
An Optimized Franz-Parisi Criterion and its Equivalence with SQ Lower Bounds

Siyu Chen

Department of Statistics and Data Science
Yale University
siyu.chen.sc3226@yale.edu

Theodor Misiakiewicz

Department of Statistics and Data Science
Yale University
theodor.misiakiewicz@yale.edu

Ilias Zadik

Department of Statistics and Data Science
Yale University
ilias.zadik@yale.edu

Peiyuan Zhang

Department of Statistics and Data Science
Yale University
peiyuan.zhang@yale.edu

Abstract

Bandeira et al. (2022) introduced the Franz-Parisi (FP) criterion for characterizing the computational hard phases in statistical detection problems. The FP criterion, based on an annealed version of the celebrated Franz-Parisi potential from statistical physics, was shown to be equivalent to low-degree polynomial (LDP) lower bounds for Gaussian additive models, thereby connecting two distinct approaches to understanding the computational hardness in statistical inference. In this paper, we propose a refined FP criterion that aims to better capture the geometric “overlap” structure of statistical models. Our main result establishes that this optimized FP criterion is equivalent to Statistical Query (SQ) lower bounds—another foundational framework in computational complexity of statistical inference. Crucially, this equivalence holds under a mild, verifiable assumption satisfied by a broad class of statistical models, including Gaussian additive models, planted sparse models, as well as non-Gaussian component analysis (NGCA), single-index (SI) models, and convex truncation detection settings. For instance, in the case of convex truncation tasks, the assumption is equivalent with the Gaussian correlation inequality (Royen, 2014) from convex geometry. In addition to the above, our equivalence not only unifies and simplifies the derivation of several known SQ lower bounds—such as for the NGCA model (Diakonikolas et al., 2017) and the SI model (Damian et al., 2024)—but also yields new SQ lower bounds of independent interest, including for the computational gaps in mixed sparse linear regression (Arpino et al., 2023) and convex truncation (De et al., 2023).

1 Introduction

Over the past decades, a central focus in statistical inference has been to understand the transition from computationally easy to hard regimes—that is, to characterize when a statistical task can be solved by polynomial-time algorithms. A key insight from this line of work is the emergence of *computational-statistical tradeoffs*: in many models, there exist broad parameter regimes where information-theoretic recovery is possible, yet no known polynomial-time algorithm succeeds.

Evidence for such tradeoffs spans multiple disciplines with varying levels of mathematical rigor. In particular, the statistical physics community has played an instrumental role by leveraging non-rigorous but highly predictive techniques to study average-case hardness. Their approach typically analyzes the geometry of solution spaces and identifies structural properties that correlate with algorithmic intractability (see [40] for a survey). Remarkably, for many statistical models, the predictions from statistical physics have been in striking agreement with the performance of the best-known polynomial-time algorithms.

Alongside these heuristic predictions, rigorous frameworks from statistics and theoretical computer science have been developed to analyze the limitations of efficient algorithms. While ruling out *all* polynomial-time algorithms would require resolving $\mathcal{P} \neq \mathcal{NP}$, substantial progress has been made by studying broad, expressive classes of polynomial-time algorithms. Two frameworks have emerged as particularly influential: *low-degree* (LD) *polynomial* lower bounds [31] and *statistical query* (SQ) lower bounds [20]. For many “nice enough” detection problems, the lower bounds derived from these frameworks align closely with the performance of the best-known polynomial-time algorithms¹. This striking consistency has motivated the formulation of the so-called *low-degree conjecture* by Hopkins [26], which posits that for sufficiently “symmetric and noisy” models, the failure of degree- $O(\log n)$ polynomials is indicative of the failure of all polynomial-time algorithms [31].

Given this context, a natural question arises: can one formally connect these two seemingly distinct approaches? At first glance, the answer appears negative, due to a fundamental mismatch in scope. Statistical physics techniques are primarily geared toward estimation problems, where the goal is to recover a hidden signal, while the rigorous frameworks discussed above—such as LD and SQ lower bounds—are focused on detection or hypothesis testing, where the task is to distinguish between the presence or absence of a signal in a noisy environment². Nevertheless, a major step towards bridging this gap was taken by Bandeira et al. (2022) [6], who introduced the *Franz-Parisi* (FP) *criterion* for computational hardness in detection tasks. Inspired by the seminal work of Franz and Parisi in spin glass theory [21], the FP criterion provides a geometric perspective on computational hardness rooted in overlap structures. Crucially, Bandeira et al. showed that for Gaussian additive models, the FP criterion is mathematically equivalent to the low-degree (LD) lower bounds, thereby establishing a rigorous link between statistical physics heuristics and formal algorithmic barriers.

Specifically, consider the following general detection problem between two distributions $\mathbb{P}$ and $\mathbb{Q}$ supported on a subset of $\mathbb{R}^n$, which in what follows we refer to as a “ $\mathbb{P}$ versus $\mathbb{Q}$ ” task. Under the *planted* distribution $\mathbb{P} = \mathbb{E}_u \mathbb{P}_u$, a signal u is drawn from a prior distribution π supported on $\Theta \subseteq \mathcal{S}^{N-1}$, and one observes m independent samples $Y_1, \dots, Y_m \sim \mathbb{P}_u$. Under the *null* distribution $\mathbb{Q}$, the samples are drawn independently from $Y_1, \dots, Y_m \sim \mathbb{Q}$. The goal in the detection task³ is to distinguish between these two hypotheses based on the observed data, that is to find a test statistics with vanishing Type I and Type II errors, as n grows. Note that the computational question then is whether such a successful test statistic exists that also terminates in polynomial-in- mn time. To characterize the hardness of detection problems from the statistical physics perspective, Bandeira et al. in [6] introduced the following notion of Franz-Parisi (FP) hardness:

Definition 1 (FP hardness). *For $D, m \in \mathbb{N}, \varepsilon > 0$, we say that a $\mathbb{P}$ versus $\mathbb{Q}$ detection task is (q, m, ε) -FP hard if*

$$\mathbf{FP:} \quad \mathbb{E} \left[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(|\langle u, v \rangle| \leq \delta(q)) \right] \leq 1 + \varepsilon, \quad \text{where} \quad (1)$$

$$\delta(q) = \sup \{ \delta > 0 : \pi^2(|\langle u, v \rangle| \geq \delta) \geq q^{-2} \}. \quad (2)$$

In the definition we denoted as customary $L_u = \frac{d\mathbb{P}_u}{d\mathbb{Q}}, u \in \Theta$, and for $f, g \in \mathcal{L}^2(\mathbb{R}^n)$, the Hilbert space $L^2(\mathbb{Q})$ of (square integrable) functions from $\mathbb{R}^n$ to $\mathbb{R}$, we use $\langle f, g \rangle_{\mathbb{Q}} = \mathbb{E}_{Y \sim \mathbb{Q}} f(Y)g(Y)$.

¹For exceptions to this correspondence, see [39] and discussion therein.

²Some recent work has extended low-degree lower bounds to estimation settings, beginning with [38], though this direction remains relatively underdeveloped.

³The associated *estimation* problem consists in recovering the planted signal u from $Y_1, \dots, Y_m \sim \mathbb{P}_u$.

We elaborate in Section A.1 on the statistical physics motivations behind this criterion, and only briefly highlight its core intuition here. The left-hand side of the FP condition integrates the function $F_{\text{ann}}(t) := \mathbb{E}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(\langle u, v \rangle = t)]$, over a $(1 - q^{-2})$ -typical region of the overlap variable t , corresponding to the constraint $|\langle u, v \rangle| \leq \delta(q)$. This function $F_{\text{ann}}(t)$ is an annealed proxy for the Franz-Parisi potential, a central object in statistical physics that has long served as a predictor of algorithmic hardness [40]. Intuitively, the Franz-Parisi potential captures the energy landscape experienced by local algorithms—such as Langevin or Glauber dynamics—whose performance is constrained by the geometry of the underlying signal space. The overlap $\langle u, v \rangle$ naturally quantifies a local “geometric” similarity between signals, making it a meaningful argument for $F_{\text{ann}}(t)$ and explaining its role within the FP criterion.

Returning to the definition of FP hardness, the parameter m corresponds to the sample size, and one should interpret q as a proxy for the required runtime. In this light, Bandeira et al. (2022) proved that, for Gaussian additive models, FP-hardness is equivalent to the failure of degree- $D = \log q$ polynomials to solve the detection task with m samples—i.e., roughly the authors of [6] showed that the problem is $(q, m, O(1))$ -FP hard if and only if it is “hard” for degree- $\log q$ polynomials to solve the detection task⁴. Hence, based on the current belief in the literature of low-degree lower bounds that a D -degree lower bound implies that the detection task requires at least e^D runtime to be solved, e.g., see [18], proving a task is $((mn)^{\omega(1)}, m, O(1))$ -FP hard for a Gaussian additive model provides rigorous evidence for polynomial-time hardness for the task.

Despite this success, the connection between the FP potential and other rigorous notions of algorithmic hardness remains limited. [6] only established a formal equivalence for Gaussian additive models and an one-sided implication for planted sparse models between the FP criterion and “low-degree” lower bounds. They further presented counterexamples where the equivalence fails entirely. In this work, our aim is to extend the Franz-Parisi criterion to rigorously characterize hardness beyond Gaussian additive models, and to clarify the scope and limitations of this framework across a broader class of statistical models.

1.1 Main Contributions

Our main contribution is to propose a slight modification of the FP-hardness criterion from [6], motivated by the observation that sticking to the Euclidean geometry assumption (and hence the “overlap” $\langle u, v \rangle$) may fail to capture the “true” hardness of some statistical models. We remark that this is an arguably natural modification, as (1) there are many statistical models for which the Euclidean geometry appears unnatural for navigating their parameter space (see Section 5 for a simple such construction), and (2) even in statistical physics settings, the Franz-Parisi potential is often considered under a more general notion of overlap [22]. Motivated by these considerations, we propose optimizing the “overlap” event inside the FP-hardness definition, subject only to a mild symmetry assumption for technical reasons. This leads to the following new criterion of FP-hardness:

Definition 2 (Generalized Franz Parisi (GFP) hardness under symmetry G). *Fix $q, m \in \mathbb{N}$, $\varepsilon > 0$ and a group G of finite order acting on the parameter space of the signal. We say a “ $\mathbb{P}$ versus $\mathbb{Q}$ ” problem is (q, m, ε) -GFP $_G$ hard if*

$$\text{GFP}_G : \inf_{\substack{A: \pi^{\otimes 2}(A) \geq 1 - q^{-2} \\ A \text{ is } G^2\text{-invariant}}} \mathbb{E} \left[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} \mathbf{1}(A) \right] \leq 1 + \varepsilon. \quad (3)$$

As in the original FP-hardness framework of [6], one should interpret q as a proxy for runtime, and therefore $((mn)^{\omega(1)}, m, O(1))$ -GFP hardness should be providing evidence of polynomial-time hardness with m samples in this framework. We highlight that the assumption on the invariance of the optimizing event under group G is made for technical reasons to enhance the applicability of our hardness criterion. We point the reader to Section 3.1 for further discussion of this assumption.

The main result of this work is that the “optimized” notion of GFP-hardness is fundamentally connected with the well-established framework of Statistical Query (SQ) hardness. The SQ framework

⁴[6] established this equivalence for $m = 1$, but the argument extends directly to general m .

was initially proposed by Kearns in [30] to capture the power of noise-tolerant algorithms. The notion of a statistical dimension proposed by [20] allowed for achieving powerful lower bounds against SQ methods, which we refer to from now on as SQ-hardness results. We employ here a slight strengthening of the notion of SQ-hardness from [20], introduced in [8].

Definition 3 (SQ hardness). *Fix $q, m \in \mathbb{N}$. We say a “ $\mathbb{P}$ versus $\mathbb{Q}$ ” detection problem is (q, m) -SQ hard if*

$$\mathbf{SQ}: \sup_{A: \pi^2(A) \geq q^{-2}} \mathbb{E} \left[\left| \langle L_u, L_v \rangle_{\mathbb{Q}} - 1 \right| \mid A \right] \leq \frac{1}{m}. \quad (4)$$

Roughly, a detection problem is (q, m) -SQ hard if any Statistical Query method succeeding at solving it with m samples requires q queries, which should be interpreted as requiring runtime q (see [8, Appendix A] for more details and motivation). Hence, proving a task is $((mn)^{\omega(1)}, m)$ -SQ hard provides evidence for polynomial-time hardness for the task.

Our main result is informally described as follows.

Theorem 1. *(Informal, GFP and SQ equivalence) Consider any $\mathbb{P}$ versus $\mathbb{Q}$ detection task which we assume (1) it satisfies a mild assumption with respect to a group G of finite order acting on the parameter space (namely Assumption 1 below), and (2) it is information-theoretically impossible to be solved with m_{IT} samples. Then the following holds for any samples size m and proxy runtime $q = m^{\Omega(1)}$.*

- *If the task is (q, m) -SQ hard, then it is also $(\Theta(q), \Theta(m), O(1))$ -GFP $_G$ -hard.*
- *If the task is $(q, m, O(1))$ -GFP $_G$ -hard, then it is also $(m^{\Theta(m_{\text{IT}})}, m^{1-o(1)})$ -SQ hard.*

Note that often in statistical tasks of interest $m_{\text{IT}} = \omega(\log n)$ (in fact, more often than not $m_{\text{IT}} = \text{poly}(n)$). Under this condition, Theorem 1 implies that a task is $((mn)^{\omega(1)}, m^{1-o(1)}, O(1))$ -GFP hard if and only if it is $((mn)^{\omega(1)}, m^{1-o(1)})$ -SQ hardness, matching the two criteria for hardness.

On top of that, as we mentioned above and discuss in Section 3.1, the required Assumption 1 on the detection task is rather mild. In fact, it turns out that it is satisfied for several models of recent interest in the community, making a strong case of how the Generalized Franz-Parisi criterion now correctly predicts the hardness phase for them. Importantly, these models include the Gaussian additive models and also greatly extend beyond them, significantly extending the key message from [6] about connecting the physics-based forms of hardness to more rigorous frameworks. We list now some of the tasks that satisfy Assumption 1.

1. *All Gaussian additive models (GAMs), under any symmetric prior, satisfy Assumption 1 with $G = \mathbb{Z}_2$ that flips the sign of the signal. Moreover, in that case the Generalized Franz-Parisi criterion is equivalent to the Franz-Parisi criterion, that is the optimizing event A in (3) is of the form $\{|\langle u, v \rangle| \leq \delta(q)\}$. Hence, Theorem 1 allows us to extend the result of [6] which proved the equivalence of FP-hardness to Low-degree hardness for GAMs, to also proving FP-hardness equivalent with SQ-hardness for these settings⁵.*
2. *All Planted Sparse Models satisfy Assumption 1 for the trivial group $G = \{\text{id}\}$. In particular, using Theorem 1 we can prove that GFP-hardness is equivalent to SQ-hardness for multiple well-studied models such as sparse phase retrieval [5], sparse regression [24, 6], (multi-sample) sparse PCA [7], and Bernoulli group testing [11]. As a corollary of this connection, we present a straightforward argument to obtain an SQ lower bound for the mixed sparse linear regression problem [5]. We remark that in [6] it has been proven that FP-hardness implies low-degree hardness for all Planted Sparse Models, but no result was presented for the other direction.*
3. *All Non-Gaussian component analysis (NGCA) models and all single-index models (under any symmetric prior) satisfy Assumption 1 with $G = \mathbb{Z}_2$. Therefore, via Theorem 1 GFP-hardness is again equivalent with SQ-hardness for these tasks.*

⁵We remark that such a connection could also be made via the results of [8], since GAMs are noise-robust.

4. *All Gaussian convex truncation models* satisfy Assumption 1 for $G = \{\text{id}\}$. In particular, interestingly Assumption 1 for these models is exactly equivalent to the celebrated Gaussian correlation inequality for convex bodies in probability theory, which was a multi-decade open problem posed in 1972 in [25] that was finally proven by Royen in 2014 [37]. Leveraging the equivalence between GFP-hardness and SQ-hardness in Theorem 1, we establish an SQ-lower bound for the convex truncation detection task. This allows us to provide, to the best of our knowledge, the first formal evidence that the current state-of-the-art polynomial-time detection method for convex truncation proposed in [15] has optimal sample complexity.

We also complement our results, with a simple example satisfying Assumption 1 where FP-hardness does *not* coincide with GFP-hardness, which we interpret as a model where the Euclidean geometry is not appropriate. We finally conclude the paper with a discussion.

For completeness, we prove in Appendix B the equivalence between GFP-hardness and low-degree (LD) polynomial hardness for noise-robust models. This result follows by combining our GFP-SQ equivalence with the equivalence between SQ-hardness and LD-hardness under noise robustness shown by Brennan et al. [8]. In particular, our GFP-hardness results for the examples presented in this paper immediately imply low-degree lower bounds in all those settings. This substantially extends the equivalence established in [6]. For clarity and readers' convenience, we also include succinct proofs of the SQ-to-LD equivalence, adapted from [8].

2 Setting and Definitions

We first recall the definition of a “**P versus Q**” task mentioned in the Introduction. Under the *planted* distribution $\mathbb{P} = \mathbb{E}_u \mathbb{P}_u$, a signal u is drawn from a prior distribution π supported on $\Theta \subseteq \mathcal{S}^{N-1}$, and one observes m independent samples $Y_1, \dots, Y_m \sim \mathbb{P}_u$. Under the *null* distribution $\mathbb{Q}$, the samples are drawn independently from $Y_1, \dots, Y_m \sim \mathbb{Q}$. The goal in the detection task⁶ is the so-called strong detection task to distinguish between these two hypotheses based on the observed data, that is to find a test statistics with vanishing Type I and Type II errors, as n grows. We will also be interested in the weak detection task, which is that the sum of type I and type II errors is at most $1 - \varepsilon$ for some fixed $\varepsilon > 0$ (not depending on n). In other words, strong detection means the test succeeds with high probability, while weak detection means the test has some non-trivial advantage over random guessing.

Throughout, we will work in the Hilbert space $L^2(\mathbb{Q})$ of (square integrable) functions $\mathbb{R}^N \rightarrow \mathbb{R}$ with inner product $\langle f, g \rangle_{\mathbb{Q}} := \mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)g(Y)]$ and corresponding norm $\|f\|_{\mathbb{Q}} := \langle f, f \rangle_{\mathbb{Q}}^{1/2}$. We will assume that $\mathbb{P}_u$ is absolutely continuous with respect to $\mathbb{Q}$ for all $u \in \text{supp}(\pi)$, use $L_u := \frac{d\mathbb{P}_u}{d\mathbb{Q}}$ to denote the likelihood ratio, and assume that $L_u \in L^2(\mathbb{Q})$ for all $u \in \text{supp}(\pi)$. The likelihood ratio between $\mathbb{P}$ and $\mathbb{Q}$ is denoted by $L := \frac{d\mathbb{P}}{d\mathbb{Q}} = \mathbb{E}_{u \sim \mu} L_u$. Observe that for m samples, we denote by $L_m = \mathbb{E}_{u \sim \mu} L_u$ the m -sample likelihood ratio. Finally, for a function $f : \mathbb{R}^N \rightarrow \mathbb{R}$ and integer $D \in \mathbb{N}$, we let $f^{\leq D}$ denote the orthogonal (w.r.t. $\langle \cdot, \cdot \rangle_{\mathbb{Q}}$) projection of f onto the subspace of polynomials of degree at most D .

