

821
822
823

824

825
826

827
828
829
830
831

832
833
834
835
836
837

838

839

840
841
842
843
844
845
846
847

848
849
850
851

852

**Supplementary Material for the paper
“An Optimized Franz-Parisi Criterion and its Equivalence with SQ Lower
Bounds”**

Contents

1	Introduction	1
1.1	Main Contributions	3
2	Main Results	5
2.1	The Assumption	5
2.2	The GFP-SQ equivalence	5
2.2.1	Simplifying GFP-hardness	5
2.2.2	The equivalence	6
3	Examples	7
3.1	Gaussian Additive Models	7
3.2	Planted Sparse Models	7
3.3	Non-Gaussian component analysis	7
3.4	Single-index Models	8
3.5	Truncated statistics	8
4	GFP-hardness is not always equal to FP-hardness	9
5	Conclusion	9
A	Additional background and discussion from the main text	23
A.1	Setting and Definitions	23
A.1.1	Low-Degree Lower Bounds	23
A.1.2	Low Samplewise Degree Lower Bounds	24
A.2	Connection of the FP criterion with statistical physics	24
A.3	Necessity of assumptions in main theorem	26
A.3.1	Necessary 1: a non-trivial information-theory threshold	26
A.3.2	Necessary: Assumption 1	27
B	Equivalence between LD, SQ, and GFP	28
B.1	Unconditional SQ hardness	28
B.2	Noise-robust models and SQ-LD equivalence	29
B.3	Equivalence of GFP, SQ, and LD hardness for noise-robust models	30
C	Details of examples and proofs	31

853	C.1 Gaussian Additive Models	31
854	C.2 Planted Sparse Models	31
855	C.2.1 Symmetric mixed sparse linear regression	31
856	C.2.2 Omitted proofs for the mSLR hardness	32
857	C.3 Non-Gaussian Component Analysis	35
858	C.4 Single-Index Models	37
859	C.5 Truncated Statistics	38
860	C.5.1 A new SQ lower bound for convex truncation	38
861	D Counterexample	40
862	E Proofs of equivalence	42
863	E.1 Proof of Theorem 2	42
864	E.2 Proof of Theorem 3	44
865	References	46

A Additional background and discussion from the main text

A.1 Setting and Definitions

We first recall the definition of a “ **\mathbb{P} versus \mathbb{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 task⁷ 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.

A.1.1 Low-Degree Lower Bounds

On top of GFP-hardness and SQ-hardness, defined in the main body, we also introduce here 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.

Definition 8 (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*

$$\text{LD}(D) := \|L_m^{\leq D}\|_{\mathbb{Q}}^2 = \left\| \left(\mathbb{E}_{u \sim \pi} L_u^{\otimes m} \right)^{\leq D} \right\|_{\mathbb{Q}}^2 = \mathbb{E}_{u, v \sim \pi} [\langle (L_u^{\otimes m})^{\leq D}, (L_v^{\otimes m})^{\leq D} \rangle_{\mathbb{Q}}]. \quad (10)$$

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].

⁷The associated *estimation* problem consists in recovering the planted signal u from $Y_1, \dots, Y_m \sim \mathbb{P}_u$.

• 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.

A.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 9. 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).

Notice that the notion of (d, k) -low degree hardness is then the natural generalization to (10). Also note 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 A.1 (Explaining Remark 2.2). 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 $\mathbb{E}_{u,v \sim \pi} [\langle L_u, L_v \rangle - 1] = \chi^2(\mathbb{P}, \mathbb{Q})$ 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 2.2 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

935

A.2 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

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

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

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

or equivalently for

$$F(t) = \mathbb{E}_{u \sim p, Y \sim \mathbb{P}_u} \left[\log \mathbb{E}_{v \sim \pi} \left(\prod_{i=1}^m \frac{\mathbb{P}_v(Y_i)}{\mathbb{Q}(Y_i)} 1(d(v, u) = t) \right) \right] , t \in [0, 1], \quad (11)$$

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,

$$\tilde{F}(t) = \mathbb{E}_{u, v \sim p, [\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle 1(d(v, u) = t)] , t \in [0, 1]. \quad (12)$$

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

$$\mathbf{FP:} \quad \mathbb{E} [\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(|\langle u, v \rangle| \leq \delta(q))] \leq 1 + \varepsilon, \quad \text{where} \quad (13)$$

$$\delta(q) = \sup\{\delta : \pi^2(\langle u, v \rangle \geq \delta) \geq q^{-2}\}, \quad (14)$$

