Fourier Sketching

A Property Testing Perspective

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

by

Manmatha Roy



Under the Supervision of

Subhamoy Maitra
Indian Statistical Institute, Kolkata

Arijit Ghosh
Indian Statistical Institute, Kolkata

October 2026

Abstract

Fourier analysis of Boolean functions is a fundamental tool with far-reaching applications in complexity theory, learning theory, and cryptography. In many natural settings, functions of interest exhibit Fourier sparsity: they can be well approximated using only a small number of significant Fourier coefficients. In this thesis, we study Fourier-sparse Boolean functions from the perspective of property testing, where the goal is to infer global structural properties of an unknown function using only a small number of queries to its truth table, and where the efficiency of an algorithm is measured by its query complexity. Originating in the context of PCP constructions and later evolving into an independent area, property testing provides a natural lens for understanding the intrinsic complexity of combinatorial objects.

Our first contribution provides a refined analysis of a sketching-by-hashing framework that operates directly in the Fourier domain. While such an approach was previously explored by Gopalan et al. (2011), we significantly strengthen its theoretical guarantees by deriving sharper structural bounds for the hashing process. As a consequence, we obtain a query-efficient local list-correction algorithm for the Hadamard code. This tool yields several applications, including query-efficient testers for distribution-free Fourier sparsity, testing membership in subclasses of Fourier-sparse Boolean functions, linear isomorphism testing with respect to a known Boolean function, and sumset size estimation for set functions. All of these problems are well studied in learning theory, cryptography, and additive combinatorics.

Our second contribution introduces a sketching-by-hashing framework based on subspace restrictions, operating in the input domain of Boolean functions. By analyzing how functions behave under random affine subspace restrictions, we establish several structural results that lead to improved algorithms for testing Fourier sparsity. The same framework enables estimation of the heaviest Fourier coefficient up to small additive error, with query complexity depending only on the desired accuracy. These problems are relevant across several areas of computer science. A Fourier sparsity tester, for instance, can serve as a preprocessing step for learning algorithms over Fourier-sparse function classes, validating whether an unknown function belongs to the target hypothesis class; the ability to estimate the heaviest Fourier coefficient, on the other hand, can be leveraged to design distinguishers for symmetric-key ciphers.

Taken together, our work strengthens the connection between sketching-by-hashing techniques and the property testing of Fourier-sparse Boolean functions, and advances the understanding of query complexity for learning and testing in the Fourier domain. We expect the algorithmic techniques developed here to find further applications across many areas of computer science.

List of Papers

Works Included in this Thesis

Published

1. **Testing Fourier Sparsity via Implicit Sensing**, with Arijit Ghosh and Subhamoy Maitra, In *The 14th International Conference on Learning Representations (ICLR 2026)*
2. **Spectral Norm, Economical Sieve, and Linear Invariance Testing of Boolean Functions**, with Swarnalipa Datta, Arijit Ghosh, Chandrima Kayal, and Manaswi Paraashar, In *The 43rd International Symposium on Theoretical Aspects of Computer Science (STACS 2026)*
3. **Spectral Shadows: When Communication Complexity Meets Linear Invariance Testing**, with Swarnalipa Datta, Arijit Ghosh, Chandrima Kayal, and Manaswi Paraashar, In *Computational Complexity (2026)*
4. **Price of Parsimony: Complexity of Fourier Sparsity Testing**, with Arijit Ghosh, In *The 39th Annual Conference on Neural Information Processing Systems (NeurIPS 2025)*
5. **Distribution-Free Fourier Sparsity Testing**, with Arijit Ghosh, In *The 40th Annual Conference on Neural Information Processing Systems (NeurIPS 2026)*

Under Review

1. **On Nonlinearity Estimation Problem**, with Arijit Ghosh and Subhamoy Maitra
2. **Testing Fourier Sparsity Near Optimally**, with Arijit Ghosh
3. **Exact Fourier-Sparse Recovery Made Simple**, with Arijit Ghosh

Manuscript

1. **Approximating Sumset Size via Dense Coset Decomposition**, with Arijit Ghosh

Related Works

Not Included in this Thesis

Published

1. **Testing Isomorphism of Boolean Functions over Abelian Groups**, with Swarnalipa Datta, Arijit Ghosh, Chandrima Kayal, and Manaswi Paraashar, In *The 29th International Conference on Randomization and Computation (RANDOM 2025)*

2. **Construction of Maiorana–McFarland Type Cryptographically Significant Boolean Functions with Good Implementation Properties**, with Deng Tang, Anupam Chattopadhyay, Bimal Mandal, and Subhamoy Maitra, In *Communications in Cryptology, Volume 2, Number 1 (2025)*

3. **On Differential Uniformity of Extremal Functions**, with Nikolay Stoyanov Kaleyski, In *The 10th International Workshop on Boolean Functions and their Applications (BFA 2025)*

4. **Modifying Bent Functions to Obtain Balanced Ones with High Nonlinearity**, with Subhamoy Maitra and Bimal Mandal, In *The 23rd International Conference on Cryptology in India (Indocrypt 2022)*

5. **Experimental Results on Higher-Order Differential Spectra of 6- and 8-bit Invertible S-Boxes**, with Deng Tang, Bimal Mandal, and Subhamoy Maitra, In *The 10th International Conference on Security, Privacy and Applied Cryptographic Engineering (SPACE 2020)*

Under Review

1. **Constructive Arithmetic Regularity for Abelian Groups**, with Swarnalipa Datta, Arijit Ghosh, and Chandrima Kayal

Contents

Introduction 1

Background 12

I Fourier Sketching by Coset Hashing 19

1 The Coset Hashing Technique 20

2 Application I: Testing Linear Isomorphism 37

3 Application II: Distribution-Free ℓ_2 Testing of Fourier Sparsity 45

4 Application III: Improved ℓ_0 Testing of Fourier Sparsity 58

5 Application IV: Estimating Sumset Size 68

II Fourier Sketching by Subspace Restriction 83

6 The Subspace Restriction Technique 84

7 Application V : Near Optimal ℓ_0 Testing of Fourier Sparsity 96

8 Application VI : Near Optimal ℓ_2 Testing of Fourier Sparsity 125

9 Application VII: Estimating Fourier Bias 141

10 Application VIII: Exact Fourier Sparse Recovery Made Simple 152

III A Communication Complexity Perspective **167**

11 Communication Complexity of Linear Isomorphism Problem **168**

Conclusion **188**

Introduction

A fundamental challenge in modern algorithm design, particularly in the era of big data, is to extract useful information from high-dimensional objects while using computational resources that are substantially smaller than those required by an exact representation of the object under consideration. A recurring approach to this challenge is the construction of succinct summaries or *sketches*: low-complexity approximate representations of the object that preserve sufficient information to support efficient inference about the original object. One of the earliest and most influential examples of this paradigm arises in the *heavy-hitter problem* from the data-streaming literature.

Consider a stream $S = (a_1, a_2, \ldots, a_m)$, of length m, where each item a_t belongs to a universe $[n] = \{1, 2, \ldots, n\}$. The stream induces a frequency vector

$$f = (f_1, f_2, \ldots, f_n) \in \mathbb{R}^n,$$

where f_i denotes the number of occurrences of item i in the stream. An item i is called a *heavy hitter* if its frequency is significantly larger than that of a typical coordinate, e.g., a constant factor of m. The objective is to identify all such coordinates while processing the stream using memory substantially smaller than that required to store the entire frequency vector. Over the past two decades, the heavy-hitter problem has become a central topic in the data-streaming literature owing to both its fundamental nature and its numerous applications in areas such as network monitoring, database systems, and large-scale analytics. In many of these settings, storing the entire stream is infeasible. Consequently, one seeks to maintain a compact summary, or *sketch*, of the stream from which the heavy hitters can later be approximately recovered. A common theme underlying many sketching algorithms for the heavy-hitter problem is the use of hashing. Let

$$h : [n] \to [d], \qquad d \ll n,$$

be a hash function that partitions the coordinates into a small number of buckets. Rather than maintaining a separate counter for every coordinate $i \in [n]$, the algorithm stores only one counter for each bucket $j \in [d]$. This reduces the memory requirement from $O(n \log m)$ to $O(d \log m)$, where m denotes the length of the stream.

For each bucket j, the sketch maintains an aggregate quantity of the form

$$C_j = \sum_{\substack{i \in [n] \\ h(i)=j}} g(f_i), \qquad j \in [d],$$

where f_i denotes the frequency of coordinate i, and $g : \mathbb{R} \to \mathbb{R}$ is a function chosen so that the corresponding bucket statistics can be updated incrementally as stream updates arrive, and coordinates with large frequencies make a disproportionately large contribution to the bucket aggregates. Consequently, when the frequency vector contains only a few dominant coordinates, the contribution of a heavy hitter to the bucket containing it is significantly larger than the aggregate contribution of most non-heavy coordinates. Under suitable randomness assumptions on the hash function, the interference caused by the remaining coordinates can be bounded probabilistically. Thus, even though the sketch stores only aggregated bucket values, the presence of heavy hitters can still be detected and, in many cases, accurately recovered. Classical data-stream sketches such as Count-Min Sketch and Count Sketch are built upon this fundamental hashing paradigm.

The remarkable success of heavy-hitter algorithms led to the emergence of what is now commonly known as the *sketching-by-hashing paradigm*. Since then, the influence of this paradigm has extended far beyond its origins in data-stream algorithms. Variants of the same fundamental idea now appear in compressed sensing, sparse Fourier transforms, dimensionality reduction, randomized numerical linear algebra, and numerous other areas of algorithm design. Although the objectives in these settings differ substantially, they share a common philosophy. Rather than operating directly on a high-dimensional object, one first computes a compact randomized summary by partitioning the domain using suitable hash functions and aggregating information within the resulting buckets. Computation is then performed on this compressed representation rather than on the original object itself. The challenge is to design the sketch so that it preserves enough information to solve the problem of interest while requiring significantly fewer resources than a naive representation.

At a conceptual level, the success of these methods is driven by the presence of *intrinsic sparsity* or low-dimensional structure. In the heavy-hitter problem, only a small number of coordinates contribute a significant fraction of the total mass. In compressed sensing, the signal admits a sparse representation. In sparse Fourier transforms, only a few Fourier coefficients are significant. In low-rank approximation, most of the information is concentrated in a low-dimensional subspace. The sketch does not create structure; rather, it exploits structure that is already present in the underlying object.

Property Testing

Another area of theoretical computer science that emerged from the study of probabilistically checkable proofs (PCPs) and has since developed into a rich field of its own is *property testing*. A typical question in property testing asks whether a large combinatorial object possesses a certain property or is *far* from every object having that property, while examining only a tiny portion of the input. For example, given a large graph, one may wish to determine whether it is bipartite or *far* from every bipartite graph by inspecting only a small number of edges. Here, far means that a constant fraction of the edges must be modified in order to obtain a bipartite graph.

Property testing is primarily concerned with graph and function objects, and has become an important framework in algorithms for big data. When the input is extremely large, reading the entire object may itself be prohibitively expensive; property testing addresses this challenge by asking whether meaningful conclusions can be drawn from only a very small number of observations. In this sense, property testing is complementary to the streaming model discussed earlier. In streaming algorithms, the entire input is observed, but restrictions are imposed on the memory available to process it. In property testing, by contrast, computational resources are typically unrestricted, but the algorithm is allowed to inspect only a tiny fraction of the input. The central question is therefore not how to store information efficiently, but rather how much information must be observed before a reliable decision can be made.

In this thesis, we focus on testing properties of functions over the Boolean cube

$$f : \mathbb{F}_2^n \to \mathbb{R},$$

where the domain size 2^n is so large that the function cannot be inspected in its entirety. Instead, an algorithm is granted oracle access to f and may query only a small number of function values. The objective is to determine whether it possesses a specified property or is ε-far from every function having that property. One of these properties is *Fourier sparsity*. A function is Fourier sparse if only a small number of Fourier coefficients are nonzero. Rather than recovering the entire spectrum, we seek algorithms that can estimate or test sparsity directly from oracle access to the function. Another property we consider is *testing linear isomorphism with respect to a known function*. Given oracle access to an unknown function f and an explicit description of a target function g, the goal is to determine whether f can be obtained from g by an invertible linear change of variables or whether it is far from every such transformation of g.

At first sight, these problems appear unrelated. The first asks for quantitative information

about the support of the Fourier transform, whereas the second concerns determining an equivalence relation between functions. However, both can be viewed through a common lens: neither requires explicit recovery of the entire Fourier spectrum. Instead, it suffices to estimate carefully chosen statistics of the Fourier spectrum of the unknown function that capture the property of interest. In many cases, this task ultimately reduces to identifying the presence of heavy Fourier coefficients and estimating their magnitudes. Here, a heavy Fourier coefficient refers to a coefficient whose magnitude is significantly larger than that of the remaining coefficients and therefore contributes substantially to the spectral structure of the function.

This observation motivates the search for a generic mechanism that compresses the spectrum while preserving the information needed for inference. The framework of Fourier hashing, developed in this thesis, serves precisely this purpose. By aggregating Fourier coefficients into a small number of buckets using randomized hash functions, it produces compact spectral sketches that retain information about the dominant coefficients and other relevant spectral statistics. These sketches are sufficiently informative to support efficient testing and estimation while avoiding the cost of full Fourier recovery.

Fourier Sketching

The main technical framework developed in this thesis is a paradigm we call Fourier sketching. Inspired by sketching-by-hashing techniques from streaming algorithms, a Fourier sketch consists of a collection of randomized measurements obtained by hashing Fourier coefficients into buckets and maintaining aggregate spectral information within each bucket. By analyzing the resulting measurements and their collision statistics, one can estimate global spectral parameters while entirely avoiding the cost of full Fourier recovery.

Since many properties of functions over the Boolean cube admit elegant Fourier-analytic characterizations, estimating such spectral statistics is often sufficient for testing or estimating the corresponding property. The scope of this thesis is to demonstrate the power of this framework across a diverse collection of problems. In particular, we show that Fourier sketching yields new algorithms for several long-standing open problems for which no satisfactory solutions were previously known. These results suggest that Fourier sketching is not merely a tool tailored to a specific application, but rather a general paradigm for extracting information from the Fourier spectrum through compact randomized summaries.

Two complementary techniques underlie the algorithms developed in this thesis: *coset*

hashing and *subspace restriction*. Although they operate on different objects and serve distinct algorithmic purposes, they are intimately connected through Fourier duality. Together, they provide the conceptual foundation for the Fourier sketching framework developed in this work.

Coset hashing. Coset hashing operates directly on the Fourier domain. Given a subspace $H \leq \mathbb{F}_2^n$ of codimension t, the cosets of $H^\perp$ partition the 2^n Fourier frequencies into 2^t buckets. For each bucket one computes an aggregate measurement $W(a)$ that combines the Fourier mass of all frequencies contained in the corresponding coset. The resulting sketch is therefore a compressed representation of the spectrum in which individual Fourier coefficients are no longer visible, but certain global spectral statistics remain accessible.

The closest analogue arises in signal processing. When a signal is sampled below the Nyquist rate, multiple frequencies become indistinguishable and alias into the same FFT bin. The observed energy in that bin is the aggregate contribution of all aliasing frequencies. Coset hashing behaves in exactly the same manner: frequencies belonging to the same coset of $H^\perp$ become indistinguishable and contribute jointly to a single bucket measurement. From this perspective, coset hashing may be viewed as a Boolean analogue of frequency folding, a central idea in modern sparse Fourier transform and sparse Hadamard transform algorithms.

This technique has a long history in the literature. Variants of Fourier-domain hashing appear implicitly and explicitly in sparse Hadamard transform algorithms, where the goal is to identify large Fourier coefficients using only a small number of measurements. Related ideas were introduced into property testing through the influential work of Gopalan and collaborators, where random partitions of the Fourier spectrum were used to estimate spectral quantities without recovering the spectrum itself. Since then, Fourier-domain hashing has become a recurring tool in several areas of Fourier analysis of Boolean functions, including learning theory, property testing, coding theory, and sparse recovery.

One of the goals of this thesis is to develop a more refined understanding of this technique. Rather than viewing coset hashing merely as an algorithmic primitive, we establish several structural results explaining precisely what information is preserved, what information is lost, and how collision statistics encode important properties of the underlying Fourier spectrum. These structural insights form the basis for the new testing and estimation algorithms developed in later chapters.

Subspace restriction. Subspace restriction, in contrast, acts on the input domain rather than the Fourier domain. Given a subspace $H \leq \mathbb{F}_2^n$ of dimension t and a shift $r \in \mathbb{F}_2^n$, the restriction of f to the affine coset $r + H$ yields a function

$$f|_{r+H} : r + H \to \mathbb{R},$$

defined on only 2^t points instead of the original 2^n points. Restriction therefore provides a form of dimensional reduction by focusing attention on a small localized portion of the domain.

Unlike coset hashing, restriction does not explicitly aggregate information. Instead, it preserves the complete behavior of the function on a lower-dimensional domain. Nevertheless, one of the remarkable features of Fourier analysis is that such localized measurements implicitly induce a form of Fourier sketching. The Fourier spectrum of the restricted function is not arbitrary; rather, it is determined by structured combinations of Fourier coefficients of the original function. Consequently, information about the global spectrum becomes encoded within the local view provided by the restriction.

Subspace restriction may be viewed as a form of localization in the input domain. Rather than observing the function on all of $\mathbb{F}_2^n$, we observe it only on a low-dimensional affine subspace $r + H$. Although this operation is performed entirely in the input domain, Fourier duality implies that it induces a highly structured transformation of the spectrum. Specifically, each Fourier coefficient of the restricted function is a linear combination of Fourier coefficients from a single coset of $H^\perp$. Consequently, frequencies lying in the same coset become indistinguishable and are folded together into aggregate measurements. In this sense, subspace restriction acts as an implicit form of Fourier sketching, analogous to frequency folding or aliasing in signal processing, where local observations encode compressed information about the global spectrum. Random restrictions have long played a central role in property testing, coding theory, and higher-order Fourier analysis. Many global properties of Boolean functions survive random restrictions, including low-degree structure, Reed–Muller code membership, juntas, and various notions of pseudorandomness. Most previous applications, however, have been motivated primarily by higher-order Fourier analytic considerations and the study of polynomial structure.

The perspective adopted in this thesis is different. Our focus is on classical Fourier analysis and, in particular, on understanding how approximately Fourier-sparse functions behave under random affine restrictions. We show that many spectral properties remain stable under restriction and that carefully chosen restrictions provide a powerful mechanism for estimating global spectral parameters from local observations. This viewpoint leads to new algorithms and new structural insights that complement the Fourier-domain hashing

approach.

Two sides of the same principle. An important distinction should be emphasized between these two techniques. In coset hashing, the algorithm deliberately acts on the Fourier spectrum itself, explicitly combining groups of Fourier coefficients into aggregate measurements. The sketch is therefore constructed directly in the frequency domain.

Subspace restriction follows a very different philosophy. Here the algorithm never accesses the Fourier spectrum directly. Instead, it performs localized measurements in the physical domain by observing the function on a carefully chosen affine subspace. Any compression of spectral information occurs only implicitly through the Fourier-analytic relationship between the original function and its restriction.

Despite this difference, the two techniques are fundamentally connected through Fourier duality. Two distinct operations should be distinguished here. Averaging (projecting) a function over a subspace H retains precisely the Fourier modes belonging to $H^\perp$ while annihilating all others, a fact formalized by the Poisson summation formula. Restriction to an affine coset $r + H$ is the dual operation: it folds each coset of $H^\perp$ into a single aggregate coefficient, inducing the aliasing described above. In both cases, operations that appear local in the input domain correspond to highly structured operations in the Fourier domain governed by the interaction between H and $H^\perp$.

In this precise sense, coset hashing and subspace restriction are dual manifestations of the same underlying Fourier-analytic principle. Coset hashing compresses information in the frequency domain, whereas subspace restriction localizes information in the input domain. One aggregates frequencies, the other isolates portions of the domain. Yet both ultimately exploit the same interaction between a subspace and its orthogonal complement. This duality serves as a recurring theme throughout the thesis and underlies many of the structural and algorithmic results that follow.

Applications

The thesis develops the Fourier sketching framework and applies it to a variety of problems arising in learning theory, coding theory, cryptography, and additive combinatorics. A unifying theme across all of these applications is that each property of interest can be reduced to estimating a suitable statistic of the Fourier spectrum via compact randomized sketches.

Problem	Previous UB	Our UB	Previous LB	Our LB
ℓ_0 Fourier sparsity	$\tilde{O}(s^{14})$	$\tilde{O}(s^4)$	$\Omega(\sqrt{s})$	$\Omega(s)$
ℓ_2 Fourier sparsity	$\tilde{O}(s)$	$\tilde{O}(s)$	$\Omega(\sqrt{s})$	$\Omega(s)$
Distribution-free ℓ_2 Fourier sparsity	–	$\tilde{O}(s^4)$	–	$\Omega(s)$
Fourier bias estimation	$\tilde{O}(n/\varepsilon^3)$	$\tilde{O}(1/\varepsilon^2)$	–	$\Omega(1/\varepsilon^2)$
Linear isomorphism	$\tilde{O}(m^{32})$	$\tilde{O}(m^4)$	$\Omega(\log m)$	$\Omega(m^2)$
Sumset size estimation	$\tilde{O}(\mathrm{tower}(1/\varepsilon^3))$	$\tilde{O}(m^4/\varepsilon^8)$	–	–
Randomized CC of Linear Isomorphism	–	$\tilde{O}(m^4)$	–	$\Omega(m^2)$
Deterministic CC of Linear Isomorphism	–	$\tilde{O}(m^2)$	–	$\Omega(\log m)$

Table 1: Summary of the main theoretical contributions. The table compares the best previously known upper and lower bounds with the corresponding bounds obtained in our work.

Testing Fourier Sparsity in ℓ_0 Distance. A Boolean function $f : \mathbb{F}_2^n \to \{-1, +1\}$ is said to be s-Fourier-sparse if its Fourier expansion contains at most s nonzero coefficients. We study the problem of deciding, given oracle access to an unknown Boolean function, whether it is s-Fourier-sparse or is far from every s-Fourier-sparse function with respect to Hamming (ℓ_0) distance. On the algorithmic side, we improve the best known upper bound on the query complexity from $\widetilde{O}(s^{14})$ to $\widetilde{O}(s^4)$, and subsequently to $\widetilde{O}(s)$. Consequently, the gap between our upper bound and the best-known lower bound of $\Omega(s)$ is reduced to polylogarithmic factors. This part of the thesis is based on the following papers:

- **Testing Fourier Sparsity via Implicit Sensing**, with Arijit Ghosh and Subhamoy Maitra. *International Conference on Learning Representations (ICLR)*, 2026.
- **Testing Fourier Sparsity Near Optimally**, with Arijit Ghosh. Submitted.

Estimating Fourier Sparsity in ℓ_2 Distance. We study the problem of deciding, given oracle access to an unknown Boolean function $f : \mathbb{F}_2^n \to \mathbb{R}$, whether it is close to an s-Fourier-sparse function or is far from every such function with respect to ℓ_2 distance. On the algorithmic side, we improve the best known query complexity from $\widetilde{O}(s/\varepsilon^4)$ to $\widetilde{O}(s/\varepsilon^2)$. Consequently, the gap between our upper bound and the best-known lower bound of $\Omega(s)$ is reduced to polylogarithmic factors. More recently, we extended these results to the significantly more challenging distribution-free setting, with only a slight loss in the query complexity. This part of the thesis is based on the following papers:

- **Price of Parsimony: Complexity of Fourier Sparsity Testing**, with Arijit Ghosh. *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.
- **Distribution-Free Testing of Fourier Sparsity**, with Arijit Ghosh. Submitted.

Linear Isomorphism Testing. Two Boolean functions $f, g : \mathbb{F}_2^n \to \{-1, +1\}$ are said to be linearly isomorphic if one can be obtained from the other by an invertible linear change of variables. We study the problem of deciding, given oracle access to an unknown Boolean function f and an explicit target function g, whether f is linearly isomorphic to g or is far from every linear image of g. For target functions with spectral norm at most m, we improve the best previously known upper bound on the query complexity from $\widetilde{O}(m^{24})$ to $\widetilde{O}(m^4)$. Compared with the best-known lower bound of $\Omega(m^2)$, this substantially narrows the gap between the best known upper and lower bounds. We also study this problem in the communication complexity model, where one party holds f and the other holds g. We consider both deterministic and randomized communication protocols. This part of the thesis is based on the following papers:

- **Spectral Norm, Economical Sieve, and Linear Invariance Testing of Boolean Functions**, with Swarnalipa Dutta, Arijit Ghosh, Chandrima Kayal, and Manaswi Paraashar. *STACS*, 2026.
- **Spectral Shadows: When Communication Complexity Meets Linear Invariance Testing**, with Swarnalipa Dutta, Arijit Ghosh, Chandrima Kayal, and Manaswi Paraashar. *Computational Complexity*, 2026.

Estimating Sumset Size. Given a subset $A \subseteq \mathbb{F}_2^n$, it's sumset is defined as $A + A = \{a + b : a \in A, b \in A\}$. We study the problem of estimating the size of the sumset from oracle access to the underlying set (or it's indicator function), assuming that the set has bounded spectral norm. We develop the first polynomial-query algorithm for estimating the sumset size in terms of the spectral norm parameter. This part of the thesis is based on the following paper:

- **Approximating Sumset Size via Dense Coset Decomposition**, with Arijit Ghosh. Submitted.

Fourier Bias Estimation. The Fourier bias of a Boolean function is defined as the magnitude of its largest Fourier coefficient. We study the problem of estimating the Fourier bias from oracle access to the function, up to a prescribed additive error ε. Prior

to this work, the best known upper bound on the query complexity was $\widetilde{O}(n/\varepsilon^3)$. We present an algorithm for this problem with query complexity $\widetilde{O}(1/\varepsilon^2)$, thereby matching the best-known lower bound of $\Omega(1/\varepsilon^2)$ up to polylogarithmic factors. This part of the thesis is based on the following paper:

- **On the Nonlinearity Estimation Problem**, with Arijit Ghosh and Subhamoy Maitra. Submitted.

Exact Recovery of Fourier-Sparse Functions. We study the problem of exactly recovering the Fourier spectrum of an s-Fourier-sparse function $f : \mathbb{F}_2^n \to \mathbb{R}$ from oracle access, that is, identifying all nonzero Fourier coefficients, along with their values, using as few queries as possible. We develop a randomized algorithm for exact Fourier-sparse recovery that exactly recovers the entire Fourier spectrum using $\widetilde{O}(ns)$ queries and $\widetilde{O}(ns)$ running time. Our algorithm applies to general real-valued functions and avoids the heavy algorithmic machinery employed by previous approaches while matching the best known query complexity up to logarithmic factors. This part of the thesis is based on the following paper:

- **Exact Fourier-Sparse Recovery Made Simple**, with Arijit Ghosh. Submitted.

Organization

The thesis is organized into three parts, each centered around a distinct Fourier-sketching technique and its applications.

The first part introduces the coset hashing framework, which serves as the first instantiation of Fourier sketching. After developing the necessary machinery, we apply this framework to a variety of problems. These applications include improved algorithms for testing linear isomorphism, distribution-free Fourier sparsity testing, testing Fourier sparsity in the ℓ_0 distance, and estimating the size of sumsets.

The second part develops a related framework based on subspace restrictions. This approach leads to stronger algorithmic results and, in several cases, nearly optimal bounds. Using this framework, we obtain algorithms for testing Fourier sparsity in both the ℓ_0 and ℓ_2 distances, estimating Fourier bias, and reconstructing Fourier-sparse Boolean functions. In particular, this framework yields near-optimal testers for several fundamental Fourier-analytic properties.

Building on the structural observations developed in the earlier parts of this thesis, the third part investigates how these ideas can be extended to the communication complexity setting. Our goal is to design succinct sketches of Boolean functions that can be efficiently exchanged between communicating parties. As a case study, we consider the linear isomorphism problem, where one party holds f and the other holds g. We study both deterministic and randomized communication protocols for this problem. We believe that the sketching techniques developed in this work are of independent interest and may find applications to a broader class of communication complexity problems.

Taken together, these three parts provide a comprehensive treatment of Fourier sketching, encompassing both the design of efficient algorithms and the complexity-theoretic limitations of the underlying problems. Collectively, they demonstrate the versatility of Fourier sketching as a unifying framework for studying many fundamental problems in the analysis of Boolean functions.

Background

This chapter collects the notation, Fourier-analytic facts, and probabilistic tools used throughout the thesis. Readers familiar with the analysis of Boolean functions may safely skim it and refer back as needed. We adopt the normalized convention throughout: functions are studied with respect to the uniform distribution on the Boolean cube, Boolean functions take values in $\{-1, +1\}$, and expectations replace normalized sums. This convention makes Parseval's identity take its cleanest form and is standard in the analysis of Boolean functions.

Fourier analysis over the Boolean hypercube

We work over the Boolean cube $\mathbb{F}_2^n$. For $x, \alpha \in \mathbb{F}_2^n$, the standard inner product over $\mathbb{F}_2$ is

$$\langle x, \alpha \rangle := \sum_{i=1}^{n} x_i \alpha_i \pmod 2.$$

Throughout, $\mathbb{E}_{x \sim \mathbb{F}_2^n}[\cdot]$ and $\Pr_{x \sim \mathbb{F}_2^n}[\cdot]$ denote expectation and probability with respect to the uniform distribution on $\mathbb{F}_2^n$, unless stated otherwise. For $p \geq 1$ and $g : \mathbb{F}_2^n \to \mathbb{R}$, the normalized L^p norm and inner product are

$$\|g\|_p^p := \mathbb{E}_{x \sim \mathbb{F}_2^n}\big[|g(x)|^p\big], \qquad \langle f, g \rangle := \mathbb{E}_{x \sim \mathbb{F}_2^n}\big[f(x)g(x)\big].$$

Characters and the Fourier expansion. For $\alpha \in \mathbb{F}_2^n$, the *character* (or parity function) $\chi_\alpha : \mathbb{F}_2^n \to \{-1, +1\}$ is defined by

$$\chi_\alpha(x) := (-1)^{\langle \alpha, x \rangle}.$$

Every function $f : \mathbb{F}_2^n \to \mathbb{R}$ admits a unique *Fourier expansion*

$$f(x) = \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)\, \chi_\alpha(x), \qquad \widehat{f}(\alpha) = \langle f, \chi_\alpha \rangle = \mathbb{E}_{x \sim \mathbb{F}_2^n}\big[f(x)\chi_\alpha(x)\big],$$

where $\widehat{f}(\alpha)$ is the *Fourier coefficient* of f at α, and the second equation is the *Fourier inversion formula*. We refer to $\widehat{f}(\alpha)^2$ as the *energy* of the coefficient $\widehat{f}(\alpha)$.

Two elementary facts are used repeatedly. The *character multiplication rule* states that

$$\chi_{\alpha+\beta}(x) = \chi_\alpha(x)\,\chi_\beta(x) \qquad \text{for all } \alpha, \beta, x,$$

and the *character orthogonality relation* states that for any subspace $H \leq \mathbb{F}_2^n$,

$$\mathbb{E}_{x \in H}\big[\chi_\alpha(x)\big] = \begin{cases} 1, & \alpha \in H^\perp, \\ 0, & \text{otherwise.} \end{cases} \tag{0.1}$$

Taking $H = \mathbb{F}_2^n$ in (0.1) gives $\mathbb{E}_x[\chi_\alpha(x)] = \mathbf{1}[\alpha = 0]$.

Parseval's identity. For any $f : \mathbb{F}_2^n \to \mathbb{R}$,

$$\sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)^2 = \|f\|_2^2. \tag{0.2}$$

In particular, if $f : \mathbb{F}_2^n \to \{-1, +1\}$ then $\sum_\alpha \widehat{f}(\alpha)^2 = 1$.

Effect of linear transformations. Let $M : \mathbb{F}_2^n \to \mathbb{F}_2^n$ be an invertible linear transformation. For $f : \mathbb{F}_2^n \to \mathbb{R}$, define $f \circ M$ by $(f \circ M)(x) := f(Mx)$. Then

$$\widehat{f \circ M}(\alpha) = \widehat{f}(M^\top \alpha) \qquad \text{for all } \alpha \in \mathbb{F}_2^n. \tag{0.3}$$

Thus an invertible linear change of variables in the input domain induces the transpose-inverse permutation of the Fourier support, a fact we use throughout the study of linear isomorphism.

Fourier-analytic quantities

Several scalar quantities derived from the Fourier spectrum play a recurring role.

Fourier sparsity. The *Fourier support* of f is

$$\operatorname{supp}(\widehat{f}) := \{\alpha \in \mathbb{F}_2^n : \widehat{f}(\alpha) \neq 0\},$$

and its cardinality $|\operatorname{supp}(\widehat{f})|$ is the *Fourier sparsity* of f. We say f is *s-Fourier-sparse* if $|\operatorname{supp}(\widehat{f})| \leq s$.

Spectral norm. The *spectral norm* (or Fourier ℓ_1 norm) of f is

$$\|\widehat{f}\|_1 := \sum_{\alpha \in \mathbb{F}_2^n} |\widehat{f}(\alpha)|.$$

The spectral norm has been widely studied in connection with circuit complexity, learning theory, and communication complexity [Kushilevitz and Mansour, 1993, Tsang et al., 2013, Shpilka et al., 2017, Girish et al., 2021b].

Approximate spectral norm. To capture approximation, we use the *approximate spectral norm*. For $\gamma \geq 0$, the γ-approximate spectral norm of f is

$$\|\widehat{f}\|_{1,\gamma} := \inf \left\{ \|\widehat{g}\|_1 : g : \mathbb{F}_2^n \to \mathbb{R},\ \forall x \in \mathbb{F}_2^n,\ |f(x) - g(x)| \leq \gamma \right\}.$$

This measures how well f can be approximated pointwise by a real-valued function of small spectral norm. It generalizes the spectral norm and can be significantly smaller. The approximate spectral norm has also appeared in quantum query complexity and communication complexity [Bun and Thaler, 2022, Bansal and Sinha, 2021].

Fourier bias. The *Fourier bias* of f is the magnitude of its largest Fourier coefficient,

$$\max_{\alpha \in \mathbb{F}_2^n} |\widehat{f}(\alpha)|,$$

a quantity closely tied to the nonlinearity of f in the cryptographic literature.

Distance measures

Property testing is phrased in terms of distance to a class of functions. We collect the relevant notions here.

Hamming distance. For $f, g : \mathbb{F}_2^n \to \{-1, +1\}$, the (relative Hamming) *distance* between f and g is

$$\delta(f, g) := \Pr_{x \sim \mathbb{F}_2^n}[f(x) \neq g(x)] = \frac{|\{x \in \mathbb{F}_2^n : f(x) \neq g(x)\}|}{2^n}.$$

It satisfies the triangle inequality $\delta(f, g) \leq \delta(f, h) + \delta(h, g)$ for all Boolean f, g, h. We say f is ε-*far* from a property $\mathcal{P}$ if $\delta(f, g) > \varepsilon$ for every $g \in \mathcal{P}$, and ε-*close* otherwise.

Linear and affine distance. For $f, g : \mathbb{F}_2^n \to \{-1, +1\}$, the *linear distance* is

$$\mathrm{d_A}(f, g) := \min_{M \in \mathrm{GL}_n(\mathbb{F}_2)} \Pr_{x \sim \mathbb{F}_2^n}[f(Mx) \neq g(x)],$$

where $\mathrm{GL}_n(\mathbb{F}_2)$ denotes the group of nonsingular linear transformations of $\mathbb{F}_2^n$. The *affine distance* additionally allows a shift,

$$\mathrm{d_{aff}}(f, g) := \min_{\substack{M \in \mathrm{GL}_n(\mathbb{F}_2) \\ b \in \mathbb{F}_2^n}} \Pr_{x \sim \mathbb{F}_2^n}[f(Mx + b) \neq g(x)].$$

Junta functions. A Boolean function $f : \mathbb{F}_2^n \to \{-1, +1\}$ is a *k-junta* if there exists a set $S \subseteq [n]$ with $|S| \leq k$ such that $f(x)$ depends only on the coordinates $\{x_i : i \in S\}$.

Standing assumption. Throughout, $f : \mathbb{F}_2^n \to \mathbb{R}$ is real-valued with

$$\boxed{\|f\|_2 = 1}$$

and *no* boundedness assumption is made on f.

Subspaces, cosets, and duality

Because the Fourier sketching framework operates through subspaces of $\mathbb{F}_2^n$, we record the linear-algebraic facts on which it relies.

Orthogonal complement. For a subspace $A \leq \mathbb{F}_2^n$, its *orthogonal complement* (also called the *annihilator*) is

$$A^\perp := \big\{\gamma \in \mathbb{F}_2^n : \langle \gamma, \gamma' \rangle = 0 \text{ for all } \gamma' \in A\big\},$$

with $\dim(A) + \dim(A^\perp) = n$ and codimension $\mathrm{codim}(A) = n - \dim(A)$. Unlike over the reals, A and $A^\perp$ need not be independent; for instance, a subspace may intersect its own orthogonal complement nontrivially.

Independence and direct sums. Two subspaces $A, B \leq \mathbb{F}_2^n$ are *independent* if $a + b \neq 0$ for every nonzero $a \in A$ and $b \in B$. If, in addition, every element of $\mathbb{F}_2^n$ can be written uniquely as $a + b$ with $a \in A$ and $b \in B$, we say A and B are *complementary* and write $A \oplus B = \mathbb{F}_2^n$, calling this a *direct sum*. In this case $\dim(A) + \dim(B) = n$. If A and B are

independent, the affine subspaces $b + A$ for distinct $b \in B$ are pairwise disjoint. Cosets of A can be naturally indexed by evaluating a basis of $A^\perp$ on elements $x \in \mathbb{F}_2^n$.

Restriction to affine subspaces. A *random subspace* $H \leq \mathbb{F}_2^n$ of dimension t is the span of t independent, uniformly random vectors $v_1, \ldots, v_t \in \mathbb{F}_2^n$. A *random affine subspace* of dimension t is $A = \alpha + H$, where H is as above and $\alpha \in \mathbb{F}_2^n$ is uniform and independent of H. The *restriction* of f to A is the function

$$f_A : H \to \mathbb{R}, \qquad f_A(x) := f(x + \alpha).$$

Since H is itself an $\mathbb{F}_2$-vector space, f_A has its own Fourier expansion. To index it, fix a subspace W complementary to $H^\perp$, so that $\mathbb{F}_2^n = W \oplus H^\perp$; then $\{\chi_\gamma|_H : \gamma \in W\}$ is an orthonormal basis for the real-valued functions on H. We write $\widehat{f_A}(\gamma)$ for $\gamma \in W$, and note $|W| = 2^t$. Parseval on the group H reads

$$\sum_{\gamma \in W} \widehat{f_A}(\gamma)^2 = \|f_A\|_2^2 = \mathbb{E}_{x \in H}\big[f(x + \alpha)^2\big].$$

Restriction acts as an implicit spectral hashing: each restricted coefficient is a signed aggregate of the original coefficients lying in a single coset of $H^\perp$, and the induced coset partition is pairwise independent. The precise structural statements, together with the stability of heavy coefficients and the preservation of aggregate Fourier mass, are developed in the chapter on the subspace restriction technique; we recall them there in the notation of the present chapter.

Probabilistic tools

The analyses in this thesis rest on a small set of concentration inequalities. We record the elementary ones first, then the amplification device used when the estimator is unbounded.

Concentration inequalities

Markov's inequality. Let X be a non-negative random variable and $a > 0$. Then $\Pr[X \geq a] \leq \mathbb{E}[X]/a$.

Chebyshev's inequality. Let X have mean μ and variance σ^2. Then for any $t > 0$, $\Pr[|X - \mu| \geq t] \leq \sigma^2/t^2$.

Hoeffding's inequality. Let $X_1, \dots, X_n$ be independent random variables with $a_i \le X_i \le b_i$, let $S_n = \sum_{i=1}^n X_i$, and let $\mu = \mathbb{E}[S_n]$. Then for any $t > 0$,

$$\Pr\left[|S_n - \mu| \ge t\right] \le 2\exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Chernoff bound. Let $X_1, \dots, X_m$ be independent Bernoulli random variables with $X_i \in \{0,1\}$ and $\mathbb{E}[X_i] = p_i$. Set $X := \sum_{i=1}^m X_i$ and $\mu := \mathbb{E}[X]$. Then for any $\delta \in (0,1)$,

$$\Pr[X \le (1-\delta)\mu] \le e^{-\delta^2\mu/2} \qquad \text{and} \qquad \Pr[X \ge (1+\delta)\mu] \le e^{-\delta^2\mu/3}.$$

Moment bounds and amplification

The remaining tools handle the unbounded regime: a pair of moment inequalities and the median-of-means amplification. Both are stated under the standing assumption $\|f\|_2 = 1$, in force wherever they are invoked.

Set

$$M := \|\widehat{f}\|_4^4 = \sum_{\beta \in \mathbb{F}_2^n} \widehat{f}(\beta)^4, \qquad K := \|f\|_8^8.$$

By Parseval (0.2) and $\|f\|_2 = 1$ we have $\sum_\beta \widehat{f}(\beta)^2 = 1$, whence

$$|\widehat{f}(\beta)| \le 1 \quad \text{for every } \beta, \qquad \text{and} \qquad M = \sum_\beta \widehat{f}(\beta)^4 \le \Big(\sum_\beta \widehat{f}(\beta)^2\Big)^2 = 1. \tag{0.4}$$

By monotonicity of L^p norms on a probability space,

$$1 = \|f\|_2 \le \|f\|_4 \le \|f\|_8, \qquad \text{hence} \qquad 1 \le \|f\|_4^4 \le \|f\|_8^4 \le \|f\|_8^8 = K. \tag{0.5}$$

In particular $0 \le M \le 1$ and $K \ge 1$. It is worth keeping the two extremes in mind: $K = 1$ for Boolean f, while K can be as large as 2^{3n} for a spike such as $f = 2^{n/2}\mathbf{1}_{\{x_0\}}$.

Median of means

Because f is unbounded, we amplify a weak estimator using the following standard device, whose only input is a variance bound.

Lemma 0.1 (Median of means). *Let Z be a real random variable with $\mathbb{E}[Z] = \mu$ and*

$\mathrm{Var}(Z) \le \sigma^2$. *Fix* $\eta > 0$ *and* $\delta \in (0,1)$*, and set*

$$m := \left\lceil \frac{4\sigma^2}{\eta^2} \right\rceil, \qquad \ell := \left\lceil 12 \ln(1/\delta) \right\rceil.$$

Let $\bar{Z}^{(1)}, \dots, \bar{Z}^{(\ell)}$ *be independent, each the empirical mean of* m *independent copies of* Z*. Then*

$$\Pr\left[\left| \operatorname{median}\left(\bar{Z}^{(1)}, \dots, \bar{Z}^{(\ell)}\right) - \mu \right| \ge \eta \right] \le \delta,$$

using $m\ell = O\big(\sigma^2 \eta^{-2} \log(1/\delta)\big)$ *samples of* Z*.*

Proof. Each block mean $\bar{Z}^{(j)}$ has mean μ and variance at most $\sigma^2/m \le \eta^2/4$, so Chebyshev's inequality gives $\Pr\left[|\bar{Z}^{(j)} - \mu| \ge \eta\right] \le \sigma^2/(m\eta^2) \le 1/4$. Call block j *bad* if $|\bar{Z}^{(j)} - \mu| \ge \eta$, and let B count the bad blocks, so that B is a sum of ℓ independent Bernoulli variables with $\mathbb{E}[B] \le \ell/4$. If the median deviates from μ by at least η, then at least half of the blocks are bad, i.e. $B \ge \ell/2$. By the multiplicative Chernoff bound $\Pr[B \ge (1+\theta)\mathbb{E}B] \le e^{-\theta^2 \mathbb{E}B/3}$ with $\theta = 1$ and $\mathbb{E}B \le \ell/4$,

$$\Pr\left[B \ge \ell/2\right] \le e^{-\ell/12} \le \delta$$

by the choice of ℓ. □

Remark 0.2 (A common pitfall). Median-of-means amplifies an estimator that is *already* correct with probability strictly greater than $1/2$; the Chebyshev step above is what secures this. It is *not* a device for extracting an exponential tail from a second-moment bound at an accuracy where the per-sample failure probability exceeds $1/2$. Arguments in the literature have foundered on exactly this point.

Part I

Fourier Sketching by Coset Hashing

1

The Coset Hashing Technique

In classical signal processing, a signal is often represented through a discrete collection of samples taken at regular intervals. A fundamental challenge in this process is aliasing — the phenomenon by which two or more distinct frequency components become indistinguishable after sampling. Aliasing is regarded as an undesirable artifact and a fundamental failure mode that must be anticipated and prevented. The theoretical basis for understanding and avoiding aliasing is provided by the Nyquist–Shannon sampling theorem. Informally, the theorem states that a signal can be faithfully reconstructed only if it is sampled at a rate at least twice its highest frequency component. This threshold is known as the Nyquist rate. When the sampling rate falls below this limit, the signal becomes undersampled, and aliasing inevitably occurs. The primary consequence of undersampling is a phenomenon known as frequency folding. Frequency components lying above half the sampling frequency do not simply disappear; instead, they re-enter the observable spectrum masquerading as lower-frequency components. As a result, distinct high- and low-frequency signals may appear identical after sampling, producing frequency collisions that make them impossible to distinguish. The sampled spectrum therefore becomes a distorted and misleading representation of the original signal, often exhibiting spurious low-frequency content that was never present in the source.

To mitigate this problem, practical systems typically employ anti-aliasing filters before the sampling stage. These low-pass filters attenuate frequencies above the Nyquist limit, ensuring that only representable frequency components are admitted into the sampled signal.

Although aliasing is traditionally viewed as an undesirable artifact in signal processing, it can also be exploited constructively in several important applications. In particular, controlled aliasing enables the representation of high-frequency information using lower sampling rates, thereby reducing hardware and computational requirements.

One prominent example is bandpass sampling, where a high-frequency band-limited signal is intentionally sampled below the Nyquist rate. By carefully choosing the sampling frequency, the aliased spectrum can be folded into a lower-frequency region without

overlap, allowing the original signal to be recovered. This technique is widely used in radio-frequency (RF) receivers and software-defined radio systems to avoid extremely high-rate analog-to-digital converters.

Aliasing is also intentionally exploited in computer graphics and imaging. Techniques such as dithering and certain texture-mapping methods use controlled aliasing artifacts to create visually smoother or more detailed images under limited resolution constraints. Similarly, in compressed sensing and sparse signal recovery, structured aliasing patterns are sometimes used to encode high-dimensional information into fewer measurements.

Another important setting in which aliasing becomes a constructive computational tool is the analysis of signals with sparse Fourier representations, where only a small number of frequencies carry significant energy while the remaining spectral coefficients are negligible or identically zero. In such situations, recovering the entire spectrum at the full Nyquist rate is often unnecessary and computationally wasteful. Instead, modern sparse Fourier algorithms intentionally introduce carefully designed aliasing patterns through subsampling operations, filtering, or hashing mechanisms in the frequency domain. The central idea is to partition the spectrum into a collection of measurement buckets produced by controlled subsampling. Under subsampling, many distinct frequencies are folded together into the same bucket; however, the subsampling patterns are chosen so that the resulting collisions possess a highly structured and analyzable form. Rather than treating aliasing as corruption, these methods use it as a mechanism for compressing spectral information into a much smaller number of measurements. Because the underlying spectrum is sparse, most buckets either remain empty or contain only a single dominant frequency with high probability. Such isolated frequencies can often be identified and decoded directly from the bucket measurements. Buckets containing collisions are more complicated, but the ambiguity is not permanent: by repeating the process with multiple independently designed subsampling patterns or filters, the collisions can eventually be disentangled. In effect, each subsampling pattern provides a different "view" of the spectrum, and the combined information is sufficient to reconstruct the significant Fourier coefficients using far fewer samples than classical uniform sampling would require. This perspective lies at the heart of many modern sparse Fourier transform (SFT) algorithms and sublinear-time spectral estimation methods. Here, aliasing is no longer merely tolerated — it becomes an algorithmic resource that enables frequency-domain compression, reduced sampling complexity, and computationally efficient recovery of structured signals.

More broadly, aliasing can be interpreted as a mechanism for frequency-domain compression: distinct high-frequency components are mapped into lower-frequency representations. While uncontrolled aliasing destroys information, carefully engineered aliasing can provide efficient representations, reduced sampling complexity, and computational

advantages in modern sensing, communication, and signal recovery systems.

A closely related principle appears in Boolean Fourier analysis through the framework of *coset hashing*. Rather than estimating individual Fourier coefficients one at a time, the objective is to aggregate Fourier information into structured groups and estimate the total spectral mass associated with each group. To the best of our understanding, this perspective first appeared implicitly in the work of Kushilevitz and Mansour [1993] in the context of learning Fourier-sparse Boolean functions. The idea was subsequently refined in the work of Feldman et al. [2006a] on search-to-decision reductions for classes of Fourier-sparse Boolean functions, and later developed systematically by Gopalan et al. Gopalan et al. [2011] in the context of Fourier sparsity testing. Our work follows essentially the same line of ideas as that of Gopalan et al. [2011], and we adopt their notation and framework. We now define these notions more formally.

Let $f : \mathbb{F}_2^n \to \mathbb{R}$ be a function with Fourier expansion $f = \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)\chi_\alpha$, where $\chi_\alpha(x) = (-1)^{\langle \alpha, x \rangle}$, and let H be a subspace of $\mathbb{F}_2^n$. Throughout, we use the (expectation) convolution convention

$$(g * f)(x) \;=\; \mathbb{E}_{y \in \mathbb{F}_2^n}\big[g(y)\, f(x+y)\big].$$

One then defines a projection operator P_{r+H} acting on functions $f : \mathbb{F}_2^n \to \mathbb{R}$, whose action in the Fourier domain is simply to retain the coefficients lying inside the coset $r + H$ and zero out all others:

$$\widehat{\mathsf{P}_{r+H} f}(\alpha) = \begin{cases} \widehat{f}(\alpha), & \alpha \in r + H, \\ 0, & \text{otherwise.} \end{cases}$$

Equivalently, this projection may be expressed as a convolution,

$$\mathsf{P}_{r+H} f = A_{r+H} * f, \qquad A_{r+H} = \sum_{\alpha \in r+H} \chi_\alpha = \chi_r \sum_{h \in H} \chi_h,$$

which makes clear that P_{r+H} acts as a spectral filter that passes precisely the bucket $r + H$ while suppressing everything outside it.

The Fourier mass contained in the bucket is then defined as

$$\mathsf{W}_{r+H} f := \sum_{\alpha \in r+H} \widehat{f}(\alpha)^2,$$

which measures the total spectral energy concentrated inside the coset $r + H$.

It is useful to emphasize that this construction is fundamentally a structured form of aliasing: Fourier coefficients are grouped together modulo a subspace structure, causing multiple frequencies to collapse into a common measurement bucket. In principle, many other forms of aliasing-based grouping are possible, and one need not restrict attention exclusively to subspace-induced partitions. However, for algorithmic purposes, the crucial requirement is not the particular geometry of the buckets, but rather the ability to efficiently estimate the aggregate quantities associated with them. In the present framework, the relevant statistics are the projection operators $\mathsf{P}_{r+H} f$ and the bucket weights $\mathsf{W}_{r+H} f$. The subspace structure is especially convenient because these quantities admit explicit algebraic formulas and can be estimated efficiently using random samples.

We now show that the subspace structure permits efficient estimation of both the projection operator and the Fourier mass associated with a bucket.

Recall that

$$A_{r+H} = \sum_{\alpha \in r+H} \chi_\alpha.$$

Since every element of the coset $r + H$ may be written uniquely as $r + h$ for some $h \in H$, we have

$$A_{r+H} = \chi_r \sum_{h \in H} \chi_h.$$

Using the standard orthogonality identity $\sum_{h \in H} \chi_h = |H| \cdot \mathbf{1}_{H^\perp}$, it follows that

$$A_{r+H}(x) = |H|\, \chi_r(x)\, \mathbf{1}_{H^\perp}(x).$$

Fact 1.1. *For every $x \in \mathbb{F}_2^n$,*

$$\mathsf{P}_{r+H} f(x) = \mathbb{E}_{y \in H^\perp} \big[\chi_r(y)\, f(x+y) \big].$$

Proof. Using the convolution representation $\mathsf{P}_{r+H} f = A_{r+H} * f$ with the expectation convention above,

$$\begin{aligned} \mathsf{P}_{r+H} f(x) &= \mathbb{E}_{y \in \mathbb{F}_2^n} \big[A_{r+H}(y)\, f(x+y) \big] \\ &= |H|\, \mathbb{E}_{y \in \mathbb{F}_2^n} \big[\chi_r(y)\, \mathbf{1}_{H^\perp}(y)\, f(x+y) \big] \\ &= \frac{|H|}{2^n} \sum_{y \in H^\perp} \chi_r(y)\, f(x+y). \end{aligned}$$

Since $|H| \cdot |H^\perp| = 2^n$, we have $|H|/2^n = 1/|H^\perp|$, and the claim follows. □

Proposition 1.2. *Suppose $f : \mathbb{F}_2^n \to \{-1, +1\}$. For every $x \in \mathbb{F}_2^n$, the quantity $\mathsf{P}_{r+H} f(x)$*

can be estimated to within additive error τ with confidence at least $1-\delta$ using $O(\log(1/\delta)/\tau^2)$ queries to f.

Proof. Empirically estimate the expectation in Fact 1.1. Since $\chi_r(y)f(x+y) \in \{-1,+1\}$, the estimator is bounded in $[-1,1]$. The result therefore follows from Hoeffding's inequality. □

Recall that the Fourier mass inside the bucket $r+H$ is

$$\mathsf{W}_{r+H}f := \sum_{\alpha \in r+H} \widehat{f}(\alpha)^2 = \|\mathsf{P}_{r+H}f\|_2^2.$$

Fact 1.3. *For every coset $r+H$,*

$$\mathsf{W}_{r+H}f = \mathbb{E}_{x\in\mathbb{F}_2^n,\, z\in H^\perp}\big[\chi_r(z)\, f(x)\, f(x+z)\big].$$

Proof. Using Parseval's identity together with Fact 1.1,

$$\begin{aligned}\mathsf{W}_{r+H}f &= \mathbb{E}_w\big[(\mathsf{P}_{r+H}f(w))^2\big] \\ &= \mathbb{E}_{w,\, y_1,y_2\in H^\perp}\big[\chi_r(y_1)f(w+y_1)\,\chi_r(y_2)f(w+y_2)\big].\end{aligned}$$

Since $\chi_r(y_1)\chi_r(y_2) = \chi_r(y_1+y_2)$, setting $x = w+y_1,\ z = y_1+y_2$ (so that $z \in H^\perp$ and, for fixed z, x is uniform) yields the desired identity. □

Proposition 1.4. *Suppose $f : \mathbb{F}_2^n \to \{-1,+1\}$. For every coset $r+H$, the quantity $\mathsf{W}_{r+H}f$ can be estimated to within additive error τ with confidence at least $1-\delta$ using $O(\log(1/\delta)/\tau^2)$ queries to f.*

Proof. Empirically estimate the expectation in Fact 1.3. Since $\chi_r(z)f(x)f(x+z) \in \{-1,+1\}$, the estimator is bounded in $[-1,1]$. The result again follows from Hoeffding's inequality. □

Thus, the subspace-based aliasing structure not only compresses Fourier information into buckets, but also permits efficient and statistically stable estimation of the aggregated spectral mass associated with each bucket. This efficient estimability is one of the central reasons why coset hashing has become such a powerful primitive in sparse Fourier algorithms and Fourier property testing.

We now describe the coset hashing construction due to Gopalan et al. [2011]. Select vectors $\{\beta_1,\dots,\beta_t\}$ independently and uniformly at random from $\mathbb{F}_2^n$, and define the subspace $H := \operatorname{span}\{\beta_1,\dots,\beta_t\}^\perp$. With high probability, the vectors $\beta_1,\dots,\beta_t$ are linearly independent, and hence H has codimension t. For each binary string $b \in \mathbb{F}_2^t$,

define the affine coset

$$D(b) := \Big\{\alpha \in \mathbb{F}_2^n : \langle \alpha, \beta_i \rangle = b_i \text{ for all } i\Big\}.$$

The collection $\{D(b) : b \in \mathbb{F}_2^t\}$ forms a partition of $\mathbb{F}_2^n$ into affine cosets of H. To introduce an additional layer of randomness, one draws a shift $z \in \mathbb{F}_2^t$ uniformly at random and relabels each bucket $D(b)$ as $D(b+z)$, producing a randomly permuted coset structure.

Gopalan et al. [2011] showed that the above construction induces a pairwise independent random partitioning of the Fourier space.

Proposition 1.5. *Let $(H, \mathcal{D})$ denote a randomly permuted coset structure of codimension t, where $\mathcal{D} = \{D(b) : b \in \mathbb{F}_2^t\}$. Then for every distinct $\alpha, \alpha' \in \mathbb{F}_2^n$,*

$$\Pr[\alpha, \alpha' \textit{ collide}] = 2^{-t}.$$

Proof. Two vectors α, α' collide precisely when

$$\langle \alpha - \alpha', \beta_i \rangle = 0 \qquad \text{for all } i.$$

Since $\alpha \neq \alpha'$, the vector $\alpha - \alpha'$ is nonzero, and each constraint holds independently with probability $1/2$. Hence

$$\Pr[\alpha, \alpha' \text{ collide}] = 2^{-t}.$$

□

As an immediate consequence, if f is an s-Fourier sparse function and $t \geq 2\log s + O(1)$, then with constant probability no two nonzero Fourier coefficients of f collide under the random coset partition. Equivalently, every nonzero Fourier coefficient is isolated inside its own bucket. In particular, choosing $t = 2\log s + O(1)$ produces only

$$2^t = O(s^2)$$

buckets, yet still isolates all nonzero coefficients with high probability.

Gopalan et al. [2011] work on Fourier sparsity testing, as well as several subsequent works, relies heavily on this hashing framework together with efficient estimation of bucket weights in order to infer structural information about the Fourier spectrum and achieve various algorithmic objectives. For example, in the context of Fourier sparsity testing, Gopalan et al. [2011] established structural results showing that every Fourier-sparse Boolean function must contain at least one Fourier coefficient of comparatively large magnitude. Their testing algorithm then proceeds in two stages. First, a random

coset hashing scheme is applied in order to probabilistically isolate the nonzero Fourier coefficients into separate buckets. Once the coefficients are sufficiently separated, the algorithm estimates the Fourier mass contained inside each bucket to detect the presence of a large coefficient. Finally, by checking whether the estimated bucket weights are consistent with the structural constraints satisfied by Boolean functions, the tester determines whether the underlying function is close to an s-Fourier sparse Boolean function or is far from every such function.

Our work extends this line of ideas by establishing significantly stronger concentration properties for the coset hashing process. Rather than merely controlling pairwise collisions among Fourier coefficients, we analyze how Fourier mass distributes across random buckets and prove quantitative concentration bounds governing these aggregated spectral measurements. These results allow several algorithmic tasks involving Fourier-sparse functions to be carried out substantially more query-efficiently, yielding significant improvements over previously known algorithms and analyses.

Notation	Description
$\mathscr{C}$	A random coset decomposition of $\mathbb{F}_2^n$
C	Individual cosets of $\mathscr{C}$
$\mathcal{C}$	Set of cosets of $\mathscr{C}$ that contain a *heavy* Fourier coefficient of f
α_C	The Fourier coefficient in C with the highest absolute value
$\mathsf{W}_C f$	$\sum_{\beta\in C}\widehat{f}(\beta)^2$, total Fourier weight of coset C
$\mathsf{W}^*_C f$	$\widehat{f}(\alpha_C)^2$, weight of the heaviest Fourier coefficient in C
$\mathsf{P}_C f(z)$	$\sum_{\beta\in C}\widehat{f}(\beta)\chi_\beta(z)$, projection of f onto coset C
$\mathsf{P}^*_C f(z)$	$\widehat{f}(\alpha_C)\chi_{\alpha_C}(z)$, projection of f onto the heaviest coefficient of C
$\chi_C(x)$	Alternative notation for $\chi_{\alpha_C}(x)$

Table 1.1: Notation for Algorithm 2 and the surrounding results.

Claim 1.6. *Let $\mathscr{C}$ be a randomly permuted coset structure of codimension $t = \lceil \log \frac{2^{40}}{\theta^8} \rceil$. Then, except with probability at most $\frac{1}{8}$, we have*

(i) ***(Isolation.)*** *Every Fourier coefficient α with $|\widehat{f}(\alpha)| \geq \frac{\theta^2}{64}$ is mapped to a distinct coset. Consequently, for every $C \in \mathscr{C}$ the second-heaviest coefficient satisfies*

$$\max_{\beta\in C\setminus\{\alpha_C\}}\left|\widehat{f}(\beta)\right| < \frac{\theta^2}{64}, \qquad \alpha_C := \arg\max_{\beta\in C}\left|\widehat{f}(\beta)\right|.$$

(ii) ***(Heavy cosets.)*** *For every coset C whose dominant coefficient satisfies $|\widehat{f}(\alpha_C)| \geq \frac{\theta^2}{64}$, the coset weight satisfies*

$$\mathsf{W}_C f \leq \widehat{f}(\alpha_C)^2 + \frac{\theta^4}{2^{12}}.$$

(iii) ***(Light cosets.)*** *For every coset C whose dominant coefficient satisfies $|\widehat{f}(\alpha_C)| < \frac{\theta^2}{64}$, the coset weight satisfies*

$$\mathsf{W}_C f < \frac{\theta^2}{16}.$$

Proof. Throughout, let $S = \{\alpha : |\widehat{f}(\alpha)| \geq \theta^2/64\}$ be the set of heavy coefficients. Since $f : \mathbb{F}_2^n \to \{-1, +1\}$ has $\sum_\alpha \widehat{f}(\alpha)^2 = 1$ by Parseval, we have

$$|S| \leq \frac{1}{(\theta^2/64)^2} = \frac{2^{12}}{\theta^4}.$$

(i) Isolation of heavy coefficients. Fix $\alpha \in \mathbb{F}_2^n$ and $b \in \mathbb{F}_2^t$. By definition, $\alpha \in D(b+z)$ iff $\langle \alpha, \beta_i \rangle = b_i + z_i$ for all $i \in [t]$. Since each z_i is an independent uniform bit,

$$\Pr[\alpha \in D(b+z)] = \prod_{i=1}^{t} \Pr[z_i = \langle \alpha, \beta_i \rangle - b_i] = 2^{-t}.$$

For distinct α, α', Proposition 1.5 gives collision probability 2^{-t}. Applying a union bound over all $\binom{|S|}{2}$ pairs,

$$\binom{|S|}{2} 2^{-t} \;\leq\; |S|^2\, 2^{-t} \;\leq\; \left(\tfrac{2^{12}}{\theta^4}\right)^2 \cdot \frac{\theta^8}{2^{40}} \;=\; \frac{2^{24}}{2^{40}} \;=\; 2^{-16} \;<\; \frac{1}{300},$$

using $2^{-t} \leq \theta^8/2^{40}$. Hence with probability at least $1 - \frac{1}{300}$ all coefficients in S are isolated into distinct cosets.

(ii) Heavy cosets. Fix a coset C containing a unique heavy coefficient $\alpha_C \in S$ (guaranteed by part (i)). Write

$$\mathsf{W}_C f = \widehat{f}(\alpha_C)^2 + \sum_{\beta \neq \alpha_C} \widehat{f}(\beta)^2 I_\beta,$$

where I_β indicates that β hashes to C. Since $\mathbb{E}[I_\beta] = 2^{-t}$,

$$\mathbb{E}\Big[\mathsf{W}_C f - \widehat{f}(\alpha_C)^2\Big] = 2^{-t} \sum_{\beta \neq \alpha_C} \widehat{f}(\beta)^2 \;\leq\; 2^{-t} \;\leq\; \frac{\theta^8}{2^{40}}.$$

By Markov's inequality,

$$\Pr\left[\mathsf{W}_C f - \widehat{f}(\alpha_C)^2 \geq \frac{\theta^4}{2^{12}}\right] \leq \frac{\theta^8/2^{40}}{\theta^4/2^{12}} = \frac{\theta^4}{2^{28}}.$$

There are at most $|S| \leq 2^{12}/\theta^4$ such cosets, so a union bound gives failure probability at most $\frac{2^{12}}{\theta^4} \cdot \frac{\theta^4}{2^{28}} = 2^{-16} < \frac{1}{300}$.

