}{6}\right)^3$ and $\ell = 9/\varepsilon$), as desired. □

3.2 Proving Theorem 2 via Two Lemmas

As discussed in the introduction, the algorithm $\mathcal{A}_{\mathsf{Matrix.Rec}}$ works by initializing a subset $S \subset \mathbb{Z}_q^m$ to the entire domain $S = \mathbb{Z}_q^m$, and proceeds to iteratively shrink S so that the chance of f agreeing with $\mathbf{A}$ *given that the input is in* S increases. Once this probability is sufficiently close to 1, $\mathbf{A}$ (or, at least, some other matrix which is almost as good) can be recovered algorithmically. For this reason, we break our algorithm $\mathcal{A}_{\mathsf{Matrix.Rec}}$ into two subroutines: $\mathcal{A}_{\mathsf{amplify}}$, which amplifies f's conditional agreement; and $\mathcal{A}_{\mathsf{output}}$ which recovers an output matrix once f's conditional agreement is sufficiently close to 1. The following lemmas summarize the correctness guarantees for these two subroutines.

Lemma 1 (The Amplification Algorithm $\mathcal{A}_{\mathsf{amplify}}$). *Let $m, n, q \in \mathbb{N}$ be integers with q prime, let $\varepsilon > 0$, and let $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ be a function. There exists a randomized algorithm $\mathcal{A}_{\mathsf{amplify}}$ which has the following syntax, running time and correctness guarantees.*

- **Syntax:** *$\mathcal{A}_{\mathsf{amplify}}$ takes no input, gets oracle access to f, and outputs the characteristic function of a set $S_{\mathsf{out}} \subset \mathbb{Z}_q^m$.*
- **Running Time:** *$\mathcal{A}_{\mathsf{amplify}}$ runs in expected $\mathsf{poly}\left(m, n, \log q, \log(1/\varepsilon)\right)$ time; the characteristic function which $\mathcal{A}_{\mathsf{amplify}}$ outputs can be computed in time $\mathsf{poly}\left(m, n, \log q\right)$ using one oracle call to f.*

- **Correctness:** *If* $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ *is such that* $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\right] \geq \varepsilon$*, then with probability at least* $\mathsf{poly}\left(\varepsilon^{\log_q(1/\varepsilon)}, m^{-\log_q(1/\varepsilon)}\right)$ *over* $S_{\mathsf{out}} \sim \mathcal{A}_{\mathsf{amplify}}^{(\varepsilon,f)}$, $\left\{\mathbf{x}\in\mathbb{Z}_q^m : f(\mathbf{x}) = \mathbf{A}\mathbf{x}\right\} \subset S_{\mathsf{out}}$ *holds, and additionally, there exists a subspace* $\mathbf{W} \subset \mathbb{Z}_q^n$ *of dimension at most* $\dim(\mathbf{W}) \leq r$ *for a parameter* $r = \mathcal{O}\left(\log_q(1/\varepsilon)\right)$ *(here the* $\mathcal{O}$ *is hiding an absolute constant, independent of all other parameters) such that*

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) \in \mathbf{A}\mathbf{x} + \mathbf{W}\middle|\mathbf{x}\in S_{\mathsf{out}}\right] \geq 1 - \frac{1}{4m}.$$

Lemma 2 (The Output Recovery Algorithm $\mathcal{A}_{\mathsf{output}}$). *Let* $m, n, q, r \in \mathbb{N}$ *be integers with* q *prime, let* $\varepsilon \geq 12q^{-m/3}$*, let* $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ *be a function, and* $S \subset \mathbb{Z}_q^m$ *a subset. There exists a randomized algorithm* $\mathcal{A}_{\mathsf{output}}$ *which has the following syntax, running time and correctness guarantees.*

- **Syntax:** $\mathcal{A}_{\mathsf{output}}$ *takes no input, gets oracle access to* f *and to the characteristic function of* S*, and outputs a matrix* $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n\times m}$*.*
- **Running Time:** $\mathcal{A}_{\mathsf{output}}$ *runs in expected* $\mathsf{poly}\left(m, n, \log q, 1/\varepsilon\right)$ *time.*
- **Correctness:** *Assume* $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ *is such that* $\left\{\mathbf{x}\in\mathbb{Z}_q^m : f(\mathbf{x}) = \mathbf{A}\mathbf{x}\right\} \subset S$*, and also that* $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) \in \mathbf{A}\mathbf{x} + \mathbf{W}\middle|\mathbf{x}\in S\right] \geq 1 - \frac{1}{4m}$ *holds for some subspace* $\mathbf{W} \subset \mathbb{Z}_q^n$ *of dimension at most* r*. Then with probability at least* $\mathsf{poly}\left(1/n, \varepsilon^r, \varepsilon^{\log_q(1/\varepsilon)}, m^{-\log_q(1/\varepsilon)}\right)$*, the output matrix* $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n\times m}$ *satisfies*

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) = \mathbf{A}_{\mathsf{out}}\mathbf{x}\right] \geq \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\right] - \varepsilon.$$

Proof of Theorem 2 Assuming Lemmas 1 and 2. The algorithm $\mathcal{A}_{\mathsf{Matrix.Rec}}$ simply calls $\mathcal{A}_{\mathsf{amplify}}$ with oracle access to f to obtain the characteristic function of a set $S \subset \mathbb{Z}_q^m$; then it calls $\mathcal{A}_{\mathsf{output}}$ with oracle access to f and to the characteristic function of S to obtain $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n\times m}$; if either subroutine fails to give output $\mathcal{A}_{\mathsf{Matrix.Rec}}$ aborts. It is clear that $\mathcal{A}_{\mathsf{Matrix.Rec}}$ has the required syntax and running time. For correctness, let $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ be such that $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\right]$ is maximal; we can assume this probability is at least δ, since otherwise any output matrix $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n\times m}$ will trivially satisfy the requirement. By Lemma 1, with probability at least $\mathsf{poly}\left(\varepsilon^{\log_q(1/\varepsilon)}, m^{-\log_q(1/\varepsilon)}\right)$, $\mathcal{A}_{\mathsf{amplify}}$ outputs $S \subset \mathbb{Z}_q^m$ such that $\left\{\mathbf{x}\in\mathbb{Z}_q^m : f(\mathbf{x}) = \mathbf{A}\mathbf{x}\right\}$ holds and also so that $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) \in \mathbf{A}\mathbf{x} + \mathbf{W}\middle|\mathbf{x}\in S\right] \geq 1 - \frac{1}{4m}$ holds for some subspace $\mathbf{W} \subset \mathbb{Z}_q^n$ of dimension at most r. By Lemma 2, when $\mathcal{A}_{\mathsf{output}}$ is run with oracle access to f and to the characteristic function of S, it outputs $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n\times m}$ such that

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) = \mathbf{A}_{\mathsf{out}}\mathbf{x}\right] \geq \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\left[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\right] - \varepsilon$$

with probability at least $\mathsf{poly}\left(1/n, \varepsilon^{\log_q(1/\varepsilon)}, m^{-\log_q(1/\varepsilon)}\right)$, using $r = \mathcal{O}\left(\log_q(1/\varepsilon)\right)$. The result follows. □

4 The Goldreich-Levin Machine

In order to make repeated use the one-dimensional Goldreich-Levin theorem of Claim 2, we analyze an algorithmic process we call the "Goldreich-Levin Machine", and denote by $\mathcal{A}_{\mathsf{GL.machine}}$.

The Goldreich-Levin Machine. Let $m, n, q \in \mathbb{N}$ be integers with q prime and let $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ be a function. $\mathcal{A}_{\mathsf{GL.machine}}$ is defined with a parameter $\delta > 0$, gets oracle access to f, takes a set $S \subset \mathbb{Z}_q^m$ as input (more precisely, $\mathcal{A}_{\mathsf{GL.machine}}$ gets oracle access to $\mathbb{1}_S$, the characteristic function of S), and outputs another set $S' \subset \mathbb{Z}_q^m$ (specifically, $\mathcal{A}_{\mathsf{GL.machine}}$ outputs the code for the characteristic function $\mathbb{1}_{S'}$). The process works by drawing $\mathbf{z} \sim \mathbb{Z}_q^n$ uniformly and then drawing $\mathbf{u} \sim \mathcal{A}_{\mathsf{GL}}^{(\delta, f_{\mathbf{z}})}$, where $\mathcal{A}_{\mathsf{GL}}$ is the one-dimensional Goldreich-Levin algorithm from Claim 2 and $f_{\mathbf{z}} : \mathbb{Z}_q^m \to \mathbb{Z}_q$ is the function

$$f_{\mathbf{z}}(\mathbf{x}) = \begin{cases} \langle \mathbf{z}, f(\mathbf{x}) \rangle, & \mathbf{x} \in S \\ \$ \sim \mathbb{Z}_q, & \mathbf{x} \notin S \end{cases} .$$

Then S' is set to $S' = S \cap \left\{ \mathbf{x} \in \mathbb{Z}_q^m : \langle \mathbf{z}, f(\mathbf{x}) \rangle = \langle \mathbf{u}, \mathbf{x} \rangle \right\}$. $\mathcal{A}_{\mathsf{GL.machine}}$ is shown formally in Figure 1.

Setup: Let $m, n, q \in \mathbb{N}$ be integers such that q is prime, let $\delta > 0$, and let $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ be a function, to which our algorithm has oracle access. Let $\mathcal{A}_{\mathsf{GL}}$ be the one-dimensional Goldreich-Levin algorithm of Claim 2.

Input: A set $S \subset \mathbb{Z}_q^m$.

Execution:

- Draw $\mathbf{z} \sim \mathbb{Z}_q^n$ uniformly.
- Compute $\mathbf{u} \sim \mathcal{A}_{\mathsf{GL}}^{(\delta, f_{\mathbf{z}})}$ where $f_{\mathbf{z}} : \mathbb{Z}_q^m \to \mathbb{Z}_q$ is the function

$$f_{\mathbf{z}}(\mathbf{x}) = \begin{cases} \langle \mathbf{z}, f(\mathbf{x}) \rangle, & \mathbf{x} \in S \\ \$ \sim \mathbb{Z}_q, & \mathbf{x} \notin S \end{cases} .$$

  If $\mathcal{A}_{\mathsf{GL}}$ fails to give output, abort and halt.

Output: Output $S' = S \cap \left\{ \mathbf{x} \in \mathbb{Z}_q^m : \langle \mathbf{z}, f(\mathbf{x}) \rangle = \langle \mathbf{u}, \mathbf{x} \rangle \right\}$.

Figure 1: The Goldreich-Levin Machine $\mathcal{A}_{\mathsf{GL.machine}}$

The Hope for Repeatedly Applying the GL Machine. Suppose that f has good agreement with some $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$. We say that $S \subset \mathbb{Z}_q^m$ *contains f's agreement with* $\mathbf{A}$ if $\left\{ \mathbf{x} \in \mathbb{Z}_q^m : f(\mathbf{x}) = \mathbf{A}\mathbf{x} \right\} \subset S$. Clearly when $S = \mathbb{Z}_q^m$ then S trivially contains f's agreement with $\mathbf{A}$. Our hope is that if $\mathcal{A}_{\mathsf{GL.machine}}$ is run repeatedly, beginning with $S = \mathbb{Z}_q^m$, and $S = S'$ is updated after each run, then S will continue to contain f's agreement with $\mathbf{A}$ and moreover, that S will shrink to the point where equality holds; *i.e.*, $S = \left\{ \mathbf{x} \in \mathbb{Z}_q^m : f(\mathbf{x}) = \mathbf{A}\mathbf{x} \right\}$. Notice that if S' is updated as $S' = S \cap \left\{ \mathbf{x} \in \mathbb{Z}_q^m : \langle \mathbf{z}, f(\mathbf{x}) \rangle = \langle \mathbf{x}, \mathbf{u} \rangle \right\}$ for vectors $(\mathbf{u}, \mathbf{z}) \in \mathbb{Z}_q^m \times \mathbb{Z}_q^n$ which satisfy $\mathbf{u} = \mathbf{A}^{\mathsf{t}}\mathbf{z}$, then S' contains f's agreement with $\mathbf{A}$ whenever S does. In general, it is too optimistic to hope to obtain $S = \left\{ \mathbf{x} \in \mathbb{Z}_q^m : f(\mathbf{x}) = \mathbf{A}\mathbf{x} \right\}$. However, our analysis tracks the probability $\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\left[f(\mathbf{x}) = \mathbf{A}\mathbf{x} \middle| \mathbf{x} \in S \right]$ in order to measure how much larger the set S is than $\left\{ \mathbf{x} \in \mathbb{Z}_q^m : f(\mathbf{x}) = \mathbf{A}\mathbf{x} \right\}$. Lemma 3 below specifies conditions under which this probability has good chances of increasing if one round of $\mathcal{A}_{\mathsf{GL.machine}}$ is executed and $S = S'$ is updated. We first set some extra notation and terminology.