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 4). 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.3 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.3.1 Necessary 1: a non-trivial information-theory threshold

First, we claim that some bound on m_{IT} is necessary for the connection between these notions of hardness; the problem should be information-theoretically impossible for some diverging number of samples. Indeed, consider the following multisample problem over graphs $n \in \mathbb{N}$ and $p = 1 - n^{-1/4}$, $k = n^{1/3+o(1)}$ the Planted Sparse Model where

- 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 a k -clique.
- Under \mathbb{Q} one sample is a sample from $G(n, p)$.

This is a variant of the classic planted clique model [29].

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

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

Moreover, the model satisfies Assumption 1. One can see this because the model is a PSM. Also, 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, the interesting fact about this PSM is that it is not SQ-hard even for $m = 1$ and $q = 1$, but it is GFP-hard even for $m = n^{\Theta(1)}$ -samples.

Not (1, 1)-SQ-hard Now for (1, 1)-SQ hardness is equivalent with $\mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|] \leq 1$. But

$$\begin{aligned} \mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|] &\geq |\langle L_u, L_u \rangle_{\mathbb{Q}} - 1| \pi^2(u = v) \\ &= (1 - n^{-1/4})^{\Theta(k^2)} \binom{n}{k} \\ &= \exp \left(\Theta(k^2 n^{-1/4}) - \Theta(k \log n) \right) \\ &= \exp \left(\Theta(n^{7/12}) \right) = \omega(1). \end{aligned}$$

($e^{n^{\Theta(1)}}$, $n^{1/8}$, $O(1)$)-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

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

1014 Hence, to prove GFP-hardness it suffices to prove that $\mathbb{E}[\langle L_u, L_v \rangle_{\mathbb{Q}}^m \mathbf{1}(|u \cap v| \leq \delta)] = O(1)$.
 1015 Therefore for $q = \exp(n^\alpha)$ for some $\alpha > 0$, and $m = n^{1/8+o(1)}$,

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

1016 for any $\alpha < 1/16$.

1017 A.3.2 Necessary: Assumption 1

1018 To connect the two notions of hardness, also some lower bound on $\min_{u,v} \langle L_u, L_v \rangle_{\mathbb{Q}}$ is necessary.
 1019 Indeed, one can easily find counterexamples when $\min_{u,v} \langle L_u, L_v \rangle_{\mathbb{Q}} = 0$.

1020 For instance, one can borrow the following model from [6, Section 4.2.] (assume for simplicity
 1021 $n \in 10\mathbb{N}$ in what follows):

- 1022 • Under \mathbb{P} , sample a $u \sim \text{Unif}(\{x \in \{-1, 1\}^n : \sum x_i = 8n/10\})$. Then under \mathbb{P}_u each
 1023 sample equals to u .
- 1024 • Under \mathbb{Q} for each sample we sample a $u \sim \text{Unif}(\{-1, 1\}^n)$ and output u .

1025 It is easy to see that for any $u, v \in \{-1, 1\}^n$, we have for all $u, v \in \{x \in \{-1, 1\}^n : \sum x_i =$
 1026 $8n/10\}$,

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

1027 **Not $(1, 1)$ -SQ hard** Clearly the model is not SQ-hard – similar to above $(1, 1)$ -SQ hard implies
 1028 $\mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|] \leq 1$, but

$$\begin{aligned} \mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|] &\geq |\langle L_u, L_u \rangle_{\mathbb{Q}} - 1| \pi^2(u = v) \\ &= (2^n - 1) \binom{n}{9n/10}^{-1} = \omega(1). \end{aligned}$$

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

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

1029 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.

B Equivalence between LD, SQ, and GFP

In this Appendix, we discuss the equivalence between GFP, SQ, and LD criteria. In Section B.1, we define an Unconditional Statistical Query (USQ) hardness criterion, which is equivalent to SQ and often appears as a convenient intermediate step in proofs. In Section B.2, we recall the result from Brennan et al. [8] that shows equivalence between SQ and LD hardness under noise robustness. Finally, we use this equivalence to state the equivalence between all three (GFP, SQ, and LD) hardness criteria in Section B.3.