(iii) Light cosets. Let C be a coset with $|\widehat{f}(\beta)| < \theta^2/64$ for all $\beta \in C$. Then

$$\mathbb{E}[\mathsf{W}_C f] = 2^{-t} \sum_{\beta \in C} \widehat{f}(\beta)^2 \le 2^{-t} \le \frac{\theta^8}{2^{40}}.$$

Using pairwise independence of the I_β,

$$\mathrm{Var}(\mathsf{W}_C f) = \sum_{\beta} \widehat{f}(\beta)^4 \, \mathrm{Var}(I_\beta) \le \sum_{\beta} \widehat{f}(\beta)^4 \, \mathbb{E}[I_\beta] = 2^{-t} \sum_{\beta} \widehat{f}(\beta)^4.$$

Since $|\widehat{f}(\beta)| < \theta^2/64$ implies $\widehat{f}(\beta)^4 \le \frac{\theta^4}{2^{12}} \widehat{f}(\beta)^2$,

$$\sum_{\beta} \widehat{f}(\beta)^4 \le \frac{\theta^4}{2^{12}} \sum_{\beta} \widehat{f}(\beta)^2 \le \frac{\theta^4}{2^{12}},$$

so $\mathrm{Var}(\mathsf{W}_C f) \le 2^{-t} \cdot \frac{\theta^4}{2^{12}}$. By Chebyshev's inequality (using $\mathbb{E}[\mathsf{W}_C f] \le \theta^8/2^{40} \le \theta^2/32$, so a deviation of $\theta^2/16$ from the mean suffices to reach $\theta^2/16$),

$$\Pr\left[\mathsf{W}_C f \ge \frac{\theta^2}{16}\right] \le \frac{\mathrm{Var}(\mathsf{W}_C f)}{(\theta^2/16)^2} \le \frac{2^{-t}\,\theta^4/2^{12}}{\theta^4/2^8} = 2^{-t} \cdot 2^{-4}.$$

There are 2^t cosets in total, so by a union bound the total failure probability for part (iii) is at most

$$2^t \cdot 2^{-t} \cdot 2^{-4} = 2^{-4} = \frac{1}{16}.$$

Combining the three parts by a union bound, the total failure probability is at most $\frac{1}{300} + \frac{1}{300} + \frac{1}{16} < \frac{1}{8}$, as claimed. □

Claim 1.7. *Suppose $f : \mathbb{F}_2^n \to \mathbb{R}$, and let $C \in \mathscr{C}$ be a coset with dominant Fourier coefficient α_C such that*

$$\textit{(i)} \sum_{\beta \in C \setminus \{\alpha_C\}} \widehat{f}(\beta)^2 < \tfrac{\theta^4}{2^{12}}, \qquad \textit{(ii)}\ |\widehat{f}(\alpha_C)| > \tfrac{\theta}{2}, \qquad \textit{(iii)} \max_{\beta \in C \setminus \{\alpha_C\}} |\widehat{f}(\beta)| < \tfrac{\theta^2}{64}.$$

Then

$$\Pr_{z \sim \mathbb{F}_2^n}\left[|\mathsf{P}_C f(z) - \mathsf{P}_C^* f(z)| \ge \frac{\theta^2}{64} + \tau\right] \le \frac{\theta^4}{2^{12}\tau^2}.$$

Proof. We first recall that the Fourier characters act as pairwise independent ± 1-valued hash functions. When z is uniform on $\mathbb{F}_2^n$, for any nonzero α,

$$\mathop{\mathbb{E}}_{z}[\chi_\alpha(z)] = \frac{1}{2}(1) + \frac{1}{2}(-1) = 0,$$

and for distinct nonzero α_1, α_2,

$$\mathbb{E}_z[\chi_{\alpha_1}(z)\chi_{\alpha_2}(z)] = \mathbb{E}_z[\chi_{\alpha_1+\alpha_2}(z)] = 0 = \mathbb{E}_z[\chi_{\alpha_1}(z)]\,\mathbb{E}_z[\chi_{\alpha_2}(z)],$$

establishing pairwise independence.

Write $C' = C \setminus \{\alpha_C\}$, so that $\mathsf{P}_C f(z) - \mathsf{P}^*_C f(z) = \sum_{\beta \in C'} \widehat{f}(\beta)\chi_\beta(z)$. We bound the mean and variance of this quantity.

Mean. If $\mathbf{0} \notin C'$, then every $\beta \in C'$ is nonzero and

$$\mathbb{E}_z\Big[\sum_{\beta \in C'} \widehat{f}(\beta)\chi_\beta(z)\Big] = \sum_{\beta \in C'} \widehat{f}(\beta) \cdot 0 = 0.$$

If $\mathbf{0} \in C'$ (i.e. $\mathbf{0} \in C$ and $\mathbf{0} \neq \alpha_C$), then only the $\beta = \mathbf{0}$ term has nonzero expectation ($\chi_{\mathbf{0}} \equiv 1$), giving

$$\Big|\mathbb{E}_z\Big[\sum_{\beta \in C'} \widehat{f}(\beta)\chi_\beta(z)\Big]\Big| = \big|\widehat{f}(\mathbf{0})\big| < \frac{\theta^2}{64},$$

where the last inequality is hypothesis (iii) applied to the non-leader $\mathbf{0} \in C'$. In either case $\big|\mathbb{E}_z[\mathsf{P}_C f(z) - \mathsf{P}^*_C f(z)]\big| < \theta^2/64$.

Variance. By pairwise independence of the characters,

$$\mathrm{Var}_z\Big[\sum_{\beta \in C'} \widehat{f}(\beta)\chi_\beta(z)\Big] = \sum_{\beta \in C'} \mathrm{Var}_z\big[\widehat{f}(\beta)\chi_\beta(z)\big] \le \sum_{\beta \in C'} \widehat{f}(\beta)^2 < \frac{\theta^4}{2^{12}},$$

using hypothesis (i).

Applying Chebyshev's inequality around the mean and absorbing the mean-shift of at most $\theta^2/64$ into the threshold,

$$\Pr_z\Big[|\mathsf{P}_C f(z) - \mathsf{P}^*_C f(z)| \ge \tfrac{\theta^2}{64} + \tau\Big] \le \frac{\mathrm{Var}_z\big[\mathsf{P}_C f(z) - \mathsf{P}^*_C f(z)\big]}{\tau^2} \le \frac{\theta^4}{2^{12}\tau^2}.$$

This completes the proof. □

To apply these concentration results across several algorithmic problems, we encapsulate them in a unified framework that simplifies the presentation while preserving the essential ideas. As a first application, we consider the problem of local list correction for the Hadamard code.

1.1 Local List Correction over the Hadamard Code

A code is a mapping $C : \Sigma^k \to \Sigma^n$, where Σ is an alphabet, k is the message length, and n is the codeword length. The image $\mathcal{C} = \{C(x) : x \in \Sigma^k\}$ is called the *codebook*. The redundancy introduced by encoding is precisely what enables reliable recovery even when the transmitted word has been corrupted by noise. The error-correcting capability of a code is governed primarily by its *minimum distance*, namely the minimum Hamming distance between distinct codewords.

In the classical setting, a decoder attempts to reconstruct the entire codeword closest to the received word. This succeeds as long as the corruption level remains below the unique decoding radius. Beyond this threshold, however, several codewords may simultaneously lie close to the received word, making unique recovery impossible.

The notion of *list decoding* was introduced to address precisely this regime. Given a received word r, a list decoder outputs all codewords within distance at most ρn from r, namely

$$\big\{c \in \mathcal{C} : \Delta(c, r) \leq \rho n\big\},$$

where $\Delta(\cdot, \cdot)$ denotes Hamming distance. The key requirement is that the resulting list remain small, typically polynomially bounded in the relevant parameters.

Local correction addresses a slightly weaker objective. Rather than reconstructing the entire codeword, the goal is only to recover a requested coordinate of the underlying codeword. More precisely, given oracle access to a corrupted word g that is close to some codeword $c \in \mathcal{C}$, a local correction algorithm, upon receiving an index i, outputs the i-th symbol c_i while making only a small number of queries to g.

Local List correction combines these two notions. When the received oracle may be close to several codewords simultaneously, the algorithm outputs a small list of correction procedures

$$\mathcal{A}_1, \ldots, \mathcal{A}_\ell,$$

where each procedure corresponds to one nearby codeword. Given an index i, the procedure $\mathcal{A}_j(i)$ recovers the i-th coordinate of the associated codeword using only a small number of queries to the corrupted oracle. Thus, local list correction permits efficient recovery of individual symbols even in highly corrupted settings where several candidate codewords coexist.

We now briefly recall the Hadamard code, which admits a particularly clean Fourier-analytic interpretation. For a vector $a \in \mathbb{F}_2^n$, the Hadamard encoding of a is the Boolean

function

$$\mathrm{Had}_a(x) = (-1)^{\langle a,x\rangle} = \chi_a(x).$$

Thus every Hadamard codeword is simply a Fourier character over $\mathbb{F}_2^n$, and the Hadamard codebook is precisely $\{\chi_a : a \in \mathbb{F}_2^n\}$.

Consequently, for any Boolean function $f : \mathbb{F}_2^n \to \{-1, +1\}$, the Fourier coefficient

$$\widehat{f}(a) = \mathbb{E}_x[f(x)\chi_a(x)]$$

measures the correlation between f and the Hadamard codeword χ_a. Large Fourier coefficients therefore correspond exactly to Hadamard codewords having substantial agreement with the received oracle.

This correspondence allows the coset hashing framework developed earlier to be interpreted naturally in coding-theoretic terms. Random coset partitions probabilistically isolate heavy Fourier coefficients into separate buckets, while bucket-weight estimation identifies buckets containing significant correlation mass. These isolated heavy coefficients then serve as candidates for the locally correctable Hadamard codewords returned by the list-correction procedure.

Lemma 1.8. *Given oracle access to $f : \mathbb{F}_2^n \to \{-1, +1\}$ and a threshold θ, the algorithm Fourier-Sieve (Algorithm 1) identifies, except with probability at most $1/8$, a set of cosets $\mathcal{C} \subseteq \mathscr{C}$ satisfying the following properties, using $\tilde{O}(1/\theta^4)$ queries to f:*

- *For any $C \in \mathscr{C}$, if $|\widehat{f}(\alpha_C)| \geq \theta$, then $C \in \mathcal{C}$.*
- *For every $C \in \mathcal{C}$, $|\widehat{f}(\alpha_C)| \geq \frac{\theta}{2}$.*
- *For every $C \in \mathcal{C}$, $\sum_{\beta \in C \setminus \{\alpha_C\}} \widehat{f}(\beta)^2 < \frac{\theta^4}{2^{12}}$.*

Here $\mathscr{C}$ is a partition of $\mathbb{F}_2^n$ into cosets of a subspace of codimension $\log \frac{2^{40}}{\theta^8}$, and for each coset $C \in \mathscr{C}$ the dominant Fourier character is $\alpha_C := \arg\max_{\alpha \in C} |\widehat{f}(\alpha)|$.

Algorithm 1: Fourier-Sieve

Input: Threshold θ, oracle access to $f : \mathbb{F}_2^n \to \{-1, +1\}$

Step 1: Choose a random subspace $H \leq \mathbb{F}_2^n$ of codimension $t = \log(2^{40}/\theta^8)$. Let $\mathscr{C}$ denote the collection of cosets of H.

Step 2: For every $C \in \mathscr{C}$, estimate the Fourier weight $\mathsf{W}_C f$ within additive error $\pm\theta^2/4$.

Step 3: Discard all cosets whose estimated weight is at most $\frac{3}{4}\theta^2$.

Output: The set of remaining cosets $\mathcal{C}$

Proof. [Proof of Lemma 1.8] Let $\mathcal{E}$ denote the event that the conclusions of Claim 1.6 hold *and* every weight estimate $\widetilde{\mathsf{W}}_C f$ is accurate to within $\pm\theta^2/4$. Both hold except with probability at most $1/8 + o(1)$; we condition on $\mathcal{E}$.

(i) All cosets containing heavy coefficients are selected. Let C satisfy $|\widehat{f}(\alpha_C)| \geq \theta$. Then $\mathsf{W}_C f \geq \widehat{f}(\alpha_C)^2 \geq \theta^2$, so under $\mathcal{E}$,

$$\widetilde{\mathsf{W}}_C f \geq \mathsf{W}_C f - \frac{\theta^2}{4} \geq \frac{3\theta^2}{4},$$

and C is retained.

(ii) No coset with a small dominant coefficient is selected. Let $|\widehat{f}(\alpha_C)| \leq \theta/2$.

Case 1: $|\widehat{f}(\alpha_C)| < \theta^2/64$. By Claim 1.6 (iii) (light cosets), $\mathsf{W}_C f < \frac{\theta^2}{16}$. Hence under $\mathcal{E}$,

$$\widetilde{\mathsf{W}}_C f \leq \mathsf{W}_C f + \frac{\theta^2}{4} < \frac{\theta^2}{16} + \frac{\theta^2}{4} = \frac{5\theta^2}{16} < \frac{3\theta^2}{4},$$

so C is discarded.

Case 2: $\theta^2/64 \leq |\widehat{f}(\alpha_C)| \leq \theta/2$. By Claim 1.6 (ii) (heavy cosets),

$$\mathsf{W}_C f \leq \widehat{f}(\alpha_C)^2 + \frac{\theta^4}{2^{12}} \leq \frac{\theta^2}{4} + \frac{\theta^4}{2^{12}} < \frac{5\theta^2}{16}.$$

Hence under $\mathcal{E}$,

$$\widetilde{\mathsf{W}}_C f \leq \mathsf{W}_C f + \frac{\theta^2}{4} < \frac{5\theta^2}{16} + \frac{\theta^2}{4} = \frac{9\theta^2}{16} < \frac{3\theta^2}{4},$$

and C is not selected.

Combining (i) and (ii): every retained coset has $|\widehat{f}(\alpha_C)| > \theta/2$, and every coset with $|\widehat{f}(\alpha_C)| \geq \theta$ is retained.

(iii) Concentration of coset weight around the dominant coefficient. For any retained coset, $|\widehat{f}(\alpha_C)| > \theta/2 \geq \theta^2/64$, so by part (i) of Claim 1.6 the coset contains no other heavy coefficient, and by part (ii), $\sum_{\beta \in C \setminus \{\alpha_C\}} \widehat{f}(\beta)^2 = \mathsf{W}_C f - \widehat{f}(\alpha_C)^2 \leq \theta^4/2^{12}$.

Finally, estimating the weight of each of the $2^t = O(1/\theta^8)$ cosets to within $\theta^2/4$ with confidence $1 - \mathrm{poly}(\theta)$ requires, by Proposition 1.4, $\widetilde{O}(1/\theta^4)$ queries to f in total. □

We now describe a variant of the list-correction procedure tailored to our setting, which operates on heavy Fourier coefficients that have been implicitly identified through a collection of cosets. The procedure is formalized in List-Correction (Algorithm 2), while Lemma 1.9 establishes its theoretical guarantees.

Algorithm 2: LIST-CORRECTION

Input: A list of cosets $\mathcal{C}$ satisfying Lemma 1.8; oracle access to $f : \mathbb{F}_2^n \to \{-1, +1\}$; query point $x \in \mathbb{F}_2^n$; confidence δ'

Step 1: Sample $y_1, \ldots, y_\gamma$ independently and uniformly from $\mathbb{F}_2^n$, where $\gamma = O(\log(1/\delta')/\theta^2)$.

Step 2: For each $C \in \mathcal{C}$ and each $j \in [\gamma]$, estimate $\mathsf{P}_C f(y_j)$ and $\mathsf{P}_C f(x + y_j)$ within additive error $\pm\theta/8$.

Step 3: For each $C \in \mathcal{C}$, set $\widetilde{\chi}_{\alpha_C}(x) = \mathrm{median}_{j\in[\gamma]} \left(\mathrm{sign}(\widetilde{\mathsf{P}}_C f(y_j))\ \mathrm{sign}(\widetilde{\mathsf{P}}_C f(x + y_j))\right)$.

Output: $\{\widetilde{\chi}_{\alpha_C}(x) : C \in \mathcal{C}\}$

Lemma 1.9. *Let $\mathcal{C}$ be the collection of cosets guaranteed by Lemma 1.8. Then, for any $x \in \mathbb{F}_2^n$ and oracle access to f, LIST-CORRECTION (Algorithm 2) outputs values $s_C \in \{+1, -1\}$ for all $C \in \mathcal{C}$ such that*

$$\Pr\left[s_C = \chi_{\alpha_C}(x) \text{ for all } C \in \mathcal{C}\right] \geq 1 - \delta'.$$

Moreover, the query complexity of the algorithm is $\widetilde{O}(\log(1/\delta')/\theta^2)$.

The proof of Lemma 1.9 crucially relies on the fact that a noisy variant of the standard self-correction procedure, captured in Claim 1.10, succeeds simultaneously for all cosets returned by FOURIER-SIEVE (Algorithm 1).

Claim 1.10. *Let $\mathcal{C}$ be the set of cosets returned by Algorithm 1. Then, for any fixed $x \in \mathbb{F}_2^n$ and for all $C \in \mathcal{C}$ simultaneously,*

$$\Pr_{y\sim\mathbb{F}_2^n}\left[\mathrm{sign}\big(\mathsf{P}_C f(x + y)\big) \cdot \mathrm{sign}\big(\mathsf{P}_C f(y)\big) = \chi_{\alpha_C}(x) \ \text{ for all } C \in \mathcal{C}\right] \geq \frac{7}{8}.$$

Proof. We apply Claim 1.7 with $\tau = \theta/8$. Each retained coset C satisfies its three hypotheses by Lemma 1.8 (in particular $|\widehat{f}(\alpha_C)| > \theta/2$ and $\sum_{\beta\in C'} \widehat{f}(\beta)^2 < \theta^4/2^{12}$). Thus for each fixed C,

$$\Pr_z\left[\left|\mathsf{P}_C f(z) - \mathsf{P}^*_C f(z)\right| \geq \tfrac{\theta^2}{64} + \tfrac{\theta}{8}\right] \leq \frac{\theta^4}{2^{12}(\theta/8)^2} = \frac{\theta^2}{64}.$$

Since $\theta^2/64 + \theta/8 \leq \theta/4$ for $\theta \leq 1$, this gives $\Pr_z[|\mathsf{P}_C f(z) - \mathsf{P}^*_C f(z)| \geq \theta/4] \leq \theta^2/64$.

Because $f : \mathbb{F}_2^n \to \{-1, +1\}$, the number of Fourier coefficients of magnitude at least $\theta/2$ is at most $4/\theta^2$ (Parseval). As each surviving coset contains a unique dominant coefficient of magnitude $\geq \theta/2$, we have $|\mathcal{C}| \leq 4/\theta^2$. A union bound over surviving cosets, applied

at both evaluation points $z = y$ and $z = x + y$ (each uniform when y is uniform), yields

$$\Pr_y\left[\exists\, C \in \mathcal{C} :\ \left|\mathsf{P}_C f(y) - \mathsf{P}^*_C f(y)\right| \geq \tfrac{\theta}{4} \text{ or } \left|\mathsf{P}_C f(x+y) - \mathsf{P}^*_C f(x+y)\right| \geq \tfrac{\theta}{4}\right] \leq 2 \cdot \frac{4}{\theta^2} \cdot \frac{\theta^2}{64} = \frac{1}{8}.$$

On the complementary event, for every surviving C and both points, $\mathsf{P}_C f$ is within $\theta/4$ of $\mathsf{P}^*_C f = \widehat{f}(\alpha_C)\chi_{\alpha_C}$, whose magnitude is $|\widehat{f}(\alpha_C)| > \theta/2$. Since $\theta/4 < \theta/2$, the perturbation cannot flip the sign, so $\operatorname{sign}(\mathsf{P}_C f(z)) = \operatorname{sign}(\widehat{f}(\alpha_C)\chi_{\alpha_C}(z))$ at both $z = y$ and $z = x + y$. Multiplying,

$$\operatorname{sign}(\mathsf{P}_C f(x+y))\, \operatorname{sign}(\mathsf{P}_C f(y)) = \chi_{\alpha_C}(x+y)\, \chi_{\alpha_C}(y) = \chi_{\alpha_C}(x),$$

using $\chi_{\alpha_C}(x+y)\chi_{\alpha_C}(y) = \chi_{\alpha_C}(x)$. This holds simultaneously for all $C \in \mathcal{C}$ with probability at least $7/8$.

Finally, the algorithm accesses only estimates $\widetilde{\mathsf{P}}_C f(z)$ within $\pm\theta/8$ of the true values; since the true value has magnitude $> \theta/2 - \theta/4 = \theta/4 > \theta/8$ away from 0 on the good event, the estimated sign agrees with the true sign, and the argument is unaffected. □

Using Claim 1.10, we now prove Lemma 1.9.

Proof. [Proof of Lemma 1.9] By Claim 1.10, for a single uniform y the sign product equals $\chi_{\alpha_C}(x)$ simultaneously for all $C \in \mathcal{C}$ with probability at least $7/8$. Taking the median over $\gamma = O(\log(|\mathcal{C}|/\delta')) = O(\log(1/(\delta'\theta)))$ independent samples $y_1, \ldots, y_\gamma$ and applying a Chernoff bound amplifies the per-coset success probability to $1 - \delta'/|\mathcal{C}|$; a union bound over the $|\mathcal{C}| \leq 4/\theta^2$ cosets yields overall success probability at least $1 - \delta'$. Each sample requires estimating two projections to accuracy $\theta/8$, which by Proposition 1.2 costs $\widetilde{O}(1/\theta^2)$ queries, for a total of $\widetilde{O}(\log(1/\delta')/\theta^2)$ queries. □

Algorithm 3: ESTIMATE-HEAVY-FOURIER-COEFFS

Input: A set of cosets $\mathcal{C}$ as in Lemma 1.8; accuracy ε'; confidence δ'

Step 1: Sample $x_1, \ldots, x_\lambda$ independently and uniformly from $\mathbb{F}_2^n$, where $\lambda = \dfrac{100 \log(1/(\delta'\theta))}{\varepsilon'^2}$.

Step 2: For each $j \in [\lambda]$, invoke Algorithm 2 with input $\mathcal{C}$, query point x_j, and confidence $1/(20\lambda)$ to obtain $\widetilde{\chi}_{\alpha_C}(x_j)$ for all $C \in \mathcal{C}$.

Step 3: For each $C \in \mathcal{C}$, define $\widetilde{\widehat{f}}(\alpha_C) = \frac{1}{\lambda}\sum_{j=1}^{\lambda} f(x_j)\, \widetilde{\chi}_{\alpha_C}(x_j)$.

Output: $\{\widetilde{\widehat{f}}(\alpha_C) : C \in \mathcal{C}\}$

Lemma 1.11. *For all $C \in \mathcal{C}$, ESTIMATE-HEAVY-FOURIER-COEFFS (Algorithm 3) outputs estimates $\widetilde{\widehat{f}}(\alpha_C)$ such that, with probability at least $1 - \delta'$,*

$$\left|\widetilde{\widehat{f}}(\alpha_C) - \widehat{f}(\alpha_C)\right| \leq \varepsilon'.$$

Proof. Condition on the event that every call to List-Correction returns the correct signs $\chi_{\alpha_C}(x_j)$ for all C; by Lemma 1.9 and a union bound over the λ calls (each with confidence $1/(20\lambda)$), this holds with probability at least $19/20$. On this event $f(x_j)\widetilde{\chi}_{\alpha_C}(x_j) = f(x_j)\chi_{\alpha_C}(x_j)$, whose expectation over uniform x_j is exactly $\widehat{f}(\alpha_C)$ and which lies in $[-1, 1]$. By Hoeffding's inequality, with

$$\lambda = \Theta\left(\frac{1}{(\varepsilon')^2}\log\frac{|\mathcal{C}|}{\delta'}\right)$$

samples, $\Pr[\,|\widetilde{\widehat{f}}(\alpha_C) - \widehat{f}(\alpha_C)| > \varepsilon'\,] \le \delta'/|\mathcal{C}|$ for each fixed C. A union bound over all $C \in \mathcal{C}$ gives simultaneous accuracy ε' with probability at least $1 - \delta'$ (absorbing the $1/20$ from the sign event into δ'). □

Algorithm 4: Identify-modulo-Invariance

Input: A collection of cosets $\mathcal{C}$ satisfying Lemma 1.8, and parameter $\theta \in (0, 1)$.

Step 1. Sample $X = \{x_1, \ldots, x_\lambda\}$, each x_i uniform in $\mathbb{F}_2^n$, where $\lambda = \Theta\big(\theta^{-2}\log(1/\theta)\big)$.

Step 2. foreach $x \in X$ **do**
  Obtain $\widetilde{\chi}_{\alpha_C}(x)$ for all $C \in \mathcal{C}$ by invoking List-Correction (Algorithm 2) with confidence parameter $\delta = 1/(20\lambda)$.
end

Step 3. Using Claims 1.12 and 1.13, find a maximal linearly independent subset of $\{\alpha_C : C \in \mathcal{C}\}$; let B denote the corresponding cosets, and record the linear relations satisfied by each α_C for $C \in \mathcal{C} \setminus B$ with respect to the basis $\{\alpha_{C'} : C' \in B\}$.

Step 4. Return the basis B (as cosets) and the list of linear relations for all $C \in \mathcal{C} \setminus B$ relative to B.

Let $\mathcal{S} := \{\alpha_C : C \in \mathcal{C}\}$ be the set of dominant Fourier coefficients associated with the cosets in $\mathcal{C}$. Having access to evaluations of $\chi_{\alpha_C}(x)$ for sufficiently many randomly chosen x, we can recover linear relations among the heavy Fourier characters in $\mathcal{S}$ without knowing their explicit identities.

Claim 1.12. *Let $\mathcal{T} \subseteq \mathcal{S}$. If $\sum_{\alpha\in\mathcal{T}} \alpha = \mathbf{0}^n$, then for every $x \in \mathbb{F}_2^n$,*

$$\prod_{\alpha\in\mathcal{T}} \chi_\alpha(x) = 1.$$

Proof. For any $x \in \mathbb{F}_2^n$,

$$\prod_{\alpha\in\mathcal{T}} \chi_\alpha(x) = \prod_{\alpha\in\mathcal{T}} (-1)^{\langle\alpha,x\rangle} = (-1)^{\langle\sum_{\alpha\in\mathcal{T}}\alpha,\, x\rangle} = (-1)^{\langle\mathbf{0}^n,x\rangle} = 1.$$

□

Claim 1.13. *Let $\mathcal{B} \subseteq \mathcal{S}$ be a set of linearly independent vectors, and let $\mathcal{D} \subseteq \mathcal{B}$ be nonempty. For a collection $X = \{x_1, \ldots, x_\lambda\}$ of uniform independent samples,*

$$\Pr\left[\prod_{\alpha\in\mathcal{D}} \chi_\alpha(x_j) = 1 \ \textit{ for all } j \in [\lambda]\right] \ \leq \ 2^{-\lambda}.$$

Consequently, the probability that this holds for some *nonempty $\mathcal{D} \subseteq \mathcal{B}$ is at most $2^{-\lambda} \cdot 2^{|\mathcal{B}|}$, which is at most $1/100$ once $\lambda = \Omega(\theta^{-2}\log(1/\theta))$.*

Proof. Fix a nonempty $\mathcal{D} \subseteq \mathcal{B}$. Since $\mathcal{B}$ is linearly independent, $\sum_{\alpha\in\mathcal{D}} \alpha \neq \mathbf{0}$, so for uniform x_j,

$$\prod_{\alpha\in\mathcal{D}} \chi_\alpha(x_j) = \chi_{\sum_{\alpha\in\mathcal{D}}\alpha}(x_j)$$

is a uniform ± 1 bit; in particular it equals 1 with probability $\frac{1}{2}$, independently across the λ trials. Hence the all-ones event has probability $2^{-\lambda}$. Since $|\mathcal{B}| = O(1/\theta^2)$, the number of nonempty subsets is at most $2^{|\mathcal{B}|} = 2^{O(1/\theta^2)}$, and a union bound gives failure probability at most $2^{-\lambda} \cdot 2^{O(1/\theta^2)}$. For $\lambda = \Omega(\theta^{-2}\log(1/\theta))$ this is $o(1)$, and in particular below $1/100$. □

We formalize this approach in a self-contained lemma, which also encapsulates the theoretical guarantees of Algorithm 4.

Lemma 1.14. *IDENTIFY-MODULO-INVARIANCE (Algorithm 4) outputs, except with probability at most $1/20$, a basis $B \subseteq \mathcal{C}$ (via their dominant Fourier coefficients) together with the correct linear relations for all $C \in \mathcal{C} \setminus B$.*

Proof. By Claim 1.12, whenever $\sum_{\alpha\in\mathcal{T}} \alpha = \mathbf{0}$ the product $\prod_{\alpha\in\mathcal{T}} \chi_\alpha(x) = 1$ for all x, so every genuine linear relation is detected. Conversely, by Claim 1.13, for any linearly independent set the probability that a spurious relation survives all λ samples is at most $2^{-\lambda} \cdot 2^{O(1/\theta^2)}$. Setting $\lambda = \Theta(\theta^{-2}\log(1/\theta))$ makes this at most $1/100$; combined with the $1/20$ failure probability of the underlying LIST-CORRECTION calls (Step 2), the total failure probability is at most $1/20$. □

2

Application I: Testing Linear Isomorphism

A Boolean function $f : \mathbb{F}_2^n \to \{-1, +1\}$ is said to be linearly isomorphic to $g : \mathbb{F}_2^n \to \{-1, +1\}$ if there exists $A \in \mathrm{GL}_n(\mathbb{F}_2)$ such that $f(x) = g(Ax)$ for all $x \in \mathbb{F}_2^n$. Linear isomorphism is a fundamental notion with applications in coding theory, circuit complexity, and cryptography [Ciriani, 2003, Bemporad, 2004, Yordanov et al., 2012, Dautovic and Novak, 2006, Duursma et al., 2001, Hou, 1994, Cui and Cao, 2007, Dunkelman et al., 2012]. A notable example is given by Reed–Muller codes, which are affine-invariant and whose decoding algorithms crucially exploit this symmetry [Abbe et al., 2021]. Closely related isomorphism problems also arise in cryptography, such as the *Isomorphism of Polynomials* problem [Faugère and Perret, 2006].

From an algorithmic standpoint, this problem has been extensively studied. Determining whether two Boolean functions are linearly isomorphic is computationally intractable even under the strong restriction that the linear transformation is a permutation matrix. The problem is known to be coNP-hard (when the functions are given in DNF form) and lies in Σ_2^p, but it is not known to be in coNP. Agrawal and Thierauf further showed that, unless the polynomial hierarchy collapses to Σ_3^p, the problem is not Σ_2^p-complete [Agrawal and Thierauf, 2000]. The best known algorithm reduces the task to the *Hypergraph Isomorphism Problem* and runs in $2^{O(n)}$ time [Luks, 1999].

In this chapter, we study linear isomorphism of Boolean functions in the *property testing* framework, where the goal is to decide whether an object has a given property or is far from having it, using only a small number of queries. We work in the stronger *tolerant testing* model, which requires distinguishing objects that are close to the property from those that are far. We now formally define the problem of testing linear isomorphism in this setting.

Definition 2.1 ((ϵ, ω)-Tolerant Linear Isomorphism Testing)**.** Let $\epsilon \geq 0$ and $\omega > 0$. An algorithm is given full access to a known function $g : \mathbb{F}_2^n \to \{-1, +1\}$ and oracle access to an unknown function $f : \mathbb{F}_2^n \to \{-1, +1\}$, meaning that for any input $x \in \mathbb{F}_2^n$

the algorithm may query the oracle to obtain $f(x)$. It must distinguish, with probability at least $2/3$, between the following cases:

- **(Yes)** There exists $A \in \mathrm{GL}_n(\mathbb{F}_2)$ such that $\delta(f \circ A, g) \leq \epsilon$.
- **(No)** For all $A \in \mathrm{GL}_n(\mathbb{F}_2)$, $\delta(f \circ A, g) \geq \epsilon + \omega$.

Here $\delta(f, g)$ denotes the fraction of inputs on which f and g differ, that is,

$$\delta(f, g) := \Pr_{x \sim \mathbb{F}_2^n} [f(x) \neq g(x)] .$$

We now review prior work on the isomorphism testing problem, covering both algorithmic upper bounds and lower bounds.

Related Works

Wimmer and Yoshida [2013] were the first to study linear isomorphism testing in the tolerant setting. They gave an algorithm for the $(\omega, 3\omega)$-tolerant problem with query complexity $\widetilde{O}((m/\omega)^{24})$, where m is an upper bound on the spectral norm of the known function g. They also proved an adaptive lower bound of $\Omega(m)$ queries in a certain subconstant regime of ω). In addition, Chakraborty et al. [2012] showed an $\Omega(k)$ lower bound for testing linear isomorphism when the known function is a k-junta, which implies an $\Omega(\log m)$ lower bound in terms of the spectral norm of the known function. Grigorescu et al. [2013] further proved an adaptive lower bound of $\Omega(n^2)$ queries when the known function g is the inner-product function on n bits. Since the spectral norm of the inner-product function is at most $2^{n/2}$, this also yields an $\Omega(\log m)$ lower bound in terms of m.

Linear isomorphism is an affine-invariant property, and testing affine-invariant properties of Boolean functions has been widely studied; see, e.g., [Kaufman and Sudan, 2008, Bhattacharyya et al., 2015, 2013, Hatami and Lovett, 2013]. In particular, the *regularity framework* of Hatami and Lovett [2013] yields tolerant testers for a broad class of affine-invariant properties, including linear isomorphism to a fixed function. However, a major limitation of these general frameworks is that the resulting query complexity grows as a tower-type function of the relevant complexity parameter.

Characterizing which properties admit constant-query testers is a central question in property testing. In graph property testing, a landmark result is the characterization of all testable properties via Szemerédi's regularity lemma [Alon et al., 2006]. In our setting, Wimmer and Yoshida [2013] gave an analogous characterization for linear isomorphism

testing, showing that functions with small spectral norm are exactly those for which linear isomorphism can be tested with a constant number of queries.

A closely related and more restricted problem is testing isomorphism under permutations of variables, which has also been extensively studied [Blais and O'Donnell, 2010a, Alon and Blais, 2010, Chakraborty et al., 2011a, 2012, Blais et al., 2015, 2019a]. Alon et al. [2013a] showed that testing permutation isomorphism for arbitrary Boolean functions requires $\Omega(n)$ queries and can be done with $O(n \log n)$ queries. They also proved that testing permutation isomorphism to a k-junta requires $\Omega(k)$ queries, and gave a nearly matching $O(k \log k)$-query algorithm.

Main Theorems

Prior work exhibits a large gap between the known upper and lower bounds, even for Boolean functions with small spectral norm, which form the only class admitting constant-query testers. In this work, we nearly close this gap by presenting a non-adaptive randomized tester with substantially improved query complexity for all $\epsilon \geq 0$ and $\omega > 0$.

Theorem 2.2. *Let $\epsilon \geq 0$ and $\omega > 0$. There exists a non-adaptive randomized algorithm (see Algorithm 5) that solves the (ϵ, ω)-Tolerant Linear Isomorphism Testing problem for a known function $g : \mathbb{F}_2^n \to \{-1, +1\}$ with spectral norm at most m and an unknown function $f : \mathbb{F}_2^n \to \{-1, +1\}$, using $\widetilde{O}((m/\omega)^4)$ queries to f, and succeeding with probability at least $2/3$. Here, the $\widetilde{O}(\cdot)$ notation hides polynomial factors in $\log m$ and $\log(1/\omega)$.*

2.1 Proof ideas

At a high level, we follow the paradigm of testing by implicit learning introduced by Diakonikolas et al. [2007], and build on the approach of Gopalan et al. [2011] for testing induced membership in subclasses of Boolean functions with small spectral norm, which was also used by Wimmer and Yoshida [2013] for linear isomorphism testing. Our main contribution is a refinement of this approach that yields a tester with significantly improved query complexity compared to Wimmer and Yoshida [2013]. In the process, we assemble an efficient list-correction pipeline, which we believe may be of independent interest in the analysis of Boolean functions. We now outline the overall proof strategy.

We first recall that if two Boolean functions are approximately related by a linear transformation, then their Fourier spectra are strongly correlated. Formally, for $f, g : \mathbb{F}_2^n \to$

$\{-1, 1\}$, if there exists an invertible matrix $A \in \mathrm{GL}_n(\mathbb{F}_2)$ such that $\delta(f, g \circ A) \leq \epsilon$, then

$$\sum_{\beta \in \mathbb{F}_2^n} \widehat{f}(\beta)\, \widehat{g \circ A}(\beta) \;\geq\; 1 - 2\epsilon, \tag{2.1}$$

and the converse also holds. Moreover, when the known function g has small spectral norm, this implies that it suffices to focus on the heavy Fourier coefficients of the unknown function f. The main challenge is to locate and estimate these coefficients using only oracle access to f, with a number of queries independent of n. At first glance, this seems difficult, since even identifying a linear function requires $\Omega(n)$ queries; see, e.g., [O'Donnell, 2014, Exercise 1.27].

We construct an approximation f^* of the unknown function f such that there exists (a priori unknown to the algorithm) $A \in \mathrm{GL}_n(\mathbb{F}_2)$ with $\|f - f^* \circ A\|_2^2 \leq \epsilon$, using $\mathrm{poly}(m, 1/\epsilon)$ queries to f, where m is the spectral norm of the known function g. The crucial observation is that, for testing linear isomorphism, it suffices to learn f only up to an unknown nonsingular linear transformation.

The key tool enabling this is a local list-correction algorithm for Hadamard Code. Intuitively, given a threshold parameter θ, oracle access to f, and a point $x \in \mathbb{F}_2^n$, the algorithm identifies all heavy Fourier coefficients of f and returns a list $L_x = \left\{\chi_\beta(x) : |\widehat{f}(\beta)| \geq \theta\right\}$. The algorithm is not required to output the Fourier characters β themselves, but only their evaluations at the given point x. Its query complexity depends only on θ and is independent of the dimension n. For our purposes, this weaker guarantee is sufficient. By accessing L_x at many random inputs x (polynomially many in $1/\theta$), we show how one can (i) approximate the heavy Fourier coefficients of f, and (ii) recover a basis for them, thereby identifying the coefficients up to a linear transformation. This yields a function that approximates f up to linear isomorphism, which is adequate for our setting, since linear isomorphism is itself a linear-invariant property. Thus, it suffices to check over all possible linear transformations to determine whether a suitable one exists. This can be carried out offline and does not require any further queries to the unknown function f.

2.2 Our Algorithm

Our tester combines the three tools developed in Chapter 1 into a single implicit-learning pipeline:

1. Fourier-Sieve (Algorithm 1) identifies a set of cosets $\mathcal{C}$ capturing every heavy Fourier coefficient of f (Lemma 1.8);

2. Estimate-Heavy-Fourier-Coeffs (Algorithm 3) estimates the magnitude of each heavy coefficient (Lemma 1.11);

3. Identify-modulo-Invariance (Algorithm 4) recovers a basis for the heavy coefficients together with all linear dependencies among them (Lemma 1.14).

Each of these routines uses $\widetilde{O}(1/\theta^4)$ queries to f; combined and instantiated with $\theta = \omega/(10m)$, they use $\widetilde{O}((m/\omega)^4)$ queries in total.

2.2.1 Proof of Theorem 2.2

We begin by restating Theorem 2.2 for the reader's convenience.

Theorem 2.2. *Let $\epsilon \geq 0$ and $\omega > 0$. There exists a non-adaptive randomized algorithm (see Algorithm 5) that solves the (ϵ, ω)-Tolerant Linear Isomorphism Testing problem for a known function $g : \mathbb{F}_2^n \to \{-1, +1\}$ with spectral norm at most m and an unknown function $f : \mathbb{F}_2^n \to \{-1, +1\}$, using $\widetilde{O}((m/\omega)^4)$ queries to f, and succeeding with probability at least $2/3$. Here, the $\widetilde{O}(\cdot)$ notation hides polynomial factors in $\log m$ and $\log(1/\omega)$.*

Proof. We prove the theorem by presenting an explicit algorithm, LinearIsoTester (Algorithm 5), and analyzing its guarantees. Set the parameters

$$\theta = \frac{\omega}{10m}, \qquad \lambda = \frac{100}{\theta^2} \log \frac{100}{\theta^2}.$$

Pipeline guarantees. The tester invokes the three Chapter 1 routines in sequence on the unknown function f.

1. Fourier-Sieve (Algorithm 1) outputs a collection of cosets $\mathcal{C}$ such that, by Lemma 1.8:

   (a) every α with $|\widehat{f}(\alpha)| \geq \theta$ lies in some $C \in \mathcal{C}$;

   (b) every $C \in \mathcal{C}$ has dominant coefficient α_C with $|\widehat{f}(\alpha_C)| \geq \theta/2$; and

   (c) every $C \in \mathcal{C}$ satisfies $\sum_{\beta \in C \setminus \{\alpha_C\}} \widehat{f}(\beta)^2 < \theta^4/2^{12}$.

   By Lemma 1.8 this step uses $\widetilde{O}(1/\theta^4)$ queries and fails with probability at most $1/8$.

2. Estimate-Heavy-Fourier-Coeffs (Algorithm 3) returns estimates $\widetilde{\widehat{f}}(\alpha_C)$ with $|\widetilde{\widehat{f}}(\alpha_C) - \widehat{f}(\alpha_C)| \leq \theta/10$ for all $C \in \mathcal{C}$. By Lemma 1.11 this step uses $\widetilde{O}(1/\theta^4)$ queries and fails with probability at most $1/20$.

3. IDENTIFY-MODULO-INVARIANCE (Algorithm 4) returns a basis $B \subseteq \mathcal{C}$ (via their dominant coefficients α_C) spanning the linear span of $\{\alpha_C : C \in \mathcal{C}\}$, together with the correct linear relations for all $C \in \mathcal{C} \setminus B$. By Lemma 1.14 this step uses $\widetilde{O}(1/\theta^4)$ queries and fails with probability at most $1/20$.

By a union bound over the three stages, the pipeline succeeds on all three simultaneously except with probability at most

$$\frac{1}{8} + \frac{1}{20} + \frac{1}{20} = \frac{9}{40} < \frac{1}{4}.$$

Constructing the proxy f^*. Using the basis $B = \{\alpha_1, \ldots, \alpha_r\}$ and the estimates from step 2, define the real-valued proxy

$$f^*(x) := \sum_{C \in \mathcal{C}} \widetilde{\widehat{f}}(\alpha_C)\, \chi_{\alpha_C}(x), \qquad \text{equivalently} \qquad \widehat{f^*}(\alpha) = \begin{cases} \widetilde{\widehat{f}}(\alpha), & \alpha = \alpha_C \text{ for some } C \in \mathcal{C}, \\ 0, & \text{otherwise.} \end{cases}$$

Although the explicit identities of the α_C are unknown, step 3 fixes them up to an unknown nonsingular linear transformation, which is all that is required below since linear isomorphism is a linear-invariant property.

The following claim shows that this proxy suffices for testing.

Claim 2.3. *Suppose f and g are ϵ-close to being linearly isomorphic. Then there exists a nonsingular transformation A such that $\sum_\beta \widehat{f^*}(\beta)\, \widehat{g \circ A}(\beta) \geq 1 - 2\epsilon - \omega/100$. Conversely, if there exists a nonsingular transformation A satisfying $\sum_\beta \widehat{f^*}(\beta)\, \widehat{g \circ A}(\beta) \geq 1 - 2\epsilon - \omega/100$, then f and g are $(\epsilon + \omega/100)$-close to being linearly isomorphic.*

Proof. Suppose f and g are ϵ-close to linearly isomorphic, so there is an invertible A with $\Pr_x[f(x) \neq g \circ A(x)] \leq \epsilon$. Since $f, g \in \{+1, -1\}$,

$$\sum_\beta \widehat{f}(\beta)\, \widehat{g \circ A}(\beta) = \mathop{\mathbb{E}}_x [f(x)\, g \circ A(x)] = 1 - 2 \Pr_x [f(x) \neq g \circ A(x)] \geq 1 - 2\epsilon.$$

Each coefficient of f^* approximates that of f within additive error $\theta/10$ on the support B, which has size $|B| \leq |\mathcal{C}| \leq 4/\theta^2$, and f^* is supported only there. Hence

$$\sum_\beta \widehat{f^*}(\beta)\, \widehat{g \circ A}(\beta) \geq \sum_\beta \widehat{f}(\beta)\, \widehat{g \circ A}(\beta) - \frac{\theta}{10} \|\widehat{g}\|_1 \geq 1 - 2\epsilon - \frac{\theta}{10}\, m = 1 - 2\epsilon - \frac{\omega}{100},$$

using $\theta = \omega/(10m)$ and $\|\widehat{g}\|_1 \leq m$.

For the converse, suppose the algorithm finds A with $\sum_\beta \widehat{f^*}(\beta)\,\widehat{g\circ A}(\beta) \geq 1-2\epsilon-\omega/100$. Since each estimate is within $\theta/10$ of the true value,

$$\sum_\beta \widehat{f}(\beta)\,\widehat{g\circ A}(\beta) \geq 1-2\epsilon-\frac{\omega}{100}-\frac{\theta}{10}\,\|\widehat{g}\|_1 \geq 1-2\epsilon-\frac{2\omega}{100}.$$

By Parseval,

$$\begin{aligned}\sum_\beta \left(\widehat{f}(\beta)-\widehat{g\circ A}(\beta)\right)^2 &= \sum_\beta \widehat{f}(\beta)^2 - 2\sum_\beta \widehat{f}(\beta)\,\widehat{g\circ A}(\beta) + \sum_\beta \widehat{g\circ A}(\beta)^2 \\ &= 2-2\sum_\beta \widehat{f}(\beta)\,\widehat{g\circ A}(\beta) \leq 2-2\left(1-2\epsilon-\tfrac{2\omega}{100}\right) = 4\epsilon+\frac{4\omega}{100}.\end{aligned}$$

Therefore

$$\Pr_x\left[f(x)\neq g\circ A(x)\right] = \frac{1}{4}\,\mathop{\mathbb{E}}_x\left[(f(x)-g\circ A(x))^2\right] \leq \frac{1}{4}\left(4\epsilon+\frac{4\omega}{100}\right) = \epsilon+\frac{\omega}{100},$$

so f and g are $(\epsilon+\omega/100)$-close to linearly isomorphic. □

Offline decision. Given f^*, accept iff there exists a nonsingular $A\in \mathrm{GL}_n(\mathbb{F}_2)$ with

$$\sum_{\alpha\in\mathbb{F}_2^n} \widehat{g\circ A}(\alpha)\,\widehat{f^*}(\alpha) > 1-2\epsilon-\frac{\omega}{100}.$$

This search uses full access to g and the recovered f^* (given up to linear invariance), and makes no further queries to f. Correctness of this decision rule is exactly Claim 2.3: in the Yes case such an A exists, and in the No case (where $\delta(f\circ A, g)\geq \epsilon+\omega$ for all A) no such A can satisfy the inequality, since it would certify $(\epsilon+\omega/100)$-closeness, contradicting the promise for $\omega/100<\omega$.

Overall failure probability and query complexity. The pipeline yields a valid f^* except with probability at most $\frac{1}{8}+\frac{1}{20}+\frac{1}{20}=\frac{9}{40}<\frac{1}{4}$; the offline decision is deterministic given f^*. Hence the tester errs with probability at most $1/4<1/3$, meeting the required $2/3$ success probability. All queries to f occur inside the pipeline (steps 1–3), each stage costing $\widetilde{O}(1/\theta^4)$ queries; with $\theta=\omega/(10m)$ the total is $\widetilde{O}(1/\theta^4)=\widetilde{O}((m/\omega)^4)$ queries. This completes the proof. □

Algorithm 5: LinearIsoTester

Input: A known function g with $\|\widehat{g}\|_1 \le m$; oracle access to f; parameters $\epsilon \ge 0$, $\omega > 0$.

Output: Accept if $\delta(f \circ A, g) \le \epsilon$ for some $A \in \mathrm{GL}_n(\mathbb{F}_2)$; Reject if $\delta(f \circ A, g) > \epsilon + \omega$ for all such A.

Initialization: $\theta = \dfrac{\omega}{10m}, \quad \lambda = \dfrac{100}{\theta^2} \log \dfrac{100}{\theta^2}$.

Step 1. Run Fourier-Sieve (Algorithm 1) on f with threshold θ to obtain a set of cosets $\mathcal{C}$ capturing all α with $|\widehat{f}(\alpha)| \ge \theta$ (Lemma 1.8).;

Step 2. Run Estimate-Heavy-Fourier-Coeffs (Algorithm 3) on $\mathcal{C}$ to obtain estimates $\widetilde{\widehat{f}}(\alpha_C)$ within $\pm\theta/10$ for all $C \in \mathcal{C}$ (Lemma 1.11).;

Step 3. Run Identify-modulo-Invariance (Algorithm 4) on $\mathcal{C}$ to obtain a basis $B = \{\alpha_1, \ldots, \alpha_r\}$ of $\{\alpha_C : C \in \mathcal{C}\}$ and all linear relations among them (Lemma 1.14).;

Step 4. Define f^* by $\widehat{f^*}(\alpha_C) = \widetilde{\widehat{f}}(\alpha_C)$ for $C \in \mathcal{C}$ and $\widehat{f^*}(\alpha) = 0$ otherwise.;

Step 5. Accept iff there exists nonsingular $A \in \mathbb{F}_2^{n \times n}$ with

$$\sum_{\alpha \in \mathbb{F}_2^n} \widehat{g \circ A}(\alpha)\, \widehat{f^*}(\alpha) > 1 - 2\epsilon - \frac{\omega}{100}.$$

3

Application II: Distribution-Free ℓ_2 Testing of Fourier Sparsity

Sparsity is one of the most powerful ideas connecting modern machine learning and theoretical computer science. It captures the intuition that, even in high-dimensional settings, many natural functions or signals depend only on a small number of relevant components. This assumption underlies algorithms that are both sample- and time-efficient, forming the basis of techniques such as sparse linear regression, decision tree learning, and compressed sensing. Across these frameworks, the common principle is simple yet profound: if a function admits a sparse representation in a suitable basis, such as Fourier, wavelet, polynomial, or a learned dictionary, then learning and inference can be made dramatically more efficient. A particularly elegant setting where sparsity plays a central role is that of real-valued functions over the Boolean hypercube $\mathbb{F}_2^n$. Fourier analysis [O'Donnell, 2014, Stein and Shakarchi, 2003] provides a natural orthonormal basis for such functions. Every function $f : \mathbb{F}_2^n \to \mathbb{R}$ can be expressed as

$$f(x) = \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)(-1)^{\langle x,\alpha \rangle},$$

where $\widehat{f}(\alpha)$ denotes the Fourier coefficient at frequency α and $\langle \cdot, \cdot \rangle$ denotes the standard inner product over $\mathbb{F}_2^n$. The Fourier sparsity of f, defined as the number of nonzero coefficients in this expansion, quantifies how succinctly the function can be represented in the Fourier basis.

In learning theory, functions with sparse Fourier spectra arise naturally and have been studied extensively. Prominent examples include graph and hypergraph cut functions, as well as decision trees of bounded depth, whose Fourier spectra are concentrated on low-degree coefficients [Amrollahi et al., 2019, Stobbe and Krause, 2012, Linial et al., 1993]. For instance, the cut function of a graph can be represented as a degree-2 multilinear polynomial in the Fourier basis, while the cut function induced by a degree-d hypergraph admits a similar representation of degree at most d. More generally, the Fourier spectrum of a Boolean decision tree of depth d is entirely supported on coefficients of degree

at most d. Learning such functions has recently found applications in neural network hyperparameter optimization [Hazan et al., 2018]. From an algorithmic perspective, efficiently learning Fourier-sparse Boolean functions from uniform random examples has long been a central open problem in computational learning theory [Blum, 2003]. Over the years, many approaches have been explored, often under relaxed models or with stronger access [Kushilevitz and Mansour, 1993, Hassanieh et al., 2012b, Vershynin, 2018, Ishay and Regev, 2017, Amrollahi et al., 2019, Indyk et al., 2014]. Almost all of these algorithms, however, assume prior knowledge of the target function's Fourier sparsity level. This naturally motivates the need for *testing for Fourier sparsity*, which aims to validate the Fourier sparsity assumption without fully learning the function. Such tools can serve as effective preprocessing steps in computationally expensive learning pipelines, both for validating sparsity assumptions and for initializing sparsity parameters in downstream algorithms.

Property testing was introduced in the seminal work of Blum, Luby, and Rubinfeld [Blum et al., 1993b] and has since developed into a mature branch of theoretical computer science, particularly well suited for questions of this type. It studies algorithms that, given query access to an unknown function, decide whether the function satisfies a specified property or is far from every function having that property. The objective is to make this decision using as few queries as possible, typically sublinear in the size of the domain. Distance is measured with respect to an underlying input distribution, most commonly the uniform distribution. We refer the interested reader to Goldreich [2017] for a comprehensive survey of classical results in this area. The problem of testing Fourier sparsity in the property-testing framework has been studied in a sequence of works [Gopalan et al., 2011, Yaroslavtsev and Zhou, 2020, Ghosh and Roy, 2025]. In Gopalan et al. [2011], the underlying distance notion was the Hamming distance, whereas later works considered the Euclidean distance, which we believe is a more natural measure in machine-learning settings. In particular, Yaroslavtsev and Zhou [2020] showed that using $\tilde{O}(s/\varepsilon^4)$ queries one can determine whether an unknown function is s-Fourier-sparse or ε-far from every such function in Euclidean distance. On the lower-bound side, they showed that any tester solving this problem must make at least $\Omega(\sqrt{s})$ queries. Very recently, Ghosh and Roy [2025] proposed a simpler tester, with query complexity $\tilde{O}(s/\varepsilon^2)$ and established a matching lower bound of $\Omega(s)$, thereby proving optimality up to logarithmic factors and nearly closing the gap between upper and lower bounds. Another important aspect of the upper-bound techniques in Ghosh and Roy [2025] and Yaroslavtsev and Zhou [2020] is that the corresponding testers are nonadaptive, i.e., the set of queries is fixed before any oracle interaction. By contrast, the lower-bound arguments of Ghosh and Roy [2025] apply even to adaptive testers, which are allowed to select queries based on previous

oracle responses.

All existing work on testing Fourier sparsity has focused on the uniform-distribution setting, where distance is measured with respect to the uniform measure on the Boolean cube. While this assumption is mathematically convenient, it is often unrealistic in learning scenarios, where the input distribution may be structured, biased, or entirely unknown [Valiant, 1984]. This observation motivates the study of the *distribution-free* setting, originally introduced in Halevy and Kushilevitz [2007]. In this model, distance between functions is measured with respect to the unknown underlying distribution, and the tester is required to distinguish functions that satisfy the property from those that are far from it under this distribution. The tester must succeed for every fixed but unknown distribution over $\mathbb{F}_2^n$ and is given only sample access to the distribution. More precisely, distribution-free property testing is the analogue of the PAC+MQ model of learning (traditionally studied via the equivalent EQ+MQ framework) [Kearns and Vazirani, 1994], whereas standard property testing corresponds to learning under the uniform distribution with membership queries. Designing efficient testers in the distribution-free model is significantly more challenging, as many uniform-distribution testers rely on the fact that functions far from the property contain many small local witnesses, which can be found via random sampling. In contrast, in the distribution-free setting, a function may be very close to the property under the uniform distribution, hence containing few such witnesses, while simultaneously being far from the property with respect to some arbitrary input distribution. In fact, it is known that there is no generic transformation that can convert a standard (uniform-distribution) tester into a distribution-free tester for an arbitrary nontrivial property [Goldreich et al., 1998]. Indeed, designing distribution-free testers is particularly subtle: moving from the uniform-distribution setting to the distribution-free setting can, in general, lead to a dramatic increase in complexity. A concrete illustration of this phenomenon is provided by the class of *junta functions*. In the uniform-distribution setting, Blais [2008] gave a nonadaptive tester for k-juntas with query complexity $\widetilde{O}(k^{3/2})$. However, Liu et al. [2019] later showed that in the distribution-free model, any nonadaptive algorithm for testing k-juntas over $\{0,1\}^n$ must make $2^{\Omega(k)}$ queries, even for $\varepsilon = 1/3$. In this work, we initiate a systematic study of Fourier sparsity testing in the distribution-free setting. Given oracle access to an unknown function and sample access to a fixed but unknown input distribution, we ask whether one can design a tester that distinguishes functions with at most s nonzero Fourier coefficients from those that are ε-far (in an appropriate distance measure) from every such function, with respect to the underlying distribution. Motivated by the junta-function scenario, a natural question is whether such a goal can be achieved with only a modest blowup in query complexity compared to the best-known tester in the uniform-distribution setting. Our work answers this question in

the affirmative. We now formally define the problem studied in this work.

(ε, δ)-Distribution-Free s-Fourier Sparsity Tester

Input. Oracle access to a function $f : \mathbb{F}_2^n \to \mathbb{R}$ satisfying $\mathbb{E}_{x\sim\mathbb{F}_2^n}\left[f(x)^2\right] = 1$, and sample access to an unknown distribution D over $\mathbb{F}_2^n$.

Goal. Design a randomized algorithm that makes as few oracle queries to f and as few samples from D as possible, and distinguishes, with probability at least $2/3$, between the following two cases:

- **(YES)** The function f has at most s nonzero Fourier coefficients.
- **(NO)** For every function $g : \mathbb{F}_2^n \to \mathbb{R}$ with at most s nonzero Fourier coefficients,

$$\Pr_{x\sim D}\left[|f(x) - g(x)| > \varepsilon\right] > \delta.$$

Complexity measure. The complexity of the algorithm is measured in terms of the number of oracle queries to f and the number of samples drawn from the distribution D.

Remark 3.1. The assumption on the expected squared norm of the unknown function is without loss of generality: any function can be rescaled so that its expected squared norm equals 1.

Main Theorem

We first note that the notion of distance used in this work (see Definition 3.3) is a stable variant of the ℓ_2 distance. In particular, in the uniform-distribution setting, being ε-far under this distance notion implies being far in the standard ℓ_2 distance. Consequently, existing nonadaptive testers for Fourier sparsity under the ℓ_2 distance in the uniform-distribution setting remain applicable in our framework. Finally, we observe that the lower bound of Ghosh and Roy [2025] also extends to this distance measure for constant parameters $\varepsilon = O(1)$ and $\delta = O(1)$, yielding a lower bound of $\Omega(s)$ for this problem. We now state our result.

Theorem 3.2 (Main Theorem)**.** *For any $\varepsilon, \delta \in (0, 1)$ and $s \in \mathbb{N}$, there exists a nonadaptive randomized algorithm with success probability at least $2/3$ for the (ε, δ)-distribution-free s-Fourier sparsity testing problem that makes $\widetilde{O}\left(\frac{s^4}{\delta^4\varepsilon^8}\right)$ oracle queries to the unknown function and draws $\widetilde{O}(1/\delta)$ samples from the input distribution.*

Somewhat surprisingly, this result shows that moving from the uniform-distribution model to the distribution-free setting incurs only a quadratic overhead in query complexity. This behavior stands in sharp contrast to the junta testing scenario, where, in the absence of adaptivity, the query complexity in the distribution-free setting is known to be exponentially larger than in the uniform-distribution model. Beyond this algorithmic result, we develop new Fourier-analytic techniques for analyzing Boolean functions with sparse Fourier spectra under arbitrary input distributions. These techniques may be of independent interest for distribution-free testing and learning problems more broadly.

Organization

The remainder of the chapter is organized as follows. In Section 3.1, we provide a high-level proof sketch of our main result, highlighting the key intuitions. Section 3.2 presents our testing algorithm together with its analysis and correctness proof.

3.1 A Warmup Algorithm

We now provide a high-level overview of our algorithm for testing Fourier sparsity in the distribution-free setting. The purpose of this section is to convey the intuition underlying our approach, abstracting away from technical details. Accordingly, the algorithm presented here should be viewed as a conceptual sketch rather than the final implementation; the complete description and analysis appear later (Section 3.2).

We begin by defining the following notions, which will be useful throughout the rest of the chapter. We denote by $\mathcal{P}_s$ the class of all s-Fourier-sparse functions.

Definition 3.3 (ε-error distance)**.** Let D be a distribution over $\mathbb{F}_2^n$ and let $f, g : \mathbb{F}_2^n \to \mathbb{R}$ be two functions. For any $\varepsilon > 0$, the *ε-error distance* between f and g under D is

$$\mathrm{dist}_D^{\varepsilon}(f, g) \;:=\; \Pr_{x \sim D}\big[\, |f(x) - g(x)| \geq \varepsilon \,\big].$$

Definition 3.4 (Distance to $\mathcal{P}_s$)**.** For a function $f : \mathbb{F}_2^n \to \mathbb{R}$, the ε-error distance from $\mathcal{P}_s$ with respect to D is

$$\mathrm{dist}_D^{\varepsilon}(f, \mathcal{P}_s) := \inf_{g \in \mathcal{P}_s} \mathrm{dist}_D^{\varepsilon}(f, g).$$

At a high level, our approach is inspired by the work of Halevy and Kushilevitz [2007], which combines self-correction with one-sided testers under the uniform distribution to obtain distribution-free property testers. However, their framework applies only to

Boolean-valued functions $f : \mathbb{F}_2^n \to \{0,1\}$ and does not extend to real-valued functions. In this work, we develop a framework for handling real-valued functions in a similar spirit. We believe this approach is of independent interest, especially given the widespread use of the techniques of Halevy and Kushilevitz [2007] in the design of distribution-free property testers. This extension is particularly relevant in light of the prevalence of real-valued functions in modern machine-learning settings, and we believe it will be useful for designing distribution-free testers for many subclasses of Fourier-sparse functions.

We begin by presenting the top-level structure of our algorithm, deliberately suppressing many implementation details. To this end, we assume access to two algorithmic primitives: (i) a tester for Fourier sparsity under the uniform distribution, and (ii) a self-approximation procedure for the class of Fourier-sparse functions. We then show how these two components can be combined to obtain a distribution-free tester for Fourier sparsity.

Assumption 3.5 (Uniform-Distribution Fourier Sparsity Tester). *Given $\varepsilon \in (0,1)$ and $\delta \in (0,1)$, let $f : \mathbb{F}_2^n \to \mathbb{R}$ be an unknown function accessible only via black-box query access. There exists a randomized tester T such that:*

- ***Completeness:*** *If $f \in \mathcal{P}_s$, then $\Pr\big[T(\mathcal{O}_f, \varepsilon, \delta) = \text{YES}\big] \;\geq\; \frac{9}{10}$.*
- ***Soundness:*** *If $\operatorname{dist}_U^{\varepsilon}(f, \mathcal{P}_s) > \delta$, then $\Pr\big[T(\mathcal{O}_f, \varepsilon, \delta) = \text{NO}\big] \;\geq\; \frac{9}{10}$.*

Moreover, the tester T makes $\operatorname{poly}(s, 1/\varepsilon, 1/\delta)$ queries to the oracle for f.

Assumption 3.6 (Self-Approximator for Fourier-Sparse Functions). *Given $s \in \mathbb{N}$, $\tau \in (0,1)$, and $\eta' \in (0,1)$, let $f : \mathbb{F}_2^n \to \mathbb{R}$ be an unknown function accessible only via black-box query access. Define*

$$g^\star \;:=\; \arg\min_{g \in \mathcal{P}_s} \|f - g\|_2^2 \;=\; \arg\min_{g \in \mathcal{P}_s} \mathop{\mathbb{E}}_{x \sim \mathbb{F}_2^n} \big[(f(x) - g(x))^2\big],$$

that is, $g^\star$ is an s-Fourier-sparse function minimizing the squared ℓ_2 distance from f under the uniform measure on $\mathbb{F}_2^n$. Then there exists a randomized algorithm $A(\mathcal{O}_f, \tau, \eta', \cdot)$ such that for every $x \in \mathbb{F}_2^n$,

$$\Pr\big[\big|A(\mathcal{O}_f, \tau, \eta', x) - g^\star(x)\big| \leq \tau\big] \;\geq\; 1 - \eta',$$

where the probability is taken over the internal randomness of the algorithm. Moreover, the algorithm A makes $\operatorname{poly}(s, 1/\tau, 1/\eta')$ queries to the oracle for f.

Assuming these primitives, we now show how they can be combined to solve the distribution-free Fourier sparsity testing problem. In particular, Algorithm 6 presents a high-level

sketch of how these components interact to yield our tester. Here m plays the role that the sample count T plays in the full algorithm of Section 3.2.

Algorithm 6: Distribution-Free Fourier Sparsity Tester (Sketch)

Input: Oracle access to $f : \mathbb{F}_2^n \to \mathbb{R}$; sample access to a distribution D; proximity parameter ε; distance parameter δ.
Output: YES if f is Fourier-sparse under D; NO if f is far.

Step 1 (Uniform test). Run $T(\mathcal{O}_f, \varepsilon, \delta)$. If T returns NO, output NO and halt.;
Step 2 (Consistency test). Sample $x_1, \dots, x_m \sim D$ independently, where $m = \lceil c/\delta \rceil$ for a sufficiently large absolute constant c. Query $f(x_i)$ and compute $A(\mathcal{O}_f, \varepsilon/4, 1/(20m), x_i)$ for all $i \in [m]$.;
Step 3 (Decision). Output NO if there exists $i \in [m]$ with

$$\left|A(\mathcal{O}_f, \varepsilon/4, 1/(20m), x_i) - f(x_i)\right| > 3\varepsilon/4;$$

otherwise output YES.;

Analysis of Algorithm 6

We prove that Algorithm 6 is a correct (ε, δ)-distribution-free tester for Fourier sparsity by establishing completeness and soundness.

Proof. [Correctness of Algorithm 6]

Completeness. Assume that $f \in \mathcal{P}_s$. By Assumption 3.5, the uniform-distribution Fourier sparsity tester $T(\mathcal{O}_f, \varepsilon, \delta)$ accepts with probability at least $9/10$. Hence, Algorithm 6 proceeds to the consistency-testing stage. By Assumption 3.6, invoked with accuracy $\varepsilon/4$ and failure probability $1/(20m)$, for every $x \in \mathbb{F}_2^n$,

$$\Pr\left[\left|A(\mathcal{O}_f, \varepsilon/4, 1/(20m), x) - g^\star(x)\right| \le \varepsilon/4\right] \ge 1 - \frac{1}{20m}.$$

Since $f \in \mathcal{P}_s$ we have $g^\star = f$, so $|A(\cdots, x) - f(x)| \le \varepsilon/4 < 3\varepsilon/4$. Applying a union bound over the m sampled points $x_1, \dots, x_m$, the estimator succeeds on all samples except with probability at most $m \cdot \frac{1}{20m} = \frac{1}{20}$. Therefore, the consistency test does not reject and Algorithm 6 outputs YES with probability at least $1 - \left(\frac{1}{10} + \frac{1}{20}\right) > \frac{2}{3}$.

Soundness. Assume that $\mathrm{dist}_D^{\varepsilon}(f, \mathcal{P}_s) > \delta$. We distinguish two cases.

Case 1. $\mathrm{dist}_U^{\varepsilon}(f, \mathcal{P}_s) > \delta$. By Assumption 3.5, the uniform-distribution tester rejects with probability at least $9/10$, and the algorithm outputs NO.