Notation. Our main algorithm uses the Goldreich-Levin machine to make progress in a variety of circumstances and we need some extra notation in order to make Lemma 3 applicable to all necessary settings. In addition to the usual integers $m, n, q \in \mathbb{N}$ with q prime and function $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$, let $r \in \mathbb{N}$ be an integer and $\mathbf{W} \subset \mathbb{Z}_q^n$ a subset of dimension $\dim(\mathbf{W}) = r$. For $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$, $S \subset \mathbb{Z}_q^m$ and $\mathbf{z} \in \mathbf{W}^\perp$, let

- $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} := \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) \in \mathbf{A}\mathbf{x} + \mathbf{W} \big| \mathbf{x} \in S\big]$;
- $\mathsf{Q}_{(\mathbf{A},S)}(\mathbf{z}) := \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[\langle \mathbf{z}, f(\mathbf{x})\rangle = \langle \mathbf{z}, \mathbf{A}\mathbf{x}\rangle \big| \mathbf{x} \in S\big]$.

As mentioned above, we say *S contains f's agreement with $\mathbf{A}$* if $\big\{\mathbf{x} \in \mathbb{Z}_q^m : f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big\} \subset S$ holds.

Lemma 3. *Let $m, n, q, r \in \mathbb{N}$ be integers with q prime, let $\sigma, \delta, \eta, \Delta > 0$, let $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ be a function, $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ a matrix, and $\mathbf{W} \subset \mathbb{Z}_q^n$ a subspace of dimension $\dim(\mathbf{W}) = r$. Suppose $\mathcal{A}_{\mathsf{GL.machine}}$ is executed with parameter δ and oracle access to f with an input set $S \subset \mathbb{Z}_q^m$ of size $|S| = \sigma q^m$ which contains f's agreement with $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$. Suppose, furthermore, that the following "Goldreich-Levin Progress Condition" (GLPC) holds:*

- **Goldreich-Levin Progress Condition:** $\Pr_{\mathbf{z} \sim \mathbf{W}^\perp}\Big[\frac{1}{q} + \frac{\delta}{\sigma} \le \mathsf{Q}_{(\mathbf{A},S)}(\mathbf{z}) \le \Delta\Big] \ge \eta$.

Then with probability at least $q^{-r}\eta\delta^2$ over $S' \sim \mathcal{A}^{(\delta,f)}_{\mathsf{GL.machine}}$, S' contains f's agreement with $\mathbf{A}$, and moreover $\mathsf{P}_{(\mathbf{A},\mathbf{W},S')} \ge \mathsf{P}_{(\mathbf{A},\mathbf{W},S)}/\Delta$ holds.

Remark. Intuitively, the GLPC ensures that with good probability over $\mathbf{z} \sim \mathbf{W}^\perp$ both of the following occur: 1) $\frac{1}{q} + \frac{\delta}{\sigma} \le \mathsf{Q}_{(\mathbf{A},S)}(\mathbf{z})$, ensuring that the call to $\mathcal{A}^{\delta,f_\mathbf{z}}_{\mathsf{GL}}$ outputs $\mathbf{u} = \mathbf{A}^\mathsf{t}\mathbf{z}$ with good probability; and 2) $\mathsf{Q}_{(\mathbf{A},S)}(\mathbf{z}) \le \Delta$ ensuring that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S')} \ge \mathsf{P}_{(\mathbf{A},\mathbf{W},S)}/\Delta$ holds with good probability since $\mathsf{P}_{(\mathbf{A},\mathbf{W},S')} = \frac{\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}}{\mathsf{Q}_{(\mathbf{A},S)}(\mathbf{z})}$ is likely.

Proof. Fix a matrix $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ and a subspace $\mathbf{W} \subset \mathbb{Z}_q^n$ of dimension $\dim(\mathbf{W}) = r$ and consider the execution of $S' \sim \mathcal{A}^{(\delta,f)}_{\mathsf{GL.machine}}(S)$ for a set $S \subset \mathbb{Z}_q^m$ of size $|S| = \sigma q^m$ which contains f's agreement with $\mathbf{A}$. Furthermore assume that the GLPC holds. With these choices fixed, let us write P, $\mathsf{Q}(\mathbf{z})$ and P' instead of $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}$, $\mathsf{Q}_{(\mathbf{A},S)}(\mathbf{z})$ and $\mathsf{P}_{(\mathbf{A},\mathbf{W},S')}$. Note that the execution of $\mathcal{A}_{\mathsf{GL.machine}}$ consists of three steps: first a random $\mathbf{z} \sim \mathbb{Z}_q^n$ is chosen, then $\mathbf{u} \sim \mathcal{A}^{(\delta,f_\mathbf{z})}_{\mathsf{GL}}$ is computed, finally $S' \subset \mathbb{Z}_q^m$ is prepared and output. Consider the following three events:

$$(1)\ \mathbf{z} \in \mathbf{W}^\perp; \quad (2)\ \tfrac{1}{q} + \tfrac{\delta}{\sigma} \le \mathsf{Q}(\mathbf{z}) \le \Delta; \quad (3)\ \mathbf{u} = \mathbf{A}^\mathsf{t}\mathbf{z}.$$

Since $\dim(\mathbf{W}) = r$, the probability of (1) is q^{-r}. Moreover, conditioned on (1) holding, the probability that (2) holds is at least η since the GLPC holds. Now, note that when (1) and (2) both hold we have

$$\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f_\mathbf{z}(\mathbf{x}) = \langle \mathbf{A}^\mathsf{t}\mathbf{z}, \mathbf{x}\rangle\big] \ge \sigma \cdot \mathsf{Q}(\mathbf{z}) + \big(1 - \sigma\big) \cdot \frac{1}{q} \ge \frac{1}{q} + \delta.$$

Therefore by Claim 2, conditioned on (1) and (2) both holding, (3) also holds with probability at least δ^2. So in summary, all three events hold with probability at least $q^{-r}\eta\delta^2$. Now, as noted above, (3) holding means that S' contains f's agreement with $\mathbf{A}$, since S does. Additionally, all three events holding means that $\mathsf{P}' = \mathsf{P}/\mathsf{Q}(\mathbf{z}) \ge \mathsf{P}/\Delta$. This completes the proof. □

4.1 An $\varepsilon^{-\mathcal{O}(n)}-$time Algorithm for Matrix Recovery

Simply by running the Goldreich-Levin machine from the previous section n times, we can already get an $\varepsilon^{-\mathcal{O}(n)}-$time algorithm for matrix recovery. Specifically, the lemma below describes an algorithm which recovers a matrix which has good agreement with f with probability $\varepsilon^{\mathcal{O}(n)}$; this can be used to recover a list which explains all of the linear agreement of f in $\varepsilon^{-\mathcal{O}(n)}$ time via the "agreement decoding implies list decoding" argument of Section 3.1. Our main algorithm runs much faster (time $\varepsilon^{-\mathcal{O}(\log_q(1/\varepsilon))}$), but will make use of the following $\varepsilon^{-\mathcal{O}(n)}-$time algorithm for small values of n.

Lemma 4. *Let $m, n, q \in \mathbb{N}$ be integers with q prime, let $\varepsilon > 0$ and let $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ be a function. There exists a randomized $\mathsf{poly}\big(n, m, \log q, \log(1/\varepsilon)\big)-$time algorithm $\mathcal{A}_{\mathsf{exp}}$ which takes no input, gets oracle access to f and outputs $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n\times m}$ such that with probability at least $\mathsf{poly}(\varepsilon^n)$,*

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_{\mathsf{out}}\mathbf{x}\big] \geq \max_{\mathbf{A}\in\mathbb{Z}_q^{n\times m}} \Big\{\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big]\Big\} - \varepsilon.$$

Remark. Lemma 4 is proved by repeatedly running the Goldreich-Levin machine (GLM), however for technical reasons having to do with how the output is prepared, we will need to be slightly more explicit about how we handle the sets which are the inputs and outputs of the GLM. Specifically, each execution of the GLM, say with input $S \subset \mathbb{Z}_q^m$, outputs $S' = S \cap \big\{\mathbf{x} \in \mathbb{Z}_q^m : \langle \mathbf{z}, f(\mathbf{x})\rangle = \langle \mathbf{u}, \mathbf{x}\rangle\big\}$ where $(\mathbf{u}, \mathbf{z}) \in \mathbb{Z}_q^m \times \mathbb{Z}_q^n$ are the vectors encountered during the computation of the GLM. In this section it will be necessary to keep track of the pairs $(\mathbf{u}, \mathbf{z})$ which occur in the various executions of the GLM. We do this by maintaining a set $\mathsf{L} \subset \mathbb{Z}_q^m \times \mathbb{Z}_q^n$, initialized to $\mathsf{L} = \emptyset$, so that at all times during the algorithm, $S = \big\{\mathbf{x} \in \mathbb{Z}_q^m : \langle \mathbf{z}, f(\mathbf{x})\rangle = \langle \mathbf{u}, \mathbf{x}\rangle \ \forall\ (\mathbf{u}, \mathbf{z}) \in \mathsf{L}\big\}$ holds. In this section only, we update the syntax of the GLM having it take the pair (L, S) as input and giving the pair (L', S') as output where S' is as usual and where $\mathsf{L}' = \mathsf{L} \cup \{(\mathbf{u}, \mathbf{z})\}$. Additionally, we extend the conclusion of the lemma to account for our new sets L and L'. Specifically, rather than concluding only that "S' contains f's agreement with $\mathbf{A}$", we add that L' is updated from L by adding a pair $(\mathbf{u}, \mathbf{z}) \in \mathbb{Z}_q^m \times \mathbb{Z}_q^n$ with $\mathbf{u} = \mathbf{A}^{\mathsf{t}}\mathbf{z}$.

Proof. Given $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$, let $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ be such that $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big]$ is maximal; we can assume this probability is at least ε since otherwise any output matrix $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n\times m}$ will satisfy the requirement. Also, assume $n \geq 2$ since if $n = 1$, then the result follows from Claim 2. The basic idea is to recover the rows of $\mathbf{A}$ one-by-one using the Goldreich-Levin machine (GLM). The algorithm $\mathcal{A}_{\mathsf{exp}}$ is shown in Figure 2, and it works as follows. A set pair (L, S) is initialized to $\mathsf{L} = \emptyset$ and $S = \mathbb{Z}_q^m$, a vector space $\mathbf{V} \subset \mathbb{Z}_q^n$ is initialized to $\mathbf{V} = \mathbb{Z}_q^n$ and a function $g : \mathbb{Z}_q^m \to \mathbf{V}$ is initialized to $g = f$. The following invariants will be maintained:

$$S = \big\{\mathbf{x} \in \mathbb{Z}_q^m : \langle \mathbf{z}, f(\mathbf{x})\rangle = \langle \mathbf{u}, \mathbf{x}\rangle \ \forall\ (\mathbf{u}, \mathbf{z}) \in \mathsf{L}\big\};\ \text{and}\ \mathbf{V} = \mathrm{Span}\big(\{\mathbf{z} \in \mathbb{Z}_q^n : (\mathbf{u}, \mathbf{z}) \in \mathsf{L}\}\big)^{\perp}.$$

Then the GLM is run k times for a randomly chosen $k \sim \{0, \ldots, n\}$ with parameter $\delta = \frac{\varepsilon^2}{16}$ on the function g. Each time the GLM is run, a set pair (L', S') is obtained and the updates

$$\mathsf{L} = \mathsf{L}';\ S' = S;\ \mathbf{V} = \mathrm{Span}\big(\{\mathbf{z} : (\mathbf{u}, \mathbf{z}) \in \mathsf{L}\}\big)^{\perp};\ g = \Pi_{\mathbf{V}} \circ f$$

are registered, where $\Pi_{\mathbf{V}} : \mathbb{Z}_q^n \to \mathbf{V}$ is the linear projection map onto $\mathbf{V}$. After the k executions of the GLM, a random matrix $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n\times m}$ which satisfies $\mathbf{A}_{\mathsf{out}}^{\mathsf{t}}\mathbf{z} = \mathbf{u}$ for all $(\mathbf{u}, \mathbf{z}) \in \mathsf{L}$ is output. For $i = 1, \ldots, k$, let $(\mathsf{L}_i, S_i, \mathbf{V}_i, g_i)$ denote the values of $(\mathsf{L}, S, \mathbf{V}, g)$ after the $i-$th execution of the GLM is

Algorithm Setup: Let $m, n, q \in \mathbb{N}$ be integers such that $n \geq 2$ and q is prime, let $\varepsilon > 0$, and let $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ be a function, to which our algorithm has oracle access. Let $\mathcal{A}_{\mathsf{GL.machine}}$ be the Goldreich-Levin machine algorithm from Section 4, with syntax modified according to the above remark.