B.1 Unconditional SQ hardness

The following hardness measure appeared, often implicitly, in several prior work (e.g., [8]):

Definition 11 (Unconditional SQ hardness). *We say a “ \mathbb{P} versus \mathbb{Q} ” detection problem is (m, t) -unconditional SQ hard for some even t if*

$$\text{USQ: } \mathbb{E} [\chi_{\mathbb{Q}}(\mathbb{P}_u, \mathbb{P}_v)^t] \leq m^{-t}. \quad (15)$$

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).

Proof of Proposition 1. This equivalence was proven in [8]. We sketch their proof below for completeness:

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 $q \geq 1$,

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

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

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

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

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

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

1059 B.2 Noise-robust models and SQ-LD equivalence

1060 This section is adapted from [8].

1061 An advantage of USQ is that it is directly related to low degree lower bounds. Recall the samplewise
1062 low degree lower bound introduced in Section A.1.2. Let's state the hardness criterion associated
1063 with this framework:

1064 **Definition 12** (Low Degree (LD) Hardness). *We say a “ \mathbb{P} versus \mathbb{Q} ” detection problem is (m, d, k, ε) -*
1065 *LD hard if*

$$1065 \quad \text{LD:} \quad \mathbb{E} [\langle (L_u^{\otimes m})^{\leq d, k}, (L_v^{\otimes m})^{\leq d, k} \rangle] \leq 1 + \varepsilon. \quad (16)$$

1066 USQ hardness is related to LD hardness with $d = \infty$, that is when we do not constrain the degree for
1067 each sample in the projection.

1068 **Proposition 2** (Equivalence between USQ and LD hardness with $d = \infty$).

- 1069 (i) *If a model is $(m, \infty, k, \varepsilon)$ -LD hard, then it is (m', k) -USQ hard with $m' = m/(k\varepsilon^{1/k})$.*
1070 (ii) *If a model is (m, k) -USQ hard, it is $(m, \infty, k, e - 1)$ -LD hard. More generally, it will be*
1071 *$(m', \infty, k, em'/m)$ -LD hard for all $m' < m$.*

1072 *Proof.* Assume that the model is $(m, \infty, k, \varepsilon)$ -LD hard. Then

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

1073 and in particular, for k even,

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

1074 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
1075 implies that

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

1076 More generally, we will have for $m' < m$

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

1077 which concludes the proof. \square

1078 Combining this equivalence of USQ and LD($d = \infty$) with the equivalence between USQ and SQ
1079 in Proposition 1, we can directly state an (unconditional) equivalence between SQ and LD($d = \infty$)
1080 hardness. In order to transfer this equivalence to LD with $d < \infty$, one can assume that the model
1081 with $d = \infty$ and $d < \infty$ are close to each other: this assumption is equivalent to being *noise-robust*
1082 (in some sense, see discussions in [8]).

1083 **Assumption 2** (Noise robustness). *We say a “ \mathbb{P} versus \mathbb{Q} ” detection problem is (d, k, δ) -noise robust*
1084 *if*

$$\|\mathbb{E}_u[(L_u^{>d})^{\otimes k}]\|_{L^2(\mathbb{Q})}^2 \leq \delta. \quad (17)$$

1085 Under this assumption, one can state the equivalence between LD and USQ:

1086 **Proposition 3** (Equivalence between LD and USQ Hardness).

- 1087 (i) *If the model is (m, t) -USQ hard, then the model is also $(m', d, k', em'/m)$ -LD hard for all*
1088 *$m' \leq m, k' \leq t$ and $d \geq 1$.*

1089 (ii) If the model is (m, d, k, ε) -LD hard and we further assume that it is (d, k, δ) -noise robust,
 1090 then the model is (m', k) -USQ hard with

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

1091 *Proof of Proposition 3.* Part (i) is directly implied by Proposition 2.(ii). For part (ii), following the
 1092 same argument as in the proof of Proposition 2.(i), we get

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

1093 Then using [8, Lemma 3.4], we obtain

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

1094 which concludes the proof. \square

1095 Then, the equivalence between LD and SQ hardness in [8] is obtained by combining Proposition 3
 1096 and Proposition 1. We state it below for completeness:

1097 **Theorem 5** (Equivalence between LD and SQ hardness).

1098 (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
 1099 $(m', d, k', em'/m)$ -LD hard for all $m' \leq m$, $k' \leq t/2$, and $d \geq 1$.