An important identity between the (squared) norm of the likelihood ratio with m samples and the *chi-squared divergence* $\chi^2(\mathbb{P}^{\otimes m} \parallel \mathbb{Q}^{\otimes m})$ is

$$\|L\|_{\mathbb{Q}}^2 = \|\mathbb{E}_{u \sim \mu} L_u\|_{\mathbb{Q}}^2 = \chi^2(\mathbb{P} \parallel \mathbb{Q}) + 1 \geq 1.$$

This quantity has the following standard implications for *information-theoretic* impossibility of testing, in the asymptotic regime $n \rightarrow \infty$. The proofs can be found in e.g. [34, Lemma 2].

- If $\|L\|_{\mathbb{Q}}^2 = O(1)$ then strong detection is impossible.
- If $\|L\|_{\mathbb{Q}}^2 = 1 + o(1)$ then weak detection is impossible.

⁶The associated *estimation* problem consists in recovering the planted signal u from $Y_1, \dots, Y_m \sim \mathbb{P}_u$.

3 Main Results

In this section, we formally present our equivalence between GFP-hardness and SQ-hardness.

3.1 The Assumption

As mentioned in the Introduction, all our results operate under a crucial assumption on the “ $\mathbb{P}$ versus $\mathbb{Q}$ ” detection task. The assumption is as follows.

Assumption 1. *Given any “ $\mathbb{P}$ versus $\mathbb{Q}$ ” task, there exists a π -preserving finite group G acting on the parameter space Θ , i.e., for all $g \in G$, $g(v) \stackrel{(d)}{=} v$ for $v \sim \pi$, such that for any sample size m for any $u, v \in \Theta$, the following “correlation” inequality holds for any $k \in \mathbb{N}$*

$$\mathbb{E}_{g, g' \sim \text{Unif}(G)} (\langle L_{g(u)}, L_{g'(v)} \rangle_{\mathbb{Q}} - 1)^k \geq 0. \quad (5)$$

We first remark that (5) is a natural condition even if G is the trivial group, $G = \{\text{id}\}$. Indeed in that case (5) asks that for all $u, v \in \Theta$,

$$\langle L_u, L_v \rangle_{\mathbb{Q}} \geq 1. \quad (6)$$

Recall that if one averages over all $(u, v) \sim \pi^{\otimes 2}$, we have by standard identities

$$\mathbb{E} \langle L_u, L_v \rangle_{\mathbb{Q}} = \mathbb{E}_{\mathbb{Q}} \|\mathbb{E}_u L_u\|_2^2 = 1 + \chi^2(\mathbb{P}, \mathbb{Q}) \geq 1.$$

Thus, (6) should be understood as a pointwise condition that is guaranteed to hold in expectation over the product measure $\pi^{\otimes 2}$ for any $\mathbb{P}, \mathbb{Q}$. While this pointwise condition turns out to be vanilla satisfied in many models (such as Planted Sparse Models or Convex Truncation settings), a slight modification of it—leading to (5)—applies more broadly. Specifically, this modified condition requires (6) to hold for a pair u, v only after performing a “small” averaging over the a group orbit that preserves the prior π . For instance, if the prior is symmetric around 0 and the group G is $\mathbb{Z}_2$, which acts by flipping the sign of the signal u , then for $k = 1$, condition (5) reduces to demonstrating that, for all u, v ,

$$\frac{1}{4} (\mathbb{E} \langle L_u, L_v \rangle_{\mathbb{Q}} + \mathbb{E} \langle L_{-u}, L_v \rangle_{\mathbb{Q}} + \mathbb{E} \langle L_u, L_{-v} \rangle_{\mathbb{Q}} + \mathbb{E} \langle L_{-u}, L_{-v} \rangle_{\mathbb{Q}}) \geq 1,$$

which is significantly less restrictive than the original pointwise condition (6). This averaging approach allows for much greater generality, making it applicable to various settings, including Gaussian additive models, single-index models, and Non-Gaussian component analysis settings.

Remark 3.1. We finally make a trivial remark that will be useful in verifying (5) in our examples in Section 4 with symmetric prior. In all of them by symmetry we have for all u, v $\langle L_u, L_{-v} \rangle_{\mathbb{Q}} = \langle L_{-u}, L_v \rangle_{\mathbb{Q}}$ and $\langle L_u, L_v \rangle_{\mathbb{Q}} = \langle L_{-u}, L_{-v} \rangle_{\mathbb{Q}}$. Using that and the trivial fact that for all $x, y \in \mathbb{R}$, if $x + y \geq 0$ then $x^k + y^k \geq 0$ for all $k \in \mathbb{N}$, we conclude that if G is either the trivial group or $\mathbb{Z}_2$ (which will be the case in all examples of Section 4) it suffices to check the case $k = 1$ in (5), and then it automatically holds for all $k \in \mathbb{N}$.

Remark 3.2. As mentioned in the previous remark, we highlight that in all our examples in Section 4 of our GFP-SQ equivalence theorem below, we either use G to be the trivial group or $\mathbb{Z}_2$. The reason we state our assumption Assumption 1 for a general finite group G is for potential further applications of our work.

3.2 The GFP-SQ equivalence

3.2.1 Simplifying GFP-hardness

We present our equivalence theorem in two steps. First, we identify an approximate optimal “overlap” event A in the definition of GFP-hardness, which simplifies GFP-hardness significantly and makes GFP-hardness easier to establish in applications. Then, we prove the equivalence between this simplified version and SQ-hardness.

Given a group G acting on the parameter space of the signal, it turns out that the approximately optimal “overlap” event A takes the form $\{\rho_G(u, v) \leq r\}$ for the following notion of “overlap” between u, v ,

$$\rho_G(u, v) = \max_{g, g' \in G} \{|\langle L_{g(u)}, L_{g'(v)} \rangle_{\mathbb{Q}} - 1|\}.$$

In particular, focusing only on such type of events we define the following version of FP-hardness.

Definition 4 (ρ_G -FP hardness). *Fix $q, m \in \mathbb{N}$, $\varepsilon > 0$ and a finite group G acting on the parameter space of the signal. We say a “ $\mathbb{P}$ versus $\mathbb{Q}$ ” problem is (q, m, ε) - ρ_G -FP hard if*

$$\rho_G\text{-FP} : \quad \mathbb{E} [\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} \cdot \mathbf{1}(\rho_G(u, v) < r(q))] \leq 1 + \varepsilon, \quad \text{where} \quad (7)$$

$$r(q) = \sup\{r : \pi^2(\rho_G(u, v) \geq r) \geq q^{-2}\}, \quad (8)$$

We prove that GFP_G -hardness is equivalent to ρ_G -FP hardness under Assumption 1.

Theorem 2. *Consider any “ $\mathbb{P}$ versus $\mathbb{Q}$ ” task that satisfies Assumption 1 for a group G . Suppose $m, q \in \mathbb{N}$ and $\varepsilon > 0$. Then the following statements hold.*

1. *If the task is (q, m, ε) - ρ_G -FP hard, then the task is also (q, m, ε) - GFP_G hard.*
2. *Assume there exists an $r > 0$ such that $\pi^2(\rho_G(u, v) < r) = 1 - q^{-2}$ and that m is even. Then, if the task is (q, m, ε) - GFP_G hard, then it is $(q, m, \frac{3}{|G|}(1 + \varepsilon) + m \cdot \chi^2(\mathbb{P}, \mathbb{Q}))$ - ρ_G -FP hard. In particular, if $m\chi^2(\mathbb{P}, \mathbb{Q}) = O(1)$, the task is $(q, m, O(1 + \varepsilon))$ - ρ_G -FP hard.*

The proof of this theorem can be found in Appendix C.1.

Remark 3.3. While the first implication is immediate to grasp, the second implication has some additional conditions we now elaborate upon. First, both the requirements of the existence of r with the desired probability mass and the parity of m are for technical convenience, and both can be easily remove with some tedious work. Second, any potential “blow-up” in the ε -term for ρ_G -FP hard depends only on $|G|$, which should be treated as constant, and the term $m \cdot \chi^2(\mathbb{P}, \mathbb{Q})$, which is an easy to compute quantity (usually $n = 1$ and $\chi^2(\mathbb{P}, \mathbb{Q})$ is an one-dimensional integral). Moreover, it is almost always of order $O(1)$ for detection tasks that are conjecturally hard with m samples. Indeed, the mathematical reason behind this is exactly that it is equal to the squared $\mathcal{L}^2$ -norm of the projection of the likelihood onto the degree-1 polynomial space, i.e., on the span of linear functions. On top of that, if the detection task is (q, m) -SQ hard for any q then it holds directly $m\chi^2(\mathbb{P}, \mathbb{Q}) = O(1)$ as well. We elaborate more on this in Remark B.1 in Section B.

3.2.2 The equivalence

As we have already proven an equivalence between GFP -hardness and ρ_G -FP hardness, it suffices to connect the latter with SQ-hardness. This is the topic of the next theorem.

Theorem 3 (SQ and ρ_G -FP Equivalence). *Suppose a “ $\mathbb{P}$ versus $\mathbb{Q}$ ” task satisfies Assumption 1 for a group G .*

1. *If the task is (q, m) -SQ hard for some q, m with $q > 2$ then, it is also $(q', m', e|G|^{-1}m'/m)$ - ρ_G -FP hard for any integers $q' < q/\sqrt{2}$ and $m' \leq m/2$.*
2. *Suppose the task is (q, m, ε) - ρ_G -FP hard for some q, m integers. Assume that there exists an $r = r(q) > 0$ such that $\pi^2(\rho_G(u, v) < r) = 1 - q^{-2}$ and m is even. Then, the model is also (q', m') -SQ hard for any even integer t with $t \leq \log q / \log m$ and any integer $q' > 0$, where*

$$m' = \frac{m}{(t(1 + \varepsilon)^{1/t} + \chi^2(\mathbb{P}^{\otimes 4t} \parallel \mathbb{Q}^{\otimes 4t}))(q')^{2/t}}.$$

In particular, if for some sample size m_{IT} , we have

(a) *(Bounded χ^2 for m_{IT} samples)*

$$\chi^2(\mathbb{P}^{\otimes m_{\text{IT}}} \parallel \mathbb{Q}^{\otimes m_{\text{IT}}}) = O(1).$$

(b) (Large enough q)

$$q \geq m^{m_{\text{IT}}}$$

then the model is $(m^{\delta m_{\text{IT}}}, \Theta(\frac{m^{1-O(\delta)}}{m_{\text{IT}}(1+\varepsilon)}))$ -SQ hard for any $\delta > 0$.

The proof of this theorem can be found in Appendix C.2. Similar to Theorem 2, the conditions on r, m of part 2 in Theorem 3 are for technical convenience and can be easily removed. As we discussed in the Introduction the assumption that there exists some sufficiently growing m_{IT} (e.g., growing super-logarithmically in n) is natural for multiple commonly studied models. We remark that the condition on the information theory threshold m_{IT} to be growing with n is also necessary, by constructing a variant of the planted clique problem which satisfies Assumption 1, it is not SQ-hard and is GFP-hard. Lastly, we also note that our introduced Assumption 1 is also necessary for the equivalence. In Section A, we discuss a counterexample not satisfying Assumption 1 that is GFP-hard, but not SQ-hard.

Remark 3.4. We note that while our bounds in the equivalence of Theorem 2 deteriorate when $|G|$ becomes large, a slightly more general equivalence between GFP and SQ, using a variant of ρ_G , can also be proven for infinite groups G under an ‘‘hypercontractivity’’ assumption on $\langle L_u, L_v \rangle_{\mathbb{Q}}$ with respect to the pair (u, v) . We omit this generalization as in all relevant examples in this work a small group action using either the trivial or 2-cyclic group suffices.

4 Examples

In this section, we discuss two popular classes of detection tasks that satisfy Assumption 1 and hence fall under our GFP-SQ equivalence. Further examples are deferred to Appendix D.

4.1 Gaussian Additive Models

A $\mathbb{P}$ versus $\mathbb{Q}$ task is a Gaussian additive model (GAM) if it satisfies:

1. Under the null model, $\mathbb{Q} = \mathcal{N}(0, I_n)$.
2. Under the planted model $\mathbb{P}_u$ (for $u \in S^{n-1}$), for some signal-to-noise ratio (SNR) $\lambda > 0$ we set

$$Y = \lambda u + Z, \quad \text{for some } Z \sim \mathbb{Q}.$$

GAMs includes multiple well-studied models in the literature, with the predominant examples being (multisample variants) of tensor PCA [36] and sparse PCA [2]. For such models, it can be straightforwardly checked (see [6, Proposition 2.3]) that for all u, v ,

$$\langle L_u, L_v \rangle_{\mathbb{Q}} = e^{\lambda^2 \langle u, v \rangle}.$$

So for instance, in the case of non-negative sparse PCA where u, v are binary k -sparse vectors in (see e.g., [4, 10]) we always have $\langle u, v \rangle \geq 0$, and therefore Assumption 1 is always satisfied for the trivial group G . On top of that, Assumption 1 remains true for any prior which is symmetric around 0; this time Assumption 1 is also always satisfied by choosing the action of $G = \mathbb{Z}_2$ which flips the sign of u . We remark that symmetric priors encompass most commonly used priors for GAMs, e.g., for tensor PCA where $u = \text{vec}(x^{\otimes r})$, $x \sim \text{Unif}(S^{d-1})$.

Lemma 1. Consider any GAM with symmetric π , i.e., $v = -v, v \sim \pi$. For any $u, v \in \text{support}(\pi)$,

$$\frac{1}{4}(\langle L_u, L_v \rangle_{\mathbb{Q}} + \langle L_{-u}, L_v \rangle_{\mathbb{Q}} + \langle L_u, L_{-v} \rangle_{\mathbb{Q}} + \langle L_{-u}, L_{-v} \rangle_{\mathbb{Q}}) \geq 1.$$

Moreover, any GAM satisfies Assumption 1 for $G = \mathbb{Z}_2$ acting by flipping the sign of u .

Proof. Notice

$$\frac{1}{4}(\langle L_u, L_v \rangle_{\mathbb{Q}} + \langle L_{-u}, L_v \rangle_{\mathbb{Q}} + \langle L_u, L_{-v} \rangle_{\mathbb{Q}} + \langle L_{-u}, L_{-v} \rangle_{\mathbb{Q}}) = \frac{1}{2}(\exp(\lambda^2 \langle u, v \rangle) + \exp(-\lambda^2 \langle u, v \rangle)) \geq 1.$$

Hence, given Remark 3.1, the conclusion follows. $\square$

Given the above lemma, we conclude the (almost) equivalence between GFP-hardness and SQ-hardness from Theorem 3.

Remark 4.1. We remark that in the symmetric prior case for a GAM and $G = \mathbb{Z}_2$ acting by flipping the sign of u , $\rho_G(u, v) = \exp(\lambda^2 |\langle u, v \rangle|)$ is an increasing function of $|\langle u, v \rangle|$. Hence, for such GAMs we conclude via Theorem 2 that FP-hardness is equivalent to GFP-hardness, and therefore also to SQ-hardness. This is in agreement with the results of [6] establishing that FP-hardness is equivalent to LD-hardness; in fact, our approach can offer an alternative proof of their result via the LD-SQ equivalence [8] and the noise robustness of GAMs (see Theorem Theorem 6).

4.2 Planted Sparse Models

In [6], the authors introduced the family of planted sparse models (PSM) and proved that FP-hardness for a PSM implies it's also low-degree hard. We start with the definition.

A $\mathbb{P}$ versus $\mathbb{Q}$ task is a planted sparse model (PSM) if it satisfies:

1. Under the null model, the one sample is given by $Y = (Y_1, \dots, Y_n) \sim \mathbb{Q}$, where each entry $Y_i, i = 1, \dots, n$ is drawn independently from some distribution $\mathbb{Q}_i, i = 1, \dots, n$ on $\mathbb{R}$.
2. Under the planted model $\mathbb{P}_u$, we associate u with a set of planted entries $\Phi_u \subset [n]$. Then on sample is generated as follows. For the entries $i \notin \Phi_u$, we draw Y_i independently from $\mathbb{Q}_i$ (which is identical as in the $\mathbb{Q}$ measure). For the entries in Φ_u we draw from an arbitrary joint distribution $P_u|_{\Phi_u}$ with the following symmetry condition: for any subset $S \subseteq \Phi_u$, the marginal distribution $P_u|_{\Phi_u}(S)$ does not depend on u but only on S , i.e. $P_u|_S = P_S$.

Multiple well-known detection models satisfy this definition, such as, a well studied model of sparse regression [24, 6], Bernoulli group testing [1, 11], sparse phase retrieval [5], as well as multi-sample variants [8] of planted clique [29] and sparse (Wigner) PCA [2].

Satisfyingly, all planted sparse models directly satisfy Assumption 1 for the trivial group. In fact this result has already been established for a different use in [6, Proposition 3.6], proving that any u, v we have $\langle L_u, L_v \rangle_{\mathbb{Q}} \geq 1$. We state here for completeness.

Lemma 2. *Consider any PSM. For any $u, v \in \text{support}(\pi)$, $\langle L_u, L_v \rangle_{\mathbb{Q}} \geq 1$. This is to say, any PSM satisfies Assumption 1 for the trivial group.*

The proof follows from [6, Proposition 3.6] for $D = 0$. Using this, one can apply our main equivalence Theorem 3 to multiple interesting planted sparse models and obtain old and new SQ-hardness results in a rather streamlined fashion. As an instantiation of this, in Appendix D.1 we prove the GFP-hardness for the mixed sparse linear regression setting studied in [5] in its conjecturally hard regime. We then use our equivalence theorem to translate it into an SQ-hardness result in the same regime. Our results complement the existing low-degree lower bound [5], providing further evidence for the hard phase of the problem.

4.2.1 Other examples

Due to space constraint, we defer the following examples to Appendix D:

- *Non-Gaussian Component Analysis (NGCA):* Assumption 1 holds with $G = \mathbb{Z}_2$ for any symmetric prior. We recover the SQ-hardness result of [16] for the uniform prior via its equivalence with GFP-hardness, and establish a new SQ lower bound under a sparse prior.
- *Single-Index Models (SIM):* Again, Assumption 1 holds with $G = \mathbb{Z}_2$. We rederive the SQ-hardness result of [12] for the uniform prior through the GFP-hardness equivalence, and prove a new SQ lower bound for sparse priors.
- *Convex truncation detection:* Here Assumption 1 holds with the trivial group. In fact this assumption is precisely equivalent to the celebrated Gaussian Correlation Inequality on convex bodies [37, 32]. Using the GFP-hardness correspondence, we derive a new SQ lower bound that matches the current state-of-the-art polynomial-time algorithm of [15].