1. Initialize: Let $\mathsf{L} \subset \mathbb{Z}_q^m \times \mathbb{Z}_q^n$, $S \subset \mathbb{Z}_q^m$, subspace $\mathbf{V} \subset \mathbb{Z}_q^n$ and $g : \mathbb{Z}_q^m \to \mathbf{V}$ be set to:

$$\mathsf{L} = \emptyset;\ S = \mathbb{Z}_q^m;\ \mathbf{V} = \mathbb{Z}_q^n;\ g = f.$$

2. The Main Loop: Draw $k \sim \{0, \ldots, n\}$, and do the following k times.

- Compute $(\mathsf{L}', S') \sim \mathcal{A}_{\mathsf{GL.machine}}^{(\delta, g)}(\mathsf{L}, S)$, where $\delta = \frac{\varepsilon^2}{16}$.
- Update $(\mathsf{L}, S) = (\mathsf{L}', S')$, and $\mathbf{V} = \mathrm{Span}\big(\{\mathbf{z} \in \mathbb{Z}_q^n : (\mathbf{u}, \mathbf{z}) \in \mathsf{L}\}\big)^{\perp}$.
- Update $g = \Pi_{\mathbf{V}} \circ f$ where $\Pi_{\mathbf{V}}$ is the projection map $\Pi_{\mathbf{V}} : \mathbb{Z}_q^n \to \mathbf{V}$.

3. Output: Draw and output a uniform $\mathbf{A}_{\mathsf{out}} \sim \mathbb{Z}_q^{n \times m}$ such that $\mathbf{A}_{\mathsf{out}}^{\mathsf{t}} \mathbf{z} = \mathbf{u}\ \forall\ (\mathbf{u}, \mathbf{z}) \in \mathsf{L}$.

Figure 2: A Matrix Recovery Algorithm with Exponentially Small Probability of Success $\mathcal{A}_{\mathsf{exp}}$

run, and let $(\mathsf{L}_0, S_i, \mathbf{V}_0, g_0) = (\emptyset, \mathbb{Z}_q^m, \mathbb{Z}_q^n, f)$. Also, let $\mathsf{P}_i(\mathbf{A}) := \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x} \big| \mathbf{x} \in S_i\big]$, so that $\mathsf{P}_0(\mathbf{A}) \geq \varepsilon$.

One subtle point is that since the GLM is run on the functions $g_i : \mathbb{Z}_q^m \to \mathbf{V}_i$ and $\dim(\mathbf{V}_i) = n - i$ (since each execution of the GLM adds a single pair to L), the first step of the GLM chooses a random $\mathbf{z} \sim \mathbb{Z}_q^{n-i}$, not $\mathbf{z} \sim \mathbb{Z}_q^n$. Technically, what is happening here is that a random $\mathbf{z}$ is being selected from the dual space of $\mathbf{V}_i$, which is $\mathbb{Z}_q^n / \mathrm{Span}\big(\{\mathbf{z} : (\mathbf{u}, \mathbf{z}) \in \mathsf{L}_i\}\big)$ since $\mathbf{V}_i = \mathrm{Span}\big(\{\mathbf{z} : (\mathbf{u}, \mathbf{z}) \in \mathsf{L}_i\}\big)^{\perp}$. Then, after the GLM recovers $\mathbf{u}$ and goes to add $(\mathbf{u}, \mathbf{z})$ to L, in order to make sure that L remains a subset of $\mathbb{Z}_q^m \times \mathbb{Z}_q^n$, the vector $\mathbf{z}' \in \mathbb{Z}_q^n$ is used instead of $\mathbf{z}$ where $\langle \mathbf{z}', \mathbf{v} \rangle = \langle \mathbf{z}, \mathbf{v} \rangle$ for all $\mathbf{v} \in \mathbf{V}_i$ and such that $\langle \mathbf{z}', \hat{\mathbf{z}} \rangle = 0$ for all $(\hat{\mathbf{u}}, \hat{\mathbf{z}}) \in \mathsf{L}_i$.

Now, let us assume that $\mathcal{A}_{\mathsf{exp}}$ got lucky with its choice of k and that k is maximal in $\{0, 1, \ldots, n\}$ such that $\mathsf{P}_i(\mathbf{A}) \geq q^{-(n-i)} + \varepsilon/2$ holds for all $i < k$ (occurs with probability $\frac{1}{n+1}$). As we explain below, Lemma 3 ensures that for each of the first k executions of the GLM, with probability at least $\varepsilon^5/2^{12}$, the set L is updated by adding a pair $(\mathbf{u}, \mathbf{z})$ with $\mathbf{u} = \mathbf{A}^{\mathsf{t}}\mathbf{z}$. In particular, this means that with probability at least $\varepsilon^{5n}/2^{12n} = \mathsf{poly}(\varepsilon^n)$, L_k agrees with $\mathbf{A}$ in the sense that $\mathbf{u} = \mathbf{A}^{\mathsf{t}}\mathbf{z}$ holds for all $(\mathbf{u}, \mathbf{z}) \in \mathsf{L}_k$. Now, the expected value of $\mathsf{P}_k(\hat{\mathbf{A}})$ over all matrices $\hat{\mathbf{A}} \in \mathbb{Z}_q^{n \times m}$ which agree with L_k is $q^{-(n-k)}$, and so with probability at least $\varepsilon/2$ over $\mathbf{A}_{\mathsf{out}}$, $\mathsf{P}_k(\mathbf{A}_{\mathsf{out}}) \geq q^{-(n-k)} - \varepsilon/2$ holds, in which case

$$\begin{aligned}\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_{\mathsf{out}}\mathbf{x}\big] &= \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[\mathbf{x} \in S_k\big] \cdot \mathsf{P}_k(\mathbf{A}_{\mathsf{out}}) \geq \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[\mathbf{x} \in S_k\big] \cdot \big(q^{-(n-k)} - \varepsilon/2\big) \\ &> \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[\mathbf{x} \in S_k\big] \cdot \big(\mathsf{P}_k(\mathbf{A}) - \varepsilon\big) \geq \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big] - \varepsilon,\end{aligned}$$

as desired. We have used that $\mathsf{P}_k(\mathbf{A}) < q^{-(n-k)} + \varepsilon/2$. Thus it remains to explain how Lemma 3 ensures that in each of the first k executions of the GLM, the set L is updated with a pair of the form $(\mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{z})$ with good probability.

So fix some $i < k$, so that $\mathsf{P}_i(\mathbf{A}) \geq q^{-(n-i)} + \varepsilon/2$, assume that $\mathbf{A}$ agrees with L_i and consider the

$(i+1)-$th execution of the GLM. Let $\mathsf{Q}_i(\mathbf{z}) := \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[\langle \mathbf{z}, f(\mathbf{x})\rangle = \langle \mathbf{z}, \mathbf{A}\mathbf{x}\rangle \big| \mathbf{x} \in S_i\big]$. Note that

$$\mathbb{E}_{\mathbf{z}\sim\mathbb{Z}_q^{n-i}}\big[\mathsf{Q}_i(\mathbf{z})\big] = \frac{1}{q} + \left(1 - \frac{1}{q}\right)\cdot \mathsf{P}_i(\mathbf{A}),$$

where by $\mathbf{z} \sim \mathbb{Z}_q^{n-i}$, we really mean that $\mathbf{z}$ is being drawn uniformly from the vector space quotient $\mathbb{Z}_q^n/\mathrm{Span}\big(\{\hat{\mathbf{z}} : (\hat{\mathbf{u}}, \hat{\mathbf{z}}) \in \mathsf{L}_i\}\big)$. We have used $\mathbb{E}_{\mathbf{z}\sim\mathbb{Z}_q^n}\big[\mathsf{Q}_i(\mathbf{z})\big] = \mathbb{E}_{\mathbf{z}\sim\mathbb{Z}_q^{n-i}}\big[\mathsf{Q}_i(\mathbf{z})\big]$, which holds because $\mathsf{Q}_i(\mathbf{z}+\mathbf{z}') = \mathsf{Q}_i(\mathbf{z})$ for all $\mathbf{z} \in \mathbb{Z}_q^n$ and $\mathbf{z}' \in \mathrm{Span}\big(\{\hat{\mathbf{z}} : (\hat{\mathbf{u}}, \hat{\mathbf{z}}) \in \mathsf{L}\}\big)$. Plugging in $\mathsf{P}_i(\mathbf{A}) \geq q^{-(n-i)} + \varepsilon/2$ and $\mathsf{Q}_i(\mathbf{0}) = 1$ gives

$$q^{-(n-i)} + \big(1 - q^{-(n-i)}\big)\cdot \mathbb{E}_{\mathbf{z}\sim\mathbb{Z}_q^{n-i}\setminus\{\mathbf{0}\}}\big[\mathsf{Q}_i(\mathbf{z})\big] \geq q^{-(n-i)} + \frac{1}{q}\cdot\big(1 - q^{-(n-i)}\big) + \frac{\varepsilon}{4},$$

which simplifies to $\mathbb{E}_{\mathbf{z}\sim\mathbb{Z}_q^{n-i}\setminus\{\mathbf{0}\}}\big[\mathsf{Q}_i(\mathbf{z})\big] \geq 1/q + \varepsilon/4$. It follows that

$$\Pr_{\mathbf{z}\sim\mathbb{Z}_q^{n-i}}\left[\mathsf{Q}_i(\mathbf{z}) \geq \frac{1}{q} + \frac{\varepsilon}{8}\right] \geq \big(1 - q^{-(n-i)}\big)\cdot \Pr_{\mathbf{z}\sim\mathbb{Z}_q^{n-i}\setminus\{\mathbf{0}\}}\left[\mathsf{Q}_i(\mathbf{z}) \geq \frac{1}{q} + \frac{\varepsilon}{8}\right] \geq \varepsilon/16.$$

Thus the Goldreich-Levin Progress Condition holds with $\mathbf{W} = \{\mathbf{0}\}$, $\Delta = 1$, $\eta = \varepsilon/16$ and $\frac{\delta}{\sigma} = \frac{\varepsilon}{8}$. Also, note $\sigma \geq \varepsilon/2$ because $\sigma = \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[\mathbf{x} \in S_i\big] \geq \mathsf{P}_0(\mathbf{A}) \geq q^{-n} + \varepsilon/2$. Therefore, by Lemma 3, the $(i+1)-$th execution of the GLM updates L_i to L_{i+1} by adding a pair of the form $(\mathbf{A}^{\mathsf{t}}\mathbf{z}, \mathbf{z})$ with probability at least $\eta\delta^2 \geq \varepsilon^5/2^{12}$, as needed. □

Our main algorithm will use the algorithm promised by this corollary as a subroutine.