1100 (ii) If the model is (m, d, k, ε) -LD hard and we further assume that it is (d, k, δ) -noise robust,
 1101 then the model is $(q, m'/q^{2/t})$ -SQ hard for all $q \geq 1$ with

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

1102 B.3 Equivalence of GFP, SQ, and LD hardness for noise-robust models

1103 Based on the SQ-LD equivalence stated in the previous section (Theorem 5) and the equivalence
 1104 between GFP and SQ (Theorem 3), we can state an equivalence between GFP and LD hardness for
 1105 noise-robust models.

1106 **Theorem 6** (LD and ρ_G -FP Equivalence). Suppose a “ \mathbb{P} versus \mathbb{Q} ” task satisfies Assumption 1 for a
 1107 group G .

1108 (i) If the model is (m, d, k, ε) -LD hard and we further assume that it is (d, k, δ) -noise robust,
 1109 then the model is $(q', m', e|G|^{-1}\tilde{m}q^{2/t}/\tilde{m})$ - ρ_G -FP hard for any integers $q \geq 1$, $q' \leq q/\sqrt{2}$,
 1110 and $m' \leq \tilde{m}/2$, with

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

1111 (ii) If a task is (q, m, ε) - ρ_G -FP hard for some q, m integers. Assume that there exists an
 1112 $r = r(q) > 0$ such that $\pi^2(\rho_G(u, v) < r) = 1 - q^{-2}$ and m is even. Then, for all even
 1113 integer $4 \leq t \leq \log(q)/\log(m)$, the model is also $(m', d, k', em'/\tilde{m})$ -LD hard for all
 1114 $m' \leq \tilde{m}$ and $k' \leq t/2$, and $d \geq 1$, where

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

1115 C Details of examples and proofs

1116 C.1 Gaussian Additive Models

1117 We here properly define what is a Gaussian additive model.

1118 A \mathbb{P} versus \mathbb{Q} task is a Gaussian additive model if it satisfies:

- 1119 1. Under the null model, $\mathbb{Q} = \mathcal{N}(0, I_n)$.
- 1120 2. Under the planted model \mathbb{P}_u , for some signal-to-noise ratio $\lambda > 0$ we set $Y = \lambda u + Z$ for
- 1121 some $Z \sim \mathbb{Q}$.

1122 The following theorem holds.

1123 **Lemma 2.** *Consider any GAM with symmetric π , i.e., $v = -v, v \sim \pi$. For any $u, v \in \text{support}(\pi)$,*

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

1124 *This is to say, any GAM satisfies Assumption 1 for $G = \mathbb{Z}_2$ acting by flipping the sign of u .*

Proof. The proof follows from the standard identity $\langle L_u, L_v \rangle_{\mathbb{Q}} = \exp(\lambda^2 \langle u, v \rangle)$ [6, Proposition 2.3] and therefore

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

1125 □

1126 C.2 Planted Sparse Models

1127 We start with the definition of a planted sparse model [6].

1128 A \mathbb{P} versus \mathbb{Q} task is a planted sparse model (PSM) if it satisfies:

- 1129 1. Under the null model, the one sample is given by $Y = (Y_1, \dots, Y_n) \sim \mathbb{Q}$, each entry
- 1130 $Y_i, i = 1, \dots, n$ is drawn independently from some distribution $\mathbb{Q}_i, i = 1, \dots, n$ on \mathbb{R} .
- 1131 2. Under the planted model \mathbb{P}_u , we associate u with a set of planted entries $\Phi_u \subset [n]$. Then on
- 1132 sample is generated as follows. For the entries $i \notin \Phi_u$, we draw Y_i independently from \mathbb{Q}_i
- 1133 (which is identical as in the \mathbb{Q} measure). For the entries in Φ_u we draw from an arbitrary
- 1134 joint distribution $P_u|_{\Phi_u}$ with the following symmetry condition: for any subset $S \subseteq \Phi_u$, the
- 1135 marginal distribution $P_u|_{\Phi_u}(S)$ does not depend on u but only on S , i.e. $P_u|_S = P_S$

1136 The following theorem holds.

1137 **Lemma 3.** *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*

1138 *satisfies Assumption 1 for the trivial group.*

1139 *Proof.* The proof follows from [6, Proposition 3.6] for $D = 0$. □

1140 Multiple well-known detection models satisfy this definition, see e.g., [6] for a well studied model of

1141 sparse regression [24], or [11] for Bernoulli group testing [1].