5 GFP-hardness is not always equal to FP-hardness

Recall that by definition, FP-hardness implies GFP-hardness. In this section, we show that the converse does not necessarily hold: we construct a $\mathbb{P}$ versus $\mathbb{Q}$ detection task that satisfies Assumption 1 and is easy under the FP criterion but hard under the GFP criterion. In particular, by using Theorem 2 and Theorem 3, the problem is also SQ-hard. Thus, while the FP criterion fails to capture the SQ-hardness in this case, our optimized GFP criterion correctly predicts it. As our initial departure from FP-hardness was that in many models the Euclidean overlap $\langle u, v \rangle$ might not be the “correct” choice, our example is carefully creating a model where the natural “overlap” $\rho_G(u, v)$ (based on Theorem 3) is not a function of the Euclidean dot product.

The $\mathbb{P}$ versus $\mathbb{Q}$ problem is defined as follows. The null model is $\mathbb{Q} = \text{Rad}(\frac{1}{2})^{\otimes(n+1)}$, i.e., each coordinate is an independent Rademacher random variable. For a signal $u \in \{0, 1\}^{n+1}$, the sample $x \sim \mathbb{P}_u$ is generated by drawing each coordinate independently according to

$$x_i = \begin{cases} +1, & \text{w.p. } \frac{1}{2} + r \cdot \frac{1-(1-\alpha) \cdot u_i}{2}, \\ -1, & \text{w.p. } \frac{1}{2} - r \cdot \frac{1-(1-\alpha) \cdot u_i}{2}, \end{cases} \quad (9)$$

where $\alpha, r \in (0, 1)$ are fixed constants to be chosen later. The following holds.

Lemma 3. *Let $u, v \in \{0, 1\}^{n+1}$. For any $u, v \in \{0, 1\}^{n+1}$, $\langle L_u, L_v \rangle_{\mathbb{Q}} = \prod_{i=0}^n (1 + r^2 \cdot \alpha^{u_i + v_i})$.*

Notice that our construction importantly ensures that the likelihood ratio inner product $\langle L_u, L_v \rangle$ is not solely a function of $\langle u, v \rangle$, but instead has a more intricate dependence on u and v . It is exactly this reason that leads to the discrepancy between GFP and FP hardness stated below.

Theorem 4. *There exist a two-point prior π on u such that, for $r = n^{-1/2}$, $\alpha = n^{-1+2\varepsilon}$, $m = n^{1-\varepsilon}$ and $D = n^\varepsilon$, where $\varepsilon > 0$ is any small constant, the following hold. The m -sample hypothesis testing problem $\mathbb{E}_{u \sim \pi} \mathbb{P}_u^{\otimes m}$ versus $\mathbb{Q}^{\otimes m}$ is $(e^{D/2}, m, \Theta(n^{-\varepsilon}))$ -GFP hard but not $(n^{-1}, m, \exp(\Theta(n^\varepsilon)))$ -FP hard. Moreover, via our equivalence theorem the model is $(e^{n^{\Theta(\varepsilon)}}, n^{1-\Theta(\varepsilon)})$ -SQ hard.*

The proof of this Theorem and the above Lemma can be found in Appendix E.

6 Conclusion

In this work, we generalize the Franz-Parisi (FP) criterion introduced by [6], motivated by the observation that the Euclidean dot product may not be the most natural geometry for all statistical task—a point partially illustrated by our example in Section 5. Our main result shows that optimizing the overlap event in the FP definition of [6] leads to a Generalized Franz-Parisi (GFP) hardness criterion, which is equivalent to SQ-hardness for models satisfying the mild Assumption 1. This assumption holds in a broad range of well-studied problems, including Gaussian additive models, planted sparse models, single-index models, and convex truncation. Our work significantly strengthens the theoretical foundation behind the (annealed) FP potential’s predictions from statistical physics, but also opens several questions:

1. (*Algorithmic implications*) Does the optimal overlap function $\rho_G(u, v)$ —as characterized in Theorem 2—yield meaningful algorithmic insights, particularly for local search or geometric methods?
2. (*The annealed potential*) Can similar equivalences be established for the original (also known as quenched) FP potential, or is the choice of the annealed version fundamental?
3. (*Interpretation of FP Area*) Why does the area under the FP curve appear to govern detection hardness? Is there some physical/algorithmic interpretation of this phenomenon?
4. (*Generalization to estimation*) Can our techniques be extended from detection to estimation tasks, for which the Franz-Parisi potential was originally introduced?

We believe these questions point toward promising future directions, with the potential to unify different approaches on the computational complexity of statistical inference.

References

- [1] Matthew Aldridge, Oliver Johnson, Jonathan Scarlett, et al. Group testing: an information theory perspective. *Foundations and Trends® in Communications and Information Theory*, 15(3-4):196–392, 2019.
- [2] Arash A Amini and Martin J Wainwright. High-dimensional analysis of semidefinite relaxations for sparse principal components. In *2008 IEEE international symposium on information theory*, pages 2454–2458. IEEE, 2008.
- [3] Gerard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Algorithmic thresholds for tensor pca. *The Annals of Probability*, 48(4):2052–2087, 2020.
- [4] Gérard Ben Arous, Alexander S Wein, and Ilias Zadik. Free energy wells and overlap gap property in sparse pca. *Communications on Pure and Applied Mathematics*, 76(10):2410–2473, 2023.
- [5] Gabriel Arpino and Ramji Venkataramanan. Statistical-computational tradeoffs in mixed sparse linear regression. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 921–986. PMLR, 2023.
- [6] Afonso S Bandeira, Ahmed El Alaoui, Samuel Hopkins, Tselil Schramm, Alexander S Wein, and Ilias Zadik. The franz-parisi criterion and computational trade-offs in high dimensional statistics. *Advances in Neural Information Processing Systems*, 35:33831–33844, 2022.
- [7] Matthew Brennan, Guy Bresler, and Wasim Huleihel. Universality of computational lower bounds for submatrix detection. In *Conference on Learning Theory*, pages 417–468. PMLR, 2019.
- [8] Matthew S Brennan, Guy Bresler, Sam Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low degree tests are almost equivalent. In *Conference on Learning Theory*, pages 774–774. PMLR, 2021.
- [9] Siyu Chen, Beining Wu, Miao Lu, Zhuoran Yang, and Tianhao Wang. Can neural networks achieve optimal computational-statistical tradeoff? an analysis on single-index model. In *The Thirteenth International Conference on Learning Representations*.
- [10] Zongchen Chen, Conor Sheehan, and Ilias Zadik. On the low-temperature mcmc threshold: the cases of sparse tensor pca, sparse regression, and a geometric rule. *arXiv preprint arXiv:2408.00746*, 2024.
- [11] Amin Coja-Oghlan, Oliver Gebhard, Max Hahn-Klimroth, Alexander S Wein, and Ilias Zadik. Statistical and computational phase transitions in group testing. In *Conference on Learning Theory*, pages 4764–4781. PMLR, 2022.
- [12] Alex Damian, Loucas Pillaud-Vivien, Jason Lee, and Joan Bruna. Computational-statistical gaps in gaussian single-index models. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 1262–1262. PMLR, 2024.
- [13] Constantinos Daskalakis, Themis Gouleakis, Chistos Tzamos, and Manolis Zampetakis. Efficient statistics, in high dimensions, from truncated samples. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 639–649. IEEE, 2018.
- [14] Constantinos Daskalakis, Themis Gouleakis, Christos Tzamos, and Manolis Zampetakis. Computationally and statistically efficient truncated regression. In *Conference on learning theory*, pages 955–960. PMLR, 2019.
- [15] Anindya De, Shivam Nadimpalli, and Rocco A Servedio. Testing convex truncation. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4050–4082. SIAM, 2023.

- [16] Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 73–84. IEEE, 2017.
- [17] Ilias Diakonikolas, Weihao Kong, and Alistair Stewart. Efficient algorithms and lower bounds for robust linear regression. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2745–2754. SIAM, 2019.
- [18] Yunzi Ding, Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Subexponential-time algorithms for sparse pca. *Foundations of Computational Mathematics*, 24(3):865–914, 2024.
- [19] Jianqing Fan, Han Liu, Zhaoran Wang, and Zhuoran Yang. Curse of heterogeneity: Computational barriers in sparse mixture models and phase retrieval. *arXiv preprint arXiv:1808.06996*, 2018.
- [20] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM (JACM)*, 64(2):1–37, 2017.
- [21] Silvio Franz and Giorgio Parisi. Recipes for metastable states in spin glasses. *Journal de Physique I*, 5(11):1401–1415, 1995.
- [22] Silvio Franz and Giorgio Parisi. Effective potential in glassy systems: theory and simulations. *Physica A: Statistical Mechanics and its Applications*, 261(3-4):317–339, 1998.
- [23] Francis Galton. An examination into the registered speeds of american trotting horses, with remarks on their value as hereditary data. *Proceedings of the Royal Society of London*, 62(379-387):310–315, 1898.
- [24] David Gamarnik and Ilias Zadik. Sparse high-dimensional linear regression. estimating squared error and a phase transition. *The Annals of Statistics*, 50(2):880–903, 2022.
- [25] S Das Gupta, Morris L Eaton, Ingram Olkin, Michael Perlman, Leonard J Savage, and Milton Sobel. Inequalities on the probability content of convex regions for elliptically contoured distributions. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971)*, volume 2, pages 241–265, 1972.
- [26] Samuel Hopkins. *Statistical Inference and the Sum of Squares Method*. PhD thesis, Cornell University, 2018.
- [27] Daniel J Hsu, Clayton H Sanford, Rocco Servedio, and Emmanouil Vasileios Vlatakis-Gkaragkounis. Near-optimal statistical query lower bounds for agnostically learning intersections of halfspaces with gaussian marginals. In *Conference on Learning Theory*, pages 283–312. PMLR, 2022.
- [28] Hidehiko Ichimura. Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of econometrics*, 58(1-2):71–120, 1993.
- [29] Mark Jerrum. Large cliques elude the metropolis process. *Random Structures & Algorithms*, 3(4):347–359, 1992.
- [30] Michael Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM (JACM)*, 45(6):983–1006, 1998.
- [31] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2019.

- [32] Rafał Łatała and Dariusz Matlak. Royen’s proof of the gaussian correlation inequality. In *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2014–2016*, pages 265–275. Springer, 2017.
- [33] Peter McCullagh. *Generalized linear models*. Routledge, 2019.
- [34] Andrea Montanari, Daniel Reichman, and Ofer Zeitouni. On the limitation of spectral methods: From the gaussian hidden clique problem to rank-one perturbations of gaussian tensors. *Advances in Neural Information Processing Systems*, 28, 2015.
- [35] Karl Pearson. On the systematic fitting of curves to observations and measurements. *Biometrika*, 1(3):265–303, 1902.
- [36] Emile Richard and Andrea Montanari. A statistical model for tensor pca. *Advances in neural information processing systems*, 27, 2014.
- [37] Thomas Royen. A simple proof of the gaussian correlation conjecture extended to multivariate gamma distributions. *arXiv preprint arXiv:1408.1028*, 2014.
- [38] Tselil Schramm and Alexander S Wein. Computational barriers to estimation from low-degree polynomials. *The Annals of Statistics*, 50(3):1833–1858, 2022.
- [39] Ilias Zadik, Min Jae Song, Alexander S Wein, and Joan Bruna. Lattice-based methods surpass sum-of-squares in clustering. In *Conference on Learning Theory*, pages 1247–1248. PMLR, 2022.
- [40] Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes], [No], or [NA].
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes]" is generally preferable to "[No]", it is perfectly acceptable to answer "[No]" provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No]" or "[NA]" is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading “NeurIPS Paper Checklist”,**
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers.**

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope?

Answer: [Yes]

Justification: Yes, the abstract and introduction make accurate claims about the paper’s contributions and scope. We describe our results accurately.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: We discuss sufficiently the limitations of our results and provide clear assumptions under which they apply.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [\[Yes\]](#)

Justification: We write clearly all the assumptions that are required for our results to hold. The proofs are all provided in the appendices.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [NA]

Justification: There are no experimental results, as this is a theory paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
 - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [NA]

Justification: There are no experimental results, as this is a theory paper.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.

- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [NA]

Justification: There are no experimental results, as this is a theory paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]

Justification: There are no experimental results, as this is a theory paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.

- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [NA]

Justification: There are no experimental results, as this is a theory paper.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: We have conducted research and presented our work in accordance with the Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [NA]

Justification: There are no direct societal impact of our work that we are aware of.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: There are no such risks that we are aware of.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: Yes, we cite all the relevant works to be acknowledged to establish our results.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets

has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [NA]

Justification: The paper does not release new assets.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: As this is a theory paper with no such crowdsourcing involved on human related subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: As this is a theory paper with no such crowdsourcing involved on human related subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: This paper does not involve the use of LLMs for our scientific methodologies.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.

Contents

1	Introduction	1
1.1	Main Contributions	3
2	Setting and Definitions	5
3	Main Results	6
3.1	The Assumption	6
3.2	The GFP-SQ equivalence	6
3.2.1	Simplifying GFP-hardness	6
3.2.2	The equivalence	7
4	Examples	8
4.1	Gaussian Additive Models	8
4.2	Planted Sparse Models	9
4.2.1	Other examples	9
5	GFP-hardness is not always equal to FP-hardness	10
6	Conclusion	10
A	Discussions on FP criterion and assumptions	23
A.1	Connection of the FP criterion with statistical physics	23
A.2	Necessity of assumptions in main theorem	24
A.2.1	Necessity of Assumption 1	24
A.2.2	Necessity of a non-trivial information-theory threshold	25
B	Equivalence between LD, SQ, and GFP	27
B.1	Low-Degree lower bounds definitions	27
B.1.1	Low-Degree Lower Bounds	27
B.1.2	Low Samplewise Degree Lower Bounds	27
B.2	Unconditional SQ hardness	28
B.3	Noise-robust models and SQ-LD equivalence	29
B.4	Equivalence of GFP, SQ, and LD hardness for noise-robust models	30
C	Proofs of main theorems	32
C.1	Proof of Theorem 2	32
C.2	Proof of Theorem 3	34
D	Details of examples and proofs	36

D.1	Symmetric mixed sparse linear regression	36
D.1.1	Proofs for mSLR	37
D.2	Non-Gaussian component analysis	39
D.2.1	Proofs for Non-Gaussian Component Analysis	40
D.3	Single-index Models	42
D.4	Truncated statistics: convex truncation	42
D.4.1	A new SQ lower bound	43
D.4.2	Proofs for Convex Truncation	43

E Details on the GFP-hardness and FP-hardness separation 45

A Discussions on FP criterion and assumptions

In this Section, we provide additional discussions on the Franz-Parisi criterion and its connection with Statistical physics, as well as, on the necessity of our assumptions in our main theorem (Theorem 3).

A.1 Connection of the FP criterion with statistical physics

We begin by discussing the connection between the Franz-Parisi (FP) criterion and statistical physics methods. For a more detailed overview and additional references, we refer the reader to [6, Section 1.3].

A natural algorithm for solving the estimation problem of recovering u from $Y = (Y_1, \dots, Y_m) \sim \mathbb{P}_u$ is to run some “local” dynamics (e.g., Langevin or Glauber dynamics) to sample from the posterior

$$\mathbb{P}(u|Y) \propto \pi(v)\mathbb{P}(Y|v) = \pi(v) \prod_{i=1}^m \mathbb{P}_v(Y_i), v \in \Theta,$$

where $Y = (Y_1, \dots, Y_m)$. In statistical physics, a powerful heuristic exists for predicting the success of local dynamics in sampling from random distributions of the form $p_Y(v)\nu(v)$, $v \in \Theta$ where ν is a reference measure and $Y \sim \mu$ is a “disorder”. The heuristic approach is to check the monotonicity of the so-called Franz-Parisi potential defined as

$$F(t) = \mathbb{E}_{u \sim \nu, Y \sim \mu} [\log \mathbb{E}_{v \sim \nu} [p_Y(v) 1(d(v, u) = t)]], t \in [0, 1],$$

where $d(\cdot, \cdot)$ is some notion of (normalized) distance between the states u, v in agreement with the operations of the local dynamics on the state space. The prediction, introduced by Franz and Parisi in [21], is that local dynamics can efficiently sample from the distribution if and only if the potential is monotonic, i.e., it lacks “bad” local minima. Remarkably, this prediction has been empirically validated across a range of problems in statistical physics, often yielding accurate forecasts of algorithmic tractability. For instance, when d is the Euclidean distance, this criterion has proven effective in the study of spin glasses [21]. Other, more intricate distance functions have also been used successfully in non-spin glass settings, such as binary fluids [22].

Now, returning to statistical estimation settings, researchers in statistical physics have applied this rule for $p_Y(v) := \mathbb{P}(Y|v) = \prod_{i=1}^m \mathbb{P}_v(Y_i)$ and $\nu := \pi$ to arrive at a prediction of success for “local” algorithms based on the geometry defined by the distance d . The prediction [40] is then based on the monotonicity of the curve

$$F(t) = \mathbb{E}_{u \sim \pi, Y \sim \mathbb{P}_u} [\log \mathbb{E}_{v \sim \pi} (\mathbb{P}(Y|v) 1[d(v, u) = t])], t \in [0, 1],$$

or equivalently for

$$F(t) = \mathbb{E}_{u \sim \pi, Y \sim \mathbb{P}_u} \left[\log \mathbb{E}_{v \sim \pi} \left(\prod_{i=1}^m \frac{\mathbb{P}_v(Y_i)}{\mathbb{Q}(Y_i)} 1(d(v, u) = t) \right) \right], t \in [0, 1], \quad (10)$$

Interestingly, when $d(\cdot, \cdot)$ is the Euclidean distance, recent mathematical works have indeed produced one-sided results linking the potential to the performance of local methods for estimation tasks in the context of the so-called Gaussian additive models (e.g., [3, 4, 6]). This connection with the choice of the Euclidean distance can be perhaps cast as natural by a well-known analogy between spin glasses and GAMs, where GAMs often take the form of "spiked" spin glass models. Now, given the above successes, both heuristic and rigorous, it is natural to conjecture a potential link between general algorithmic hardness and the monotonicity of $F(t)$. However, this connection remains unproven in general, and known counterexamples exist. For instance, in sparse tensor PCA [10], there are regimes where the FP potential is non-monotonic (suggesting hardness), but some polynomial-time methods do succeed.

Despite the above issue, [6] used the Franz-Parisi potential to arrive at a different criterion, but now for algorithmic hardness of detection. Following an application of Jensen's inequality described in [6, Section 1.3] one gets the following "annealed" upper bound for any $t \in [0, 1]$ $F(t) \leq \log \tilde{F}(t)$ for,

$$\tilde{F}(t) = \mathbb{E}_{u,v \sim p, [\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \mathbf{1}[d(v, u) = t]], t \in [0, 1]. \quad (11)$$

Then by focusing on the Euclidean distance d (or equivalently the Euclidean dot product $\langle u, v \rangle$) they suggested the Franz-Parisi (FP) criterion Definition 1, restated here.