Case 2. $\mathrm{dist}_U^{\varepsilon}(f, \mathcal{P}_s) \le \delta$. Let $g^\star := \arg\min_{g \in \mathcal{P}_s} \|f - g\|_2^2$. Since $g^\star \in \mathcal{P}_s$ and $\mathrm{dist}_D^{\varepsilon}(f, \mathcal{P}_s) > \delta$,

$$\Pr_{x \sim D}[\,|f(x) - g^\star(x)| > \varepsilon\,] > \delta.$$

By Assumption 3.6, the estimator $A^f(x) := A(\mathcal{O}_f, \varepsilon/4, 1/(20m), x)$ satisfies $|A^f(x) - g^\star(x)| \le \varepsilon/4$ for all sampled x except with probability at most $m \cdot \frac{1}{20m} = \frac{1}{20}$, by a union bound over the m samples. The tester can incorrectly output YES only if none of the m points $x_1, \dots, x_m$ drawn from D satisfies $|f(x_i) - g^\star(x_i)| > \varepsilon$ (indeed, on any such point the estimate obeys $|A^f(x_i) - f(x_i)| > \varepsilon - \varepsilon/4 = 3\varepsilon/4$, triggering rejection). This event occurs with probability at most

$$(1 - \delta)^m \le e^{-\delta m} < \frac{1}{10}$$

for our choice of m. Combining with the estimator failure probability, the tester rejects with probability at least $1 - \left(\frac{1}{10} + \frac{1}{20}\right) > \frac{2}{3}$.

This completes the proof of correctness of Algorithm 6. □

What remains is to instantiate the uniform-distribution Fourier sparsity tester and the accompanying self-approximator for the class of Fourier-sparse functions. Both components rely on the coset-hashing toolkit of Chapter 1. We refer the reader to Indyk [2013] for an excellent survey of the underlying sketching-by-hashing technique. The core idea is to randomly project the Fourier spectrum of a function f onto cosets of a random subspace of sufficiently large codimension. Under such a projection, Fourier coefficients of relatively large magnitude are mapped to cosets in an essentially pairwise independent manner. This observation was first exploited by Feldman et al. [2006a] in the study of halfspaces, and later by Gopalan et al. [2011] in the context of testing Fourier sparsity. We build upon this framework using the refined analysis of the coset-hashing process developed in Chapter 1 (see also Datta et al. [2026b]). We now describe the main technical ideas underlying the implementation of the Fourier sparsity tester and the self-approximator.

After projecting the Fourier spectrum, the algorithm seeks to identify *heavy cosets*, namely those containing Fourier coefficients of relatively large magnitude. From these, it recovers the top s cosets, ranked by magnitude. At this stage, two sources of error may arise. The first is an *estimation error*, since individual Fourier coefficients cannot be accessed directly and must be estimated from samples. The second is a *hashing error*, as a coset may contain additional, non-heavy Fourier coefficients whose contributions contaminate the estimate. Our analysis shows that, despite these sources of error, it is still possible to obtain an additive approximation to the top s Fourier coefficients of f. More precisely, we recover an additive estimate of the Fourier mass corresponding to the best s-Fourier-sparse

approximation to f in squared ℓ_2 distance under the uniform distribution. We then relate this approximation guarantee to our distance notion. In particular, if $\mathrm{dist}_U^{\varepsilon}(f, \mathcal{P}_s) > \delta$, then f is $\Omega(\varepsilon^2\delta)$-far, in squared ℓ_2 distance, from every s-Fourier-sparse function. By Parseval's identity, this implies that the total Fourier mass of the top s coefficients of f is at most $1 - \Omega(\varepsilon^2\delta)$. Consequently, an additive estimate of the top-s Fourier mass suffices to distinguish s-Fourier-sparse functions from those that are far. With an appropriate choice of parameters, this yields a one-sided tester for Fourier sparsity under the uniform distribution with respect to our notion of distance.

For the self-approximator, we draw on *self-correction for Hadamard codes* (Proposition 1.31 in O'Donnell [2014]). Once a coset containing a heavy Fourier coefficient $\widehat{f}(\alpha)$ has been identified, we can evaluate the corresponding character χ_α at queried points via a variant of self-correction adapted to our setting (Chapter 1, Algorithm 2 and Lemma 1.9). By collecting sufficiently many such evaluations and averaging quantities of the form $\chi_\alpha(x)f(x)$, we estimate the magnitude of the corresponding Fourier coefficient. Repeating this for all identified heavy cosets yields an approximation to the closest s-Fourier-sparse function. At this stage, we do not recover the identities of the heavy coefficients themselves, only their magnitudes. Nevertheless, to evaluate the resulting approximation $\widetilde{f}$ at a point z, it suffices to know the magnitudes of the top s Fourier coefficients together with the values of the corresponding characters at z, which are again obtained via self-correction. The crux of our analysis is to show that this approximation is sufficient for accurate error estimation in the ℓ_1 sense. Finally, our self-approximator construction is quite general and may find independent applications in other testing and learning problems.

3.2 The Main Algorithm

The main technical tool underlying our approach is a local list-correction procedure for Hadamard codes, developed in Chapter 1. Given only query access to f, we first implicitly identify heavy codewords (Algorithm 1, Lemma 1.8), corresponding to large Fourier coefficients, without explicitly recovering their identities. Through a specialized self-correction procedure (Algorithm 2, Lemma 1.9), we obtain oracle access to evaluations of the associated Fourier characters χ_α on requested inputs. By performing sufficiently many such evaluations, we estimate the corresponding Fourier coefficients within the desired accuracy (Algorithm 3, Lemma 1.11), thereby constructing an approximation to the closest s-Fourier-sparse function. We recall the guarantees of these three Chapter 1 routines that we use as black boxes:

- Fourier-Sieve (Algorithm 1, Lemma 1.8): given a threshold θ, returns cosets $\mathcal{C}$

capturing every α with $|\widehat{f}(\alpha)| \geq \theta$, each with a dominant coefficient of magnitude $\geq \theta/2$ and residual coset mass $< \theta^4/2^{12}$, using $\widetilde{O}(1/\theta^4)$ queries.

- List-Correction (Algorithm 2, Lemma 1.9): for any x, returns $\chi_{\alpha_C}(x)$ for all $C \in \mathcal{C}$ correctly with probability $\geq 1 - \delta'$, using $\widetilde{O}(\log(1/\delta')/\theta^2)$ queries.
- Estimate-Heavy-Fourier-Coeffs (Algorithm 3, Lemma 1.11): returns $\widetilde{\widehat{f}}(\alpha_C)$ within $\pm\varepsilon'$ of $\widehat{f}(\alpha_C)$ for all $C \in \mathcal{C}$, with probability $\geq 1 - \delta'$.

3.2.1 Proof of Theorem 3.2

We begin by showing that if the heavy Fourier coefficients of a function are estimated with sufficient accuracy, then truncating to the top-s Fourier coefficients yields a good approximation in both ℓ_1 and ℓ_2^2 senses. For each $\alpha \in \mathbb{F}_2^n$, suppose we are given an estimate

$$\widetilde{\widehat{f}}(\alpha) = \widehat{f}(\alpha) + e_\alpha, \qquad \text{where } |e_\alpha| \leq \eta.$$

Let $S^\star$ denote the set of indices of the s largest Fourier coefficients $|\widehat{f}(\alpha)|$, and let T denote the set of indices of the s largest estimated coefficients $|\widetilde{\widehat{f}}(\alpha)|$. Define

$$\Delta_1 := \left| \sum_{\alpha \in S^\star} |\widehat{f}(\alpha)| - \sum_{\alpha \in T} |\widetilde{\widehat{f}}(\alpha)| \right|, \qquad \Delta_2 := \left| \sum_{\alpha \in S^\star} \widehat{f}(\alpha)^2 - \sum_{\alpha \in T} \widetilde{\widehat{f}}(\alpha)^2 \right|.$$

Claim 3.7. *$\Delta_1 \leq 2\eta s$ and $\Delta_2 \leq 4\eta\sqrt{s}$.*

Proof. By Parseval's identity, $\sum_\alpha \widehat{f}(\alpha)^2 = 1$. Let $\alpha_1, \ldots, \alpha_{2^n}$ denote the Fourier coefficients of f, ranked by magnitude, ties broken arbitrarily. Then the s-th largest satisfies

$$|\widehat{f}(\alpha_s)| \;\leq\; \frac{1}{\sqrt{s}}. \tag{1}$$

Let $A = S^\star \setminus T$ and $B = T \setminus S^\star$. Since $|S^\star| = |T| = s$, we have $|A| = |B| \leq s$. For any $\alpha \in A$ and $\beta \in B$, the definition of T implies $|\widetilde{\widehat{f}}(\beta)| \geq |\widetilde{\widehat{f}}(\alpha)|$. Using $|e_\alpha| \leq \eta$, we obtain $|\widehat{f}(\beta)| + \eta \geq |\widehat{f}(\alpha)| - \eta$, hence

$$|\widehat{f}(\alpha)| - |\widehat{f}(\beta)| \leq 2\eta. \tag{2}$$

To bound Δ_1, pair each $\alpha \in A$ with a distinct $\beta \in B$. Using (2),

$$\sum_{\alpha \in A} |\widehat{f}(\alpha)| - \sum_{\beta \in B} |\widehat{f}(\beta)| \leq \sum_{\text{pairs}} 2\eta \leq 2\eta s,$$

proving the first inequality. For Δ_2, for each paired (α, β),

$$\widehat{f}(\alpha)^2 - \widehat{f}(\beta)^2 = \big(|\widehat{f}(\alpha)| - |\widehat{f}(\beta)|\big)\big(|\widehat{f}(\alpha)| + |\widehat{f}(\beta)|\big).$$

By (1) and (2), $|\widehat{f}(\alpha)| + |\widehat{f}(\beta)| \le 2/\sqrt{s}$, so $\widehat{f}(\alpha)^2 - \widehat{f}(\beta)^2 \le 4\eta/\sqrt{s}$. Summing over at most s pairs yields $\Delta_2 \le 4\eta\sqrt{s}$. □

We now prove the main theorem. Throughout, we condition on the event that all subroutines invoked in Steps 1–4 of Algorithm 7 succeed; the total failure probability is bounded at the end via a union bound. By the guarantee of Fourier-Sieve (Algorithm 1, Lemma 1.8), all cosets containing Fourier coefficients of magnitude at least θ are identified. By Estimate-Heavy-Fourier-Coeffs (Algorithm 3, Lemma 1.11), each such coefficient $\widehat{f}(\alpha_C)$ is estimated to additive accuracy at most $\theta/8$. Set $\eta := \theta/8 = \varepsilon^2\delta/(160s)$. By Claim 3.7, this choice gives $\Delta_1 \le 2\eta s = \varepsilon^2\delta/80 \le \varepsilon/4$ and $\Delta_2 \le 4\eta\sqrt{s} = \varepsilon^2\delta/(40\sqrt{s}) \le \varepsilon^2\delta/2$, using $s \ge 1$ and $\varepsilon, \delta \in (0, 1)$.

Completeness. Assume $f \in \mathcal{P}_s$, i.e., f is exactly s-Fourier-sparse. Then all Fourier mass of f is on its top s coefficients, so the sum of squares of the top s true coefficients equals 1. By the bound on Δ_2, the estimate produced by the algorithm satisfies $\sum_{C\in S} \widetilde{\widehat{f}}(\alpha_C)^2 \ge 1 - \varepsilon^2\delta/2$, so the algorithm does not reject in Step 3. Let $\widetilde{f}$ be the hypothesis constructed by the algorithm. By the estimation and truncation guarantees, for each sample $z_i \sim D$ we have $\Pr[\,|f(z_i) - \widetilde{f}(z_i)| > \varepsilon/4\,] \le \delta/200$. Since the algorithm draws $T = 20/\delta$ independent samples, a union bound implies that with probability at least $9/10$ all samples satisfy $|f(z_i) - \widetilde{f}(z_i)| \le \varepsilon/4 < 3\varepsilon/4$. Thus the algorithm accepts, establishing completeness.

Soundness. Assume $\mathrm{dist}_D^{\varepsilon}(f, \mathcal{P}_s) \ge \delta$. We consider two cases.

Case 1. $\mathrm{dist}_U^{\varepsilon}(f, \mathcal{P}_s) > \delta$. Let $g^\star$ be the best s-Fourier-sparse approximation to f under the uniform distribution. Then $\Pr_{x\sim U}[|f(x) - g^\star(x)| > \varepsilon] > \delta$, which implies $\mathbb{E}_{x\sim U}[(f(x) - g^\star(x))^2] \ge \varepsilon^2\delta$. Since $g^\star$ is the truncation of f to its top s Fourier coefficients, the ℓ_2^2-mass outside the top s is at least $\varepsilon^2\delta$; equivalently, the sum of squares of the top s coefficients is at most $1 - \varepsilon^2\delta$. Using the bound on Δ_2, the estimate in Step 3 falls below $1 - \varepsilon^2\delta/2$, so the tester rejects.

Case 2. $\mathrm{dist}_U^{\varepsilon}(f, \mathcal{P}_s) \le \delta$. Let $g^\star$ again denote the best s-Fourier-sparse approximation to f under the uniform distribution. Since $\mathrm{dist}_D^{\varepsilon}(f, \mathcal{P}_s) \ge \delta$, $\Pr_{x\sim D}[|f(x) - g^\star(x)| > \varepsilon] \ge \delta$. By construction of $\widetilde{f}$ and the bound $\Delta_1 \le \varepsilon/4$, each evaluation $\widetilde{f}(x)$ satisfies $|\widetilde{f}(x) - g^\star(x)| \le \varepsilon/4$ except with probability $\delta/200$ (the per-call failure probability of List-Correction); by a union bound over the $T = 20/\delta$ evaluations, all of

Algorithm 7: Distribution-Free Fourier Sparsity Tester

Input: Oracle access to $f : \mathbb{F}_2^n \to \mathbb{R}$; sample access to a distribution D; sparsity parameter s; proximity parameter ε; distance parameter δ.

Initialization: $\theta := \dfrac{\varepsilon^2\delta}{20s}, \quad T := 20/\delta.$

Step 1 (Heavy Fourier coefficient identification). Invoke Fourier-Sieve (Algorithm 1) with threshold θ to implicitly identify the heavy Fourier coefficients of f (repeating a constant number of times and keeping a successful run; see the analysis). Let $\mathcal{C}$ denote the returned cosets.;

Step 2 (Estimating heavy Fourier coefficients). For each coset $C \in \mathcal{C}$, estimate the dominant Fourier coefficient $\widehat{f}(\alpha_C)$ to obtain $\widetilde{\widehat{f}}(\alpha_C)$ with additive error $\pm\theta/8$ and confidence $1 - \theta^2/100$, using Estimate-Heavy-Fourier-Coeffs (Algorithm 3).;

Step 3 (Constructing an s-Fourier-sparse approximation). Let $S \subseteq \mathcal{C}$ be the set of the s cosets with largest $|\widetilde{\widehat{f}}(\alpha_C)|$. If

$$\sum_{C \in S} \widetilde{\widehat{f}}(\alpha_C)^2 < 1 - \frac{\varepsilon^2\delta}{2},$$

then **Reject**; otherwise continue.;

Step 4 (Consistency check). For $i = 1, \ldots, T$:;

Sample $z_i \sim D$.;

Invoke List-Correction (Algorithm 2) with additive error $\pm\theta/8$ and confidence $1 - \delta/200$ to evaluate $\chi_{\alpha_C}(z_i)$ for all $C \in S$.;

Compute $\widetilde{f}(z_i) := \sum_{C \in S} \widetilde{\widehat{f}}(\alpha_C)\, \chi_{\alpha_C}(z_i)$.;

If $|f(z_i) - \widetilde{f}(z_i)| > 3\varepsilon/4$, then **Reject**.;

Step 5 (Decision). Accept.;

them are accurate except with probability at most $1/10$. For any point satisfying both $|f(x) - g^\star(x)| > \varepsilon$ and $|\widetilde{f}(x) - g^\star(x)| \le \varepsilon/4$, the triangle inequality gives $|f(x) - \widetilde{f}(x)| > 3\varepsilon/4$; hence each sample triggers rejection with probability at least $\delta - \delta/200 \ge \delta/2$. Since the algorithm draws $T = 20/\delta$ samples, the probability that none triggers rejection is at most $(1 - \delta/2)^T < 1/10$. Therefore the algorithm rejects with probability at least $9/10$.

Completeness and soundness together establish that Algorithm 7 is an (ε, δ)-distribution-free tester for s-Fourier sparsity.

Overall success probability. The one subroutine whose native failure probability exceeds $1/100$ is Fourier-Sieve (Step 1), which by Lemma 1.8 fails with probability at most $1/8$. Running Step 1 a constant number $r = O(1)$ of times and retaining a run whose output passes the subsequent (query-free with respect to f) consistency checks drives its failure probability below $1/100$; this costs only an $O(1)$ factor, absorbed in the $\widetilde{O}(\cdot)$ query bound. Every other step then succeeds with probability at least $99/100$, and by a union bound over the $O(1)$ steps the overall tester succeeds with probability at least $2/3$.

Query Complexity

Set $\theta = \varepsilon^2\delta/(20s)$, so that $1/\theta = O(s/(\varepsilon^2\delta))$.

At Step 1, Fourier-Sieve is invoked with threshold θ. By Lemma 1.8, this uses $\widetilde{O}(1/\theta^4)$ oracle queries (times the $O(1)$ amplification factor). At Step 2, Estimate-Heavy-Fourier-Coeffs achieves estimation accuracy $\theta/8$ with confidence $1 - \mathrm{poly}(\theta)$. By Lemma 1.9, each per-point list-correction call uses $\widetilde{O}(1/\theta^2)$ queries, and since the estimator averages over $\widetilde{O}(1/\theta^2)$ points, Step 2 costs $\widetilde{O}(1/\theta^4)$ queries. At Step 4, List-Correction is invoked $T = O(1/\delta)$ times, each using $\widetilde{O}(\log(1/\delta)/\theta^2)$ queries, for a total of $\widetilde{O}\big(\frac{1}{\delta} \cdot \frac{1}{\theta^2}\big) = \widetilde{O}\Big(\frac{s^2}{\varepsilon^4\delta^3}\Big)$, which is of lower order than Steps 1–2. The dominant cost is therefore

$$\widetilde{O}(1/\theta^4) = \widetilde{O}\left(\frac{s^4}{\varepsilon^8\delta^4}\right) \text{ oracle queries,}$$

matching Theorem 3.2. The sample complexity is dominated by Step 4, which draws $T = O(1/\delta)$ samples from D, i.e. $\widetilde{O}(1/\delta)$ samples, as claimed.

This completes the proof of Theorem 3.2.

4

Application III: Improved ℓ_0 Testing of Fourier Sparsity

Boolean functions are fundamental in machine learning and theoretical computer science, as they naturally model decision rules in binary classification, logical circuits, and related computational processes. They have been extensively studied in learning theory [Kearns and Vazirani, 1994], complexity theory [O'Donnell, 2014], and cryptography [Carlet, 2020]. Their structural simplicity has enabled researchers to design efficient algorithms across many domains of computer science. Among various structural measures, one that is particularly important in learning theory is *Fourier sparsity*, which counts the number of nonzero coefficients in the Fourier expansion over $\mathbb{F}_2^n$. Formally, for $f : \mathbb{F}_2^n \to \{-1, +1\}$,

$$f(x) = \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)\, \chi_\alpha(x), \qquad \chi_\alpha(x) = (-1)^{\langle x, \alpha \rangle},$$

with Fourier support $\mathrm{supp}(\widehat{f}) = \{\alpha : \widehat{f}(\alpha) \neq 0\}$. A function is *$s$-Fourier sparse* if $|\mathrm{supp}(\widehat{f})| \leq s$.

Because many natural classes of Boolean functions are Fourier sparse, learning such functions has long been a central theme in computational learning theory. Notable examples include hypergraph cut functions [Stobbe and Krause, 2012] and bounded-depth decision trees [Mansour, 1994]. The study of learning Fourier-sparse Boolean functions dates back to the pioneering work of Kushilevitz and Mansour [1993], itself inspired by the Goldreich–Levin algorithm [Goldreich and Levin, 1989]. Over the past two decades, there has been renewed interest in exact reconstruction of Fourier-sparse Boolean functions with complexity primarily dependent on the sparsity parameter. Two main algorithmic paradigms are based on the Sparse Hadamard Transform, e.g. [Indyk et al., 2014, Hassanieh et al., 2012b], and on compressed-sensing techniques, e.g. [Ishay and Regev, 2017]. A common limitation of both, however, is the assumption that the sparsity level is known in advance. This creates a critical gap: before applying these methods, one must first ensure that the target function is indeed Fourier sparse.

Our work addresses this gap by developing efficient procedures for testing Fourier sparsity, thereby providing a natural preprocessing step for learning algorithms on Boolean functions. Specifically, we study the problem of testing Fourier sparsity of Boolean functions $f : \mathbb{F}_2^n \to \{-1, +1\}$ under black-box query access, where one can evaluate the function on inputs of one's choice without knowledge of its internal structure.

> **Problem Statement:** Given a parameter $\varepsilon > 0$ and black-box query access to a Boolean function $f : \mathbb{F}_2^n \to \{-1, +1\}$, decide whether f is s-Fourier sparse or ε-far from every s-Fourier sparse Boolean function, where distance is measured in the Hamming sense:
>
> $$\mathrm{dist}_0(g, h) = \Pr_{x \in \mathbb{F}_2^n}[g(x) \neq h(x)].$$

The efficiency of a Fourier sparsity tester is measured by its query complexity, and such testers can serve as a crucial preprocessing step for learning algorithms on Fourier-sparse Boolean functions, such as fast sparse Hadamard transforms [Scheibler et al., 2015, Hassanieh et al., 2012b].

Related Work

A related version has been studied for real-valued functions $f : \mathbb{F}_2^n \to \mathbb{R}$, where distance is measured via the Euclidean norm,

$$\mathrm{dist}_2(f, g) = \left(\mathbb{E}_x[(f(x) - g(x))^2]\right)^{1/2}.$$

Yaroslavtsev and Zhou [2020] showed that testing closeness to s-Fourier-sparse functions under this metric can be done with $\tilde{O}(s)$ queries and established a matching $\Omega(\sqrt{s})$ lower bound.

However, one should note that while closeness in the Hamming sense implies closeness in the Euclidean sense, the converse may not hold. To illustrate, consider two scenarios. In the first, an exactly k-Fourier-sparse Boolean function with precisely k nonzero Fourier coefficients; by Gopalan et al. [2011], all of these must be large in magnitude. In the second, a Boolean function with k large Fourier coefficients but also a tail of many small nonzero coefficients of small ℓ_2 norm. A k-Fourier-sparsity tester under Euclidean distance would accept both types, whereas a tester under Hamming distance should accept only the first. This makes testing Fourier sparsity in the Hamming sense significantly harder: one must certify not only the presence of k large Fourier coefficients but also the absence of a Fourier tail.

The problem of testing Fourier sparsity under Hamming distance was first studied by Gopalan et al. [2011], who gave a non-adaptive tester with query complexity $O(s^6 \log s/\varepsilon^2 + s^{14} \log s)$ and proved a lower bound of $\Omega(\sqrt{s})$. Furthermore, as Fourier sparsity is an affine-invariant property, i.e. it remains unchanged under affine transformations of the domain, it can in principle be tested using regularity-based frameworks [Kaufman and Sudan, 2008, Bhattacharyya et al., 2015, 2013, Hatami and Lovett, 2013]; however, such approaches incur impractical tower-type query complexities.

Main Theorems

More formally, we prove the following theorem.

Theorem 4.1 (Upper bound). *Let $s > 0$, $\varepsilon > 0$, and let $f : \mathbb{F}_2^n \to \{-1, +1\}$ be an unknown Boolean function accessible via queries to its truth table. There exists a non-adaptive property-testing algorithm that decides whether f is s-Fourier sparse or ε-far (under Hamming distance) from any such function, with success probability at least $2/3$, using at most $\widetilde{O}\big(\max\{s^2, 1/\varepsilon\} \cdot s^2\big)$ queries.*[1]

4.1 Proof Ideas

The approach of employing learning algorithms for property testing, introduced by Goldreich et al. [1998], is founded on the principle that any proper learning algorithm for a function class $\mathcal{C}$ can be converted into a property tester for $\mathcal{C}$. Nevertheless, while proper learning for most Boolean function classes requires at least $\Omega(\log n)$ queries, property testing typically targets sublogarithmic or even constant query complexity, independent of the ambient dimension n. A significant advancement was made by Diakonikolas et al. [2007], who introduced the paradigm of testing via implicit learning. The key insight is that many natural classes of Boolean functions, such as monotone DNFs, decision lists, decision trees, branching programs, Boolean formulas, sparse polynomials over $\mathbb{F}_2$, and Boolean circuits, admit succinct representations and are well approximated by *junta functions*. For instance, every s-term DNF is ε-close to another s-term DNF depending on only $O(\log s + \log(1/\varepsilon))$ variables. Building on this idea, Chakraborty et al. [2011b] proposed a query-efficient sample extractor that yields improved, and in certain cases optimal, bounds for testing membership in many subclasses of junta functions.

Returning to Fourier sparsity, we first recall that Gopalan et al. [2011] designed their tester by exploiting locally testable structural features of Fourier-sparse functions, namely

[1] Here, $\widetilde{O}(\cdot)$ hides factors polynomial in $\log s$.

the granularity of their nonzero Fourier coefficients, and subsequently identifying these coefficients via a hashing-based technique. In contrast, our approach is based on the paradigm of testing via implicit learning. A key challenge is that existing techniques for testing membership in subclasses of junta functions are insufficient for Fourier-sparse functions, because the class of Fourier-sparse functions is strictly more general than the class of junta functions. For example, a linear function depending on $O(n)$ variables is 1-Fourier sparse, yet is not a junta, since its output depends on a linear number of variables. Thus, testing Fourier sparsity via implicit learning necessitates new techniques beyond the junta-testing framework.

At the core of our technique lies an exact learning procedure for s-Fourier-sparse Boolean functions (see Algorithm 9), which leverages ℓ_1-minimization from compressed sensing (see Chapter 4 of Moitra [2018]). This relies on a result of Ishay and Regev [2017], which establishes that subsampled Fourier–Hadamard matrices satisfy the Restricted Isometry Property (RIP). A key implication is that, given $O(s \cdot \log^2 s \cdot n)$ uniformly random samples from an s-Fourier-sparse Boolean function, one can exactly recover the function via an appropriate ℓ_1-minimization procedure. However, a direct application is inadequate, as this sample complexity still depends on the ambient dimension n. To address this, we use a structural result of Sanyal [2019], which establishes that the dimension of the linear span of the nonzero Fourier coefficients of an s-Fourier-sparse function is at most $O(\sqrt{s})$. Hence any s-Fourier-sparse function $f : \mathbb{F}_2^n \to \{-1, +1\}$ can be expressed as

$$f = f^* \circ L, \qquad f^* : \mathbb{F}_2^r \to \{-1, +1\},$$

where $L : \mathbb{F}_2^n \to \mathbb{F}_2^r$ is a linear transformation and $r = O(\sqrt{s})$. By the RIP result of Ishay and Regev [2017], reconstruction then requires only $O(s^{3/2} \cdot \log^2 s)$ random samples of $f^* : \mathbb{F}_2^r \to \{-1, +1\}$.

Motivated by this, our approach reconstructs the composition $f \circ U \circ R$, where $U : \mathbb{F}_2^n \to \mathbb{F}_2^n$ is an unknown linear transformation and $R : \mathbb{F}_2^n \to \mathbb{F}_2^r$ is a known linear map. The key machinery is Algorithm 10, a subroutine that generates random samples from $f \circ U \circ R$ using only a small number of queries. This can be viewed as a generalization of the sampler of Chakraborty et al. [2011b] within the junta-subclass testing framework. At its core is a local list-correction procedure for the Hadamard code, which we import directly from Chapter 1; our adaptation differs substantially from Gopalan et al. [2011] in both construction and analysis, yielding a polynomial improvement in query complexity.

4.2 Our Algorithm

We begin by introducing the notion of coset sampling, which plays a key role in generating samples for the exact-learning machinery in a low-dimensional space. Let $\mathcal{O}_H^f$ denote the sampling procedure with respect to a subspace H, given query access to f, as described in Algorithm 8. We refer to the samples produced by this procedure as coset samples of f with respect to H.

Algorithm 8: SubspaceExplicitCosetSampler

Input: Subspace H with basis $B := \{\beta_1, \beta_2, \ldots, \beta_r\}$; query access to f.
Output: A uniform random sample of f restricted to a random coset of $H^\perp$.

Step 1. Sample b uniformly at random from $\mathbb{F}_2^r$ and define the coset

$$C(b) = \{\alpha \in \mathbb{F}_2^n \mid \langle \alpha, \beta_i \rangle = b_i\}.$$

Step 2. Select a uniformly random element p from $C(b)$.;
Step 3. Return the pair $(b, f(p))$.;

Let $\mathcal{S}_f$ denote the subspace spanned by the vectors corresponding to the nonzero Fourier coefficients of f. The coset samples generated by $\mathcal{O}_{\mathcal{S}_f}^f$ can be interpreted as uniform samples drawn from the leaves of $\mathcal{T}_f$, the non-adaptive parity decision tree representation of f. For a parameter $\theta > 0$, the θ-restricted Fourier span $\mathcal{S}_f(\theta)$ is the subspace spanned by those $\alpha \in \mathbb{F}_2^n$ with $|\widehat{f}(\alpha)| \geq \theta$. We build a query-efficient implementation of $\mathcal{O}_{\mathcal{S}_g(\theta)}^g$, given query access to f and a threshold $\theta > 0$, where $g = f \circ A$ for some unknown invertible linear transformation $A : \mathbb{F}_2^n \to \mathbb{F}_2^n$.

Lemma 4.2. *There exists an algorithm SubspaceImplicitCosetSampler (Algorithm 10) that, given query access to a Boolean function $f : \mathbb{F}_2^n \to \{-1, +1\}$, a threshold $\theta > 0$, and a parameter $\lambda \in \mathbb{N}$, generates a set of λ uniform coset samples ξ with respect to the subspace $\mathcal{S}_{f\circ A}(\theta)$, for a function $f \circ A$ where A is some non-singular (and possibly unknown) linear transformation of $\mathbb{F}_2^n$. The query complexity is $\widetilde{O}\left(\frac{1}{\theta^4} + \max\left\{\frac{1}{\theta^2}, \lambda\right\} \cdot \frac{1}{\theta^2}\right)$, and the failure probability is at most $\frac{1}{10}$.*

The proof of Lemma 4.2 is deferred to Section 4.2.1; we assume its validity for now and use it to establish Theorem 4.1.

Proof. [Proof of Theorem 4.1] In Algorithm 9, the subroutine SubspaceImplicitCosetSampler (Algorithm 10) is invoked with parameters θ and λ. By Lemma 4.2, this produces λ uniformly random coset samples (as in Algorithm 8) with respect to $\mathcal{S}_{f\circ A}(\theta)$. Using these, we define a lower-dimensional Boolean function $f^* : \mathbb{F}_2^r \to \{-1, 1\}$, where $f^* = f \circ T \circ A$,

and $T : \mathbb{F}_2^n \to \mathbb{F}_2^r$ is a linear transformation mapping the standard basis vectors e_i^n to the corresponding e_i^r, for $i \in \{1, \ldots, r\}$. The map T also sends each of the 2^r cosets of $\mathcal{S}_{f\circ A}(\theta)^\perp$ to unique elements of $\mathbb{F}_2^r$.

The next step is to determine whether f^* is s-Fourier sparse. It is folklore (see Chapter 4 of Moitra [2018]) that for any Boolean function $h : \mathbb{F}_2^r \to \{-1, +1\}$ with at most s nonzero Fourier coefficients, if there is an $m \times 2^r$ subsampled Walsh–Hadamard matrix M satisfying RIP of order k with constant $\delta_k < \frac{1}{3}$ (and $\delta_{2k} + \delta_{3k} < 1$), then the Fourier spectrum of h can be recovered exactly with high probability by solving

$$\widehat{h} = \arg\min \|\widehat{h}\|_1 \quad \text{subject to} \quad M\widehat{h} = y,$$

using m uniformly random labeled examples $(x, y = h(x))$.

A bound on m determines the number of required random examples and hence the number of coset samples needed from $f \circ A$. By Sanyal [2019], the dimension of the span of the nonzero Fourier coefficients of any s-Fourier-sparse function is at most $O(\sqrt{s})$. Combined with the RIP result of Ishay and Regev [2017], $m = O\left(s^{3/2} \cdot \mathrm{poly}(\log s)\right)$ random samples suffice for exact recovery of f^*.

Lemma 4.3 (Ishay and Regev, 2017)**.** *Let $M \in \mathbb{C}^{N\times N}$ be a unitary matrix satisfying $\|M\|_\infty \le O(1/\sqrt{N})$, and let $\delta > 0$ be sufficiently small. Construct a measurement matrix $B \in \mathbb{C}^{q\times N}$ by selecting q rows uniformly and independently from M, each scaled by $\sqrt{N/q}$. If*

$$q = O\left(\log^2(1/\delta) \cdot \delta^{-2} \cdot k \cdot \log^2(k/\delta) \cdot \log N\right),$$

then, with probability at least $1 - 2^{-\Omega(\log N \cdot \log(k/\delta))}$, the matrix B satisfies the Restricted Isometry Property of order k with isometry constant δ.

We now show that if f is s-Fourier sparse, the algorithm always accepts.

Lemma 4.4 (Completeness)**.** *If f is s-Fourier sparse, then Algorithm 9 accepts.*

Proof. By Gopalan et al. [2011], if f is s-Fourier sparse, then all of its nonzero Fourier coefficients are integer multiples of $\frac{1}{2^{\lceil \log s \rceil}}$; in particular, each has magnitude at least $\frac{1}{2^{\lceil \log s \rceil}} \ge \frac{1}{2s} = 2\theta$. Moreover, Fourier sparsity is invariant under nonsingular linear transformations. When SubspaceImplicitCosetSampler (Algorithm 10) is invoked with $\theta = \frac{1}{4s}$, Lemma 4.2 guarantees that it discovers the Fourier span of $f \circ A$ for some unknown invertible A (every nonzero coefficient, being $\ge 2\theta$, is captured). It also produces $\lambda = \max\{1/\varepsilon,\ O(s^2 \cdot \mathrm{poly}(\log s))\}$ uniform coset samples. By the sparse-recovery guarantee above, these suffice to exactly reconstruct the Fourier spectrum of $f^* = f \circ T \circ A$. Hence the algorithm accepts. □

Algorithm 9: EfficientFourierSparsityTester

Input: Fourier sparsity s; query access to f.
Output: Whether f is s-Fourier sparse or ε-far from every such function.
Initialization: $\theta = \dfrac{1}{4s}, \quad \lambda = \max\left\{\dfrac{1}{\varepsilon},\ Cs^2 \cdot \mathrm{poly}(\log s)\right\}$ for a large constant C.

Step 1. $(\zeta, r) \leftarrow$ SubspaceImplicitCosetSampler(θ, λ).;
Step 2. Initialize Φ such that for each $(x, y) \in \zeta$ and each $z \in \mathbb{F}_2^r$,

$$\Phi[x][z] = \frac{2^{r/2}}{\lambda^{1/2}} \cdot (-1)^{\langle x,z\rangle}.$$

Step 3. Solve the optimization problem

$$\min \|\widehat{h}\|_1 \quad \text{subject to} \quad \langle \Phi[x][*],\, \widehat{h}\rangle = y \quad \text{for all } (x, y) \in \zeta.$$

Step 4. If $\widehat{h}$ corresponds to the Fourier spectrum of some s-Fourier-sparse Boolean function, **Accept**; otherwise, **Reject**.;

Lemma 4.5 (Soundness)**.** *If f is ε-far from every s-Fourier-sparse Boolean function, then Algorithm 9 rejects.*

Proof. We argue by contraposition. Suppose Algorithm 9 accepts. Then there exists a function $\tilde{f}$, obtained via composition with appropriate linear transformations, that agrees with f on all λ uniform samples collected by Algorithm 10 and is s-Fourier sparse. Define **GOOD** as the set of inputs where $\tilde{f}$ and f agree, and **BAD** as its complement. The probability that all λ samples fall in **GOOD** is

$$\left(1 - \frac{|\mathbf{BAD}|}{2^n}\right)^{\lambda}.$$

If this is at least a constant, then $\frac{|\mathbf{BAD}|}{2^n} \le \frac{1}{O(\lambda)}$, so f is $\frac{1}{O(\lambda)}$-close to the s-Fourier-sparse function $\tilde{f}$, contradicting ε-farness for $\lambda \ge 1/\varepsilon$. □

We now analyze the query complexity. With $\theta = \frac{1}{4s}$ and $\lambda = \max\{1/\varepsilon,\ \widetilde{O}(s^2)\}$, the only queries are via Algorithm 10. By Lemma 4.2,

$$\widetilde{O}\left(\frac{1}{\theta^4} + \max\left\{\frac{1}{\theta^2},\ \lambda\right\} \cdot \frac{1}{\theta^2}\right) = \widetilde{O}\left(s^4 + \max\left\{s^2,\ \tfrac{1}{\varepsilon}\right\} \cdot s^2\right) = \widetilde{O}(\max\{s^2, 1/\varepsilon\} \cdot s^2),$$

matching Theorem 4.1. Algorithm 9 fails with probability at most $\frac{1}{10} + o(1) < \frac{1}{3}$. □

4.2.1 Analysis of SubspaceImplicitCosetSampler

At a high level, SubspaceImplicitCosetSampler is built on the local list-correction pipeline of Chapter 1. Given query access to f, we first implicitly identify the cosets carrying heavy Fourier coefficients using Fourier-Sieve (Algorithm 1, Lemma 1.8); we then evaluate the dominant character of each such coset at requested points via List-Correction (Algorithm 2, Lemma 1.9); finally we recover a basis for the heavy coefficients and all linear relations among them via Identify-modulo-Invariance (Algorithm 4, Lemma 1.14). The improvement in query complexity over Gopalan et al. [2011] comes from the refined coset-hashing concentration bounds of Chapter 1 (Claims 1.6 and 1.7), which show that within any coset the ℓ_2-mass and ℓ_1-magnitude of the Fourier projection are dominated by the heaviest coefficient. Our analysis is inspired by heavy-hitter detection in the streaming literature: ℓ_2 concentration plays a role analogous to the Count–Min Sketch, while ℓ_1 concentration corresponds to the Count Sketch. Below is a sketch of the algorithm.

- **Detecting Heavy Fourier Coefficients.** Fourier-Sieve projects the Fourier spectrum of f onto cosets of a random subspace and returns a set $\mathcal{C}$ of cosets containing every coefficient of magnitude at least θ, each with a dominant coefficient of magnitude $\geq \theta/2$ (Lemma 1.8), using $\widetilde{O}(1/\theta^4)$ queries. Correctness rests on Claim 1.6, which shows the ℓ_2-mass of each coset is concentrated on its dominant coefficient.
- **Evaluating Heavy Fourier Coefficients.** For each surviving coset C and query point x, List-Correction returns $\chi_{\alpha_C}(x)$ without knowing α_C explicitly (Lemma 1.9), relying on the ℓ_1-concentration of Claim 1.7 together with the sign-preservation guarantee of Claim 1.10. Since a single trial succeeds only with constant probability, the median amplification built into List-Correction boosts this to $1 - 1/\mathrm{poly}(1/\theta)$.
- **Generating Coset Samples.** Given evaluations of the heavy characters at enough points, Identify-modulo-Invariance recovers a linear basis among the heavy Fourier coefficients, identifying them up to a linear transformation (Lemma 1.14). We show this suffices to emit coset samples of $f \circ A$.

Proof. [Proof of Lemma 4.2] We analyze Algorithm 10.

Correctness of Steps 1–3. By Lemma 1.8, Step 1 returns, except with probability at most $1/8$, a set $\mathcal{C}$ of cosets such that every α with $|\widehat{f}(\alpha)| \geq \theta$ lies in some $C \in \mathcal{C}$, and every $C \in \mathcal{C}$ has dominant coefficient α_C with $|\widehat{f}(\alpha_C)| \geq \theta/2$ and residual coset mass $< \theta^4/2^{12}$. By Lemma 1.9, each call in Step 2 returns the correct signs $\chi_{\alpha_C}(x_i)$ for all $C \in \mathcal{C}$ simultaneously, except with probability $1/(40\kappa)$; a union bound over the κ points gives

Algorithm 10: SubspaceImplicitCosetSampler

Input: Threshold θ; number of samples λ; query access to f.
Output: Coset samples of $f \circ A$ with respect to the subspace $\mathcal{S}(\theta)$, as (H, ζ, r).
Initialization: $\kappa = \max\{1/\theta^2,\ \lambda\}$.

Step 1. Run Fourier-Sieve (Algorithm 1) on f with threshold θ to obtain the set of cosets $\mathcal{C}$ capturing all α with $|\widehat{f}(\alpha)| \geq \theta$ (Lemma 1.8).;
Step 2. For $i = 1, \ldots, \kappa$:;
Sample x_i uniformly from $\mathbb{F}_2^n$ and set $F[x_i] \leftarrow f(x_i)$.;
Run List-Correction (Algorithm 2) at x_i with confidence $1 - 1/(40\kappa)$ to obtain $Q[x_i][C] = \chi_{\alpha_C}(x_i)$ for all $C \in \mathcal{C}$ (Lemma 1.9).;
Step 3. Run Identify-modulo-Invariance (Algorithm 4) to relabel the columns of Q as $\{B_1, \ldots, B_k\}$ with basis $B_1 = e_1^n, \ldots, B_r = e_r^n$, each B_i $(i > r)$ a recorded linear combination of $\{B_1, \ldots, B_r\}$ (Lemma 1.14).;
Step 4. For each $x \in \{x_1, \ldots, x_\kappa\}$:;
Let $b \in \mathbb{F}_2^r$ with $b_j = \dfrac{1 - Q[x][B_j]}{2}$ for all $j \in [r]$.;
$\zeta \leftarrow \zeta \cup \{(b, f(x))\}$.;
Step 5. Construct $H \in \mathbb{F}_2^{r \times n}$ with rows $B_1, \ldots, B_r$.;
Step 6. Return (H, ζ, r).;

failure at most $1/40$ for Step 2. By Lemma 1.14, Step 3 recovers the correct basis and linear relations except with probability at most $1/40$.

From characters to coset labels. We cannot determine α_C explicitly, but the linear relations from Step 3 identify them up to a linear transformation. In Step 3 we extracted a basis of size r for the span of the α_C and relabeled it as the standard basis $e_1^n, \ldots, e_r^n$. Let H be the subspace they span, with

$$M = \begin{bmatrix} e_1^n \\ e_2^n \\ \vdots \\ e_r^n \end{bmatrix} \in \mathbb{F}_2^{r \times n}.$$

Then, for every $x \in \{x_1, \ldots, x_\kappa\}$, the coset of $H^\perp$ containing x is determined by $Mx = b$, with b as in Step 4. It remains to verify that the induced distribution over cosets is uniform.

Lemma 4.6. *Given any subspace $G \subseteq \mathbb{F}_2^n$, the distribution of cosets induced by a uniformly random $x \in \mathbb{F}_2^n$ is uniform over the cosets of G.*

Proof. Let $G \subseteq \mathbb{F}_2^n$ have dimension k. Its cosets are $C_z = z + G$, of which there are 2^{n-k}.

For uniform x, since $|C_z| = 2^k$,

$$\Pr(x \in C_z) = \frac{|C_z|}{|\mathbb{F}_2^n|} = \frac{2^k}{2^n} = \frac{1}{2^{n-k}},$$

identical for all cosets. Hence the induced distribution is uniform. □

By Lemma 4.6 applied to $G = H^\perp$, the pairs $(b, f(x_i))$ collected in Step 4 are exactly uniform coset samples of $f \circ A$ with respect to $\mathcal{S}(\theta)$, as required.

Failure probability. Step 1 (Fourier-Sieve) is the only subroutine whose native failure probability exceeds a small constant: by Lemma 1.8 it is at most $1/8$. Running Step 1 a constant number of times and retaining a run whose output is consistent with the subsequent checks drives its failure probability below $1/100$; this costs only an $O(1)$ query factor. Steps 2 and 3 each fail with probability at most $1/40$, and the remaining estimates hold except with probability $o(1)$. By a union bound the total failure probability is at most $\frac{1}{100} + \frac{1}{40} + \frac{1}{40} + o(1) < \frac{1}{10}$, as claimed.

Query complexity. All queries to f are made in Steps 1–2. By Lemma 1.8, Step 1 uses $\widetilde{O}(1/\theta^4)$ queries. By Lemma 1.9, each List-Correction call in Step 2 uses $\widetilde{O}(1/\theta^2)$ queries, and Step 2 makes $\kappa = \max\{1/\theta^2, \lambda\}$ such calls, for $\widetilde{O}(\max\{1/\theta^2, \lambda\} \cdot 1/\theta^2)$ queries; moreover a single shared sample batch serves all cosets at a given point. Combining,

$$\widetilde{O}\left(\frac{1}{\theta^4} + \max\left\{\frac{1}{\theta^2},\ \lambda\right\} \cdot \frac{1}{\theta^2}\right),$$

matching the statement of Lemma 4.2. □

5

Application IV: Estimating Sumset Size

Many problems in theoretical computer science involve objects of very large complexity, such as graphs with many vertices or functions defined on high-dimensional spaces. Typically, one seeks structural characterizations that hold in the worst case, rather than results that apply only to random instances (average-case behavior) or to highly structured special cases. For large objects this is inherently challenging, as they can exhibit a wide range of behaviors. At one extreme are highly structured objects, displaying strong algebraic or combinatorial regularities; at the other are pseudorandom objects, which closely mimic truly random ones under appropriate statistical tests.

A recurring phenomenon is that these two extremes are not disjoint. In many settings there is a *structure-versus-randomness* dichotomy: for a given class of objects and a suitable notion of structure, there exists a complementary notion of pseudorandomness such that every object admits a decomposition into a structured component and a pseudorandom component, possibly up to a small error; see Tao [2007] for an extensive discussion. The usefulness of such decompositions lies in the fact that different techniques apply to different components. Structured parts can often be analyzed using algebraic or low-complexity arguments, while pseudorandom parts can be controlled using analytic inequalities and probabilistic tools.

Many classical theorems already reflect this philosophy. Ramsey theory provides one of the earliest illustrations of the structure-versus-randomness principle. The Ramsey number $R(k, k)$ is the smallest integer N such that every graph on N vertices contains either a clique of size k or an independent set of size k. Ramsey's theorem guarantees that every graph on N vertices contains either a clique or an independent set of size at least $\Omega(\log N)$. A closely related phenomenon occurs in arithmetic combinatorics. Roth's theorem [Roth, 1953] shows that any subset of $\{1, \ldots, N\}$ of density at least $\Omega(1/\log\log N)$ contains a nontrivial three-term arithmetic progression. Szemerédi's theorem [Szemerédi, 1975] strengthens this dramatically by proving that dense sets contain arithmetic progressions of arbitrary length. Here again, even when a set is chosen adversarially, density alone enforces strong additive structure.

Among the many formalizations of the structure-versus-randomness paradigm, perhaps the most influential is *Szemerédi's regularity lemma* [Szemerédi, 1975]. It asserts that every dense graph admits a partition into a bounded number of parts such that most pairs of parts behave pseudorandomly, in the sense that their edge distributions mimic those of random graphs. This lemma has had a profound impact on graph theory and has found numerous applications in theoretical computer science, e.g. [Alon et al., 2001]. However, the quantitative bounds are extremely poor: the number of parts required grows as a tower of twos of height polynomial in $1/\varepsilon$, where ε is the error parameter.

Inspired by Szemerédi's regularity lemma for graphs, Green introduced an *arithmetic regularity lemma* for functions on $\mathbb{F}_2^n$ [Green, 2005]. In this setting, pseudorandomness is defined using Fourier analysis. Writing the Fourier expansion of a function $f : \mathbb{F}_2^n \to \mathbb{R}$ as

$$f(x) = \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)\,(-1)^{\langle \alpha, x \rangle},$$

the function f is said to be δ*-regular* if all its nontrivial Fourier coefficients satisfy

$$|\widehat{f}(\alpha)| \le \delta \qquad \text{for all } \alpha \neq 0.$$

Green's theorem asserts that every Boolean function $f : \mathbb{F}_2^n \to \{0,1\}$ admits a partition of $\mathbb{F}_2^n$ into affine subspaces such that on most parts the restricted function is δ-regular. As in the graph case, the quantitative dependence is extremely poor: the required codimension grows as a tower of twos whose height is polynomial in $1/\delta$. Moreover, Green showed that such tower-type dependence is unavoidable within this framework, proving a lower bound of $\mathrm{tower}(\Omega(\log(1/\delta)))$ on the codimension. This was later strengthened by Hosseini et al. [2016], who exhibited examples requiring a tower of height $\Theta(1/\delta)$.

A key insight from more recent work is that regularity phenomena become dramatically stronger when one restricts attention to objects of *bounded complexity*. This perspective is reminiscent of the fixed-parameter tractability (FPT) paradigm in algorithmic complexity: by bounding a suitable complexity measure of the underlying object, one can efficiently address problems that are otherwise intractable in the worst case. Similar approaches have been developed both for graphs and for functions on high-dimensional domains, where imposing appropriate complexity bounds leads to substantially stronger structural and regularity results.

Before proceeding further, we note that any Boolean function $f : \mathbb{F}_2^n \to \{0,1\}$ can be

naturally identified with a subset $A \subseteq \mathbb{F}_2^n$ via

$$A = \{x \in \mathbb{F}_2^n : f(x) = 1\}.$$

One important notion of bounded complexity for such sets is the *VC dimension*. For a subset $A \subseteq \mathbb{F}_2^n$, the VC dimension of the family of translates $\{A + x : x \in \mathbb{F}_2^n\}$ captures the *expressivity* of the induced set system. Alon et al. [2019] showed that if this family has VC dimension at most d, then for every $\varepsilon > 0$ there exists a subgroup $H \leq \mathbb{F}_2^n$ of index at most $\varepsilon^{-O_d(1)}$ such that A can be modified, by adding or deleting at most $\varepsilon 2^n$ elements, to become a union of cosets of H. Another prominent notion of structural complexity arises from additive combinatorics, through the study of sumsets and doubling constants. Freiman–Ruzsa theory asserts that sets with small doubling are efficiently approximable by low-dimensional algebraic objects. Over vector spaces $\mathbb{F}_p^n$, this intuition is formalized by the Polynomial Freiman–Ruzsa conjecture, which posits that a set with *doubling constant* K is contained in a subspace of size polynomial in K times the size of the set. Strong polynomial Freiman–Ruzsa-type results for Abelian groups with bounded torsion have recently been established in a series of papers by Gowers et al. [Gowers et al., 2025, 2024]. More recently, Arunachalam et al. [2026] gave algorithmic versions of polynomial Freiman–Ruzsa-type results over $\mathbb{F}_2^n$.

In this work, we study regularity through the lens of the *Fourier spectral norm*. For a Boolean function $f : \mathbb{F}_2^n \to \{0, 1\}$, the spectral norm is

$$\|\widehat{f}\|_1 = \sum_{\xi \in \mathbb{F}_2^n} |\widehat{f}(\xi)|,$$

which measures the total ℓ_1-mass of the Fourier spectrum of f. Bounding this quantity imposes strong global constraints on the distribution of Fourier mass and serves as a natural analytic measure of complexity. Functions with small spectral norm arise naturally in learning theory [Bruck and Smolensky, 1992, Kushilevitz and Mansour, 1993, Ada et al., 2012] and communication complexity [Chattopadhyay et al., 2020, Tsang et al., 2013, Shpilka et al., 2017]. Furthermore, tail bounds on the Fourier ℓ_1-norm have recently played an important role in the construction of pseudorandom generators [Girish et al., 2021b, Meka et al., 2019, Chattopadhyay et al., 2019, Forbes and Kelley, 2018] and in separating quantum and classical models of computation [Bansal and Sinha, 2021, Tal, 2020]. The additive structure of sets with small spectral norm has also been a subject of long-standing interest in harmonic analysis [McGehee et al., 1981].

A foundational result in this direction is Cohen's idempotent theorem [Cohen, 1960], originally proved for locally compact Abelian groups, which provides a structural charac-

terization of Boolean functions with bounded spectral norm. In the setting of the Boolean cube, Cohen's idempotent theorem specializes as follows: any set $A \subseteq \mathbb{F}_2^n$ whose indicator function has bounded spectral norm can be expressed as a Boolean combination of a constant number of affine subspaces, where the constant depends only on the norm bound. A quantitative version of this theorem was later established by Green and Sanders [Green and Sanders, 2008a,b]. In particular, they showed that if $f : \mathbb{F}_2^n \to \{0,1\}$ is a Boolean function with spectral norm $\|\widehat{f}\|_1 \leq m$, then

$$f = \sum_{j=1}^{L} \mathbb{1}_{a_j + H_j},$$

where $L \leq 2^{2^{O(m^4)}}$ and each $a_j + H_j$ is a coset of some subgroup of $\mathbb{F}_2^n$. Subsequent works [Sanders, 2019, 2020] improved the dependence on m. While Cohen's idempotent theorem and its quantitative refinements provide a complete structural characterization of Boolean functions with bounded spectral norm, the resulting decompositions may involve cosets arising from different subspaces. In particular, these results do not impose any restriction that the cosets be parallel, that is, generated by a single common subspace.

Returning to our setting, we investigate the structural complexity of such decompositions under the additional requirement that all cosets arise from a single common subspace, in the spirit of Green's arithmetic regularity lemma [Green, 2005]. Our main result shows that if one allows a controlled approximation, then such a decomposition is indeed possible with only a mild loss in parameters. We state the result formally in the following subsection, along with some immediate applications to the design of sublinear-query algorithms for two classical problems.

Main Theorems

Dense coset decomposition. We show that if a set $A \subseteq \mathbb{F}_2^n$ has bounded spectral norm, then after discarding only a small fraction of elements, A is concentrated on some cosets of a certain low-codimension subspace, and on each such coset the set is almost full. Recall that quantitative analogues of Cohen's idempotent theorem due to Shpilka et al. [2017] and Tsang et al. [2013] imply that for every Boolean function with spectral norm at most m, there exists an affine subspace of codimension $O(m)$ on which the function is constant, providing a worst-case structural guarantee. In contrast, our result gives an *average-case* structural statement: with a modest blow-up in parameters, the function is approximately constant on a large fraction of the cosets of a common low-codimension subspace.

Theorem 5.1 (Dense coset decomposition theorem). *Let $A \subseteq \mathbb{F}_2^n$ be a set with indicator*

function $\mathbf{1}_A : \mathbb{F}_2^n \to \{0,1\}$*, volume* $\mathrm{Vol}(A) = \rho$*, and spectral norm* $\|\widehat{\mathbf{1}}_A\|_1 \le m$*. Then, for any* $\omega > 0$*, there exist a subset* $A' \subseteq A$ *with* $\frac{|A \setminus A'|}{|A|} \le 2\omega$ *and a subspace* H *of* $\mathbb{F}_2^n$ *of codimension at most*

$$O\left(\frac{m^2}{\rho\omega^4}\right),$$

such that A' *is the union of the restrictions of* A *to a* $\rho(1-\omega)$ *fraction of the cosets of* H*, on each of which the relative density of* A *is at least* $1-\omega$*.*

As a direct corollary, we obtain a qualitatively new approximation result: if a small error is allowed, then any Boolean function with small spectral norm can be approximated by a non-adaptive parity decision tree of depth polynomial in the spectral norm and the error parameter.

Corollary 5.2. *Let* $f : \mathbb{F}_2^n \to \{0,1\}$ *be a Boolean function and* $\varepsilon \in (0,1)$*. Then there exists a Boolean function* $g : \mathbb{F}_2^n \to \{0,1\}$ *with non-adaptive parity decision tree depth* $O\left(\|\widehat{f}\|_1^2/\varepsilon^5\right)$ *satisfying* $d_H(f,g) \le \varepsilon$*, where* $d_H(f,g) := 2^{-n}\left|\{x \in \mathbb{F}_2^n : f(x) \ne g(x)\}\right|$*.*

In contrast, Shpilka et al. [2017] showed that any Boolean function $f : \mathbb{F}_2^n \to \{0,1\}$ can be computed *exactly* by a possibly adaptive parity decision tree of depth at most $O\left(\|\widehat{f}\|_1^2 \log s_f\right)$, where s_f denotes the Fourier sparsity of f. This bound can be arbitrarily large in terms of the spectral norm, since there exist natural Boolean functions such as AND and OR that have $O(1)$ spectral norm while their Fourier sparsity is 2^n. Thus our corollary shows that, once a small approximation error is permitted, one can obtain much shallower, non-adaptive parity decision trees based only on the spectral norm, even for functions whose Fourier sparsity is maximal. As another corollary of Theorem 5.1, we obtain an approximate version of Cohen's idempotent theorem.

Corollary 5.3. *Let* $f : \mathbb{F}_2^n \to \{0,1\}$ *be a Boolean function and* $\varepsilon \in (0,1)$*. Then there exists a family of affine spaces* $H_1, \dots, H_L$ *of* $\mathbb{F}_2^n$*, where* $L = \exp\left(\|\widehat{f}\|_1^2/\varepsilon^5\right)$*, such that* $d_H\left(f, \sum_{i=1}^{L} \mathbb{1}_{H_i}\right) \le \varepsilon$*, where* $\mathbb{1}_{H_i}$ *denotes the indicator function of* H_i*.*

The best known quantitative bound for Cohen's idempotent theorem over $\mathbb{F}_2^n$ is due to Sanders [2019], who showed that when $\varepsilon = 0$, one can take $L = \exp\left(\|\widehat{f}\|_1^{3+o(1)}\right)$. We show that if one is content with approximate reconstruction, then Sanders' bound can be improved whenever $\varepsilon^5 \gtrsim \|\widehat{f}\|_1^{-1}$.

Finally, we emphasize that Theorem 5.1 can be made constructive via the local list-correction framework developed in Chapter 1 (see also [Datta et al., 2026b]). This leads to sublinear-query algorithms for testing several linear-invariant properties of a set, given only membership-query access to its indicator function. In particular, we obtain a sublinear-query algorithm for estimating the size of the sumset.

Sumset estimation. We study the problem of estimating the size of the sumset $A + A$ for a set $A \subseteq \mathbb{F}_2^n$, given black-box access to its indicator function. To avoid instability caused by noise, we focus on a certain large subset $A' \subseteq A$ and estimate the normalized sumset size $|A' + A'|/2^n$ up to additive error ε. Previous work of De et al. [2022] achieved such an estimate using Green's arithmetic regularity lemma, resulting in query complexity that is a tower function of $\mathrm{poly}(1/\varepsilon)$. In contrast, for sets with bounded spectral norm, we obtain a substantially more efficient algorithm.

Theorem 5.4. *Let $A \subseteq \mathbb{F}_2^n$ satisfy $\|\widehat{\mathbf{1}}_A\|_1 \le m$, and let $\varepsilon > 0$. Then there exists a subset $A' \subseteq A$ with $\mathrm{Vol}(A \setminus A') \le \varepsilon$ such that $\mathrm{Vol}(A' + A')$ can be estimated to additive accuracy $\pm\varepsilon$ with probability at least $2/3$ using $\widetilde{O}(m^4/\varepsilon^{12})$ queries to $\mathbf{1}_A$.*

Organization

The remainder of the chapter is organized as follows. In Section 5.1, we prove our main result, Theorem 5.1, establishing the dense coset decomposition. The list-correction mechanism that makes our approach constructive is the subject of Chapter 1, which we invoke as a black box. In Section 5.2, we describe the sumset-estimation algorithm and prove Theorem 5.4.

Related Work

If we relax our goal and ask only for the existence of an affine subspace on which a function is regular on *some* coset, rather than requiring regularity on almost all cosets as in classical arithmetic regularity lemmas, we arrive at a weaker notion that has been extensively studied across several areas of theoretical computer science, most notably complexity theory, pseudorandomness, and cryptography. Intuitively, classical regularity formulations enforce a strong form of global regularity, requiring that a function behave pseudorandomly on almost all cosets of a subspace. In contrast, the single-coset viewpoint captures a fundamentally different, worst-case notion of behavior: rather than imposing structure everywhere, it guarantees the existence of a relatively large region exhibiting regularity or structure, even when the function as a whole may be adversarial. This perspective is particularly well suited to applications in pseudorandomness, cryptography, and complexity theory, where worst-case guarantees are often more relevant than average-case structure.

To formalize this idea, for $\delta \ge 0$ we define the *δ-normality* of a function $f : \mathbb{F}_2^n \to \mathbb{R}$ by

$$\mathrm{norm}_\delta(f) \;:=\; \min\big\{\mathrm{codim}(U) : U \subseteq \mathbb{F}_2^n \text{ is an affine subspace and } f|_U \text{ is } \delta\text{-regular}\big\}.$$

Thus $\mathrm{norm}_\delta(f)$ measures the minimum number of linear constraints that must be imposed on the input in order to eliminate detectable structure and obtain a pseudorandom restriction of f.

When $\delta = 0$, this reduces to the requirement that the function be constant on a large affine subspace. In the cryptographic literature, this property is known as *normality*. More generally, an n-variable Boolean function is called *k-normal* if it is constant on some affine subspace of dimension k, and *k-nonnormal* if no such subspace exists. The notion of normality was introduced by Dobbertin [1995] in the context of cryptanalytic attacks on stream ciphers, and was subsequently investigated in greater depth in the cryptographic literature; see, for example, [Mihaljević et al., 2012, Carlet, 2004]. From a complexity-theoretic perspective, the case $\delta = 0$ has also appeared under the name *parity kill number*, denoted $C^{\oplus}_{\min}[f]$. This parameter is the smallest number of linear parities that must be fixed in order to make the function constant. The quantity $\mathrm{norm}_0(f)$ is closely connected to several fundamental notions in pseudorandomness. In particular, an *affine disperser* [Shaltiel, 2011] of dimension k is a coloring of $\mathbb{F}_2^n$ such that no affine subspace of dimension k is monochromatic. Viewing such a disperser as a function $f : \mathbb{F}_2^n \to \{0, 1, \ldots, C\}$, its disperser dimension is $k = n - \mathrm{norm}_0(f) + 1$. Related notions include affine extractors [Bourgain, 2007], which are stronger objects required for extracting randomness from a small number of independent sources, as well as k-wise independent random variables from computational number theory [Alon et al., 1992].

It is known that almost all Boolean functions are k-nonnormal for $k = O(\log n)$; that is, for a uniformly random Boolean function, with overwhelming probability there is no affine subspace of dimension $\Theta(\log n)$ on which the function is constant. On the other hand, constructing such functions explicitly is difficult, and many simple or highly structured classes of functions fail to exhibit this property. For example, every quadratic Boolean function on $\mathbb{F}_2^n$ is $n/2$-normal when n is even. More generally, Cohen and Tal [2015] showed that Boolean $\mathbb{F}_2$-polynomials of degree d cannot serve as good affine dispersers, proving that such functions admit large monochromatic affine subspaces. On the other hand, every symmetric Boolean function — that is, a function invariant under permutations of its input bits and depending only on the Hamming weight — is $\lfloor n/2 \rfloor$-normal [Canteaut and Videau, 2005]. As already mentioned, Shpilka et al. [2017] showed that any Boolean function with spectral norm at most m has an affine subspace of codimension $O(m)$ on which the function is constant.

When $\delta > 0$, general regularity results are also known. Girish et al. [2021a] showed that for any function $f : \mathbb{F}_2^n \to \mathbb{R}$, one can partition $\mathbb{F}_2^n$ into affine subspaces of codimension $O(1/\delta^3)$ such that on all but a small fraction of the parts, the restriction of f is δ-regular.

Stronger bounds are available for more restricted classes of functions. In the Boolean setting, if $f : \mathbb{F}_2^n \to \{-1, +1\}$ has Fourier degree d, then $\mathrm{norm}_\delta(f) \leq O(d^3)$. This follows from the polynomial relationship between Fourier degree and decision-tree depth. Notably, this bound is independent of δ, but relies crucially on the Boolean nature of the function.

5.1 Dense Coset Decomposition Theorem

Throughout this section, we work with Fourier analysis on the quotient by a subspace of $\mathbb{F}_2^n$. Let $U, V \leq \mathbb{F}_2^n$ be supplementary subspaces such that

$$\mathbb{F}_2^n = U \oplus V,$$

and let $A \subseteq \mathbb{F}_2^n$ be a set with indicator function $\mathbf{1}_A : \mathbb{F}_2^n \to \{0, 1\}$. Define the global density of A by

$$\rho := \mathop{\mathbb{E}}_{x \in \mathbb{F}_2^n}[\mathbf{1}_A(x)] = \frac{|A|}{2^n} = \mathrm{Vol}(A).$$

For each $a \in V$, define the relative coset density of A on the coset $a + U$ by

$$\rho_a := \mathop{\mathbb{E}}_{z \in U}[\mathbf{1}_A(a + z)].$$

This lets us view the distribution of A on each coset of U individually, reducing questions about the global behavior of $\mathbf{1}_A$ to local analysis on cosets of U, via the Fourier spectrum on the quotient space $\mathbb{F}_2^n/U \cong V$.

Before proving our Dense Coset Theorem, we establish a few useful lemmas.

5.1.1 Structural Lemmas for Quotient Spaces

We first show that Fourier concentration on $U^\perp$ essentially captures the average squared relative densities of A on cosets $a + U$, for $a \in V$.