1142 C.2.1 Symmetric mixed sparse linear regression

1143 The symmetric mixed sparse linear regression (mSLR) setting, is a \mathbb{P} versus \mathbb{Q} detection task defined

1144 as follows. Given $k, n \in \mathbb{N}$ with $k \leq n$ and $\sigma^2 > 0$, we have:

1145 • Under the planted model, we first sample $u \sim \pi$ uniformly from set $u \in \{0, 1\}^n$ with
 1146 $\|u\|_0 = k$. Then, the sample $(x_i, y_i) \sim \mathbb{P}_u$ is generated by

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

1147 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
 1148 also denote $\text{SNR} := k/\sigma^2$.

1149 • Under the null model, the sample $(y_i, x_i) \sim \mathbb{Q}$ is generated by $y_i \sim \mathcal{N}(0, 1)$ and indepen-
 1150 dently $x_i \sim \mathcal{N}(0, I_n)$.

1151 To see that mSLR is a PSM, set for any u , $\phi_u := \text{support}(u) \cup \{n+1\}$, that is the coordinates of
 1152 $((x_i)_j)_{j \in \text{support}(u)}$ and of y_i . Then it is easy to confirm that for any subset $S \subseteq \Phi_u$, the marginal
 1153 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
 1154 alter this distribution.

1155 It is known that the information theory sample size threshold of the problem is

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

1156 see e.g., [19]. Also in [5] it was proven that in the similar mSLR setting where u 's coordinates can
 1157 take values in $\{-1, 0, 1\}$, if

$$m = \tilde{o}\left(\frac{(\text{SNR} + 1)^2}{\text{SNR}^2} k^2\right)$$

1158 then the problem is $O(\log n)$ -degree hard. Here we prove that with sample size $\tilde{o}\left(\frac{(\text{SNR}+1)^2}{\text{SNR}^2} k^2\right)$ the
 1159 problem is also GFP-hard, and hence via Theorem 3 also SQ-hard. Our result holds under a very
 1160 mild assumption on SNR being not exponential in k . Interestingly, the proof is relatively short.

1161 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
 1162 Gaussian measure.

1163 **Lemma 4.** *For any sample size m and any u, v binary k -sparse vector, the following holds for the*
 1164 *mSLR model:*

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

1165 Using Lemma 4 one can prove the GFP-hardness and the SQ-hardness.

1166 **Theorem 7.** *If $n^{\Omega(1)} = k^2 = o(n)$ then for any $m = o\left(\frac{k}{\log\left(\frac{\text{SNR}^2}{2\text{SNR}+1} + 1\right)}\right)$, it holds*

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

1167 *Moreover, for any constant $T > 0$, for any $m = O\left(\frac{(\text{SNR}+1)^2}{(\log n)^{2T+2}} k^2\right)$ and $q = \Theta(\exp((\log n)^T))$ then*
 1168 *mSLR is $(q, m, O(1))$ -GFP hard.*

1169 *In particular, if $\text{SNR} \leq e^{k^{1-\alpha}}$, for some $\alpha > 0$, then from Theorem 3, it is also*
 1170 *$(e^{\Theta((\log n)^T)}, \left(\frac{(\text{SNR}+1)^2}{\text{SNR}^2} k^2\right)^{1-o(1)})$ -SQ hard.*

1171 C.2.2 Omitted proofs for the mSLR hardness

1172 *Proof of Lemma 4.* 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
 1173 on the case $m = 1$.

1174 Let $Y = Y_1, X = X_1$. By definition and Bayes' rule,

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

1175 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

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

1176 and so

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

1177 Now a standard integration argument using the MGF of the χ^2 distribution (see e.g., the proof of [6,

1178 Proposition 6.8.] for an almost identical argument) gives for any u, v binary k -sparse vectors,

$$\langle L_u, L_v \rangle_{\mathbb{Q}} = \frac{\lambda^2}{2(2\lambda^2-1)^{1/2}} \mathbb{E}_{X \sim \mathbb{Q}} \left(\exp\left(\frac{1}{2\sigma^2(2\lambda^2-1)} [(1-\lambda^2)(\langle X, u \rangle^2 + \langle X, v \rangle^2) + 2\lambda^2\langle X, u \rangle\langle X, v \rangle]\right) \right) \quad (18)$$

$$+ \exp\left(\frac{1}{2\sigma^2(2\lambda^2-1)} [(1-\lambda^2)(\langle X, u \rangle^2 + \langle X, v \rangle^2) - 2\lambda^2\langle X, u \rangle\langle X, v \rangle]\right) \quad (19)$$