Definition 5 (FP hardness). *We say a problem is (q, m, ε) -FP hard if*

$$\mathbf{FP}: \quad \mathbb{E} [\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(|\langle u, v \rangle| \leq \delta(q))] \leq 1 + \varepsilon, \quad \text{where} \quad (12)$$

$$\delta(q) = \sup\{\delta : \pi^2(\langle u, v \rangle \geq \delta) \geq q^{-2}\}, \quad (13)$$

Notice that the FP criterion says that to check for "hardness" of detection one should integrate the annealed FP curve is an $(1 - q^{-2})$ -typical overlap t -region. Moreover, as we elaborated in the Introduction, one should understand q in the above definition as a proxy for q run-time. In that light, [6] roughly proved that for any GAMs is a $(q, m, O(1))$ -FP hard if and only if $D = \log q$ -degree polynomials fail to detect between $\mathbb{P}$ and $\mathbb{Q}$ with m samples. We remark that, albeit this is an equivalence for detection, this is a first-of-a-kind result for GAMs as it is mathematical connection between the FP curve and a rigorous form of hardness. However, [6] also presented counterexamples where this equivalence breaks down when we move away from GAMs.

The central idea of this work is **to optimize over the integration region** in the FP criterion, rather than fixating on the Euclidean dot product. This leads us to propose the Generalized Franz-Parisi (GFP) criterion (see Definition 2). Our motivation arises from the observation that while the Euclidean distance is natural for GAMs (and spin glass models), it may be inappropriate in other statistical settings (see Section 5). This echoes insights from statistical physics, where non-Euclidean distances are used in models beyond spin glasses [22]. Satisfyingly, this generalization enables a broad equivalence with statistical query (SQ) lower bounds, as shown in Theorem 3.

A.2 Necessity of assumptions in main theorem

In this section, we comment on the necessity of our assumptions for the GFP-hardness and SQ-hardness equivalence.

A.2.1 Necessity of Assumption 1

We first show that there is a strong separation between GFP-hardness and SQ-hardness unless some non trivial lower bound on $\min_{u,v} \langle L_u, L_v \rangle_{\mathbb{Q}}$ is assumed, providing support for our Assumption 1. Indeed, we present here a (very) simple counterexample when one allows for $\min_{u,v} \langle L_u, L_v \rangle_{\mathbb{Q}} = 0$.

Let us define a variant of the following model from [6, Section 4.2.] (assume for simplicity n is a multiple of 10 in what follows):

- Under $\mathbb{P}$, we first sample a $u \sim \text{Unif}(\{x \in \{-1, 1\}^n : \sum x_i = 8n/10\})$. Then under $\mathbb{P}_u$ each sample always equals to u , i.e., $\mathbb{P}_u$ is the Dirac measure on u .

- Under $\mathbb{Q}$ for each sample, we sample $u \sim \text{Unif}(\{-1, 1\}^n)$ and output u .

It is easy to see that for all $u, v \in \{x \in \{-1, 1\}^n : \sum x_i = 8n/10\}$,

$$\langle L_u, L_v \rangle_{\mathbb{Q}} = 2^n \mathbf{1}(u = v).$$

We first prove that the task is *not even* $(1, 1)$ -SQ hard. Indeed $(1, 1)$ -SQ hard implies $\mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|] \leq 1$, but

$$\mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|] \geq |\langle L_u, L_u \rangle_{\mathbb{Q}} - 1| \pi^2(u = v) \quad (14)$$

$$= (2^n - 1) \binom{n}{9n/10}^{-1} = \omega(1). \quad (15)$$

On the contrary, we now show that the task is $(e^{n^{\Theta(1)}}, \infty, O(1))$ -GFP-hard, more specifically we prove that it is $(m, q, O(1))$ -GFP-hard for any sample size m and $q = \binom{n}{9n/10}^{1/2} = e^{n^{\Theta(1)}}$. Indeed, we have $\pi^2(u \neq v) = 1 - q^{-2}$ and therefore to prove (m, q) -GFP hardness it suffices

$$\mathbb{E}[\langle L_u, L_v \rangle_{\mathbb{Q}}^m \mathbf{1}(u \neq v)] = O(1).$$

But in fact it even holds $\mathbb{E}[\langle L_u, L_v \rangle_{\mathbb{Q}}^m \mathbf{1}(u \neq v)] = 0$ completing this proof.

A.2.2 Necessity of a non-trivial information-theory threshold

The second assumption that our equivalence operates under is that m_{IT} is non-trivial. We now claim that some non-trivial lower bound on m_{IT} is also necessary for the connection between the notions of GFP-hardness and SQ-hardness, even under Assumption 1.

Indeed, consider the following multisample problem over graphs. Let $n \in \mathbb{N}$, $p = 1 - n^{-1/4}$ and $k = n^{1/3+o(1)}$.

- Under $\mathbb{P}$ we choose a u being a k -clique in K_n , chosen uniformly at random (we see u as a k -vertex set). Then under $\mathbb{P}_u$ one sample consists of the union of a $G(n, p)$ with the k -clique on u .
- Under $\mathbb{Q}$ one sample is a sample from $G(n, p)$.

We note this is a multi-sample variant of the classic planted clique model [29], but on the (very) dense regime, which has been recently used to establish circuit lower bounds for the model in [?].

Now, even for $m = 1$ the problem is information-theoretically solvable; indeed, under $\mathbb{P}$ there is always a k -clique, while under $\mathbb{Q}$ there is no k -clique with probability $1 - o(1)$, and a brute force method can distinguish the two cases. Indeed, by a union bound the probability there is a k -clique under $\mathbb{Q}$ is at most

$$n^k (1 - n^{-1/4})^{\binom{k}{2}} \leq \exp\left(n^{1/3} \log n - \Theta(n^{2/3-1/4})\right) = \exp\left(-\Theta(n^{5/12})\right) = o(1).$$

Moreover, the model satisfies Assumption 1. One can see this because the model is a PSM and use Lemma 2. Alternatively, one can just directly observe that for any u , we have $L_u(G) = \mathbf{1}(u \text{ is a } k\text{-clique in } G) p^{-\binom{k}{2}}$. Hence, for any u, v

$$\langle L_u, L_v \rangle_{\mathbb{Q}} = p^{-\binom{|u \cap v|}{2}} = (1 - n^{-1/4})^{-\binom{|u \cap v|}{2}} \geq 1.$$

Now, we prove that this PSM is not SQ-hard even for $m = 1$ and $q = 1$, but it is GFP-hard even for $m = n^{\Theta(1)}$ -samples.

We first prove that the task is *not* $(1, 1)$ -SQ-hard. Notice that $(1, 1)$ -SQ hardness is equivalent with the condition $\mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|] \leq 1$. But in this context

$$\mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|] \geq |\langle L_u, L_u \rangle_{\mathbb{Q}} - 1| \pi^2(u = v)$$

$$\begin{aligned}
&= (1 - n^{-1/4})^{\Theta(k^2)} \binom{n}{k} \\
&= \exp \left(\Theta(k^2 n^{-1/4}) - \Theta(k \log n) \right) \\
&= \exp \left(\Theta(n^{5/12}) \right) = \omega(1).
\end{aligned}$$

On the contrary, the task is $(e^{n^{\Theta(1)}}, n^{1/8}, O(1))$ -GFP-hard. Fix m samples. Notice that for i.i.d. $u, v \sim \pi$, the overlap $|u \cap v|$ follows an (n, k, k) -Hypergeometric. Hence by [6, Lemma 6.6.] for any $q > 0$ if $\delta = \log(k^2 q) > 0$ it holds

$$\pi^2(|u \cap v| \geq \delta) \leq k(k^2/n)^\delta \leq k2^{-\delta} = q^{-2}.$$

Hence, to prove GFP-hardness it suffices to prove that $\mathbb{E}[\langle L_u, L_v \rangle_{\mathbb{Q}}^m \mathbf{1}(|u \cap v| \leq \delta)] = O(1)$. Therefore for $q = \exp(n^\alpha)$ for some $\alpha > 0$, and $m = n^{1/8+o(1)}$,

$$\begin{aligned}
\mathbb{E}[\langle L_u, L_v \rangle_{\mathbb{Q}}^m \mathbf{1}(|u \cap v| \leq \delta)] &\leq \mathbb{E}[(1 - n^{-1/4})^{-m \binom{|u \cap v|}{2}} \mathbf{1}(|u \cap v| \leq \delta)] \\
&\leq (1 - n^{-1/4})^{-m \binom{\delta}{2}} \\
&= (1 - n^{-1/4})^{-\Theta(m(\log(kq^2))^2)} \\
&\leq \exp \left(\Theta(mn^{-1/4}(\log(kq^2))^2) \right) \\
&= \exp \left(n^{-1/8+2\alpha} \right) = O(1),
\end{aligned}$$

where the last equality hold say for any $0 < \alpha < 1/16$. Hence, choosing say $q = e^{n^{1/32}}$ concludes the proof.

B Equivalence between LD, SQ, and GFP

In this Appendix, we discuss the equivalence between GFP and low-degree (LD) polynomial hardness. This result is obtained by combining the GFP-SQ equivalence from Section 3 with the equivalence between SQ and LD hardness under noise robustness proved by Brennan et al. [8]. We recall and provide a succinct proof of Brennan et al.’s result for completeness.

B.1 Low-Degree lower bounds definitions

We start by recalling the definition of a low-degree lower bound. The definition is based on the *low-degree likelihood ratio* $L^{\leq D}$, where we recall that $L^{\leq D}$ denotes the projection of the likelihood ratio onto the subspace of degree-at-most- D polynomials.

B.1.1 Low-Degree Lower Bounds

The following is the standard definition of Low-Degree hardness as originally stated, for example, in [26].

Definition 6 (Low-Degree Likelihood Ratio). *For m samples, define the squared norm of the degree- D likelihood ratio (also called the “low-degree likelihood ratio”) to be the quantity*

$$\text{LD}(D) := \|L_m^{\leq D}\|_{\mathbb{Q}}^2 = \left\| (\mathbb{E}_{u \sim \pi} L_u^{\otimes m})^{\leq D} \right\|_{\mathbb{Q}}^2 = \mathbb{E}_{u, v \sim \pi} [\langle (L_u^{\otimes m})^{\leq D}, (L_v^{\otimes m})^{\leq D} \rangle_{\mathbb{Q}}] . \quad (16)$$

For some increasing sequence $D = D_n$, we say that the hypothesis testing problem above is hard for the degree- D likelihood or simply D -degree hard if $\text{LD}(D) = O(1)$.

While we direct the reader to [6, Section 1.2] a relation between the Low-degree likelihood ratio and the performance of low-degree algorithms we highlight some key conjectures in the community.

- We expect the class of degree- D polynomials to be as powerful as all $\exp(\tilde{\Theta}(D))$ -time tests (which is the runtime needed to naively evaluate the polynomial term-by-term). Thus, if $\text{LD}(D) = O(1)$ (or $1 + o(1)$), we take this as evidence that strong (or weak, respectively) detection requires runtime $e^{\tilde{\Omega}(D)}$; see Hypothesis 2.1.5 of [26].
- On a finer scale, we expect the class of degree- $O(\log n)$ polynomials to be at least as powerful as all polynomial-time tests. Thus, if $\text{LD}(D) = O(1)$ (or $1 + o(1)$) for some $D = \omega(\log n)$, we take this as evidence that strong (or weak, respectively) detection cannot be achieved in polynomial time; see Conjecture 2.2.4 of [26].

We emphasize that the above statements are not true in general (see for instance [39] for some discussion of counterexamples) and depend on the choice of $\mathbb{P}$ and $\mathbb{Q}$, yet remarkably often appear to hold up for a broad class of distributions arising in high-dimensional statistics.

B.1.2 Low Samplewise Degree Lower Bounds

In multisample settings like ours, a similar notion of “samplewise” low degree lower bounds have been considered in [8].

Definition 7. For $d, k \in \mathbb{N} \cup \{\infty\}$ a function $f : (\mathbb{R}^n)^{\otimes m} \rightarrow \mathbb{R}$ has *samplewise degree* (d, k) if it can be written as a linear combination of functions which have degree at most d in each x_i and non-zero degree in at most k of the x_i ’s (if $d < \infty$ the function is therefore a polynomial).

Let’s state the hardness criterion associated with this samplewise low degree polynomials:

Definition 8 (Low Degree (LD) Hardness). *We say a “ $\mathbb{P}$ versus $\mathbb{Q}$ ” detection problem is (m, d, k, ε) -LD hard if*

$$\text{LD:} \quad \mathbb{E} [\langle (L_u^{\otimes m})^{\leq d, k}, (L_v^{\otimes m})^{\leq d, k} \rangle] \leq 1 + \varepsilon. \quad (17)$$

Notice that this notion of (d, k) -low degree hardness is the natural generalization to (16). As a point of comparison, dk -degree polynomials contain all (d, k) -degree polynomials and (d, d) -degree polynomials contain all d -degree polynomial.

Remark B.1 (Explaining Remark 3.3). A nice property of the low samplewise-degree degree projection is that it is easy to relate it to d -degree projections. Indeed, using a binomial expansion argument (see [8, Claim 3.3.]),

$$\|L_m^{\leq(d,k)}\|_{\mathbb{Q}}^2 = \mathbb{E}_{u,v \sim \pi} \left[\langle (L_u^{\otimes m})^{\leq(d,k)}, (L_v^{\otimes m})^{\leq(d,k)} \rangle_{\mathbb{Q}} \right] = \sum_{t=0}^m \binom{m}{t} \mathbb{E}_{u,v \sim \pi} [(\langle L_u^{\leq d}, L_v^{\leq d} \rangle_{\mathbb{Q}} - 1)^t].$$

In particular, if $k = 1, d = \infty$, since $\mathbb{E}_{u,v \sim \pi} [\langle L_u, L_v \rangle_{\mathbb{Q}} - 1] = \chi^2(\mathbb{P}, \mathbb{Q})$ we have

$$\|L_m^{\leq(\infty,1)}\|_{\mathbb{Q}}^2 = 1 + m\chi^2(\mathbb{P}, \mathbb{Q}).$$

In particular, notice that the condition $m\chi^2(\mathbb{P}, \mathbb{Q}) = O(1)$ discussed in Theorem 2 and Remark 3.3 is equivalent with a samplewise $(\infty, 1)$ -degree lower bound for the task, i.e., a lower bound against function that are linear combination of functions of one sample at a time. In [8] the authors prove that SQ lower bounds are (almost) equivalent with sample-wise degree lower bounds, therefore it is perhaps no surprise that the condition $m\chi^2(\mathbb{P}, \mathbb{Q}) = O(1)$ can be also explained as a (very) weak consequence of any SQ lower bounds against m samples. Indeed, assume a $\mathbb{P}$ versus $\mathbb{Q}$ detection problem is (q, m) -SQ hard for any q (even $q = 1$). Then setting $A = \text{support}(\pi)^{\otimes 2}$ we have that it must hold $m\mathbb{E}_{u,v \sim \pi} [|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|] \leq 1$ and therefore

$$m\chi^2(\mathbb{P}, \mathbb{Q}) = m\mathbb{E}_{u,v \sim \pi} [\langle L_u, L_v \rangle_{\mathbb{Q}} - 1] \leq 1.$$

B.2 Unconditional SQ hardness

Before stating the equivalence between the above LD-hardness criterion and SQ-hardness, we define an *Unconditional Statistical Query* (USQ) hardness criterion, which is equivalent to SQ and often appears as a convenient intermediate step in proofs. This hardness measure appeared, often implicitly, in several prior work (e.g., [8]):

Definition 9 (Unconditional SQ hardness). *We say a “ $\mathbb{P}$ versus $\mathbb{Q}$ ” detection problem is (m, t) -unconditional SQ hard for some even t if*

$$\text{USQ: } \mathbb{E} [\chi_{\mathbb{Q}}(\mathbb{P}_u, \mathbb{P}_v)^t] \leq m^{-t}. \quad (18)$$

The USQ criterion removes the conditioning on event A from the SQ criterion, which makes it much easier to manipulate. USQ hardness is essentially equivalent to SQ hardness as stated in the next proposition:

Proposition 1 (Equivalence USQ and SQ hardness).

- (i) *If a model is (m, t) -USQ hard, then it is $(q, m/q^{2/t})$ -SQ hard for all integers $q \geq 1$.*
- (ii) *If a model is $(q, m/q^{2/t})$ -SQ hard for all integers $q \geq 1$, then it is (m', t') -USQ hard for all $t' < t$ and $m' \leq m \cdot 2^{-1/t}(t - t')^{1/t'}$.*

For simplicity, for $t \geq 4$, we can set $t' = t/2$ and $m' = m$ in Proposition 1.(ii). Proposition 1 was proven in [8]. We provide a succinct proof for completeness.

Proof of Proposition 1. USQ hardness implies SQ hardness. By Hölder’s inequality,

$$\begin{aligned} \mathbb{E} [|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1| \mid A] &\leq \frac{(\mathbb{E} [|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^t])^{1/t} \cdot (\mathbb{E} [\mathbf{1}[(u, v) \in A]])^{1-1/t}}{\pi^2(A)} \\ &= \left(\frac{\mathbb{E} [|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^t]}{\pi^2(A)} \right)^{1/t}. \end{aligned}$$

Assuming that we have (m, t) -USQ hardness, this implies that for any $q \geq 1$,

$$\sup_{A: \pi^2(A) \geq q^{-2}} \mathbb{E} [|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1| | A] \leq \frac{q^{2/t}}{m},$$

which establishes the $(q, m/q^{2/t})$ -SQ hardness.

SQ hardness implies USQ hardness. For convenience, introduce the random variable $X = |\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|$ with $(u, v) \sim \pi^2$. Assume that we have $(q, m/q^{2/t})$ -SQ hardness for all $q \geq 1$. In particular, for all A , we have

$$\mathbb{E}[X|A] \leq \frac{1}{\pi^2(A)^{1/t} m}.$$

Using [8, Fact 4.3], we have for every $t > t' > 0$,

$$\mathbb{E}[X^{t'}] \leq \left(2 \sup_A \pi^2(A) \cdot \mathbb{E}[X|A]^t \right)^{t'/t} \cdot \frac{t}{t-t'} \leq \frac{2^{t'/t}}{m^{t'/t}} \cdot \frac{t}{t-t'},$$

which establishes (t', m') -USQ hardness for any $t' < t$ and $m' = m \cdot 2^{-1/t}(t-t')^{1/t'}$. $\square$

B.3 Noise-robust models and SQ-LD equivalence

An advantage of USQ is that it is directly related to Low Degree lower bounds: USQ hardness is equivalent to LD hardness with $d = \infty$, that is, with no degree-constraint on each sample in the projection.

Proposition 2 (Equivalence between USQ and LD hardness with $d = \infty$).

- (i) If a model is $(m, \infty, k, \varepsilon)$ -LD hard, then it is (m', k) -USQ hard with $m' = m/(k\varepsilon^{1/k})$.
- (ii) If a model is (m, k) -USQ hard, it is $(m, \infty, k, e-1)$ -LD hard. More generally, it will be $(m', \infty, k, em'/m)$ -LD hard for all $m' < m$.