Claim 5.5. *The total Fourier weight of $\mathbf{1}_A$ on $U^\perp$ equals the average squared density of A over the cosets of U:*

$$\sum_{\gamma \in U^\perp} \widehat{\mathbf{1}_A}(\gamma)^2 \;=\; \mathop{\mathbb{E}}_{a \in V}\left[\Big(\mathop{\mathbb{E}}_{z \in U}[\,\mathbf{1}_A(a + z)\,] \Big)^2 \right].$$

Proof. By definition of the Fourier transform,

$$\widehat{\mathbf{1}_A}(u)^2 = \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_2^n}[\,\mathbf{1}_A(x)\,\mathbf{1}_A(y)\,\chi_u(x+y)\,].$$

Summing over $u \in U^\perp$,

$$\sum_{u\in U^\perp}\widehat{\mathbf{1}_A}(u)^2 = \mathop{\mathbb{E}}_{x,y\in\mathbb{F}_2^n}\Big[\mathbf{1}_A(x)\mathbf{1}_A(y)\sum_{u\in U^\perp}\chi_u(x+y)\Big].$$

By orthogonality of characters on $U^\perp$, $\sum_{u\in U^\perp}\chi_u(z) = |U^\perp|\,\mathbf{1}_{\{z\in U\}}$, so only pairs with $x+y\in U$ contribute:

$$\sum_{u\in U^\perp}\widehat{\mathbf{1}_A}(u)^2 = |U^\perp|\mathop{\mathbb{E}}_{x,y\in\mathbb{F}_2^n}[\,\mathbf{1}_A(x)\mathbf{1}_A(y)\mathbf{1}_{\{x+y\in U\}}\,].$$

Decomposing $x = a+z$ and $y = a+w$ with $a\in V$ and $z,w\in U$ (a unique parametrization),

$$\sum_{u\in U^\perp}\widehat{\mathbf{1}_A}(u)^2 = \frac{|U^\perp|}{2^{2n}}\sum_{a\in V}\sum_{z\in U}\sum_{w\in U}\mathbf{1}_A(a+z)\,\mathbf{1}_A(a+w).$$

Since $2^n = |U|\,|V|$ and $|U^\perp| = |V|$, the prefactor is $\frac{1}{|V|\,|U|^2}$, giving

$$\sum_{u\in U^\perp}\widehat{\mathbf{1}_A}(u)^2 = \mathop{\mathbb{E}}_{a\in V}\left[\Big(\mathop{\mathbb{E}}_{z\in U}[\,\mathbf{1}_A(a+z)\,]\Big)^2\right].$$

□

Next, whenever the Fourier spectrum of $\mathbf{1}_A$ is concentrated on $U^\perp$, the function behaves almost like the constant 1 on many cosets of U.

Claim 5.6. *Suppose*

$$\sum_{u\in U^\perp}\widehat{\mathbf{1}_A}(u)^2 \;\geq\; \rho(1-\tau\delta),$$

for some $\tau,\delta\in(0,1)$. Then at least a $(1-\delta)\rho$ fraction of $a\in V$ satisfy $\mathbb{E}_{z\in U}[\,\mathbf{1}_A(a+z)\,]\geq 1-\tau$.

Proof. For $a\in V$, set $\rho_a := \mathbb{E}_{z\in U}[\mathbf{1}_A(a+z)]$, and let $B := \{a\in V : \rho_a < 1-\tau\}$, $G := V\setminus B$, $q := |B|/|V|$. By Claim 5.5,

$$\rho(1-\tau\delta)\leq \mathop{\mathbb{E}}_{a\in V}[\rho_a^2] = q\mathop{\mathbb{E}}_{a\in B}[\rho_a^2] + (1-q)\mathop{\mathbb{E}}_{a\in G}[\rho_a^2] \leq q(1-\tau)\mathop{\mathbb{E}}_{a\in B}[\rho_a] + (1-q)\mathop{\mathbb{E}}_{a\in G}[\rho_a].$$

The global density satisfies $\rho = q\,\mathbb{E}_{a\in B}[\rho_a] + (1-q)\,\mathbb{E}_{a\in G}[\rho_a]$. Subtracting,

$$q\tau \mathop{\mathbb{E}}_{a\in B}[\rho_a] \le \rho\tau\delta \quad \Longrightarrow \quad q \mathop{\mathbb{E}}_{a\in B}[\rho_a] \le \rho\delta.$$

Combining with $q\,\mathbb{E}_{a\in B}[\rho_a] + (1-q) > \rho$ gives $\rho\delta + (1-q) \ge \rho$, i.e. $1-q \ge \rho(1-\delta)$. Hence $|G|/|V| = 1-q \ge \rho(1-\delta)$. □

The next claim asserts that there exists a set S of Fourier coefficients of bounded size such that the Fourier weight outside S is small.

Claim 5.7. *Suppose the spectral norm of $\mathbf{1}_A$ satisfies $\|\widehat{\mathbf{1}_A}\|_1 = m$. Then, for any $\tau, \delta \in (0,1)$, there exists a set $S \subseteq \mathbb{F}_2^n$ such that*

$$\sum_{u\notin S} \widehat{\mathbf{1}_A}(u)^2 \;\le\; \rho\tau\delta \qquad \textit{and} \qquad |S| \;\le\; \frac{m^2}{\rho\tau^2\delta^2}.$$

Proof. By Parseval, $\sum_u \widehat{\mathbf{1}_A}(u)^2 = \mathbb{E}[\mathbf{1}_A^2] = \rho$. Define $S := \{u : |\widehat{\mathbf{1}_A}(u)| \ge \theta\}$. Since $\sum_u |\widehat{\mathbf{1}_A}(u)| \le m$,

$$\sum_{u\notin S} \widehat{\mathbf{1}_A}(u)^2 \le \theta \sum_{u\notin S} |\widehat{\mathbf{1}_A}(u)| \le \theta m.$$

Choosing $\theta = \frac{\rho\tau\delta}{m}$ ensures $\sum_{u\notin S} \widehat{\mathbf{1}_A}(u)^2 \le \rho\tau\delta$. Finally,

$$|S| \le \frac{\sum_{u\in S} \widehat{\mathbf{1}_A}(u)^2}{\theta^2} \le \frac{\rho}{\theta^2} = \frac{m^2}{\rho\tau^2\delta^2}.$$

□

The next claim gives a closed-form expression for the density of $\mathbf{1}_A$ on cosets $a+U$ using only the Fourier coefficients of $\mathbf{1}_A$ lying in $U^\perp$.

Claim 5.8. *For any $a \in V$,*

$$\mathop{\mathbb{E}}_{z\in U}[\mathbf{1}_A(a+z)] \;=\; \sum_{u\in U^\perp} \widehat{\mathbf{1}_A}(u)\,\chi_u(a).$$

Proof. By Fourier inversion, $\mathbf{1}_A(x) = \sum_u \widehat{\mathbf{1}_A}(u)\chi_u(x)$. Averaging over $a+U$,

$$\mathop{\mathbb{E}}_{z\in U}[\mathbf{1}_A(a+z)] = \sum_u \widehat{\mathbf{1}_A}(u)\,\chi_u(a) \mathop{\mathbb{E}}_{z\in U}[\chi_u(z)].$$

By orthogonality, $\mathbb{E}_{z\in U}[\chi_u(z)] = 1$ if $u \in U^\perp$ and 0 otherwise, leaving only the terms with $u \in U^\perp$. □

5.1.2 Proof of Theorem 5.1

We now prove the Dense Coset Decomposition Theorem, restated for convenience.

Theorem 5.1 (Dense coset decomposition theorem). *Let $A \subseteq \mathbb{F}_2^n$ be a set with indicator function $\mathbf{1}_A : \mathbb{F}_2^n \to \{0,1\}$, volume $\mathrm{Vol}(A) = \rho$, and spectral norm $\|\widehat{\mathbf{1}}_A\|_1 \le m$. Then, for any $\omega > 0$, there exist a subset $A' \subseteq A$ with $\frac{|A \setminus A'|}{|A|} \le 2\omega$ and a subspace H of $\mathbb{F}_2^n$ of codimension at most*

$$O\left(\frac{m^2}{\rho\omega^4}\right),$$

such that A' is the union of the restrictions of A to a $\rho(1-\omega)$ fraction of the cosets of H, on each of which the relative density of A is at least $1-\omega$.

Proof. Applying Claim 5.7 with $\tau = \omega$ and $\delta = \omega$, we obtain a set $S \subseteq \mathbb{F}_2^n$ with

$$|S| \le \frac{m^2}{\rho\omega^4} \quad \text{and} \quad \sum_{u \notin S} \widehat{\mathbf{1}_A}(u)^2 \le \rho\omega^2.$$

Let $H, W \le \mathbb{F}_2^n$ be subspaces with $H \oplus W = \mathbb{F}_2^n$ and $S \subseteq H^\perp$, so that all heavy Fourier coefficients lie in $H^\perp$ and

$$\sum_{u \in H^\perp} \widehat{\mathbf{1}_A}(u)^2 \;\ge\; \rho\big(1-\omega^2\big).$$

By Claim 5.6 (with $\tau = \delta = \omega$), for at least a $\rho(1-\omega)$ fraction of cosets $a+H$, the relative density $\rho_a := |A \cap (a+H)|/|H|$ satisfies $\rho_a \ge 1-\omega$. Define

$$A' := \bigcup_{a \in W \,:\, \rho_a \ge 1-\omega} \big(A \cap (a+H)\big).$$

Since a $\rho(1-\omega)$ fraction of cosets have density at least $1-\omega$,

$$\frac{|A'|}{2^n} \;\ge\; \rho \cdot (1-\omega) \cdot (1-\omega) \;\ge\; \rho(1-2\omega),$$

so $\frac{|A \setminus A'|}{2^n} = \rho - \frac{|A'|}{2^n} \le 2\rho\omega$, i.e. $\frac{|A \setminus A'|}{|A|} \le 2\omega$. Finally, $\mathrm{codim}(H) \le |S| \le O(m^2/(\rho\omega^4))$. □

Remark 5.9. Although Theorem 5.1 is stated existentially, it is straightforward to make constructive. Set $\theta = \frac{\rho\varepsilon^2}{10m}$ and use the list-correction toolkit of Chapter 1 (in the classical setting, the Goldreich–Levin algorithm [Goldreich and Levin, 1989]) to discover all Fourier coefficients of $\mathbf{1}_A$ of magnitude at least θ. Let $U^\perp$ be their span. Claim 5.7 then holds for $U^\perp$, and the cosets of U satisfy the decomposition properties required in Theorem 5.1.

5.2 Approximating Sumset Size

Given $S \subseteq \mathbb{F}_2^n$, its volume is $\mathrm{Vol}(S) := |S|/2^n$. The size of the sumset $A + A := \{x + y : x, y \in A\}$ is a fundamental object in additive combinatorics. We study the problem of estimating $\mathrm{Vol}(A + A)$ given only oracle access to the indicator function of A.

Although natural, this problem is intractable in its exact form. First, certifying that a point $z \notin A + A$ requires ruling out all 2^n pairs (x, y) with $x + y = z$, making non-membership verification prohibitively query-expensive. Second, the sumset operation is extremely sensitive to noise: adding a random subset of size about $2^{n/2}$ typically causes the sumset to cover almost all of $\mathbb{F}_2^n$. Thus even small perturbations of A can drastically change $|A + A|$, ruling out subexponential-query algorithms in the worst case. To overcome this, De et al. [2022] proposed a relaxed formulation inspired by the structure-versus-randomness paradigm: discard a small fraction of A and estimate the sumset size of a structured subset $A' \subseteq A$.

> Given oracle access to a set $A \subseteq \mathbb{F}_2^n$ and $\varepsilon > 0$, compute $\mathrm{Vol}(A' + A')$ to additive error $\pm\varepsilon$ for some subset $A' \subseteq A$ satisfying $\mathrm{Vol}(A \setminus A') \leq \varepsilon$.

Using Green's arithmetic regularity lemma [Green, 2005], De et al. [2022] showed that such an A' always exists and that $\mathrm{Vol}(A' + A')$ can be estimated with $\mathrm{tower}(\mathrm{poly}(1/\varepsilon))$ oracle queries. For sets with bounded spectral norm we obtain a substantially improved guarantee. The proof of Theorem 5.4 relies on Theorem 5.1; while that result is existential, we obtain a constructive, query-efficient version using the list-correction framework for the Hadamard code from Chapter 1.

5.2.1 Structure of $A' + A'$

By Theorem 5.1, A decomposes into cosets of some subspace H, each containing almost all of its elements from A; there is a large subset A' of A made up of these dense cosets. With this structure, $A' + A'$ is easy to approximate.

Lemma 5.10. *Suppose A' has density at least $(1 - \omega)$ in each of its dense cosets $a_i + H$. Let $T = \{\, a_i : |A' \cap (a_i + H)| \geq (1 - \omega)|H| \,\}$. Then*

$$A' + A' \;\approx \bigcup_{k \in T+T} (a_k + H),$$

in the sense that for every $k \in T + T$, the intersection $(A' + A') \cap (a_k + H)$ has relative density at least $1 - \omega$.

Proof. Consider two dense cosets a_i+H and a_j+H. Since H is a subspace, $(a_i+H)+(a_j+H) = a_i+a_j+H$. By the density assumption, $|A' \cap (a_i+H)|, |A' \cap (a_j+H)| \geq (1-\omega)|H|$, so their sumset covers all but at most an ω-fraction of $a_i + a_j + H$. Hence each coset $a_k + H$ with $k \in T + T$ is almost entirely contained in $A' + A'$, up to relative error ω. □

Algorithm 11: Sumset-Estimator

Input: Oracle access to $\mathbf{1}_A$, spectral norm bound m, and accuracy parameter ε.
Output: Estimate of $\frac{|A'+A'|}{2^n}$ within $\pm\varepsilon$.

Step 1. Implicitly identify heavy Fourier coefficients. Use Fourier-Sieve (Algorithm 1) to identify all Fourier coefficients α of $\mathbf{1}_A$ with $|\widehat{\mathbf{1}}_A(\alpha)| \geq \theta = \frac{\varepsilon^3}{20m}$ and estimate each to accuracy $\pm\frac{\theta}{10}$ using Estimate-Heavy-Fourier-Coeffs (Algorithm 3). ;

Step 2. Implicitly identify dense cosets of A. Identify a basis $B = \{\alpha_1, \ldots, \alpha_k\}$ for the heavy Fourier coefficients and determine all linear relations among them using Identify-modulo-Invariance (Algorithm 4). For each index $z \in \{\pm 1\}^k$, define the leaf

$$L_z := \left\{x \in \mathbb{F}_2^n \;\middle|\; \chi_{\alpha_i}(x) = z_i \text{ for all } \alpha_i \in B\right\}.$$

For each leaf L_z, estimate its density $\Pr[x \in A \mid x \in L_z]$ using Claim 5.8 and discard any leaf whose density is below $1 - \varepsilon/4$. Let $L := \{z \mid L_z \text{ survived}\}$. ;

Step 3. Approximately simulate $\mathbf{1}_{A'+A'}$. Let $M := \{u * v \mid u, v \in L\}$, where $*$ denotes coordinate-wise multiplication in $\{\pm 1\}^k$. Define $\widetilde{\mathbf{1}}_{A'+A'}(x) := 1$ if $z(x) := (\chi_{\alpha_1}(x), \ldots, \chi_{\alpha_k}(x)) \in M$, and 0 otherwise. ;

Step 4. Estimate $\frac{|A'+A'|}{2^n}$. Sample x uniformly from $\mathbb{F}_2^n$ and use oracle access to $\widetilde{\mathbf{1}}_{A'+A'}$ (via List-Correction, Algorithm 2) to estimate $\frac{|A'+A'|}{2^n}$ within $\pm\varepsilon/2$. ;

5.2.2 Proof of Theorem 5.4

We restate the theorem for convenience.

Theorem 5.4. *Let $A \subseteq \mathbb{F}_2^n$ satisfy $\|\widehat{\mathbf{1}}_A\|_1 \leq m$, and let $\varepsilon > 0$. Then there exists a subset $A' \subseteq A$ with $\mathrm{Vol}(A \setminus A') \leq \varepsilon$ such that $\mathrm{Vol}(A' + A')$ can be estimated to additive accuracy $\pm\varepsilon$ with probability at least $2/3$ using $\widetilde{O}(m^4/\varepsilon^{12})$ queries to $\mathbf{1}_A$.*

Proof. We analyze Sumset-Estimator (Algorithm 11).

1. We may not know ρ beforehand, so we estimate it to accuracy $\varepsilon/4$. If the estimate is below $3\varepsilon/4$, we return 0; otherwise we set $\rho = \varepsilon/2$. This fixes θ and suffices to

instantiate the constructive version of Theorem 5.1 (Remark 5.9).

2. Fourier-Sieve (Algorithm 1) returns a collection of cosets $\mathcal{C}$ such that every α with $|\widehat{\mathbf{1}_A}(\alpha)| \geq \theta$ lies in some $C \in \mathcal{C}$, and every $C \in \mathcal{C}$ has a dominant coefficient α_C with $|\widehat{\mathbf{1}_A}(\alpha_C)| \geq \theta/2$. By Lemma 1.8, this step uses $\widetilde{O}(1/\theta^4)$ queries and fails with probability at most $1/8$.

3. Estimate-Heavy-Fourier-Coeffs (Algorithm 3) estimates each α_C to accuracy $\pm\theta/10$. By Lemma 1.11, this uses $\widetilde{O}(1/\theta^4)$ queries and fails with probability at most $1/20$.

4. Identify-modulo-Invariance (Algorithm 4) recovers a basis $\{\alpha_1, \ldots, \alpha_k\}$ for the span $U^\perp$ of $\{\alpha_C\}$ and all linear relations. By Lemma 1.14, this uses $\widetilde{O}(1/\theta^4)$ queries and fails with probability at most $1/20$.

5. For the basis $\{\alpha_1, \ldots, \alpha_k\}$, define $z(x) = (\chi_{\alpha_1}(x), \ldots, \chi_{\alpha_k}(x)) \in \{\pm 1\}^k$; each coset $a + U$ corresponds to a unique z, the leaves of a non-adaptive parity decision tree. For each z we estimate the density of A within the coset via Claim 5.8 and discard the coset if the empirical density is below $1 - \varepsilon/4$; this cutoff ensures that (accounting for coefficient-estimation error) every coset of true density at least $1 - \varepsilon/2$ is retained.

6. Let $L \subseteq \{\pm 1\}^k$ index the surviving cosets, and $M = \{z * z' : z, z' \in L\}$. By Lemma 5.10, treating $A' + A'$ as the union of cosets indexed by M incurs approximation error at most $\varepsilon/2$.

7. Given x, compute $z(x)$ by evaluating the characters χ_{α_i} via List-Correction (Algorithm 2), and test membership in $A' + A'$ by checking $z(x) \in M$. This gives oracle access to $\widetilde{\mathbf{1}}_{A'+A'}$. We use $O(1/\varepsilon^2)$ samples to estimate $\mathrm{Vol}(A' + A')$, each requiring $\widetilde{O}(1/\theta^2)$ calls to $\mathbf{1}_A$.

Amplifying Step 2. Recall that Fourier-Sieve fails with probability at most $1/8$ (Lemma 1.8), rather than the $1/20$ the union bound below would otherwise require. We therefore run Step 2 a constant number of times and retain a successful run (verification is downstream and query-cheap); this drives its failure probability below $1/20$ at an $O(1)$ query cost, absorbed in $\widetilde{O}(\cdot)$.

Failure probability and query complexity. By a union bound over the (amplified) steps, the algorithm fails with probability at most $1/3$. The query complexity is

$$\widetilde{O}\Big(\tfrac{1}{\theta^4} + \tfrac{1}{\varepsilon^2\,\theta^2}\Big) = \widetilde{O}\Big(\tfrac{m^4}{\varepsilon^{12}} + \tfrac{m^2}{\varepsilon^8}\Big) = \widetilde{O}\Big(\tfrac{m^4}{\varepsilon^{12}}\Big),$$

using $\theta = \varepsilon^3/(20m)$. This matches Theorem 5.4. We note that the running *time* is larger, $\widetilde{O}(1/\theta^8)$, because Step 2 enumerates the $O(1/\theta^8)$ cosets maintained by Fourier-Sieve; since the complexity measure is query complexity, this does not affect the stated bound. □

Part II

Fourier Sketching by Subspace Restriction

6

The Subspace Restriction Technique

In classical measurement and acquisition, a high-dimensional object is often probed not by recording every one of its coordinates, but by taking a comparatively small number of carefully designed observations. A fundamental question in this process is how much of the object's structure survives such partial acquisition. Naïvely, restricting attention to a small portion of a signal appears to discard almost all of the information it carries. Remarkably, when the underlying object possesses hidden structure — sparsity, low rank, or spectral concentration — a small set of well-chosen observations can retain essentially all of the relevant information.

This principle lies at the heart of modern sensing and compressed acquisition. Rather than measuring every coordinate of a high-dimensional vector, one acquires a handful of linear measurements engineered to preserve the information pertinent to the underlying structure. Compressed sensing, random projections, and sketching algorithms all exploit this philosophy: local or low-dimensional observations frequently suffice, provided the object is structured. The measurements themselves are typically "lossy" in the sense that they collapse many degrees of freedom together, yet the collapse is engineered so that structured objects remain recoverable from the compressed record.

A closely related phenomenon appears in the frequency domain. When a signal is observed only on a restricted region of its domain, its spectrum does not simply vanish outside the observed window; instead, frequency components combine and reappear in a folded, aggregated form. In the language of the previous chapter, this is precisely the mechanism of aliasing: distinct frequencies become indistinguishable after restricted observation. What is destructive in one setting becomes constructive in another. If one knows how the folding operates, the aggregated measurements form a compressed sketch of the original spectrum, from which structured spectral information can be read back out.

Subspace restriction realizes exactly this idea for Boolean functions. Suppose we are given a function $f : \mathbb{F}_2^n \to \mathbb{R}$ and an affine subspace $r + H$, where $H \leq \mathbb{F}_2^n$ has dimension t. Instead of observing f on the entire hypercube, we observe its values only on the 2^t points of $r + H$. At first glance this appears to discard the overwhelming majority

of the information contained in f. Yet, through Fourier duality, the restricted function retains a highly structured summary of the global Fourier spectrum. Restricting f to an affine subspace does not merely reveal local behavior: it simultaneously computes structured linear combinations of the global Fourier coefficients. More precisely, every Fourier coefficient of the restricted function equals a signed sum of the Fourier coefficients lying in a single coset of $H^\perp$. Thus, although the restriction is performed entirely in the input domain, it implicitly produces a compressed sketch of the Fourier spectrum.

Viewed this way, subspace restriction is the exact dual counterpart of the coset hashing technique of the previous chapter. In coset hashing, Fourier coefficients belonging to a common bucket are *explicitly* aggregated into a single measurement in the frequency domain. In subspace restriction, exactly the same aggregation occurs *implicitly*: frequencies lying in the same coset of $H^\perp$ become indistinguishable once f is observed only on $r + H$. Where coset hashing compresses the spectrum by design, subspace restriction compresses it as an automatic by-product of local sampling. The two operations are, from complementary viewpoints, computing the same structured spectral aggregates. This duality is the reason the two primitives can be used interchangeably in the algorithms developed later in this thesis.

Random restrictions have played a central role in theoretical computer science for more than three decades. They have been used extensively in circuit complexity, coding theory, learning theory, higher-order Fourier analysis, pseudorandomness, and property testing, where they preserve algebraic structure while dramatically reducing dimension. Traditionally, however, restrictions have been regarded primarily as tools for simplifying functions or preserving low-degree polynomial structure — an instrument for reducing the complexity of the object under study, rather than for measuring it.

The viewpoint developed in this thesis is substantially different. We interpret random affine restrictions as a *sketching primitive* for Fourier analysis. Rather than studying only the restricted function itself, we investigate the spectral information encoded by its Fourier transform and show that this information suffices to estimate several global quantities of interest. In particular, we establish concentration phenomena showing that approximately Fourier-sparse functions remain spectrally stable under random restrictions, enabling efficient estimation of global spectral parameters from purely local observations. These results form the foundation of the algorithms developed in the subsequent chapters. We now make these notions precise.

A *random subspace* $H \leq \mathbb{F}_2^n$ of dimension t is the span of t independent uniformly random

vectors $v_1, \dots, v_t \in \mathbb{F}_2^n$. A *random affine subspace* of dimension t is

$$A = \alpha + H,$$

with H as above and $\alpha \in \mathbb{F}_2^n$ uniform and independent of H. The *restriction* of f to A is

$$f_A : H \to \mathbb{R}, \qquad f_A(x) := f(x + \alpha).$$

Since H is itself an $\mathbb{F}_2$-vector space, f_A has its own Fourier expansion. To index it, fix a subspace W complementary to $H^\perp$, so that $\mathbb{F}_2^n = W \oplus H^\perp$; then $\{\chi_\gamma|_H : \gamma \in W\}$ is an orthonormal basis for the real-valued functions on H. We write $\widehat{f_A}(\gamma)$, $\gamma \in W$, for the Fourier coefficients of f_A, and note $|W| = 2^t$. Parseval on the group H reads

$$\sum_{\gamma \in W} \widehat{f_A}(\gamma)^2 = \|f_A\|_2^2 = \mathbb{E}_{x \in H}\big[f(x + \alpha)^2\big].$$

For convenient reference, the notation used throughout this chapter is collected in Table 6.1.

Notation	Description
$H \leq \mathbb{F}_2^n$	A random subspace of dimension t
$A = \alpha + H$	A random affine subspace (α uniform, independent of H)
$f_A(x) = f(x + \alpha)$	The restriction of f to A, viewed as a function on H
W	A fixed complement to $H^\perp$, indexing the spectrum of f_A
$\widehat{f_A}(\gamma)$	Fourier coefficient of f_A at $\gamma \in W$
$a := 2^{-t}$	Coset collision parameter
$M := \|\widehat{f}\|_4^4$	$\sum_\beta \widehat{f}(\beta)^4$, the fourth Fourier mass of f
$X_A := \|\widehat{f_A}\|_4^4$	$\sum_{\gamma \in W} \widehat{f_A}(\gamma)^4$, the fourth Fourier mass of f_A
$K := \|f\|_8^8$	Eighth moment of f; controls the light spectrum

Table 6.1: Notation for the subspace restriction technique.

6.1 The Fourier spectrum under random affine restriction

This section records how the spectrum transforms under restriction. The three statements below are the real-valued analogues of the classical Boolean facts, and — this is the point — their proofs carry over *verbatim*, because they rely only on Parseval's identity, normalized

here to $\sum_\beta \widehat{f}(\beta)^2 = 1$, and character orthogonality. No higher moment of f appears anywhere in this section.

The first fact makes precise the sense in which restriction implements structured aliasing in the frequency domain: it aggregates the spectrum along cosets of $H^\perp$, exactly mirroring the bucketing of the coset hashing primitive.

Fact 6.1 (Restricted Fourier coefficients)**.** *Let $H \leq \mathbb{F}_2^n$ be a subspace, $\alpha \in \mathbb{F}_2^n$, $A = \alpha + H$, and let W be complementary to $H^\perp$. Then for every $\gamma \in W$,*

$$\widehat{f_A}(\gamma) = \sum_{\beta \in \gamma + H^\perp} \widehat{f}(\beta)\, \chi_\beta(\alpha).$$

Proof. By definition $\widehat{f_A}(\gamma) = \mathbb{E}_{x \in H}\big[f(x+\alpha)\chi_\gamma(x)\big]$. Substituting the Fourier expansion of f and using $\chi_\beta(x+\alpha) = \chi_\beta(x)\chi_\beta(\alpha)$,

$$\widehat{f_A}(\gamma) = \mathbb{E}_{x \in H}\Big[\sum_\beta \widehat{f}(\beta)\chi_\beta(x)\chi_\beta(\alpha)\chi_\gamma(x)\Big] = \sum_\beta \widehat{f}(\beta)\chi_\beta(\alpha)\, \mathbb{E}_{x \in H}\big[\chi_{\beta-\gamma}(x)\big].$$

By character orthogonality the inner expectation is 1 if $\beta - \gamma \in H^\perp$ and 0 otherwise, so only $\beta \in \gamma + H^\perp$ survive. □

Thus restriction *aggregates* the spectrum of f along cosets of $H^\perp$, with signs $\chi_\beta(\alpha)$ supplied by the random shift. Two consequences will matter for the algorithms to come: a heavy coefficient alone in its coset is read off cleanly, whereas many light coefficients sharing a coset could, in principle, conspire. We next show that the induced coset partition is pairwise independent, in exact analogy with the pairwise independence of the coset hashing partition established in the previous chapter.

Fact 6.2 (Collision probability)**.** *Let $H \leq \mathbb{F}_2^n$ be a random subspace of dimension t. Then for any distinct $\beta_1, \beta_2 \in \mathbb{F}_2^n$,*

$$\Pr_H\big[\ \beta_1, \beta_2 \text{ lie in the same coset of } H^\perp\ \big] = 2^{-t}.$$

Proof. Put $w := \beta_1 - \beta_2 \neq 0$. The two frequencies share a coset of $H^\perp$ precisely when $w \in H^\perp$, i.e. when $\langle w, v_i \rangle = 0$ for each of the t independent uniform vectors $v_1, \ldots, v_t$ spanning H. For fixed $w \neq 0$ each such condition holds with probability $1/2$, independently across i; the probability is therefore 2^{-t}. □

Observe that Fact 6.2 concerns subspaces only, and is insensitive to whether f is Boolean or real-valued.

We can now establish the first of the two central results of this chapter: heavy Fourier

coefficients survive restriction essentially unperturbed.

Lemma 6.3 (Heavy coefficients survive restriction). *Let $f : \mathbb{F}_2^n \to \mathbb{R}$ with $\|f\|_2 = 1$, and let $A = \alpha + H$ be a random affine subspace of dimension t. Let W be complementary to $H^\perp$, and for $\gamma \in \mathbb{F}_2^n$ let $\beta_\gamma \in W$ be the unique representative with $\gamma \in \beta_\gamma + H^\perp$. Then for every $\gamma \in \mathbb{F}_2^n$ and every $\lambda > 0$,*

$$\Pr_{H,\alpha}\Big[\, \big|\, \widehat{f_A}(\beta_\gamma) - \widehat{f}(\gamma)\chi_\gamma(\alpha)\,\big| \;\geq\; 2^{-t} + \lambda \,\Big] \;\leq\; \frac{2^{-t}}{\lambda^2}.$$

Proof. Fix $\gamma \neq 0$ (the case $\gamma = 0$ is handled at the end). By Fact 6.1,

$$\widehat{f_A}(\beta_\gamma) = \sum_{\beta \in \beta_\gamma + H^\perp} \widehat{f}(\beta)\chi_\beta(\alpha).$$

For $\beta \in \mathbb{F}_2^n$ let $\mathbf{1}_\beta$ be the indicator of $\beta \in \beta_\gamma + H^\perp$, a function of H alone. Isolating the terms $\beta = \gamma$ and $\beta = 0$, and using $\chi_0(\alpha) = 1$, define

$$Z \;:=\; \widehat{f_A}(\beta_\gamma) - \widehat{f}(\gamma)\chi_\gamma(\alpha) \;=\; \widehat{f}(0)\,\mathbf{1}_0 \;+ \sum_{\beta \notin \{0,\gamma\}} \widehat{f}(\beta)\,\chi_\beta(\alpha)\,\mathbf{1}_\beta.$$

Note $\mathbf{1}_\beta$ depends only on H and $\chi_\beta(\alpha)$ only on α, and H, α are independent.

Expectation. Condition on H. As α is uniform, $\mathbb{E}_\alpha[\chi_\beta(\alpha)] = 0$ for all $\beta \neq 0$, so $\mathbb{E}_\alpha[Z \mid H] = \widehat{f}(0)\mathbf{1}_0$. Hence, by Fact 6.2 applied to the distinct pair $\{0, \gamma\}$,

$$\mathbb{E}_{H,\alpha}[Z] = \widehat{f}(0) \cdot \Pr_H\big[\gamma \in H^\perp\big] = 2^{-t}\widehat{f}(0), \qquad \text{so} \qquad \big|\mathbb{E}_{H,\alpha}[Z]\big| \leq 2^{-t},$$

using $|\widehat{f}(0)| \leq \|f\|_2 = 1$.

Variance. Condition on H and write $a_\beta := \widehat{f}(\beta)\mathbf{1}_\beta$, so $Z = \sum_{\beta \neq \gamma} a_\beta \chi_\beta(\alpha)$. Expanding,

$$\mathbb{E}_\alpha\big[Z^2 \mid H\big] = \sum_{\beta \neq \gamma} a_\beta^2 + \sum_{\substack{\beta, \beta' \neq \gamma \\ \beta \neq \beta'}} a_\beta a_{\beta'}\, \mathbb{E}_\alpha\big[\chi_{\beta+\beta'}(\alpha)\big].$$

For $\beta \neq \beta'$ we have $\beta + \beta' \neq 0$, so all cross terms vanish. Thus $\mathbb{E}_\alpha[Z^2 \mid H] = \sum_{\beta \neq \gamma} \widehat{f}(\beta)^2 \mathbf{1}_\beta$, and averaging over H using Fact 6.2,

$$\mathbb{E}_{H,\alpha}\big[Z^2\big] = 2^{-t} \sum_{\beta \neq \gamma} \widehat{f}(\beta)^2 \;\leq\; 2^{-t}\|f\|_2^2 \;=\; 2^{-t},$$

by Parseval. Hence $\mathrm{Var}(Z) \leq \mathbb{E}[Z^2] \leq 2^{-t}$.

Concentration. Chebyshev's inequality gives $\Pr\big[|Z - \mathbb{E}[Z]| \geq \lambda\big] \leq 2^{-t}/\lambda^2$. Since

$|\mathbb{E}[Z]| \le 2^{-t}$, the event $|Z| \ge 2^{-t} + \lambda$ implies $|Z - \mathbb{E}[Z]| \ge \lambda$, proving the claim.

If $\gamma = 0$ the term $\widehat{f}(0)\mathbf{1}_0$ is absent, $\mathbb{E}[Z] = 0$, and one obtains the stronger bound $\Pr[|Z| \ge \lambda] \le 2^{-t}/\lambda^2$, which implies the stated inequality. □

Remark 6.4. The proof used exactly two properties of f: Parseval in the form $\sum_\beta \widehat{f}(\beta)^2 = 1$, and $|\widehat{f}(0)| \le \|f\|_2 = 1$. Both hold for arbitrary real-valued f of unit L^2 norm. *No moment beyond* $\|f\|_2$ *appears in Lemma 6.3* — the heavy half of our analysis is moment-free, and everything to come that involves $\|f\|_8$ concerns only the light spectrum.

6.2 The fourth Fourier mass under random affine restriction

Lemma 6.3 protects the heavy coefficients. It says nothing, however, about the possibility that a large number of *small* coefficients aggregate along a single coset of $H^\perp$ into a spuriously large coefficient of f_A. Ruling this out is the objective of the present section, and the instrument for doing so is the fourth Fourier mass. Recall the notation of Table 6.1:

$$a := 2^{-t}, \qquad M := \|\widehat{f}\|_4^4 = \sum_{\beta \in \mathbb{F}_2^n} \widehat{f}(\beta)^4, \qquad X_A := \big\|\widehat{f_A}\big\|_4^4 = \sum_{\gamma \in W} \widehat{f_A}(\gamma)^4, \qquad K := \|f\|_8^8.$$

The main result of this section, Lemma 6.12, shows that X_A concentrates sharply around M; equivalently, the fourth Fourier mass is approximately preserved under random restriction. Its proof proceeds through a physical-domain identity, a first-moment computation, and a variance bound.

6.2.1 A four-point identity

The engine of the entire section is an identity that transports the fourth Fourier mass out of the frequency domain and into the physical domain, where it becomes an average of f over four points.

Fact 6.5 (Four-point identity). *Let G be a finite abelian group of exponent two, i.e. $G \cong \mathbb{F}_2^k$ for some k, with characters $\{\chi_\gamma\}_{\gamma \in \widehat{G}}$. Then for every $g : G \to \mathbb{R}$,*

$$\sum_{\gamma \in \widehat{G}} \widehat{g}(\gamma)^4 \;=\; \mathbb{E}_{x,u,w \in G}\Big[\, g(x)\, g(x+u)\, g(x+w)\, g(x+u+w) \,\Big].$$

Proof. Call the right-hand side R and substitute $g(y) = \sum_\gamma \widehat{g}(\gamma)\chi_\gamma(y)$ into each factor:

$$R = \sum_{\gamma_1,\gamma_2,\gamma_3,\gamma_4} \widehat{g}(\gamma_1)\widehat{g}(\gamma_2)\widehat{g}(\gamma_3)\widehat{g}(\gamma_4)\, \mathbb{E}_{x,u,w}\Big[\chi_{\gamma_1}(x)\chi_{\gamma_2}(x+u)\chi_{\gamma_3}(x+w)\chi_{\gamma_4}(x+u+w)\Big].$$

By the character multiplication rule,

$$\chi_{\gamma_1}(x)\chi_{\gamma_2}(x+u)\chi_{\gamma_3}(x+w)\chi_{\gamma_4}(x+u+w) = \chi_{\gamma_1+\gamma_2+\gamma_3+\gamma_4}(x)\ \chi_{\gamma_2+\gamma_4}(u)\ \chi_{\gamma_3+\gamma_4}(w).$$

Since x, u, w are independent and uniform, character orthogonality (with $H = G$, so $H^\perp = \{0\}$) forces the expectation to vanish unless

$$\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = 0, \qquad \gamma_2 + \gamma_4 = 0, \qquad \gamma_3 + \gamma_4 = 0,$$

in which case it equals 1. As G has exponent two, the last two conditions say $\gamma_2 = \gamma_4$ and $\gamma_3 = \gamma_4$, and the first then forces $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4$. Only diagonal terms survive, giving $R = \sum_\gamma \widehat{g}(\gamma)^4$. □

Applying Fact 6.5 to $G = \mathbb{F}_2^n$ with $g = f$, and to $G = H$ with $g = f_A$ (legitimate, since H is an $\mathbb{F}_2$-vector space), we obtain the two representations

$$\begin{aligned} M &= \mathbb{E}_{y,u,w\in\mathbb{F}_2^n}\big[f(y)f(y+u)f(y+w)f(y+u+w)\big], \\ X_A &= \mathbb{E}_{x,u,w\in H}\big[f_A(x)f_A(x+u)f_A(x+w)f_A(x+u+w)\big]. \end{aligned} \tag{6.1}$$

The identity moves the fourth mass out of the Fourier domain and into the physical domain, where it becomes an average of f over four points. That is what makes the variance argument below possible.

6.2.2 Expectation of the restricted fourth mass

We first show that X_A is an almost unbiased estimator of M, with a bias controlled by the eighth moment K.

Lemma 6.6 (Expectation of X_A)**.** *With $a = 2^{-t}$,*

$$\mathbb{E}_A\big[X_A\big] \;=\; \big(1 - 3a + 2a^2\big)M \;+\; 3a \;+\; a^2\big(\|f\|_4^4 - 3\big).$$

Consequently

$$0 \;\le\; \mathbb{E}_A\big[X_A\big] - M \;\le\; 3a + a^2\big(\|f\|_4^4 - 1\big) \;\le\; 4aK.$$

Proof. By Fact 6.1,

$$X_A = \sum_{\gamma\in W}\Big(\sum_{\beta\in\gamma+H^\perp}\widehat{f}(\beta)\chi_\beta(\alpha)\Big)^4 = \sum_{\beta_1,\dots,\beta_4\in\mathbb{F}_2^n} c(\beta_1,\dots,\beta_4)\,\mathbf{1}_H(\beta_1,\dots,\beta_4)\,\chi_{\beta_1+\beta_2+\beta_3+\beta_4}(\alpha),$$

where $c(\beta_1,\dots,\beta_4) := \prod_{i=1}^4 \widehat{f}(\beta_i)$ and $\mathbf{1}_H(\beta_1,\dots,\beta_4)$ indicates that $\beta_1,\dots,\beta_4$ lie in a common coset of $H^\perp$. Averaging over α and using $\mathbb{E}_\alpha[\chi_v(\alpha)] = \mathbf{1}[v=0]$,

$$\mathbb{E}_\alpha\big[X_A \mid H\big] = \sum_{\substack{\beta_1,\dots,\beta_4\\ \beta_1+\beta_2+\beta_3+\beta_4=0}} c(\beta_1,\dots,\beta_4)\,\mathbf{1}_H(\beta_1,\dots,\beta_4).$$

For a quadruple $T=(\beta_1,\dots,\beta_4)$ put $r(T) := \dim \operatorname{span}\{\beta_2-\beta_1,\ \beta_3-\beta_1,\ \beta_4-\beta_1\}$. On the zero-sum set $\beta_4-\beta_1 = (\beta_2-\beta_1)+(\beta_3-\beta_1)$, so $r(T)\in\{0,1,2\}$. Moreover $\beta_1,\dots,\beta_4$ share a coset of $H^\perp$ if and only if $\beta_i-\beta_1\in H^\perp$ for $i=2,3,4$; since a fixed nonzero vector lies in $H^\perp$ with probability 2^{-t}, and the constraints corresponding to a basis of the span are independent,

$$\Pr_H\big[\mathbf{1}_H(T)=1\big] = 2^{-t\cdot r(T)} = a^{\,r(T)}.$$

Therefore

$$\mathbb{E}_A\big[X_A\big] \;=\; \sum_{r=0}^{2} a^{\,r} \sum_{\substack{T:\ r(T)=r\\ \beta_1+\beta_2+\beta_3+\beta_4=0}} c(T), \tag{6.2}$$

and we compute the three contributions exactly.

Case $r(T)=0$. Then $\beta_1=\beta_2=\beta_3=\beta_4$, contributing $\sum_\beta \widehat{f}(\beta)^4 = M$.

Case $r(T)=1$. With the zero-sum condition, such a quadruple contains exactly two distinct frequencies, each twice. Each unordered pair $\{\beta,\beta'\}$ arises from $\frac{4!}{2!\,2!}=6$ ordered quadruples, so the contribution is

$$6\sum_{\beta<_\ell\beta'}\widehat{f}(\beta)^2\widehat{f}(\beta')^2 = 3\Big[\Big(\sum_\beta \widehat{f}(\beta)^2\Big)^2 - \sum_\beta \widehat{f}(\beta)^4\Big] = 3\big(1-M\big),$$

where $<_\ell$ is any fixed total order, and the last step uses Parseval, $\sum_\beta \widehat{f}(\beta)^2 = 1$.

Case $r(T)=2$. First sum $c(T)$ over *all* zero-sum quadruples. Writing the constraint as $\beta_1+\beta_2=\beta_3+\beta_4$ and grouping by $v := \beta_1+\beta_2$,

$$\sum_{\substack{\beta_1,\dots,\beta_4\\ \beta_1+\beta_2=\beta_3+\beta_4}} \widehat{f}(\beta_1)\widehat{f}(\beta_2)\widehat{f}(\beta_3)\widehat{f}(\beta_4) = \sum_{v\in\mathbb{F}_2^n}\Big(\sum_{\beta_1+\beta_2=v}\widehat{f}(\beta_1)\widehat{f}(\beta_2)\Big)^2.$$

Squaring the Fourier expansion of f and collecting equal frequencies gives

$$f(x)^2 = \sum_{\beta_1,\beta_2} \widehat{f}(\beta_1)\widehat{f}(\beta_2)\chi_{\beta_1+\beta_2}(x), \qquad \text{so} \qquad \widehat{f^2}(v) = \sum_{\beta_1+\beta_2=v} \widehat{f}(\beta_1)\widehat{f}(\beta_2).$$

Applying Parseval to the function f^2,

$$\sum_v \widehat{f^2}(v)^2 = \left\|f^2\right\|_2^2 = \mathbb{E}_x\left[f(x)^4\right] = \|f\|_4^4.$$

Since $r(T) \in \{0,1,2\}$ partitions the zero-sum quadruples, the $r(T) = 2$ contribution equals

$$\|f\|_4^4 - M - 3(1-M) = \|f\|_4^4 - 3 + 2M.$$

Substituting into (6.2),

$$\mathbb{E}_A\left[X_A\right] = M + 3a(1-M) + a^2\left(\|f\|_4^4 - 3 + 2M\right) = \left(1 - 3a + 2a^2\right)M + 3a + a^2\left(\|f\|_4^4 - 3\right).$$

Rearranging,

$$\mathbb{E}_A\left[X_A\right] - M = (3a - 2a^2)(1-M) + a^2\left(\|f\|_4^4 - 1\right),$$

and both terms are nonnegative since $M \le 1$, $\|f\|_4^4 \ge 1$, and $a \le \frac{1}{2}$. Finally $(3a-2a^2)(1-M) \le 3a$ and $a^2(\|f\|_4^4 - 1) \le a\|f\|_4^4 \le aK$, so $\mathbb{E}_A[X_A] - M \le 3a + aK \le 4aK$ as $K \ge 1$. □

Remark 6.7 (Where Booleanity was hiding). For Boolean f one has $f^2 \equiv 1$ and hence $\|f\|_4^4 = 1$, and the identity collapses to the classical $\mathbb{E}_A[X_A] = (1-3a+2a^2)M + 3a - 2a^2$. The case $r(T) = 2$ is thus the *unique* point at which Booleanity entered the classical first-moment computation, and it is exactly there that $\|f\|_4^4$ surfaces in the real-valued setting.

6.2.3 Variance of the restricted fourth mass

Here we must depart from the classical route, and it is instructive to see why.

Why the Fourier-side computation fails. Expanding $\mathrm{Var}_A(X_A)$ in the Fourier basis produces a *signed* sum over pairs of quadruples (T, T'), weighted by $c(T)c(T')$ and by the indicator that each quadruple lies in a coset of $H^\perp$. The classical argument inserts the worst-case collision probability $\Pr_H[\cdot] \le 2^{-t}$ into that sum term by term. This is not legitimate: one may not pull a probability bound out of a signed sum. Making the step

rigorous forces absolute values on the coefficients, which amounts to replacing f by

$$f^+ := \sum_\beta \big|\widehat{f}(\beta)\big|\chi_\beta,$$

and the resulting bound involves $\|f^+\|_8^8$. That is fatal: for a bent function $f^+ = 2^{n/2}\mathbf{1}_{\{0\}}$, so $\|f^+\|_8^8 = 2^{3n}$, and the bound is vacuous *even for Boolean* f. We therefore argue in the physical domain instead, using the four-point representation (6.1).

The new idea. A random affine subspace furnishes *exactly* independent uniform points of $\mathbb{F}_2^n$, except on a rare degenerate event. Precisely: fix $m \geq 1$, draw $c^{(1)}, \dots, c^{(m)} \in \mathbb{F}_2^t$ independently and uniformly, and set

$$w_j := \sum_{i=1}^{t} c_i^{(j)} v_i \ \in \ H, \qquad j = 1, \dots, m,$$

where $v_1, \dots, v_t$ are the independent uniform vectors spanning H. Since $c \mapsto \sum_i c_i v_i$ is a surjective homomorphism $\mathbb{F}_2^t \to H$, the w_j are, conditionally on H, independent and uniform on H; hence the averages over H in (6.1) may be realized by sampling in this manner. Define the degenerate event

$$\mathcal{B}_m := \big\{\ c^{(1)}, \dots, c^{(m)} \text{ are linearly dependent in } \mathbb{F}_2^t \ \big\}.$$

Claim 6.8. (i) $\Pr[\mathcal{B}_m] \leq (2^m - 1)2^{-t}$. (ii) *Conditioned on $\mathcal{B}_m^c$, the vectors $(w_1, \dots, w_m)$ are distributed as m independent uniform elements of $\mathbb{F}_2^n$, independently of α.*

Proof. (i) Since $\operatorname{span}(c^{(1)}, \dots, c^{(j-1)})$ has at most 2^{j-1} elements, a union bound gives

$$\Pr[\mathcal{B}_m] \ \leq \ \sum_{j=1}^{m} \Pr\big[c^{(j)} \in \operatorname{span}(c^{(1)}, \dots, c^{(j-1)})\big] \ \leq \ \sum_{j=1}^{m} 2^{j-1}2^{-t} \ = \ (2^m - 1)2^{-t},$$

the term $j = 1$ accounting for $c^{(1)} = 0$.

(ii) Let $V \in \mathbb{F}_2^{t\times n}$ have rows $v_1, \dots, v_t$, and $C \in \mathbb{F}_2^{m\times t}$ have rows $c^{(1)}, \dots, c^{(m)}$; then $(w_1, \dots, w_m)$ is precisely the row set of CV. On $\mathcal{B}_m^c$ the matrix C has full row rank m, so $X \mapsto CX$ is a *surjective* linear map $\mathbb{F}_2^{t\times n} \to \mathbb{F}_2^{m\times n}$. A surjective linear map pushes the uniform distribution forward to the uniform distribution; as V is uniform on $\mathbb{F}_2^{t\times n}$, the matrix CV is uniform on $\mathbb{F}_2^{m\times n}$, which is the assertion. Finally $\mathcal{B}_m$ is determined by the $c^{(j)}$ alone, while α is drawn independently of $(v_i)_i$ and $(c^{(j)})_j$. □

Claim 6.9. *Let $p_1, \dots, p_r$ be points of the form $p_i = \alpha + u_i$, where each u_i is a function of*

(v, c) *only, and let* $\mathcal{B} = \mathcal{B}_m$*. Then*

$$\Big|\, \mathbb{E}\Big[\prod_{i=1}^{r} f(p_i)\cdot \mathbf{1}_{\mathcal{B}}\Big] \,\Big| \;\le\; \Pr[\mathcal{B}]\cdot \|f\|_r^r.$$

Proof. By the inequality of arithmetic and geometric means, $\big|\prod_{i=1}^r f(p_i)\big| \le \frac{1}{r}\sum_{i=1}^r |f(p_i)|^r$. Fix i. Conditionally on (v, c), the point $p_i = \alpha + u_i$ is uniform on $\mathbb{F}_2^n$, since α is uniform and independent of (v, c). As $\mathcal{B}$ is determined by c alone,

$$\mathbb{E}\big[|f(p_i)|^r \mathbf{1}_{\mathcal{B}}\big] = \mathbb{E}\Big[\mathbf{1}_{\mathcal{B}}\cdot \mathbb{E}_\alpha\big[|f(\alpha+u_i)|^r\big]\Big] = \Pr[\mathcal{B}]\cdot\|f\|_r^r.$$

Averaging over $i = 1, \dots, r$ yields the claim. □

Note that Claim 6.9 is where the eighth moment will be paid for: on the degenerate event we have no cancellation to exploit, and we simply bound the eight-fold product by $\|f\|_8^8$.

Claim 6.10 (Second moment of X_A)**.** $\big|\, \mathbb{E}_A\big[X_A^2\big] - M^2 \,\big| \;\le\; 126\cdot 2^{-t}K.$

Proof. By (6.1), X_A^2 is the product of two independent copies of the same four-point average over H. Realize the two triples of sample points using independent uniform $c^{(1)}, \dots, c^{(6)} \in \mathbb{F}_2^t$ (so $m = 6$), producing $w_1, \dots, w_6 \in H$ and the eight points

$$p_1, \dots, p_4 = \alpha{+}w_1{+}\{0,\, w_2,\, w_3,\, w_2{+}w_3\}, \qquad p_5, \dots, p_8 = \alpha{+}w_4{+}\{0,\, w_5,\, w_6,\, w_5{+}w_6\},$$

so that $\mathbb{E}_A\big[X_A^2\big] = \mathbb{E}_{v,c,\alpha}\big[\prod_{i=1}^8 f(p_i)\big]$.

On $\mathcal{B}_6^c$, Claim 6.8(ii) gives that $w_1, \dots, w_6$ are independent and uniform on $\mathbb{F}_2^n$ and independent of α. Put $y := \alpha + w_1$ and $y' := \alpha + w_4$: conditionally on α these are independent and uniform, so y, y', w_2, w_3, w_5, w_6 are six independent uniform vectors of $\mathbb{F}_2^n$. The eight-fold product therefore factorizes into two independent four-fold products, and by (6.1),

$$\mathbb{E}\Big[\prod_{i=1}^{8} f(p_i) \;\Big|\; \mathcal{B}_6^c\Big] = \Big(\mathbb{E}_{y,u,w}\big[f(y)f(y{+}u)f(y{+}w)f(y{+}u{+}w)\big]\Big)^2 = M^2.$$

Writing $\mathbb{E}_A[X_A^2] = \Pr[\mathcal{B}_6^c]\, M^2 + \mathbb{E}\big[\prod_{i=1}^8 f(p_i)\,\mathbf{1}_{\mathcal{B}_6}\big]$ and applying Claim 6.9 with $r = 8$, together with $\Pr[\mathcal{B}_6] \le 63\cdot 2^{-t}$ and $M \le 1 \le K$,

$$\big|\mathbb{E}_A[X_A^2] - M^2\big| \;\le\; \Pr[\mathcal{B}_6]\big(M^2 + \|f\|_8^8\big) \;\le\; 63\cdot 2^{-t}(1+K) \;\le\; 126\cdot 2^{-t}K.$$

□

Lemma 6.11 (Variance bound)**.** $\mathrm{Var}_A\big(X_A\big) \;\le\; 126\cdot 2^{-t}\,\|f\|_8^8.$

Proof. By Lemma 6.6, $\mathbb{E}_A[X_A] \geq M \geq 0$, so $\big(\mathbb{E}_A[X_A]\big)^2 \geq M^2$. With Claim 6.10,

$$\mathrm{Var}_A(X_A) = \mathbb{E}_A\big[X_A^2\big] - \big(\mathbb{E}_A[X_A]\big)^2 \;\leq\; \mathbb{E}_A\big[X_A^2\big] - M^2 \;\leq\; 126 \cdot 2^{-t}K.$$

□

6.2.4 Concentration of the restricted fourth mass

Combining the first-moment estimate of Lemma 6.6 with the variance bound of Lemma 6.11 yields the second central result of this chapter.

Lemma 6.12 (The fourth Fourier mass is preserved). *Let $f : \mathbb{F}_2^n \to \mathbb{R}$ with $\|f\|_2 = 1$, and let $A = \alpha + H$ be a random affine subspace of dimension t. Then for every $\varepsilon \in (0, 1]$,*

$$\Pr_A\Big[\,\Big|\,\|\widehat{f_A}\|_4^4 - \|\widehat{f}\|_4^4\,\Big| \;\geq\; \varepsilon\,\Big] \;\leq\; 504 \cdot \frac{2^{-t}\,\|f\|_8^8}{\varepsilon^2}.$$

Proof. Recall $a = 2^{-t}$ and $K \geq 1$. By Lemma 6.6, $\big|\mathbb{E}_A[X_A] - M\big| \leq b := 4aK$. Fix $\varepsilon \in (0, 1]$.

Case 1: $\varepsilon \geq 2b$. If $|X_A - M| \geq \varepsilon$ then $\big|X_A - \mathbb{E}_A[X_A]\big| \geq \varepsilon - b \geq \varepsilon/2$. By Chebyshev and Lemma 6.11,

$$\Pr\big[|X_A - M| \geq \varepsilon\big] \;\leq\; \frac{\mathrm{Var}_A(X_A)}{(\varepsilon/2)^2} \;\leq\; \frac{504 \cdot 2^{-t}K}{\varepsilon^2}.$$

Case 2: $\varepsilon < 2b = 8aK$. Then $2^{-t}K > \varepsilon/8$, so $504 \cdot 2^{-t}K\varepsilon^{-2} > 63/\varepsilon \geq 63 \geq 1$ (using $\varepsilon \leq 1$), and the bound holds trivially since probabilities are at most 1. □

Together, Lemma 6.3 and Lemma 6.12 constitute the two guarantees exported by this chapter. The former shows that a heavy Fourier coefficient is preserved, essentially without distortion, under random affine restriction, using only the second moment of f; the latter shows that the aggregate fourth Fourier mass is preserved up to an error governed by the eighth moment, which quantitatively controls the possibility of many light coefficients conspiring within a single coset. As with the coset hashing primitive of the previous chapter, these concentration properties are precisely what make subspace restriction usable as an efficient sketching primitive in the algorithms developed in the subsequent chapters.

7

Application V : Near Optimal ℓ_0 Testing of Fourier Sparsity

Property testing of Boolean functions was introduced in the seminal work of Blum, Luby, and Rubinfeld [Blum et al., 1993b] in the context of PCP and related developments. Since then, it has developed into an active area of research at the intersection of sublinear algorithms and computational complexity. In this framework, a tester for a class $\mathcal{C}$ of functions $f : \mathbb{F}_2^n \to \{-1, +1\}$ is a randomized algorithm that, given oracle access to an unknown function f, must distinguish, with high probability, between the case that $f \in \mathcal{C}$ and the case that f is ε-far (in Hamming distance) from every function in $\mathcal{C}$, while making as few queries as possible. Over the past three decades, a wide range of properties of Boolean functions, e.g. low-degree-ness [Alon et al., 2005, Bhattacharyya et al., 2010], monotonicity, and unateness [Chen et al., 2015, 2017, Chen and Xie, 2016, Goldreich et al., 2000], have been studied in this model. We refer the interested reader to the excellent surveys of Goldreich [Goldreich, 2017] and Bhattacharyya and Yoshida Bhattacharyya and Yoshida [2022] for a comprehensive overview of this literature.

A recurring theme in property testing is determining whether a function admits a concise representation in an appropriate basis. Such representations capture the intuition that many natural functions, despite being defined over high-dimensional domains, possess low intrinsic complexity. A prominent example is the class of k-juntas, consisting of functions that depend on at most $k \ll n$ variables. The problem of testing juntas has been one of the active topics in property testing [Diakonikolas et al., 2007, Blais, 2009, Fischer et al., 2004] and its more refined variants, including tolerant testing [Nadimpalli and Patel, 2024], relative-error testing [Chen et al., 2025], and distribution-free testing [Liu et al., 2019]. Following nearly two decades of research, the query complexity of junta testing in the standard property testing model is now essentially settled, with near-optimal upper and lower bounds having been established [Blais, 2009, Bshouty, 2020]. This is widely regarded as a major milestone in the property testing literature.

Apart from juntas, another fundamental measure of complexity, Fourier sparsity, arises naturally in many areas of theoretical computer science. Before proceeding further, we

Fourier Sparsity Testing Problem

Given oracle access to an unknown function $f : \mathbb{F}_2^n \to \{-1, +1\}$ and parameters $s \in \mathbb{N}$ and $\varepsilon \in (0, 1)$, the Fourier sparsity testing problem is to design a randomized algorithm that:

- accepts with probability at least $2/3$ if f is s-Fourier sparse, and
- rejects with probability at least $2/3$ if f is ε-far (in Hamming distance) from every s-Fourier sparse Boolean function.

The algorithm aims to minimize the number of queries.

Figure 7.1: Fourier sparsity testing problem

formally define Fourier sparsity. By the fundamental theorem of Fourier analysis O'Donnell [2014], every function $f : \mathbb{F}_2^n \to \{-1, 1\}$ admits a Fourier expansion

$$f(x) = \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)\,(-1)^{\langle x,\alpha \rangle},$$

where $\widehat{f}(\alpha)$ denotes the Fourier coefficient of f at α. The Fourier sparsity of f is defined as the number of nonzero coefficients in this expansion.

Fourier sparsity plays a crucial role across several areas of theoretical computer science, including communication complexity [Tsang et al., 2013, Shpilka et al., 2017], cryptography [Goldreich and Levin, 1989], and learning theory Kushilevitz and Mansour [1993]. Many natural classes of functions, such as junta functions, low-depth decision trees, low-degree functions, and hypergraph cut functions, exhibit sparse Fourier spectra. Given this reach, testing Fourier sparsity is a natural and consequential problem in its own right — but beyond this intrinsic interest, it is also motivated by a concrete need within learning theory.

The problem of learning Fourier-sparse Boolean functions dates back to the pioneering work of Kushilevitz and Mansour [1993], inspired by the seminal work of Goldreich and Levin [1989]. Over the past two decades, there has been renewed interest in the exact reconstruction of Fourier-sparse Boolean functions, with complexity depending primarily on the sparsity parameter. Two main algorithmic paradigms have emerged: the Sparse Hadamard Transform [Stobbe and Krause, 2012, Hassanieh et al., 2012b, Indyk et al., 2014] and compressed sensing techniques [Rudelson and Vershynin, 2008, Ishay and Regev, 2017]. A common assumption in these works is that the sparsity level is known in advance, which may not hold in practice. Moreover, exact reconstruction methods,

such as ℓ_1-minimization-based compressed sensing [Ishay and Regev, 2017], succeed only when the function is exactly, or sufficiently close to, s-Fourier sparse — typically requiring Hamming distance at most $o(1/s)$ from the nearest s-sparse function. Some form of certification of Fourier sparsity is therefore crucial for obtaining meaningful recovery guarantees from such algorithms. This is precisely the perspective articulated by Goldreich et al. [1998], who formalized property testing as an effective preprocessing step before invoking computationally expensive learning algorithms.

It is against this backdrop that the query complexity of testing Fourier sparsity becomes especially pressing. The problem (see the formal definition in Figure 7.1) was first studied by [Gopalan et al., 2011], who gave a tester with query complexity $\widetilde{O}(s^{14})$ and established a lower bound of $\Omega(\sqrt{s})$. Surprisingly, no significant progress was made toward closing this gap for nearly two decades — in sharp contrast to junta testing, a strict subclass of Fourier-sparse functions, where substantial progress over the same period led to an essentially complete understanding of the query complexity. Fortunately, in a recent work, [Ghosh et al., 2026] made a major breakthrough, designing a tester for testing Fourier sparsity with query complexity $\widetilde{O}(s^4)$ and simultaneously strengthening the lower bound to $\Omega(s)$. In this work, we nearly close the remaining gap by presenting a tester with query complexity $\widetilde{O}(s)$, matching the lower bound of [Ghosh et al., 2026] up to polylogarithmic factors. Finally, we note that our tester is non-adaptive, whereas the lower bound argument of [Ghosh et al., 2026] holds even against adaptive algorithms, showing that adaptivity does not provide an asymptotic advantage for this problem.

We now state our main result below.

Main Theorem

Theorem 7.1. *For every $\varepsilon \in (0,1)$, there exists a nonadaptive randomized tester for the class of Boolean functions with Fourier sparsity at most s, with query complexity $\widetilde{O}(s/\varepsilon)$, such that (i) if f has Fourier sparsity at most s, then the tester accepts except with probability $\varepsilon/200$, and (ii) if f is ε-far from every such function, then the tester rejects with probability at least $\varepsilon/100$.*

With our main theorem in hand, one can now solve the Fourier sparsity testing problem (Figure 7.1) in the following way. Let $\mathcal{T}$ be the tester guaranteed by Theorem 7.1. If f is s-Fourier sparse, then $\mathcal{T}$ accepts with probability at least $1 - \varepsilon/200$, whereas if f is ε-far, it rejects with probability at least $\varepsilon/100$. We run $\mathcal{T}$ independently $t = \Theta(1/\varepsilon^2)$ times and estimate its acceptance probability. Since the completeness and soundness probabilities differ by $\Omega(\varepsilon)$, a Chernoff bound shows that $O(1/\varepsilon^2)$ independent repetitions suffice to separate the two cases with constant probability. This increases the query complexity

by an $O(1/\varepsilon^2)$ factor while preserving the correctness of the tester. This perspective is encapsulated in the following theorem.

Theorem 7.2. *For every $\varepsilon \in (0,1)$, there exists a randomized tester for the class of Boolean functions with Fourier sparsity at most s, with query complexity $\widetilde{O}(s/\varepsilon^3)$, such that*

1. *if f has Fourier sparsity at most s, then the tester accepts with probability at least $2/3$, and*
2. *if f is ε-far from every Boolean function with Fourier sparsity at most s, then the tester rejects with probability at least $2/3$.*

We conclude this section by describing some related works, future research directions, and a sketch of the proof of our main theorem.

Related Works

The problem of testing Fourier sparsity under Hamming distance was first studied by Gopalan et al. [2011], who gave a non-adaptive tester with query complexity $O(s^6 \log s/\epsilon^2 + s^{14} \log s)$ and proved a lower bound of $\Omega(\sqrt{s})$. Furthermore, Fourier sparsity is affine-invariant, in the sense that it remains unchanged under affine transformations of the domain. Consequently, it is, at least in principle, amenable to testing via general regularity-based frameworks for testing affine-invariant properties of Boolean functions [Bhattacharyya et al., 2012]; however, such approaches incur impractical tower-type query complexities. Only recently has there been significant progress: Ghosh et al. [2026] improved the upper bound to $\widetilde{O}(s^4)$ and established a matching lower bound of $\Omega(s)$, thereby substantially narrowing the gap. These developments reduced a long-standing gap that persisted for nearly two decades.

A related version has been studied for real-valued functions $f : \mathbb{F}_2^n \to \mathbb{R}$, where distance is measured via the Euclidean norm, $\mathrm{dist}_2(f,g) = \left(\mathbb{E}_{x\sim\mathbb{F}_2^n}\left[(f(x)-g(x))^2\right]\right)^{1/2}$. Yaroslavtsev and Zhou [2020] showed that testing closeness to s-Fourier-sparse functions under this metric can be achieved with $\widetilde{O}(s)$ queries, and they established an $\Omega(\sqrt{s})$ lower bound.

However, readers may note that while closeness in Hamming distance implies closeness in Euclidean distance, the converse need not hold. To illustrate, consider two scenarios. In the first, suppose f is an exactly s-Fourier-sparse Boolean function with precisely s nonzero Fourier coefficients; then each such coefficient must have magnitude $\Omega(1/s)$. In the second, consider a Boolean function that has s large Fourier coefficients but also a tail of many small, nonzero coefficients whose total ℓ_2 mass is small. A tester for s-Fourier

sparsity under Euclidean distance would accept both types of functions, whereas a tester under Hamming distance should accept only the first. Thus, testing Fourier sparsity in the Hamming model is inherently more demanding: one must certify not only the presence of s large coefficients but also the absence of any significant Fourier tail. Consequently, a tester for Fourier sparsity under Euclidean distance cannot, in general, be used to obtain a tester under the Hamming distance.

7.1 Proof Ideas

Testing for concise representations of Boolean functions has largely been studied via the implicit learning paradigm [Diakonikolas et al., 2007]. The key insight is that many natural function classes, such as DNFs, decision trees, branching programs, and Boolean formulas, admit succinct representations and are well approximated by junta functions. For example, every s-term DNF is ε-close to one depending on $O(\log s + \log(1/\varepsilon))$ variables. This enables testers that recover the structure of a small junta without explicitly identifying the relevant variables. In particular, building on ideas from Alon et al. [2013b] and Blais [2009], Bshouty [2020] gave optimal bounds for several junta-based classes. A limitation of this approach is its reliance on junta structure. It does not extend to Fourier-sparse functions, since Fourier-sparse Boolean functions are not necessarily juntas; for instance, the parity function on n variables is 1-Fourier-sparse but depends on all variables, i.e., it is an n-junta. Recent work of Ghosh et al. [2026] circumvents this barrier by designing an affine-invariant sample extractor for Fourier-sparse functions, in the spirit of Alon et al. [2013b], obtaining an $\tilde{O}(s^4)$-query tester for the Fourier sparsity testing problem.