1179 Now of course the pair $(\langle X, u \rangle, \langle X, v \rangle)$ follows a bivariate Gaussian law with variances equals to k
 1180 and covariance $\langle u, v \rangle$. Hence, some standard manipulations (see again the proof of [6, Proposition
 1181 6.8.] for an almost identical argument) allow us to derive that for $Z \in \mathbb{R}^{1 \times 3}$ with i.i.d. $\mathcal{N}(0, 1)$
 1182 entries and

$$t := \frac{1}{2\sigma^2(2\lambda^2-1)} = \frac{1}{2\sigma^2(2k/\sigma^2+1)} = \frac{1}{4k+2\sigma^2}. \quad (20)$$

1183 it holds

$$\langle L_u, L_v \rangle_{\mathbb{Q}} = \frac{\lambda^{2m}}{2(2\lambda^2-1)^{m/2}} \mathbb{E}_Z (\exp(t\langle M_1, Z^\top Z \rangle) + \exp(t\langle M_2, Z^\top Z \rangle)) \quad (21)$$

1184 where for $\ell := \langle u, v \rangle$,

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

1185 and

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

1186 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 ,

$$\begin{aligned} u_1^\top &= (0 \quad 1 \quad -1) & \lambda_1 &= (1-2\lambda^2)(k-\ell) \\ u_2^\top &= (\sqrt{k-\ell} \quad -\sqrt{\ell} \quad -\sqrt{\ell}) & \lambda_2 &= 0 \\ u_3^\top &= (2\sqrt{\ell} \quad \sqrt{k-\ell} \quad \sqrt{k-\ell}) & \lambda_3 &= k+\ell. \end{aligned} \quad (22)$$

1187 and for M_2 ,

$$\begin{aligned} u_1^\top &= (0 \quad 1 \quad -1) & \lambda_1 &= k-\ell \\ u_2^\top &= (\sqrt{k-\ell} \quad -\sqrt{\ell} \quad -\sqrt{\ell}) & \lambda_2 &= 0 \\ u_3^\top &= (2\sqrt{\ell} \quad \sqrt{k-\ell} \quad \sqrt{k-\ell}) & \lambda_3 &= (1-2\lambda^2)(k+\ell). \end{aligned} \quad (23)$$

1188 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

1189 A.5.] for $B(U) = \mathbb{R}^{n \times m}$, we have

$$\langle L_u, L_v \rangle_{\mathbb{Q}} = \frac{\lambda^2}{2(2\lambda^2-1)^{1/2}} (\det(I_3 - 2tM_1)^{-1/2} + \det(I_3 - 2tM_2)^{-1/2}). \quad (24)$$

1190 Using (20) and (22), (23) the eigenvalues of the matrices $I_3 - 2tM_1, I_3 - 2tM_2$ are

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

1191 and

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

1192 Since $\lambda^2 = k/\sigma^2 + 1$ we have

$$\begin{aligned} \frac{\lambda^2}{\sqrt{2\lambda^2 - 1}} \det(I_3 - 2tM_1)^{-1/2} &= \lambda^2 \left[\left(2\lambda^2 - 1 - \frac{k + \ell}{\sigma^2}\right) \left(1 + \frac{k - \ell}{\sigma^2}\right) \right]^{-1/2} \\ &= \frac{\frac{k}{\sigma^2} + 1}{1 + \frac{k - \ell}{\sigma^2}} \\ &= \left(1 - \frac{\ell}{k + \sigma^2}\right)^{-1}. \end{aligned}$$

1193 and by symmetry,

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

1194 Combining the above,

$$\langle L_u, L_v \rangle_{\mathbb{Q}} = \frac{1}{2} \left(\left(1 - \frac{\ell}{k + \sigma^2}\right)^{-1} + \left(1 + \frac{\ell}{k + \sigma^2}\right)^{-1} \right) \quad (25)$$

$$= \left(1 - \left(\frac{\ell}{k + \sigma^2}\right)^2\right)^{-1} \quad (26)$$

$$\leq \exp\left(\frac{m\ell^2}{(k + \sigma^2)^2 - \ell^2}\right), \quad (27)$$

1195 where for the last inequality we used that $\log x \geq 1 - 1/x$, for $x > 0$. \square