Proof of Proposition 2. We follow the proof in [8]. Assume that the model is $(m, \infty, k, \varepsilon)$ -LD hard. Then

$$\|\mathbb{E}_u[(L_u^{\otimes m})^{\leq \infty, k}] - 1\|_{\mathbb{Q}}^2 = \sum_{s=1}^k \binom{m}{s} \mathbb{E}_{u,v}[\chi_{\mathbb{Q}}(\mathbb{P}_u, \mathbb{P}_v)^s] \leq \varepsilon,$$

and in particular, for k even,

$$\|\mathbb{E}_u[(L_u^{\otimes m})^{\leq \infty, k}] - 1\|_{\mathbb{Q}}^2 - \|\mathbb{E}_u[(L_u^{\otimes m})^{\leq \infty, k-1}] - 1\|_{\mathbb{Q}}^2 = \binom{m}{k} \mathbb{E}_{u,v}[\chi_{\mathbb{Q}}(\mathbb{P}_u, \mathbb{P}_v)^k] \leq \varepsilon.$$

This implies that $\mathbb{E}_{u,v}[\chi_{\mathbb{Q}}(\mathbb{P}_u, \mathbb{P}_v)^k] \leq \varepsilon / \binom{m}{k} \leq \varepsilon k^k / m^k$. On the other hand, (m, k) -USQ hardness implies that

$$\|\mathbb{E}_u[(L_u^{\otimes m})^{\leq \infty, k}] - 1\|_{\mathbb{Q}}^2 \leq \sum_{s=1}^k \frac{m^s}{s!} \mathbb{E}_{u,v}[\chi_{\mathbb{Q}}(L_u, L_v)^s] \leq (e-1).$$

More generally, we will have for $m' < m$

$$\|\mathbb{E}_u[(L_u^{\otimes m'})^{\leq \infty, k}] - 1\|_{\mathbb{Q}}^2 \leq \sum_{s=1}^k \frac{(m')^s}{s!} \mathbb{E}_{u,v}[\chi_{\mathbb{Q}}(L_u, L_v)^s] \leq \sum_{s=1}^k \frac{1}{s!} \left(\frac{m'}{m} \right)^s \leq e \frac{m'}{m},$$

which concludes the proof. $\square$

Combining this equivalence of USQ and LD($d = \infty$) with the equivalence between USQ and SQ in Proposition 1, we can directly state an (unconditional) equivalence between SQ and LD($d = \infty$) hardness. In order to transfer this equivalence to LD with $d < \infty$, one can assume that the model with $d = \infty$ and $d < \infty$ are close to each other: this assumption is equivalent to being *noise-robust* (in some sense, see discussions in [8]).

Assumption 2 (Noise robustness). We say a “ $\mathbb{P}$ versus $\mathbb{Q}$ ” detection problem is (d, k, δ) -noise robust if

$$\|\mathbb{E}_u[(L_u^{>d})^{\otimes k}]\|_{L^2(\mathbb{Q})}^2 \leq \delta. \quad (19)$$

Under this assumption, one can state the equivalence between LD and USQ:

Proposition 3 (Equivalence between LD and USQ Hardness).

- (i) If the model is (m, t) -USQ hard, then the model is also $(m', d, k', em'/m)$ -LD hard for all $m' \leq m$, $k' \leq t$ and $d \geq 1$.
- (ii) If the model is (m, d, k, ε) -LD hard and we further assume that it is (d, k, δ) -noise robust, then the model is (m', k) -USQ hard with

$$m' = \frac{m}{m\delta^{1/k} + k\varepsilon^{1/k}}.$$

Proof of Proposition 3. Part (i) is directly implied by Proposition 2.(ii). For part (ii), following the same argument as in the proof of Proposition 2.(i), we get

$$\mathbb{E}[|\langle L_u^{\leq d}, L_v^{\leq d} \rangle - 1|^k] \leq \varepsilon \frac{k^k}{m^k}.$$

Then using [8, Lemma 3.4], we obtain

$$\begin{aligned} \mathbb{E}[|\langle L_u, L_v \rangle - 1|^k]^{1/k} &\leq \mathbb{E}[|\langle L_u^{\leq d}, L_v^{\leq d} \rangle - 1|^k]^{1/k} + \mathbb{E}[|\langle L_u^{>d}, L_v^{>d} \rangle|^k]^{1/k} \\ &\leq \varepsilon^{1/k} \frac{k}{m} + \delta^{1/k} = \frac{k\varepsilon^{1/k} + m\delta^{1/k}}{m}, \end{aligned}$$

which concludes the proof. $\square$

Then, the equivalence between LD and SQ hardness in [8] is obtained by combining Proposition 3 and Proposition 1. We state it below for completeness:

Theorem 5 (Equivalence between LD and SQ hardness).

- (i) If the model is $(q, m/q^{2/t})$ -SQ hard for all $q \geq 1$ (with $t \geq 4$, for simplicity), then it is $(m', d, k', em'/m)$ -LD hard for all $m' \leq m$, $k' \leq t/2$, and $d \geq 1$.
- (ii) If the model is (m, d, k, ε) -LD hard and we further assume that it is (d, k, δ) -noise robust, then the model is $(q, m'/q^{2/t})$ -SQ hard for all $q \geq 1$ with

$$m' = \frac{m}{m\delta^{1/k} + k\varepsilon^{1/k}}.$$

B.4 Equivalence of GFP, SQ, and LD hardness for noise-robust models

Based on the SQ-LD equivalence stated in the previous section (Theorem 5) and the equivalence between GFP and SQ (Theorem 3), we can state an equivalence between GFP and LD hardness for noise-robust models.

Theorem 6 (LD and ρ_G -FP Equivalence). Suppose a “ $\mathbb{P}$ versus $\mathbb{Q}$ ” task satisfies Assumption 1 for a group G .

- (i) If the model is (m, d, k, ε) -LD hard and we further assume that it is (d, k, δ) -noise robust, then the model is $(q', m', e|G|^{-1}\tilde{m}q^{2/t}/\tilde{m})$ - ρ_G -FP hard for any integers $q \geq 1$, $q' \leq q/\sqrt{2}$, and $m' \leq \tilde{m}/2$, with

$$\tilde{m} = \frac{m}{m\delta^{1/k} + k\varepsilon^{1/k}}.$$

(ii) *If a task is (q, m, ε) - ρ_G -FP hard for some q, m integers. Assume that there exists an $r = r(q) > 0$ such that $\pi^2(\rho_G(u, v) < r) = 1 - q^{-2}$ and m is even. Then, for all even integer $4 \leq t \leq \log(q)/\log(m)$, the model is also $(m', d, k', em'/\tilde{m})$ -LD hard for all $m' \leq \tilde{m}$ and $k' \leq t/2$, and $d \geq 1$, where*

$$\tilde{m} = \frac{m}{t(1 + \varepsilon)^{1/t} + \chi^2(\mathbb{P}^{\otimes 4t} \parallel \mathbb{Q}^{\otimes 4t})}.$$

Note that the implication of GFP hardness to LD hardness is unconditional. LD hardness with $d = \infty$ implies GFP hardness, while for $d < \infty$, this implication holds under the noise robustness assumption.

C Proofs of main theorems

C.1 Proof of Theorem 2

Proof of Theorem 2. It is clear that (q, m, ε) - ρ_G -FP hard implies (q, m, ε) -GFP $_G$ hard as for the event

$$A := \{\rho_G(u, v) < r(q)\},$$

it clearly holds $\pi^{\otimes 2}(A) \geq 1 - q^{-2}$ and, since G is a group, $G^{\otimes 2}(A) = A$. Hence,

$$\inf_{\substack{A: \pi^{\otimes 2}(A) \geq 1 - q^{-2} \\ G^{\otimes 2}(A) = A}} \mathbb{E} \left[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} \mathbf{1}(A) \right] \leq \mathbb{E} \left[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} \cdot \mathbf{1}(\rho_G(u, v) < r(q)) \right] \leq 1 + \varepsilon,$$

implying the desired result.

We now focus on the other direction. By decomposing the likelihood ratio inner product, we obtain

$$\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} = (\langle L_u, L_v \rangle_{\mathbb{Q}} - 1 + 1)^m = \sum_{t=0}^m \binom{m}{t} \cdot (\langle L_u, L_v \rangle_{\mathbb{Q}} - 1)^t. \quad (20)$$

Taking expectation over the prior $\pi^{\otimes 2}$ conditioned on *any* event A satisfying $G^{\otimes 2}(A) = A$ and $\pi^2(A) = 1 - q^{-2}$, we have

$$\begin{aligned} \mathbb{E}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} | A] &= \sum_{t=0}^m \binom{m}{t} \cdot \mathbb{E}[(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1)^t | A] \\ &= \sum_{t=1}^m \binom{m}{t} \cdot \left(\mathbb{E}_{g \sim \text{Unif}(G)} \mathbb{E}[(\langle L_{g(u)}, L_{g(v)} \rangle_{\mathbb{Q}} - 1)^t | A] \right) + 1 \\ &\geq \sum_{t=1}^{\lfloor m/2 \rfloor} \binom{m}{2t} \cdot \left(\mathbb{E}[\mathbb{E}_{g \sim \text{Unif}(G)} (\langle L_{g(u)}, L_{g(v)} \rangle_{\mathbb{Q}} - 1)^{2t} | A] \right) + 1. \end{aligned}$$

where in the second equality, we used that G is a π -preserving transformation, and for the inequality we use Assumption 1.

Clearly for all $t \geq 0$,

$$\mathbb{E}_{g \sim \text{Unif}(G)} (\langle L_{g(u)}, L_{g(v)} \rangle_{\mathbb{Q}} - 1)^{2t} \geq |G|^{-1} \rho_G(u, v)^{2t}. \quad (21)$$

Therefore, we further conclude that

$$\mathbb{E}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} - 1 | A] \geq |G|^{-1} \sum_{t=1}^{\lfloor m/2 \rfloor} \binom{m}{2t} \cdot \mathbb{E}[\rho_G(u, v)^{2t} | A]. \quad (22)$$

Recall that $r(q)$ satisfies

$$\pi^2((u, v) : \rho_G(u, v) \leq r(q)) = \pi^2(A) = 1 - q^{-2}.$$

Hence, by definition of $r(q)$ we have

$$|G|^{-1} \sum_{t=1}^{\lfloor m/2 \rfloor} \binom{m}{2t} \cdot \mathbb{E}[\rho_G(u, v)^{2t} | A] \geq |G|^{-1} \sum_{t=1}^{\lfloor m/2 \rfloor} \binom{m}{2t} \cdot \mathbb{E}[\rho_G(u, v)^{2t} | \rho_G(u, v) \leq r(q)] \quad (23)$$

$$\geq |G|^{-1} \sum_{t=1}^{\lfloor m/2 \rfloor} \binom{m}{2t} \cdot \mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^{2t} | \rho_G(u, v) \leq r(q)]. \quad (24)$$

In addition, we notice that for each even order $2t+1$ with $t = 1, \dots, \lfloor m/2 \rfloor - 1$, it holds by Lemma 4 that

$$\frac{\binom{m}{2t+1} \cdot |\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^{2t+1}}{\sqrt{\binom{m}{2t} \cdot |\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^{2t} \cdot \binom{m}{2t+2} \cdot |\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^{2t+2}}} = \frac{\binom{m}{2t+1}}{\sqrt{\binom{m}{2t} \cdot \binom{m}{2t+2}}} \leq 2.$$

Therefore, using the inequality $2\sqrt{ab} \leq a + b$ for $a, b \geq 0$, we obtain

$$\binom{m}{2t+1} \cdot |\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^{2t+1} \leq \binom{m}{2t} \cdot |\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^{2t} + \binom{m}{2t+2} \cdot |\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^{2t+2}.$$

Consequently, the right-hand-side of (24) can be further lower bounded by

$$|G|^{-1} \sum_{t=1}^{\lfloor m/2 \rfloor} \binom{m}{2t} \cdot \mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^{2t} \mid \rho_G(u, v) \leq r(q)] \quad (25)$$

$$\geq \frac{|G|^{-1}}{3} \sum_{t=2}^m \binom{m}{t} \cdot \mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^t \mid \rho_G(u, v) \leq r(q)] \quad (26)$$

$$\geq \frac{|G|^{-1}}{3} \sum_{t=2}^m \binom{m}{t} \cdot \mathbb{E}[(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1)^t \mid \rho_G(u, v) \leq r(q)]. \quad (27)$$

Combining (22), (24), and (27), with the condition of (q, m, ε) -GFP $_{\mathcal{T}}$ hardness, we obtain

$$\sum_{t=2}^m \binom{m}{t} \cdot \mathbb{E}[(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1)^t \mid \rho_G(u, v) \leq r(q)] \leq 3|G| \cdot \mathbb{E}[(\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} - 1) \mid A]. \quad (28)$$

Again, by the definition of $r(A)$ it follows that

$$\sum_{t=2}^m \binom{m}{t} \cdot \mathbb{E}[(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1)^t \cdot \mathbf{1}(\rho_G(u, v) \leq r(q))] \leq 3|G| \mathbb{E}[(\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} - 1) \mathbf{1}(A)]$$

and therefore by (20)

$$\begin{aligned} & \mathbb{E}[(\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} - 1) \mathbf{1}(\rho_G(u, v) \leq r(q))] \\ & \leq 3|G| \mathbb{E}[(\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} - 1) \mathbf{1}(A)] + m \mathbb{E}[(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1) \cdot \mathbf{1}(\rho_G(u, v) \leq r(q))] \end{aligned}$$

Next, we aim to upper bound the first order term, namely $m \cdot \mathbb{E}[(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1) \cdot \mathbf{1}(\rho_G(u, v) \leq r(q))]$. Note that $A' := \{(u, v) : \rho_G(u, v) \leq r(q)\}$ is also $G^{\otimes 2}$ -invariant. Hence, employing also Assumption 1 we also have

$$\begin{aligned} & \mathbb{E}[(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1) \cdot \mathbf{1}(\rho_G(u, v) \leq r(q))] \\ & = \mathbb{E}[(\mathbb{E}_{g \sim \text{Unif}(G)} \langle L_g(u), L_g(v) \rangle_{\mathbb{Q}} - 1) \cdot \mathbf{1}(\rho_G(u, v) \leq r(q))] \\ & \leq \mathbb{E}[(\mathbb{E}_{g \sim \text{Unif}(G)} \langle L_g(u), L_g(v) \rangle_{\mathbb{Q}} - 1)] \\ & = \mathbb{E}[(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1)] \\ & = \chi^2(\mathbb{P}, \mathbb{Q}). \end{aligned}$$

Therefore,

$$\mathbb{E}[(\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} - 1) \mathbf{1}(\rho_G(u, v) \leq r(q))] \leq 3|G| \mathbb{E}[(\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} - 1) \mathbf{1}(A)] + m \chi^2(\mathbb{P}, \mathbb{Q}).$$

from which the result follows. $\square$

Lemma 4. For any $t \in \{1, 2, \dots, n-1\}$ and $n \geq 3$, we have

$$\frac{\binom{n}{t}^2}{\binom{n}{t-1} \cdot \binom{n}{t+1}} \leq 4. \quad (29)$$

Proof. Note that by the successive ratio between binomial coefficients, we have

$$\frac{\binom{n}{t}^2}{\binom{n}{t-1} \cdot \binom{n}{t+1}} = 1 + \frac{1+n}{t(n-t)} \leq 1 + \frac{1+n}{n-1} = 2 + \frac{2}{n-1} \leq 4. \quad (30)$$

This completes the proof. $\square$

C.2 Proof of Theorem 3

Proof of Theorem 3. SQ implies ρ_G -FP. We have that

$$\sup_{A: \pi^2(A) \geq q^{-2}} \mathbb{E} [|\langle L_u, L_v \rangle - 1| \mid A] \leq \frac{1}{m}. \quad (31)$$

Now as G is π -preserving that easily implies that for any A such that $G^{\otimes 2}(A) = A$, that

$$\sup_{A: \pi^2(A) \geq q^{-2}} \mathbb{E} [\rho_G(u, v) \mid A] \leq |G| \sup_{A: \pi^2(A) \geq q^{-2}} \mathbb{E} [\mathbb{E}_{g \sim \text{Unif}(G)} |\langle L_{g(u)}, L_{g(v)} \rangle - 1| \mid A] \leq \frac{|G|}{m}.$$

Hence for any $r > 0$ if we set $A_r = \{\rho_G(u, v) \geq r\}$ since $G^{\otimes 2}(A) = A$ we conclude that $\pi^2(A_r) \geq q^{-2}$ implies $r \leq |G|/m$. Recall that $r(q) > 0$ satisfies $\pi^2(A_{r(q)}) \geq q^{-2}$. In particular, $r(q) \leq |G|/m$, and therefore for any $m' \leq m/2$,

$$\begin{aligned} & \mathbb{E} [\langle L_u^{\otimes m'}, L_v^{\otimes m'} \rangle_{\mathbb{Q}} \cdot \mathbf{1}(\rho_G(u, v) \leq r(q))] \\ &= \mathbb{E} [(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1 + 1)^{m'} \cdot \mathbf{1}(\rho_G(u, v) \leq r(q))] \\ &\leq \mathbb{E} [(\rho_G(u, v) + 1)^{m'} \cdot \mathbf{1}(\rho_G(u, v) \leq r(q))] \end{aligned} \quad (32)$$

$$\leq (r(q) + 1)^{m'} \leq (|G|/m + 1)^{m'} \leq 1 + e|G|m'/m. \quad (33)$$

This concludes the $(q, m', e|G|m'/m)$ - ρ_G -FP hardness.