Corollary 1. *Let $m, n, q \in \mathbb{N}$ be integers with $m \geq 18$, q prime, and let $\varepsilon > 0$ be such that $\varepsilon \geq 4q^{-m/3}$ holds. Let $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ be a function, and let $T \subset \mathbb{Z}_q^m$ be a subset. There exists a randomized $\mathsf{poly}\big(n, m, \log(q), \log(1/\varepsilon)\big)-$time algorithm which takes no input, gets oracle access to f and to the indicator function $\mathbb{1}_T$, and outputs $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n\times m}$ such that with probability at least $\mathsf{poly}(\varepsilon^n)$,*

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_{\mathsf{out}}\mathbf{x} \,\&\, \mathbf{x} \in T\big] \geq \max_{\mathbf{A}\in\mathbb{Z}_q^{n\times m}}\Big\{\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x} \,\&\, \mathbf{x} \in T\big]\Big\} - \varepsilon.$$

Proof. Assume $n \geq 2$, since if $n = 1$, the corollary follows from Claim 2. Invoke the algorithm $\mathcal{A}_{\mathsf{exp}}$ from Lemma 4 with parameter $\varepsilon/2$ on the function $g : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ defined by

$$g(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \mathbf{x} \in T \\ \$ \sim \mathbb{Z}_q^n, & \mathbf{x} \notin T \end{cases},$$

and let $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n\times m}$ be the output. Note that for any $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$,

$$\begin{aligned}\mathsf{P}_{f,T}(\mathbf{A}) := \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x} \,\&\, \mathbf{x} \in T\big] &= \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[g(\mathbf{x}) = \mathbf{A}\mathbf{x} \,\&\, \mathbf{x} \in T\big] \\ &= \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[g(\mathbf{x}) = \mathbf{A}\mathbf{x}\big] - \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[g(\mathbf{x}) = \mathbf{A}\mathbf{x} \,\&\, \mathbf{x} \notin T\big] \\ &=: \mathsf{P}_g(\mathbf{A}) - \mathsf{P}_{g,\overline{T}}(\mathbf{A}).\end{aligned}$$

With probability at least $\mathsf{poly}(\varepsilon^n)$,

$$\mathsf{P}_{f,T}(\mathbf{A}) - \mathsf{P}_{f,T}(\mathbf{A}_{\mathsf{out}}) \leq \Big(\mathsf{P}_g(\mathbf{A}) - \mathsf{P}_g(\mathbf{A}_{\mathsf{out}})\Big) + \Big|\mathsf{P}_{g,\overline{T}}(\mathbf{A}) - \mathsf{P}_{g,\overline{T}}(\mathbf{A}_{\mathsf{out}})\Big| \leq \varepsilon$$

holds, since $\mathsf{P}_g(\mathbf{A}) - \mathsf{P}_g(\mathbf{A}_{\mathsf{out}}) \leq \varepsilon/2$ holds with probability $\mathsf{poly}(\varepsilon^n)$ over the random coins of $\mathcal{A}_{\mathsf{exp}}$ by Lemma 4, and since $\big|\mathsf{P}_{g,\overline{T}}(\mathbf{A}) - \overline{\tau}q^{-n}\big| \leq \varepsilon/4$ and $\big|\mathsf{P}_{g,\overline{T}}(\mathbf{A}_{\mathsf{out}}) - \overline{\tau}q^{-n}\big| \leq \varepsilon/4$ both hold with high probability over the randomness of g by Claim 3, where $\overline{\tau} = \big|\mathbb{Z}_q^m \setminus T\big|/q^m$. The result follows. □

5 The Amplification Step

Lemma 1 (Restated). *Let $m, n, q \in \mathbb{N}$ be integers with q prime, let $\varepsilon > 0$, and let $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ be a function. There exists a randomized algorithm $\mathcal{A}_{\mathsf{amplify}}$ which has the following syntax, running time and correctness guarantees.*

- **Syntax:** *$\mathcal{A}_{\mathsf{amplify}}$ takes no input, gets oracle access to f, and outputs the characteristic function of a set $S_{\mathsf{out}} \subset \mathbb{Z}_q^m$.*
- **Running Time:** *$\mathcal{A}_{\mathsf{amplify}}$ runs in expected $\mathsf{poly}\big(m, n, \log q, \log(1/\varepsilon)\big)$ time; the characteristic function which $\mathcal{A}_{\mathsf{amplify}}$ outputs can be computed in time $\mathsf{poly}\big(m, n, \log q\big)$ using one oracle call to f.*
- **Correctness:** *If $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ is such that $\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big] \geq \varepsilon$, then with probability at least $\mathsf{poly}\big(\varepsilon^{\log_q(1/\varepsilon)}, m^{-\log_q(1/\varepsilon)}\big)$ (over the random coins of $\mathcal{A}_{\mathsf{amplify}}$), S_{out} contains f's agreement with $\mathbf{A}$, and moreover, there exists a subspace $\mathbf{W} \subset \mathbb{Z}_q^n$ of dimension at most $\dim(\mathbf{W}) \leq r$ for a parameter $r = \mathcal{O}\big(\log_q(1/\varepsilon)\big)$ such that*

$$\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) \in \mathbf{A}\mathbf{x} + \mathbf{W} \big| \mathbf{x} \in S_{\mathsf{out}}\big] \geq 1 - \frac{1}{4m}.$$

Proof. The algorithm initializes $S = \mathbb{Z}_q^m$, chooses random $k_1, k_2 \sim \{0, 1, \ldots, T\}$ for a sufficiently large integer $T \in \mathbb{N}$ (specified later), and does the following for $i = 1, \ldots, k_1 + k_2$:

- Compute $S' \sim \mathcal{A}_{\mathsf{GL.machine}}^{(\delta_i, f)}(S)$ and update $S' = S$, where $\delta_i = \begin{cases} \varepsilon^2/1000, & i \leq k_1 \\ \varepsilon \cdot \min\{.01, q^{-2}\}, & i > k_1 \end{cases}$.

After running the GLM $k_1 + k_2$ times, the set $S_{\mathsf{out}} = S$ is output. It is clear that $\mathcal{A}_{\mathsf{amplify}}$ satisfies the required syntax and running time guarantees. Towards proving correctness, assume $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ is such that $\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big] \geq \varepsilon$. Note that at the start of the algorithm, S trivially contains f's agreement with $\mathbf{A}$ since $S = \mathbb{Z}_q^m$ is initialized. The proof of correctness works by maintaining a subspace $\mathbf{W} \subset \mathbb{Z}_q^n$ and keeping track of the probability

$$\mathsf{P}_{(\mathbf{A}, \mathbf{W}, S)} := \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) \in \mathbf{A}\mathbf{x} + \mathbf{W} \big| \mathbf{x} \in S\big]$$

throughout the execution of the algorithm. We initialize $\mathbf{W}$ to $\mathbf{W} = \{\mathbf{0}\}$ so that $\mathsf{P}_{(\mathbf{A}, \mathbf{W}, S)} \geq \varepsilon$ holds at the start of the algorithm. Our hope is that by repeatedly applying the GLM, we will gradually increase $\mathsf{P}_{(\mathbf{A}, \mathbf{W}, S)}$ from ε all the way to $1 - \frac{1}{4m}$ while maintaining that S contains f's agreement with $\mathbf{A}$. Crucial to this plan is Lemma 3, which promises that if S contains f's agreement with $\mathbf{A}$ and additionally if the Goldreich-Levin Progress Condition holds for parameters $\delta, \Delta, \eta > 0$:

$$\Pr_{\mathbf{z} \sim \mathbf{W}^\perp}\left[\frac{1}{q} + \frac{\delta}{\varepsilon} \leq \mathsf{Q}_{(\mathbf{A}, S)}(\mathbf{z}) \leq \Delta\right] \geq \eta, \tag{GLPC}$$

where $\mathsf{Q}_{(\mathbf{A}, S)}(\mathbf{z}) := \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[\langle \mathbf{z}, f(\mathbf{x})\rangle = \langle \mathbf{z}, \mathbf{A}\mathbf{x}\rangle \big| \mathbf{x} \in S\big]$, then with probability $q^{-\dim(\mathbf{W})} \cdot \delta^2 \eta$ over $S' \sim \mathcal{A}_{\mathsf{GL.machine}}^{(\delta, f)}(S)$, S' contains f's agreement with $\mathbf{A}$, and additionally, $\mathsf{P}_{(\mathbf{A}, \mathbf{W}, S')} \geq \mathsf{P}_{(\mathbf{A}, \mathbf{W}, S)}/\Delta$.

With this high level plan in place, our proof splits into two parts which are handled somewhat differently. The goal of the first part of the proof is to amplify $\mathsf{P}_{(\mathbf{A}, \mathbf{W}, S)}$ from ε to $\frac{1}{q} + \gamma$ where $\gamma > 0$ is

$$\gamma = \begin{cases} .01, & q \in \{2, 3, 5, 6\} \\ \frac{1}{q^2}, & q \geq 11 \end{cases}.$$

The goal of the second part is to amplify $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}$ the rest of the way from $\frac{1}{q}+\gamma$ to $1-\frac{1}{4m}$. The key difference between the two settings is that when $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq \frac{1}{q}+\gamma$, (GLPC) is guaranteed to hold for some choice of (δ,Δ,η), whereas this is not true when $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} < \frac{1}{q}+\gamma$. The following claims summarize the two different parts of our proof.

Claim 4. *Let notations be as above, let $\Delta > 0$ be*

$$\Delta = \begin{cases} .99, & q \in \{2,3,5,7\} \\ q^{-1/4}, & q \geq 11 \end{cases},$$

and let $r \in \mathbb{N}$ be the minimal integer such that $\varepsilon/\Delta^r \geq \frac{1}{q}+\gamma$. Note that $r = \mathcal{O}\big(\log_q(1/\varepsilon)\big)$. Suppose $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ is such that $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big] \geq \varepsilon$. Then there exists $k_1 \in \{0,\ldots,r\}$ such that if the GLM is run k_1 times with parameter $\delta = \varepsilon^2/1000$, then with probability at least $\mathsf{poly}\big(\varepsilon^{\log_q(1/\varepsilon)}\big)$, the resulting $S \subset \mathbb{Z}_q^m$ contains f's agreement with $\mathbf{A}$ and moreover is such that there exists a subspace $\mathbf{W} \subset \mathbb{Z}_q^n$ of dimension at most $\dim(\mathbf{W}) \leq r$ such that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq \frac{1}{q}+\gamma$.

Claim 5. *Let notations be as above. Suppose $S \subset \mathbb{Z}_q^m$ contains f's agreement with $\mathbf{A}$ and also that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq \frac{1}{q}+\gamma$ holds for some subspace $\mathbf{W} \subset \mathbb{Z}_q^n$ of dimension at most r. Then there exists a $k_2 \in \{0,\ldots,2\log_q(4m)\}$ such that if the GLM is run repeatedly k_2 times with parameter $\delta = \gamma\varepsilon$, then with probability at least $\mathsf{poly}\big(m^{-\log_q(1/\varepsilon)}\big)$, the resulting S also contains f's agreement with $\mathbf{A}$ and is moreover such that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq 1-\frac{1}{4m}$.*

Notice that these claims combine to easily prove Lemma 1. Suppose $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ is such that $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big] \geq \varepsilon$ holds, and let $T = \max\{r, 2\log_q(4m)\}$. Then if $\mathcal{A}_{\mathsf{amplify}}$ chooses the k_1 specified by Claim 4 (this k_1 is chosen with probability at least $\frac{1}{T+1}$), then with probability at least $\mathsf{poly}\big(\varepsilon^{\log_q(1/\varepsilon)}\big)$, the $S \subset \mathbb{Z}_q^m$ obtained after running the GLM with parameter $\delta = \varepsilon^2/1000$, k_1 times will contain f's agreement with $\mathbf{A}$ and be such that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq \frac{1}{q}+\gamma$ for some subspace $\mathbf{W} \subset \mathbb{Z}_q^n$ of dimension at most r. If $\mathcal{A}_{\mathsf{amplify}}$ then chooses the k_2 specified by Claim 5 (occurs with probability at least $\frac{1}{T+1}$), then with probability at least $\mathsf{poly}\big(m^{-\log_q(1/\varepsilon)}\big)$ the $S \subset \mathbb{Z}_q^m$ obtained after running the GLM k_2 times with parameter $\delta = \gamma\varepsilon$ contains f's agreement with $\mathbf{A}$ and is such that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq 1-\frac{1}{4m}$. If we put everything together, we get that with probability at least $\mathsf{poly}\big(\varepsilon^{\log_q(1/\varepsilon)}, m^{-\log_q(1/\varepsilon)}\big)$, the set $S_{\mathsf{out}} \subset \mathbb{Z}_q^m$ output by $\mathcal{A}_{\mathsf{amplify}}$ contains f's agreement with $\mathbf{A}$ and is such that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq 1-\frac{1}{4m}$ for some subspace $\mathbf{W} \subset \mathbb{Z}_q^n$ of dimension at most r, as desired. □

5.1 Proof of Claim 4

Proof. Recall $r \in \mathbb{N}$ is the minimal integer such that $\varepsilon/\Delta^r \geq \frac{1}{q}+\gamma$ for the parameters $\gamma,\Delta > 0$ specified above. We amplify $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}$ from ε to $\frac{1}{q}+\gamma$ in at most r stages where in each stage we hope to increase $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}$ by a multiplicative factor of at least Δ^{-1}. We do this in one of two ways, depending on whether the following version of the Goldreich-Levin Progress Condition holds:

$$\Pr_{\mathbf{z}\sim\mathbf{W}^\perp}\left[\frac{1}{q}+\frac{\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}}{1000} \leq \mathsf{Q}_{(\mathbf{A},S)}(\mathbf{z}) \leq \Delta\right] \geq \frac{\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}}{1000}. \tag{GLPC}$$

Clearly, when (GLPC) holds, progress can be made with good probability by running $\mathcal{A}_{\mathsf{GL.machine}}$ with $\delta = \varepsilon^2/1000$, so that $\frac{\delta}{\varepsilon} = \frac{\varepsilon}{1000} \leq \frac{\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}}{1000}$. In this case, Lemma 3 says that if $S \subset \mathbb{Z}_q^m$ contains f's agreement with $\mathbf{A}$ then with probability at least $\frac{\varepsilon^3}{q^r\cdot 10^6}$ over $S' \sim \mathcal{A}^{(\delta,f)}_{\mathsf{GL.machine}}(S)$, S' also contains f's

agreement with $\mathbf{A}$ and is such that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S')} \geq \mathsf{P}_{(\mathbf{A},\mathbf{W},S)}/\Delta$. The crux of the proof lies in establishing the following Claim which allows us to also make progress when (GLPC) does not hold.