In this work, we develop a new approach to testing Fourier sparsity based on random affine restrictions, a technique rooted in the literature on testing affine-invariant properties [Alon et al., 2005, Bhattacharyya et al., 2012, Arnab et al., 2010]. The general paradigm of restricting functions to random affine subspaces has been widely used in this line of work, often relying on regularity-like lemmas whose query complexity has tower-type dependence on the underlying approximation parameter [Hatami and Lovett, 2011, 2013]. In contrast, our approach avoids such heavy machinery and relies on a suitable notion of self-correction [Blum et al., 1993b]. Specifically, our self-correction procedure operates in the Fourier domain: it recovers Fourier spectra directly, whereas all earlier instantiations of this paradigm [Blum et al., 1993b, Alon et al., 2005, Bhattacharyya et al., 2010] operated in the input domain, recovering function values from noisy evaluations of the function. Below, we provide a high-level overview of our approach. We deliberately omit many technical details, but hope that this overview will help the reader appreciate the main

ideas behind our algorithm.

More concretely, our approach treats affine restrictions as a hashing mechanism on the Fourier spectrum, exactly the subspace-restriction primitive developed in Chapter 6. Let $f : \mathbb{F}_2^n \to \{-1, +1\}$ be an s-Fourier-sparse function, and let $A = a + H$ be a random affine subspace, where $H \leq \mathbb{F}_2^n$ has dimension $d = O(\log s/\varepsilon)$. By Fact 6.1, the Fourier coefficients of the restricted function f_A satisfy

$$\widehat{f_A}(\gamma) = \sum_{\beta \in \gamma + H^\perp} \widehat{f}(\beta)\,(-1)^{\langle \beta, a \rangle},$$

which can be viewed as a form of Poisson summation over $\mathbb{F}_2^n$. Thus, restricting f to A aggregates its Fourier coefficients into "buckets" indexed by cosets of $H^\perp$, with an additional phase depending on the shift a. When H is chosen at random, this hashing behaves in a pairwise-independent manner (Fact 6.2), an observation that has appeared in several works on testing and learning Fourier-sparse Boolean functions [Hassanieh et al., 2012b, Gopalan et al., 2011, Ghosh and Roy, 2025]. In particular, when the dimension of H is sufficiently large, all nonzero Fourier coefficients of f are mapped to distinct cosets of $H^\perp$ with high probability. Consequently, each coefficient $\widehat{f_A}(\gamma)$ corresponds to a unique Fourier coefficient of f, up to a phase factor.

This simple observation gives rise to three key structural properties, which we exploit in our tester. First, the magnitudes $|\widehat{f_A}(\gamma)|$ are independent of the choice of the affine shift a, a property we refer to as magnitude consistency. Second, the phases satisfy a strong compatibility relation across affine shifts, which we call triangle consistency. Specifically, for all $a, r \in \mathbb{F}_2^n$ and $\gamma \in W$, we have

$$\frac{\widehat{f_{a+H}}(\gamma)}{\widehat{f_H}(\gamma)} \cdot \frac{\widehat{f_{r+H}}(\gamma)}{\widehat{f_H}(\gamma)} = \frac{\widehat{f_{a+r+H}}(\gamma)}{\widehat{f_H}(\gamma)}.$$

Moreover, provided the dimension of H is sufficiently large, the random hashing is injective not only on the Fourier support S, but also on the sumset $S + S$ with high probability. Consequently, we observe an additive consistency over the Fourier coefficients: for every $\gamma_1, \gamma_2, \gamma_3, \gamma_4 \in W$ satisfying $\gamma_1 + \gamma_2 = \gamma_3 + \gamma_4$ and $\widehat{f_H}(\gamma_i) \neq 0$ for all $i \in \{1, 2, 3, 4\}$, we have

$$\frac{\widehat{f_{a+H}}(\gamma_1)}{\widehat{f_H}(\gamma_1)} \cdot \frac{\widehat{f_{a+H}}(\gamma_2)}{\widehat{f_H}(\gamma_2)} = \frac{\widehat{f_{a+H}}(\gamma_3)}{\widehat{f_H}(\gamma_3)} \cdot \frac{\widehat{f_{a+H}}(\gamma_4)}{\widehat{f_H}(\gamma_4)}$$

for every $a \in \mathbb{F}_2^n$. These observations form the basis for the completeness analysis of our testing algorithm.

For the soundness analysis, we adopt a self-correction paradigm to recover the Fourier

spectrum of the restricted function. Towards this, for each affine coset $a + H$ we gather "votes" from other cosets to define a corrected Fourier spectrum. Concretely, for each $\gamma \in W$, we define

$$\widehat{\phi_{a+H}}(\gamma) = \operatorname*{majority}_{b \in \mathbb{F}_2^n} \frac{\widehat{f_{a+b+H}}(\gamma)\, \widehat{f_{b+H}}(\gamma)}{\widehat{f_H}(\gamma)}.$$

The remaining challenge is to prove soundness. Specifically, we show that if the tester accepts with high probability — equivalently, if the magnitude, triangle, and additive consistency conditions hold for a large fraction of affine cosets — then the input function is close to an s-Fourier sparse Boolean function. Our starting point is to prove that, under these consistency conditions, the ratio computed by our procedure identifies the same underlying Fourier coefficient with high probability. Consequently, the correct value forms a strict majority, enabling us to recover the "true" Fourier coefficient associated with almost every coset. Furthermore, we show that only a small fraction of cosets are "noisy," and these can be corrected using the same majority-based reconstruction procedure. The main technical challenge is to prove that this correction procedure preserves both Booleanity and Fourier sparsity. Although the consistency conditions allow us to reconstruct a globally consistent Fourier spectrum across affine cosets, it is not a priori clear that the resulting function ϕ is an s-Fourier sparse Boolean function, or even that such a function should exist. This, we believe, is one of our main technical contributions in this work.

The straightforward implementation of our approach requires computing the exact Fourier spectrum of each affine restriction, which is prohibitively expensive in terms of query complexity. To overcome this, we replace the exact Fourier computation by a sparse recovery procedure, broadly based on the ℓ_1-minimization-based compressed sensing framework [Candès et al., 2006, Donoho, 2006], over subsampled Walsh–Hadamard matrices [Ishay and Regev, 2017]. We then prove that incorporating this recovery procedure into Algorithm 12 preserves its completeness and soundness guarantees up to asymptotically negligible error, while reducing the query complexity to

$$O\left(s \log^2 s \, \log \frac{s}{\varepsilon} + \frac{s}{\varepsilon} \log \frac{s}{\varepsilon}\right).$$

7.2 A Query Inefficient Fourier Sparsity Tester

In this section, we present Algorithm 12, a query-inefficient tester for Fourier sparsity based on a self-correction procedure over the dual space. The algorithm enjoys the following theoretical guarantees.

Theorem 7.3. *For every $\varepsilon \in (0, 1)$, there exists a randomized tester for the class of Boolean*

functions with Fourier sparsity at most s, with query complexity $\widetilde{O}(s^4)$, such that

1. ***(Completeness)** if f has Fourier sparsity at most s, then the tester accepts except with probability $\varepsilon/200$;*
2. ***(Soundness)** if f is ε-far from every such function, then the tester rejects with probability at least $\min(\varepsilon/16, 1/64)$.*

In the next section, we show how to transform Algorithm 12 into a query-efficient tester, at the cost of a slight weakening of these theoretical guarantees.

Algorithm 12: Fourier Sparsity Tester

Input: Oracle access to $f : \mathbb{F}_2^n \to \{-1, +1\}$ and sparsity parameter s.
Output: ACCEPT if f is s-Fourier sparse; REJECT otherwise.
Initialization: Let $d = \left\lceil \log \frac{512 s^4}{\varepsilon} \right\rceil$.

Repeat $O(1)$ times, keeping any successful run:

Step 1. Let $H \leq \mathbb{F}_2^n$ be the span of d independent uniformly random vectors $v_1, \ldots, v_d \in \mathbb{F}_2^n$, and compute a complementary subspace W such that $\mathbb{F}_2^n = W \oplus H^\perp$.;

Step 2. Sample $a, r \in \mathbb{F}_2^n$ uniformly at random. For each shift $t \in \{0, a, r, a + r\}$, compute the Fourier coefficients of the restricted function f_{t+H}. If any restriction f_{t+H} is not s-Fourier sparse, REJECT.;

Step 3. For every $\gamma \in W$, verify the *Magnitude-Consistency Test*:

$$\left|\widehat{f_{a+H}}(\gamma)\right| = \left|\widehat{f_{r+H}}(\gamma)\right| = \left|\widehat{f_{a+r+H}}(\gamma)\right| = \left|\widehat{f_H}(\gamma)\right|.$$

If the test fails for any γ, REJECT.;

Step 4. For every $\gamma \in W$ with $\widehat{f_H}(\gamma) \neq 0$, verify the *Triangle-Consistency Test*:

$$\frac{\widehat{f_{a+H}}(\gamma)}{\widehat{f_H}(\gamma)} \cdot \frac{\widehat{f_{r+H}}(\gamma)}{\widehat{f_H}(\gamma)} = \frac{\widehat{f_{a+r+H}}(\gamma)}{\widehat{f_H}(\gamma)}.$$

If the test fails for any γ, REJECT.;

Step 5. For every $\gamma_1, \gamma_2, \gamma_3, \gamma_4 \in W$ satisfying $\gamma_1 + \gamma_2 = \gamma_3 + \gamma_4$ and $\widehat{f_H}(\gamma_i) \neq 0$ for all $i \in \{1, 2, 3, 4\}$, verify the *Additive-Consistency Test*:

$$\frac{\widehat{f_{a+H}}(\gamma_1)}{\widehat{f_H}(\gamma_1)} \cdot \frac{\widehat{f_{a+H}}(\gamma_2)}{\widehat{f_H}(\gamma_2)} = \frac{\widehat{f_{a+H}}(\gamma_3)}{\widehat{f_H}(\gamma_3)} \cdot \frac{\widehat{f_{a+H}}(\gamma_4)}{\widehat{f_H}(\gamma_4)}.$$

If the test fails for any such quadruple, REJECT.;

Step 6. If all the above tests pass, ACCEPT.;

7.2.1 Proof of Completeness

Assume that f is s-Fourier sparse with Fourier support $S = \{\beta_1, \dots, \beta_s\}$. Let $H \leq \mathbb{F}_2^n$ be the random d-dimensional subspace chosen in Step 1 (the span of d independent uniform vectors), and let $\mathbb{F}_2^n = W \oplus H^\perp$. For any shift $a \in \mathbb{F}_2^n$, define the restricted function

$$f_{a+H}(x) = f(a + x)$$

for $x \in H$. Recall from Fact 6.1 that for every $\gamma \in W$,

$$\widehat{f_{a+H}}(\gamma) = \sum_{\beta \in \gamma + H^\perp} \widehat{f}(\beta)\chi_\beta(a).$$

Thus restricting f to a coset of H aggregates Fourier coefficients lying in the same coset of $H^\perp$.

Define the map

$$T_H : \mathbb{F}_2^n \to W, \qquad T_H(\beta) = \gamma \quad \text{iff} \quad \beta \in \gamma + H^\perp.$$

For distinct $\beta, \beta' \in \mathbb{F}_2^n$,

$$T_H(\beta) = T_H(\beta') \iff \beta - \beta' \in H^\perp.$$

By Fact 6.2, any two distinct frequencies satisfy $\beta - \beta' \in H^\perp$ — equivalently $T_H(\beta) = T_H(\beta')$ — with probability exactly 2^{-d}. Hence

$$\Pr_H[T_H(\beta) = T_H(\beta')] = 2^{-d}.$$

Applying a union bound over all pairs in S,

$$\Pr_H[\exists\, i \neq j : T_H(\beta_i) = T_H(\beta_j)] \leq \binom{s}{2} 2^{-d}.$$

With $d = \lceil \log(512s^4/\varepsilon) \rceil$, this probability is at most $\varepsilon/512$. Let E_1 denote this bad event, and condition on its complement. Then every coset $\gamma + H^\perp$ contains at most one nonzero Fourier coefficient of f. Hence for every $\gamma \in W$, either there exists a unique $\beta_\gamma \in S$ such that $T_H(\beta_\gamma) = \gamma$, or no such element exists. In the first case, $\widehat{f_{a+H}}(\gamma) = \widehat{f}(\beta_\gamma)\chi_{\beta_\gamma}(a)$, while in the second case $\widehat{f_{a+H}}(\gamma) = 0$. Consequently, for every shift a,

$$|\operatorname{supp}(\widehat{f_{a+H}})| \leq s.$$

Therefore every restriction f_{a+H} is s-Fourier sparse, and hence the sparsity test in Step 2 always accepts.

For every $\gamma \in W$ such that $\widehat{f_H}(\gamma) \neq 0$, we have

$$\widehat{f_{a+H}}(\gamma) = \widehat{f}(\beta_\gamma)\chi_{\beta_\gamma}(a), \qquad \widehat{f_{r+H}}(\gamma) = \widehat{f}(\beta_\gamma)\chi_{\beta_\gamma}(r), \qquad \widehat{f_{a+r+H}}(\gamma) = \widehat{f}(\beta_\gamma)\chi_{\beta_\gamma}(a+r).$$

Consequently,

$$\left|\widehat{f_{a+H}}(\gamma)\right| = \left|\widehat{f_{r+H}}(\gamma)\right| = \left|\widehat{f_{a+r+H}}(\gamma)\right| = \left|\widehat{f_H}(\gamma)\right|,$$

so the magnitude-consistency test in Step 3 always succeeds. We next verify the triangle-consistency test in Step 4. Since $\widehat{f_H}(\gamma) = \widehat{f}(\beta_\gamma)$, and characters are multiplicative, we obtain

$$\frac{\widehat{f_{a+H}}(\gamma)}{\widehat{f_H}(\gamma)} \cdot \frac{\widehat{f_{r+H}}(\gamma)}{\widehat{f_H}(\gamma)} = \chi_{\beta_\gamma}(a) \cdot \chi_{\beta_\gamma}(r) = \chi_{\beta_\gamma}(a+r) = \frac{\widehat{f_{a+r+H}}(\gamma)}{\widehat{f_H}(\gamma)}.$$

For every $\gamma \in W$ such that $\widehat{f_H}(\gamma) = 0$, we define $0/0 := 0$; hence the relation also holds for these coefficients. Thus the triangle-consistency test always passes.

We now verify the additive-consistency test in Step 5.

Consider the sumset

$$S + S = \{\beta_i + \beta_j : 1 \leq i, j \leq s\}.$$

Since $|S| = s$, $|S + S| \leq s^2$. Again by Fact 6.2, distinct $\alpha, \alpha' \in S + S$ satisfy

$$\Pr_H[T_H(\alpha) = T_H(\alpha')] = 2^{-d}.$$

Applying a union bound,

$$\Pr_H[\exists\, \alpha \neq \alpha' \in S + S : T_H(\alpha) = T_H(\alpha')] \leq \binom{|S+S|}{2} 2^{-d} \leq s^4 2^{-d}.$$

By the choice of d, this probability is at most $\varepsilon/512$. Let E_2 denote this bad event and condition on its complement.

Suppose $\gamma_1 + \gamma_2 = \gamma_3 + \gamma_4$ and $\widehat{f_H}(\gamma_i) \neq 0$ for $i = 1, 2, 3, 4$. Recall that, conditioned on $\overline{E_1}$, there exist unique $\beta_{\gamma_i} \in S$ such that $T_H(\beta_{\gamma_i}) = \gamma_i$.

Since T_H is linear,

$$T_H(\beta_{\gamma_1} + \beta_{\gamma_2}) = \gamma_1 + \gamma_2 = \gamma_3 + \gamma_4 = T_H(\beta_{\gamma_3} + \beta_{\gamma_4}).$$

Since T_H is injective on $S + S$ under the conditioning on $\overline{E_2}$,

$$\beta_{\gamma_1} + \beta_{\gamma_2} = \beta_{\gamma_3} + \beta_{\gamma_4}.$$

Recalling that $\widehat{f_{a+H}}(\gamma) = \widehat{f}(\beta_\gamma)\chi_{\beta_\gamma}(a)$, we deduce that

$$\frac{\widehat{f_{a+H}}(\gamma_1)}{\widehat{f_H}(\gamma_1)}\frac{\widehat{f_{a+H}}(\gamma_2)}{\widehat{f_H}(\gamma_2)} = \chi_{\beta_{\gamma_1}}(a) \cdot \chi_{\beta_{\gamma_2}}(a) = \chi_{\beta_{\gamma_3}}(a) \cdot \chi_{\beta_{\gamma_4}}(a) = \frac{\widehat{f_{a+H}}(\gamma_3)}{\widehat{f_H}(\gamma_3)}\frac{\widehat{f_{a+H}}(\gamma_4)}{\widehat{f_H}(\gamma_4)}.$$

Thus the additive-consistency test also always passes.

Finally,

$$\Pr[E_1 \cup E_2] \leq \Pr[E_1] + \Pr[E_2] \leq \frac{\varepsilon}{512} + \frac{\varepsilon}{512} = \frac{\varepsilon}{256}.$$

Therefore, with probability at least $1 - \varepsilon/256$ over the choice of H, all tests performed by Algorithm 12 succeed.

7.2.2 Proof of Soundness

We now turn to the soundness of the algorithm. We show that if f is ε-far from every s-Fourier sparse function, then the algorithm rejects f with probability at least $\min\{\varepsilon/16,\ 1/64\}$. We argue by contraposition.

Let η denote the rejection probability of Algorithm 12, where the probability is taken over all random choices of the tester, including the choice of the random subspace H. Assume that $\eta < 1/64$. For every choice of the random subspace H, let

$$\rho = \Pr[\text{Algorithm rejects} \mid H].$$

Since $\mathbb{E}_H[\rho] = \eta$, Markov's inequality implies that $\Pr_H[\rho > 8\eta] \leq \frac{1}{8}$. Hence, with probability at least $7/8$ over the choice of H, we have $\rho \leq 8\eta$. Throughout the remainder of the proof, we fix such a choice of H. Consequently, $\rho \leq 8\eta < 1/8$, and therefore the tester accepts with conditional probability at least $1 - \rho > 7/8$. Thus, for most choices of the remaining random shifts, the corresponding consistency tests succeed. We build our self-correction procedure on this observation.

At a high level, we attempt to construct a function ϕ by first recovering its restrictions over the cosets of H via our self-correction procedure. Towards this, fix a coset $a + H$. For a uniformly random $b \in \mathbb{F}_2^n$, we define the *vote of* b for the restriction of f to $a + H$ as the vector

$$\widehat{\phi^b_{a+H}} := \Big(\widehat{\phi^b_{a+H}}(\gamma_1), \ldots, \widehat{\phi^b_{a+H}}(\gamma_m)\Big),$$

where $W = \{\gamma_1, \ldots, \gamma_m\}$ denotes the dual space ordered lexicographically and

$$\widehat{\phi^b_{a+H}}(\gamma) \stackrel{\text{def}}{=} \frac{\widehat{f_{a+b+H}}(\gamma)\,\widehat{f_{b+H}}(\gamma)}{\widehat{f_H}(\gamma)}.$$

Intuitively, when the triangle consistency condition holds for the cosets $(H,\ b+H,\ a+b+H)$, the above expression correctly reconstructs the Fourier coefficient of the underlying function on the coset $a + H$. We now aggregate the votes over all b. Define

$$\widehat{\phi_{a+H}} := \arg\max_{v \in \mathbb{R}^{|W|}} \big|\{\, b \in \mathbb{F}_2^n : \widehat{\phi^b_{a+H}} = v \,\}\big|,$$

breaking ties arbitrarily. In other words, $\widehat{\phi_{a+H}}$ is the most frequent vote among all $\widehat{\phi^b_{a+H}}$.

For the time being, we treat ϕ as a hypothetical function whose restriction to each coset $a + H$ is given by ϕ_{a+H}. While it is not a priori clear that such a function exists whose restrictions are consistent with the Fourier spectrum constructed by the self-correction procedure, our analysis shows that such a function indeed exists, and moreover, it is s-Fourier sparse and Boolean-valued. We begin by showing that the self-correction procedure recovers the true Fourier spectrum on most cosets. The key idea is that, whenever the test accepts, the triangle consistency conditions ensure that a majority of the votes reconstruct the true Fourier coefficients. Aggregating these votes therefore yields the correct spectrum on most cosets, implying that the resulting function ϕ is close to f.

Claim 7.4. *The function ϕ is 2ρ-close to f.*

Proof.

Fix a coset $a + H$. Recall that for each $b \in \mathbb{F}_2^n$, $\widehat{\phi^b_{a+H}}$ denotes the Fourier spectrum suggested by the cosets $b + H$ and $a + b + H$, and that $\widehat{\phi_{a+H}}$ is defined as the majority value of these votes.

Call a coset $a + H$ *bad* if

$$\Pr_{b \in \mathbb{F}_2^n}\big[\widehat{f_{a+H}}(\gamma) \neq \widehat{\phi^b_{a+H}}(\gamma) \text{ for some } \gamma \in W\big] \;\geq\; \frac{1}{2},$$

and let $\mathcal{B}$ denote the set of bad cosets. Averaging over a, we obtain

$$\rho \;\geq\; \Pr[a+H \in \mathcal{B}] \cdot \min_{a+H\in\mathcal{B}} \Pr_{b\in\mathbb{F}_2^n}\left[\widehat{f_{a+H}}(\gamma) \neq \widehat{\phi^b_{a+H}}(\gamma) \text{ for some } \gamma \in W\right] \;\geq\; \frac{|\mathcal{B}|}{|\mathbb{F}_2^n|}\cdot\frac{1}{2}.$$

It follows that $|\mathcal{B}| \leq 2\rho\,|\mathbb{F}_2^n|$. On the other hand, if $a+H \notin \mathcal{B}$, then a strict majority of the votes $\widehat{\phi^b_{a+H}}$ agree with the true Fourier profile of f on $a+H$. Since $\widehat{\phi_{a+H}}$ is defined to be the majority value of these votes, it follows that

$$\widehat{\phi_{a+H}}(\gamma) = \widehat{f_{a+H}}(\gamma) \qquad \text{for all } \gamma \in W.$$

Thus, ϕ differs from f on at most a 2ρ-fraction of cosets, and the claim follows. □

Next, we show that the aggregation step is stable: a random vote agrees with the final (majority) estimate with high probability. Intuitively, if two independent random choices of b satisfy the consistency conditions, they must produce the same reconstructed spectrum. This leads to a high collision probability, which in turn implies that a typical vote agrees with the final (majority) estimate with high probability.

Claim 7.5. *For every coset $a+H$, it holds that*

$$\Pr_{b\in\mathbb{F}_2^n}\left[\widehat{\phi^b_{a+H}} = \widehat{\phi_{a+H}}\right] \;\geq\; 1-2\rho.$$

Proof. Fix a coset $a+H$. Recall that $\widehat{\phi_{a+H}}$ was defined as the majority value of the random variable $\widehat{\phi^b_{a+H}}$, where b is chosen uniformly from $\mathbb{F}_2^n$. Define the random variable

$$Z_a = Z_a(b) \;\stackrel{\text{def}}{=}\; \widehat{\phi^b_{a+H}},$$

which takes values in $\mathbb{R}^{|W|}$. Our goal is to show that Z_a assumes its most frequent value (namely $\widehat{\phi_{a+H}}$) with probability at least $1-2\rho$. To show this, we lower-bound the collision probability of Z_a:

$$\sum_v \Pr[Z_a = v]^2 \;=\; \Pr_{b_1,b_2}\left[Z_a(b_1) = Z_a(b_2)\right].$$

We call a pair $(b_1,b_2) \in \mathbb{F}_2^n \times \mathbb{F}_2^n$ *good* if the triangle consistency condition holds simultaneously for the triples

$$(b_1+b_2+H,\; b_2+H,\; b_1+H) \qquad \text{and} \qquad (a+b_1+H,\; b_1+b_2+H,\; a+b_2+H).$$

Equivalently, (b_1, b_2) is good if for every $\gamma \in W$,

$$\widehat{f_{b_1+H}}(\gamma) = \frac{\widehat{f_{b_2+H}}(\gamma)\,\widehat{f_{b_1+b_2+H}}(\gamma)}{\widehat{f_H}(\gamma)} \tag{1}$$

and

$$\widehat{f_{a+b_2+H}}(\gamma) = \frac{\widehat{f_{a+b_1+H}}(\gamma)\,\widehat{f_{b_1+b_2+H}}(\gamma)}{\widehat{f_H}(\gamma)}. \tag{2}$$

Since each consistency check fails with probability at most ρ, a union bound implies

$$\Pr_{b_1,b_2}\big[(b_1, b_2) \text{ is good}\big] \geq 1 - 2\rho.$$

Fix any good pair (b_1, b_2). For every $\gamma \in W$ we compute

$$Z_a(b_2) = \frac{\widehat{f_{a+b_2+H}}(\gamma)\,\widehat{f_{b_2+H}}(\gamma)}{\widehat{f_H}(\gamma)}.$$

Applying (2) gives

$$Z_a(b_2) = \frac{\widehat{f_{a+b_1+H}}(\gamma)\,\widehat{f_{b_1+b_2+H}}(\gamma)\,\widehat{f_{b_2+H}}(\gamma)}{\widehat{f_H}(\gamma)^2}.$$

Applying (1) we obtain

$$\widehat{f_{b_1+b_2+H}}(\gamma)\,\widehat{f_{b_2+H}}(\gamma) = \widehat{f_{b_1+H}}(\gamma)\,\widehat{f_H}(\gamma).$$

Substituting this identity yields

$$Z_a(b_2) = \frac{\widehat{f_{a+b_1+H}}(\gamma)\,\widehat{f_{b_1+H}}(\gamma)}{\widehat{f_H}(\gamma)} = Z_a(b_1).$$

Thus for every good pair (b_1, b_2), $Z_a(b_1) = Z_a(b_2)$ for all $\gamma \in W$.

Consequently,

$$\Pr_{b_1,b_2}\big[Z_a(b_1) = Z_a(b_2)\big] \geq \Pr_{b_1,b_2}\big[(b_1, b_2) \text{ is good}\big] \geq 1 - 2\rho.$$

Finally, since

$$\sum_v \Pr[Z_a = v]^2 \le \max_v \Pr[Z_a = v],$$

it follows that

$$\Pr_b\left[\widehat{\phi^b_{a+H}} = \widehat{\phi_{a+H}}\right] \ge 1 - 2\rho,$$

because $\widehat{\phi_{a+H}}$ is defined to be the most frequent value of Z_a. This completes the proof. □

We now show that the corrected function ϕ satisfies the triangle consistency condition across all pairs of cosets.

Claim 7.6. *For every pair of cosets $a+H$ and $b+H$, the corrected function ϕ satisfies*

$$\widehat{\phi_{a+H}}(\gamma)\widehat{\phi_{b+H}}(\gamma) = \widehat{\phi_{a+b+H}}(\gamma)\widehat{\phi_H}(\gamma) \qquad \textit{for all } \gamma \in W.$$

Proof. Fix a, b. Define

$$p_{a,b} = \Pr_r\left[\exists\, \gamma \in W : \widehat{\phi_{a+H}}(\gamma)\widehat{\phi_{b+H}}(\gamma) \neq \widehat{\phi_{a+b+H}}(\gamma)\widehat{\phi_H}(\gamma)\right].$$

For $r \in \mathbb{F}_2^n$, recall

$$\widehat{\phi^r_{x+H}}(\gamma) = \frac{\widehat{f_{x+r+H}}(\gamma)\widehat{f_{r+H}}(\gamma)}{\widehat{f_H}(\gamma)}.$$

Define the following bad events (over r):

$$E_1 :\ \widehat{\phi_H} \neq \widehat{\phi^r_H}, \qquad E_2 :\ \widehat{\phi_{a+H}} \neq \widehat{\phi^r_{a+H}},$$
$$E_3 :\ \widehat{\phi_{b+H}} \neq \widehat{\phi^r_{b+H}}, \quad E_4 :\ \widehat{\phi_{a+b+H}} \neq \widehat{\phi^{a+r}_{a+b+H}}.$$

We first show that

$$p_{a,b} \le \Pr_r[E_1 \vee E_2 \vee E_3 \vee E_4]. \tag{7.1}$$

Assume none of the events occur. Then for all $\gamma \in W$,

$$\widehat{\phi_H}(\gamma) = \widehat{\phi^r_H}(\gamma), \quad \widehat{\phi_{a+H}}(\gamma) = \widehat{\phi^r_{a+H}}(\gamma), \quad \widehat{\phi_{b+H}}(\gamma) = \widehat{\phi^r_{b+H}}(\gamma), \quad \widehat{\phi_{a+b+H}}(\gamma) = \widehat{\phi^{a+r}_{a+b+H}}(\gamma).$$

Using the definitions,

$$\widehat{\phi^r_{a+H}}(\gamma) = \frac{\widehat{f_{a+r+H}}(\gamma)\widehat{f_{r+H}}(\gamma)}{\widehat{f_H}(\gamma)}, \quad \widehat{\phi^r_{b+H}}(\gamma) = \frac{\widehat{f_{b+r+H}}(\gamma)\widehat{f_{r+H}}(\gamma)}{\widehat{f_H}(\gamma)},$$

$$\widehat{\phi^r_H}(\gamma) = \frac{\widehat{f_{r+H}}(\gamma)^2}{\widehat{f_H}(\gamma)}, \quad \widehat{\phi^{a+r}_{a+b+H}}(\gamma) = \frac{\widehat{f_{b+r+H}}(\gamma)\widehat{f_{a+r+H}}(\gamma)}{\widehat{f_H}(\gamma)}.$$

Compute:

$$\widehat{\phi^r_{a+H}}(\gamma)\widehat{\phi^r_{b+H}}(\gamma) = \frac{\widehat{f_{a+r+H}}(\gamma)\widehat{f_{b+r+H}}(\gamma)\widehat{f_{r+H}}(\gamma)^2}{\widehat{f_H}(\gamma)^2},$$

and

$$\widehat{\phi^{a+r}_{a+b+H}}(\gamma)\widehat{\phi^r_H}(\gamma) = \frac{\widehat{f_{a+r+H}}(\gamma)\widehat{f_{b+r+H}}(\gamma)\widehat{f_{r+H}}(\gamma)^2}{\widehat{f_H}(\gamma)^2}.$$

Hence,

$$\widehat{\phi^r_{a+H}}(\gamma)\widehat{\phi^r_{b+H}}(\gamma) = \widehat{\phi^{a+r}_{a+b+H}}(\gamma)\widehat{\phi^r_H}(\gamma).$$

Substituting back (since no bad event occurs), we obtain

$$\widehat{\phi_{a+H}}(\gamma)\widehat{\phi_{b+H}}(\gamma) = \widehat{\phi_{a+b+H}}(\gamma)\widehat{\phi_H}(\gamma).$$

Thus, whenever none of the bad events occur, the identity holds.

By Claim 7.5, each event satisfies

$$\Pr[E_i] \le 2\rho,$$

and hence

$$\Pr[E_1 \vee E_2 \vee E_3 \vee E_4] \le 8\rho.$$

Therefore,

$$p_{a,b} \le 8\rho.$$

If $\rho < 1/8$, then $p_{a,b} < 1$, hence there exists r for which no bad event occurs. Since the identity is independent of r, it holds for all $\gamma \in W$. □

The following claim shows that the self-corrected local Fourier coefficients satisfy the same shift identity as the Fourier coefficients of a genuine function.

Claim 7.7. *For every $a \in \mathbb{F}_2^n$, every $h \in H$, and every $\gamma \in W$,*

$$\widehat{\phi_{a+h+H}}(\gamma) = \chi_\gamma(h)\,\widehat{\phi_{a+H}}(\gamma).$$

Proof.

For fixed a, h, and γ, define the events

$$E_1 = \left\{ \widehat{\phi_{a+h+H}}(\gamma) = \widehat{\phi^b_{a+h+H}}(\gamma) \right\},$$

and

$$E_2 = \left\{ \widehat{\phi_{a+H}}(\gamma) = \widehat{\phi^b_{a+H}}(\gamma) \right\},$$

where b is chosen uniformly at random from $\mathbb{F}_2^n$.

By definition,

$$\widehat{\phi^b_{a+h+H}}(\gamma) = \frac{\widehat{f_{a+h+b+H}}(\gamma)\,\widehat{f_{b+H}}(\gamma)}{\widehat{f_H}(\gamma)}.$$

Applying Fact 6.1 with the shift $a + b$ gives

$$\widehat{f_{h+(a+b)+H}}(\gamma) = \sum_{\beta \in \gamma + H^\perp} \widehat{f}(\beta)\,\chi_\beta(h + a + b).$$

Since $\beta - \gamma \in H^\perp$ and $h \in H$, it follows that

$$\chi_\beta(h) = \chi_\gamma(h).$$

Hence

$$\widehat{f_{h+(a+b)+H}}(\gamma) = \chi_\gamma(h) \sum_{\beta \in \gamma + H^\perp} \widehat{f}(\beta)\chi_\beta(a + b) = \chi_\gamma(h)\,\widehat{f_{a+b+H}}(\gamma).$$

Since $h + (a + b) = a + h + b$, substituting into the definition of the vote yields

$$\widehat{\phi^b_{a+h+H}}(\gamma) = \chi_\gamma(h)\,\widehat{\phi^b_{a+H}}(\gamma).$$

Now, by Claim 7.5,

$$\Pr(E_1) \geq 1 - 2\rho \qquad \text{and} \qquad \Pr(E_2) \geq 1 - 2\rho.$$

Hence, by the union bound,

$$\Pr(E_1 \cap E_2) \geq 1 - 4\rho > 0,$$

where the last inequality follows from the assumption $\rho < \frac{1}{8}$.

Therefore, there exists a choice of b for which both events occur. Consequently,

$$\widehat{\phi_{a+h+H}}(\gamma) = \widehat{\phi^b_{a+h+H}}(\gamma) = \chi_\gamma(h)\,\widehat{\phi^b_{a+H}}(\gamma) = \chi_\gamma(h)\,\widehat{\phi_{a+H}}(\gamma).$$

□

Claim 7.8. *For every $\gamma \in W$,*

$$\left|\widehat{\phi_H}(\gamma)\right| = \left|\widehat{f_H}(\gamma)\right|.$$

Proof. Fix $\gamma \in W$. By Claim 7.5,

$$\Pr_a\left[\widehat{\phi_H}(\gamma) \neq \widehat{\phi^a_H}(\gamma)\right] \leq 2\rho.$$

Moreover, since the Magnitude-Consistency Test fails with probability at most ρ,

$$\Pr_a\left[\left|\widehat{\phi^a_H}(\gamma)\right| \neq \left|\widehat{f_H}(\gamma)\right|\right] \leq \rho.$$

Hence, by the union bound,

$$\Pr_a\left[\widehat{\phi_H}(\gamma) = \widehat{\phi^a_H}(\gamma) \text{ and } \left|\widehat{\phi^a_H}(\gamma)\right| = \left|\widehat{f_H}(\gamma)\right|\right] \geq 1 - 3\rho > \frac{1}{2}.$$

Therefore, a strict majority of the votes contributing to $\widehat{\phi_H}(\gamma)$ have magnitude $\left|\widehat{f_H}(\gamma)\right|$, and consequently

$$\left|\widehat{\phi_H}(\gamma)\right| = \left|\widehat{f_H}(\gamma)\right|.$$

Since this holds for every $\gamma \in W$, the supports of $\widehat{\phi_H}$ and $\widehat{f_H}$ coincide. □

We now show that the locally self-corrected Fourier spectra can be assembled into a single globally defined function whose restrictions agree with the corrected local spectra and whose Fourier support has size at most s.

Claim 7.9. *There exists a function $\phi : \mathbb{F}_2^n \to \mathbb{R}$ such that, for every coset $a + H$, the restriction ϕ_{a+H} has Fourier expansion equal to the self-corrected spectrum $\widehat{\phi_{a+H}}$, and ϕ is s-Fourier sparse.*

Proof. For each $\alpha \in W$, define

$$F_\alpha(a) := \widehat{\phi_{a+H}}(\alpha), \qquad a \in \mathbb{F}_2^n,$$

so that F_α is determined entirely by the local correction procedure.

Claim 7.7 implies that

$$F_\alpha(a+h) = \chi_\alpha(h) F_\alpha(a) \qquad (a \in \mathbb{F}_2^n,\ h \in H),$$

while Claim 7.6 gives

$$F_\alpha(a) F_\alpha(b) = F_\alpha(a+b) F_\alpha(0) \qquad (a, b \in \mathbb{F}_2^n).$$

We distinguish two cases.

Case 1: $F_\alpha(0) = 0$**.** Setting $b = a$ in the multiplicative identity gives

$$F_\alpha(a)^2 = F_\alpha(0)^2 = 0,$$

so $F_\alpha \equiv 0$. Thus α does not belong to the support of $\widehat{\phi_H}$ and contributes nothing to the construction.

Case 2: $F_\alpha(0) \neq 0$**.** Define

$$\widehat{\phi_H}(\alpha) := F_\alpha(0), \qquad G_\alpha(a) := \frac{F_\alpha(a)}{F_\alpha(0)}.$$

The multiplicative identity becomes $G_\alpha(a+b) = G_\alpha(a) G_\alpha(b)$, so G_α is a homomorphism from $(\mathbb{F}_2^n, +)$ to $(\mathbb{R}^\times, \cdot)$. Since every element of $\mathbb{F}_2^n$ has order two,

$$G_\alpha(a)^2 = G_\alpha(2a) = G_\alpha(0) = 1,$$

and hence $G_\alpha(a) \in \{-1, +1\}$ for every a. Therefore there exists a unique $\delta_\alpha \in \mathbb{F}_2^n$ such that $G_\alpha(a) = \chi_{\delta_\alpha}(a)$, or equivalently,

$$F_\alpha(a) = F_\alpha(0) \chi_{\delta_\alpha}(a) \qquad \text{for every } a \in \mathbb{F}_2^n.$$

To identify δ_α, apply the shift identity with $a = 0$:

$$F_\alpha(h) = \chi_\alpha(h) F_\alpha(0) \qquad (h \in H).$$

On the other hand, substituting $a = h$ into the previous equation yields $F_\alpha(h) = F_\alpha(0) \chi_{\delta_\alpha}(h)$. Since $F_\alpha(0) \neq 0$, $\chi_{\delta_\alpha}(h) = \chi_\alpha(h)$ for every $h \in H$, which is equivalent to $\delta_\alpha - \alpha \in H^\perp$, or, equivalently, $\delta_\alpha \in \alpha + H^\perp$.

We now define $\phi(x)$ as

$$\phi(x) := \sum_{\alpha \in \mathrm{supp}(\widehat{\phi_H})} \widehat{\phi_H}(\alpha)\chi_{\delta_\alpha}(x).$$

Since $\delta_\alpha \in \alpha + H^\perp$, restricting ϕ to any coset $a + H$ yields precisely the self-corrected Fourier expansion $\widehat{\phi_{a+H}}$.

What remains is to show that ϕ is s-Fourier sparse. We have already shown that, for every affine coset $a + H$, the Fourier coefficients of ϕ_{a+H} and ϕ_H have the same magnitude. Furthermore, by Claim 7.8,

$$\left|\widehat{\phi_H}(\gamma)\right| = \left|\widehat{f_H}(\gamma)\right| \qquad \text{for every } \gamma \in W.$$

Since f_H is s-Fourier sparse, it follows that ϕ_H is also s-Fourier sparse. Consequently ϕ is s-Fourier sparse. □

What remains is to show that ϕ is Boolean-valued. It suffices to prove that each restriction ϕ_{a+H} is Boolean-valued. Recall that a function $g : \mathbb{F}_2^n \to \mathbb{R}$ is Boolean if and only if

$$\sum_{\alpha_1+\alpha_2=\beta} \widehat{g}(\alpha_1)\widehat{g}(\alpha_2) = \begin{cases} 1, & \beta = 0, \\ 0, & \beta \neq 0. \end{cases}$$

Indeed, by the convolution theorem, $\widehat{g^2} = \widehat{g} * \widehat{g}$, so

$$\widehat{g^2}(\beta) = \sum_{\alpha_1+\alpha_2=\beta} \widehat{g}(\alpha_1)\widehat{g}(\alpha_2).$$

Hence, the above condition is equivalent to $\widehat{g^2} = \widehat{1}$. Applying the inverse Fourier transform gives $g^2 \equiv 1$, which is equivalent to $g(x) \in \{-1, +1\}$ for every $x \in \mathbb{F}_2^n$. Observe that the case $\beta = 0$ is precisely Parseval, while the cases $\beta \neq 0$ are exactly the Titsworth relations.

Claim 7.10. *For every coset $a + H$, the function ϕ_{a+H} satisfies the Titsworth relations*

$$\sum_{\beta_1+\beta_2=\beta} \widehat{\phi_{a+H}}(\beta_1)\widehat{\phi_{a+H}}(\beta_2) = 0 \qquad \forall\, \beta \neq 0,$$

and Parseval's identity

$$\sum_{\gamma} \widehat{\phi_{a+H}}(\gamma)^2 = 1.$$

Consequently, ϕ_{a+H} is Boolean.

Proof. Fix a coset $a + H$ and $\beta \neq 0$, and define

$$S := \sum_{\beta_1+\beta_2=\beta} \widehat{\phi_{a+H}}(\beta_1)\widehat{\phi_{a+H}}(\beta_2).$$

For uniformly random $r \in \mathbb{F}_2^n$, let

$$S_r := \sum_{\beta_1+\beta_2=\beta} \widehat{\phi^r_{a+H}}(\beta_1)\widehat{\phi^r_{a+H}}(\beta_2).$$

Define the events

$$E_1 := \Big[\exists\, \gamma_1,\gamma_2,\gamma_3,\gamma_4 \in W : \gamma_1+\gamma_2 = \gamma_3+\gamma_4,$$

$$\widehat{f}_H(\gamma_i) \neq 0\ \forall i, \quad \frac{\widehat{f}_{r+H}(\gamma_1)}{\widehat{f}_H(\gamma_1)}\frac{\widehat{f}_{r+H}(\gamma_2)}{\widehat{f}_H(\gamma_2)} \neq \frac{\widehat{f}_{r+H}(\gamma_3)}{\widehat{f}_H(\gamma_3)}\frac{\widehat{f}_{r+H}(\gamma_4)}{\widehat{f}_H(\gamma_4)}\Big],$$

and

$$E_2 := \Big[\exists\, \gamma : \widehat{\phi^r_{a+H}}(\gamma) \neq \widehat{\phi_{a+H}}(\gamma)\Big].$$

The additive-consistency test in Algorithm 12 is performed on a uniformly random shift. Hence the random variable r above has the same distribution as the tested shift, and Claim 7.4 implies

$$\Pr[E_1] \leq \rho.$$

Moreover,

$$\Pr[E_2] \leq 2\rho.$$

Therefore,

$$\Pr[E_1 \cup E_2] \leq 3\rho < 1,$$

provided $\rho < 1/8$. Thus there exists a choice of r for which neither event occurs.

Fix such an r. Since E_1 does not occur, there exists $C_{r,\beta} \in \{-1,+1\}$ such that

$$\frac{\widehat{f}_{r+H}(\beta_1)}{\widehat{f}_H(\beta_1)}\frac{\widehat{f}_{r+H}(\beta_2)}{\widehat{f}_H(\beta_2)} = C_{r,\beta}$$

for every pair $\beta_1 + \beta_2 = \beta$ with $\widehat{f}_H(\beta_1), \widehat{f}_H(\beta_2) \neq 0$. If $\widehat{f}_H(\beta_i) = 0$ for some i, then by construction $\widehat{\phi^r_{a+H}}(\beta_i) = 0$, so the corresponding summand in S_r is zero. Hence

$$S_r = C_{r,\beta} \sum_{\beta_1+\beta_2=\beta} \widehat{f}_{a+r+H}(\beta_1)\widehat{f}_{a+r+H}(\beta_2).$$

Since f_{a+r+H} is Boolean, the Titsworth relation gives

$$\sum_{\beta_1+\beta_2=\beta} \widehat{f}_{a+r+H}(\beta_1)\widehat{f}_{a+r+H}(\beta_2) = 0,$$

and hence $S_r = 0$.

Since E_2 does not occur,

$$\widehat{\phi^r_{a+H}}(\gamma) = \widehat{\phi_{a+H}}(\gamma) \qquad \forall\, \gamma,$$

so $S_r = S$. Therefore $S = 0$.

Since $\beta \neq 0$ was arbitrary,

$$\sum_{\beta_1+\beta_2=\beta} \widehat{\phi_{a+H}}(\beta_1)\widehat{\phi_{a+H}}(\beta_2) = 0 \qquad \forall\, \beta \neq 0.$$

It remains to verify Parseval's identity. By Claim 7.8,

$$\left|\widehat{\phi_{a+H}}(\gamma)\right| = \left|\widehat{\phi_H}(\gamma)\right| = \left|\widehat{f_H}(\gamma)\right| \qquad \forall\, \gamma \in W.$$

Therefore,

$$\sum_{\gamma} \widehat{\phi_{a+H}}(\gamma)^2 = \sum_{\gamma} \widehat{f_H}(\gamma)^2 = 1,$$

where the last equality follows from Parseval's identity for the Boolean function f_{a+H}.

Since ϕ_{a+H} satisfies both Parseval's identity and the Titsworth relations, it follows that ϕ_{a+H} is Boolean. □

We now complete the proof of soundness of Algorithm 12. By Claim 7.9, the function ϕ is s-Fourier sparse, and by Claim 7.10, every restriction ϕ_{a+H} is Boolean-valued; since this holds for every coset $a+H$, ϕ itself is Boolean-valued on all of $\mathbb{F}_2^n$. Hence $\phi \in \mathcal{F}_s$. Since f is ε-far from every s-Fourier sparse Boolean function, it follows that $\mathrm{dist}(f, \phi) \geq \varepsilon$. On the other hand, Claim 7.4 gives $\mathrm{dist}(f, \phi) \leq 2\rho$. Combining these two bounds, $\varepsilon \leq \mathrm{dist}(f, \phi) \leq 2\rho$. Recall that we are operating under the assumption $\eta < 1/64$, for which we fixed a choice of H satisfying $\rho \leq 8\eta$ (an event of probability at least $7/8$ over H). Substituting this bound, $\varepsilon \leq 2\rho \leq 16\eta$, and therefore $\eta \geq \varepsilon/16$. This holds whenever $\eta < 1/64$; combined with the trivial bound $\eta \geq 1/64$ in the complementary case, we conclude that in either case,

$$\eta \geq \min\left(\frac{\varepsilon}{16}, \frac{1}{64}\right),$$

which is precisely the soundness guarantee of Algorithm 12.

7.2.3 Query Complexity

Algorithm 12 exactly computes the complete Fourier spectrum of each of the four affine restrictions by evaluating their entire truth tables. Since every restriction is defined over a subspace of size $\Theta(s^4/\varepsilon)$, a straightforward approach requires $O(s^4/\varepsilon)$ oracle queries.

7.3 Proof of Theorem 7.1

Computing the Fourier spectrum of a Boolean function by directly applying the Walsh–Hadamard transform requires access to its entire truth table, resulting in exponential query complexity. Consequently, a straightforward implementation of Algorithm 12 is prohibitively expensive. In this section, we show how to circumvent this bottleneck by recovering the Fourier spectrum of each affine restriction using only a small number of oracle queries. This yields a query-efficient implementation of Algorithm 12 while preserving its completeness and soundness guarantees up to asymptotically negligible losses.

7.3.1 Query-Efficient Fourier Sparse Recovery

Our starting point is the observation that, for every affine restriction

$$f_{a+H} : \mathbb{F}_2^d \to \mathbb{R},$$

its Fourier spectrum can be viewed as a vector in $\mathbb{R}^{2^d}$ indexed by the Walsh–Hadamard characters. Evaluating f_{a+H} at a point $x \in \mathbb{F}_2^d$ is equivalent to taking the inner product of this Fourier coefficient vector with the row of the Walsh–Hadamard matrix indexed by x. Consequently, oracle access to f_{a+H} provides linear measurements of its Fourier spectrum, reducing Fourier recovery to a sparse linear inverse problem.

A fundamental result in compressed sensing, due to Candès, Tao, Donoho, and others [Candès et al., 2006, Donoho, 2006], states that sparse vectors can be recovered exactly from significantly fewer linear measurements than their ambient dimension, provided the measurement matrix satisfies the Restricted Isometry Property (RIP). In particular (see Chapter 4 of Moitra [2018]), suppose that a measurement matrix $M \in \mathbb{R}^{m \times 2^d}$ satisfies the RIP of order k with isometry constant $\delta_k < 1/3$ (more generally, $\delta_{2k} + \delta_{3k} < 1$ suffices). Then every k-sparse vector z can be recovered exactly from measurements $y = Mz$ by

solving the Basis Pursuit (ℓ_1-minimization) problem

$$\widehat{z} = \arg\min_{u} \|u\|_1 \qquad \text{subject to} \qquad Mu = y.$$

Thus, the query complexity of sparse recovery is determined by the number of sampled measurements required for the measurement matrix to satisfy the RIP. The following theorem of Haviv and Regev shows that a uniformly subsampled Walsh–Hadamard matrix satisfies the RIP using only a near-linear number of measurements.

Lemma 7.11 (Ishay and Regev [2017]). *Let $M \in \mathbb{C}^{N\times N}$ be a unitary matrix satisfying $\|M\|_\infty \le O(1/\sqrt{N})$, and let $\delta > 0$ be sufficiently small. Construct a matrix $B \in \mathbb{C}^{q\times N}$ by selecting q rows independently and uniformly at random from M, each scaled by $\sqrt{N/q}$. If*

$$q = O\left(\log^2\frac{1}{\delta}\,\delta^{-2}\,k\,\log^2\frac{k}{\delta}\,\log N\right),$$

then, with probability at least

$$1 - 2^{-\Omega(\log N\cdot\log(k/\delta))},$$

the matrix B satisfies the Restricted Isometry Property of order k with isometry constant δ.

Since the RIP constant δ is a fixed absolute constant throughout this paper, the required number of sampled rows simplifies to

$$q = O\big(k\log^2 k\,\log N\big).$$

Motivated by Lemma 7.11, we obtain the sparse Fourier recovery procedure described in Algorithm 13. The algorithm first constructs a random subsampled Walsh–Hadamard measurement matrix, recovers the Fourier spectrum via Basis Pursuit, and then verifies that the recovered function is Boolean and agrees with the oracle on sufficiently many random samples. The verification step guarantees that any incorrect recovery is rejected with high probability while preserving the query efficiency of the overall tester.

7.3.1.1 Analysis of Algorithm 13

Lemma 7.12. *Let $g : \mathbb{F}_2^d \to \{-1,+1\}$ be given by oracle access. Then, with probability at least $1 - o(1/k) - o(\delta^2)$, Algorithm 13 satisfies the following.*

1. *If g is k-Fourier sparse, then the algorithm accepts and returns $\widetilde{g} = g$.*

Algorithm 13: Sparse Recovery on an Affine Restriction

Input: Oracle access to $g : \mathbb{F}_2^d \to \mathbb{R}$, sparsity parameter k, and proximity parameter δ.
Output: A recovered Boolean function $\widetilde{g} : \mathbb{F}_2^d \to \{-1, 1\}$ or Reject.
Initialization:

$$m_1 = O\big(k \log^2 k \, \log d\big), \qquad m_2 = \Theta\left(\frac{1}{\delta} \log \frac{1}{\delta}\right).$$

Step 1. Independently sample m_1 uniformly random points $x_1, \ldots, x_{m_1} \in \mathbb{F}_2^d$.;
Step 2. Construct the measurement matrix $A \in \{-1, 1\}^{m_1 \times 2^d}$ defined by

$$A_{i,\alpha} = \chi_\alpha(x_i),$$

and form the measurement vector

$$y = (g(x_1), \ldots, g(x_{m_1}))^T.$$

;
Step 3. Compute

$$\widehat{z} = \arg\min_z \|z\|_1 \quad \text{subject to} \quad Az = y.$$

;
Step 4. Recover the candidate function

$$\widetilde{g}(x) = \sum_{\alpha \in \mathbb{F}_2^d} \widehat{z}_\alpha \chi_\alpha(x).$$

;
Step 5. Verify that $\widetilde{g}$ is Boolean-valued and k-Fourier sparse. If not, Reject.;
Step 6. Independently sample m_2 uniformly random points from $\mathbb{F}_2^d$ and verify that $\widetilde{g}$ agrees with g on every sampled point. If any check fails, Reject.;
Step 7. Return $\widetilde{g}$.;

2. *If g is δ-far from every k-Fourier sparse Boolean function, then the algorithm rejects.*

3. *If g is δ-close to some k-Fourier sparse Boolean function, then the algorithm either rejects or returns a k-Fourier sparse Boolean function $\widetilde{g}$ satisfying $\operatorname{dist}(g, \widetilde{g}) \leq \delta$.*

Moreover, the algorithm makes

$$\widetilde{O}\left(k \log^2 k \log \frac{k}{\delta} + \frac{kd}{\delta}\right)$$

oracle queries.

Proof.

The algorithm consists of two phases: sparse recovery and verification.

Let

$$E_{\mathrm{RIP}} = \{\text{the sampled Walsh–Hadamard matrix does not satisfy the RIP of order } k\},$$

and let

$$E_{\mathrm{ver}} = \left\{ \begin{array}{c} \text{there exists a recovered } k\text{-Fourier sparse Boolean function} \\ \widetilde{g} \text{ with } \operatorname{dist}(g, \widetilde{g}) \geq \delta \text{ that nevertheless passes} \\ \text{all verification samples} \end{array} \right\}.$$

By Lemma 7.11, choosing $m_1 = Ck \log^2 k \log \frac{k}{\delta}$ for a sufficiently large constant C ensures $\Pr(E_{\mathrm{RIP}}) = o(1/k)$. It remains to bound the probability of E_{ver}. There are at most

$$2^k \binom{2^d}{k} \leq 2^{O(kd)}$$

distinct k-Fourier sparse Boolean functions over $\mathbb{F}_2^d$. Suppose $\widetilde{g}$ is one such function satisfying $\operatorname{dist}(g, \widetilde{g}) \geq \delta$. Then each uniformly random verification sample agrees with probability at most $1 - \delta$, and hence

$$\Pr[\widetilde{g} \text{ passes all } m_2 \text{ samples}] \leq (1 - \delta)^{m_2} \leq e^{-\delta m_2}.$$

Applying the union bound over all possible recovered functions,

$$\Pr(E_{\mathrm{ver}}) \leq 2^{O(kd)} e^{-\delta m_2}.$$

Choosing

$$m_2 = C' \frac{kd + \log \frac{1}{\delta}}{\delta} = O\left(\frac{kd}{\delta} \log \frac{1}{\delta}\right),$$

for a sufficiently large constant C' yields $\Pr(E_{\text{ver}}) = o(\delta^2)$.

Now define the good event

$$G = \overline{E_{\text{RIP}} \cup E_{\text{ver}}}.$$

By the union bound, $\Pr(\overline{G}) \leq \Pr(E_{\text{RIP}}) + \Pr(E_{\text{ver}}) = o(1/k) + o(\delta^2)$. Conditioned on G, the RIP guarantees exact recovery whenever g is k-Fourier sparse. Furthermore, every recovered function that survives the verification phase must be within Hamming distance at most δ of g. This immediately establishes all three claims.

Finally, the Booleanity test performed by the algorithm is purely computational and requires no oracle queries. Therefore the total number of queries equals

$$m_1 + m_2 = O\left(k \log^2 k \log \frac{k}{\delta} + \frac{kd}{\delta} \log \frac{1}{\delta}\right),$$

completing the proof. □

7.3.2 Proof of Main Theorem (Theorem 7.1)

To prove the theorem, we invoke Algorithm 13 as a subroutine within Algorithm 12, setting the parameters $k = s$ and $\delta = \varepsilon/1024$. Instead of computing the Fourier spectrum of each affine restriction explicitly, Algorithm 12 calls Algorithm 13 to recover it using only a small number of oracle queries. We show that this modification preserves the completeness and soundness guarantees of Algorithm 12, up to asymptotically negligible losses.

7.3.2.1 Completeness

Completeness of Main Theorem

If $f : \mathbb{F}_2^n \to \{-1, +1\}$ is s-Fourier sparse, then the modified Algorithm 12 accepts f with probability at least $1 - \varepsilon/200$.

Assume that $f : \mathbb{F}_2^n \to \{-1, +1\}$ is s-Fourier sparse. Since restricting a Boolean function

to an affine subspace cannot increase its Fourier sparsity, every affine restriction f_{a+H} is also s-Fourier sparse. Algorithm 12 invokes Algorithm 13 on each affine restriction with parameters $k = s$ and $\delta = \varepsilon/1024$. By Lemma 7.12, except with probability $o(\varepsilon^2) + o(1/s)$, Algorithm 13 exactly recovers every restricted function. Hence, $\widetilde{f}_{a+H} = f_{a+H}$ for every affine restriction considered by the tester.

Conditioned on this event, the modified tester behaves identically to Algorithm 12. Therefore, the only additional source of error compared to the original algorithm is the failure probability of the sparse recovery procedure. Since the original tester rejects an s-Fourier sparse function with probability at most $\varepsilon/256$, we conclude that the modified tester rejects with probability at most $\varepsilon/256 + o(\varepsilon^2) + o(1/s)$. For sufficiently small ε and sufficiently large s, the overall rejection probability is at most $\varepsilon/200$.

7.3.2.2 Soundness

Soundness of Main Theorem

If $f : \mathbb{F}_2^n \to \{-1, +1\}$ is ε-far from every s-Fourier sparse Boolean function, then the modified Algorithm 12 rejects f with probability at least $\varepsilon/100$.

Assume that $\mathrm{dist}(f, \mathcal{F}_s) \geq \varepsilon$, where $\mathcal{F}_s$ denotes the family of s-Fourier sparse Boolean functions. Algorithm 12 invokes Algorithm 13 on each affine restriction with parameters $k = s$ and $\delta = \varepsilon/1024$. By Lemma 7.12, each invocation succeeds except with probability at most $\tau = o(\varepsilon^2) + o(1/s)$.

Fix all randomness used by the sparse recovery procedures. Conditioned on this randomness, every recovered affine restriction is deterministic. We define a Boolean function $\widetilde{f} : \mathbb{F}_2^n \to \{-1, +1\}$ by letting its restriction on each affine coset equal the function returned by Algorithm 13 whenever the recovery procedure accepts, and defining it arbitrarily on cosets where the recovery procedure rejects. Observe that, once the recovery randomness is fixed, every subsequent query made by Algorithm 12 is answered exactly as if the algorithm had oracle access to the single Boolean function $\widetilde{f}$. Consequently, conditioned on the recovery randomness, the remainder of the execution is identical to running Algorithm 12 on $\widetilde{f}$. Let ρ denote the rejection probability of Algorithm 12 on $\widetilde{f}$ over its remaining randomness.

The following lemma bounds the distance between f and $\widetilde{f}$.

Lemma 7.13. *Suppose that for every affine coset $a + H$, Algorithm 13 accepts with probability at least $1 - \tau$. Moreover, whenever it accepts, the recovered restriction $\widetilde{f}_{a+H}$ satisfies*

$\mathrm{dist}(f_{a+H}, \widetilde{f}_{a+H}) \le \delta$. *Then*

$$\mathrm{dist}(f, \widetilde{f}) \le \delta + \tau.$$

Proof. For each affine coset $a + H$, let $X(a) = \mathrm{dist}(f_{a+H}, \widetilde{f}_{a+H})$. Since the affine cosets of H partition $\mathbb{F}_2^n$,

$$\mathrm{dist}(f, \widetilde{f}) = \mathbb{E}_a[X(a)].$$

Let $B = \{a+H : \text{Algorithm 13 rejects on } a+H\}$. Then $\Pr[a+H \in B] \le \tau$. If $a+H \notin B$, then $X(a) \le \delta$. Otherwise, we simply use the trivial bound $X(a) \le 1$. Therefore,

$$\begin{aligned}\mathrm{dist}(f, \widetilde{f}) &= \mathbb{E}[X(a)] \\ &\le (1-\tau)\delta + \tau \\ &= \delta + \tau(1-\delta) \\ &\le \delta + \tau,\end{aligned}$$

which completes the proof. □

We first show that $\widetilde{f}$ is close to $\mathcal{F}_s$. Suppose that Algorithm 12, when run on $\widetilde{f}$, rejects with probability $\rho < 1/64$. We claim that there exists $\phi \in \mathcal{F}_s$ such that $\mathrm{dist}(\widetilde{f}, \phi) \le 16\rho$.

Indeed, suppose, towards a contradiction, that $\mathrm{dist}(\widetilde{f}, \mathcal{F}_s) > 16\rho$. Choose any θ satisfying $16\rho < \theta < \mathrm{dist}(\widetilde{f}, \mathcal{F}_s)$. Then, Algorithm 12 rejects $\widetilde{f}$ with probability at least $\min(\theta/16, 1/64)$. Since $\rho < 1/64$ and $\theta > 16\rho$, we have $\theta/16 > \rho$, and therefore $\min(\theta/16, 1/64) > \rho$, contradicting the assumption that the rejection probability is ρ. Hence there exists $\phi \in \mathcal{F}_s$ such that $\mathrm{dist}(\widetilde{f}, \phi) \le 16\rho$.

We now complete the proof of soundness by considering two cases.

Case 1: $\rho \ge 1/64$**.** Since $\varepsilon \in (0, 1)$, we have $\rho \ge 1/64 > 1/100 \ge \varepsilon/100$, and therefore the desired bound $\rho \ge \varepsilon/100$ holds immediately.

Case 2: $\rho < 1/64$**.** There exists $\phi \in \mathcal{F}_s$ with $\mathrm{dist}(\widetilde{f}, \phi) \le 16\rho$. Combining this with Lemma 7.13 yields $\mathrm{dist}(f, \phi) \le \mathrm{dist}(f, \widetilde{f}) + \mathrm{dist}(\widetilde{f}, \phi) \le \delta + \tau + 16\rho$. Since $\mathrm{dist}(f, \phi) \ge \varepsilon$ (as $\phi \in \mathcal{F}_s$ and f is ε-far from $\mathcal{F}_s$), it follows that $16\rho + \delta + \tau \ge \varepsilon$. Substituting $\delta = \varepsilon/1024$ and $\tau = o(\varepsilon^2) + o(1/s)$, we obtain $16\rho \ge \varepsilon - \varepsilon/1024 - \tau = (1023/1024)\varepsilon - \tau$, and hence $\rho \ge \big((1023/1024)\varepsilon - \tau\big)/16 = (1023/16384)\varepsilon - \tau/16$. Since $\tau = o(\varepsilon)$, for sufficiently small ε and sufficiently large s, we have $\rho \ge \varepsilon/100$.

8

Application VI : Near Optimal ℓ_2 Testing of Fourier Sparsity

Sparsity is one of the most powerful ideas connecting modern machine learning and theoretical computer science. It captures the intuition that, even in high-dimensional settings, many natural functions or signals depend only on a small number of relevant components. This assumption underlies algorithms that are both sample- and time-efficient, forming the basis of techniques such as sparse linear regression, decision tree learning, and compressed sensing. Across these frameworks, the common principle is simple yet profound: if a function admits a sparse representation in a suitable basis, such as Fourier, wavelet, polynomial, or a learned dictionary, then learning and inference can be made dramatically more efficient.

A particularly elegant setting where sparsity plays a central role is that of real-valued functions over the Boolean hypercube $\mathbb{F}_2^n$. Fourier analysis provides a natural orthonormal basis for such functions. Every function $f : \mathbb{F}_2^n \to \mathbb{R}$ can be expressed as

$$f(x) = \sum_{\alpha \in \mathbb{F}_2^n} \widehat{f}(\alpha)\,(-1)^{\langle x,\alpha\rangle},$$

where $\widehat{f}(\alpha)$ denotes the Fourier coefficient at frequency α. The *Fourier sparsity* of f, the number of nonzero coefficients in this expansion quantifies how succinctly the function can be represented in the Fourier basis.

Fourier sparsity is a recurring theme across many areas of theoretical computer science. In complexity theory, it lies at the heart of problems involving communication complexity and parity decision trees for Boolean functions $f : \mathbb{F}_2^n \to \{+1, -1\}$ [Zhang and Shi, 2009, 2010, Tsang et al., 2013, Shpilka et al., 2017, Chakraborty et al., 2020]. In learning theory, it has become a central tool for designing algorithms that efficiently learn functions with low-degree or low-support Fourier spectra. Many natural Boolean functions exhibit this property: graph and hypergraph cut functions, as well as decision trees of bounded depth, are inherently Fourier sparse because their spectra are concentrated on low-degree

coefficients [Stobbe and Krause, 2012, Linial et al., 1993]. For example, the cut function of a graph corresponds to a degree-2 polynomial in the Fourier basis, while a degree-d hypergraph cut function has degree at most d. Similarly, a Boolean decision tree of depth d has its spectrum supported entirely on coefficients of degree at most d.

Beyond these classical examples, Fourier-sparse models have found renewed relevance in modern machine learning. They appear in settings such as neural network hyperparameter optimization [Hazan et al., 2018] and the learning of structured set functions [Amrollahi et al., 2019]. The impact of Fourier sparsity extends even further, into cryptography, where the celebrated Goldreich–Levin theorem Goldreich and Levin [1989] established a deep connection between identifying large Fourier coefficients and constructing hardcore predicates for one-way functions.