ρ_G -FP hardness implies SQ-hardness. Suppose we have (q, m, ε) - ρ_G -FP hardness

$$\mathbb{E} [\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} \cdot \mathbf{1}(\rho_G(u, v) < r(q))] \leq 1 + \varepsilon, \quad \text{where} \quad \pi^2(\rho_G(u, v) \geq r(q)) = q^2.$$

By definition 4, we have that

$$\begin{aligned} 1 + \varepsilon &\geq \mathbb{E} [(\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} - 1) \cdot \mathbf{1}(\rho_G(u, v) < r(q))] \\ &= \mathbb{E} \left[\sum_{t=1}^m \binom{m}{t} \cdot (\langle L_u, L_v \rangle_{\mathbb{Q}} - 1)^t \cdot \mathbf{1}(\rho_G(u, v) < r(q)) \right] \\ &= \mathbb{E} \left[\sum_{t=1}^m \binom{m}{t} \mathbb{E}_{g \sim \text{Unif}(G)} [(\langle L_{g(u)}, L_{g(v)} \rangle_{\mathbb{Q}} - 1)^t] \cdot \mathbf{1}(\rho_G(u, v) < r(q)) \right], \end{aligned}$$

where the first inequality holds by the definition of the ρ_G -FP hardness and the second equality holds by using the elementary $\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} = (\langle L_u, L_v \rangle_{\mathbb{Q}} - 1 + 1)^m$. The last equality holds by using that G is π -measure preserving. As crucially $\mathbb{E}_{g \sim \text{Unif}(G)} [(\langle L_{g(u)}, L_{g(v)} \rangle_{\mathbb{Q}} - 1)^t] \geq 0$ for all integers $t \geq 0$, we have

$$\begin{aligned} & \mathbb{E} \left[\sum_{t=1}^m \binom{m}{t} \mathbb{E}_{g \sim \text{Unif}(G)} [(\langle L_{g(u)}, L_{g(v)} \rangle_{\mathbb{Q}} - 1)^t] \cdot \mathbf{1}(\rho_G(u, v) < r(q)) \right] \\ &\geq \mathbb{E} \left[\sum_{\substack{t \leq m, \\ t \text{ even}}} \binom{m}{t} \cdot \mathbb{E}_{g \sim \text{Unif}(G)} [(\langle L_{g(u)}, L_{g(v)} \rangle_{\mathbb{Q}} - 1)^t] \cdot \mathbf{1}(\rho_G(u, v) < r(q)) \right] \\ &\geq \max_{\substack{1 \leq t \leq m, \\ t \text{ even}}} \mathbb{E} \left[\binom{m}{t} \cdot (\langle L_u, L_v \rangle_{\mathbb{Q}} - 1)^t \cdot \mathbf{1}(\rho_G(u, v) < r(q)) \right]. \end{aligned}$$

Hence, combining the two for all even t , with $1 \leq t \leq m$,

$$\max_{\substack{1 \leq t \leq m, \\ t \text{ even}}} \mathbb{E} [(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1)^t \cdot \mathbf{1}(\rho_G(u, v) < r(q))] \leq \frac{1 + \varepsilon}{\binom{m}{t}}. \quad (34)$$

Therefore, we have for any even t with $t \leq m$ that

$$\begin{aligned}\mathbb{E}[(\langle L_u, L_v \rangle - 1)^t] &= \mathbb{E}[(\langle L_u, L_v \rangle - 1)^t \mathbf{1}(\rho_G(u, v) < r(q))] + \mathbb{E}[(\langle L_u, L_v \rangle - 1)^t \mathbf{1}(\rho_G(u, v) \geq r(q))] \\ &\leq \frac{1 + \varepsilon}{\binom{m}{t}} + \mathbb{E}[(\langle L_u, L_v \rangle - 1)^{2t}]^{1/2} \cdot q^{-1} \\ &\leq \left(\frac{t(1 + \varepsilon)^{1/t}}{m} + \frac{\chi^2(\mathbb{P}^{\otimes 4t} \parallel \mathbb{Q}^{\otimes 4t})^{1/2t}}{q^{1/t}} \right)^t.\end{aligned}$$

where in the first inequality, we use the Cauchy-Schwarz inequality for the second term and the fact that $\pi^2(\rho_G(u, v) \geq r(q)) \leq q^2$. In the second term, we use the elementary $\binom{m}{t} \geq (m/t)^t$.

Now focusing on $t \leq \log q / \log m$ we further have

$$\mathbb{E}[(\langle L_u, L_v \rangle - 1)^t] \leq \left(\frac{t(1 + \varepsilon)^{1/t} + \chi^2(\mathbb{P}^{\otimes 4t} \parallel \mathbb{Q}^{\otimes 4t})}{m} \right)^t. \quad (35)$$

Hence the model is $(\frac{m}{t(1+\varepsilon)^{1/t} + \chi^2(\mathbb{P}^{\otimes 4t} \parallel \mathbb{Q}^{\otimes 4t})}, t)$ -USQ hard. By Proposition 1 we conclude for any $q' > 0$ that the model is $(q', \frac{m(q')^{-2/t}}{t(1+\varepsilon)^{1/t} + \chi^2(\mathbb{P}^{\otimes 4t} \parallel \mathbb{Q}^{\otimes 4t})})$ -SQ hard. The second part follows by setting $t = (\log m)^s$ and $q' = e^{\delta(\log m)^{s+1}}$. $\square$

D Details of examples and proofs

D.1 Symmetric mixed sparse linear regression

The symmetric mixed sparse linear regression (mSLR) setting, is a $\mathbb{P}$ versus $\mathbb{Q}$ detection task defined as follows. Given $k, n \in \mathbb{N}$ with $k \leq n$ and $\sigma^2 > 0$, we have:

- Under the planted model, we first sample $u \sim \pi$ uniformly from set $u \in \{0, 1\}^n$ with $\|u\|_0 = k$. Then, the sample $(x_i, y_i) \sim \mathbb{P}_u$ is generated by

$$y_i = (k + \sigma)^{-1} [z_i \odot \langle x_i, u \rangle + (1 - z_i) \odot \langle x_i, -u \rangle + w_i],$$

for independent $w_i \sim \mathcal{N}(0, \sigma^2)$, $x_i \sim \mathcal{N}(0, I_n)$ and $z_i \sim \text{Bern}(1/2)$. Following [5] we also denote $\text{SNR} := k/\sigma^2$.

- Under the null model, the sample $(y_i, x_i) \sim \mathbb{Q}$ is generated by $y_i \sim \mathcal{N}(0, 1)$ and independently $x_i \sim \mathcal{N}(0, I_n)$.

To see that mSLR is a PSM, set for any u , $\phi_u := \text{support}(u) \cup \{n+1\}$, that is the coordinates of $((x_i)_j)_{j \in \text{support}(u)}$ and of y_i . Then it is easy to confirm that for any subset $S \subseteq \Phi_u$, the marginal distribution $P_u|_{\phi_u}(S)$ does not depend on u but only on S ; the choice of $\text{support}(u) \setminus S$ does not alter this distribution.

It is known that the information theory sample size threshold of the problem is

$$m_{\text{STATS}} = \tilde{\Theta} \left(\frac{k}{\log(\frac{\text{SNR}^2}{2\text{SNR}+1} + 1)} \right),$$

see e.g., [19]. Also in [5] it was proven that in the similar mSLR setting where u 's coordinates can take values in $\{-1, 0, 1\}$, if

$$m \leq m_{\text{ALG}} = \tilde{\Theta} \left(\frac{(\text{SNR} + 1)^2}{\text{SNR}^2} k^2 \right),$$

then the problem is $O(\log n)$ -degree hard. Here we prove that with sample size $m \leq (m_{\text{ALG}})^{1-o(1)}$ the problem is also GFP-hard, and hence via Theorem 3 also SQ-hard. Our result holds under a very mild assumption on SNR being not exponential in k . Interestingly, the proof is relatively short.

The first step is to calculate the inner product $\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}}$ which accounts to a calculation over the Gaussian measure.

Lemma 5. *For any sample size m and any u, v binary k -sparse vector, the following holds for the mSLR model:*

$$\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} = \left(1 - \left(\frac{\langle u, v \rangle}{k + \sigma^2} \right)^2 \right)^{-m} \leq \exp \left(\frac{m \langle u, v \rangle^2}{(k + \sigma^2)^2 - \langle u, v \rangle^2} \right).$$

Using Lemma 5 one can prove the GFP-hardness, and therefore the SQ-hardness.

Theorem 7. *If $n^{\Omega(1)} = k = o(n^{1/2})$ then for any $m = o \left(\frac{k}{\log(\frac{\text{SNR}^2}{2\text{SNR}+1} + 1)} \right)$, it holds*

$$\chi^2(\mathbb{P}^{\otimes m}, \mathbb{Q}^{\otimes m}) = 1 + o(1).$$

Moreover, for any constant $T > 1$, for any $m = O \left(\frac{(\text{SNR}+1)^2 k^2}{\text{SNR}^2 (\log n)^{2T+2}} \right)$ and $q = e^{\Theta((\log n)^T)}$, the mSLR task is $(q, m, O(1))$ -GFP hard. In particular, if $\text{SNR} \leq e^{k^{1-\alpha}}$ for some $\alpha > 0$, then for any $T > 1$ the mSLR task is $(e^{\Theta((\log n)^T)}, (\frac{(\text{SNR}+1)^2}{\text{SNR}^2} k^2)^{1-o(1)})$ -SQ hard.

The proof of this Theorem can be found in Appendix D.

D.1.1 Proofs for mSLR

Proof of Lemma 5. Let $\lambda = \sqrt{k/\sigma^2 + 1}$ and since $\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} = (\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}})^m$ we focus on the case $m = 1$.

Let $Y = Y_1, X = X_1$. By definition and Bayes' rule,

$$L_u = L_u(X, Y) = \frac{\mathbb{P}(Y|X, u)}{\mathbb{Q}(Y)}$$

Under $\mathbb{Q}$ we have $\lambda\sigma Y \sim \mathcal{N}(0, \lambda^2\sigma^2)$, while under $\mathbb{P}$ conditional on (X, u) we have

$$\lambda\sigma Y = \sqrt{k + \sigma^2}Y \sim \frac{1}{2}\mathcal{N}(\langle X, u \rangle, \sigma^2) + \frac{1}{2}\mathcal{N}(-\langle X, u \rangle, \sigma^2),$$

and so

$$\begin{aligned} L_u &= \frac{\mathbb{P}(Y|X, u)}{\mathbb{Q}(Y)} \\ &= \frac{1}{2}\lambda \exp\left(-\frac{1}{2\sigma^2}(\lambda\sigma Y - \langle X, u \rangle)^2 + \frac{1}{2\lambda^2\sigma^2}(\lambda\sigma Y)^2\right) \\ &\quad + \frac{1}{2}\lambda \exp\left(-\frac{1}{2\sigma^2}(\lambda\sigma Y + \langle X, u \rangle)^2 + \frac{1}{2\lambda^2\sigma^2}(\lambda\sigma Y)^2\right) \\ &= \frac{\lambda^m}{2} \left\{ \exp\left(-\frac{\lambda^2 - 1}{2}Y^2 + \frac{\lambda}{\sigma}Y\langle X, u \rangle - \frac{1}{2\sigma^2}\langle X, u \rangle^2\right) \right. \\ &\quad \left. + \exp\left(-\frac{\lambda^2 - 1}{2}Y^2 - \frac{\lambda}{\sigma}Y\langle X, u \rangle - \frac{1}{2\sigma^2}\langle X, u \rangle^2\right) \right\}. \end{aligned}$$

Now a standard integration argument using the MGF of the χ^2 distribution (see e.g., the proof of [6, Proposition 6.8.] for an almost identical argument) gives for any u, v binary k -sparse vectors,

$$\langle L_u, L_v \rangle_{\mathbb{Q}} = \frac{\lambda^2}{2(2\lambda^2 - 1)^{1/2}} \mathbb{E}_{X \sim \mathbb{Q}} \left(\exp\left(\frac{1}{2\sigma^2(2\lambda^2 - 1)} [(1 - \lambda^2)(\langle X, u \rangle^2 + \langle X, v \rangle^2) + 2\lambda^2\langle X, u \rangle\langle X, v \rangle]\right) \right) \quad (36)$$

$$+ \exp\left(\frac{1}{2\sigma^2(2\lambda^2 - 1)} [(1 - \lambda^2)(\langle X, u \rangle^2 + \langle X, v \rangle^2) - 2\lambda^2\langle X, u \rangle\langle X, v \rangle]\right) \quad (37)$$

Now of course the pair $(\langle X, u \rangle, \langle X, v \rangle)$ follows a bivariate Gaussian law with variances equals to k and covariance $\langle u, v \rangle$. Hence, some standard manipulations (see again the proof of [6, Proposition 6.8.] for an almost identical argument) allow us to derive that for $Z \in \mathbb{R}^{1 \times 3}$ with i.i.d. $\mathcal{N}(0, 1)$ entries and

$$t := \frac{1}{2\sigma^2(2\lambda^2 - 1)} = \frac{1}{2\sigma^2(2k/\sigma^2 + 1)} = \frac{1}{4k + 2\sigma^2}. \quad (38)$$

it holds

$$\langle L_u, L_v \rangle_{\mathbb{Q}} = \frac{\lambda^{2m}}{2(2\lambda^2 - 1)^{m/2}} \mathbb{E}_Z (\exp(t\langle M_1, Z^\top Z \rangle) + \exp(t\langle M_2, Z^\top Z \rangle)) \quad (39)$$

where for $\ell := \langle u, v \rangle$,

$$M_1 = M_1(\ell) := \begin{pmatrix} 2\ell & \sqrt{\ell(k - \ell)} & \sqrt{\ell(k - \ell)} \\ \sqrt{\ell(k - \ell)} & (1 - \lambda^2)(k - \ell) & \lambda^2(k - \ell) \\ \sqrt{\ell(k - \ell)} & \lambda^2(k - \ell) & (1 - \lambda^2)(k - \ell) \end{pmatrix}.$$

and

$$M_2 = M_2(\ell) := \begin{pmatrix} 2(1 - 2\lambda^2)\ell & (1 - 2\lambda^2)\sqrt{\ell(k - \ell)} & (1 - 2\lambda^2)\sqrt{\ell(k - \ell)} \\ (1 - 2\lambda^2)\sqrt{\ell(k - \ell)} & (1 - \lambda^2)(k - \ell) & -\lambda^2(k - \ell) \\ (1 - 2\lambda^2)\sqrt{\ell(k - \ell)} & -\lambda^2(k - \ell) & (1 - \lambda^2)(k - \ell) \end{pmatrix}.$$

The eigendecompositions of M_1, M_2 of the form $\sum_{i=1}^3 \lambda_i \frac{u_i u_i^\top}{\|u_i\|^2}$ are, first for M_1 ,

$$\begin{aligned} u_1^\top &= (0 \quad 1 \quad -1) & \lambda_1 &= (1 - 2\lambda^2)(k - \ell) \\ u_2^\top &= (\sqrt{k - \ell} \quad -\sqrt{\ell} \quad -\sqrt{\ell}) & \lambda_2 &= 0 \\ u_3^\top &= (2\sqrt{\ell} \quad \sqrt{k - \ell} \quad \sqrt{k - \ell}) & \lambda_3 &= k + \ell. \end{aligned} \quad (40)$$

and for M_2 ,

$$\begin{aligned} u_1^\top &= (0 \quad 1 \quad -1) & \lambda_1 &= k - \ell \\ u_2^\top &= (\sqrt{k - \ell} \quad -\sqrt{\ell} \quad -\sqrt{\ell}) & \lambda_2 &= 0 \\ u_3^\top &= (2\sqrt{\ell} \quad \sqrt{k - \ell} \quad \sqrt{k - \ell}) & \lambda_3 &= (1 - 2\lambda^2)(k + \ell). \end{aligned} \quad (41)$$

As $t < 1/(4k)$ we have $2t \max\{\|M_1\|_{\text{op}}, \|M_2\|_{\text{op}}\} < (k + \ell)/2k \leq 1$. Hence, using [6, Lemma A.5.] for $B(U) = \mathbb{R}^{n \times m}$, we have

$$\langle L_u, L_v \rangle_{\mathbb{Q}} = \frac{\lambda^2}{2(2\lambda^2 - 1)^{1/2}} (\det(I_3 - 2tM_1)^{-1/2} + \det(I_3 - 2tM_2)^{-1/2}). \quad (42)$$

Using (38) and (40), (41) the eigenvalues of the matrices $I_3 - 2tM_1, I_3 - 2tM_2$ are

$$\{1, 1 - 2t(k + \ell), 1 - 2t(1 - 2\lambda^2)(k - \ell)\} = \left\{1, 1 - \frac{k + \ell}{\sigma^2(2\lambda^2 - 1)}, 1 + \frac{k - \ell}{\sigma^2}\right\}$$

and

$$\{1, 1 - 2t(k - \ell), 1 - 2t(1 - 2\lambda^2)(k + \ell)\} = \left\{1, 1 - \frac{k - \ell}{\sigma^2(2\lambda^2 - 1)}, 1 + \frac{k + \ell}{\sigma^2}\right\}.$$

Since $\lambda^2 = k/\sigma^2 + 1$ we have

$$\frac{\lambda^2}{\sqrt{2\lambda^2 - 1}} \det(I_3 - 2tM_1)^{-1/2} = \lambda^2 \left[\left(2\lambda^2 - 1 - \frac{k + \ell}{\sigma^2}\right) \left(1 + \frac{k - \ell}{\sigma^2}\right) \right]^{-1/2} \quad (43)$$

$$= \frac{\frac{k}{\sigma^2} + 1}{1 + \frac{k - \ell}{\sigma^2}} \quad (44)$$

$$= \left(1 - \frac{\ell}{k + \sigma^2}\right)^{-1}. \quad (45)$$

and by symmetry,

$$\frac{\lambda^2}{\sqrt{2\lambda^2 - 1}} \det(I_3 - 2tM_2)^{-1/2} = \left(1 + \frac{\ell}{k + \sigma^2}\right)^{-1}. \quad (46)$$

Combining the above,

$$\langle L_u, L_v \rangle_{\mathbb{Q}} = \frac{1}{2} \left(\left(1 - \frac{\ell}{k + \sigma^2}\right)^{-1} + \left(1 + \frac{\ell}{k + \sigma^2}\right)^{-1} \right) \quad (47)$$

$$= \left(1 - \left(\frac{\ell}{k + \sigma^2}\right)^2\right)^{-1} \quad (48)$$

$$\leq \exp\left(\frac{m\ell^2}{(k + \sigma^2)^2 - \ell^2}\right), \quad (49)$$

where for the last inequality we used that $\log x \geq 1 - 1/x$, for $x > 0$. $\square$

Proof of Theorem 7. We have from the first part of Lemma 5,

$$\chi^2(\mathbb{P}^{\otimes m}, \mathbb{Q}^{\otimes m}) - 1 = \mathbb{E}_{u, v \sim \pi} \langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} \leq \mathbb{E}_{u, v \sim \pi} \left(1 - \frac{\langle u, v \rangle^2}{(k + \sigma^2)^2}\right)^{-m} \quad (50)$$

But, in this setting $\langle u, v \rangle$ follows an Hypergeometric distribution with parameters n, k, k . Hence, by [6, Lemma 6.6],

$$\chi^2(\mathbb{P}^{\otimes m}, \mathbb{Q}^{\otimes m}) - 1 \leq \sum_{\ell=0}^k \left(1 - \frac{\ell^2}{(k + \sigma^2)^2}\right)^{-m} \left(\frac{k^2}{n - k}\right)^\ell \quad (51)$$

$$\leq \sum_{\ell=0}^{\lfloor k/2 \rfloor} \exp\left(\frac{m\ell^2}{(k + \sigma^2)^2 - \ell^2}\right) e^{-\ell \log(\frac{n-k}{k^2})} + \sum_{\ell=\lfloor k/2 \rfloor}^k \left(1 - \frac{k^2}{(k + \sigma^2)^2}\right)^{-m} e^{-\ell \log(\frac{n-k}{k^2})} \quad (52)$$

$$\leq \sum_{\ell=0}^{\lfloor k/2 \rfloor} e^{\ell m / (k - \ell) - \ell \log(\frac{n-k}{k^2})} + k \left(\frac{(\frac{k}{\sigma^2})^2}{2\frac{k}{\sigma^2} + 1} + 1\right)^m e^{-\Theta(k \log(\frac{n-k}{k^2}))}, \quad (53)$$

where for the last inequality we used $\log x \geq 1 - 1/x$ for $x > 0$.