Claim 6. *Assume* $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} < \frac{1}{q} + \gamma$ *and that (GLPC) does not hold. Then there exists* $\mathbf{w} \in \mathbb{Z}_q^n \setminus \mathbf{W}$ *such that* $\mathsf{P}_{(\mathbf{A},\mathbf{W}',S)} \geq \mathsf{P}_{(\mathbf{A},\mathbf{W},S)}/\Delta$*, where* $\mathbf{W}' = \mathbf{W} + \mathrm{Span}(\mathbf{w})$.

In total, after at most r stages of increasing $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}$ by a multiplicative factor of Δ^{-1} (either by shrinking S using the GLM and invoking Lemma 3 or by increasing $\mathbf{W}$ via Claim 6), we will have that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq \varepsilon/\Delta^r \geq \frac{1}{q} + \gamma$ holds with probability $\left(\frac{\varepsilon^3}{q^r \cdot 10^6}\right)^{k_1} = \mathsf{poly}\left(\varepsilon^{\log_q(1/\varepsilon)}\right)$, where $k_1 \leq r$ is the number of stages during which (GLPC) held, and so progress was attempted by running $\mathcal{A}_{\mathsf{GL.machine}}$. Since Claim 6 was invoked at most r times, we also have $\dim(\mathbf{W}) \leq r$. This proves Claim 4, and so it remains only to prove Claim 6. We do this in Section 5.3. □

5.2 Proof of Claim 5

Proof. Assume $S \subset \mathbb{Z}_q^m$ contains f's agreement with $\mathbf{A}$ and is such that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq \frac{1}{q} + \gamma$ for some subspace $\mathbf{W} \subset \mathbb{Z}_q^n$ of dimension at most $\dim(\mathbf{W}) \leq r$. Recall that we say the Goldreich-Levin Progress Condition holds for parameters $\delta, \Delta, \eta > 0$ if

$$\Pr_{\mathbf{z} \sim \mathbf{W}^\perp}\left[\frac{1}{q} + \frac{\delta}{\varepsilon} \leq \mathsf{Q}_{(\mathbf{A},S)}(\mathbf{z}) \leq \Delta\right] \geq \eta, \tag{GLPC}$$

and that Lemma 3 ensures that if $S \subset \mathbb{Z}_q^m$ contains f's agreement with $\mathbf{A}$ and if (GLPC) holds, then with probability at least $q^{-r}\delta^2\eta$ over $S' \sim \mathcal{A}^{(\delta,f)}_{\mathsf{GL.machine}}(S)$, S' also contains f's agreement with $\mathbf{A}$ and $\mathsf{P}_{(\mathbf{A},\mathbf{W},S')} \geq \mathsf{P}_{(\mathbf{A},\mathbf{W},S)}/\Delta$. Amplifying $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}$ from $\frac{1}{q}+\gamma$ to $1-\frac{1}{4m}$ simply involves invoking Lemma 3 repeatedly for different parameter choices. The following claim specifies the parameters we will use for this part, asserting that (GLPC) holds for each of them. We prove Claim 7 below, after the current proof.

Claim 7. *Fix* $\delta = \gamma\varepsilon$*, and write* P *instead of* $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}$ *for shorthand. We have all of the following.*

1. *If* $\mathsf{P} \geq 1 - q^{-t}$ *for* $\left\{\begin{array}{ll}(1a) & t \geq 2, q \geq 2\\ (1b) & t \geq 1, q \geq 3\end{array}\right\}$*, then (GLPC) holds for* $(\Delta, \eta) = \left(\frac{\mathsf{P}}{1-q^{-(t+1/2)}}, .05\right)$.
2. *If* $\mathsf{P} \geq \frac{1}{q} + \gamma$*, then (GLPC) holds for* $(\Delta, \eta) = \left(2\mathsf{P}, \frac{1}{q} + \gamma\right)$.
3. *If* $\mathsf{P} \geq 1/2$*, then (GLPC) holds for* $(\Delta, \eta) = \left(\frac{\mathsf{P}}{1-q^{-1}}, \frac{1}{q+1}\right)$.
4. *If* $q = 2$ *and* $\left\{\begin{array}{ll}(4a) & .51 \leq \mathsf{P} \leq .6\\ (4b) & .6 \leq \mathsf{P} \leq .68\\ (4c) & .67 \leq \mathsf{P} \leq .73\\ (4d) & .73 \leq \mathsf{P} \leq .75\end{array}\right\}$ *then (GLPC) holds for* $(\Delta, \eta) = \left\{\begin{array}{ll}(4a) & (.85, .02)\\ (4b) & (.89, .04)\\ (4c) & (.9, .02)\\ (4d) & (.9, .1)\end{array}\right\}$.

Using Claim 7, we can amplify $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}$ from $\frac{1}{q} + \gamma$ to $1 - \frac{1}{4m}$ just by repeatedly running the GLM. For example, suppose S contains f's agreement with $\mathbf{A}$ and that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq 1 - q^{-2}$ holds. Then Point 1a of Claim 7 says that (GLPC) holds for the specified (δ, Δ, η). By Lemma 3, this means that with probability at least $\frac{\gamma^2\varepsilon^2}{q^r} \cdot (.05)$ over $S' \sim \mathcal{A}^{(\gamma\varepsilon,f)}_{\mathsf{GL.machine}}(S)$, S' still contains f's agreement with $\mathbf{A}$, and additionally $\mathsf{P}_{(\mathbf{A},\mathbf{W},S')} \geq 1 - q^{-2.5}$. Repeating this combination of Point 1a of Claim 7 and Lemma 3 at most $2\log_q(4m) - 4$ times will result in $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq 1 - \frac{1}{4m}$ with probability at least

$\left[\frac{\gamma^2\varepsilon^2}{q^r}\cdot(.05)\right]^{2\log_q(4m)-4}$. To get $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}$ from $\frac{1}{q}+\gamma$ to $1-q^{-2}$ we use the other points of Claim 7 in combination with Lemma 3. When $q\geq 3$, we use:

- point 2 to amplify from $\frac{1}{q}+\gamma$ to $\frac{1}{2}$ with probability at least $\frac{\gamma^2\varepsilon^2}{q^r}\cdot\left(\frac{1}{q}+\gamma\right)$;
- point 3 to amplify from $\frac{1}{2}$ to $1-q^{-1}$ with probability at least $\frac{\gamma^2\varepsilon^2}{q^r}\cdot\frac{1}{q+1}$;
- point 1b twice to amplify from $1-q^{-1}$ to $1-q^{-2}$ with probability at least $\frac{\gamma^4\varepsilon^4}{q^{2r}}\cdot(.05)^2$.

When $q=2$ this does not work because amplifying from $\frac{1}{q}+\gamma$ to $\frac{1}{2}$ to $1-q^{-1}$ is not making progress. Instead, when $q=2$ we use point 4 to amplify from $.51=\frac{1}{2}+\gamma$ to

$$\left(\frac{1}{2}+\gamma\right)\cdot\frac{1}{\prod_{i=1}^4\Delta_i}=\frac{.51}{(.85)(.89)(.9)(.9)}>.75=1-2^{-2},$$

with probability at least $\frac{\gamma^8\varepsilon^8}{q^{4r}}\cdot(.02)(.04)(.02)(.1)$. In total, by running the GLM at most $2\log_q(4m)$ times we will amplify $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}$ from $\frac{1}{q}+\gamma$ all the way to $1-\frac{1}{4m}$ with probability at least

$$\left[\frac{\gamma^2\varepsilon^2}{q^r}\cdot(.05)\right]^{2\log_q(4m)}\cdot\min\left\{\left(\frac{1}{q}+\gamma\right)\cdot\frac{1}{q+1}\cdot(.05)^2,(.02)(.04)(.02)(.1)\right\},$$

which is $\mathsf{poly}\big(m^{-\log_q(1/\varepsilon)},m^{-r}\big)=\mathsf{poly}\big(m^{-\log_q(1/\varepsilon)}\big)$. □

Proof of Claim 7. Write P and $\mathsf{Q}(\mathbf{z})$ instead of $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}$ and $\mathsf{Q}_{(\mathbf{A},S)}(\mathbf{z})$ for shorthand. When $\delta=\gamma\varepsilon$, the GLPC is:

$$\Pr_{\mathbf{z}\sim\mathbf{W}^\perp}\left[\frac{1}{q}+\gamma\leq\mathsf{Q}(\mathbf{z})\leq\Delta\right]\geq\eta. \tag{GLPC}$$

Since $\mathsf{Q}(\mathbf{z})\geq\mathsf{P}$ holds for all $\mathbf{z}\in\mathbf{W}^\perp$, if $\mathsf{P}\geq\frac{1}{q}+\gamma$, then the lower bound of the GLPC trivially holds, and so (GLPC) simplifies to $\Pr_{\mathbf{z}\sim\mathbf{W}^\perp}\big[\mathsf{Q}(\mathbf{z})\leq\Delta\big]\geq\eta$. Let $\mathsf{R}:=\Pr_{\mathbf{z}\sim\mathbf{W}^\perp}\big[\mathsf{Q}(\mathbf{z})\leq\Delta\big]$; we must show that $\mathsf{R}\geq\eta$ holds in all cases. We have $\mathsf{P}+(1-\mathsf{P})\cdot\frac{1}{q}=\mathbb{E}_{\mathbf{z}\sim\mathbf{W}^\perp}\big[\mathsf{Q}(\mathbf{z})\big]\geq\mathsf{R}\cdot\mathsf{P}+(1-\mathsf{R})\cdot\Delta$, which rearranges to

$$\mathsf{R}\geq 1-\frac{1-\mathsf{P}}{q(\Delta-\mathsf{P})}. \tag{$\dagger$}$$

For point 1, plug $\Delta-\mathsf{P}=\mathsf{P}\big(\frac{1}{1-q^{-(t+1/2)}}-1\big)\geq\mathsf{P}\cdot q^{-(t+1/2)}$ into ($\dagger$) and get $\mathsf{R}\geq 1-\frac{1}{\sqrt{q}}\cdot\frac{1}{1-q^{-t}}\geq.05$, which holds if either $t\geq 2$ and $q\geq 2$ or if $t\geq 1$ and $q\geq 3$. For point 2, plug $\Delta-\mathsf{P}=\mathsf{P}$ into ($\dagger$):

$$\mathsf{R}-\frac{1}{q}-\gamma\geq 1-\frac{1}{q}-\gamma-\frac{1-\mathsf{P}}{q\mathsf{P}}\geq\left(1-\frac{1}{q}-\gamma\right)\cdot\left(1-\frac{1}{1+q\gamma}\right)>0.$$

For point 3, we plug $q(\Delta-\mathsf{P})=q\mathsf{P}\big(\frac{1}{1-q^{-1}}-1\big)\geq\mathsf{P}\cdot\frac{q+1}{q}$ into ($\dagger$) to get $\mathsf{R}\geq 1-\frac{q(1-\mathsf{P})}{(q+1)\mathsf{P}}\geq\frac{1}{q+1}$. Finally, all of the sub-parts of part 4 are proved exactly the same way, namely by plugging in $q=2$, the lower bound for P on the top and the upper bound for P on the bottom. For example, 4a is established as follows, we omit the others as they are analogous:

$$\mathsf{R}\geq 1-\frac{1-\mathsf{P}}{q(\Delta-\mathsf{P})}\geq 1-\frac{1-.51}{2(.85-.6)}=1-\frac{.49}{.5}=.02.$$