1196 *Proof of Theorem 7.* We have from the first part of Lemma 4,

$$\chi^2(\mathbb{P}^{\otimes m}, \mathbb{Q}^{\otimes m}) - 1 = \mathbb{E}_{u, v \sim \pi} \langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} \leq \mathbb{E}_{u, v \sim \pi} \left(1 - \frac{\langle u, v \rangle^2}{(k + \sigma^2)^2}\right)^{-m} \quad (28)$$

1197 But, in this setting $\langle u, v \rangle$ follows an Hypergeometric distribution with parameters n, k, k . Hence, by
1198 [6, Lemma 6.6],

$$\begin{aligned} \chi^2(\mathbb{P}^{\otimes m}, \mathbb{Q}^{\otimes m}) - 1 &\leq \sum_{\ell=0}^k \left(1 - \frac{\ell^2}{(k + \sigma^2)^2}\right)^{-m} \left(\frac{k^2}{n - k}\right)^{\ell} \\ &\leq \sum_{\ell=0}^{\lfloor k/2 \rfloor} \exp\left(\frac{m\ell^2}{(k + \sigma^2)^2 - \ell^2}\right) e^{-\ell \log(\frac{n-k}{k^2})} + \sum_{\ell=\lfloor k/2 \rfloor}^k \left(1 - \frac{k^2}{(k + \sigma^2)^2}\right)^{-m} e^{-\ell \log(\frac{n-k}{k^2})} \\ &\leq \sum_{\ell=0}^{\lfloor k/2 \rfloor} e^{\ell m / (k - \ell) - \ell \log(\frac{n-k}{k^2})} + k \left(\frac{(\frac{k}{\sigma^2})^2}{2\frac{k}{\sigma^2} + 1} + 1\right)^m e^{-\Theta(k \log(\frac{n-k}{k^2}))}, \end{aligned}$$

1199 where for the last inequality we used $\log x \geq 1 - 1/x$ for $x > 0$.

1200 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 as $\langle L_u, L_v \rangle_{\mathbb{Q}}$ is an increasing function of $\langle u, v \rangle^2$ we know that for any $q > 0$ there exists $\delta_0(q) > 0$ such that $\{\rho_{\text{id}}(u, v) \geq r(q)\} = \{\langle u, v \rangle^2 \geq \delta_0(q)\}$. But, notice that for any $q = e^{(\log n)^T}$ if we choose $\delta := \log(kq^2) = \Theta((\log n)^{T+1})$, then 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, $\delta_0 \leq \delta$. Moreover, from the second part of Lemma 4,

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

The SQ-hardness follows from 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 $\alpha > 0$. \square

C.3 Non-Gaussian Component Analysis

Borrowing terminology from single-index models [12], we define the *generative exponent* of the NGCA model:

Definition 13 (Generative exponent). *The generative exponent of the NGCA model in Definition 5 is defined as*

$$s^* := \min\{s \geq 1 : \lambda_s \neq 0\}, \quad \text{with} \quad \lambda_s := \mathbb{E}_{z \sim \mu}[h_s(z)],$$

where h_s is the (normalized) degree- s Hermite polynomial.

We can write the Hermite expansion of the likelihood function

$$L_u(x) = 1 + \sum_{s=s^*}^{\infty} \lambda_s h_s(\langle u, x \rangle) \quad (\text{in } L^2(\mathbb{Q})),$$

and the inner-product of the likelihood ratios for any $u, v \in \mathcal{S}^{n-1}$

$$\langle L_u, L_v \rangle = 1 + \sum_{s=s^*}^{\infty} \lambda_s^2 \cdot (\langle u, v \rangle)^s, \quad (29)$$

where we used that $\mathbb{E}[h_s(\langle u, x \rangle) h_k(\langle v, x \rangle)] = \delta_{ks} \langle u, v \rangle^s$. It is not true that for any u, v it holds $\langle L_u, L_v \rangle \geq 1$, as in Planted Sparse Models. Yet, our main result still applies for all NGCA models as Assumption 1 is satisfied, but now under the action of the group of order 2.

Lemma 5 (Nonnegativity of NGC Model). *Consider any NGCA model in Definition 5. For any $(u, v) \in \mathcal{S}^{n-1}$ and $k \in \mathbb{N}$*

$$\mathbb{E}_{\varepsilon_1, \varepsilon_2 \sim \text{Rad}(1/2)^{\otimes 2}} [\chi_{\mathbb{Q}}(\mathbb{P}_{\varepsilon_1 u}, \mathbb{P}_{\varepsilon_2 v})^k] \geq 1.$$

This is to say, any NGC model satisfies Assumption 1 for the group \mathbb{Z}_2 .