Since $k^2 = o(n)$, $m = o(\frac{k}{\log(\frac{\text{SNR}^2}{2\text{SNR}+1} + 1)})$, $\text{SNR} = k/\sigma^2$ we have for large enough n ,

$$k \left(\frac{(\frac{k}{\sigma^2})^2}{2\frac{k}{\sigma^2} + 1} + 1\right)^m e^{-\Theta(k \log(\frac{n-k}{k^2}))} = e^{-\Theta(k \log(\frac{n-k}{k^2}))} = o(1).$$

Moreover, since $k^2 = o(n)$, $m = o(k)$ we have for large enough n ,

$$\sum_{\ell=0}^{\lfloor k/2 \rfloor} e^{\ell m / (k - \ell) - \ell \log(\frac{n-k}{k^2})} \leq \sum_{\ell=0}^{\lfloor k/2 \rfloor} e^{2\ell m / k - \ell \log(\frac{n-k}{k^2})} \leq \sum_{\ell=0}^{\lfloor k/2 \rfloor} e^{-\ell \log(\frac{n-k}{k^2})/2} = 1 + o(1).$$

Now, fix any $T > 1$. Notice $\langle L_u, L_v \rangle_{\mathbb{Q}}$ is an increasing function of $\langle u, v \rangle$. Hence, for any $q > 0$ there exists $\delta_0(q) > 0$ such that $\{\rho_{\text{id}}(u, v) \geq r(q)\} = \{\langle u, v \rangle \geq \delta_0(q)\}$. Moreover, from the tail of $\langle u, v \rangle$ which is an (n, k, k) Hypergeometric distribution, there exists some $q = q_T = e^{\Theta((\log n)^T)}$ for which there exists $r(T)$ with $\pi(\{\rho_{\text{id}}(u, v) \geq r(T)\}) = q^{-2}$.

Let us then fix $q = q_T$. Notice that if we choose $\delta := \log(kq^2) = \Theta((\log n)^{T+1})$, then we have for large enough n , by [6, Lemma 6.6] $\pi^{\otimes 2}(\langle u, v \rangle \geq \delta) \leq k(\frac{k^2}{n})^\delta \leq k2^{-\delta} = 1/q^2$. Hence, $\delta_0 \leq \delta$. Combining the above with the second part of Lemma 5,

$$\begin{aligned} \mathbb{E}_{u, v \sim \pi}(\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} \mathbf{1}(\langle u, v \rangle \leq \delta_0)) &\leq \mathbb{E}_{u, v \sim \pi} \exp\left(\frac{m \langle u, v \rangle^2}{(k + \sigma^2)^2 - \langle u, v \rangle^2}\right) \mathbf{1}(\langle u, v \rangle \leq \delta) \\ &\leq \exp\left(\frac{m \delta^2}{(k + \sigma^2)^2 - \delta^2}\right) \\ &= \exp\left(m \Theta\left(\frac{(\log n)^{2(T+1)}}{k^2(1 + \text{SNR}^{-1})^2}\right)\right) = O(1), \end{aligned}$$

for any $m = O(\frac{k^2(1 + \text{SNR}^{-1})^2}{(\log n)^{2(T+1)}})$. Hence, the model is $(q = q_T, m = \Theta(\frac{k^2(1 + \text{SNR}^{-1})^2}{(\log n)^{2(T+1)}}), O(1))$ - ρ_{id} -hard. Using now Theorem 3 for $m_{\text{IT}} = (\log n)^T$, which is permissible to use using our χ^2 bound and that $\text{SNR} \leq e^{k^{1-\alpha}}$ for some $\alpha > 0$, we conclude for all $T > 1$, the $(e^{\Theta((\log n)^{T-1})}, (k^2(1 + \text{SNR}^{-1})^2)^{1-o(1)})$ -SQ hardness. The result follows. $\square$

D.2 Non-Gaussian component analysis

The following model was introduced in [16] to capture the complexity of learning Gaussian mixtures.

Definition 10 (Non-Gaussian component analysis model). A “ $\mathbb{P}$ versus $\mathbb{Q}$ ” detection problem is a Non-Gaussian component (NGCA) model if:

- There exists $\mu \in \mathcal{P}(\mathbb{R})$ such that, under the planted hypothesis $\mathbb{P}_u$ with $u \in \mathcal{S}^{n-1}$, we sample $x \sim \mathcal{N}(0, I_n)$ and replace the component $\langle x, u \rangle \cdot u$ by $z \cdot u$ where $z \sim \mu$ independently;

- Under the null model, we sample $x \sim \mathcal{N}(0, I_n)$.

In other words, an NGCA model is an isotropic Gaussian distribution with a non-Gaussian marginal in direction u . The SQ-hardness for NGCA models established also in [16] has been of big importance in proving many recent SQ-hardness for learning tasks, such as for robust estimation of Gaussians [17] and robust linear regression [17] among others.

Interestingly, we can also connect all NGCA models with GFP-hardness for $G = \mathbb{Z}_2$. By a direct Hermite expansion, we can decompose the likelihood function (in $L^2(\mathbb{Q})$)

$$L_u(x) = 1 + \sum_{i=s^*}^{\infty} \nu_i h_i(\langle u, x \rangle), \quad \nu_i := \mathbb{E}_{z \sim \mu}[h_i(z)],$$

where h_i is the (normalized) degree- i Hermite polynomial. Here we denoted $s^* > 0$ the first non-zero coefficient $\nu_i \neq 0$, that is, the smallest moment of μ that disagrees with $N(0, 1)$ moments (we call s^* the generative exponent of the NGCA model). The inner-product of the likelihood ratios is given for all $u, v \in \mathcal{S}^{n-1}$ by

$$\langle L_u, L_v \rangle = 1 + \sum_{i=s^*}^{\infty} \nu_i^2 \cdot (\langle u, v \rangle)^i, \quad (54)$$

where we used that $\mathbb{E}[h_s(\langle u, x \rangle) h_k(\langle v, x \rangle)] = \delta_{ks} \langle u, v \rangle^s$. Similar to GAMs, using again the group $G = \mathbb{Z}_2$ acting on flipping the sign of each parameter, we get

$$\mathbb{E}_{g, g' \sim \text{Unif}(G)} (\langle L_{g(u)}, L_{g'(v)} \rangle_{\mathbb{Q}} - 1) = \sum_{i=s^*, i \text{ even}}^{\infty} \nu_i^2 \cdot (\langle u, v \rangle)^i \geq 0,$$

concluding that NGCA satisfy Assumption 1 with $G = \mathbb{Z}_2$. Hence, based on our equivalence, for any NGCA model the SQ-hardness is equivalent with the GFP-hardness for any symmetric prior (that is $\pi(-u) = \pi(u)$). We illustrate this equivalence for two standard priors: the uniform prior $\pi = \text{Unif}(\mathcal{S}^{n-1})$ and the k -sparse prior $\pi = \text{Unif}(\{u \in \pm \frac{1}{\sqrt{k}} \{0, 1\}^n : \|u\|_0 = k\})$.

Theorem 8 (GFP-hardness of NGCA, uniform prior). *Consider a NGCA model with generative exponent s^* and the uniform prior $\pi = \text{Unif}(\mathcal{S}^{n-1})$. For any $\varepsilon \in (0, 1/2)$, the NGCA model is $(\exp(\Theta(n^\varepsilon)), m, O(1))$ -GFP hard with*

$$m = \frac{1}{\nu_{s^*}^2} n^{s^*/2 - \Theta(\varepsilon)}.$$

Moreover, via our equivalence theorem, the model is $(\exp(n^{\Theta(\varepsilon)}), m^{1-\Theta(\varepsilon)})$ -SQ hard.

Theorem 9 (GFP-hardness of NGCA, sparse prior). *Consider a NGCA model with generative exponent s^* and the k -sparse prior $\pi = \text{Unif}(\{u \in \pm \frac{1}{\sqrt{k}} \{0, 1\}^n : \|u\|_0 = k\})$. For any $\varepsilon \in (0, 1/2)$ so that $k = n^{\Omega(\varepsilon)}$, the NGCA model is $(\exp(\Theta(n^\varepsilon)), m, O(1))$ -GFP hard with*

$$m = \frac{1}{\nu_{s^*}^2} \min(n^{s^*/2 - \Theta(\varepsilon)}, k^{s^*} n^{-\Theta(\varepsilon)}).$$

Moreover, via our equivalence theorem, the model is $(\exp(n^{\Theta(\varepsilon)}), m^{1-\Theta(\varepsilon)})$ -SQ hard.

The SQ lower bound in Theorem 8 was proven in [16] by a direct argument: here, we prove this SQ-hardness via equivalence to GFP-hardness. The sparse prior was not considered previously and we include it to illustrate the broad applicability of our equivalence.

D.2.1 Proofs for Non-Gaussian Component Analysis

Proof of Theorem 8. Let us prove that the model is $\rho_{\mathbb{Z}_2}$ -FP hard and conclude using the implication in Theorem 2.1. Note that $\rho_{\mathbb{Z}_2}(u, v) = \langle L_u, L_{\text{sign}(\langle u, v \rangle)v} \rangle - 1$, that is

$$\rho_{\mathbb{Z}_2}(u, v) = \sum_{s=s^*}^{\infty} \nu_s^2 |\langle u, v \rangle|^s,$$

and $\rho_{\mathbb{Z}_2}(u, v)$ is an increasing function of $|\langle u, v \rangle|$. Thus, $\rho_{\mathbb{Z}_2}$ -FP hardness is equivalent to FP hardness. Using that $\langle u, v \rangle$ under the uniform prior is distributed as the first coordinate of $z \sim \text{Unif}(S^{d-1})$, we get

$$\pi(|\langle u, v \rangle| \geq \kappa) \leq 2\exp(-cn\kappa^2),$$

for some universal constant $c > 0$. For simplicity denote $\rho = |\langle u, v \rangle|$. Using that $\nu_s^2 \leq 1$ for any $s \in \mathbb{N}$ by Jensen's inequality, we can write

$$\langle L_u, L_v \rangle - 1 \leq \sum_{s \geq s^*} \nu_s^2 \rho^s \leq \nu_{s^*}^2 \rho^{s^*} + \rho^{s^*+1} \sum_{s \geq s^*+1} \rho^s \leq \rho^{s^*} \left(\nu_{s^*}^2 + \frac{\rho}{1-\rho} \right),$$

so that for $\rho = o_n(1)$ and n large enough, $\langle L_u, L_v \rangle - 1 \leq 2\nu_{s^*}^2 \rho^{s^*}$. We deduce that for $\kappa = n^{-1/2+\varepsilon}$, we have

$$\begin{aligned} \mathbb{E}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} \cdot \mathbf{1}(|\langle u, v \rangle| < \kappa)] &\leq 1 + \sum_{j=1}^m \binom{m}{j} \mathbb{E}[(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1)^j \cdot \mathbf{1}(|\langle u, v \rangle| < \kappa)] \\ &\leq 1 + \sum_{j=1}^m (2m\nu_{s^*}^2 \kappa^{s^*})^j. \end{aligned}$$

Thus we deduce that the problem is $(\exp(\Theta(n^\varepsilon)), m, \Theta(1))$ -GFP hard with $m = n^{s^*/2-\Theta(\varepsilon)}/\nu_{s^*}^2$.

To use the equivalence with SQ, we need to compute the χ^2 -divergence, that is

$$\mathbb{E}[\langle L_u, L_v \rangle^{4t}] = \mathbb{E}[(\langle L_u, L_v \rangle - 1)^{4t}] = 1 + \sum_{j=1}^{4t} \binom{4t}{j} \mathbb{E}[(\langle L_u, L_v \rangle - 1)^j].$$

Let us bound

$$\begin{aligned} \mathbb{E}[(\langle L_u, L_v \rangle - 1)^j] &= \mathbb{E}[(\langle L_u, L_v \rangle - 1)^j \cdot \mathbf{1}(|\rho| \leq \kappa)] + \mathbb{E}[(\langle L_u, L_v \rangle - 1)^j \cdot \mathbf{1}(|\rho| > \kappa)] \\ &\leq (2\nu_{s^*}^2 \kappa^{s^*})^j + M^j \exp(-cn^{2\varepsilon}), \end{aligned}$$

where we denoted $M = \|L_u\|_{\mathbb{Q}}^2 - 1 = O(\exp(n^{\varepsilon/2}))$. Thus

$$\mathbb{E}[(\langle L_u, L_v \rangle - 1)^j] = 1 + \sum_{j=1}^{4t} (8t\nu_{s^*}^2 \kappa^{s^*})^j + (4tM)^j \exp(-cn^{2\varepsilon}) = O(1),$$

where we used that $t \log(t) = \tilde{\Theta}(n^{\varepsilon/2})$ by assumption. We can therefore apply Theorem 3 with $q' = \exp(n^{\varepsilon/2})$ and $t = n^{\varepsilon/2}$ (so that $t \leq \log(q)/\log(m) = \tilde{\Theta}(n^\varepsilon)$). The model is (q', m') -SQ hard with

$$m' = \frac{m}{(t(1+\varepsilon)^{1/t} + \chi^2(\mathbb{P}^{\otimes 4t} \parallel \mathbb{Q}^{\otimes 4t}))(q')^{2/t}} = \Theta(m/t) = m^{1-\Theta(\varepsilon)},$$

which concludes the proof. $\square$

Proof of Theorem 9. The proof proceeds similarly as the proof of Theorem 8. The main difference is the new tail bound on $\langle u, v \rangle$ given in Lemma 6. We now set $\kappa = n^\varepsilon \max(n^{-1/2}, k^{-1})$, so that

$$\pi(|\langle u, v \rangle| \geq \kappa) \leq 2\exp(-cn^\varepsilon).$$

With this modification, the rest of the proof is identical and we omit it. $\square$

Lemma 6 (Tail bound for sparse prior). *Let u, v be independently sampled from the prior $\pi = \text{Unif}(\{u \in \pm \frac{1}{\sqrt{k}}\{0, 1\}^n : \|u\|_0 = k\})$. Then for any $t \geq 0$, we have*

$$\pi^2(\langle u, v \rangle \geq t) \leq \exp(-c \min\{nt^2, kt\}), \quad (55)$$

for some universal constant $c > 0$.

D.3 Single-index Models

Another extremely popular class of models in statistics dating back to the 80s [33, 28] are the so-called single-index models.

Definition 11 (Single-index model). A “ $\mathbb{P}$ versus $\mathbb{Q}$ ” detection problem is a Single-index model if:

- There exists a distribution $\mu \in \mathcal{P}(\mathbb{R} \times \mathbb{R})$ such that, under the planted hypothesis $\mathbb{P}_u$, we sample $x \sim \mathcal{N}(0, I_n)$ and $y \sim \mu(\cdot | z_u)$, where $z_u := \langle x, u \rangle$;
- Under the null model, we sample $x \sim \mathcal{N}(0, I_n)$ and $y \sim \mu_y$, where μ_y is the marginal distribution of μ .

Also, all single index models satisfy Assumption 1 for $G = \mathbb{Z}_2$. Indeed, if s^* is the generative exponent of the model [12], following [12] we know that an Hermite expansion gives for some $s^* \in \mathbb{N}$ (s^* is called the generative exponent) that for all $u, v \in \mathcal{S}^{n-1}$,

$$\langle L_u, L_v \rangle_{\mathbb{Q}} = 1 + \sum_{i=s^*}^{\infty} \lambda_i^2 \cdot (\langle u, v \rangle)^i, \quad \lambda_i := \|\zeta_i(Y)\|_{\mu_y}, \quad \zeta_i(y) := \mathbb{E}[h_s(z)|y].$$

From this point on, the argument is identical as in the case of NGCA, including the nonnegativity with $G = \mathbb{Z}_2$ as well as the examples of GFP-hardness with uniform and sparse priors. For completeness, we state separate theorems for single-index models:

Theorem 10 (GFP-hardness of SI models, uniform prior). Consider a SI model with generative exponent s^* and the uniform prior $\pi = \text{Unif}(\mathcal{S}^{n-1})$. For any $\varepsilon \in (0, 1/2)$, the SI model is $(\exp(\Theta(n^\varepsilon)), m, O(1))$ -GFP hard with

$$m = \frac{1}{\lambda_{s^*}^2} n^{s^*/2 - \Theta(\varepsilon)}.$$

Moreover, via our equivalence theorem, the model is $(\exp(n^{\Theta(\varepsilon)}), m^{1-\Theta(\varepsilon)})$ -SQ hard.

Theorem 11 (GFP-hardness of SI models, sparse prior). Consider a SI model with generative exponent s^* and the k -sparse prior $\pi = \text{Unif}(\{u \in \pm \frac{1}{\sqrt{k}} \{0, 1\}^n : \|u\|_0 = k\})$. For any $\varepsilon \in (0, 1/2)$ so that $k = n^{\Omega(\varepsilon)}$, the SI model is $(\exp(\Theta(n^\varepsilon)), m, O(1))$ -GFP hard with

$$m = \frac{1}{\lambda_{s^*}^2} \min(n^{s^*/2 - \Theta(\varepsilon)}, k^{s^*} n^{-\Theta(\varepsilon)}).$$

Moreover, via our equivalence theorem, the model is $(\exp(n^{\Theta(\varepsilon)}), m^{1-\Theta(\varepsilon)})$ -SQ hard.

The SQ lower bounds in Theorem 10 and Theorem 11 were proven in [12] and [9] via direct argument. Here, we obtain these bounds via the equivalence of the SQ-hardness and GFP-hardness.

D.4 Truncated statistics: convex truncation

Learning from truncated data has been a topic of interest since the late 1800s and the pioneering works of Galton [23] and Pearson [35]. Interestingly, there has been some recent line of works on truncated statistics tasks that seeks to revisit these old questions from a computational viewpoint, see e.g., [13], [14] and references therein. In this line of recent work, the problem of detecting a convex truncation in Gaussian noise has been proposed.

Definition 12. Fix $\alpha \in (0, 1)$. A hypothesis testing “ $\mathbb{P}$ versus $\mathbb{Q}$ ” problem is called an α -Convex Truncation model if it satisfies:

1. Under the null hypothesis $\mathbb{Q}$, $x \sim N(0, I_n)$.
2. Under the planted hypothesis $\mathbb{P}_K$, $x \sim N(0, I_n) | K$ where K is a symmetric convex body with Gaussian volume at most $1 - \alpha$.

Interestingly, also all α -Convex Truncation models satisfy Assumption 1 for the trivial group G . Perhaps this fact is even more interesting because it turns out Assumption 1 is exactly equivalent with the celebrated Gaussian Correlation Inequality on convex bodies [37, 32].

Lemma 7. *Consider an α -convex truncated model in Definition 12. For any K, K' two symmetric convex bodies of Gaussian volume $1 - \alpha$, it holds $\langle L_K, L_{K'} \rangle_{\mathbb{Q}} \geq 1$. This is to say, and α -Convex Truncated Model satisfies Assumption 1 for the trivial group.*

Proof. For any K , it holds $L_K(x) = 1(x \in K)/\mathbb{Q}(K)$, $x \in \mathbb{R}^n$. Hence,

$$\langle L_K, L_{K'} \rangle = \frac{\mathbb{Q}(K \cap K')}{\mathbb{Q}(K)\mathbb{Q}(K')}.$$

But the so-called Gaussian correlation inequality for symmetric convex bodies in convex geometry [37] states exactly that for any symmetric convex bodies K, K' it holds $\mathbb{Q}(K \cap K') \geq \mathbb{Q}(K)\mathbb{Q}(K')$ yielding the result. $\square$

Now, for the α -Convex truncation models, the state-of-the-art polynomial-time algorithms require $O(n/\alpha^2)$ samples [15], and the best known information-theoretic lower bound is $\Omega(n/\alpha)$ samples [15]. Using the GFP-hardness to SQ-hardness framework we prove that for some prior on K , it is SQ-hard to distinguish with $\tilde{o}(n/\alpha^2)$ samples, providing evidence that the polynomial-time method from [15] cannot be improved.