□

5.3 Proof of Claim 6

In this section, $(\mathbf{A}, \mathbf{W}, S)$ will all be fixed, so we write P instead of $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)}$ to simplify notations. Recall also that for $\mathbf{z} \in \mathbb{Z}_q^n$, the quantity $\mathsf{Q}_{(\mathbf{A},S)}(\mathbf{z})$ was defined as $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[\langle \mathbf{z}, f(\mathbf{x})\rangle = \langle \mathbf{z}, \mathbf{A}\mathbf{x}\rangle \big| \mathbf{x} \in S\big]$. Claim 6 is proved using Fourier analysis on the function $\mathsf{Q} : \mathbb{Z}_q^n \to \mathbb{R}$,

$$\mathsf{Q}(\mathbf{z}) = \begin{cases} \mathsf{Q}_{(\mathbf{A},S)}(\mathbf{z}), & \mathbf{w} \in \mathbf{W}^\perp \\ 0, & \mathbf{w} \notin \mathbf{W}^\perp \end{cases}.$$

Claim 8 (Fourier Coefficients of Q). *For all $\mathbf{w} \in \mathbb{Z}_q^n$,*

$$q^{\dim(\mathbf{W})+1} \cdot \hat{\mathsf{Q}}(\mathbf{w}) = \begin{cases} \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) - \mathbf{A}\mathbf{x} \in \big(\mathbf{W} + \mathrm{Span}(\mathbf{w})\big) \setminus \mathbf{W} \big| \mathbf{x} \in S\big], & \mathbf{w} \notin \mathbf{W} \\ 1 + (q-1)\mathsf{P}, & \mathbf{w} \in \mathbf{W} \end{cases}$$

Proof. By definition, for all $\mathbf{w} \in \mathbb{Z}_q^n$,

$$q^{\dim(\mathbf{W})} \cdot \hat{\mathsf{Q}}(\mathbf{w}) = \mathbb{E}_{\mathbf{z}\sim\mathbf{W}^\perp}\Big[\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[\langle \mathbf{z}, f(\mathbf{x}) - \mathbf{A}\mathbf{x}\rangle = 0 \big| \mathbf{x} \in S\big] \cdot \omega^{-\langle \mathbf{z},\mathbf{w}\rangle}\Big],$$

since $\mathsf{Q}(\mathbf{z}) = 0$ for $\mathbf{z} \notin \mathbf{W}^\perp$. When $\mathbf{w} \in \mathbf{W}$, the quantity $\omega^{-\langle \mathbf{z},\mathbf{w}\rangle}$ vanishes from the right hand side giving $q^{\dim(\mathbf{W})} \cdot \hat{\mathsf{Q}}(\mathbf{w}) = \mathsf{P} + \frac{1}{q} \cdot (1 - \mathsf{P})$ in this case. When $\mathbf{w} \notin \mathbf{W}$, consider the quantity $\mathsf{val}(\mathbf{x}, \mathbf{w})$ for $\mathbf{x} \in \mathbb{Z}_q^m$ defined by

$$\mathsf{val}(\mathbf{x}, \mathbf{w}) := \mathbb{E}_{\mathbf{z}\sim\mathbf{W}^\perp}\Big[\mathbb{1}_{\mathbf{z}\perp f(\mathbf{x})-\mathbf{A}\mathbf{x}} \cdot \omega^{-\langle \mathbf{z},\mathbf{w}\rangle}\Big],$$

so that $q^{\dim(\mathbf{W})} \cdot \hat{\mathsf{Q}}(\mathbf{w}) = \mathbb{E}_{\mathbf{x}\sim S}\big[\mathsf{val}(\mathbf{x}, \mathbf{w})\big]$. Note that if $f(\mathbf{x}) \in \mathbf{A}\mathbf{x} + \mathbf{W}$ then $\mathbf{z} \perp f(\mathbf{x}) - \mathbf{A}\mathbf{x}$ trivially holds for all $\mathbf{z} \in \mathbf{W}^\perp$, in which case Claim 1 says that $\mathsf{val}(\mathbf{x}, \mathbf{w}) = \mathbb{E}_{\mathbf{z}\sim\mathbf{W}^\perp}\big[\omega^{-\langle \mathbf{z},\mathbf{w}\rangle}\big] = 0$. On the other hand, when $f(\mathbf{x}) \notin \mathbf{A}\mathbf{x} + \mathbf{W}$, Claim 1 gives

$$\mathsf{val}(\mathbf{x}, \mathbf{w}) = \frac{1}{q} \cdot \mathbb{E}_{\substack{\mathbf{z}\sim\mathbf{W}^\perp \\ \mathbf{z}\perp f(\mathbf{x})-\mathbf{A}\mathbf{x}}}\Big[\omega^{-\langle \mathbf{z},\mathbf{w}\rangle}\Big] = \begin{cases} 1/q, & \mathbf{w} \in \mathbf{W} + \mathrm{Span}\big(f(\mathbf{x}) - \mathbf{A}\mathbf{x}\big) \\ 0, & \mathbf{w} \notin \mathbf{W} + \mathrm{Span}\big(f(\mathbf{x}) - \mathbf{A}\mathbf{x}\big) \end{cases}$$

For $\mathbf{w} \notin \mathbf{W}$, this gives

$$\begin{aligned} q^{\dim(\mathbf{W})} \cdot \hat{\mathsf{Q}}(\mathbf{w}) &= \frac{1}{q} \cdot \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\Big[f(\mathbf{x}) \notin \mathbf{A}\mathbf{x} + \mathbf{W} \;\&\; \mathbf{w} \in \mathbf{W} + \mathrm{Span}\big(f(\mathbf{x}) - \mathbf{A}\mathbf{x}\big)\Big|\mathbf{x} \in S\Big] \\ &= \frac{1}{q} \cdot \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\Big[f(\mathbf{x}) - \mathbf{A}\mathbf{x} \in \big(\mathbf{W} + \mathrm{Span}(\mathbf{w})\big) \setminus \mathbf{W}\Big|\mathbf{x} \in S\Big], \end{aligned}$$

proving the claim. □

We are now ready to prove Claim 6, restated here for convenience.

Claim 6 (Restated). *Assume $\mathsf{P} < \frac{1}{q} + \gamma$ and furthermore that*

$$\Pr_{\mathbf{z}\sim\mathbf{W}^\perp}\left[\frac{1}{q} + \frac{\mathsf{P}}{1000} \le \mathsf{Q}(\mathbf{z}) \le \Delta\right] < \frac{\mathsf{P}}{1000}. \tag{+}$$

Then there exists $\mathbf{w} \in \mathbb{Z}_q^n \setminus \mathbf{W}$ such that $\mathsf{P}' \ge \mathsf{P}/\Delta$, where $\mathsf{P}' := \mathsf{P}_{(\mathbf{A},\mathbf{W}',S)}$ for $\mathbf{W}' = \mathbf{W} + \mathrm{Span}(\mathbf{w})$.

Proof. Claim 8 translates Claim 6 into a statement about the Fourier coefficients of Q. Indeed, note that if $\mathbf{w} \in \mathbb{Z}_q^n \setminus \mathbf{W}$ and $\mathbf{W}' = \mathbf{W} + \mathrm{Span}(\mathbf{w})$ then

$$\mathsf{P}' = \mathsf{P} + \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\Big[f(\mathbf{x}) - \mathbf{A}\mathbf{x} \in \big(\mathbf{W} + \mathrm{Span}(\mathbf{w})\big) \setminus \mathbf{W}\Big|\mathbf{x} \in S\Big] = \mathsf{P} + q^{\dim(\mathbf{W})+1} \cdot \hat{\mathsf{Q}}(\mathbf{w}).$$

Therefore, it suffices to show that under the hypotheses of Claim 6, there exists $\mathbf{w} \in \mathbb{Z}_q^n \setminus \mathbf{W}$ such that $q^{\dim(\mathbf{W})+1} \cdot \hat{\mathsf{Q}}(\mathbf{w}) \geq \mathsf{P} \cdot \big(1/\Delta - 1\big)$; *i.e.*, we need to show that Q has a heavy Fourier coefficient outside of $\mathbf{W}$. This is what we prove below, however our argument is confounded by the fact that Δ and γ have different values for the small primes than they do for the large primes. For this reason, we first present the proof skeleton which works for all q, then we zoom in to complete the proof in all cases for q.

Let $\mu := \mathbb{E}_{\mathbf{z}\sim\mathbf{W}^\perp}\big[\mathsf{Q}(\mathbf{z})\big] = \frac{1}{q} + \big(1 - \frac{1}{q}\big) \cdot \mathsf{P}$. We have

$$\begin{aligned}\Pr_{\mathbf{z}\sim\mathbf{W}^\perp}\big[\mathsf{Q}(\mathbf{z}) \geq \Delta\big] &= \Pr_{\mathbf{z}\sim\mathbf{W}^\perp}\left[\mathsf{Q}(\mathbf{z}) \geq \frac{1}{q} + \frac{\mathsf{P}}{1000}\right] - \Pr_{\mathbf{z}\sim\mathbf{W}^\perp}\left[\frac{1}{q} + \frac{\mathsf{P}}{1000} \leq \mathsf{Q}(\mathbf{z}) \leq \Delta\right] \\ &\geq \left(1 - \frac{1}{q} - \frac{1}{1000}\right) \cdot \mathsf{P} - \frac{\mathsf{P}}{1000} = \big(.998 - 1/q\big) \cdot \mathsf{P},\end{aligned}$$

using $(+)$. Let $\Phi := (\Delta - \mu)^2\big(.998 - 1/q\big)\mathsf{P}$ be shorthand. Note,

$$\Phi \leq (\Delta - \mu)^2 \cdot \Pr_{\mathbf{z}\sim\mathbf{W}^\perp}\big[(\mathsf{Q}(\mathbf{z}) - \mu)^2 \geq (\Delta - \mu)^2\big] \leq \mathbb{E}_{\mathbf{z}\sim\mathbf{W}^\perp}\big[\mathsf{Q}(\mathbf{w})^2\big] - \mu^2$$

by Markov's inequality. Also, we have

$$\mathbb{E}_{\mathbf{z}\sim\mathbf{W}^\perp}\big[\mathsf{Q}(\mathbf{z})^2\big] = q^{\dim(\mathbf{W})} \cdot \mathbb{E}_{\mathbf{z}\sim\mathbb{Z}_q^n}\big[\mathsf{Q}(\mathbf{z})^2\big] = q^{\dim(\mathbf{W})} \cdot \sum_{\mathbf{w}\in\mathbb{Z}_q^n} \hat{\mathsf{Q}}(\mathbf{w})^2,$$

since $\mathsf{Q}(\mathbf{z}) = 0$ for $\mathbf{z} \notin \mathbf{W}^\perp$ and using Parseval (and that $|\hat{\mathsf{Q}}(\mathbf{w})| = \hat{\mathsf{Q}}(\mathbf{w})$ since the Fourier coefficients of Q are all positive reals by Claim 8). Plugging in $q^{\dim(\mathbf{W})} \cdot \hat{\mathsf{Q}}(\mathbf{w}) = \mu$ for all $\mathbf{w} \in \mathbf{W}$ gives

$$\Phi \leq q^{\dim(\mathbf{W})} \cdot \sum_{\mathbf{w}\notin\mathbf{W}} \hat{\mathsf{Q}}(\mathbf{w})^2 \leq q^{\dim(\mathbf{W})} \cdot \max_{\mathbf{w}\notin\mathbf{W}}\Big\{\hat{\mathsf{Q}}(\mathbf{w})\Big\} \cdot \sum_{\mathbf{w}\in\mathbb{Z}_q^n} \hat{\mathsf{Q}}(\mathbf{w}) = q^{\dim(\mathbf{W})} \cdot \max_{\mathbf{w}\notin\mathbf{W}}\Big\{\hat{\mathsf{Q}}(\mathbf{w})\Big\},$$

since $\sum_{\mathbf{w}\in\mathbb{Z}_q^n} \hat{\mathsf{Q}}(\mathbf{w}) = \mathsf{Q}(\mathbf{0}) = 1$. So we have shown that there exists $\mathbf{w} \in \mathbb{Z}_q^n \setminus \mathbf{W}$ such that

$$q^{\dim(\mathbf{W})+1} \cdot \hat{\mathsf{Q}}(\mathbf{w}) \geq q\Phi,$$

and so it just remains to show that we have set things up so that $q\Phi \geq \mathsf{P} \cdot \big(1/\Delta - 1\big)$ for all q. We do this separately for the cases when $q \geq 11$ and when $q \in \{2, 3, 5, 7\}$.