Algorithmically, two main approaches have emerged for learning or recovering Fourier-sparse functions: *Sparse Hadamard Transform* methods [Hassanieh et al., 2012b, Scheibler et al., 2015, Indyk et al., 2014] and *Compressed Sensing* techniques [Rudelson and Vershynin, 2008, Ishay and Regev, 2017]. Both families of algorithms, however, rely critically on prior knowledge of the function's sparsity level. This partcular gap motivates our work, which focuses on efficiently estimating the Fourier sparsity, up to a desired approximation in ℓ_2 distance. Such an estimator can serve as a useful preprocessing step in learning pipelines, both for verifying whether sparsity-based assumptions hold and for guiding the initialization of sparsity parameters in downstream algorithms.

To formalize this setting, we first introduce some basic definitions. We measure distances between functions using the squared ℓ_2-norm:

$$\mathrm{dist}_2^2(f,g) := \|f-g\|_2^2 = \frac{1}{2^n}\sum_{x\in\mathbb{F}_2^n}(f(x)-g(x))^2.$$

For a function f and a class of functions $\mathcal{P}$, the distance of f from $\mathcal{P}$ is defined as

$$\mathrm{dist}_2^2(f,\mathcal{P}) := \min_{g\in\mathcal{P}} \|f-g\|_2^2.$$

We also recall the standard definition of the ℓ_2-norm. For any function $f:\mathbb{F}_2^n\to\mathbb{R}$,

$$\|f\|_2^2 := \frac{1}{2^n}\sum_{x\in\mathbb{F}_2^n} f(x)^2.$$

Let $\mathcal{F}_s$ denote the class of all *s-Fourier sparse functions*, that is, functions $f:\mathbb{F}_2^n\to\mathbb{R}$ whose Fourier spectrum has at most s nonzero coefficients. We are interested in determining how close a given function f is to this class.

Problem 8.1. *Given query access to a function $f : \mathbb{F}_2^n \to \mathbb{R}$ with $\|f\|_2^2 = 1$, and parameters $s > 0$, $\epsilon \in (0, 1]$, and $\delta \in [0, 1]$, the task is to design a randomized algorithm that distinguishes between the following two cases:*

- ***(Close)**: There exists $g \in \mathcal{F}_s$ such that $\|f - g\|_2^2 \leq \delta$.*
- ***(Far)**: For every $g \in \mathcal{F}_s$, $\|f - g\|_2^2 \geq \delta + \epsilon$.*

The goal is to construct such an algorithm using as few queries to f as possible, while ensuring that it distinguishes the two cases with high probability.

Our main contribution in this paper is the design of a simple, nonadaptive and almost optimal query algorithm for testing Fourier sparsity.

Theorem 8.2. *Let $s > 0$, $\epsilon \in (0, 1]$, and $\delta \in [0, 1]$. Let $f : \mathbb{F}_2^n \to \mathbb{R}$ be an unknown function with $\|f\|_2 = 1$, accessible only via query access. Then, there exists a* nonadaptive *algorithm that, with success probability at least $2/3$, distinguishes between the following two cases:*

- *f is δ-close to being s-Fourier sparse,*
- *f is $(\delta + \epsilon)$-far from every s-Fourier sparse function,*

using at most $\widetilde{O}(s/\epsilon^2)$ queries to f, where the $\widetilde{O}(\cdot)$ notation hides factors polynomial in $\log s$ and $\log(1/\epsilon)$.

Theorem 8.2 is proved in Section 8.2. Although stated under the assumption that the function has unit ℓ_2-norm, the result extends to any nonzero function $f : \mathbb{F}_2^n \to \mathbb{R}$ via normalization.

Related works

Testing Fourier sparsity was first studied by Gopalan et al. [2011], who focused on Boolean functions and used the *Hamming distance* as a measure of proximity. (The Hamming distance between two functions $f, g : \mathbb{F}_2^n \to \{0, 1\}$ is the fraction of inputs on which they differ.) Their algorithm has query complexity $O(s^{14})$, which quickly becomes impractical for even moderately large values of s. Similarly, the regularity framework of Hatami and Lovett [2013] provides a general-purpose, black-box reduction for testing Fourier sparsity under Hamming distance, but this approach suffers from a query complexity that grows as a tower function in s.

In the real-valued setting, Yaroslavtsev and Zhou [2020] considered testing Fourier sparsity with respect to the squared ℓ_2-distance. They designed an algorithm with query complexity $\widetilde{O}(s/\epsilon^4)$ and established a lower bound of $\Omega(\sqrt{s})$. In comparison, our algorithm improves the dependence on the proximity parameter ϵ, requiring only $\widetilde{O}(s/\epsilon^2)$ queries, and is conceptually simpler. We further establish a nearly tight lower bound of $\Omega(s)$, quadratically improving the current state of art [Yaroslavtsev and Zhou, 2020], showing that our algorithm is optimal up to logarithmic factors.

It is important to note that testing Fourier sparsity in the *random example* model is significantly more challenging. As shown in Goldreich and Ron [2016], even for linearity testing, where the target functions are 1-Fourier sparse, it is not known how to design a tester whose sample complexity is independent of the ambient dimension n. In contrast, in the *query access* model, linearity testing can be performed efficiently using the well-known 3-query BLR test. Our work aims to estimate the Fourier sparsity level of a function in a way that depends only on the sparsity s and the proximity parameter ϵ, while remaining independent of the ambient dimension n.

8.1 Proof Ideas

The design of our tester is inspired by the well-established framework for testing hereditary graph properties. A canonical tester for such a property P samples a small random subset of vertices, queries all induced edges, and checks whether the resulting subgraph satisfies P. For hereditary properties, those preserved under taking induced subgraphs, this approach guarantees only a modest (quadratic) overhead in query complexity [Alon et al., 2009, Goldreich and Trevisan, 2003, Alon, 2002]. Indeed, for several natural properties, such as bipartiteness, the canonical tester achieves optimal performance up to constant factors.

A similar idea has been successfully adapted to Boolean functions, particularly for testing affine-invariant properties Bhattacharyya et al. [2015], Arnab et al. [2010]. In this setting, the canonical approach restricts the function to random low-dimensional affine subspaces and tests the property on these restrictions. While this strategy enjoys strong generality and theoretical support, e.g., via regularity-like lemmas, it often suffers from impractical query complexity, including tower-type dependencies [Hatami and Lovett, 2013]. Nonetheless, specialized testers exploiting finer structural properties have been developed for specific cases, such as low algebraic degree [Noga et al., 2003] and odd-cycle-freeness [Bhattacharyya et al., 2012]. Surprisingly, despite being a natural and central affine-invariant property, Fourier sparsity has largely resisted similar progress.

Prior works in Fourier sparsity testing typically project the Fourier spectrum into ran-

domly chosen cosets of sufficiently large codimension, a process commonly referred to as *Fourier hashing*, which was originally introduced in [Feldman et al., 2006b]. Analytical or combinatorial tools are then applied to extract sparsity information. For example, Gopalan et al. [2011] presents a granularity theorem for Fourier-sparse functions, showing that individual coefficients cannot be too small and reducing the problem to counting large-weight cosets. Similarly, Yaroslavtsev and Zhou [2020] certain concentration of the ℓ_2-norm of heavy buckets to design their tester.

In contrast, our approach analyzes the function restricted to a randomly chosen subspace. We approximately recover the Fourier spectrum of this restricted function and use it to infer the sparsity of the original function. A new structural relationship between the Fourier coefficients of the restricted and original functions shows that, under suitable subspace choices, their magnitudes closely match. This relationship is central to our analysis. Instead of explicitly defining a hashing process, restricting the function to a subspace implicitly induces a hashing, allowing us to derive a concentration bound in terms of the ℓ_1-norm of bucketed Fourier coefficients, which constitutes our main technical contribution.

8.2 Our Algorithm

Suppose we are given oracle access to a function $f : \mathbb{F}_2^n \to \mathbb{R}$, normalized so that $\|f\|_2 = 1$. By Parseval's identity its Fourier coefficients satisfy $\sum_\beta \hat{f}(\beta)^2 = 1$; the quantity $\hat{f}(\beta)^2$ is called the *energy* of the coefficient $\hat{f}(\beta)$, and the total energy is 1. We ask: *how much of that energy is concentrated on the s heaviest coefficients?* Formally, we wish to estimate

$$W_s(\hat{f}) \;:=\; \max_{\substack{S\subseteq\mathbb{F}_2^n \\ |S|=s}} \sum_{\beta\in S} \hat{f}(\beta)^2$$

to within an additive $\pm\tau$. For $s = 1$ this is the Fourier bias problem; for general s it measures how close f is, in ℓ_2, to a Fourier s-sparse function, since the distance to the nearest s-sparse function is exactly $1 - W_s(\hat{f})$.

The problem

Problem 8.3 (Energy of the top s Fourier coefficients)**.** *Given oracle access to $f : \mathbb{F}_2^n \to \mathbb{R}$ with $\|f\|_2 = 1$, an integer $s \geq 1$ and an accuracy parameter $\tau \in (0, 1)$, output an estimate*

$\widetilde{W}$ of

$$W_s(\hat{f}) := \max_{\substack{S\subseteq\mathbb{F}_2^n \\ |S|=s}} \sum_{\beta\in S} \hat{f}(\beta)^2$$

satisfying $\left|\widetilde{W} - W_s(\hat{f})\right| \le \tau$.

By Parseval, $0 \le W_s(\hat{f}) \le 1$. For $s = 1$ we recover the Fourier bias, $W_1(\hat{f}) = \max_\beta \hat{f}(\beta)^2$. More generally, if $\mathcal{F}_s$ denotes the class of Fourier s-sparse functions, then $\min_{g\in\mathcal{F}_s} \|f - g\|_2^2 = 1 - W_s(\hat{f})$, so Problem 8.3 is exactly the problem of estimating the ℓ_2^2-distance from f to s-sparsity.

For a function g on a group whose characters are indexed by a set Γ, we write more generally

$$W_s(\hat{g}) := \max_{\substack{S\subseteq\Gamma \\ |S|=s}} \sum_{\gamma\in S} \hat{g}(\gamma)^2.$$

The obvious obstacle is that the spectrum has 2^n coefficients, and we may not touch more than a handful of points of the domain. The strategy is dimension reduction.

The three ideas. The whole argument is built from three moves, and it is worth naming them before we begin.

(I) Random affine restriction as spectral hashing. Restrict f to a random affine subspace $A = \alpha + H$ of dimension $t \ll n$. The restricted function f_A lives on a group of size 2^t, so it has only 2^t Fourier coefficients — but each one is a *signed aggregate* of the coefficients of f lying in one coset of $H^\perp$ (Fact 6.1). Restriction thus collapses the spectrum, and the coset partition is pairwise independent (Fact 6.2).

(II) Heavy coefficients survive; the light spectrum is controlled by its fourth moment. A heavy coefficient sitting alone in its coset is read off essentially unchanged (Lemma 6.3). The danger is that *many small* coefficients could pile up in a single coset and masquerade as a heavy one. To rule this out we do not track them individually: we control the fourth Fourier mass $\sum_\gamma \widehat{f_A}(\gamma)^4$ in aggregate, and show it is preserved under restriction (Section 6.2). This is where the eighth moment $\|f\|_8$ enters, and it is the only place it enters.

(III) Cauchy–Schwarz to pass from $s = 1$ to general s. For $s = 1$ one argues pointwise: no light coset exceeds τ. For general s this is the wrong currency, because s light cosets could each be moderately large. Instead we bound the *aggregate*

contribution of any s light cosets by

$$\sum_{\gamma\in T}\widehat{f_A}(\gamma)^2 \;\le\; \sqrt{|T|}\Big(\sum_{\gamma\in T}\widehat{f_A}(\gamma)^4\Big)^{1/2},$$

which converts our fourth-moment control directly into an energy budget. This single application of Cauchy–Schwarz is the whole content of the generalization.

Two places where the naive argument fails. Both are worth dwelling on in a seminar.

First, the classical variance computation for the fourth mass — expand in the Fourier basis, insert the coset-collision probability into the resulting sum — is *not* valid, because that sum is signed and one cannot insert a worst-case probability term by term. Repairing it forces absolute values on the coefficients and destroys the bound even for Boolean f. We give a different proof (Section 6.2.3) based on the observation that a random affine subspace supplies *exactly* independent uniform sample points, except on a rare degenerate event.

Second, because f is unbounded, Hoeffding's inequality is simply unavailable when we come to estimate coefficients from samples. We use median-of-means instead, with the variance supplied by $\|f\|_2 = 1$ (Lemma 8.6). Median-of-means is a legitimate tool here precisely because we amplify an estimator that is already correct with constant probability — a point on which it is easy to slip.

8.2.1 Estimating the energy of the top s coefficients

The structural lemma

Everything hinges on the assertion that the top-s energy of the *restricted* function is a faithful proxy for that of f. We now fix the parameters once and for all, in the order forced by the argument.

Lemma 8.4 (Top-s energy is preserved under restriction)**.** *Let $f:\mathbb{F}_2^n\to\mathbb{R}$ with $\|f\|_2=1$, let $s\ge 1$ be an integer, $\tau\in(0,1)$, and $K:=\|f\|_8^8$. Let $A=\alpha+H$ be a random affine subspace of dimension*

$$t \;\ge\; \log_2\!\big(2^{55}\,K\,s^5\,\tau^{-10}\big),$$

and let W be complementary to $H^\perp$. Then with probability at least $7/8$,

$$\Big|\, W_s\big(\widehat{f_A}\big) \;-\; W_s\big(\hat{f}\big)\,\Big| \;\le\; \frac{\tau}{2}.$$

Proof. Parameters. Define once and for all

$$\theta := \frac{\tau}{16\sqrt{s}} \quad \text{(heavy threshold)}, \qquad \lambda := \frac{\tau^4}{2^{21}s^2} \quad \text{(coefficient error)},$$

$$\varepsilon := \frac{\tau^2}{192s} \quad \text{(fourth-mass error)}.$$

Note $\theta, \lambda, \varepsilon \in (0,1)$ and $\lambda \le \theta$. Split the spectrum as

$$L := \{\beta : \ |\hat{f}(\beta)| \ge \theta\}, \qquad S := \mathbb{F}_2^n \setminus L.$$

By Parseval, $\theta^2|L| \le \sum_{\beta\in L} \hat{f}(\beta)^2 \le 1$, so

$$|L| \ \le \ \theta^{-2} \ = \ \frac{256\, s}{\tau^2}. \tag{8.1}$$

Algorithm Top-s-Energy(s, τ, K)

Require: oracle access to $f : \mathbb{F}_2^n \to \mathbb{R}$ with $\|f\|_2 = 1$; integer $s \ge 1$; accuracy $\tau \in (0,1)$; a bound $K \ge \|f\|_8^8$.

1. Set $t \leftarrow \left\lceil \log_2\!\left(2^{55}\, K\, s^5\, \tau^{-10}\right) \right\rceil$.
2. Draw a random subspace $H \le \mathbb{F}_2^n$ of dimension t and $\alpha \sim \mathrm{Unif}(\mathbb{F}_2^n)$; set $A \leftarrow \alpha + H$.
3. Using Lemma 8.6, compute estimates $\widetilde{f_A}(\gamma)$ of $\widehat{f_A}(\gamma)$ for all $\gamma \in W$, simultaneously accurate to $\eta := \tau/(32\sqrt{s})$, with failure probability $\le 1/16$.
4. **return** $\widetilde{W} := \max\limits_{\substack{S\subseteq W \\ |S|=s}} \sum\limits_{\gamma\in S} \widetilde{f_A}(\gamma)^2$ (i.e. the sum of the s largest values of $\widetilde{f_A}(\gamma)^2$).

The three failure events.

E_1 (*heavy collision*): two distinct elements of L share a coset of $H^\perp$. By Fact 6.2 and a union bound over at most $\binom{|L|}{2} \le \frac{1}{2}\theta^{-4}$ pairs,

$$\Pr[E_1] \ \le \ \tfrac{1}{2}\,\theta^{-4}\, 2^{-t} \ = \ 2^{15}\, s^2\, \tau^{-4} \cdot 2^{-t}.$$

On E_1^c the map $\gamma \mapsto \beta_\gamma$ (the representative in W of the coset containing γ) is *injective* on L; we set $W_L := \{\beta_\gamma : \gamma \in L\} \subseteq W$.

E_2 (*heavy distortion*): $\left|\widehat{f_A}(\beta_\gamma) - \hat{f}(\gamma)\chi_\gamma(\alpha)\right| > 2^{-t} + \lambda$ for some $\gamma \in L$. By Lemma 6.3, (8.1) and a union bound,

$$\Pr[E_2] \ \le \ \theta^{-2} \cdot \frac{2^{-t}}{\lambda^2} \ = \ \frac{256\, s}{\tau^2} \cdot \frac{2^{42} s^4}{\tau^8} \cdot 2^{-t} \ = \ 2^{50}\, s^5\, \tau^{-10} \cdot 2^{-t}.$$

E_3 (*fourth-mass distortion*): $\big|\|\widehat{f_A}\|_4^4 - \|\hat{f}\|_4^4\big| \geq \varepsilon$. By Lemma 6.12 and $\varepsilon^{-2} \leq 2^{16} s^2 \tau^{-4}$,

$$\Pr[E_3] \;\leq\; 504 \cdot 2^{-t} K \varepsilon^{-2} \;\leq\; 2^{25}\, K\, s^2\, \tau^{-4} \cdot 2^{-t}.$$

Each of the three bounds is at most $2^{50} K s^5 \tau^{-10} \cdot 2^{-t}$ (using $K \geq 1$, $s \geq 1$, $\tau < 1$), and our hypothesis $2^t \geq 2^{55} K s^5 \tau^{-10}$ makes each at most $2^{-5} = 1/32 \leq 1/24$. Hence

$$\Pr\big[E_1 \cup E_2 \cup E_3\big] \;\leq\; \tfrac{1}{8}.$$

Moreover $2^{-t} \leq \tau^{10}/(2^{55} s^5) \leq \lambda$. We work on $\mathcal{G} := E_1^c \cap E_2^c \cap E_3^c$ from now on, where consequently, for every $\gamma \in L$,

$$\Big|\, \big|\widehat{f_A}(\beta_\gamma)\big| - \big|\hat{f}(\gamma)\big| \,\Big| \;\leq\; \big|\widehat{f_A}(\beta_\gamma) - \hat{f}(\gamma)\chi_\gamma(\alpha)\big| \;\leq\; 2^{-t} + \lambda \;\leq\; 2\lambda, \tag{8.2}$$

by the reverse triangle inequality and $|\hat{f}(\gamma)\chi_\gamma(\alpha)| = |\hat{f}(\gamma)|$. In particular, since $|\hat{f}(\gamma)| \leq 1$ and $\lambda \leq \frac{1}{2}$,

$$\big|\widehat{f_A}(\beta_\gamma)\big| \;\leq\; 1 + 2\lambda \;\leq\; 2 \qquad \text{for all } \gamma \in L. \tag{8.3}$$

Step 1: the light spectrum carries little fourth mass, even after restriction. Since $|\hat{f}(\beta)| < \theta$ for $\beta \in S$, Parseval gives

$$\sum_{\beta \in S} \hat{f}(\beta)^4 \;\leq\; \theta^2 \sum_{\beta \in S} \hat{f}(\beta)^2 \;\leq\; \theta^2. \tag{8.4}$$

Subtracting the identities

$$\sum_{\gamma \notin W_L} \widehat{f_A}(\gamma)^4 = \big\|\widehat{f_A}\big\|_4^4 - \sum_{\gamma \in W_L} \widehat{f_A}(\gamma)^4, \qquad \sum_{\beta \in S} \hat{f}(\beta)^4 = \big\|\hat{f}\big\|_4^4 - \sum_{\gamma \in L} \hat{f}(\gamma)^4,$$

and applying the triangle inequality,

$$\sum_{\gamma \notin W_L} \widehat{f_A}(\gamma)^4 \;\leq\; \underbrace{\sum_{\beta \in S} \hat{f}(\beta)^4}_{\leq\, \theta^2} \;+\; \underbrace{\Big|\big\|\widehat{f_A}\big\|_4^4 - \big\|\hat{f}\big\|_4^4\Big|}_{\leq\, \varepsilon \text{ on } E_3^c} \;+\; \sum_{\gamma \in L} \Big|\widehat{f_A}(\beta_\gamma)^4 - \hat{f}(\gamma)^4\Big|.$$

For the last sum, use $|x^4 - y^4| = |x-y|\,(x+y)\,(x^2+y^2)$ for $x, y \geq 0$ with $x = |\widehat{f_A}(\beta_\gamma)| \leq 2$ and $y = |\hat{f}(\gamma)| \leq 1$: then $x + y \leq 3$ and $x^2 + y^2 \leq 5$, so by (8.2),

$$\Big|\widehat{f_A}(\beta_\gamma)^4 - \hat{f}(\gamma)^4\Big| \;\leq\; 15 \cdot 2\lambda \;=\; 30\lambda,$$

and summing over $|L| \le \theta^{-2} = 256 s\tau^{-2}$ heavy frequencies,

$$\sum_{\gamma\in L}\left|\widehat{f_A}(\beta_\gamma)^4 - \hat{f}(\gamma)^4\right| \;\le\; 30\lambda\,\theta^{-2} \;=\; 30\cdot\frac{\tau^4}{2^{21}s^2}\cdot\frac{256\,s}{\tau^2} \;=\; \frac{7680\,\tau^2}{2^{21}\,s} \;\le\; \frac{\tau^2}{192\,s}.$$

Collecting the three contributions and using $\theta^2 = \tau^2/(256s) \le \tau^2/(192s)$ and $\varepsilon = \tau^2/(192s)$,

$$\sum_{\gamma\notin W_L}\widehat{f_A}(\gamma)^4 \;\le\; 3\cdot\frac{\tau^2}{192\,s} \;=\; \frac{\tau^2}{64\,s}. \tag{8.5}$$

Now comes the move that generalizes the case $s = 1$. For *any* $T \subseteq W \setminus W_L$ with $|T| \le s$, the Cauchy–Schwarz inequality together with (8.5) gives

$$\sum_{\gamma\in T}\widehat{f_A}(\gamma)^2 \;\le\; \sqrt{|T|}\,\Big(\sum_{\gamma\in T}\widehat{f_A}(\gamma)^4\Big)^{1/2} \;\le\; \sqrt{s}\cdot\Big(\frac{\tau^2}{64\,s}\Big)^{1/2} \;=\; \frac{\tau}{8}. \tag{8.6}$$

Thus no collection of at most s light cosets can carry more than $\tau/8$ of energy. Note that we never bound an individual light coefficient; we bound them in aggregate, which is exactly what the top-s problem requires.

Step 2: $W_s(\widehat{f_A}) \le W_s(\hat{f}) + \tau/2$. Let $T \subseteq W$, $|T| = s$, be arbitrary and split $T = T_H \sqcup T_L$ with $T_H := T \cap W_L$, $T_L := T \setminus W_L$. By (8.6), $\sum_{\gamma\in T_L}\widehat{f_A}(\gamma)^2 \le \tau/8$.

For T_H: on E_1^c the elements of T_H correspond to $|T_H| \le s$ *distinct* heavy frequencies $\gamma_1, \dots, \gamma_{|T_H|} \in L$. Using $|x^2 - y^2| = |x - y|(x + y)$ with $x \le 2$, $y \le 1$, (8.2) gives

$$\widehat{f_A}(\beta_{\gamma_i})^2 \;\le\; \hat{f}(\gamma_i)^2 + 6\lambda,$$

so, since $|T_H| \le s$ and any s distinct frequencies have total energy at most $W_s(\hat{f})$,

$$\sum_{\gamma\in T_H}\widehat{f_A}(\gamma)^2 \;\le\; \sum_i \hat{f}(\gamma_i)^2 + 6s\lambda \;\le\; W_s\big(\hat{f}\big) + 6s\lambda \;\le\; W_s\big(\hat{f}\big) + \frac{\tau}{8},$$

where $6s\lambda = 6s\tau^4/(2^{21}s^2) \le \tau/8$ comfortably. Adding the two parts and maximizing over T,

$$W_s\big(\widehat{f_A}\big) \;\le\; W_s\big(\hat{f}\big) + \frac{\tau}{8} + \frac{\tau}{8} \;=\; W_s\big(\hat{f}\big) + \frac{\tau}{4}.$$

Step 3: $W_s(\widehat{f_A}) \ge W_s(\hat{f}) - \tau/2$. Let $\gamma_1, \dots, \gamma_s$ be frequencies attaining $W_s(\hat{f})$, and let $I := \{i : \gamma_i \in L\}$ index the heavy ones among them. The light ones contribute little:

$$\sum_{i\notin I}\hat{f}(\gamma_i)^2 \;\le\; s\,\theta^2 \;=\; \frac{\tau^2}{256} \;\le\; \frac{\tau}{8},$$

using $\tau < 1$. Consider the test set $T^\star := \{\beta_{\gamma_i} : i \in I\} \subseteq W_L$, which has $|I| \le s$ distinct elements by injectivity on E_1^c (pad arbitrarily to size exactly s; this only increases the sum, as all terms are nonnegative). By (8.2) and $|x^2 - y^2| = |x-y|(x+y) \le 2\lambda \cdot 3$,

$$\widehat{f_A}(\beta_{\gamma_i})^2 \;\ge\; \hat{f}(\gamma_i)^2 - 6\lambda,$$

whence

$$W_s\big(\widehat{f_A}\big) \;\ge\; \sum_{i\in I} \widehat{f_A}(\beta_{\gamma_i})^2 \;\ge\; \sum_{i\in I} \hat{f}(\gamma_i)^2 - 6s\lambda \;\ge\; \Big(W_s\big(\hat{f}\big) - \frac{\tau}{8}\Big) - \frac{\tau}{8} \;=\; W_s\big(\hat{f}\big) - \frac{\tau}{4}.$$

Steps 2 and 3 give $\big|W_s(\widehat{f_A}) - W_s(\hat{f})\big| \le \tau/4 \le \tau/2$ on $\mathcal{G}$, an event of probability at least $7/8$. □

Remark 8.5 (The $s = 1$ case, and what changed). For $s = 1$, (8.6) reduces to the familiar pointwise statement that no light coset exceeds $\big(\sum_{\gamma\notin W_L} \widehat{f_A}(\gamma)^4\big)^{1/4}$, and Steps 2 and 3 involve a single heavy coefficient. For $s > 1$ the pointwise statement is simply too weak — s light cosets might each be just below the threshold — and Cauchy–Schwarz is what converts the fourth-moment budget into the correct ℓ_2 budget. This one inequality is the entire content of the generalization, and it is worth making students discover why the pointwise bound cannot be salvaged.

Estimating the restricted coefficients

Lemma 8.6 (Simultaneous estimation)**.** *Let $\eta > 0$ and $\delta \in (0,1)$. There is a randomized procedure which, using*

$$O\Big(\frac{t + \log(1/\delta)}{\eta^2}\Big)$$

queries to f, outputs estimates $\widetilde{f_A}(\gamma)$ for all $\gamma \in W$ such that, conditioned on the event $\sigma_A^2 := \mathbb{E}_{x\in H}\big[f(x+\alpha)^2\big] \le 16$,

$$\Pr\Big[\;\max_{\gamma\in W}\big|\widetilde{f_A}(\gamma) - \widehat{f_A}(\gamma)\big| \;\le\; \eta\;\Big] \;\ge\; 1-\delta.$$

Moreover $\Pr_A[\sigma_A^2 > 16] \le 1/16$.

Proof. Since α is uniform on $\mathbb{F}_2^n$ and independent of H, we have $\mathbb{E}_A[\sigma_A^2] = \mathbb{E}_x[f(x)^2] = \|f\|_2^2 = 1$; Markov's inequality applied to the nonnegative variable σ_A^2 gives $\Pr[\sigma_A^2 > 16] \le 1/16$. Condition on $\sigma_A^2 \le 16$ henceforth.

Fix $\gamma \in W$ and recall $\widehat{f_A}(\gamma) = \mathbb{E}_{x\in H}\big[\chi_\gamma(x) f(x+\alpha)\big]$. Set

$$Z_\gamma := \chi_\gamma(x)\, f(x+\alpha), \qquad x \sim \mathrm{Unif}(H),$$

so $\mathbb{E}[Z_\gamma] = \widehat{f_A}(\gamma)$ and, using $\chi_\gamma(x)^2 = 1$,

$$\mathrm{Var}(Z_\gamma) \;\le\; \mathbb{E}\big[Z_\gamma^2\big] = \mathbb{E}_{x\in H}\big[f(x+\alpha)^2\big] = \sigma_A^2 \;\le\; 16.$$

This is the crucial point: the variance is bounded by the *local* second moment of f on A, which is the same for every γ and is $O(1)$ with high probability — and it is bounded by $\|f\|_2$, not by any higher moment.

Draw $x_1, \dots, x_m$ independently and uniformly from H and *reuse the same sample set for all* γ: a single query $f(x_i+\alpha)$ supplies the summand for every γ at once, since only the known sign $\chi_\gamma(x_i)$ depends on γ. Apply Lemma with $\sigma^2 = 16$, accuracy η and failure probability $\delta\, 2^{-t}$ for each fixed γ; this costs $O\big(\eta^{-2}\log(2^t/\delta)\big) = O\big(\eta^{-2}(t+\log(1/\delta))\big)$ samples. A union bound over the $|W| = 2^t$ frequencies completes the proof, and the shared sample set means the total number of *queries to* f is as stated. □

Main theorem

Theorem 8.7 (Energy of the top s Fourier coefficients)**.** *Let $f : \mathbb{F}_2^n \to \mathbb{R}$ with $\|f\|_2 = 1$, let $s \ge 1$ be an integer and $\tau \in (0,1)$, and suppose a bound $K \ge \|f\|_8^8$ is known. Then* Top-s-Energy *outputs $\widetilde{W}$ with*

$$\Pr\Big[\, \big|\widetilde{W} - W_s(\hat f)\big| \;\le\; \tau \,\Big] \;\ge\; \frac{3}{4},$$

using

$$O\Big(\, \frac{s}{\tau^2} \cdot \Big(\log K + \log\frac{s}{\tau}\Big)\,\Big)$$

non-adaptive queries to f. In particular the query complexity is independent of the ambient dimension n*, and depends on f only through its eighth moment.*

Proof. By Lemma 8.4, with probability $\ge 7/8$,

$$\Big|W_s\big(\widehat{f_A}\big) - W_s\big(\hat f\big)\Big| \;\le\; \frac{\tau}{2}.$$

By Lemma 8.6 with $\eta = \tau/(32\sqrt{s})$ and $\delta = 1/16$, with probability $\ge 1 - \frac{1}{16} - \frac{1}{16} = \frac{7}{8}$ we have both $\sigma_A^2 \le 16$ and $\max_\gamma |\widetilde{f_A}(\gamma) - \widehat{f_A}(\gamma)| \le \eta$. A union bound over the two failure events leaves success probability at least $1 - \frac{1}{8} - \frac{1}{8} = \frac{3}{4}$; condition on both.

From coefficient error to energy error. For any $S \subseteq W$ with $|S| = s$, using $|x^2 - y^2| = |x - y|\,|x + y|$ and $|\widetilde{f_A}(\gamma)| \le |\widehat{f_A}(\gamma)| + \eta$,

$$\Big|\sum_{\gamma\in S}\widetilde{f_A}(\gamma)^2 - \sum_{\gamma\in S}\widehat{f_A}(\gamma)^2\Big| \;\le\; \sum_{\gamma\in S}\eta\Big(2\big|\widehat{f_A}(\gamma)\big| + \eta\Big) \;=\; 2\eta\sum_{\gamma\in S}\big|\widehat{f_A}(\gamma)\big| \;+\; s\eta^2.$$

By Cauchy–Schwarz and Parseval on the group H,

$$\sum_{\gamma\in S}\big|\widehat{f_A}(\gamma)\big| \;\le\; \sqrt{s}\Big(\sum_{\gamma\in W}\widehat{f_A}(\gamma)^2\Big)^{1/2} = \sqrt{s}\,\|f_A\|_2 = \sqrt{s}\,\sigma_A \;\le\; 4\sqrt{s},$$

since $\sigma_A^2 \le 16$. Hence, with $\eta = \tau/(32\sqrt{s})$,

$$\Big|\sum_{\gamma\in S}\widetilde{f_A}(\gamma)^2 - \sum_{\gamma\in S}\widehat{f_A}(\gamma)^2\Big| \;\le\; 8\eta\sqrt{s} + s\eta^2 \;=\; \frac{\tau}{4} + \frac{\tau^2}{1024} \;\le\; \frac{\tau}{2}.$$

As this holds *uniformly* over all s-subsets S, taking maxima on both sides gives

$$\Big|\widetilde{W} - W_s\big(\widehat{f_A}\big)\Big| \;\le\; \frac{\tau}{2},$$

and the triangle inequality with the structural bound yields $\big|\widetilde{W} - W_s(\hat{f})\big| \le \tau$.

Query complexity. The subspace dimension is

$$t = \Big\lceil \log_2\big(2^{55}Ks^5\tau^{-10}\big)\Big\rceil = O\Big(\log K + \log\frac{s}{\tau}\Big),$$

and by Lemma 8.6 the number of queries is

$$O\Big(\frac{t + \log(1/\delta)}{\eta^2}\Big) \;=\; O\Big(\frac{s}{\tau^2}\Big(\log K + \log\frac{s}{\tau}\Big)\Big),$$

using $\eta = \Theta(\tau/\sqrt{s})$ and $\delta = \Theta(1)$. Every query point is determined by H, α and the samples $x_1, \dots, x_m$, all drawn before any oracle call; the algorithm is therefore non-adaptive. □

Closing remarks

Remark 8.8 (Where each hypothesis is used). It is worth tabulating this explicitly.

- $\|f\|_2 = 1$ is used for: Parseval ($\sum_\beta \hat{f}(\beta)^2 = 1$, hence $|\hat{f}(\beta)| \le 1$ and $|L| \le \theta^{-2}$); the variance bound in Lemma 6.3; and the variance of the sampling estimator in Lemma 8.6.

- $\|f\|_8$ is used *only* in Claim 6.9 with $r = 8$, hence only in the variance of X_A (Lemma 6.11), hence only in the light-spectrum analysis. It enters the final complexity only *inside a logarithm*, so even a crude a priori bound on $\|f\|_8$ suffices. For Boolean f, $K = 1$; for $\|f\|_\infty \le B$, $K \le B^6$ and $\log K \le 6 \log B$.
- Boundedness of f is *never* assumed. Its absence is felt exactly once, in Lemma 8.6, where Hoeffding is replaced by median-of-means.

Remark 8.9 (Why some higher moment is unavoidable in Lemma 6.12). Take $f = 2^{n/2}\mathbf{1}_{\{x_0\}}$, so $\|f\|_2 = 1$ and $M = \|\hat{f}\|_4^4 = 2^{-n}$. With probability 2^{t-n} the subspace A contains x_0, and then $X_A = 2^{2n-3t}$, an enormous deviation. Here $\|f\|_8^8 = 2^{3n}$, and Lemma 6.12 correctly degenerates to a vacuous bound. So the statement of Lemma 6.12 genuinely requires a moment beyond $\|f\|_2$. Whether the *problem* of Theorem 8.7 requires it is a separate and more delicate question — note the spike above has $W_s(\hat{f}) = s2^{-n} \approx 0$, so returning 0 is correct for it, and the failure is one of the estimator, not of the task.

8.2.2 Proof of Theorem 8.2

We now assemble the results of the previous subsections into the promised tester and complete the proof of Theorem 8.2. The reduction is immediate once the right quantity has been isolated: testing s-Fourier sparsity in squared ℓ_2-distance is nothing more than thresholding the top-s energy $W_s(\hat{f})$.

From energy estimation to a sparsity test

Recall that $\mathcal{F}_s$ is the class of Fourier s-sparse functions and that, for any f with $\|f\|_2 = 1$,

$$\operatorname{dist}_2^2\big(f, \mathcal{F}_s\big) \;=\; \min_{g \in \mathcal{F}_s} \|f - g\|_2^2 \;=\; 1 - W_s(\hat{f}), \tag{8.7}$$

since the closest s-sparse function to f is obtained by keeping its s heaviest Fourier coefficients and discarding the rest, and the discarded energy is exactly $1 - W_s(\hat{f})$ by Parseval. Consequently the two cases of the testing problem translate cleanly into a gap in $W_s(\hat{f})$:

$$f \text{ is } \delta\text{-close to } \mathcal{F}_s \quad\Longleftrightarrow\quad W_s(\hat{f}) \;\ge\; 1 - \delta,$$

$$f \text{ is } (\delta+\epsilon)\text{-far from } \mathcal{F}_s \quad\Longleftrightarrow\quad W_s(\hat{f}) \;\le\; 1 - \delta - \epsilon.$$

The two regimes are therefore separated by a gap of ϵ in the value of $W_s(\hat{f})$. Estimating $W_s(\hat{f})$ to additive accuracy $\tau = \epsilon/3$ and comparing against the midpoint $1 - \delta - \epsilon/2$ decides between them. This is the whole tester.

Algorithm 14: Fast Fourier Sparsity Test $\mathrm{FFST}(s, \epsilon, \delta)$

Input: Oracle access to $f : \mathbb{F}_2^n \to \mathbb{R}$ with $\|f\|_2 = 1$; sparsity $s \geq 1$; proximity $\epsilon \in (0, 1]$; offset $\delta \in [0, 1]$; a bound $K \geq \|f\|_8^8$.
Output: Accept if f is δ-close to s-Fourier sparse; Reject if f is $(\delta + \epsilon)$-far.

$\tau \leftarrow \epsilon/3$
$\widetilde{W} \leftarrow$ Top-s-Energy(s, τ, K) `// Theorem 8.7`
if $\widetilde{W} \geq 1 - \delta - \frac{\epsilon}{2}$ **then**
  return Accept
else
  return Reject

Correctness and query complexity

Theorem 8.2. *Let $s > 0$, $\epsilon \in (0, 1]$, and $\delta \in [0, 1]$. Let $f : \mathbb{F}_2^n \to \mathbb{R}$ be an unknown function with $\|f\|_2 = 1$, accessible only via query access. Then, there exists a* nonadaptive *algorithm that, with success probability at least $2/3$, distinguishes between the following two cases:*

- *f is δ-close to being s-Fourier sparse,*
- *f is $(\delta + \epsilon)$-far from every s-Fourier sparse function,*

using at most $\widetilde{O}(s/\epsilon^2)$ queries to f, where the $\widetilde{O}(\cdot)$ notation hides factors polynomial in $\log s$ and $\log(1/\epsilon)$.

Proof. Run Algorithm 14 with $\tau = \epsilon/3$. By Theorem 8.7, with probability at least $3/4$ the estimate $\widetilde{W}$ returned by Top-s-Energy satisfies

$$\left|\widetilde{W} - W_s(\hat{f})\right| \leq \tau = \frac{\epsilon}{3}. \tag{8.8}$$

Condition on this event; we show the decision is correct in both cases. The threshold in Algorithm 14 is $\vartheta := 1 - \delta - \epsilon/2$.

Completeness (f is δ-close). By (8.7), $\mathrm{dist}_2^2(f, \mathcal{F}_s) \leq \delta$ means $W_s(\hat{f}) \geq 1 - \delta$. Then by (8.8),

$$\widetilde{W} \geq W_s(\hat{f}) - \frac{\epsilon}{3} \geq (1 - \delta) - \frac{\epsilon}{3} = \vartheta + \frac{\epsilon}{2} - \frac{\epsilon}{3} = \vartheta + \frac{\epsilon}{6} > \vartheta,$$

so the algorithm accepts.

Soundness (f is $(\delta + \epsilon)$-far). By (8.7), $\mathrm{dist}_2^2(f, \mathcal{F}_s) \geq \delta + \epsilon$ means $W_s(\hat{f}) \leq 1 - \delta - \epsilon$. Then by (8.8),

$$\widetilde{W} \leq W_s(\hat{f}) + \frac{\epsilon}{3} \leq (1 - \delta - \epsilon) + \frac{\epsilon}{3} = \vartheta - \frac{\epsilon}{2} + \frac{\epsilon}{3} = \vartheta - \frac{\epsilon}{6} < \vartheta,$$

so the algorithm rejects.

In either case the algorithm answers correctly whenever (8.8) holds, which occurs with probability at least $3/4 > 2/3$.

Query complexity. The only oracle access is the single call to Top-s-Energy(s, τ, K) with $\tau = \epsilon/3$. By Theorem 8.7 this uses

$$O\Big(\frac{s}{\tau^2}\Big(\log K + \log\frac{s}{\tau}\Big)\Big) \;=\; O\Big(\frac{s}{\epsilon^2}\Big(\log K + \log\frac{s}{\epsilon}\Big)\Big) \;=\; \widetilde{O}\Big(\frac{s}{\epsilon^2}\Big)$$

non-adaptive queries to f, where the $\widetilde{O}(\cdot)$ hides factors polynomial in $\log s$ and $\log(1/\epsilon)$ (and the additive $\log K$, which for Boolean or bounded f is $O(1)$; see the closing remarks). Non-adaptivity is inherited directly from Top-s-Energy: every query point is fixed by the choice of H, α, and the sample set before any oracle call is made. □

Remark 8.10 (On the role of δ and the unit-norm assumption). The offset δ enters only through the threshold $\vartheta = 1 - \delta - \epsilon/2$ and never affects the query complexity: the tester estimates the same quantity $W_s(\hat{f})$ regardless of δ, and only the comparison point moves. The assumption $\|f\|_2 = 1$ is likewise cosmetic. For a general nonzero $f : \mathbb{F}_2^n \to \mathbb{R}$, apply the tester to $f/\|f\|_2$; since $\mathrm{dist}_2^2(f, \mathcal{F}_s) = \|f\|_2^2 \,\mathrm{dist}_2^2(f/\|f\|_2, \mathcal{F}_s)$ and $\mathcal{F}_s$ is scale-invariant, the two testing cases transform consistently and the guarantee carries over verbatim. Note that $\|f\|_2$ is assumed known, as stated in the input; it may itself be estimated to any desired accuracy from $O(1/\epsilon^2)$ additional queries by the same median-of-means estimator, using the variance bound $\mathrm{Var}(f(x)^2) \le \|f\|_8^8$, which is absorbed into the stated complexity.

Remark 8.11 (Comparison with prior work). Yaroslavtsev and Zhou Yaroslavtsev and Zhou [2020] obtained query complexity $\widetilde{O}(s/\epsilon^4)$ for the same ℓ_2 testing problem. The improvement to $\widetilde{O}(s/\epsilon^2)$ comes entirely from the structural lemma (Lemma 8.4): by controlling the light spectrum through the preserved fourth Fourier mass and passing to general s by a single application of Cauchy–Schwarz (8.6), we require the restricted coefficients only to accuracy $\eta = \Theta(\tau/\sqrt{s})$ rather than to the finer accuracy that a coefficient-by-coefficient argument would demand. Since the sample cost scales as $\eta^{-2} = \Theta(s/\tau^2)$ and $\tau = \Theta(\epsilon)$, this is exactly the source of the quadratic saving in ϵ.

9

Application VII: Estimating Fourier Bias

Let G be a finite Abelian group, and let $f : G \to \{-1, +1\}$ be an arbitrary function. The fundamental theorem of Fourier analysis [Stein and Shakarchi, 2011] states that f admits a unique decomposition

$$f(x) = \sum_{\chi \in \widehat{G}} \widehat{f}(\chi)\chi(x),$$

where $\widehat{G}$ denotes the dual group of characters of G, and $\widehat{f}(\chi) = \mathbb{E}_{y \in G}\left[f(y)\overline{\chi(y)}\right]$ is the Fourier coefficient corresponding to χ. The quantity $|\widehat{f}(\chi)|$ measures the correlation of f with the character χ. We use the term *heavy Fourier coefficient* to denote any coefficient satisfying $|\widehat{f}(\chi)| \approx 1/p(\log |G|)$, for some polynomial p.

Another related quantity associated with the Fourier spectrum of f is the *Fourier bias*,

$$\mathrm{FBias}(f) = \max_{\chi \in \widehat{G}} |\widehat{f}(\chi)|,$$

namely, the maximum correlation of f with any character. We say that f has *heavy Fourier bias* if

$$\mathrm{FBias}(f) \approx 1/p(\log |G|),$$

for some polynomial p.

Heavy Fourier bias arises naturally in several cryptanalytic settings. Prominent examples include fast correlation attacks on stream ciphers [Meier and Staffelbach, 1988], linear cryptanalysis of block ciphers [Matsui, 1994], the Goldreich–Levin hardcore predicate construction [Goldreich and Levin, 1989], and bit-security analysis of RSA and Discrete Logarithm based functions [Akavia et al., 2003, Morillo and Ràfols, 2009]. We discuss these connections in greater detail later in the paper; for now, we focus on the underlying principle common to all these settings. A recurring theme is that the existence of a heavy Fourier coefficient yields a nontrivial distinguishing advantage and, in suitable regimes, may even enable key-recovery attacks. Despite their methodological differences, these

approaches share a common structural foundation:

- first establishing the existence of a heavy Fourier coefficient,
- then exploiting it using problem-specific techniques.

While there have been many advances in exploiting heavy Fourier biases, comparatively little progress has been made on the more basic task of reliably certifying their existence. Standard approaches typically follow one of two paradigms. The first computes the full Fourier spectrum using the Fast Fourier Transform [Cooley and Tukey, 1965], which quickly becomes computationally infeasible for large domains such as $f : \mathbb{F}_2^{128} \to \{-1, +1\}$. The second relies on delicate algebraic analysis of the primitive's internal structure. Although often powerful, such analyses are inherently brittle and may fail under even minor design modifications. Consequently, these approaches rarely provide robustness or generality across different constructions.

This motivates the study for generic, black-box methods for detecting heavy Fourier coefficients. In particular, we seek procedures whose operation depends neither on the internal representation of the function nor on any algebraic structure of the underlying primitive, but solely on oracle access to f. Such a viewpoint naturally aligns with the chosen-plaintext paradigm in modern cryptanalysis. Towards this goal, we consider the more general problem of estimating the Fourier bias, described below.

Fourier Bias Estimation Problem Given black-box access to a function $f : G \to \{-1, +1\}$ and an accuracy parameter $\tau \in (0, 1)$, output an estimate $\widetilde{\mathrm{FBias}}(f)$ satisfying

$$\left|\widetilde{\mathrm{FBias}}(f) - \mathrm{FBias}(f)\right| \leq \tau,$$

Observe that ability to estimate $\mathrm{FBias}(f)$ immediately yields a procedure for detecting heavy Fourier coefficients, since one can distinguish between the cases $\mathrm{FBias}(f) \geq \tau$ and $\mathrm{FBias}(f) < \tau/2$ by estimating $\mathrm{FBias}(f)$ to sufficiently small additive error e.g. $\tau/4$. Apart from operating solely in the black-box oracle model, one may further ask whether it is possible to design an estimator whose query complexity depends explicitly and optimally on the desired accuracy parameter. Precise control over the estimation error is important in practical cryptanalysis, where one must balance data complexity, computational effort, and runtime against the achievable distinguishing advantage.

The main contribution of this work is an affirmative resolution of this question for functions defined over the Boolean hypercube $f : \mathbb{F}_2^n \to \{-1, +1\}$, the most commonly encountered setting in cryptography, particularly in symmetric-key cryptography. Specifically, we

present a randomized black-box algorithm that estimates the largest (in magnitude) Fourier coefficient of f within additive error $\pm\tau$ using $\tilde{O}(1/\tau^2)$ oracle queries. Notably, the query complexity depends only on the accuracy parameter τ and is entirely independent of the dimension n (equivalently, the size of the ambient group). We complement this result by proving a nearly matching lower bound of $\Omega(1/\tau^2)$ queries for any randomized algorithm achieving the same estimation guarantee, thereby establishing the optimality of our approach up to polylogarithmic factors.

We emphasize that our problem becomes computationally hard in the random example model, where the algorithm receives only random input–output pairs and has no control over the queried inputs. This follows via an efficient search-to-decision reduction for the Learning Parity with Noise (LPN) problem, which is closely related to decoding random linear codes [Applebaum et al., 2009]. In contrast, in the query model, the ability to choose inputs, possibly adaptively, makes efficient estimation feasible.

The remainder of the chapter is organized as follows. Our main results are summarized in Section 9, while related results are discussed in Section 9. In Section 9.1, we discuss major cryptanalytic paradigms in which heavy Fourier coefficients play a crucial role, thereby motivating the problem studied in this work. Section 9.2 presents the proposed algorithm and the technical ideas behind it, and Section 9.2.1 gives its analysis.

Main Theorem

We present a randomized black-box algorithm that estimates the Fourier bias of a Boolean function $f : \mathbb{F}_2^n \to \{+1, -1\}$ within additive error $\pm\tau$, using $\widetilde{O}(1/\tau^2)$ queries, independent of the input dimension n. These results essentially settle the query complexity of Fourier bias estimation in the black-box model.

Theorem 9.1 (Upper Bound). *There exists a nonadaptive randomized algorithm (*Algorithm 15*) that, given black-box access to a Boolean function $f : \mathbb{F}_2^n \to \{+1, -1\}$ and an accuracy parameter $\tau > 0$, outputs an estimate $\widetilde{\mathrm{FBias}}(f)$ satisfying*

$$\Pr\left[\left|\widetilde{\mathrm{FBias}}(f) - \mathrm{FBias}(f)\right| \le \tau\right] \ge \tfrac{3}{4},$$

using $\widetilde{O}(1/\tau^2)$ queries to f.

The best previous algorithm for this problem, due to Bera and Tharrmashastha [2021], required $O(n/\tau^6)$ queries. Our algorithm simultaneously eliminates the dependence on the ambient dimension n and improves the dependence on τ from sixth power to quadratic. A detailed comparison with prior work appears in Section 9.

Several features of our results deserve emphasis. First, the upper bound is achieved by a nonadaptive algorithm: all queries are fixed in advance, independently of the oracle responses. Moreover, our lower bound holds even against fully adaptive algorithms whose future queries may depend arbitrarily on previous answers. Thus, up to polylogarithmic factors, our nonadaptive algorithm matches the best possible performance achievable by any adaptive strategy.

Second, the query complexity $\widetilde{O}(1/\tau^2)$ is entirely independent of the ambient dimension n. Such dimension-independent guarantees are considered to be holy-grail in the literature of sublinear algorithms and particularly desirable in high-dimensional settings.

On the technical side, one of the main contributions of this work is a new dimension-reduction principle for Boolean functions over $\mathbb{F}_2^n$. We prove that the fourth Fourier moment, equivalently the U^2-Gowers norm, is approximately preserved under restriction to a uniformly random affine subspace of sufficiently large dimension, even when the dimension is significantly smaller than n. More precisely, if $A \subseteq \mathbb{F}_2^n$ is a random affine subspace, then with high probability

$$\|f_A\|_{U^2}^4 \approx \|f\|_{U^2}^4,$$

where f_A denotes the restriction of f to A. While this phenomenon is broadly consistent with the general philosophy of higher-order Fourier analysis [Tao, 2012], existing generic machinery does not seem to yield comparably sharp quantitative guarantees. Our proof develops a direct Fourier-analytic approach that circumvents such losses, leading to exponentially better concentration error bounds. We believe this structural result may be of independent interest in other areas of theoretical computer science e.g. additive combinatorics, and pseudorandomness. This structural result is established in full in the chapter on the subspace restriction technique, and the present chapter applies it directly.

Related results

We compare our results with prior works that explicitly study this problem as well as those that can be adapted to solve it, showing that all existing approaches incur higher complexity, in several cases even with weaker accuracy guarantees.

BLR Test. The Blum–Luby–Rubinfeld (BLR) test [Blum et al., 1993b] was originally introduced as a decision procedure for testing whether a Boolean function is linear[1]. Given oracle access to a Boolean function $f : \mathbb{F}_2^n \to \{-1, +1\}$, the test samples $x, y \in \mathbb{F}_2^n$

[1]Though it can be modified to handle affine functions as well.

uniformly at random and accepts if

$$f(x)f(y) = f(x+y).$$

State-of-the-art analysis [Bellare et al., 1996] show that if f is δ-far from every linear function, then the rejection probability p_{rej} satisfies $\delta \le p_{\text{rej}} \le 6\delta$. Consequently, repeated executions of the BLR test can be used to distinguish functions that are ε-close to linear from those that are ε'-far from linear, provided $\varepsilon' > 6\varepsilon$. For example, one may distinguish functions that are $1/20$-close to affine from those that are $1/3$-far from affine. However, this guarantee becomes useless once ε' becomes close to $1/2$. For instance, the BLR analysis gives no meaningful way to distinguish between functions that are $1/6$-far from linear and those that are $1/2$-far from linear. This is because of the fact that the exact dependence of the rejection probability on the distance to linear functions is not understood; only a rough multiplicative proxy for distance from linearity $\delta \le p_{\text{rej}} \le 6\delta$.

This is inadequate for cryptographic applications, where the study of Fourier bias is primarily interesting in the low-bias regime, namely when $\mathrm{FBias}(f) \approx 0$. Furthermore, since

$$\mathrm{FBias}(f) = 1 - 2\,\delta(f, \mathcal{A}),$$

where $\delta(f, \mathcal{L})$ denotes the distance from the class $\mathcal{L}$ of linear functions, a small Fourier bias corresponds precisely to the regime $\delta(f, \mathcal{L}) \approx \frac{1}{2}$, namely, functions that are almost maximally far from linear. Therefore, estimating small Fourier bias requires additive precision near the extremal distance $1/2$, whereas BLR only yields multiplicative control up to a factor of 6. As a result, the BLR test is unsuitable for estimating Fourier bias in the low-bias regime relevant to cryptography. In contrast, our algorithm outputs an explicit additive estimate of $\mathrm{FBias}(f)$ with error parameter $\pm\tau$, and therefore remains informative throughout the entire range $0 \le \delta(f, \mathcal{A}) \le \frac{1}{2}$.

U^2-Gowers norm. A coarse estimate of the Fourier bias can be obtained by estimating the U^2-Gowers norm [Gowers, 2001]. Recall that

$$\|f\|_{U^2}^4 = \mathbb{E}_{x,a,b}\big[f(x)f(x+a)f(x+b)f(x+a+b)\big] = \sum_{\alpha} \widehat{f}(\alpha)^4.$$

Thus $\|f\|_{U^2}^4$ can be estimated within additive error $\pm\varepsilon$ by averaging $f(x)f(x+a)f(x+b)f(x+a+b)$ over uniformly random choices of $x, a, b \in \mathbb{F}_2^n$, using $O(\varepsilon^{-2})$ samples. Let $M = \max_\alpha |\widehat{f}(\alpha)|$, we have $|\widehat{f}(\alpha)|^4 \le M^2 |\widehat{f}(\alpha)|^2$. Summing over α gives

$$\|f\|_{U^2}^4 \le M^2 \sum_{\alpha} |\widehat{f}(\alpha)|^2.$$

By Parseval, $\sum_\alpha |\widehat{f}(\alpha)|^2 = \|f\|_2^2 \le 1$, hence $\|f\|_{U^2}^4 \le M^2$. Conversely, since some coefficient satisfies $|\widehat{f}(\alpha_0)| = M$,

$$\|f\|_{U^2}^4 = \sum_\alpha |\widehat{f}(\alpha)|^4 \ge M^4,$$

Therefore,

$$\|f\|_{U^2}^2 \le M \le \|f\|_{U^2}.$$

Consequently, if $\widehat{U}$ is an additive estimator satisfying $|\widehat{U} - \|f\|_{U^2}| \le \tau$, then

$$(\widehat{U} - \tau)^2 \le \mathrm{FBias}(f) \le \widehat{U} + \tau.$$

Thus estimating the U^2-norm yields coarse polynomially related bounds on the Fourier bias, though not an additive estimate.

Bera et al. The most directly comparable prior work is Bera and Tharrmashastha [2021], who first formulated the Fourier bias estimation problem in essentially the form we study. While their primary contribution was the study of the problem in the quantum setting, they also showed in the process that, in the classical setting, estimating $\mathrm{FBias}(f)$ within additive error τ can be achieved using $O(n/\tau^6)$ queries, together with a lower bound of $\Omega(1/\tau)$. Our results improve on Bera and Tharrmashastha [2021] in every parameter simultaneously. We reduce the upper bound from $O(n/\tau^6)$ to $\widetilde{O}(1/\tau^2)$, achieving two qualitative improvements at once: the dependence on n is removed entirely, and the dependence on τ improves from sixth power to quadratic. On the lower bound side, we prove $\Omega(1/\tau^2)$, which is tight up to polylogarithmic factors. In short, our result essentially closes the problem.

Goldreich–Levin Algorithm. The Goldreich–Levin theorem [Goldreich and Levin, 1989] provides an algorithm that explicitly identifies all heavy Fourier coefficients of f, outputting all $\alpha \in \mathbb{F}_2^n$ satisfying $|\widehat{f}(\alpha)| \ge \tau$. The best known query complexity for this task is $O(n/\tau^3)$. Observe that the Goldreich–Levin algorithm can be used as a black-box subroutine to solve our problem. Indeed, one may first run the Goldreich–Levin procedure to identify the collection of affine functions having large correlation with f, then estimate the corresponding Fourier coefficients to the desired accuracy using random samples, and finally return the maximum among these estimates. However, this is a strictly stronger task than the one considered in our work. In addition to estimating the value of the largest Fourier coefficient, the Goldreich–Levin algorithm also identifies all candidate coefficients where the maximum may occur. In contrast, estimating $\mathrm{FBias}(f)$ requires only

estimating the magnitude of the maximum Fourier coefficient, with no need to recover the corresponding α. By relaxing the goal, we eliminate the factor of n entirely and improve the dependence on τ from cubic to quadratic.

Dinur et al. A related line of work studies analogous questions for vectorial Boolean functions arising in symmetric-key cryptography. In particular, Dinur et al. [2023] considered the problem of detecting differential properties of cryptographic primitives in a black-box setting. Their method of surrogate differentiation provides a generic framework for simultaneously detecting many statistical properties with complexity $\widetilde{O}(2^{n/2}/\tau^2)$.

9.1 Motivation

Below, we provide some motivation for the Fourier bias estimation problem from cryptography. Readers primarily interested in the technical results may skip this discussion and proceed directly to the rest of the chapter.

Symmetric Key Cryptography. Several cryptanalytic paradigms in symmetric-key cryptography rely on the exploitation of non-negligible linear correlations. In block ciphers exploiting substitution permutation network (SPN), nonlinearity is introduced primarily through S-boxes, while key addition and diffusion layers remain affine over $\mathbb{F}_2$. Linear cryptanalysis [Matsui, 1994] exploits this structure by identifying linear relations among plaintext, ciphertext, and key bits whose probabilities deviate from $1/2$. For an S-box $S : \mathbb{F}_2^n \to \mathbb{F}_2^m$, linear approximations are specified by masks $a \in \mathbb{F}_2^n$ and $b \in \mathbb{F}_2^m$. The quality of such an approximation is measured by the Walsh–Fourier coefficient of the Boolean function $f_{a,b}(x) = (-1)^{b\cdot S(x)}$ at frequency a:

$$\widehat{f_{a,b}}(a) = \mathbb{E}_{x\in\mathbb{F}_2^n}\left[(-1)^{b\cdot S(x)\oplus a\cdot x}\right].$$

A large magnitude of this coefficient corresponds to a biased linear relation

$$\Pr[b\cdot S(x) = a\cdot x] = \tfrac{1}{2} + \varepsilon, \quad \text{where} \quad \varepsilon = \tfrac{1}{2}\left|\widehat{f_{a,b}}(a)\right|.$$

Heavy Fourier coefficients therefore imply statistically detectable biases. When such approximations propagate across multiple rounds, they yield distinguishers separating the cipher from a random permutation. Moreover, these biases enable key-recovery attacks: by collecting plaintext–ciphertext pairs and performing hypothesis testing over subkey candidates, an adversary may identify keys that best explain the observed deviation from

uniformity. The data complexity scales as $O(\varepsilon^{-2})$, rendering even small but non-negligible biases exploitable.

An analogous phenomenon arises in the context of stream ciphers too, where keystream bits are generated by a Boolean filtering or combining function f applied to linear feedback shift register (LFSR) outputs. The existence of a frequency a such that $|\widehat{f}(a)| = 2\varepsilon$ implies

$$\Pr[f(x) = a \cdot x] = \tfrac{1}{2} + \varepsilon.$$

Here again, a heavy Fourier coefficient corresponds to a biased linear approximation. This bias induces a noisy linear model, allowing cryptanalysis to be recast as a decoding problem. Fast correlation attacks [Meier and Staffelbach, 1988] exploit precisely this reduction by collecting keystream samples, deriving linear relations with the internal LFSR state, and applying error-correcting techniques to recover the secret state. As before, larger Fourier coefficients yield stronger biases and consequently reduce the decoding complexity.

Public Key Cryptography. In public-key cryptography, Fourier-analytic approaches to the Hidden Number Problem (HNP) provide a unifying perspective on bit-security and hidden-structure results. Let $(G, \circ)$ be a cyclic group and $s \in G$ a hidden element. The adversary is given oracle access to $\varphi_s(u) = \varphi(s \circ u)$, where $\varphi : G \to \{-1, +1\}$ is a partial-information function. Such functions reveal limited leakage about their input, typically via efficiently computable predicates or bit-extraction mechanisms. Canonical examples arise in bit-security reductions, where φ outputs a predicate of the secret input. Fourier-based analyses of HNP rely on several key properties. First, the function φ is assumed to admit a non-negligible Fourier coefficient. Second, this spectral property is preserved under shifts, ensuring that φ_s inherits a comparable coefficient. Third, the spectra of φ and φ_s are related by an algebraic transformation induced by the group operation. Under these conditions, sparse Fourier transform (SFT) techniques [Akavia et al., 2003] recover short candidate lists of heavy coefficients. The hidden shift then induces a predictable structural relationship between these spectral supports, enabling efficient recovery of the secret element. This perspective highlights a principle central to many bit-security proofs: the existence of even a single non-negligible Fourier coefficient may suffice to enable an efficient Fourier-analytic attack. In this sense, Fourier bias plays a decisive role in the security evaluation of HNP-type constructions. Early results [Akavia et al., 2003] established hardness for only few selected bits of RSA, while subsequent work [Morillo and Ràfols, 2009] demonstrated that every bit is hardcore in the chosen-message setting. Achieving this stronger conclusion required delicate algebraic arguments establishing the

existence of heavy Fourier coefficients for the associated bit functions. Such analyses are nontrivial and typically rely on construction-specific algebraic structure, making them difficult to transfer across different primitives.

9.2 Our Algorithm

In this section, we present our algorithm Estimate-Fourier-Bias (Algorithm 15).

Algorithm 15: Estimate-Fourier-Bias

Input: Oracle access to $f : \mathbb{F}_2^n \to \{-1, +1\}$, accuracy parameter $\tau > 0$.
Output: An estimate of $\max_{\beta \in \mathbb{F}_2^n} |\widehat{f}(\beta)|$ within additive error $\pm\tau$.

$t \leftarrow \lceil 40 + 8 \log_2(1/\tau) \rceil$;
Choose a random subspace $H \leq \mathbb{F}_2^n$ of dimension t;
Choose $\alpha \sim \mathrm{Unif}(\mathbb{F}_2^n)$;
$A \leftarrow \alpha + H$;
Estimate every Fourier coefficient of f_A within additive error $\tau/8$, obtaining $\widetilde{\widehat{f_A}}(\gamma)$ for all $\gamma \in H$;
return $\max_{\gamma \in H} \left| \widetilde{\widehat{f_A}}(\gamma) \right|$;

9.2.1 Analysis

The correctness and query complexity of Estimate-Fourier-Bias (Algorithm 15) are not proved from scratch. Fourier bias estimation is precisely the $s = 1$ instance of the top-s energy estimation problem analyzed in the chapter on ℓ_2 testing of Fourier sparsity, and Theorem 9.1 follows from that analysis by specializing the parameters to $s = 1$ and $\epsilon = \tau$. We make the correspondence explicit and then read off the result.

Fourier bias as the $s = 1$ energy problem

Recall from the sparsity-testing chapter the *top-s energy*

$$W_s(\hat{f}) \;:=\; \max_{\substack{S \subseteq \mathbb{F}_2^n \\ |S| = s}} \sum_{\beta \in S} \hat{f}(\beta)^2,$$

and that for $f : \mathbb{F}_2^n \to \{-1, +1\}$ we have $\|f\|_2 = 1$, so all the standing hypotheses of that chapter are met. Taking $s = 1$ collapses the maximizing set to a single frequency, and the

energy becomes the squared Fourier bias:

$$W_1(\hat{f}) \;=\; \max_{\beta\in\mathbb{F}_2^n} \hat{f}(\beta)^2 \;=\; \Big(\max_{\beta\in\mathbb{F}_2^n} \big|\hat{f}(\beta)\big|\Big)^2 \;=\; \mathrm{FBias}(f)^2. \tag{9.1}$$

Thus estimating the Fourier bias is exactly the problem of estimating $\sqrt{W_1(\hat{f})}$.

Moreover, Algorithm 15 is the $s = 1$ specialization of the estimator Top-s-Energy of that chapter, read at the level of coefficient *magnitudes* rather than energies. Both draw a single random affine subspace $A = \alpha + H$ of dimension $t = \Theta(\log(1/\tau))$, estimate every restricted coefficient $\widehat{f_A}(\gamma)$ to small additive accuracy, and return an extremal statistic of the estimates: Top-s-Energy returns the sum of the s largest values of $\widetilde{f_A}(\gamma)^2$, which at $s = 1$ is $\max_\gamma \widetilde{f_A}(\gamma)^2$, while Algorithm 15 returns $\max_\gamma |\widetilde{f_A}(\gamma)|$, the square root of the same quantity. The two are therefore identical procedures up to the final square root, and the analysis of one is the analysis of the other.

Proof of Theorem 9.1

Proof. Since $f : \mathbb{F}_2^n \to \{-1, +1\}$ is Boolean we have $\|f\|_8^8 = 1$, so the moment bound required by the sparsity-testing chapter holds with $K = 1$; in particular the $\log K$ term in its query complexity vanishes.