D.4.1 A new SQ lower bound

To apply our framework, we focus on the following prior on K , a variant of which has been studied in [15] to prove their information-theoretic lower bound of $\Omega(n/\alpha)$ samples. To define it we let

$$K = K_v = \{x \in \mathbb{R}^d : |\langle x, v \rangle| \leq \kappa\},$$

for any $v \in \text{Unif}(\{-1/\sqrt{d}, 1/\sqrt{d}\}^d)$. Here, we choose $\kappa = \kappa(\alpha, d)$ is such that the Gaussian measure of each K_v is $1 - \alpha$. Then our prior is uniform among K_v , $v \sim \text{Unif}(\{-1/\sqrt{d}, 1/\sqrt{d}\}^d)$. We refer to the α -convex truncation setting with this prior as the “ α -Slice Convex Truncation” model.

We first point out that for any $m = \omega(n/\alpha)$, detection with m samples is always possible in the α -Slice Convex Truncation model from a time-inefficient method. Indeed, one can brute-force search for some $v \in \{-1/\sqrt{d}, 1/\sqrt{d}\}^d$ for which it holds: for all $i = 1, 2, \dots, m$, $|\langle x_i, v \rangle| \leq \kappa$. Under $\mathbb{P}$, there always exists such a vector v and hence the brute force search algorithm will find it with probability 1. Under $\mathbb{Q}$ though a direct union bound gives that such a v exists only with probability at most $2^d(1 - \alpha)^m = o(1)$ for any $m = \omega(d/\alpha)$. Hence, the algorithm can detect with probability $1 - o(1)$. In that context, we prove the following result.

Theorem 12 (ρ_{Id} -FP- and SQ-hardness of Convex Truncation). *Let $n \in \mathbb{N}$ growing and arbitrary $\alpha = \alpha_n \in (0, 1)$. There exists a universal constant $C > 0$ and a prior π on the convex bodies K of Gaussian volume $1 - \alpha$ such that for any $q \in \mathbb{N}$ with $q = e^{o(\alpha n)}$, the α -Convex Truncation model under π is $(q, \frac{Cn}{\alpha^2 \log(1/\alpha)^{3/2} \log q})$ - ρ_{Id} -FP-hard.*

In particular, for any constant $T > 0$ if $\alpha = \omega(\frac{(\log n)^T}{n})$ then the α -Convex Truncation model under π is $(e^{\Theta((\log n)^T)}, \Theta(\frac{n}{\alpha^2 \log(1/\alpha)^{3/2} (\log n)^{2T+1}}))$ -SQ hard.

Satisfyingly the proof of this result is also relatively short. The proof of this Theorem can be found in Appendix D.

D.4.2 Proofs for Convex Truncation

Proof of Theorem 12. Observe that for $L_u := L_{K_u}$ we have via standard Hermite expansion (identical to the argument in [15, Line (32), proof of Claim 24]),

$$\langle L_u, L_v \rangle_{\mathbb{Q}} = \frac{\mathbb{Q}(K_u \cap K_v)}{(1 - \alpha)^2} = 1 + (1 - \alpha)^{-2} \langle u, v \rangle^2 \left[\sum_{i=1}^{\infty} f_{2i}^2 \langle u, v \rangle^{2(i-1)} \right], \quad (56)$$

where f_i is the i -th Hermite weight of $\mathbf{1}(x \in [-\kappa, \kappa])$, $x \in \mathbb{R}$ for κ such that $\Phi(\kappa) = 1 - \alpha/2$ where Φ is the CDF of a standard Gaussian.

Now, conveniently, the authors [27] have already studied the Hermite mass of indicators of symmetric intervals around 0. Indeed, applying [27, Lemma 27] for $j = 2$, $\theta = k$ imply that

$$f_2^2 = O(\kappa \phi(\kappa)^2),$$

where ϕ is the PDF of a standard Gaussian. But observe that by standard tail bounds $\kappa = O(\sqrt{\log(1/\alpha)})$ and from the Mill's ratio bound $\phi(\kappa) = \Theta((1 - \Phi(\kappa))\kappa)$. Combining the above we conclude

$$f_2^2 = O\left(\alpha^2 \log(1/\alpha)^{3/2}\right).$$

Parseval's identity gives $\sum_{i>0} f_i^2 = \alpha(1 - \alpha) \leq \alpha$, and hence for some constant $C > 0$

$$\langle L_u, L_v \rangle_{\mathbb{Q}} \leq 1 + C \left((1 - \alpha)^{-2} \langle u, v \rangle^2 \left(\alpha^2 \log(1/\alpha)^{3/2} + \langle u, v \rangle^2 \alpha \right) \right).$$

Now, notice that from (56), $\langle L_u, L_v \rangle_{\mathbb{Q}}$ is an increasing function of $\langle u, v \rangle^2$. Hence, for any $q > 0$ there exists $\delta_0(q) > 0$ such that $\{\rho_{\text{id}}(u, v) \geq r(q)\} = \{\langle u, v \rangle^2 \geq \delta_0(q)\}$. From Hoeffding's inequality we have that for some constant $C' > 0$ if $\delta = C' \frac{\log q}{n}$ then $\pi^2(\langle u, v \rangle^2 \geq \delta) \leq q^{-2}$. Hence $\delta_0(q) \leq \delta = C' \frac{\log q}{n}$.

Combining the above we have that for any $q = e^{o(\alpha n)}$,

$$\begin{aligned} \mathbb{E}[\langle L_u, L_v \rangle_{\mathbb{Q}}^m \mathbf{1}(\langle u, v \rangle^2 \leq \delta_0)] &\leq \left[1 + C \left((1 - \alpha)^{-2} \delta_0 (\alpha^2 \log(1/\alpha)^{3/2} + \delta_0 \alpha) \right) \right]^m \\ &\leq \left[1 + C \left((1 - \alpha)^{-2} C' \frac{\log q}{n} (\alpha^2 \log(1/\alpha)^{3/2} + C' \frac{\log q}{n} \alpha) \right) \right]^m \\ &\leq \left[1 + 2C \left(C' \frac{\log q}{n(1 - \alpha)^2} (\alpha^2 \log(1/\alpha)^{3/2}) \right) \right]^m \\ &= O(1), \end{aligned}$$

as long as $m = O(d/(\alpha^2 \log(1/\alpha)^{3/2} \log q))$. So we conclude the $(q, \Theta(n/(\alpha^2 \log(1/\alpha)^{3/2} \log q)))$ - ρ_{id} -FP-hard for any $q = e^{o(\alpha n)}$.

Now via an identical proof to [15, Theorem 23] we have for any $m = o(n/\alpha)$ that

$$\chi^2(\mathbb{P}^{\otimes m}, \mathbb{Q}^{\otimes m}) = O(1).$$

In particular, for any constant $T > 0$,

$$\chi^2(\mathbb{P}^{\otimes (\log n)^T}, \mathbb{Q}^{\otimes (\log n)^T}) = O(1).$$

Finally, notice that again since $\langle L_u, L_v \rangle_{\mathbb{Q}}$ is a strictly increasing function of $\langle u, v \rangle^2$, and $\langle u, v \rangle$ is a sum of iid Rademacher random variables we conclude via standard Central Limit Theorem arguments that for any $T > 0$ there exists $q = q(T) = e^{\Theta((\log n)^{T+1})}$ for which for some $r' = r'(T)$, $r = r(T) > 0$ it holds that $\pi^2(\langle u, v \rangle^2 \geq r) = \pi^2(\rho_{\text{id}}(u, v) \geq r') = q^{-2}$.

Hence, for any $T > 0$ we can apply our equivalence Theorem 3 for $m_{\text{IT}} = (\log n)^T$, $q = q(T + 1)$ (so $\log q = \Theta((\log n)^{T+1})$), and appropriate $t = \Theta((\log n)^T)$ to conclude the $(e^{\Theta((\log n)^T)}, \Theta(\frac{n}{\alpha^2 \log(1/\alpha)^{3/2} (\log n)^{2T+1}}))$ -SQ hardness of the task. $\square$

E Details on the GFP-hardness and FP-hardness separation

Below, we provide details on the counterexample described in Section 5.

Proof of Lemma 3. By definition, for any $u \in \{0, 1\}^{n+1}$,

$$\begin{aligned} L_u(x) &= \prod_{i=0}^n \left(\mathbf{1}(x_i = 1) \cdot \frac{1/2 + r \cdot \frac{1-(1-\alpha) \cdot u_i}{2}}{1/2} + \mathbf{1}(x_i = -1) \cdot \frac{1/2 - r \cdot \frac{1-(1-\alpha) \cdot u_i}{2}}{1/2} \right) \\ &= \prod_{i=0}^n (1 + rx_i \cdot [1 - (1 - \alpha) \cdot u_i]). \end{aligned}$$

For any $u, v \in \{0, 1\}^{n+1}$, the inner product $\langle L_u, L_v \rangle$ satisfies

$$\begin{aligned} \langle L_u, L_v \rangle &= \mathbb{E}_{x \sim \mathbb{Q}} \left[\prod_{i=0}^n (1 + rx_i \cdot [1 - (1 - \alpha) \cdot u_i]) (1 + rx_i \cdot [1 - (1 - \alpha) \cdot v_i]) \right] \\ &= \prod_{i=0}^n \mathbb{E}_{x_i \sim \text{Rad}(1/2)} (1 + rx_i \cdot [1 - (1 - \alpha) \cdot u_i]) (1 + rx_i \cdot [1 - (1 - \alpha) \cdot v_i]) \\ &= \prod_{i=0}^n (1 + r^2 \cdot (1 - (1 - \alpha) \cdot u_i)(1 - (1 - \alpha) \cdot v_i)) \end{aligned}$$

Denote $a_i = 1 + r^2 \cdot (1 - (1 - \alpha) \cdot u_i)(1 - (1 - \alpha) \cdot v_i)$. When $u_i = v_i = 0$, we have $a_i = 1 + r^2$; when $u_i = v_i = 1$, $a_i = 1 + r^2 \cdot \alpha^2$; when there is exactly one 1 and one 0 in u_i, v_i , we get $a_i = 1 + r^2 \cdot \alpha$. We deduce that $a_i = 1 + r^2 \cdot \alpha^{u_i + v_i}$ and the lemma follows. $\square$

Let us consider the m -sample version of the hypothesis testing problem. The null hypothesis is then $\mathbb{Q}^{\otimes m}$ and the alternative hypothesis is $\mathbb{E}_{u \sim \pi} \mathbb{P}_u^{\otimes m}$, where u is sampled from the following two-point prior π :

$$u = \begin{cases} (1, 0, \dots, 0), & \text{w.p. } \rho, \\ (0, 1, \dots, 1), & \text{w.p. } 1 - \rho. \end{cases} \quad (57)$$

We abbreviate these vectors as $u_1 = (1, 0, \dots, 0)$ and $u_2 = (0, 1, \dots, 1)$ for convenience. Using Lemma 3, it holds that

$$\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle = \prod_{i=0}^n (1 + r^2 \cdot \alpha^{u_i + v_i})^m. \quad (58)$$

Let's next show that this problem is GFP hard but FP easy. Note that $\langle L_u, L_v \rangle \geq 1$ for all u, v and therefore the model verifies Assumption 1 for the trivial group.

Theorem 13. *For the two-point prior π in (58) with $\rho = \exp(-n^\varepsilon/2)$, and for $r = n^{-1/2}$, $\alpha = n^{-1+2\varepsilon}$, $m = n^{1-\varepsilon}$ and $D = n^\varepsilon$, where $\varepsilon > 0$ is any small constant, the following hold. The m -sample hypothesis testing problem $\mathbb{E}_{u \sim \pi} \mathbb{P}_u^{\otimes m}$ versus $\mathbb{Q}^{\otimes m}$ is $(e^{D/2}, m, \Theta(n^{-\varepsilon}))$ -GFP hard but not $(n^{-1}, m, \exp(\Theta(n^\varepsilon)))$ -FP hard. Moreover, via our equivalence theorem the model is $(e^{n^{\Theta(\varepsilon)}}, n^{1-\Theta(\varepsilon)})$ -SQ hard.*

Proof of Theorem 13. Let us first show it is FP easy. Define $\delta := \delta(n^{-1/2})$ the supremum over δ such that $\pi^2(\langle u, v \rangle \geq \delta) \geq 1/n$. We observe when $u \neq v$, then we must have $\langle u, v \rangle = 0 < \delta$ by the choice of the two points prior with $\langle u_1, u_2 \rangle = 0$. Therefore,

$$\pi^2(u \neq v) = 2\rho(1 - \rho) \leq 2e^{-n^\varepsilon/2} \ll n^{-1} \leq \pi^2(\langle u, v \rangle \geq \delta). \quad (59)$$

We deduce the following lower bound

$$\mathbb{E}_{u,v}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(\langle u, v \rangle < \delta)] \geq \mathbb{E}_{u,v}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(u \neq v)]$$

$$= \pi^2[u \neq v] \cdot \mathbb{E}_{u,v}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle | u \neq v]$$

When conditioned on $u \neq v$, we get

$$\mathbb{E}_{u,v}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle | u \neq v] = (1 + \alpha r^2)^{(n+1)m},$$

by applying Eq. (58), with $u_i + v_i = 1$ for all $0 \leq i \leq n$. Inserting the parameters stated in the lemma, we obtain

$$\begin{aligned} \mathbb{E}_{u,v}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(\langle u, v \rangle < \delta)] &\geq 2\rho(1 - \rho) \cdot (1 + \alpha r^2)^{(n+1)m} \\ &\geq \exp(-\tfrac{1}{2}n^\varepsilon) \cdot (1 + n^{-2+2\varepsilon})^{(n+1)n^{1-\varepsilon}} \\ &= \Omega(1) \cdot \exp(-\tfrac{1}{2}n^\varepsilon) \cdot \exp(n^\varepsilon) \geq \Omega(1) \cdot \exp(\tfrac{1}{2}n^\varepsilon). \end{aligned}$$

This shows that under our parameter choice, the task is $(n^{-1/2}, m, \exp(\Theta(n^\varepsilon)))$ -FP easy.

Let us now show that this model is GFP hard. We will prove that the model is ρ_{Id} -FP hard and conclude using the implication Theorem 2.1. Under the trivial group, we have $\rho_{\text{Id}}(u, v) = \langle L_u, L_v \rangle_{\mathbb{Q}} - 1$. From Eq. (58), the m -sample inner product of likelihood ratio is given for $u = v = u_1$ by

$$\langle L_{u_1}^{\otimes m}, L_{u_1}^{\otimes m} \rangle = (1 + \alpha^2 r^2)^m \cdot (1 + r^2)^{mn} \quad (60)$$

and for $u = v = u_2$ by

$$\langle L_{u_2}^{\otimes m}, L_{u_2}^{\otimes m} \rangle = (1 + \alpha^2 r^2)^{nm} \cdot (1 + r^2)^m. \quad (61)$$

Because $\alpha \ll 1$, it is not hard to notice

$$\langle L_{u_1}^{\otimes m}, L_{u_2}^{\otimes m} \rangle < \langle L_{u_2}^{\otimes m}, L_{u_2}^{\otimes m} \rangle < \langle L_{u_1}^{\otimes m}, L_{u_1}^{\otimes m} \rangle. \quad (62)$$

From the definition of π , it holds that

$$\pi^2(\{u = u_2, v = u_2\} \cup \{u \neq v\}) = 1 - e^{-n^\varepsilon} \geq 1 - q^{-2}, \quad (63)$$

using that we set $q = \exp(D/2)$. Combining with Eq. (62), we conclude that the event $\{\rho(u, v) \leq r(q)\} \subset \{u = u_2, v = u_2\} \cup \{u \neq v\}$. This allows us to estimate the upper bound as

$$\begin{aligned} \mathbb{E}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(r \leq r(q))] &\leq \mathbb{E}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(\{u = u_2, v = u_2\} \cup \{u \neq v\})] \\ &\leq (1 + \alpha^2 r^2)^{nm} \cdot (1 + r^2)^m. \end{aligned} \quad (64)$$

Inserting our choice of parameters, we obtain

$$\begin{aligned} \mathbb{E}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(r \leq r(q))] &\leq (1 + n^{-3+4\varepsilon})^{n^{2-\varepsilon}} \cdot (1 + n^{-1})^{n^{1-\varepsilon}} \\ &\leq \exp(n^{-1+3\varepsilon} + n^{-\varepsilon}) \leq 1 + 2n^{-\varepsilon}. \end{aligned} \quad (65)$$

Thus, the model is $(e^{D/2}, m, \Theta(n^{-\varepsilon}))$ - ρ_{Id} -FP hard, and therefore $(e^{D/2}, m, \Theta(n^{-\varepsilon}))$ -GFP hard.

Finally, let us use the SQ-GFP equivalence in Theorem 3 to show that the model is also SQ hard, with parameters $q' = e^{n^{\varepsilon/2}}$ and $t = n^{\varepsilon/2}$ (where indeed $t \leq \log(q)/\log(m) = \tilde{\Theta}(n^\varepsilon)$). To apply the theorem, we need to compute the χ^2 -divergence. Denoting $X = \langle L_u^{\otimes 4t}, L_v^{\otimes 4t} \rangle$ with $t = n^{\varepsilon/2}$,

$$\begin{aligned} \chi^2(\mathbb{P}^{\otimes 4t}, \mathbb{Q}^{\otimes 4t}) + 1 &= \pi^2(u = u_1, v = u_1) \cdot \mathbb{E}[X | u = u_1, v = u_1] + \pi^2(u \neq v) \cdot \mathbb{E}[X | u \neq v] \\ &\quad + \pi^2(u = u_2, v = u_2) \cdot \mathbb{E}[X | u = u_2, v = u_2] \\ &= (1 - \rho)^2 \cdot (1 + \alpha^2 r^2)^{4nt} \cdot (1 + r^2)^{4t} + \rho^2 \cdot (1 + \alpha^2 r^2)^{4t} \cdot (1 + r^2)^{4nt} \\ &\quad + 2\rho(1 - \rho) \cdot (1 + \alpha^2 r^2)^{(n+1)4t} \\ &\leq (1 + n^{-3+4\varepsilon})^{n^{1+\varepsilon}} \cdot (1 + n^{-1})^{n^\varepsilon} \\ &\quad + e^{-n^\varepsilon} \cdot (1 + n^{-3+4\varepsilon})^{n^\varepsilon} \cdot (1 + n^{-1})^{n^{1+\varepsilon/2}} + 2n^{-\varepsilon} \cdot (1 + n^{-2+3\varepsilon})^{2n^{1+\varepsilon}} \\ &\leq 1 + 4n^{-1+\varepsilon}. \end{aligned}$$

Thus, we obtain

$$m' = \frac{m}{(t(1 + \varepsilon)^{1/t} + \chi^2(\mathbb{P}^{\otimes 4t} \parallel \mathbb{Q}^{\otimes 4t}))(q')^{2/t}} = m^{1-\Theta(\varepsilon)},$$

and we deduce the model is $(e^{D/2}, m^{1-\Theta(\varepsilon)})$ -SQ hard. $\square$