• **Case 1 ($q \geq 11$).** In this case we have $\Delta = q^{-1/4}$, $\big(.998 - 1/q\big) \geq .9$, and $\mu = \frac{1}{q} + \big(1 - \frac{1}{q}\big)\mathsf{P} < \frac{2}{q}$, since $\mathsf{P} < \frac{1}{q} + \gamma$ for $\gamma = \frac{1}{q^2}$. It follows that $q\Phi \geq .9\big(q^{1/2} - 4q^{-1/4} + 4q^{-1}\big) \geq q^{1/4} - 1$, where the final bound holds because the function $\psi(x) = 1 + .9\big(x^{1/2} - 4x^{-1/4} + 4x^{-1}\big) - x^{1/4}$ is positive for $x \geq 6.6$.

• **Case 2 ($q \in \{2, 3, 5, 7\}$).** In this case we have $\Delta = .99$ and $\mu = \frac{1}{q} + \big(1 - \frac{1}{q}\big)\mathsf{P} < \frac{2}{q} - \frac{1}{q^2} + .01$, since $\mathsf{P} < \frac{1}{q} + \gamma$ for $\gamma = .01$. The quantity $(\Delta - \mu)^2$ is thus at least $(.05, .17, .38, .5)$ when $q = (2, 3, 5, 7)$. One can now simply check that $1 + q(\Delta - \mu)^2(.998 - 1/q) \geq 1/\Delta$ holds for $q = 2, 3, 5, 7$ using the crude bound $.998 - 1/q \geq .49$ (crude, at least, when $q > 2$). □

6 Preparing the Output

Lemma 2 (Restated). *Let $m, n, q, r \in \mathbb{N}$ be integers with q prime, let $\varepsilon \geq 12q^{-m/3}$, let $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ be a function, and $S \subset \mathbb{Z}_q^m$ a subset. There exists a randomized algorithm $\mathcal{A}_{\mathsf{output}}$ which has the following syntax, running time and correctness guarantees.*

- **Syntax:** *$\mathcal{A}_{\mathsf{output}}$ takes no input, gets oracle access to f and to the characteristic function of S, and outputs a matrix $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n \times m}$.*
- **Running Time:** *$\mathcal{A}_{\mathsf{output}}$ runs in expected $\mathsf{poly}\big(m, n, \log q, 1/\varepsilon\big)$ time.*
- **Correctness:** *If $\mathbf{A} \in \mathbb{Z}_q^{n \times m}$ is such that S contains f's agreement with $\mathbf{A}$, and if there exists a subspace $\mathbf{W} \subset \mathbb{Z}_q^n$ such that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq 1 - \frac{1}{4m}$, then the output matrix $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n \times m}$ satisfies*

$$\Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_{\mathsf{out}}\mathbf{x}\big] \geq \Pr_{\mathbf{x} \sim \mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big] - \varepsilon,$$

with probability at least $\mathsf{poly}\big(1/n, \varepsilon^{\dim(\mathbf{W})}, \varepsilon^{\log_q(1/\varepsilon)}, m^{-\log_q(1/\varepsilon)}\big)$.

Algorithm Setup: Let $m, n, q \in \mathbb{N}$ be integers such that q is prime, let $\varepsilon > 0$, and let $f : \mathbb{Z}_q^m \to \mathbb{Z}_q^n$ be a function, to which our algorithm has oracle access. Let $\mathcal{A}_{\mathsf{exp}}$ be the matrix recovery algorithm of Section 4.1, Corollary 1 which has exponentially small probability of success.

Input: A set $S \subset \mathbb{Z}_q^m$ (formally, $\mathcal{A}_{\mathsf{output}}$ gets oracle access to $\mathbb{1}_S$).

Part 1: Initialize a linearly independent set $\mathcal{B} \subset \mathbb{Z}_q^m$ and a vector subspace $\mathbf{X} \subset \mathbb{Z}_q^m$ to $\mathcal{B} = \emptyset$ and $\mathbf{X} = \{\mathbf{0}\}$; the invariant $\mathbf{X} = \mathrm{Span}(\mathcal{B})$ will be maintained.

(a) Choose $k \sim \{1, \ldots, m\}$ and do the following k times:
- draw $\mathbf{x} \sim \mathbb{Z}_q^m$;
- if $\mathbf{x} \notin S$ or if $\mathbf{x} \in \mathbf{X}$, reject $\mathbf{x}$ and resample (*i.e.*, go back one instruction);
- if $\mathbf{x} \in S$ and $\mathbf{x} \notin \mathbf{X}$, update $\mathcal{B} = \mathcal{B} \cup \{\mathbf{x}\}$ and $\mathbf{X} = \mathbf{X} + \mathrm{Span}(\mathbf{x})$.

(b) Choose $k \sim \{0, 1, \ldots, m - |\mathcal{B}|\}$, and do the following k times:
- draw $\mathbf{x} \sim \mathbb{Z}_q^m$ and set $\mathcal{B} = \mathcal{B} \cup \{\mathbf{x}\}$ and $\mathbf{X} = \mathbf{X} + \mathrm{Span}(\mathbf{x})$.

Part 2: Let $\mathbf{A}' \in \mathbb{Z}_q^{n \times m}$ be any matrix such that $\mathbf{A}'\mathbf{x} = f(\mathbf{x})$ holds for all $\mathbf{x} \in \mathcal{B}$. Initialize a vector space $\mathbf{V} \subset \mathbb{Z}_q^n$ to $\mathbf{V} = \{\mathbf{0}\}$, choose $k \sim \{0, 1, \ldots, n\}$, and do the following k times.
- Draw $\mathbf{x} \sim \mathbf{X}$, let $\mathbf{v} = f(\mathbf{x}) - \mathbf{A}'\mathbf{x} \in \mathbb{Z}_q^n$ and update $\mathbf{V} = \mathbf{V} + \mathrm{Span}(\mathbf{v})$.

Part 3: Let $T \subset \mathbf{X}$ be the set of $\mathbf{x} \in \mathbf{X}$ such that $f(\mathbf{x}) - \mathbf{A}'\mathbf{x} \in \mathbf{V}$, and let $g : T \to \mathbf{V}$ be the function $g(\mathbf{x}) = f(\mathbf{x}) - \mathbf{A}'\mathbf{x}$. Call $\mathcal{A}_{\mathsf{exp}}$ on g with parameter $\varepsilon/3$ and subset $T \subset \mathbf{X}$, and let $\mathbf{A}'_{\mathsf{out}} \in \mathbb{Z}_q^{n \times m}$ be the output.

Output: Output any $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n \times m}$ such that $\mathbf{A}_{\mathsf{out}}\mathbf{x} = (\mathbf{A}' + \mathbf{A}'_{\mathsf{out}})\mathbf{x}$ holds $\forall\, \mathbf{x} \in \mathbf{X}$.

Figure 3: The Output Preparation Algorithm $\mathcal{A}_{\mathsf{output}}$

Proof. The algorithm $\mathcal{A}_{\mathsf{output}}$ is shown in Figure 3. It is clear that $\mathcal{A}_{\mathsf{output}}$ satisfies the required syntax and running time guarantees. Assume $\mathbf{A} \in \mathbb{Z}_q^{n\times m}$ and $S \subset \mathbb{Z}_q^m$ are such that the input set S contains f's agreement with $\mathbf{A}$ and moreover that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq 1 - \frac{1}{4m}$ holds for some subspace $\mathbf{W} \subset \mathbb{Z}_q^n$ of dimension at most r, and let $\mathbf{A}_{\mathsf{out}} \in \mathbb{Z}_q^{n\times m}$ be the output matrix when $\mathcal{A}_{\mathsf{output}}$ is run on input S. To prove correctness, we specify notions of "success" for the different parts of the algorithm and observe that when success occurs in each part, we have

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big] \leq \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_{\mathsf{out}}\mathbf{x}\big] + \varepsilon.$$

The following are the notions of success for the various parts of the algorithm.

- **Success in Part 1:** Let $\mathcal{B} \subset \mathbb{Z}_q^m$ and $\mathbf{X} \subset \mathbb{Z}_q^m$ be the linearly independent set and subspace constructed in Part 1 of $\mathcal{A}_{\mathsf{output}}$. We say *success occurs in part 1* if the following both hold:
  - $f(\mathbf{x}) \in \mathbf{A}\mathbf{x} + \mathbf{W}$ for all $\mathbf{x} \in \mathcal{B}$;
  - $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big|\mathbf{x} \notin \mathbf{X}\big] \leq \varepsilon/3$.
- **Success in Part 2:** Let $\mathbf{A}' \in \mathbb{Z}_q^{n\times m}$ be the matrix, and $\mathbf{V} \subset \mathbb{Z}_q^n$ the vector space computed during Part 2 of $\mathcal{A}_{\mathsf{output}}$. We say *success occurs in part 2* if $\mathbf{V} \subset \mathbf{W}$ and

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) \in \mathbf{A}'\mathbf{x} + \mathbf{W} \setminus \mathbf{V}\big|\mathbf{x} \in \mathbf{X}\big] \leq \varepsilon/3.$$

- **Success in Part 3:** Let $g : \mathbf{X} \to \mathbf{V}$ be the function, $T = \big\{\mathbf{x} \in \mathbf{X} : f(\mathbf{x}) \in \mathbf{A}'\mathbf{x} + \mathbf{V}\big\}$ the subset, and $\mathbf{A}'_{\mathsf{out}} \in \mathbb{Z}_q^{n\times m}$ the matrix computed in Part 3. We say *success occurs in part 3* if

$$\Pr_{\mathbf{x}\sim\mathbf{X}}\big[g(\mathbf{x}) = \mathbf{A}'_{\mathsf{out}}\mathbf{x}\ \&\ \mathbf{x} \in T\big] \geq \max_{\hat{\mathbf{A}}\in\mathbb{Z}_q^{n\times m}}\Big\{\Pr_{\mathbf{x}\sim\mathbf{X}}\big[g(\mathbf{x}) = \hat{\mathbf{A}}\mathbf{x}\ \&\ \mathbf{x} \in T\big]\Big\} - \varepsilon/3.$$

Success in Part 3 occurs with probability $\mathsf{poly}(\varepsilon^r)$ by Corollary 1, since $\varepsilon \geq 12q^{-m/3}$. Claims 9 and 10 below ensure that success also occurs in Parts 1 and 2 with good probability. These claims are proved below, outside the current proof.

Claim 9. *Assume $S \subset \mathbb{Z}_q^m$ contains f's agreement with $\mathbf{A}$ and that $\mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \geq 1 - \frac{1}{4m}$ for some subspace $\mathbf{W} \subset \mathbb{Z}_q^n$ of dimension at most r. Then success occurs in Part 1 of $\mathcal{A}_{\mathsf{output}}$ with probability at least $\mathsf{poly}\big(m^{-\log_q(1/\varepsilon)}, \varepsilon^{\log_q(1/\varepsilon)}\big)$.*

Claim 10. *Consider an execution of $\mathcal{A}_{\mathsf{output}}$ where success occurred in Part 1. Then success also occurs in Part 2 with probability $\mathsf{poly}\big(1/n, \varepsilon^r\big)$.*

It follows that with probability at least $\mathsf{poly}\big(1/n, \varepsilon^r, \varepsilon^{\log_q(1/\varepsilon)}, m^{-\log_q(1/\varepsilon)}\big)$, success occurs in all three parts of $\mathcal{A}_{\mathsf{output}}$, in which case

$$\begin{aligned}\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big] &\leq q^{-d}\cdot\mathsf{P}^{(1)} + \varepsilon/3 \leq q^{-d}\cdot\mathsf{P}^{(2)} + 2\varepsilon/3 \leq q^{-d}\cdot\mathsf{P}^{(3)} + \varepsilon \\ &\leq \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_{\mathsf{out}}\mathbf{x}\big] + \varepsilon,\end{aligned}$$

holds, where $d = m - \dim(\mathbf{X})$, and

- $\mathsf{P}^{(1)} := \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big|\mathbf{x} \in \mathbf{X}\big]$;
- $\mathsf{P}^{(2)} := \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\ \&\ f(\mathbf{x}) \in \mathbf{A}'\mathbf{x} + \mathbf{V}\big|\mathbf{x} \in \mathbf{X}\big]$;

· $\mathsf{P}^{(3)} := \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_{\mathsf{out}}\mathbf{x} \,\&\, f(\mathbf{x}) \in \mathbf{A}'\mathbf{x} + \mathbf{V} \big| \mathbf{x} \in \mathbf{X}\big]$.