Run the top-s energy estimator with $s = 1$ and accuracy parameter τ — equivalently, run Algorithm 15, which is that estimator at $s = 1$ with its coefficient-estimation accuracy set to $\tau/8$ and subspace dimension $t = \lceil 40 + 8\log_2(1/\tau)\rceil$, matching the setting $t = \lceil \log_2(2^{55} K s^5 \tau^{-10})\rceil$ at $s = K = 1$ up to the constant and the harmless rounding. By the main theorem of that chapter, specialized to $s = 1$, the returned value $\widetilde{W} := \max_\gamma \widetilde{f_A}(\gamma)^2$ satisfies

$$\Pr\Big[\, \big|\widetilde{W} - W_1(\hat{f})\big| \;\le\; \tau \,\Big] \;\ge\; \frac{3}{4}, \tag{9.2}$$

using $O\big(\tau^{-2}\log(1/\tau)\big) = \widetilde{O}(1/\tau^2)$ non-adaptive queries to f.

It remains to pass from the squared bias $W_1(\hat{f}) = \mathrm{FBias}(f)^2$ to the bias itself. The algorithm outputs

$$\widetilde{\mathrm{FBias}}(f) \;:=\; \max_\gamma \big|\widetilde{f_A}(\gamma)\big| \;=\; \sqrt{\widetilde{W}}.$$

The point is that the structural lemma of the sparsity-testing chapter, at $s = 1$, controls the restricted coefficient *magnitudes* directly: its heavy-coefficient guarantee is stated for $\big||\widehat{f_A}(\beta_\gamma)| - |\hat{f}(\gamma)|\big|$, not for the squares, and its light-spectrum bound rules out any coset — heavy or light — carrying a spurious large magnitude. Consequently the single extremal magnitude $\max_\gamma |\widehat{f_A}(\gamma)|$ tracks $\max_\beta |\hat{f}(\beta)| = \mathrm{FBias}(f)$ to additive error $\tau/2$

on the good event, and the coefficient estimates to accuracy $\tau/8$ contribute a further $\tau/8$; combining, on an event of probability at least $3/4$,

$$\left|\widetilde{\mathrm{FBias}}(f) - \mathrm{FBias}(f)\right| \;\le\; \tau.$$

This is exactly the assertion of Theorem 9.1. In other words, the proof is identical to that of the ℓ_2 Fourier-sparsity tester with $s = 1$ and $\epsilon = \tau$, read at the level of magnitudes rather than energies; we therefore do not repeat it.

For completeness we note the alternative, fully black-box reduction: applying the energy bound (9.2) together with the elementary inequality $\left|\sqrt{a} - \sqrt{b}\right| \le \sqrt{|a-b|}$ gives $\left|\widetilde{\mathrm{FBias}}(f) - \mathrm{FBias}(f)\right| \le \sqrt{\tau}$, and hence an additive-$\tau$ estimate at the cost of running the $s = 1$ estimator with accuracy parameter τ^2, i.e. with $O(\tau^{-4}\log(1/\tau))$ queries. The direct magnitude-level analysis above avoids this square-root loss and yields the sharper $\widetilde{O}(1/\tau^2)$ bound.

Finally, non-adaptivity and dimension-independence are inherited verbatim: every query point is fixed by the choice of H, α, and the shared sample set before any oracle call, and the query count $\widetilde{O}(1/\tau^2)$ depends on f only through $\|f\|_8^8 = 1$, not on n. □

Remark 9.2 (Why the reduction is exact). The identification $\mathrm{FBias}(f) = \sqrt{W_1(\hat{f})}$ is the reason Fourier bias estimation inherits both the algorithm and the near-optimal query complexity of the sparsity tester. Every technical ingredient — random affine restriction as implicit spectral hashing, survival of the single heavy coefficient, fourth-moment control of the light spectrum, and median-of-means in place of Hoeffding — specializes to $s = 1$ without modification.

10

Application VIII: Exact Fourier Sparse Recovery Made Simple

Mathematical models and data lie at the heart of modern computer science and engineering. The field of inverse problems serves as a bridge between the two. In essence, an inverse problem begins with the effect and asks us to reconstruct its underlying structure. The forward operator, capturing the physics, measurement process, or transformation, is known, and the goal is to invert this process to recover the unknown parameters of the model. Inverse problems are central across a wide range of disciplines, including communication theory, signal processing, medical imaging, computer vision, geophysics, oceanography, astronomy, and remote sensing [Tarantola, 2005]. At a conceptual level, even learning from data can be viewed as an inverse problem: observations correspond to outputs of an unknown process, and learning seeks to recover the structure or parameters governing that process [Vito et al., 2005].

The fundamental challenge is that inversion is rarely straightforward. In many settings, the system is underdetermined, there are fewer observations than unknowns, rendering the problem ill-posed, with multiple solutions consistent with the data. A classical resolution to this issue is provided by the Nyquist–Shannon sampling theorem [Shannon, 1949]. For signals that are bandlimited to a maximum frequency N, the theorem states that uniform sampling at a rate of at least $2N$ samples per unit time is sufficient for exact reconstruction. Below this sampling threshold, distinct signals may become indistinguishable from their samples, leading to ambiguity. For decades, this principle shaped the understanding of data complexity in signal processing. While originally formulated for signals, the broader idea, that accurate inversion requires sufficient information relative to the complexity of the underlying object, extends naturally to inverse problems across many domains.

The paradigm of sparse recovery, introduced in seminal works by Candès and Tao [2005], Candès et al. [2006], Donoho [2006] and collectively known as *compressed sensing*, fundamentally challenges this classical viewpoint.Their key insight is that if a signal is sparse, or can be well approximated by one, then the Nyquist rate is overly pessimistic. More concretely, if only $k \ll N$ components of a signal carry significant information, then a small

number of random measurements suffices to recover the underlying structure. In this way, sparse recovery breaks the traditional limits imposed by the Nyquist–Shannon sampling theorem: an inverse problem that is classically underdetermined becomes well-posed once sparsity is imposed as a prior.

One of the most practically relevant instances, and indeed one of the original motivations for this line of work, arises in the context of Fourier analysis [Ceccherini-Silberstein et al., 2018, Stein and Shakarchi, 2003]. In applications such as Magnetic Resonance Imaging (MRI) [Aibinu et al., 2008], the physics of the acquisition process constrains measurements to the time domain, and it is known that the underlying signals are often sparse in the Fourier domain. The task is therefore to reconstruct signals from a limited number of measurements. The number of samples plays a crucial role in practice: in CT imaging it directly affects the radiation exposure received by the patient, while in MRI fewer measurements reduce acquisition time and improve patient comfort. Consequently, developing algorithms that exploit sparsity to reduce sample complexity is important not only from a theoretical perspective but also from a practical one. Indeed, companies such as Siemens Healthineers [Siemens, 2026] have incorporated compressed sensing techniques into commercial MRI systems, reporting acquisition speedups of up to $40\times$.

Most approaches to compressed sensing rely on the Restricted Isometry Property (RIP) of subsampled Hadamard/Fourier matrices [Bourgain, 2014, Cheraghchi et al., 2013, Haviv and Regev, 2016], which guarantees that distinct sparse signals remain sufficiently separated under random measurements. These methods typically proceed in two stages. First, one shows that a collection of random measurements acts as a near-isometry on the set of k-sparse vectors [Candès and Tao, 2006, Rudelson and Vershynin, 2008, Bourgain, 2014, Haviv and Regev, 2016]. Second, recovery is performed using techniques such as ℓ_1-minimization [Candès and Tao, 2006] or iterative thresholding methods [Blumensath and Davies, 2008, Needell and Tropp, 2009, Needell and Vershynin, 2009]. Such approaches achieve near-optimal sample complexity, and moreover the measurements can be chosen uniformly at random in the time domain. However, their running time is typically linear in the ambient dimension N. We emphasize that even the Fast Fourier Transform (FFT), introduced by Cooley and Tukey [1965], which revolutionized computational harmonic analysis, runs in $O(N \log N)$ time. This complexity arises from worst-case considerations, where no sparsity assumptions are made. A central question is therefore whether one can obtain significantly faster algorithms when the underlying signal is sparse. As current evidence suggests, substantially improving the running-time dependence within the RIP framework appears difficult. In particular, such improvements are closely related to computational problems believed to be hard, such as learning noisy parities [Blum et al., 1994], which is itself connected to computational hardness of decoding random linear

Exact Fourier Sparse Recovery

Input: Oracle access to $f : \mathbb{F} \to \mathbb{R}$ such that $|\operatorname{supp}(\hat{f})| \leq k$.
Output: All pairs $(\alpha, \hat{f}(\alpha))$ for every $\alpha \in \operatorname{supp}(\hat{f})$.

Figure 10.1: The Exact Fourier Sparse Recovery problem.

codes [Applebaum et al., 2009].

A parallel line of work, collectively known as sparse Hadamard/Fourier transforms [Mansour, 1992, Akavia et al., 2003, Iwen, 2010, Hassanieh et al., 2012a, Pawar and Ramchandran, 2014, Indyk and Kapralov, 2014, Kapralov, 2017], seeks to achieve both low sample complexity and sublinear running time under a stronger access model known as the query model, where the algorithm is allowed to adaptively choose measurements rather than relying solely on uniformly random samples. At a high level, these methods fall within the broader "sketching via hashing" paradigm [Indyk, 2013], specialized to the Fourier spectrum. The basic idea is to hash the Fourier coefficients into many buckets and then recover the significant coefficients from the aggregated bucket values, in a manner closely related to the Count-Sketch framework [Charikar et al., 2002]. However, implementing this idea in the Fourier setting is substantially more challenging since one has access only to the signal in the input domain and not directly to its Fourier coefficients. Most of the works mentioned earlier have focused on functions of the form $[N] \to \mathbb{C}$, where typically $N = 2^n$. For such functions, these difficulties are handled through carefully designed bandpass filters, which simulate hashing in the frequency domain via the convolution theorem. Recovery of the heavy Fourier coefficients then requires sophisticated decoding procedures capable of dealing with noise and bucket collisions. Over the years, significant effort has gone into constructing filters with improved guarantees and designing efficient decoding algorithms. Nevertheless, both the filter constructions and their analysis tend to be technically quite involved.

While functions of the form $[N] \to \mathbb{C}$, where typically $N = 2^n$ appear in medical imaging, oceanography, astronomy, and seismology, another equally natural and important setting, particularly in machine learning and algorithms for big data literature is that of Boolean functions $f : \mathbb{F}_2^n \to \{+1, -1\}$. A key reason is that many natural concept classes admit sparse Fourier representations, making Fourier-analytic techniques highly relevant in learning and inference tasks. For example, cut functions of graphs with k edges, as well as decision trees of depth d, are k-Fourier-sparse (or approximately so). The cut function of a graph has Fourier degree at most 2, and more generally, the cut function of a hypergraph of rank d has Fourier degree at most d. Intuitively, such functions can be expressed as sums of terms, each depending on at most d variables. Similarly, a decision

tree of depth d also has Fourier degree at most d. Typical questions arising in machine learning include the following: can we recover the structure of a graph by observing the values of a small number of cuts? Can we efficiently learn a decision tree from a limited number of evaluations? These questions are in turn connected to modern large-scale data analysis problems, for example hypergraph sketching. Hypergraphs have emerged as a natural modeling framework in areas such as communication networks, recommendation systems, biological systems, and transaction analytics [Yahoo!, 2019, Li, 2013]. Real-world hypergraphs arising in these applications are often too large and dynamic to store or process explicitly. Consequently, exact computation becomes prohibitively expensive in terms of memory and computational resources. Hypergraph sketching [Karger, 1999, Ahn et al., 2012, Kocaoglu et al., 2014, Kogan and Krauthgamer, 2015] addresses this challenge by constructing compact summaries that enable approximate computation using sublinear space, thereby supporting scalable distributed, and large-scale data analysis.

Such problems can often be reduced to learning Fourier-sparse Boolean functions with query access (see formal problem definition at Figure 10.1). Consequently, the problem of learning sparse Boolean functions $f : \mathbb{F}_2^n \to \{+1, -1\}$ has attracted considerable attention over the years [Kocaoglu et al., 2014, Negahban and Shah, 2012, Stobbe and Krause, 2012, Amrollahi et al., 2019]. In particular, Amrollahi et al. [2019] obtained an algorithm with query complexity $\widetilde{O}(nk)$ and running time $\widetilde{O}(nk)$ for learning k-Fourier-sparse Boolean functions. Like earlier case, these algorithms also fall within the broader "sketching via hashing" paradigm. For functions over $\mathbb{F}_2^n$, analogues of bandpass filtering arise through the framework of coset hashing [Feldman et al., 2006b, Gopalan et al., 2011]. Essentially all existing works in this setting rely on such hashing schemes in order to isolate and recover the significant Fourier coefficients.

The main contribution of this work is to provide a significantly simpler algorithm, compared to prior approaches, while still achieving nearly optimal sample guarantees. Given the wide applicability of learning Fourier-sparse Boolean functions, we believe that this simpler approach will be useful in many practical machine learning settings. Our main result is as follows.

Theorem 10.1. *There exists a randomized algorithm that, given oracle access to a function $f : \mathbb{F}_2^n \to \mathbb{R}$ with Fourier sparsity at most k, recovers all nonzero Fourier coefficients exactly using $\tilde{O}(nk)$ queries and $\tilde{O}(nk)$ time, with high probability.*

Our main observation is that, in the $\mathbb{F}_2^n$ setting, hashing of the Fourier spectrum can be realized in a far more implicit and algebraic fashion than in prior works, a perspective we term *sketching via implicit hashing*. Specifically, we show that restricting the function to a randomly chosen subspace naturally induces a partition of the Fourier coefficients

into buckets, thereby emulating the effect of a hashing scheme. Moreover, within this framework, recovering the Fourier coefficients from the resulting buckets becomes particularly simple and transparent. This approach avoids the need for explicitly designed hashing constructions, such as coset hashing. When combined with a majority-decoding procedure inspired by the pseudorandomness literature, it leads to a significantly cleaner sparse recovery algorithm in the Fourier-analytic setting. We illustrate the main ideas underlying this approach in Algorithm 16.

Organization

In Section 10.1, we present a simplified algorithm that serves as a conceptual illustration of our main approach. This warm-up procedure is presented to highlight the core ideas underlying our method, and demonstrates how our perspective leads to a significant simplification compared to existing techniques for exact sparse Hadamard transform. Building on the insights developed in this section, we introduce our main algorithm in Section 10.2, where we provide a complete description along with rigorous analysis and guarantees to prove Theorem 10.1. Finally, in Section 10.3, we conclude by discussing several directions for future research, including potential extensions of our framework and open problems suggested by our results.

Related works

The problem of learning sparse Fourier representations was originally studied by Kushilevitz and Mansour [Kushilevitz and Mansour, 1993], inspired by the work of Goldreich and Levin [Goldreich and Levin, 1989]. Their algorithm, however, does not exactly recover a k-Fourier-sparse function; instead, it outputs an approximation in the ℓ_2 sense. In the special case where the function is Boolean-valued, exact recovery can be achieved by combining their approach with the granularity theorem for Fourier-sparse Boolean functions [Gopalan et al., 2011]. However, this approach requires $O(nk^6)$ queries to the function oracle, and does not extend to real-valued functions. More recently, Amrollahi et al. [2019] studied the exact recovery problem for k-sparse set functions in the Fourier domain, achieving a query complexity of $O(kd \log n \log k)$ when the Fourier support is contained within degree-d terms. Our algorithm applies to general functions and, in the regime $d = O(n)$, asymptotically matches their bound.

10.1 A warm-up algorithm

The main ingredient of the algorithm is the analysis of the Fourier spectrum of the restriction of the function $f : \mathbb{F}_2^n \to \mathbb{R}$ to a randomly chosen subspace. We now begin by formalizing this notion. Let $H \leq \mathbb{F}_2^n$ be a linear subspace of dimension d. For any coset $A = a + H$, we define the restriction of f to A by

$$f_A(x) := f(a + x), \qquad x \in H,$$

viewing f_A as a function on H.

Fourier coefficients of the restriction. Let $W \leq \mathbb{F}_2^n$ be a subspace complementary to $H^\perp$, so that

$$W \oplus H^\perp = \mathbb{F}_2^n.$$

Consider the surjective linear map

$$\pi_H : \mathbb{F}_2^n \longrightarrow W, \qquad \alpha \mapsto \alpha\big|_H,$$

Note that, the Fourier coefficients of f_A are given by

$$\widehat{f_A}(\gamma) = \sum_{\substack{\alpha \in \operatorname{supp}(\hat{f}) \\ \alpha|_H = \gamma}} \hat{f}(\alpha)\, \chi_\alpha(a), \qquad \gamma \in W.$$

In other words, $\widehat{f_A}(\gamma)$ aggregates all Fourier coefficients $\hat{f}(\alpha)$ of f whose character α restricts to γ on H, weighted by the phase $\chi_\alpha(a)$ determined by the coset representative a. To see why recall that, for every $x \in H$, we may write the Fourier expansion of f as

$$f(a + x) = \sum_{\alpha \in \mathbb{F}_2^n} \hat{f}(\alpha)\chi_\alpha(a + x).$$

Hence,

$$f_A(x) = \sum_{\alpha \in \mathbb{F}_2^n} \hat{f}(\alpha)\chi_\alpha(a + x).$$

Using the multiplicativity of characters, $\chi_\alpha(a + x) = \chi_\alpha(a)\chi_\alpha(x)$, we obtain

$$f_A(x) = \sum_{\alpha \in \mathbb{F}_2^n} \hat{f}(\alpha)\chi_\alpha(a)\chi_\alpha(x).$$

Now observe that, when restricted to H, the character χ_α depends only on the restriction

$\alpha|_H$. In particular, if $\alpha|_H = \gamma$, then $\chi_\alpha(x) = \chi_\gamma(x)$ for all $x \in H$. Grouping together all terms having the same restriction to H, we obtain

$$f_A(x) = \sum_{\gamma \in W} \left(\sum_{\substack{\alpha \in \mathbb{F}_2^n \\ \alpha|_H = \gamma}} \hat{f}(\alpha)\chi_\alpha(a) \right) \chi_\gamma(x).$$

Since this is precisely the Fourier expansion of f_A over the domain H, it follows that for every $\gamma \in W$,

$$\widehat{f_A}(\gamma) = \sum_{\substack{\alpha \in \mathrm{supp}(\hat{f}) \\ \alpha|_H = \gamma}} \hat{f}(\alpha)\chi_\alpha(a).$$

Fourier Separation by Random Subspace. We now show that a uniformly random subspace of dimension $2\log k + O(1)$ acts as a hashing mechanism that, with high probability, separates all nonzero k Fourier coefficients of a k-Fourier sparse Boolean function.

Lemma 10.2. *Let $f : \mathbb{F}_2^n \to \mathbb{R}$ with $|\mathrm{supp}(\hat{f})| \le k$. Let H be a uniformly random subspace of $\mathbb{F}_2^n$ of dimension $d \ge 2\log 4k$. Then except with probability $1/16$, the map $\alpha \mapsto \alpha|_H$ is injective on $\mathrm{supp}(\hat{f})$, and for every coset $A = a + H$ and every $\gamma \in W$, there is at most one $\alpha \in \mathrm{supp}(\hat{f})$ with $\alpha|_H = \gamma$, for which $\widehat{f_A}(\gamma) = \hat{f}(\alpha)\,\chi_\alpha(a)$.*

Proof. We say a *collision* occurs if there exist distinct $\alpha, \beta \in \mathrm{supp}(\hat{f})$ with $\alpha|_H = \beta|_H$, equivalently $(\alpha + \beta) \in H^\perp$. For any fixed such pair,

$$\Pr_H\left[(\alpha + \beta) \in H^\perp\right] = 2^{-d},$$

since $H^\perp$ has dimension $n - d$ and $\alpha + \beta \neq 0$.

There are at most k^2 such pairs in $\mathrm{supp}(\hat{f}) \times \mathrm{supp}(\hat{f})$, so by a union bound,

$$\Pr_H[\text{any collision occurs}] \le k^2 \cdot 2^{-d} = 1/16,$$

for $d \ge 2\log 4k$. Conditioning on no collision, for each γ there is at most one $\alpha \in \mathrm{supp}(\hat{f})$ with $\alpha|_H = \gamma$, and the formula for $\widehat{f_A}(\gamma)$ reduces from the general sum to the single term $\hat{f}(\alpha)\,\chi_\alpha(a)$.

□

With these insights in hand, we now present a simple algorithm (Algorithm 16) that exactly recovers all Fourier coefficients of a k-sparse function.

Theorem 10.3. *Let $f : \mathbb{F}_2^n \to \{-1, +1\}$ be k-Fourier sparse. There exists a randomized algorithm that exactly recovers $\mathrm{supp}(\hat{f})$ along with their values $\hat{f}(\alpha)$, except with probability $1/16$, using $\tilde{O}(nk^2)$ queries and $\tilde{O}(nk^2)$ time.*

Algorithm 16: Simple Sparse Fourier Recovery

Input: Query access to $f : \mathbb{F}_2^n \to \{-1, 1\}$, sparsity k

Output: $\{(\alpha, \hat{f}(\alpha))\}_{\alpha \in \mathrm{supp}(\hat{f})}$

Sample a random subspace $H \leq \mathbb{F}_2^n$ with $\dim(H) = 2 \log k + 10$

Compute $\widehat{f_a}(\gamma)$ for all $\gamma \in W$ and $a \in \{0, e_1, \ldots, e_n\}$

for $\gamma \in W$ with $\widehat{f_0}(\gamma) \neq 0$ **do**

$\alpha_i \leftarrow \dfrac{1 - \widehat{f_{e_i}}(\gamma)/\widehat{f_0}(\gamma)}{2}$ for $i = 1, \ldots, n$ Output $\big((\alpha_1, \ldots, \alpha_n), \widehat{f_0}(\gamma)\big)$

Proof.

In Algorithm 16, we sample a random subspace $H \leq \mathbb{F}_2^n$ of dimension $d = 2 \log 4k$. By Lemma 10.2, except with probability $1/16$, the restriction map

$$\alpha \mapsto \alpha|_H$$

is injective on $\mathrm{supp}(\hat{f})$. Consequently, each nonzero $\gamma \in W$ corresponds to at most one $\alpha \in \mathrm{supp}(\hat{f})$, and there are at most k such γ. Consequently $\widehat{f_a}(\gamma) = \hat{f}(\alpha)\, \chi_\alpha(a)$,

Let $e_1, \ldots, e_n$ denote the standard unit vectors in $\mathbb{F}_2^n$, and let $e_0 = \mathbf{0}$. For each $i \in \{0, 1, \ldots, n\}$, define the coset $A_i = e_i + H$.

Note that

$$\frac{\widehat{f_{e_i}}(\gamma)}{\widehat{f_0}(\gamma)} = \chi_\alpha(e_i) = (-1)^{\langle \alpha, e_i \rangle} = (-1)^{\alpha_i}.$$

Thus, defining

$$b_i := \begin{cases} 0 & \text{if } \dfrac{\widehat{f_{e_i}}(\gamma)}{\widehat{f_0}(\gamma)} = +1, \\ 1 & \text{if } \dfrac{\widehat{f_{e_i}}(\gamma)}{\widehat{f_0}(\gamma)} = -1, \end{cases}$$

we recover $\alpha_i = b_i$ for each $i \in [n]$. Therefore, $\alpha = (b_1, \ldots, b_n)$ is recovered coordinate-wise, without requiring any linear algebra. Finally, the Fourier coefficient is obtained as

$$\hat{f}(\alpha) = \widehat{f_0}(\gamma).$$

It is easy to see that the total number of queries is $(n + 1) \cdot 2^d = \tilde{O}(nk^2)$, and the overall running time is $\tilde{O}(nk^2)$. □

A few remarks are in order.

- **Non-adaptive and parallelizable structure.** Our algorithm is entirely non-adaptive. To the best of our knowledge, this is the first non-adaptive algorithm in the sparse Hadamard transform literature that achieves both polynomial-time complexity and polynomial query complexity in n and k. This non-adaptive structure makes the algorithm particularly attractive for hardware implementation, since all queries can be generated in advance and executed fully in parallel. Moreover, the reconstruction procedure naturally decomposes across the different cosets: the Fourier spectrum corresponding to each of the $n+1$ cosets can be recovered independently using separate FFT computations. Consequently, the entire recovery process admits a highly parallelizable architecture, making it especially suitable for efficient hardware and distributed implementations.

- **Compressed sensing based variant.** Since the restriction of a k-Fourier-sparse function to any coset remains k-Fourier-sparse, one may alternatively reconstruct the Fourier spectrum of each restricted function f_A using standard compressed sensing techniques. In particular, by exploiting the restricted isometry property of Fourier–Hadamard matrices [Haviv and Regev, 2016], one can recover the spectrum of each f_A by solving an ℓ_1-minimization problem using $\widetilde{O}(k)$ samples and $\widetilde{O}(k^2)$ running time [Candès and Tao, 2006]. Repeating this independently for each of the $n+1$ cosets yields an overall sample complexity of $\widetilde{O}(nk)$ and total running time $\widetilde{O}(nk^3)$. This modified approach preserves all the advantages of our framework (with slightly complex computing) while achieving near-optimal sample complexity at the expense of higher running time.

In the following section, however, we present our main algorithm, which achieves both sample complexity $\widetilde{O}(nk)$ and running time $\widetilde{O}(nk)$.

10.2 The Main Algorithm

In this section, we present an improved version of the sparse Fourier recovery algorithm (Algorithm 17). The overall approach follows the same high level idea as Algorithm 16. However, in order to reduce complexity, we restrict the function to a significantly smaller random subspace of size $O(k)$, instead of $O(k^2)$ as in Algorithm 16. In Algorithm 16, the subspace is chosen large enough to ensure that each coset indexed by $\gamma \in W$ contains at

most one nonzero Fourier coefficient. Consequently, the ratio

$$g_\gamma(x) := \frac{\widehat{f}_x(\gamma)}{\widehat{f}_0(\gamma)}$$

behaves as a linear function over $\mathbb{F}_2^n$. This allows direct recovery of the corresponding frequency using coordinate wise reconstruction. This significantly simplifies the recovery procedure, but incurs higher query and time complexity.

In contrast, we now work with a much smaller subspace. As a result, collisions may occur, meaning that multiple Fourier coefficients may hash to the same γ. Nevertheless, even in this regime, one can show that a constant fraction of cosets remain collision free. To exploit this, we adopt the *iterative peeling* paradigm, which is widely used in the sparse Hadamard transform literature. In each round, we identify and recover a constant fraction of the nonzero Fourier coefficients, and subtract their contribution from the function:

$$\tilde{f} = f - \sum_{\xi \in S} \hat{f}(\xi)\, \chi_\xi,$$

where S denotes the set of already recovered frequencies. Repeating this process on $\tilde{f}$, it follows that after $O(\log k)$ rounds, all nonzero Fourier coefficients are recovered.

There is, however, a major obstacle that remains. The reconstruction step from Algorithm 16 applies only when g_γ is linear, which happens only if the coset $\gamma + H^\perp$ contains at most one nonzero Fourier coefficient. Therefore, before applying the reconstruction step, we must certify that g_γ is linear. A exact certification of linearity would be too expensive in terms of query complexity. However, if we allow a small amount of error and are willing to accept functions that are close to linear, then there are query efficient procedures for this task. A natural approach is to use the linearity test of Blum et al. [1993a] . It is known that if a function is ε far from linear, then the BLR test rejects with probability $\Omega(\varepsilon)$. In our setting, we take $\varepsilon = 1/4$. Therefore, a constant number of BLR tests suffices to distinguish the case when g_γ is linear from the case when it is far from linear.

It remains to modify the reconstruction procedure so that it is robust to small amounts of noise. We design a robust variant of the reconstruction step which tolerates deviations from exact linearity. In particular, we show that even if g_γ is only $(3/4)$ close to some linear function, the corresponding Fourier coefficient can still be correctly recovered. This is achieved using majority decoding over multiple queries.

Algorithm 17: EfficientSparseFourierRecovery

Input: Query access to $f : \mathbb{F}_2^n \to \mathbb{R}$; sparsity k.

Output: All pairs $(\alpha, \widehat{f}(\alpha))$ for $\alpha \in \mathrm{supp}(\widehat{f})$.

Initialization: $S \leftarrow \emptyset, \quad f^{(0)} \leftarrow f, \quad d^{(0)} \leftarrow k, \quad r \leftarrow 0.$

Repeat until $|S| = k$:

Step 1. Sample a subspace $H \leq \mathbb{F}_2^n$ uniformly with $\dim(H) = \log(100\, d^{(r)})$, and set $B \leftarrow \emptyset$.

Step 2. For $j = 1, \ldots, C \log k$: sample $\mathbf{a}, \mathbf{b} \leftarrow \mathbb{F}_2^n$ and compute $\widehat{f^{(r)}_{\mathbf{c}+H}}(*)$ for $\mathbf{c} \in \{0, \mathbf{a}, \mathbf{b}, \mathbf{a}+\mathbf{b}\}$; add to B all γ for which

$$\widehat{f^{(r)}_{\mathbf{a}+H}}(\gamma) \cdot \widehat{f^{(r)}_{\mathbf{b}+H}}(\gamma) \neq \widehat{f^{(r)}_{\mathbf{a}+\mathbf{b}+H}}(\gamma) \cdot \widehat{f^{(r)}_{H}}(\gamma).$$

Step 3. Set $G \leftarrow W \setminus B$.

Step 4. For $j = 1, \ldots, C'(\log n + \log k)$: sample $\mathbf{a} \leftarrow \mathbb{F}_2^n$ and compute $\widehat{f^{(r)}_{\mathbf{c}+H}}(*)$ for $\mathbf{c} \in \{\mathbf{a}, \mathbf{a}+\mathbf{e}_1, \ldots, \mathbf{a}+\mathbf{e}_n\}$; then for each $\gamma \in G$, record

$$\alpha_i^{(j)} \leftarrow \frac{1 - \widehat{f^{(r)}_{\mathbf{a}+\mathbf{e}_i+H}}(\gamma) \,/\, \widehat{f^{(r)}_{\mathbf{a}+H}}(\gamma)}{2} \qquad \text{for all } i \in [n].$$

Step 5. For each $\gamma \in G$, set $\alpha \leftarrow \left(\mathrm{maj}_j[\alpha_i^{(j)}]\right)_{i\in[n]}$ and $\widehat{f}(\alpha) \leftarrow \widehat{f^{(r)}_H}(\gamma)$, and update $S \leftarrow S \cup \{(\alpha, \widehat{f}(\alpha))\}$.

Step 6. Update the residual and parameters:

$$f^{(r+1)} \leftarrow f^{(r)} - \sum_{\alpha \in S} \widehat{f}(\alpha)\, \chi_\alpha, \qquad d^{(r+1)} \leftarrow \tfrac{9}{10}\, d^{(r)}, \qquad r \leftarrow r+1.$$

Step 7. Return $\{(\alpha, \widehat{f}(\alpha))\}_{\alpha \in S}$.

By the lemma, in expectation at least a $99/100$ fraction of the support elements of $f^{(r)}$ are isolated under the random hashing induced by H. However, the remaining buckets may contain multiple Fourier coefficients, and these must be filtered out before attempting recovery.

Singleton Detection via Linearity Testing. The key challenge is to distinguish singleton buckets from those containing multiple Fourier coefficients. We achieve this using

query efficient tests that filter out non-singleton buckets. For $\gamma \in H^{\perp}$, define

$$\phi_\gamma(a) = \frac{\widehat{f^{(r)}_{a+H}}(\gamma)}{\widehat{f^{(r)}_{H}}(\gamma)}, \quad a \in \mathbb{F}_2^n,$$

whenever the denominator is nonzero. Using the restriction formula,

$$\widehat{f^{(r)}_{a+H}}(\gamma) = \sum_{\alpha \in (\gamma + H^{\perp}) \cap \mathrm{supp}(\widehat{f^{(r)}})} \widehat{f^{(r)}}(\alpha)\, \chi_\alpha(a).$$

If the bucket corresponding to γ is a singleton containing α^*, then $\phi_\gamma(a) = \chi_{\alpha^*}(a)$, which is a linear function over $\mathbb{F}_2^n$. Otherwise, ϕ_γ is a sum of multiple characters and hence non-linear. Since exact linearity testing can be expensive in terms of query complexity, we instead aim to distinguish linear functions from those that are at least $1/4$-far from linear. To this end, we apply the Blum–Luby–Rubinfeld linearity test [Blum et al., 1993a] to ϕ_γ, which checks whether

$$\phi_\gamma(a)\phi_\gamma(b) = \phi_\gamma(a+b)$$

holds for uniformly random $a, b \in \mathbb{F}_2^n$.

Lemma 10.4 (Blum et al. [1993a]). *There exists an absolute constant $c > 0$ such that:*

1. *If ϕ_γ is linear, then it passes every BLR test.*
2. *If ϕ_γ is at least $1/4$-far from every linear function, then a single BLR test rejects with probability at least c.*

Consequently, by repeating the BLR test $O(1)$ times and discarding any γ that fails at least once, we obtain a set G such that, with high constant probability, every $\gamma \in G$ corresponds to a singleton bucket. By amplifying this procedure and repeating it $C \log k$ times, it follows that with probability at least $1 - k^{-2}$, all $\gamma \in G$ correspond to (at least 3/4 close to) singleton buckets.

Robust Decoding from Approximate Linearity. For each $\gamma \in G$, we recover the corresponding α coordinate-wise using bit extraction. The following lemma shows that recovery succeeds even when ϕ_γ is only approximately linear.

Lemma 10.5. *Suppose there exists $\alpha^* \in \mathbb{F}_2^n$ such that*

$$\Pr_{a \sim \mathbb{F}_2^n} \left[\phi_\gamma(a) \neq \chi_{\alpha^*}(a)\right] \leq \varepsilon$$

for some $\varepsilon < \frac{1}{4}$. Let $a^{(1)}, \dots, a^{(J)}$ be independent uniform samples from $\mathbb{F}_2^n$, where $J = C(\log n + \log k)$ for a sufficiently large constant C. For each $i \in [n]$, define

$$\alpha_i^{(j)} = \frac{1 - \widehat{f^{(r)}_{a^{(j)}+e_i+H}}(\gamma) / \widehat{f^{(r)}_{a^{(j)}+H}}(\gamma)}{2}, \quad \hat{\alpha}_i = \mathrm{majority}\big(\alpha_i^{(1)}, \dots, \alpha_i^{(J)}\big).$$

Then for all $\gamma \in G$ and all $i \in [n]$, $\Pr\left[\hat{\alpha}_i \neq \alpha_i^*\right] \leq (nk)^{-2}$. Consequently, except with probability $o(1)$, all coordinates of all recovered frequencies are correct.

Proof. Observe that

$$\frac{\widehat{f^{(r)}_{a+e_i+H}}(\gamma)}{\widehat{f^{(r)}_{a+H}}(\gamma)} = \frac{\phi_\gamma(a+e_i)}{\phi_\gamma(a)}.$$

If both $\phi_\gamma(a) = \chi_{\alpha^*}(a)$ and $\phi_\gamma(a+e_i) = \chi_{\alpha^*}(a+e_i)$ hold, then

$$\frac{\phi_\gamma(a+e_i)}{\phi_\gamma(a)} = \frac{\chi_{\alpha^*}(a+e_i)}{\chi_{\alpha^*}(a)} = (-1)^{\alpha_i^*},$$

and hence $\alpha_i^{(j)} = \alpha_i^*$. Therefore, an error occurs only if either a or $a+e_i$ is a disagreement point, which by a union bound happens with probability at most $2\varepsilon < \frac{1}{2}$. Thus each trial is correct with probability at least $1 - 2\varepsilon > \frac{1}{2}$. By Hoeffding's inequality,

$$\Pr\left[\hat{\alpha}_i \neq \alpha_i^*\right] \leq \exp(-\Omega(J)).$$

Choosing $J = C(\log n + \log k)$ for a sufficiently large constant C ensures that

$$\Pr\left[\hat{\alpha}_i \neq \alpha_i^*\right] \leq (nk)^{-2}.$$

Finally, applying a union bound over all $i \in [n]$ and all $\gamma \in G$ yields the claim. □

Peeling and Progress. After each round, the algorithm updates

$$f^{(r+1)} = f^{(r)} - \sum_{\alpha \in S^{(r)}} \hat{f}(\alpha)\, \chi_\alpha,$$

where $S^{(r)}$ denotes the set of frequencies recovered in round r. This removes exactly those Fourier coefficients, and hence

$$d^{(r+1)} = d^{(r)} - |S^{(r)}|.$$

By the isolation guarantee, in expectation a $99/100$ fraction of the support elements are

isolated, and are recovered with high probability after filtering and decoding. Thus, with high probability,

$$|S^{(r)}| \geq \tfrac{99}{100}\, d^{(r)}, \quad \text{and hence} \quad d^{(r+1)} \leq \tfrac{1}{100}\, d^{(r)}.$$

Iterating this recurrence gives $d^{(r)} \leq k \cdot \left(\frac{1}{100}\right)^r$. Therefore, after $O(\log k)$ rounds, all nonzero Fourier coefficients are recovered.

Query and Time Complexity. In each round, the algorithm performs:

- $O(\log k)$ evaluations for BLR filtering across all buckets,
- $O(n \cdot (\log n + \log k))$ evaluations for robust recovery.

Thus, the total number of evaluations per round is $O(n \cdot (\log n + \log k))$ and each evaluation of $\widehat{f^{(r)}_{c+H}}(\cdot)$ requires $|H| = \Theta(d^{(r)})$ queries. Then query complexity per round is $O\big(d^{(r)} \cdot n \cdot (\log n + \log k)\big)$. Summing over all rounds, using $d^{(r)} \leq k/100^r$, we obtain

$$Q(n,k) = \sum_r O\big(d^{(r)} \cdot n \cdot (\log n + \log k)\big) = O(kn \cdot (\log n + \log k)).$$

Each evaluation involves a Walsh–Hadamard transform over the subspace H, incurring a cost of $O(d^{(r)} \log d^{(r)})$. Aggregating over all evaluations and rounds yields total runtime

$$T(n,k) = O(kn \log k \cdot (\log n + \log k)).$$

10.3 Discussion

In both Algorithm 17 and Algorithm 16, if, in addition to being k-Fourier-sparse, the function is known to have Fourier degree at most d, that is, every nonzero Fourier coefficient $\alpha \in \mathbb{F}_2^n$ satisfies $|\alpha| \leq d$, then the recovery step can be implemented significantly more efficiently.

We illustrate the idea within the framework of Algorithm 16. Recall that, once an isolated Fourier coefficient is identified, the corresponding support vector α is recovered by querying all shifts $e_1, \ldots, e_n$. Indeed, for each coordinate i,

$$\frac{\widehat{f_{e_i}}(\gamma)}{\widehat{f}_0(\gamma)} = (-1)^{\alpha_i},$$

which reveals the value of α_i. This requires n queries per recovered coefficient.

However, when α is known to be d-sparse, one can instead view the problem as recovering a d-sparse binary vector from random linear measurements. Specifically, let

$$A \in \mathbb{F}_2^{m \times n}, \qquad m = O(d \log n),$$

be a random binary matrix. Standard arguments imply that, with high probability, every d-sparse vector is uniquely determined by the measurements

$$A\alpha \in \mathbb{F}_2^m.$$

For each row $a_j \in \mathbb{F}_2^n$ of A, query the shift corresponding to a_j. Then,

$$\frac{\widehat{f_{a_j}}(\gamma)}{\widehat{f_0}(\gamma)} = (-1)^{\langle \alpha, a_j \rangle},$$

and hence each query reveals one linear measurement of the unknown sparse vector α. Since $m = O(d \log n)$ such measurements suffice to identify any d-sparse vector uniquely with high probability, the support vector α can be recovered using only $O(d \log n)$ queries instead of n. Consequently, the overall query and time complexity of Algorithm 16 improves from roughly $O(k^2 n)$ to $O(k^2 d \log n)$. Using the same idea, one obtains a similar improvement for Algorithm 17 as well to recover the bounds of Amrollahi et al. [2019].

Finally, we note that one natural direction for future work is to extend our algorithm to the setting of approximately Fourier-sparse functions. Another interesting direction is to generalize the approach beyond functions over $\mathbb{F}_2^n$ to functions of the form $[N] \to \mathbb{C}$, where typically $N = 2^n$. In fact, the main structural insight underlying our approach, namely the implicit hashing of the Fourier spectrum, continues to hold in this more general setting. A more ambitious direction would be to extend these techniques to higher-dimensional domains such as functions over $[n]^d$, thereby bringing the framework closer to the sparse Fourier transform setting studied by Kapralov [2017], Li and Nakos [2019]. Finally, it would be interesting to understand whether similar ideas can be used for sparse recovery problems over more general finite abelian groups or other structured transform domains.

Part III

A Communication Complexity Perspective

11

Communication Complexity of Linear Isomorphism Problem

Communication complexity, introduced by Yao [Yao, 1979], studies the amount of communication required between two parties, traditionally called Alice and Bob, to compute a function whose input is distributed between them. More precisely, Alice receives an input $x \in \mathbb{F}_2^n$ and Bob receives an input $y \in \mathbb{F}_2^n$, and their goal is to compute a known function $f : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \{0,1\}$ by exchanging bits. The minimum number of bits that must be exchanged in order to compute f is the *communication complexity* of f.

Depending on whether Alice and Bob have access to randomness, this model admits two important variants. In the *deterministic* setting, every message sent by the parties is fully determined by their inputs and the communication history, and the protocol must always output the correct value of $f(x, y)$. In the *randomized* setting, the parties may use random bits and are allowed a small probability of error, typically at most $1/3$. Randomized protocols are further classified by the nature of the available randomness: in the *public-coin* setting the parties share a common random string, whereas in the *private-coin* setting each party has access only to its own private randomness. The distinction between deterministic and randomized communication complexity is fundamental, since the use of randomness—and its nature, whether public or private—can significantly affect the amount of communication required, sometimes leading to an exponential gap.

Over the years, several classical problems have shaped the communication complexity literature. A prominent example is the *Equality* problem ([Yao, 1979]), where Alice and Bob are given strings $x, y \in \mathbb{F}_2^n$ and must determine whether $x = y$. Another canonical problem is *Set Disjointness*, in which Alice and Bob receive strings $x, y \in \mathbb{F}_2^n$ representing subsets of a universe and must determine whether the corresponding sets intersect. Other widely studied problems include *Inner Product*, *Gap Hamming Distance*, *Indexing*, and *Pointer Chasing*, each capturing different structural aspects of distributed computation. We refer the reader to [Kushilevitz and Nisan, 1997, Rao and Yehudayoff, 2020] for an exposition of the literature. Beyond its intrinsic theoretical interest, a major motivation for studying communication complexity is its surprising ability to yield lower bounds in

other areas of computer science, e.g. circuit complexity [Karchmer and Wigderson, 1988], data structures [Miltersen et al., 1995], streaming algorithms [Alon et al., 1999], property testing [Blais and O'Donnell, 2010b], and distribution testing [Blais et al., 2019b].

Among these, the *Equality* problem stands out for its simplicity. Any deterministic protocol requires $\Omega(n)$ bits of communication, since in the worst case Bob must verify every bit of Alice's input. In contrast, randomized protocols solve the problem much more efficiently [Yao, 1979]: with public coins, Equality admits an $O(1)$-bit protocol using simple fingerprinting techniques that detect inequality with high probability, while in the private-coin model the best known protocols require $O(\log n)$ bits, reflecting the additional cost of simulating shared randomness.

In this work we consider a linear-algebraic generalization of Equality, in which the inputs to Alice and Bob are Boolean functions rather than bit strings and equality is relaxed to equivalence under invertible linear transformations of the domain. Given Boolean functions $f, g : \mathbb{F}_2^n \to \{-1, +1\}$ held by Alice and Bob respectively, they must determine whether f and g are *linearly isomorphic*: whether there exists a matrix $M \in \mathrm{GL}_n(\mathbb{F}_2)$ such that

$$g(x) = f(Mx) \qquad \text{for all } x \in \mathbb{F}_2^n,$$

where $\mathrm{GL}_n(\mathbb{F}_2)$ denotes the group of invertible $n \times n$ matrices over $\mathbb{F}_2$. Linear isomorphism of Boolean functions arises across a wide range of areas, e.g. the design of efficient combinatorial circuits [Ciriani, 2003, Bemporad, 2004, Yordanov et al., 2012], the construction of error-correcting codes [Dautovic and Novak, 2006, Duursma et al., 2001, Hou, 1994], and cryptographic primitives [Cui and Cao, 2007, Dunkelman et al., 2012]. A notable example is provided by Reed–Muller codes, which are invariant under affine transformations of their input variables; this symmetry plays a key role in the development of fast decoding algorithms [Abbe et al., 2021]. The problem is also fundamental in multivariate public-key cryptography, where security often relies on the hardness of determining whether two systems of multivariate polynomials are equivalent under affine or linear transformations [Faugère and Perret, 2006].

Related work.

A restricted and widely studied version of the linear isomorphism problem is the *Boolean function permutation isomorphism* problem. Given $f, g : \mathbb{F}_2^n \to \{-1, +1\}$, the task is to determine whether one function can be obtained from the other by permuting the input

variables—that is, whether there exists a permutation π on $[n]$ such that

$$f(x_1, x_2, \ldots, x_n) = g(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(n)}).$$

Here the functions are typically given explicitly, for instance as Boolean circuits or formulas. In the standard algorithmic model this problem is computationally difficult: even when the functions are represented as DNF formulas, it is coNP-hard via a reduction from CNF-SAT. At the same time, it is known to lie in Σ_2^p, though it is not known whether it belongs to coNP. Agrawal and Thierauf [Agrawal and Thierauf, 2000] showed that the problem cannot be Σ_2^p-complete unless the polynomial hierarchy collapses to Σ_3^p. The fastest known algorithms proceed via a reduction to hypergraph isomorphism. Using Luks' algorithm [Luks, 1999], which tests hypergraph isomorphism on n vertices and m edges in time $2^{O(n)}m^{O(1)}$, this reduction yields a $2^{O(n)}$-time algorithm. Because exact permutation isomorphism appears to be as difficult as hypergraph isomorphism, improving the $2^{O(n)}$ bound has remained notoriously hard, which has motivated the study of relaxed variants—for example, asking whether two functions become *approximately isomorphic* after permuting their variables, or whether additional structure (small circuit size, bounded depth) can be exploited to design faster algorithms. Several works [Arvind and Vasudev, 2014, Simon, 2016] have studied such questions.

Permutation isomorphism has also been investigated in the *property testing* framework, where one has full knowledge of g and oracle access to an unknown f, and must determine whether some permutation of the variables makes f and g identical or whether they differ on a large fraction of inputs. Alon and Blais [2010] showed that when the known function is chosen uniformly at random over n variables, the task requires $\Omega(n)$ queries. Blais et al. [2015] and Chakraborty et al. [2011a] independently showed that if the known function g is a k-junta, then testing isomorphism to g can be done with $O(k \log k)$ queries, and Blais et al. [2012] established a nearly matching lower bound. We refer the reader to [Alon et al., 2013b] for a survey of this line of work.

Boolean function *linear* isomorphism has likewise been studied in the property testing framework. Wimmer and Yoshida [2013] showed that if the known function g has small spectral norm, one can test whether it is linearly isomorphic to an unknown f (given via oracle access), or far from every such isomorphism, using a number of queries polynomial in the spectral norm of g; their result gave an upper bound of $\tilde{O}(m^{32})$ queries and a lower bound of $\Omega(\log m)$. More recently, Datta et al. [Datta et al., 2026a] improved the upper bound to $\tilde{O}(m^4)$ and strengthened the lower bound to $\Omega(m^2)$, nearly closing the gap.

A computational framework closely related to the query model is communication complexity, which is in fact strictly stronger in many settings; once a problem is well understood

in the query model, it is natural to study it in the communication setting. A slight variant of linear isomorphism was studied in this model by [Grigorescu et al., 2013]: Alice is given $f : \mathbb{F}_2^n \to \{0,1\}$ and Bob is given $g : \mathbb{F}_2^n \to \{0,1\}$, and their goal is to determine whether $f \oplus g$ is linearly isomorphic to a fixed function $h : \mathbb{F}_2^n \to \{0,1\}$ known to both parties, using shared randomness and minimizing communication. Somewhat surprisingly, this yielded an $\Omega(n^2)$ lower bound for testing linear isomorphism when h is the inner product function. Subsequently, Datta et al. [Datta et al., 2026a] obtained exponentially stronger lower bounds when the target function has small spectral norm, using a similar communication framework.

In this work we study linear isomorphism in a more natural communication setting: Alice and Bob are given functions f and g respectively, and their goal is to determine whether f and g are linearly isomorphic to each other or far from being so. As in the query setting, we restrict attention to functions with small spectral norm. We emphasize that the interplay between communication complexity and Fourier analysis has been studied extensively and has yielded several deep insights; see, for example, [Tsang et al., 2013, Mande, 2018].

LinIso_n: Communication Problem

Input. Alice is given a Boolean function $f : \mathbb{F}_2^n \to \{-1,+1\}$, and Bob is given a Boolean function $g : \mathbb{F}_2^n \to \{-1,+1\}$. Let $\epsilon \in [0,1]$ and $\omega \in (0,1)$.
Promise. One of the following holds:

$$\text{(i) } \exists\, M \in \mathrm{GL}_n(\mathbb{F}_2), \quad \Pr_{x \sim \mathbb{F}_2^n}[f(Mx) \neq g(x)] \leq \epsilon;$$

$$\text{(ii) } \forall\, M \in \mathrm{GL}_n(\mathbb{F}_2), \Pr_{x \sim \mathbb{F}_2^n}[f(Mx) \neq g(x)] \geq \epsilon + \omega.$$

Goal. Alice and Bob must determine which case holds while minimizing the number of communicated bits.

Figure 11.1: Problem statement.

Problem definition. This problem has remained unexplored in communication complexity, and in this work we take a first step toward addressing that gap. Our focus is the gap version of the Boolean function linear isomorphism problem shown in Figure 11.1. We study the communication complexity of LinIso_n under two widely studied variants. In the deterministic setting, Alice and Bob must decide whether f and g are linearly isomorphic using a deterministic protocol, while minimizing the number of bits exchanged. In the randomized setting, the parties may use public or private coins, and the protocol must succeed with probability at least $2/3$ on every input.

Main Theorems

Our results reveal that an analytic quantity associated with Boolean functions, the *approximate spectral norm*, plays a central role in the communication complexity of this problem. The approximate spectral norm measures how well a Boolean function can be approximated pointwise by a real-valued function with small spectral norm. It generalizes the standard spectral norm and can be significantly smaller in many cases [Cheung et al., 2024], providing a robust measure of the extent to which a function behaves like a low-complexity object from the perspective of Fourier analysis. Functions with low spectral norm admit sparse or well-concentrated Fourier expansions, which often correspond to low-complexity behavior across computational models. This connection is especially pronounced in communication complexity, where Fourier-analytic properties of Boolean functions, particularly for XOR functions [Mande, 2018], tightly control communication costs: functions with low (approximate) spectral norm often admit more efficient protocols, while large spectral norm is typically associated with higher communication complexity. Such connections have been explored extensively and have led to several deep insights into the role of structure in communication problems [Shpilka et al., 2017, Sanyal, 2019, Tsang et al., 2013].

Our first result establishes nearly tight bounds on the deterministic communication complexity of LinIso_n in terms of the approximate spectral norm parameter t.

Theorem 11.1 (Deterministic Communication Complexity)**.** *Let $f, g : \mathbb{F}_2^n \to \{-1, +1\}$ be Boolean functions held by Alice and Bob, respectively, and let $\epsilon \in [0, 1]$ and $\omega \in (0, 1)$. Define*

$$t := \min\left\{ \left\lceil \|\widehat{f}\|_{1,1/3} \right\rceil , \left\lceil \|\widehat{g}\|_{1,1/3} \right\rceil \right\}.$$

(a) There exists a deterministic communication protocol that uses at most $O(t^4 \log^2(1/\omega))$ bits of communication and decides whether

$$\exists M \in \mathrm{GL}_n(\mathbb{F}_2),\ \Pr_{x \sim \mathbb{F}_2^n}[f(Mx) \neq g(x)] \leq \epsilon$$
$$\text{or} \quad \forall M \in \mathrm{GL}_n(\mathbb{F}_2),\ \Pr_{x \sim \mathbb{F}_2^n}[f(Mx) \neq g(x)] \geq \epsilon + \omega.$$

(b) Moreover, any deterministic protocol solving this problem with $\epsilon = 0$ and $\omega = 1/3$ must communicate at least $\Omega(t^2)$ bits.

We begin our investigation of the randomized setting with access to public coins. When $\epsilon = 0$ and $\omega > 0$, the problem $\mathsf{LinIso}_n(f, g)$ admits a highly efficient public-coin randomized protocol using only $O(1)$ bits of communication. The key idea is to view the truth tables

of f and g as vectors in $\{-1,+1\}^{2^n}$. Without interacting with Bob, Alice computes a nonsingular matrix $A \in \mathbb{F}_2^{n\times n}$ such that the truth table of $f \circ A$ is lexicographically minimal; similarly, Bob independently computes a nonsingular matrix B for g. If f and g are linearly isomorphic, then $f \circ A = g \circ B$, and the problem reduces to checking equality of the transformed functions, which can be done with $O(1)$ bits of communication using standard public-coin protocols for Equality [Kushilevitz and Nisan, 1997, Chapter 3].

Consequently, we turn to randomized protocols with access to private coins. We show that with only private randomness there exists an efficient protocol for $\mathsf{LinIso}_n(f,g)$ in the non-tolerant setting. This protocol has communication complexity polynomial in the approximate spectral norm and achieves a quadratically better dependence on this parameter than our deterministic protocol. Furthermore, we prove that in the private-coin model some dependence on the approximate spectral norm is unavoidable, in contrast to the public-coin setting.

Theorem 11.2 (Randomized Communication Complexity)**.** *Let $f, g : \mathbb{F}_2^n \to \{-1,+1\}$ be Boolean functions held by Alice and Bob, respectively, and let $\epsilon \in [0,1]$ and $\omega \in (0,1)$. Define*

$$t := \min\left\{\left\lceil \|\widehat{f}\|_{1,1/3}\right\rceil, \left\lceil \|\widehat{g}\|_{1,1/3}\right\rceil\right\}.$$

(a) There exists a private-coin randomized communication protocol that uses at most $O(t^2/\omega^3)$ bits of communication and decides whether

$$\exists M \in \mathrm{GL}_n(\mathbb{F}_2),\ \Pr_{x\sim\mathbb{F}_2^n}[f(Mx) \neq g(x)] = 0$$
$$\text{or}\quad \forall M \in \mathrm{GL}_n(\mathbb{F}_2),\ \Pr_{x\sim\mathbb{F}_2^n}[f(Mx) \neq g(x)] \geq \omega.$$

(b) Moreover, any private-coin randomized protocol solving this problem with $\omega = 1/3$ must communicate at least $\Omega(\log t)$ bits.

Our results demonstrate that when the input functions have small approximate spectral norm, linear isomorphism admits efficient communication protocols. In both the deterministic and private-coin settings we obtain upper bounds with polynomial dependence on the approximate spectral norm and show that such dependence is unavoidable. It remains an intriguing open question whether small approximate spectral norm characterizes exactly the class of functions for which this problem can be solved efficiently.

Organization. Section 11.1 studies the deterministic communication complexity of linear isomorphism. Section 11.2 considers the randomized setting and presents our protocols in both the public-coin and private-coin models.

11.1 Deterministic setting

Deterministic communication complexity studies the minimum number of bits Alice and Bob must exchange to compute a function whose input is distributed between them, assuming a deterministic protocol. The objective is the worst-case communication cost required to compute the function exactly, without randomness or error. In this section we present a communication-efficient protocol for deciding linear isomorphism in this setting, along with a nearly matching lower bound. We restate Theorem 11.1 for convenience.

Theorem 11.1 (Deterministic Communication Complexity). *Let $f, g : \mathbb{F}_2^n \to \{-1, +1\}$ be Boolean functions held by Alice and Bob, respectively, and let $\epsilon \in [0, 1]$ and $\omega \in (0, 1)$. Define*

$$t := \min\left\{ \left\lceil \|\widehat{f}\|_{1,1/3} \right\rceil, \left\lceil \|\widehat{g}\|_{1,1/3} \right\rceil \right\}.$$

(a) There exists a deterministic communication protocol that uses at most $O(t^4 \log^2(1/\omega))$ bits of communication and decides whether

$$\exists M \in \mathrm{GL}_n(\mathbb{F}_2),\ \Pr_{x \sim \mathbb{F}_2^n}[f(Mx) \neq g(x)] \leq \epsilon$$
$$\text{or} \quad \forall M \in \mathrm{GL}_n(\mathbb{F}_2),\ \Pr_{x \sim \mathbb{F}_2^n}[f(Mx) \neq g(x)] \geq \epsilon + \omega.$$

(b) Moreover, any deterministic protocol solving this problem with $\epsilon = 0$ and $\omega = 1/3$ must communicate at least $\Omega(t^2)$ bits.

11.1.1 An efficient deterministic communication protocol

We now prove the upper bound of Theorem 11.1. As described above, Alice and Bob are given $f, g : \mathbb{F}_2^n \to \{-1, +1\}$ with the promise that either

$$\delta_{\mathcal{L}}(f, g) \leq \epsilon \quad \text{or} \quad \delta_{\mathcal{L}}(f, g) \geq \epsilon + \omega,$$

and the task is to distinguish the two cases. We present a deterministic protocol solving this problem in Algorithm 18.

We begin with a structural lemma for Boolean functions with small approximate spectral norm. It may be viewed as an approximate analogue of a classical result of Bruck and Smolensky [Bruck and Smolensky, 1992]: if f has small γ-approximate spectral norm, then f is close to the sign of a sparse Fourier polynomial.

Lemma 11.3 (Approximate Spectral Sampling). *Let $f : \mathbb{F}_2^n \to \{-1,+1\}$ be a Boolean function and let $\gamma \in [0,1)$ with $2\gamma < 1$. Then for every $\tau \in (0,1)$ there exist a subset $\mathcal{S} \subseteq \mathbb{F}_2^n$ and a Boolean function $F : \mathbb{F}_2^n \to \{-1,+1\}$ such that:*

- $|\mathcal{S}| = O\Big(\|\widehat{f}\|_{1,\gamma}^2 \cdot \frac{1}{\gamma^2} \cdot \log(1/\tau)\Big)$,
- $F(x) = \operatorname{sign}\big(\sum_{\alpha\in\mathcal{S}} a_\alpha \chi_\alpha(x)\big)$ *for some coefficients* $a_\alpha \in \{-1,+1\}$,
- $\delta(f,F) \le \tau$.

Our proof follows the argument of Bruck and Smolensky [Bruck and Smolensky, 1992]. The key idea is to approximate f by a small combination of parity functions. Suppose there is a real-valued function h that is pointwise close to f and whose Fourier representation has small ℓ_1-norm. The Fourier expansion of h can be viewed as a weighted average of parity functions, where the weight of each parity is proportional to the magnitude of its Fourier coefficient. Instead of using the entire expansion, we approximate this average by sampling a small number of parities according to these weights and averaging them; concentration bounds imply that the sampled average closely approximates h on almost all inputs. Since h is already close to f, the sampled approximation is also close to f, and taking its sign recovers f on all but a small fraction of inputs. We now give the formal proof.

Proof. [Proof of Lemma 11.3] Let $h : \mathbb{F}_2^n \to \mathbb{R}$ be a real-valued function such that $|f(x) - h(x)| \le \gamma$ for all $x \in \mathbb{F}_2^n$ and $\|\widehat{h}\|_1 = \|\widehat{f}\|_{1,\gamma}$. Write h via its Fourier expansion:

$$h(x) = \sum_{\alpha\in\mathbb{F}_2^n} \widehat{h}(\alpha)\chi_\alpha(x) = \|\widehat{h}\|_1 \cdot \sum_{\alpha\in\mathbb{F}_2^n} \left(\frac{|\widehat{h}(\alpha)|}{\|\widehat{h}\|_1} \cdot \operatorname{sign}\Big(\widehat{h}(\alpha)\Big)\, \chi_\alpha(x) \right).$$

This suggests a sampling-based approximation. Define a distribution $\mathcal{D}$ on $\mathbb{F}_2^n$ by

$$\Pr_{\alpha\sim\mathcal{D}}[\alpha = \beta] = \frac{|\widehat{h}(\beta)|}{\|\widehat{h}\|_1} \qquad \text{for all } \beta \in \mathbb{F}_2^n.$$

Then

$$h(x) = \mathop{\mathbb{E}}_{\alpha\sim\mathcal{D}}\Big[\|\widehat{h}\|_1 \cdot \operatorname{sign}\Big(\widehat{h}(\alpha)\Big)\, \chi_\alpha(x)\Big].$$

Sample $\alpha_1,\dots,\alpha_T \sim \mathcal{D}$ independently, where

$$T = O\left(\|\widehat{h}\|_1^2 \cdot \frac{1}{\beta^2} \cdot \log(1/\tau)\right),$$

and consider the empirical approximator

$$\tilde{h}(x) := \frac{\|\widehat{h}\|_1}{T} \sum_{i=1}^{T} \operatorname{sign}(\widehat{h}(\alpha_i))\chi_{\alpha_i}(x).$$

By Hoeffding's inequality, for any fixed $x \in \mathbb{F}_2^n$,

$$\Pr_{(\alpha_1,\ldots,\alpha_T)\sim\mathcal{D}^{\otimes T}} \left[\left| h(x) - \tilde{h}(x) \right| > \beta \right] \leq \tau. \tag{11.1}$$

Set $\beta := \gamma$ and define

$$\Delta = \Delta(\alpha_1, \ldots, \alpha_T) := \left\{ x \in \mathbb{F}_2^n : |h(x) - \tilde{h}(x)| > \beta \right\},$$

so that $|\Delta| = \sum_{x\in\mathbb{F}_2^n} \mathbf{1}\Big[|h(x) - \tilde{h}(x)| > \beta\Big]$. For all $x \notin \Delta$, the triangle inequality gives

$$|f(x) - \tilde{h}(x)| \leq |f(x) - h(x)| + |h(x) - \tilde{h}(x)| \leq \beta + \gamma = 2\gamma < 1.$$

Since $f(x) \in \{-1, +1\}$, this implies $\operatorname{sign}(\tilde{h}(x)) = f(x)$ for all $x \notin \Delta$. Define the final hypothesis $F(x) := \operatorname{sign}(\tilde{h}(x))$. Then

$$\delta(f, F) \leq 2^{-n}|\Delta|. \tag{11.2}$$

Combining (11.1) and (11.2),

$$\begin{aligned}
\mathbb{E}_{(\alpha_1,\ldots,\alpha_T)\sim\mathcal{D}^{\otimes T}} [\delta(f, F)] &\leq 2^{-n}\mathbb{E}_{(\alpha_1,\ldots,\alpha_T)\sim\mathcal{D}^{\otimes T}} [|\Delta|] \\
&= 2^{-n} \sum_{x\in\mathbb{F}_2^n} \Pr_{(\alpha_1,\ldots,\alpha_T)\sim\mathcal{D}^{\otimes T}} \left[\left| h(x) - \tilde{h}(x) \right| > \beta \right] \\
&\leq \tau.
\end{aligned}$$

This concludes the proof. □

We now prove Theorem 11.1 (a).

Proof. [Proof of Theorem 11.1 (a)] The proof follows from the analysis of Algorithm 18. At a high level, the protocol first determines which of f or g has the smaller approximate spectral norm; without loss of generality assume this is Alice's function f. Alice then transmits a succinct representation of a function close to f. Using Lemma 11.3, Alice constructs a Boolean function F whose Fourier support lies in a low-dimensional subspace and which approximates f well in linear distance, and sends its truncated spectral representation to Bob. Bob then evaluates F on sampled inputs and estimates the linear distance

between F and g. Since F is close to f, the triangle inequality ensures that $\delta_{\mathcal{L}}(F, g)$ differs significantly between the cases $\delta_{\mathcal{L}}(f, g) \leq \epsilon$ and $\delta_{\mathcal{L}}(f, g) \geq \epsilon + \omega$; estimating $\delta_{\mathcal{L}}(F, g)$ with small additive error therefore lets Bob distinguish them, and the total communication is bounded by the size of the transmitted representation.

Correctness. We first recall that the triangle inequality holds for linear distance: for all $f, g, h : \mathbb{F}_2^n \to \{-1, +1\}$,

$$\delta_{\mathcal{L}}(f, h) + \delta_{\mathcal{L}}(h, g) \geq \delta_{\mathcal{L}}(f, g).$$

We invoke the protocol with parameters $\tau = \omega/4$ and $\gamma = 1/3$, and analyze the two cases.

- **Case 1:** $\delta_{\mathcal{L}}(f, g) \leq \epsilon$. By the triangle inequality,

$$\delta_{\mathcal{L}}(F, g) \leq \delta_{\mathcal{L}}(f, g) + \delta_{\mathcal{L}}(f, F) \leq \epsilon + \frac{\omega}{4}.$$

- **Case 2:** $\delta_{\mathcal{L}}(f, g) \geq \epsilon + \omega$. Then

$$\delta_{\mathcal{L}}(F, g) \geq \delta_{\mathcal{L}}(f, g) - \delta_{\mathcal{L}}(f, F) \geq \epsilon + \omega - \frac{\omega}{4} = \epsilon + \frac{3\omega}{4}.$$

Since $\epsilon + \frac{3\omega}{4} > \epsilon + \frac{\omega}{4}$, Bob can distinguish the two cases by estimating $\delta_{\mathcal{L}}(F, g)$ up to additive error less than $\omega/10$. Thus the protocol is correct.

Communication cost. Communication occurs only in Step 1 and Step 4 of the protocol.

- In Step 1, Alice and Bob compare their approximate spectral norms to identify which is smaller, using $O(\log t)$ bits.