Indeed, the first inequality holds because success occurs in Part 1:

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big] = q^{-d}\cdot\mathsf{P}^{(1)} + \big(1-q^{-d}\big)\cdot\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x}\big|\mathbf{x}\notin\mathbf{X}\big] \le q^{-d}\cdot\mathsf{P}^{(1)} + \varepsilon/3.$$

The second inequality follows from

$$\mathsf{P}^{(1)} = \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}\mathbf{x} \,\&\, f(\mathbf{x}) \in \mathbf{A}'\mathbf{x} + \mathbf{W}\big|\mathbf{x}\in\mathbf{X}\big] \le \mathsf{P}^{(2)} + \varepsilon/3,$$

which holds when Parts 1 and 2 are successful because in this case $(\mathbf{A}-\mathbf{A}')\mathbf{x} \in \mathbf{W}$ holds for all $\mathbf{x}\in\mathbf{X}$, and additionally $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) \in \mathbf{A}'\mathbf{x}+\mathbf{W}\setminus\mathbf{V}\big|\mathbf{x}\in\mathbf{X}\big] \le \varepsilon/3$. The third inequality holds when Part 3 is successful since in this case

$$\mathsf{P}^{(2)} = \Pr_{\mathbf{x}\sim\mathbf{X}}\big[g(\mathbf{x}) = (\mathbf{A}-\mathbf{A}')\mathbf{x} \,\&\, \mathbf{x}\in T\big] \le \Pr_{\mathbf{x}\sim\mathbf{X}}\big[g(\mathbf{x}) = \mathbf{A}'_{\mathsf{out}}\mathbf{x} \,\&\, \mathbf{x}\in T\big] + \varepsilon/3 = \mathsf{P}^{(3)} + \varepsilon/3.$$

Finally, the fourth inequality holds because

$$q^{-d}\cdot\mathsf{P}^{(3)} = \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_{\mathsf{out}}\mathbf{x} \,\&\, f(\mathbf{x}) \in \mathbf{A}'\mathbf{x}+\mathbf{V} \,\&\, \mathbf{x}\in\mathbf{X}\big] \le \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) = \mathbf{A}_{\mathsf{out}}\mathbf{x}\big].$$

This completes the proof of Lemma 2 and it remains only to prove Claims 9 and 10. □

Proof of Claim 9. Let $\hat{S} = \big\{\mathbf{x}\in S : f(\mathbf{x}) \in \mathbf{A}\mathbf{x}+\mathbf{W}\big\} \subset \mathbb{Z}_q^m$, and let $\sigma := |S|/q^m$ and $\hat{\sigma} := |\hat{S}|/q^m$. Note $\sigma \ge \varepsilon$, as S contains f's agreement with $\mathbf{A}$, and $\hat{\sigma} = (1-\zeta)\sigma$, where $1-\zeta = \mathsf{P}_{(\mathbf{A},\mathbf{W},S)} \ge 1 - \frac{1}{4m}$. The two quantities of interest for Part 1a are

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[\mathbf{x}\in S\setminus\mathbf{X}\big];\ \text{and}\ \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[\mathbf{x}\in\hat{S}\big|\mathbf{x}\in S\setminus\mathbf{X}\big],$$

since these dictate the chances that the random sample $\mathbf{x}\sim\mathbb{Z}_q^m$ is kept, and that $f(\mathbf{x})\in\mathbf{A}\mathbf{x}+\mathbf{W}$ given that it is kept. Note that as long as the dimension of the subspace $\mathbf{X}\subset\mathbb{Z}_q^m$ is not too large, these probabilities are very close to σ and $1-\zeta$, respectively. Specifically, let $m-d = \dim(\mathbf{X})$ and assume $q^{-d} \le \sigma\zeta$ holds. Then we have

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[\mathbf{x}\in S\big|\mathbf{x}\notin\mathbf{X}\big] = \frac{\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}[\mathbf{x}\in S\setminus\mathbf{X}]}{\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}[\mathbf{x}\notin\mathbf{X}]} \ge \frac{\sigma q^m - q^{m-d}}{q^m - q^{m-d}} \ge \frac{\sigma q^m - \sigma\zeta q^m}{q^m} = \sigma(1-\zeta);$$

and similarly,

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[\mathbf{x}\in\hat{S}\big|\mathbf{x}\in S\setminus\mathbf{X}\big] = \frac{\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}[\mathbf{x}\in\hat{S}\,\&\,\mathbf{x}\notin\mathbf{X}]}{\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}[\mathbf{x}\in S\,\&\,\mathbf{x}\notin\mathbf{X}]} \ge \frac{\hat{\sigma}q^m - q^{m-d}}{\sigma q^m} \ge \frac{\hat{\sigma}q^m - \sigma\zeta q^m}{\sigma q^m} = 1-2\zeta.$$

In particular, this means that as long as the k chosen during Step 1a is maximal such that $q^{-(m-k)} \le \sigma\zeta$ (occurs with probability $1/m$), the Step 1a loop terminates in expected $\mathsf{poly}\big(m,n,\log q, 1/\sigma\big)$ time and with $\mathcal{B}\subset T$ with probability at least $1-2\zeta m \ge \frac{1}{2}$, by the union bound. Note that $\mathcal{B}\subset T$ means exactly that $f(\mathbf{x})\in\mathbf{A}\mathbf{x}+\mathbf{W}$ holds for all $\mathbf{x}\in\mathcal{B}$.

Now, suppose that the random $k\sim\{0,1,\ldots,m-|\mathcal{B}|\}$ chosen during Step 1b is maximal such that $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x})\in\mathbf{A}\mathbf{x}+\mathbf{W}\big|\mathbf{x}\notin\mathbf{X}_j\big] \ge \varepsilon/3$ holds for all $j\le k$, where $\mathbf{X}_j$ denotes the vector space $\mathbf{X}$ at the beginning of the $j-$th execution of the loop in Step 1b. Note if we get lucky in this way with our choice of k (occurs with probability at least $1/m$), then $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x})\in\mathbf{A}\mathbf{x}+\mathbf{W}\big|\mathbf{x}\notin\mathbf{X}\big] < \varepsilon/3$ holds upon exiting Step 1b. Additionally, since $\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x})\in\mathbf{A}\mathbf{x}+\mathbf{W}\big|\mathbf{x}\notin\mathbf{X}\big] \ge \varepsilon/3$ holds at all

times during the execution of Step 1b, for each $\mathbf{x} \sim \mathbb{Z}_q^m$ chosen during the loop, the chance that $\mathbf{x} \notin \mathbf{X}$ and $f(\mathbf{x}) \in \mathbf{A}\mathbf{x} + \mathbf{W}$ hold is at least

$$\Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[\mathbf{x} \notin \mathbf{X}\big] \cdot \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) \in \mathbf{A}\mathbf{x} + \mathbf{W} | \mathbf{x} \notin \mathbf{X}\big] \geq \frac{\varepsilon}{6}.$$

Therefore, with probability at least $\big(\varepsilon/6\big)^{m-|\mathcal{B}|} \geq \big(\varepsilon/6\big)^{\log_q(1/\sigma)+\log_q(1/\zeta)} = \mathsf{poly}\big(m^{-\log_q(1/\varepsilon)}, \varepsilon^{\log_q(1/\varepsilon)}\big)$, $\mathcal{B} \subset \hat{S}$ holds upon exiting Step 1b. □

Proof of Claim 10. Let $\mathbf{V}_i$ for $i = 1, \ldots, k$ denote the vector space $\mathbf{V}$ after the $i-$th execution of the loop in Part 2, let $\mathbf{V}_0 = \{\mathbf{0}\}$, and let $\rho_i := \Pr_{\mathbf{x}\sim\mathbb{Z}_q^m}\big[f(\mathbf{x}) \in \mathbf{A}'\mathbf{x} + \mathbf{W} \setminus \mathbf{V}_i \big| \mathbf{x} \in \mathbf{X}\big]$. Assume we got lucky with our choice of $k \sim \{0, \ldots, n\}$ and that k is maximal such that $\rho_i \geq \varepsilon/3$ for all $i < k$ (happens with probability $\frac{1}{n+1}$). Note in this case, $\rho_k < \varepsilon/3$ holds upon exiting Part 2. Also, since $\rho_i \geq \varepsilon/3$ holds for all $i < k$, each time through the loop in Part 2 with probability at least $\varepsilon/3$, the $\mathbf{x} \sim \mathbf{X}$ drawn will be such that $\mathbf{v} = f(\mathbf{x}) - \mathbf{A}'\mathbf{x} \in \mathbf{W} \setminus \mathbf{V}_i$. Thus, with probability at least $\big(\varepsilon/3\big)^k \leq \big(\varepsilon/3\big)^r$, $\mathbf{V}_k \subset \mathbf{W}$ holds. □

References

[AGGS22] Vahid R. Asadi, Alexander Golovnev, Tom Gur, and Igor Shinkar. Worst-case to average-case reductions via additive combinatorics. In Stefano Leonardi and Anupam Gupta, editors, *STOC '22: 54th Annual ACM SIGACT Symposium on Theory of Computing, Rome, Italy, June 20 - 24, 2022*, pages 1566–1574. ACM, 2022.

[AGS03] Adi Akavia, Shafi Goldwasser, and Shmuel Safra. Proving hard-core predicates using list decoding. In *44th Symposium on Foundations of Computer Science (FOCS 2003), 11-14 October 2003, Cambridge, MA, USA, Proceedings*, pages 146–157. IEEE Computer Society, 2003.

[BBW18] László Babai, Timothy J. F. Black, and Angela Wuu. List-decoding homomorphism codes with arbitrary codomains. In Eric Blais, Klaus Jansen, José D. P. Rolim, and David Steurer, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2018, August 20-22, 2018 - Princeton, NJ, USA*, volume 116 of *LIPIcs*, pages 29:1–29:18. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *J. Comput. Syst. Sci.*, 47(3):549–595, 1993.

[Blu83] Manuel Blum. Coin flipping by telephone a protocol for solving impossible problems. *SIGACT News*, 15(1):23–27, 1983.

[DGKS08] Irit Dinur, Elena Grigorescu, Swastik Kopparty, and Madhu Sudan. Decodability of group homomorphisms beyond the johnson bound. In Cynthia Dwork, editor, *Proceedings of the 40th Annual ACM Symposium on Theory of Computing, Victoria, British Columbia, Canada, May 17-20, 2008*, pages 275–284. ACM, 2008.

[DHK+19] Irit Dinur, Prahladh Harsha, Tali Kaufman, Inbal Livni Navon, and Amnon Ta-Shma. List decoding with double samplers. In Timothy M. Chan, editor, *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2019, San Diego, California, USA, January 6-9, 2019*, pages 2134–2153. SIAM, 2019.

[GKS06] Elena Grigorescu, Swastik Kopparty, and Madhu Sudan. Local decoding and testing for homomorphisms. In Josep Díaz, Klaus Jansen, José D. P. Rolim, and Uri Zwick, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 9th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, APPROX 2006 and 10th International Workshop on Randomization and Computation, RANDOM 2006, Barcelona, Spain, August 28-30 2006, Proceedings*, volume 4110 of *Lecture Notes in Computer Science*, pages 375–385. Springer, 2006.

[GL89] Oded Goldreich and Leonid A. Levin. A hard-core predicate for all one-way functions. In David S. Johnson, editor, *Proceedings of the 21st Annual ACM Symposium on Theory of Computing, May 14-17, 1989, Seattle, Washigton, USA*, pages 25–32. ACM, 1989.

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999.

[IJK09] Russell Impagliazzo, Ragesh Jaiswal, and Valentine Kabanets. Approximate list-decoding of direct product codes and uniform hardness amplification. *SIAM J. Comput.*, 39(2):564–605, 2009.

[KM93] Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the fourier spectrum. *SIAM J. Comput.*, 22(6):1331–1348, 1993.

[KT00] Jonathan Katz and Luca Trevisan. On the efficiency of local decoding procedures for error-correcting codes. In F. Frances Yao and Eugene M. Luks, editors, *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, May 21-23, 2000, Portland, OR, USA*, pages 80–86. ACM, 2000.

[Sam07] Alex Samorodnitsky. Low-degree tests at large distances. In David S. Johnson and Uriel Feige, editors, *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*, pages 506–515. ACM, 2007.

[San12] Tom Sanders. On the bogolyubov-ruzsa lemma. *Analysis & PDE*, 5(3):627–655, 2012.

[Sud97] Madhu Sudan. Decoding of reed solomon codes beyond the error-correction bound. *J. Complex.*, 13(1):180–193, 1997.

[Tre03] Luca Trevisan. List-decoding using the XOR lemma. In *44th Symposium on Foundations of Computer Science (FOCS 2003), 11-14 October 2003, Cambridge, MA, USA, Proceedings*, pages 126–135. IEEE Computer Society, 2003.

[TW14] Madhur Tulsiani and Julia Wolf. Quadratic goldreich-levin theorems. *SIAM J. Comput.*, 43(2):730–766, 2014.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il