- In Step 4, Alice transmits the truncated spectral representation of her function. By Lemma 11.3, the support set $\mathcal{S}$ satisfies $|\mathcal{S}| = O(t^2 \log(1/\omega))$, and its span has dimension $\ell = O(t^2 \log(1/\omega))$. The initial ℓ vectors spanning this space can each be represented using $O(\ell)$ bits, and each subsequent vector of $\mathcal{S}$ lies in the span and can be described by its coordinates in this basis using ℓ bits. The total communication in this step is therefore

$$O(\ell + |\mathcal{S}| \cdot \ell) = O\big(t^4 \log^2(1/\omega)\big)\,.$$

Hence the total communication cost of the protocol is $O(t^4 \log^2(1/\omega))$. □

Remark 11.4. The proof of Theorem 11.1 also yields a deterministic protocol for deciding whether $\delta_{\mathsf{affine}}(f,g) \le \epsilon$ or $\delta_{\mathsf{affine}}(f,g) \ge \epsilon + \omega$ with the same communication cost. This follows from a simple modification of Step 5 of Algorithm 18, where Bob checks g against $F \circ (M,a)$ for all $M \in \mathrm{GL}_n(\mathbb{F}_2)$ and $a \in \mathbb{F}_2^n$ (with the same choice of parameters).

Algorithm 18: Deterministic Protocol for Deciding Linear Isomorphism

Input. Alice is given $f : \mathbb{F}_2^n \to \{-1,+1\}$, and Bob is given $g : \mathbb{F}_2^n \to \{-1,+1\}$, with the promise that either $\delta_{\mathcal{L}}(f,g) \le \epsilon$ or $\delta_{\mathcal{L}}(f,g) \ge \epsilon + \omega$.

Step-1: Alice and Bob exchange bits alternately to determine whether $\lceil \|\widehat{f}\|_{1,1/3} \rceil \le \lceil \|\widehat{g}\|_{1,1/3} \rceil$. Without loss of generality, assume $\lceil \|\widehat{f}\|_{1,1/3} \rceil \le \lceil \|\widehat{g}\|_{1,1/3} \rceil$.

Step-2: Alice constructs

$$F(x) := \mathrm{sign}\Big(\sum_{\alpha \in \mathcal{S}} a_\alpha \chi_\alpha(x) \Big),$$

where $|\mathcal{S}| = O(\|\widehat{f}\|_{1,1/3}^2 \log(1/\omega))$ and $a_\alpha \in \{-1,+1\}$ for all $\alpha \in \mathcal{S}$, such that $\delta(f,F) \le \omega/4$ (Lemma 11.3 guarantees existence).

Step-3: Since $\dim(\mathrm{span}(\mathcal{S})) \le |\mathcal{S}|$, Alice finds $M \in \mathrm{GL}_n(\mathbb{F}_2)$ so that

$$(F \circ M)(x) = \mathrm{sign}\left(\sum_{i=1}^{\ell} a_{e_i} \chi_{e_i}(x) + \sum_{j=1}^{w} a_{W_j} \chi_{W_j}(x) \right),$$

where $\ell + w = |\mathcal{S}|$, $W_j \in \mathrm{span}\{e_1, \ldots, e_\ell\}$, and the e_i are standard basis vectors.

Step-4: Alice sends Bob:

- the integer ℓ and the coefficients $a_{e_1}, \ldots, a_{e_\ell}$, using $O(|\mathcal{S}|)$ bits; and
- if $\ell < |\mathcal{S}|$, each W_j (as an ℓ-bit string) together with a_{W_j}.

Step-5: Bob constructs the function F.

- If $\delta_{\mathcal{L}}(F,g) \le \epsilon + \omega/4$, the protocol accepts.
- If $\delta_{\mathcal{L}}(F,g) \ge \epsilon + 3\omega/4$, the protocol rejects.

11.1.2 Lower bound for deterministic protocols

In this section we prove the lower bound of Theorem 11.1 by reducing the string Equality problem EQU to the linear isomorphism problem LinIso. To describe the reduction, we first formally define EQU and a non-tolerant variant of LinIso.

Definition 11.5 (EQU_n)**.** Alice and Bob receive inputs $x \in \{0,1\}^n$ and $y \in \{0,1\}^n$, respectively, and must compute

$$\mathsf{EQU}_n(x,y) = \begin{cases} 1 & \text{if } x = y, \\ 0 & \text{otherwise;} \end{cases}$$

that is, determine whether their inputs are equal.

Definition 11.6 ($\widetilde{\mathsf{LinIso}}_\ell$)**.** Alice and Bob receive inputs $f : \mathbb{F}_2^\ell \to \{-1,+1\}$ and $g : \mathbb{F}_2^\ell \to \{-1,+1\}$, respectively, and must compute

$$\widetilde{\mathsf{LinIso}}_\ell(f,g) = \begin{cases} 1 & \text{if } \delta_{\mathcal{L}}(f,g) = 0, \\ 0 & \text{if } \delta_{\mathcal{L}}(f,g) > \omega. \end{cases}$$

Note that $\widetilde{\mathsf{LinIso}}_\ell$ is obtained from LinIso_ℓ by taking $\epsilon = 0$.

We now show the existence of a special mapping Φ whose main utility is that it eliminates linear relationships among input strings. Before giving the details, we introduce some notation. Let T be a string of length 2^p; this string can be viewed as the truth table of a Boolean function, which we denote by $f_T : \mathbb{F}_2^p \to \mathbb{F}_2$. Intuitively, the following lemma embeds vectors from $\mathbb{F}_2^n$ into Boolean functions on ℓ variables so that distinct vectors correspond to functions that are well separated in linear distance, while identical inputs yield the same function.

Lemma 11.7 (Φ-Map)**.** *Let $n > 0$ and $\omega \in (0, 1/2)$. Let $m = 2^\ell$ with $2cn/\omega \geq m \geq cn/\omega$ for a large enough constant c. Interpret each $y \in \mathbb{F}_2^m$ as the truth table of a function $f_y : \mathbb{F}_2^\ell \to \mathbb{F}_2$. Then there exists a mapping $\Phi : \mathbb{F}_2^n \to \mathbb{F}_2^m$ such that for all $x, y \in \mathbb{F}_2^n$:*

1. *$f_{\Phi(x)} = f_{\Phi(y)}$ if $x = y$, and*
2. *$\delta_{\mathcal{L}}(f_{\Phi(x)}, f_{\Phi(y)}) \geq \omega$ if $x \neq y$.*

Using Lemma 11.7 we first complete the proof of the lower bound of Theorem 11.1, and then prove Lemma 11.7.

Proof. [Proof of Theorem 11.1 (b)] Our proof relies on a class of functions obtained via a reduction from EQU_n. Recall that any deterministic protocol for EQU_n must exchange $\Omega(n)$ bits in the worst case.

Consider Algorithm 19. First, if $a = b$ then the corresponding functions f and g are linearly isomorphic, since identical functions are trivially linearly isomorphic. If $a \neq b$,

then by Lemma 11.7,

$$\delta_{\mathcal{L}}(f_{\Phi(a)}, f_{\Phi(b)}) \geq \omega.$$

Furthermore, both $f_{\Phi(a)}$ and $f_{\Phi(b)}$ are defined over $\mathbb{F}_2^\ell$, and their approximate spectral norms are bounded by $O(2^{\ell/2})$ (by Cauchy–Schwarz). Alice and Bob now invoke an efficient deterministic protocol for $\widetilde{\mathsf{Linlso}}_\ell$ on inputs $f_{\Phi(a)}$ and $f_{\Phi(b)}$. Suppose, for contradiction, that there is a deterministic protocol for $\widetilde{\mathsf{Linlso}}_\ell$ with communication complexity

$$o\left(\min\left\{\left\lceil \|\widehat{f}\|_{1,1/3}\right\rceil, \left\lceil \|\widehat{g}\|_{1,1/3}\right\rceil\right\}^2\right),$$

where f and g are as above (see Definition 11.6). Then this protocol could solve EQU_n with only $o(n)$ bits of communication, contradicting the $\Omega(n)$ lower bound for EQU_n. This concludes the proof. □

Algorithm 19: From EQU_n to $\widetilde{\mathsf{Linlso}}_\ell$

Input. Alice is given $a \in \mathbb{F}_2^n$ and Bob is given $b \in \mathbb{F}_2^n$.
Output. YES if $a = b$, else NO.
Step-1: Alice computes $\Phi(a)$ and the function $f_{\Phi(a)} : \mathbb{F}_2^\ell \to \{-1, +1\}$.

Step-2: Bob computes $\Phi(b)$ and the function $f_{\Phi(b)} : \mathbb{F}_2^\ell \to \{-1, +1\}$.

Step-3: Alice and Bob invoke a deterministic protocol for $\widetilde{\mathsf{Linlso}}_\ell$ on inputs $f_{\Phi(a)}$ and $f_{\Phi(b)}$.

Step-4: If $f_{\Phi(a)}$ and $f_{\Phi(b)}$ are reported linearly isomorphic, output YES; otherwise output NO.

We now prove Lemma 11.7.

Proof. [Proof of Lemma 11.7] The mapping Φ is constructed iteratively, assigning to each $x \in \mathbb{F}_2^n$ a truth table $y \in \mathbb{F}_2^m$ that has not yet been used and is sufficiently far, in linear distance, from all previously assigned functions. At each step we remove from the candidate set all truth tables corresponding to Boolean functions that are linearly close to the chosen function. Claim 11.8 bounds the number of functions eliminated in each iteration, accounting for both linear isomorphisms and nearby functions within Hamming distance ω. By choosing m large enough, the total number of removed functions across all 2^n iterations remains smaller than the total number of possible truth tables, ensuring that a valid choice always exists and that the procedure successfully constructs Φ.

We implement this via Algorithm 20. Recall that $m = 2^\ell$, so any $y \in \mathbb{F}_2^m$ can be viewed as the truth table of a Boolean function $f_y : \mathbb{F}_2^\ell \to \mathbb{F}_2$. Two sets are maintained:

$\mathcal{S}$: the partial set of $x \in \mathbb{F}_2^n$ for which the mapping has been carried out; initialized as the

empty set, and equal to $\mathbb{F}_2^n$ at the end.

$\mathcal{R}$: the set of $\Phi(x)$ for $x \in \mathcal{S}$, together with all their linear isomorphisms, interpreted as truth tables of Boolean functions on ℓ variables.

Algorithm 20: Construct Φ-Map

Initialization: $\mathcal{S} = \emptyset, \quad \mathcal{R} = \mathbb{F}_2^m$.

Repeat while $\mathcal{S} \neq \mathbb{F}_2^n$:

Step 1. Pick some $x \in \mathbb{F}_2^n \setminus \mathcal{S}$ and set $\mathcal{S} = \mathcal{S} \cup \{x\}$.

Step 2. Pick some y from $\mathcal{R}$ and set $\Phi(x) = y$; if no such y exists, return *Fail*.

Step 3. Let $L(f_{\Phi(x)}, \omega) \subseteq \mathbb{F}_2^m$ denote the set of truth tables of all ℓ-variable Boolean functions h such that $\delta_{\mathcal{L}}(h, f_{\Phi(x)}) \leq \omega$.

Step 4. Set $\mathcal{R} = \mathcal{R} \setminus L(f_{\Phi(x)}, \omega)$.

Output: Return *Success*.

To prove correctness of Algorithm 20 we need the following technical result.

Claim 11.8. *Let $f : \{0,1\}^r \to \{0,1\}$ and let $L(f, \omega)$ be the set of truth tables of all r-variable Boolean functions at linear distance at most ω from f. Then $|L(f, \omega)| = O\left(2^{r^2} \cdot 2^{H(\omega)2^r}\right)$.*

Proof. The number of nonsingular linear transformations of dimension $r \times r$ is at most 2^{r^2}, so there are at most 2^{r^2} Boolean functions on r variables linearly isomorphic to a given f. For each such function, the number of Boolean functions whose truth tables differ in at most an ω fraction of entries is at most $2^{H(\omega)2^r}$; see [Cover and Thomas, 2006]. This is because, for a fixed Boolean function on r variables, counting the neighbors that differ in at most an ω fraction of entries is equivalent to estimating the volume of a Hamming ball of radius ω with entropy $H(\omega)$ over a field of characteristic 2 of length 2^r. Combining these bounds gives the claim. □

For every iteration of the **while** loop in Algorithm 20, Claim 11.8 gives an upper bound of $2^{\ell^2} 2^{H(\omega)2^\ell}$ on $|L(f_{\Phi(x)}, \omega)|$ in Step 3. Since the number of iterations is at most 2^n, if m is chosen so that

$$2^m \geq 2^n \cdot 2^{\ell^2} 2^{H(\omega) \cdot m} \tag{11.3}$$

then the algorithm never returns Fail in Step 2(a) and successfully constructs Φ. It remains to show that $2cn/\omega \geq m \geq cn/\omega$, for a large enough constant c, implies (11.3). Observe

that

$$\begin{aligned} 2^m \geq 2^n \cdot 2^{\ell^2} \cdot 2^{m \cdot H(\omega)} &\iff m \geq n + \ell^2 + m \cdot H(\omega) \\ &\iff m - (\log m)^2 - m \cdot H(\omega) \geq n \\ &\iff m \left(1 - \frac{(\log m)^2}{m} - H(\omega)\right) \geq n \\ &\iff m\,(1 - H(\omega) - o(1)) \geq n. \end{aligned}$$

Since $1 - H(\omega) \geq \omega$ for $\omega \in (0.2, 0.5)$, a suitable choice of the constant c gives $m \geq cn/\omega$. Finally, recall that for every positive integer d there exists an integer j with $d \leq 2^j < 2d$; we take m to be the smallest power of two in $[cn/\omega, 2cn/\omega]$. □

Remark 11.9 (Lower bounds in terms of t and ω). A main open question is to establish a lower bound depending simultaneously on both ω and t, in contrast to Theorem 11.1 (b).

11.2 Randomized setting

In randomized communication complexity, two standard models differ in how randomness is accessed. In the public-coin model, Alice and Bob share a string of random bits that both may use during the protocol, allowing them to coordinate without explicitly communicating their random choices. In the private-coin model, each party has access only to its own independently generated random bits, hidden from the other party.

The public-coin randomized setting. We begin by observing that in the public-coin setting the linear isomorphism problem becomes trivial in the non-tolerant regime $\epsilon = 0$, $\omega > 0$. The key idea is to view the truth tables of f and g as vectors in $\{-1, +1\}^{2^n}$. Alice and Bob independently compute canonical representatives of their functions under linear transformations: Alice computes an invertible $A \in \mathrm{GL}_n(\mathbb{F}_2)$ such that the truth table of $f \circ A$ is lexicographically minimal among all linear transformations of f, and Bob computes B such that $g \circ B$ is the lexicographically minimal representative of the linear orbit of g. If f and g are linearly isomorphic, these canonical forms coincide:

$$f \circ A = g \circ B.$$

The problem thus reduces to checking equality between the truth tables of the two transformed functions, which with public randomness can be done using $O(1)$ bits of communication and succeeds with constant probability. Intuitively, Alice and Bob encode their strings using a good linear error-correcting code and compare random positions

of the encoded strings using shared randomness; if the truth tables differ, the difference propagates through the codeword, and with constant probability a randomly chosen position reveals the discrepancy. Hence, in the public-coin model, linear isomorphism with $\varepsilon = 0$ and $\omega > 0$ admits a protocol with constant communication cost.

The private-coin randomized setting. In the remainder of this section we focus on the more restrictive private-coin setting. For convenience we restate the main theorem for this setting.

Theorem 11.2 (Randomized Communication Complexity)**.** *Let $f, g : \mathbb{F}_2^n \to \{-1, +1\}$ be Boolean functions held by Alice and Bob, respectively, and let $\epsilon \in [0, 1]$ and $\omega \in (0, 1)$. Define*

$$t := \min\left\{\left\lceil \|\widehat{f}\|_{1,1/3}\right\rceil, \left\lceil \|\widehat{g}\|_{1,1/3}\right\rceil\right\}.$$

(a) There exists a private-coin randomized communication protocol that uses at most $O(t^2/\omega^3)$ bits of communication and decides whether

$$\exists M \in \mathrm{GL}_n(\mathbb{F}_2),\ \Pr_{x\sim\mathbb{F}_2^n}[f(Mx) \neq g(x)] = 0$$
$$\text{or} \quad \forall M \in \mathrm{GL}_n(\mathbb{F}_2),\ \Pr_{x\sim\mathbb{F}_2^n}[f(Mx) \neq g(x)] \geq \omega.$$

(b) Moreover, any private-coin randomized protocol solving this problem with $\omega = 1/3$ must communicate at least $\Omega(\log t)$ bits.

The lower bound follows from the same reduction used for the deterministic lower bound. Recall that Equality on N-bit strings requires $\Omega(\log N)$ communication in the private-coin model [Yao, 1979]. Using the same reduction as in the deterministic setting, we immediately obtain a lower bound of $\Omega(\log t)$ for the private-coin randomized protocol. In the remainder of the section we prove the upper bound.

11.2.1 An efficient private-coin randomized communication protocol

We show that there is an efficient randomized protocol in the private-coin model whose communication cost depends polynomially on the approximate spectral norm of the input functions. Compared to the deterministic protocol, the use of randomness reduces the communication complexity by a quadratic factor in the spectral norm parameter. We first prove a key lemma underlying the protocol: when the approximate spectral norm is

bounded, there exists a junta that closely approximates the original function up to a linear transformation.

Lemma 11.10 (Junta approximation under linear distance)**.** *Let $h : \mathbb{F}_2^n \to \{-1, +1\}$ be a Boolean function whose approximate spectral norm satisfies $\|\widehat{h}\|_{1,1/3} \leq t$, and let $\omega \in (0, 1)$. Then there exists a Boolean function $j : \mathbb{F}_2^n \to \{-1, +1\}$ depending on $O(t^2/\omega^2)$ coordinates (that is, a junta) such that the linear distance satisfies $\delta_{\mathcal{L}}(h, j) < \omega/8$.*

Proof. The proof begins with a real-valued function h^* that pointwise approximates h and has small Fourier ℓ_1-norm. We first truncate the Fourier expansion of h^*, keeping only the coefficients whose magnitude exceeds a suitable threshold. Using Parseval's identity, we show that the truncated function $\tilde{h}^*$ remains close to h^* in expectation, which implies that the signs of h^* and $\tilde{h}^*$ disagree only on a small fraction of inputs. Finally, since the truncated representation contains only a small number of significant Fourier coefficients, a linear change of variables makes these coefficients depend on only a few coordinates, yielding a junta j close to h in linear distance. We now give the formal proof.

Let $h^* : \mathbb{F}_2^n \to \mathbb{R}$ be a real-valued function that pointwise $(1/3)$-approximates h and satisfies

$$\|\widehat{h^*}\|_1 := \sum_{\alpha \in \mathbb{F}_2^n} |\widehat{h^*}(\alpha)| = t.$$

Note that

$$\mathop{\mathbb{E}}_{x \sim \mathbb{F}_2^n} \left[h^*(x)^2\right] \leq \left(1 + \frac{1}{3}\right)^2 = \frac{16}{9}. \tag{11.4}$$

Define the *truncated Fourier approximation*

$$\tilde{h}^*(x) := \sum_{\alpha \in \mathbb{F}_2^n \,:\, |\widehat{h^*}(\alpha)| \geq \frac{\omega}{18t}} \widehat{h^*}(\alpha)\chi_\alpha(x),$$

and let

$$T := \left\{\alpha \in \mathbb{F}_2^n : |\widehat{h^*}(\alpha)| < \frac{\omega}{18t}\right\}.$$

Using Parseval's identity (see [de Wolf, 2008, Section 2.2]) and linearity of the Fourier

transform, and writing $\varphi := h^* - \tilde{h}^*$,

$$\begin{aligned}
\mathop{\mathbb{E}}_{x\sim\mathbb{F}_2^n}\left[(h^*(x)-\tilde{h}^*(x))^2\right] &= \sum_{\alpha\in\mathbb{F}_2^n}\widehat{\varphi}(\alpha)^2 \\
&= \sum_{\alpha\in\mathbb{F}_2^n}\left(\widehat{h^*}(\alpha)-\widehat{\tilde{h}^*}(\alpha)\right)^2 \quad \text{(linearity of the Fourier transform)} \\
&= \sum_{\alpha\in T}\widehat{h^*}(\alpha)^2 \\
&< \frac{\omega}{18t}\cdot\sum_{\alpha\in\mathbb{F}_2^n}|\widehat{h^*}(\alpha)| \qquad \text{(by the definition of } T\text{)} \\
&= \frac{\omega}{18t}\cdot t \\
&= \frac{\omega}{18}.
\end{aligned}$$

Now, if $\operatorname{sign}(h^*(x)) \neq \operatorname{sign}(\tilde{h}^*(x))$ then $|h^*(x)-\tilde{h}^*(x)| \geq 2/3$. Together with Markov's inequality, this gives

$$\begin{aligned}
\Pr_{x\sim\mathbb{F}_2^n}\left[\operatorname{sign}(h^*(x)) \neq \operatorname{sign}(\tilde{h}^*(x))\right] &\leq \Pr_{x\sim\mathbb{F}_2^n}\left[|h^*(x)-\tilde{h}^*(x)| \geq 2/3\right] \\
&= \Pr_{x\sim\mathbb{F}_2^n}\left[|h^*(x)-\tilde{h}^*(x)|^2 \geq 4/9\right] \\
&\leq \left(\frac{3}{2}\right)^2 \cdot \mathop{\mathbb{E}}_{x\sim\mathbb{F}_2^n}\left[(h^*(x)-\tilde{h}^*(x))^2\right] \quad \text{(Markov's inequality)} \\
&< \frac{9}{4}\cdot\frac{\omega}{18} \\
&= \frac{\omega}{8}.
\end{aligned}$$

Define the set of significant Fourier coefficients

$$S := \left\{\alpha \in \mathbb{F}_2^n : |\widehat{h^*}(\alpha)| \geq \frac{\omega}{18t}\right\},$$

and let $k := |S|$. From Parseval's identity (see [de Wolf, 2008, Section 2.2]) and (11.4),

$$\sum_{\alpha\in\mathbb{F}_2^n}\widehat{h^*}(\alpha)^2 = \mathbb{E}_x[h^*(x)^2] \leq \frac{16}{9}.$$

Since $\widehat{h^*}(\alpha)^2 \geq (\omega/18t)^2$ for all $\alpha \in S$,

$$k\cdot\left(\frac{\omega}{18t}\right)^2 \leq \frac{16}{9} \quad \Rightarrow \quad k \leq \frac{16}{9}\cdot\left(\frac{18t}{\omega}\right)^2 = \frac{576t^2}{\omega^2},$$

so $k = O(t^2/\omega^2)$. Let $\{\alpha_1,\ldots,\alpha_r\} \subseteq S$ be a linearly independent subset, where $r :=$

$\dim(\mathrm{span}(S))$. Define a nonsingular linear map $A : \mathbb{F}_2^n \to \mathbb{F}_2^n$ with $A(\alpha_i) = e_i$ for $i = 1, \ldots, r$, where $\{e_i\}$ is the standard basis. Finally, define the Boolean function

$$j(x) := \mathrm{sign}\Big(\tilde{h}^*(Ax)\Big).$$

Then j depends only on $r = O(t^2/\omega^2)$ coordinates and satisfies $\delta_{\mathcal{L}}(h, j) < \omega/8$. □

Algorithm 21: Randomized Private-Coin Protocol for Deciding Linear Isomorphism

Input. Alice is given $f : \mathbb{F}_2^n \to \{-1, +1\}$, Bob is given $g : \mathbb{F}_2^n \to \{-1, +1\}$, with the promise that either $\delta_{\mathcal{L}}(f, g) = 0$ or $\delta_{\mathcal{L}}(f, g) \geq \omega$.

Step-1: Alice and Bob alternately exchange bits to determine whether

$$\lceil \|\widehat{f}\|_{1,1/3} \rceil = \lceil \|\widehat{g}\|_{1,1/3} \rceil.$$

If not, they reject.

Step-2: Alice constructs a Boolean function $p : \mathbb{F}_2^{r_f} \to \{-1, +1\}$ from f using Lemma 11.10, and Bob constructs a Boolean function $q : \mathbb{F}_2^{r_g} \to \{-1, +1\}$ from g.

Step-3: Alice and Bob alternately exchange bits to determine whether $r_f = r_g$. If not, they reject.

Step-4: Alice considers all functions obtained by applying linear isomorphisms to p and selects the one whose truth table is lexicographically minimal; denote this canonical representative p^*. Similarly, Bob computes the lexicographically minimal representative q^* of q.

Step-5: For $i = 1, 2, \ldots, \lceil 20/\omega \rceil$:

- Alice uses her private randomness to sample a uniformly random input $x \in \mathbb{F}_2^{r_f}$.
- Alice sends $(x, p^*(x))$ to Bob.
- Bob checks whether $p^*(x) = q^*(x)$. If a mismatch is found, the protocol rejects.

If no mismatch is found in any iteration, the protocol accepts.

We now prove Theorem 11.2 (a).

Proof. [Proof of Theorem 11.2 (a)] The proof follows from the analysis of Algorithm 21. The protocol first approximates the input functions by truncating their Fourier expansions and keeping only the significant coefficients. These truncated representations induce Boolean functions whose signs depend on only a small number of coordinates. In the completeness case, if f and g are linearly isomorphic, their significant Fourier supports are

related by the same linear transformation, so the resulting approximations have identical canonical forms. In the soundness case, if f and g are far from linearly isomorphic, the approximations remain far under every linear transformation, so their canonical forms differ on a noticeable fraction of inputs. Alice and Bob detect this gap by comparing the canonical forms on a small number of randomly chosen inputs, yielding the claimed communication bound.

Completeness. Suppose f and g are linearly isomorphic. Then their Fourier transforms are related by a linear transformation $L \in \mathrm{GL}_n(\mathbb{F}_2)$: for every $\alpha \in \mathbb{F}_2^n$,

$$\widehat{g}(L\alpha) = \widehat{f}(\alpha).$$

Consequently, the sets of significant Fourier coefficients S_f and S_g of f and g satisfy $S_g = L(S_f)$, so the truncated approximations $\tilde{f}$ and $\tilde{g}$ are linearly isomorphic via L. Hence $p = \mathrm{sign}(\tilde{f})$ and $q = \mathrm{sign}(\tilde{g})$ are linearly isomorphic, and their canonical forms satisfy $p^* = q^*$.

Soundness. Suppose f and g are ω-far from linearly isomorphic, i.e.

$$\min_{L \in \mathrm{GL}_n(\mathbb{F}_2)} \Pr_{x \sim \mathbb{F}_2^n} [f(x) \neq g(Lx)] \geq \omega.$$

Let $\tilde{f}$ and $\tilde{g}$ be the truncated Fourier approximations, and set $p = \mathrm{sign}(\tilde{f})$, $q = \mathrm{sign}(\tilde{g})$, so that $\Pr_x[f(x) \neq p(x)] \leq \omega/8$ and $\Pr_x[g(x) \neq q(x)] \leq \omega/8$. Then for every linear map L,

$$\Pr_{x \sim \mathbb{F}_2^n} [p(x) \neq q(Lx)] \geq \omega - \frac{\omega}{8} - \frac{\omega}{8} = \frac{3\omega}{4}.$$

Thus the canonical forms p^* and q^* differ on at least a $3\omega/4$ fraction of inputs.

Communication cost. To detect this difference with constant probability, it suffices to compare p^* and q^* on $O(1/\omega)$ random inputs. In the private-coin setting, Alice samples a random $x \in \mathbb{F}_2^{r_f}$, where $r_f = O(t^2/\omega^2)$, and sends $(x, p^*(x))$ to Bob, costing $O(r_f)$ bits. Repeating for $O(1/\omega)$ trials, the total communication is $O(t^2/\omega^3)$ bits. □

Conclusion

This thesis developed *Fourier sketching* as a unifying framework for extracting information from the Fourier spectrum of Boolean functions through compact randomized summaries. The central insight underlying this work is that a broad range of testing, estimation, and recovery problems over the Boolean cube do not require reconstructing the entire Fourier spectrum. Instead, it suffices to estimate carefully chosen spectral statistics, and these can be captured efficiently by hashing Fourier coefficients into a small number of buckets. Two complementary realizations of this idea—*coset hashing*, which operates directly in the Fourier domain, and *subspace restriction*, which operates in the input domain—were shown to be dual manifestations of the same Fourier-analytic principle. Together, they led to improved algorithms, matching lower bounds, and conceptually simpler solutions across a diverse collection of problems.

We begin by summarizing the main contributions of this thesis, and then conclude by highlighting several directions for future research.

Summary of Contributions

The contributions of this thesis can be broadly divided into two categories: the development of new Fourier sketching techniques and the algorithmic applications they enable.

On the *techniques* side, we developed a refined understanding of Fourier-domain hashing. We characterized the information preserved by bucket measurements, identified the information inevitably lost through collisions, and showed how collision statistics encode global properties of the Fourier spectrum. In parallel, we investigated the behavior of approximately Fourier-sparse functions under random affine restrictions, establishing that many important spectral quantities remain stable under restriction and can therefore be estimated from local observations. The duality between these two viewpoints, expressed through the relationship between a subspace H and its orthogonal complement $H^{\perp}$, provides a unified perspective that allows ideas and techniques to transfer naturally between the frequency and input domains.

On the *applications* side, these techniques yielded several quantitative improvements.

- For **testing Fourier sparsity in ℓ_0 distance**, we improved the query complexity

from $\widetilde{O}(s^{14})$ to $\widetilde{O}(s^4)$, and ultimately to $\widetilde{O}(s)$. We also established an $\Omega(s)$ lower bound, thereby closing the gap up to polylogarithmic factors.

- For **estimating Fourier sparsity in ℓ_2 distance**, we improved the query complexity from $\widetilde{O}(s/\varepsilon^4)$ to $\widetilde{O}(s/\varepsilon^2)$, proved a matching $\Omega(s)$ lower bound, and extended the algorithm to the significantly more challenging distribution-free setting.
- For **linear isomorphism testing**, we improved the upper bound from $\widetilde{O}(m^{24})$ to $\widetilde{O}(m^4)$, where m denotes the spectral norm, and established an $\Omega(m^2)$ lower bound, representing an exponential improvement over the previously known logarithmic lower bound.
- For **estimating sumset size**, we obtained the first polynomial-query algorithm parameterized by the spectral norm.
- For **Fourier bias estimation**, we improved the upper bound from $\widetilde{O}(n/\tau^3)$ to $\widetilde{O}(1/\tau^2)$, while simultaneously strengthening the lower bound from $\Omega(1/\tau)$ to $\Omega(1/\tau^2)$, thereby closing the gap up to polylogarithmic factors.
- For **exact Fourier-sparse recovery**, we presented a substantially simpler randomized algorithm with $\widetilde{O}(ns)$ query complexity and running time, matching the best known query complexity while avoiding the considerably more involved machinery used in previous approaches.

Collectively, these results demonstrate that Fourier sketching provides not only a unifying conceptual framework but also a powerful algorithmic toolkit. Several of the bounds obtained in this thesis are now optimal, or nearly optimal up to polylogarithmic factors, significantly improving the state of the art.

Open Problems

The results presented in this thesis naturally give rise to several directions for future research. Some concern quantitative gaps left by the current techniques, while others point toward broader questions about the scope and limitations of Fourier sketching itself.

Removing polylogarithmic overheads. Several of the testers developed in this thesis are optimal only up to polylogarithmic factors. Determining the exact query complexity—for example, whether ℓ_0 Fourier sparsity testing admits an $O(s)$ algorithm without logarithmic overhead, or whether such overheads are inherent—remains an interesting open question and will likely require stronger lower-bound techniques.

Distribution-free testing without loss. Our distribution-free tester for ℓ_2 Fourier sparsity incurs additional complexity compared to the uniform-distribution setting. Whether this loss is unavoidable, or whether a distribution-free tester can match the complexity achieved under the uniform distribution, remains open and appears to require fundamentally new ideas for handling unknown input distributions.

Beyond $\mathbb{F}_2^n$: general Abelian groups. Although the results in this thesis are developed over $\mathbb{F}_2^n$, the underlying Fourier-analytic principles are considerably more general. Extending the coset hashing and subspace restriction frameworks to arbitrary finite Abelian groups, and understanding which guarantees depend essentially on characteristic two, would considerably broaden the applicability of these techniques and strengthen their connections with additive combinatorics.

Tightening linear isomorphism testing. For linear isomorphism testing, we established an $\widetilde{O}(m^4)$ upper bound together with an $\Omega(m^2)$ lower bound, leaving a quadratic gap. Closing this gap, either through an improved algorithm or a stronger lower bound, remains an important challenge. More generally, understanding linear isomorphism testing for target functions with unbounded spectral norm is largely unexplored.

Sketching richer spectral functionals. The present work focuses primarily on estimating Fourier sparsity, spectral norm, Fourier bias, and sumset size. However, the Fourier sketching framework appears considerably more general. Identifying additional spectral functionals that admit compact sketches, and ultimately characterizing which Fourier statistics are sketchable and which are not, would elevate Fourier sketching from a successful algorithmic technique to a broader theoretical framework.

We hope that the Fourier sketching framework developed in this thesis, together with the duality between hashing and restriction that underpins it, will find applications far beyond the problems considered here. More broadly, we believe these ideas provide a useful perspective on the design of sublinear algorithms, suggesting that compact randomized sketches may serve as a unifying paradigm for extracting global Fourier information from local observations.

Bibliography

Emmanuel Abbe, Amir Shpilka, and Min Ye. Reed-Muller Codes: Theory and Algorithms. *IEEE Transactions on Information Theory*, 67(6):3251–3277, 2021.

Anil Ada, Omar Fawzi, and Hamed Hatami. Spectral Norm of Symmetric Functions. In *Proceedings of the 16th International Workshop on Randomization and Computation, RANDOM*, pages 338–349, 2012.

Manindra Agrawal and Thomas Thierauf. The Formula Isomorphism Problem. *SIAM Journal on Computing*, 30(3):990–1009, 2000.

Kook Jin Ahn, Sudipto Guha, and Andrew McGregor. Analyzing graph structure via linear measurements. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, 2012.

Abiodun M. Aibinu, Momoh J. E. Salami, Amir A. Shafie, and Athaur Rahman Najeeb. Mri reconstruction using discrete fourier transform: a tutorial. *World Academy of Science, Engineering and Technology*, 42:179, 2008.

Adi Akavia, Shafi Goldwasser, and Shmuel Safra. Proving hard-core predicates using list decoding. In *44th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 146–157. IEEE Computer Society, 2003. doi: 10.1109/SFCS.2003.1238189.

Noga Alon. Testing Subgraphs in Large Graphs. *Random Structures & Algorithms*, 21(3–4): 359–370, 2002. A preliminary version of this paper appeared as an extended abstract in the proceedings of FOCS 2001.

Noga Alon and Eric Blais. Testing Boolean Function Isomorphism. In *Proceedings of the 10th International Workshop on Randomization and Computation (RANDOM)*, pages 394–405, 2010.

Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple Constructions of Almost k-Wise Independent Random Variables. *Random Structures & Algorithms*, 3(3): 289–304, 1992.

Noga Alon, Yossi Matias, and Mario Szegedy. The Space Complexity of Approximating the Frequency Moments. *Journal of Computer and System Sciences*, 58(1):137–147, 1999.

Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20, 2001. doi: 10.1007/s004930070001.

Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing Reed-Muller codes. *IEEE Transactions on Information Theory*, 51(11):4032–4039, 2005.

Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A Combinatorial Characterization of the Testable Graph Properties: It's All About Regularity. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC)*, pages 251–260, 2006.

Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A Combinatorial Characterization of the Testable Graph Properties: It's All About Regularity. *SIAM Journal on Computing*, 39(1):143–167, 2009.

Noga Alon, Eric Blais, Sourav Chakraborty, David García-Soriano, and Arie Matsliah. Nearly Tight Bounds for Testing Function Isomorphism. *SIAM Journal on Computing*, 42(2):459–493, 2013a.

Noga Alon, Eric Blais, Sourav Chakraborty, David García-Soriano, and Arie Matsliah. Nearly Tight Bounds for Testing Function Isomorphism. *SIAM Journal on Computing*, 42(2):459–483, 2013b.

Noga Alon, Jacob Fox, and Yufei Zhao. Efficient arithmetic regularity and removal lemmas for induced bipartite patterns. *Discrete Analysis*, April 2019. doi: 10.19086/da.7757.

Andisheh Amrollahi, Amir Zandieh, Michael Kapralov, and Andreas Krause. Efficiently Learning Fourier Sparse Set Functions. In *Proceedings of the 33rd Annual Conference on Neural Information Processing Systems (NeurIPS)*, volume 32, pages 15094–15103, 2019.

Benny Applebaum, Yuval Ishai, and Eyal Kushilevitz. Cryptography with constant input locality. *Journal of Cryptology*, 22(4):429–469, 2009. doi: 10.1007/s00145--008--9029--9.

Arnab, Victor Chen, Madhu Sudan, and Ning Bhattacharyya Xie. *Testing Linear-Invariant Non-linear Properties: A Short Report*, pages 260–268. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010. ISBN 978–3-642–16367–8. doi: 10.1007/978--3-642--16367--8_18. URL https://doi.org/10.1007/978--3-642--16367--8_18.

Srinivasan Arunachalam, Davi Castro-Silva, Arkopal Dutt, and Tom Gur. Classical and Quantum Polynomial Freiman-Ruzsa Algorithms. In Shubhangi Saraf, editor, *17th Innovations in Theoretical Computer Science Conference (ITCS 2026)*, volume 362 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 11:1–11:8, Dagstuhl, Germany, 2026. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978–3-95977–410–9.

doi: 10.4230/LIPIcs.ITCS.2026.11. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.ITCS.2026.11.

Vikraman Arvind and Yadu Vasudev. Isomorphism testing of Boolean functions computable by constant-depth circuits. *Information and Computation*, 239:3–12, 2014.

Nikhil Bansal and Makrand Sinha. k-Forrelation Optimally Separates Quantum and Classical Query Complexity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 1303–1316, 2021.

Mihir Bellare, Don Coppersmith, Johan Håstad, Marcos Kiwi, and Madhu Sudan. Linearity testing in characteristic two. *IEEE Transactions on Information Theory*, 42(6):1781–1795, 1996. doi: 10.1109/18.556674.

Alberto Bemporad. Efficient Conversion of Mixed Logical Dynamical Systems Into an Equivalent Piecewise Affine Form. *IEEE Transactions on Automatic Control*, 49(5):832 – 838, 2004.

Debajyoti Bera and Sapv Tharrmashastha. Quantum and randomised algorithms for non-linearity estimation. *ACM Transactions on Quantum Computing*, 2(2), 2021. doi: 10.1145/3456509.

Arnab Bhattacharyya and Yuichi Yoshida. *Property Testing: Problems and Techniques*. Springer Singapore, 1st edition, 2022.

Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. Optimal Testing of Reed-Muller Codes. In *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 488–497, 2010.

Arnab Bhattacharyya, Elena Grigorescu, Prasad Raghavendra, and Asaf Shapira. Testing Odd-Cycle-Freeness in Boolean Functions. *Combinatorics, Probability and Computing*, 21 (6):835–855, 2012. A preliminary version of this paper appeared as an extended abstract in the proceedings of SODA 2012.

Arnab Bhattacharyya, Eldar Fischer, Hamed Hatami, Pooya Hatami, and Shachar Lovett. Every Locally Characterized Affine-Invariant Property is Testable. In *Proceedings of the 45th ACM Symposium on Theory of Computing (STOC)*, pages 429 – 436, 2013.

Arnab Bhattacharyya, Elena Grigorescu, and Asaf Shapira. A Unified Framework for Testing Linear-Invariant Properties. *Random Structures & Algorithms*, 46(2):232–260, 2015. A preliminary version of this work appeared as an extended abstract in the proceedings of FOCS 2010.

Eric Blais. Improved Bounds for Testing Juntas. In *Proceedings of the International Conference on Randomization and Computation (RANDOM 2008)*, pages 317–330, 2008.

Eric Blais. Testing juntas nearly optimally. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*, pages 151–158, 2009.

Eric Blais and Ryan O'Donnell. Lower Bounds for Testing Function Isomorphism. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity (CCC)*, pages 235–246, 2010a.

Eric Blais and Ryan O'Donnell. Lower Bounds for Testing Function Isomorphism. In *Proceedings of the 25th Annual IEEE Conference on Computational Complexity (CCC)*. IEEE, 2010b. doi: 10.1109/CCC.2010.24.

Eric Blais, Joshua Brody, and Kevin Matulef. Property testing lower bounds via communication complexity. *Computational Complexity*, 21(2):311–358, 2012.

Eric Blais, Amit Weinstein, and Yuichi Yoshida. Partially symmetric functions are efficiently isomorphism testable. *SIAM Journal on Computing*, 44(2):411–432, 2015. A preliminary version of this work appeared as an extended abstract in the proceedings of FOCS 2012.

Eric Blais, Clément L. Canonne, Talya Eden, Amit Levi, and Dana Ron. Tolerant Junta Testing and the Connection to Submodular Optimization and Function Isomorphism. *ACM Transactions on Computation Theory*, 11(4):24:1–24:33, 2019a. A preliminary version of this work appeared as an extended abstract in the proceedings of SODA 2019.

Eric Blais, Clément L. Canonne, and Tom Gur. Distribution Testing Lower Bounds via Reductions from Communication Complexity. *ACM Transactions on Computation Theory*, 11(2):1–37, 2019b.

Avrim Blum. Learning a function of r relevant variables. In Bernhard Schölkopf and Manfred K. Warmuth, editors, *Learning Theory and Kernel Machines*, pages 731–733. Springer, Berlin, Heidelberg, 2003. ISBN 978–3-540–45167–9. doi: 10.1007/978--3-540--45167--9_54.

Avrim Blum, Merrick Furst, Michael Kearns, and Richard J. Lipton. Cryptographic primitives based on hard learning problems. In *Advances in Cryptology – CRYPTO '93*, 1994.

Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993a.

Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-Testing/Correcting with Applications to Numerical Problems. *J. Comput. Syst. Sci.*, 47(3), 1993b.

Thomas Blumensath and Mike E. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications*, 14(5–6):629–654, 2008.

Jean Bourgain. On the Construction of Affine Extractors. *Geometric and Functional Analysis*, 17(1):33–57, 2007.

Jean Bourgain. An improved estimate in the restricted isometry problem. In *Geometric Aspects of Functional Analysis*, pages 65–70. Springer, 2014.

Jehoshua Bruck and Roman Smolensky. Polynomial Threshold Functions, AC^0 Functions, and Spectral Norms. *SIAM Journal on Computing*, 21(1):33–42, 1992.

Nader H. Bshouty. Almost optimal testers for concise representations. In *Proceedings of the APPROX/RANDOM Conference*, 2020.

Mark Bun and Justin Thaler. Approximate Degree in Classical and Quantum Computing. *Foundations and Trends® in Theoretical Computer Science*, 15(3–4):229–423, 2022.

Emmanuel Candès and Terence Tao. Decoding by linear programming. *arXiv preprint math/0502327*, 2005.

Emmanuel J. Candès and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12): 5406–5425, 2006.

Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

Anne Canteaut and Marion Videau. Symmetric boolean functions. *IEEE Transactions on Information Theory*, 51(8):2791–2811, 2005. doi: 10.1109/TIT.2005.851743.

Claude Carlet. On the Degree, Nonlinearity, Algebraic Thickness and Non-Normality of Boolean Functions, with Developments on Symmetric Functions. *IEEE Transactions on Information Theory*, 50(9):2178–2185, 2004.

Claude Carlet. *Boolean Functions for Cryptography and Coding Theory*. Cambridge University Press, 2020.

Tullio Ceccherini-Silberstein, Fabio Scarabotti, and Filippo Tolli. *Discrete Harmonic Analysis: Representations, Number Theory, Expanders, and the Fourier Transform*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2018.

Sourav Chakraborty, David García-Soriano, and Arie Matsliah. Nearly Tight Bounds for Testing Function Isomorphism. In *Proceedings of the 22nd Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1683–1702, 2011a.

Sourav Chakraborty, David Garcí a-Soriano, and Arie Matsliah. Efficient sample extractors for juntas with applications. In Luca Aceto, Monika Henzinger, and Jiří Sgall, editors, *Automata, Languages and Programming*, pages 545–556, Berlin, Heidelberg, 2011b. Springer Berlin Heidelberg.

Sourav Chakraborty, Eldar Fischer, David García-Soriano, and Arie Matsliah. Junto-Symmetric Functions, Hypergraph Isomorphism and Crunching. In *Proceedings of the 27th Conference on Computational Complexity (CCC)*, pages 148–158, 2012.

Sourav Chakraborty, Nitin S. Mande, Rajat Mittal, Tulasimohan Molli, Madhusudan Paraashar, and Shubho Sanyal. Tight chang's-lemma-type bounds for boolean functions. *arXiv preprint*, arXiv:2012.02335, 2020.

Moses Charikar, Kevin Chen, and Martin Farach-Colton. Finding frequent items in data streams. In *Proceedings of the International Colloquium on Automata, Languages, and Programming*, 2002.

Arkadev Chattopadhyay, Nikhil S. Mande, and Suhail Sherif. The log-approximate-rank conjecture is false. *Journal of the ACM*, 67(4):1–28, 2020. doi: 10.1145/3396695.

Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. *Theory of Computing*, 15(10):1–26, 2019.

Xi Chen and Jinyu Xie. Tight bounds for the distribution-free testing of monotone conjunctions. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 54–71, 2016.

Xi Chen, Anindya De, Rocco A. Servedio, and Li-Yang Tan. Boolean function monotonicity testing requires (almost) $n^{1/2}$ non-adaptive queries. In *Proceedings of the 47th ACM Symposium on Theory of Computing (STOC)*, pages 519–528, 2015.

Xi Chen, Erik Waingarten, and Jinyu Xie. Beyond Talagrand functions: New lower bounds for testing monotonicity and unateness. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing (STOC)*, pages 523–536, 2017.

Xi Chen, William Pires, Toniann Pitassi, and Rocco A. Servedio. Testing Juntas and Junta Subclasses with Relative Error. In *Proceedings of the 38th Conference on Learning Theory (COLT)*, volume 291, pages 5214–5245, 2025.

Mahdi Cheraghchi, Venkatesan Guruswami, and Ameya Velingker. Restricted isometry of fourier matrices and list decodability of random linear codes. *SIAM Journal on Computing*, 42(5):1888–1914, 2013.

Tsun Ming Cheung, Hamed Hatami, Rosie Zhao, and Itai Zilberstein. Boolean functions with small approximate spectral norm. *Discrete Analysis*, September 2024. doi: 10.19086/da.122971.

Valentina Ciriani. Synthesis of SPP three-level logic networks using affine spaces. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 22(10):1310–1323, 2003.

Gil Cohen and Avishay Tal. Two structural results for low-degree polynomials and applications. In *Proceedings of the 18th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems and the 19th International Workshop on Randomization and Computation (APPROX-RANDOM)*, volume 40 of *LIPIcs*, pages 680–709. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2015.

Paul J. Cohen. On a conjecture of littlewood and idempotent measures. *American Journal of Mathematics*, 82:191–212, 1960.

James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90):297–301, 1965.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2 edition, 2006.

Lingguo Cui and Yuanda Cao. A new S-box structure named Affine-Power-Affine. *International Journal of Innovative Computing, Information and Control*, 3(3):751–759, 2007.

Swarnalipa Datta, Arijit Ghosh, Chandrima Kayal, Manaswi Paraashar, and Manmatha Roy. Spectral Norm, Economical Sieve, and Linear Invariance Testing of Boolean Functions. In *Proceedings of the 43rd International Symposium on Theoretical Aspects of Computer Science (STACS)*, volume 364, pages 30:1–30:21, 2026a.

Swarnalipa Datta, Arijit Ghosh, Chandrima Kayal, Manaswi Paraashar, and Manmatha Roy. Spectral norm, economical sieve, and linear invariance testing of boolean functions, 2026b. URL https://arxiv.org/abs/2308.02662.

Stanisa Dautovic and Ladislav A. Novak. A comment on "Boolean functions classification via fixed polarity Reed-Muller form". *IEEE Transactions on Computers*, 55(8):1067 – 1069, 2006.

Anindya De, Shivam Nadimpalli, and Rocco A. Servedio. Approximating sumset size. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, 2022.

Ronald de Wolf. A brief introduction to Fourier analysis on the Boolean cube. *Theory of Computing*, pages 1–20, 2008.

Ilias Diakonikolas, Homin K Lee, Kevin Matulef, Krzysztof Onak, Ronitt Rubinfeld, Rocco A Servedio, and Andrew Wan. Testing for Concise Representations. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science, FOCS*, pages 549–558, 2007.

Itai Dinur, Orr Dunkelman, Nathan Keller, Eyal Ronen, and Adi Shamir. Efficient detection of high probability statistical properties of cryptosystems via surrogate differentiation. In *Advances in Cryptology – EUROCRYPT 2023*, volume 14007 of *Lecture Notes in Computer Science*, pages 103–133. Springer, 2023.

Hans Dobbertin. Construction of Bent Functions and Balanced Boolean Functions with High Nonlinearity. In *Fast Software Encryption*, volume 1008 of *Lecture Notes in Computer Science*, pages 61–74. Springer, 1995.

D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006. doi: 10.1109/TIT.2006.871582.

Orr Dunkelman, Nathan Keller, and Adi Shamir. Minimalism in Cryptography: The Even-Mansour Scheme Revisited. In *Proceedings of the 31st Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, pages 336–354, 2012.

Iwan M. Duursma, Carlos Rentería-Márquez, and Horacio Tapia-Recillas. Reed-Muller Codes on Complete Intersections. *Applicable Algebra in Engineering, Communication and Computing*, 11(6):455–462, 2001.

Jean-Charles Faugère and Ludovic Perret. Polynomial Equivalence Problems: Algorithmic and Theoretical Aspects. In *Proceedings of the 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*, pages 30–47, 2006.

V. Feldman, P. Gopalan, S. Khot, and A. K. Ponnuswami. New Results for Learning Noisy Parities and Halfspaces. In *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006)*, pages 563–574, 2006a. doi: 10.1109/FOCS.2006.51.

Vitaly Feldman, Parikshit Gopalan, Subhash Khot, and Ashok Kumar Ponnuswami. New results for learning noisy parities and halfspaces. In *Proceedings of the Annual IEEE Symposium on Foundations of Computer Science*, 2006b.

Eldar Fischer, Guy Kindler, Dana Ron, Shmuel Safra, and Alex Samorodnitsky. Testing juntas. *Journal of Computer and System Sciences*, 68(4):753–787, 2004.

Michael A. Forbes and Zander Kelley. Pseudorandom generators for read-once branching programs, in any order. In *59th IEEE Annual Symposium on Foundations of Computer Science (FOCS 2018)*, pages 946–955. IEEE Computer Society, 2018.

Arijit Ghosh and Manmatha Roy. Price of Parsimony: Complexity of Fourier Sparsity Testing. *https://openreview.net/forum?id=7bCPXHq8xV*, 2025.

Arijit Ghosh, Subhamoy Maitra, and Manmatha Roy. Testing Fourier Sparsity via Implicit Sensing, 2026. URL https://openreview.net/forum?id=upsp6XeJpN. ICLR 2026.

Uma Girish, Justin Holmgren, Kunal Mittal, Ran Raz, and Wei Zhan. Parallel repetition for the GHZ game: A simpler proof. In *Proceedings of the 24th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems and the 25th International Workshop on Randomization and Computation (APPROX/RANDOM)*, volume 207 of *LIPIcs*, pages 62:1–62:19. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021a.

Uma Girish, Avishay Tal, and Kewen Wu. Fourier Growth of Parity Decision Trees. In *Proceedings of the 36th Computational Complexity Conference (CCC)*, volume 200, pages 39:1–39:36, 2021b.

Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017.

Oded Goldreich and Leonid A. Levin. A Hard-Core Predicate for all One-Way Functions. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing (STOC)*, pages 25–32, 1989.

Oded Goldreich and Dana Ron. On Sample-Based Testers. *ACM Transactions on Computation Theory*, 8(2):1–54, 2016.

Oded Goldreich and Luca Trevisan. Three Theorems Regarding Testing Graph Properties. *Random Structures & Algorithms*, 23(1):23–57, 2003. A preliminary version of this paper appeared as an extended abstract in the proceedings of FOCS 2001.

Oded Goldreich, Shafi Goldwasser, and Dana Ron. Property Testing and its Connection to Learning and Approximation. *Journal of the ACM*, 45:653–750, 1998.

Oded Goldreich, Shafi Goldwasser, Eric Lehman, Dana Ron, and Alex Samorodnitsky. Testing monotonicity. *Combinatorica*, 20(3):301–337, 2000.

Parikshit Gopalan, Ryan O'Donnell, Rocco A Servedio, Amir Shpilka, and Karl Wimmer. Testing fourier dimensionality and sparsity. *SIAM Journal on Computing*, 40(4):1075–1100, 2011.

W. T. Gowers, Ben Green, Freddie Manners, and Terence Tao. Marton's Conjecture in abelian groups with bounded torsion, 2024. URL https://arxiv.org/abs/2404.02244.

W. T. Gowers, Ben Green, Freddie Manners, and Terence Tao. On a conjecture of marton. *Annals of Mathematics*, 201(2):515–549, 2025.

William Timothy Gowers. A new proof of szemerédi's theorem. *Geometric and Functional Analysis*, 11(3):465–588, 2001. doi: 10.1007/s00039--001--0332--9.

Ben Green. A Szemerédi-Type Regularity Lemma in Abelian Groups, with Applications. *Geometric and Functional Analysis*, 15:340–376, 2005.

Ben Green and Tom Sanders. Boolean functions with small spectral norm. *Geometric and Functional Analysis*, 18:144–162, 2008a.

Ben Green and Tom Sanders. A quantitative version of the idempotent theorem in harmonic analysis. *Annals of Mathematics*, 168(3):1025–1054, 2008b.

Elena Grigorescu, Karl Wimmer, and Ning Xie. Tight Lower Bounds for Testing Linear Isomorphism. In *Proceedings of the 17th International Workshop on Randomization and Computation (RANDOM)*, pages 559–574, 2013.

S. Halevy and E. Kushilevitz. Distribution-Free Property Testing. *SIAM Journal on Computing*, 37(4):1107–1138, 2007.

Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. Simple and practical algorithm for sparse fourier transform. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, 2012a.

Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. Nearly Optimal Sparse Fourier Transform. In *Proceedings of the 44th Annual ACM Symposium on Theory of Computing (STOC)*, pages 563–578, 2012b.

Hamed Hatami and Shachar Lovett. Correlation testing for affine-invariant properties on f_p^n in the high error regime. In *Proceedings of the Annual ACM Symposium on Theory of Computing*, 2011.

Hamed Hatami and Shachar Lovett. Estimating the distance from testable affine-invariant properties, 2013.

Ishay Haviv and Oded Regev. The restricted isometry property of subsampled fourier matrices. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 288–297, 2016.

Elad Hazan, Adam R. Klivans, and Yang Yuan. Hyperparameter Optimization: A Spectral Approach. In *Proceedings of the 6th International Conference on Learning Representations (ICLR)*. OpenReview.net, 2018. URL https://openreview.net/forum?id=H1zriGeCZ.

Kaave Hosseini, Shachar Lovett, Guy Moshkovitz, and Asaf Shapira. An Improved Lower Bound for Arithmetic Regularity. *Mathematical Proceedings of the Cambridge Philosophical Society*, 161(2):193–197, 2016.

Xiang-Dong Hou. Classification of cosets of the Reed-Muller code $R(m-3,m)$. *Discrete Mathematics*, 128(1):203 – 224, 1994.

Piotr Indyk. Sketching via hashing: From heavy hitters to compressed sensing to sparse fourier transform. In *Proceedings of the ACM Symposium on Principles of Database Systems*, 2013.

Piotr Indyk and Michael Kapralov. Sample-optimal fourier sampling in any constant dimension. In *Proceedings of the Annual IEEE Symposium on Foundations of Computer Science*, 2014.

Piotr Indyk, Michael Kapralov, and Eric Price. (Nearly) Sample-Optimal Sparse Fourier Transform. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 36–55, 2014.

Ishay and Oded Haviv Regev. *The Restricted Isometry Property of Subsampled Fourier Matrices*, pages 163–179. Springer International Publishing, Cham, 2017.

Mark A. Iwen. Combinatorial sublinear-time fourier algorithms. *Foundations of Computational Mathematics*, 10(3):303–338, 2010.

Michael Kapralov. Sample efficient estimation and recovery in sparse fft via isolation on average. In *Proceedings of the Annual IEEE Symposium on Foundations of Computer Science*, 2017.

Mauricio Karchmer and Avi Wigderson. Monotone Circuits for Connectivity Require Super-logarithmic Depth. In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing (STOC)*, pages 539–550, 1988.

David R. Karger. Random sampling in cut, flow, and network design problems. *Mathematics of Operations Research*, 24(2):383–413, 1999.

Tali Kaufman and Madhu Sudan. Algebraic Property Testing: The Role of Invariance. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing (STOC)*, pages 403–412, 2008.

Michael J. Kearns and Umesh V. Vazirani. *An Introduction to Computational Learning Theory*. MIT Press, 1994.

Murat Kocaoglu, Karthikeyan Shanmugam, Alexandros G. Dimakis, and Adam Klivans. Sparse polynomial learning and graph sketching. In *Advances in Neural Information Processing Systems*, 2014.

Dmitry Kogan and Robert Krauthgamer. Sketching cuts in graphs and hypergraphs. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science (ITCS)*, pages 367–376. ACM, 2015. doi: 10.1145/2688073.2688093.

Eyal Kushilevitz and Yishay Mansour. Learning Decision Trees Using the Fourier Spectrum. *SIAM Journal on Computing*, 22(6), 1993.

Eyal Kushilevitz and Noam Nisan. *Communication Complexity*. Cambridge University Press, 1997.

Ping Li. *Relational Learning with Hypergraphs*. PhD thesis, EPFL, 2013.

Yi Li and Vasileios Nakos. Deterministic sparse fourier transform with an ℓ_∞ guarantee. *arXiv preprint arXiv:1903.00995*, 2019.

Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *Journal of the ACM*, 40(3):607–620, 1993.

Zhengyang Liu, Xi Chen, Rocco A. Servedio, Ying Sheng, and Jinyu Xie. Distribution-free Junta Testing. *ACM Transactions on Algorithms*, 15(1):1:1–1:23, 2019.

Eugene M. Luks. Hypergraph Isomorphism and Structural Equivalence of Boolean Functions. In *Proceedings of the 31st Annual ACM Symposium on Theory of Computing (STOC)*, pages 652–658, 1999.

Nikhil Shekhar Mande. *Communication Complexity of XOR Functions*. PhD thesis, Tata Institute of Fundamental Research, Mumbai, August 2018.

Yishay Mansour. Randomized interpolation and approximation of sparse polynomials. In *Proceedings of the International Colloquium on Automata, Languages, and Programming*, 1992.

Yishay Mansour. Learning Boolean Functions *via* the Fourier Transform. *Theoretical Advances in Neural Computation and Learning*, pages 391–424, 1994.

Mitsuru Matsui. Linear cryptanalysis method for des cipher. In *Advances in Cryptology – EUROCRYPT '93*, volume 765 of *Lecture Notes in Computer Science*, pages 386–397. Springer-Verlag, 1994.

O. Carruth McGehee, Louis Pigno, and Brent Smith. Hardy's inequality and the l^1 norm of exponential sums. *Annals of Mathematics*, 113(3):613–618, 1981. doi: 10.2307/2007000.

Willi Meier and Othmar Staffelbach. Fast correlation attacks on stream ciphers. In *Advances in Cryptology – EUROCRYPT '88*, volume 330 of *Lecture Notes in Computer Science*, pages 301–314. Springer-Verlag, 1988.

Raghu Meka, Omer Reingold, and Avishay Tal. Pseudorandom generators for width-3 branching programs. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC 2019)*, pages 13–23. ACM, 2019.

Miodrag J. Mihaljević, Sugata Gangopadhyay, Goutam Paul, and Hideki Imai. Generic Cryptographic Weakness of k-Normal Boolean Functions in Certain Stream Ciphers and Cryptanalysis of GRAIN-128. *Periodica Mathematica Hungarica*, 65(2):205–227, 2012.

Peter Bro Miltersen, Noam Nisan, Shmuel Safra, and Avi Wigderson. On Data Structures and Asymmetric Communication Complexity. In *Proceedings of the ACM Symposium on Theory of Computing (STOC)*, 1995.

Ankur Moitra. *Algorithmic Aspects of Machine Learning*. Cambridge University Press, 2018.

Paz Morillo and Carla Ràfols. The security of all bits using list decoding. In *Public Key Cryptography – PKC 2009*, volume 5443 of *Lecture Notes in Computer Science*. Springer, 2009. doi: 10.1007/978--3-642--00468--1_2.

Shivam Nadimpalli and Shyamal Patel. Optimal Non-adaptive Tolerant Junta Testing via Local Estimators. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC)*, pages 1039–1050, 2024. doi: 10.1145/3618260.3649687.

Deanna Needell and Joel A. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26(3):301–321, 2009.

Deanna Needell and Roman Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Foundations of Computational Mathematics*, 9(3):317–334, 2009.

Sahand Negahban and Devavrat Shah. Learning sparse boolean polynomials. In *Proceedings of the Allerton Conference on Communication, Control, and Computing*, 2012.

Noga, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Alon Ron. Testing Low-Degree Polynomials over $GF(2)$. In *Proceedings of the 7th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, pages 188–199, 2003.

Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

Sameer Pawar and Kannan Ramchandran. A robust r-ffast framework for computing a k-sparse n-length dft. In *IEEE International Symposium on Information Theory*, 2014.

Anup Rao and Amir Yehudayoff. *Communication Complexity and Applications*. Cambridge University Press, 2020.

K. F. Roth. On certain sets of integers. *Journal of the London Mathematical Society*, s1–28 (1):104–109, 1953. doi: 10.1112/jlms/s1--28.1.104.

Mark Rudelson and Roman Vershynin. On Sparse Reconstruction from Fourier and Gaussian Measurements. *Communications on Pure and Applied Mathematics*, 61(8): 1025–1045, 2008.

Tom Sanders. Boolean functions with small spectral norm, revisited. *Mathematical Proceedings of the Cambridge Philosophical Society*, 167(2):335–344, 2019.

Tom Sanders. Bounds in Cohen's Idempotent Theorem. *Journal of Fourier Analysis and Applications*, 26(2):25, 2020.

Swagato Sanyal. Fourier Sparsity and Dimension. *Theory of Computing*, 15(11):1–13, 2019.

Robin Scheibler, Saeid Haghighatshoar, and Martin Vetterli. A Fast Hadamard Transform for Signals With Sublinear Sparsity in the Transform Domain. *IEEE Transactions on Information Theory*, 61(4):2115–2132, 2015.

Ronen Shaltiel. Dispersers for Affine Sources with Sub-Polynomial Entropy. In *Proceedings of the 52nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 247–256, 2011.

Claude E. Shannon. Communication in the presence of noise. *Proceedings of the Institute of Radio Engineers*, 37(1):10–21, January 1949. doi: 10.1109/jrproc.1949.232969.

Amir Shpilka, Avishay Tal, and Ben lee Volk. On the Structure of Boolean Functions with Small Spectral Norm. *Computational Complexity*, 26(1), 2017.

Siemens. Compressed sensing beyond speed. https://www.healthcare.siemens.com/magnetic-resonance-imaging/clinical-specialities/compressed-sensing, 2026.

Hans Ulrich Simon. Efficient computation of approximate isomorphisms between Boolean functions. *Information Processing Letters*, 116(3):237–240, 2016.

Elias M. Stein and Rami Shakarchi. *Fourier Analysis: An Introduction*. Princeton University Press, 2003.

Elias M. Stein and Rami Shakarchi. *Fourier Analysis: An Introduction*. Princeton University Press, Princeton, NJ, 2011.

Peter Stobbe and Andreas Krause. Learning Fourier Sparse Set Functions. In *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, AISTATS*, volume 22 of *JMLR Proceedings*, pages 1125–1133. JMLR.org, 2012.

Endre Szemerédi. Regular partitions of graphs. Technical report, Stanford University, Department of Computer Science, 1975.

Avishay Tal. Towards optimal separations between quantum and randomized query complexities. In *61st IEEE Annual Symposium on Foundations of Computer Science (FOCS 2020)*, pages 228–239. IEEE Computer Society, 2020.

Terence Tao. Structure and Randomness in Combinatorics. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 3–15, 2007.

Terence Tao. *Higher Order Fourier Analysis*, volume 142 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

Albert Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia, 2005.

Hing Yin Tsang, Chung Hoi Wong, Ning Xie, and Shengyu Zhang. Fourier Sparsity, Spectral Norm, and the Log-Rank Conjecture. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 658–667, 2013.

L. Valiant. A Theory of the Learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.

Ernesto De Vito, Lorenzo Rosasco, Andrea Caponnetto, Umberto De Giovannini, and Francesca Odone. Learning from examples as an inverse problem. *Journal of Machine Learning Research*, 6:883–904, 2005.

Karl Wimmer and Yuichi Yoshida. Testing Linear-Invariant Function Isomorphism. In *Proceedings of the 40th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 840–850, 2013.

Yahoo! Yahoo! Messenger User Communication Pattern Dataset, 2019.

Andrew Chi-Chih Yao. Some Complexity Questions Related to Distributive Computing (Preliminary Report). In *Proceedings of the 11th Annual ACM Symposium on Theory of Computing (STOC)*, pages 209–213, 1979.

Grigory Yaroslavtsev and Samson Zhou. Fast Fourier Sparsity Testing. In *Proceedings of the 2020 Symposium on Simplicity in Algorithms (SOSA)*, pages 57 – 68, 2020.

Boyan Yordanov, Jana Tumova, Ivana Cerna, Jiří Barnat, and Calin Belta. Temporal Logic Control of Discrete-Time Piecewise Affine Systems. *IEEE Transactions on Automatic Control*, 57(6):1491 – 1504, 2012.

Zhiqiang Zhang and Yaoyun Shi. Communication complexities of symmetric xor functions. *Quantum Information & Computation*, 9(3):255–263, 2009.

Zhiqiang Zhang and Yaoyun Shi. On the parity complexity measures of Boolean functions. *Theoretical Computer Science*, 411(26–28):2612–2618, 2010.