Proof. Note that $\langle L_{-u}, L_{-v} \rangle = \langle L_u, L_v \rangle$ and $\langle L_{-u}, L_v \rangle = \langle L_u, L_{-v} \rangle$. Hence, we only need to show the result for $k = 1$, as $X + Y \geq 0$ implies $X^k + Y^k \geq 0$ for all $k \geq 1$. We have simply

$$\mathbb{E}_{\varepsilon_1, \varepsilon_2} \langle L_{\varepsilon_1 u}, L_{\varepsilon_2 v} \rangle - 1 = \sum_{s=s^*}^{\infty} \lambda_s^2 \cdot \mathbb{E}_{\varepsilon_1, \varepsilon_2} [(\langle \varepsilon_1 u, \varepsilon_2 v \rangle)^s] \geq 0, \quad (30)$$

as for all odd s by symmetry $\mathbb{E}_{\varepsilon_1, \varepsilon_2} (\langle \varepsilon_1 u, \varepsilon_2 v \rangle)^s = 0$. \square

1226 Hence, under symmetric prior (that is $\pi(-u) = \pi(u)$), we have equivalence between GFP and SQ
 1227 hardness by our main theorem. We illustrate this equivalence for two standard priors: the uniform
 1228 prior $\pi = \text{Unif}(\mathcal{S}^{n-1})$ and the k -sparse prior $\pi = \text{Unif}(\{u \in \pm \frac{1}{\sqrt{k}}\{0, 1\}^n : \|u\|_0 = k\})$.

1229 **Theorem 8** (GFP hardness of NGCA, uniform prior). *Consider a NGCA model with generative*
 1230 *exponent s^* and the uniform prior $\pi = \text{Unif}(\mathcal{S}^{n-1})$. For any $\varepsilon \in (0, 1/2)$, the NGCA model is*
 1231 *$(\exp(\Theta(n^\varepsilon)), m, O(1))$ -GFP hard with*

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

1232 *Moreover, via our equivalence theorem, the model is $(\exp(n^{\Theta(\varepsilon)}), m^{1-\Theta(\varepsilon)})$ -SQ hard.*

1233 **Theorem 9** (GFP hardness of NGCA, sparse prior). *Consider a NGCA model with generative*
 1234 *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 $\varepsilon \in (0, 1/2)$*
 1235 *so that $k = n^{\Omega(\varepsilon)}$, the NGCA model is $(\exp(\Theta(n^\varepsilon)), m, O(1))$ -GFP hard with*

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

1236 *Moreover, via our equivalence theorem, the model is $(\exp(n^{\Theta(\varepsilon)}), m^{1-\Theta(\varepsilon)})$ -SQ hard.*

1237 The SQ lower bound in Theorem 8 was proven in [16] by a direct argument: here, we prove this SQ
 1238 hardness via equivalence to GFP hardness. The sparse prior was not considered previously and we
 1239 include it to illustrate the broad applicability of our equivalence. We prove these two theorems below.

1240 *Proof of Theorem 8.* Let us prove that the model is $\rho_{\mathbb{Z}_2}$ -FP hard and conclude using the implication
 1241 in Theorem 2.1. Note that $\rho_{\mathbb{Z}_2}(u, v) = \langle L_u, L_{\text{sign}(\langle u, v \rangle)} \rangle - 1$, that is

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

1242 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.
 1243 Using that $\langle u, v \rangle$ under the uniform prior is distributed as the first coordinate of $z \sim \text{Unif}(\mathcal{S}^{d-1})$, we
 1244 get

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

1245 for some universal constant $c > 0$. For simplicity denote $\rho = |\langle u, v \rangle|$. Using that $\lambda_s^2 \leq 1$ for any
 1246 $s \in \mathbb{N}$ by Jensen's inequality, we can write

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

1247 so that for $\rho = o_n(1)$ and n large enough, $\langle L_u, L_v \rangle - 1 \leq 2\lambda_{s^*}^2 \rho^{s^*}$. We deduce that for $\kappa = n^{-1/2+\varepsilon}$,
 1248 we have

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

1249 Thus we deduce that the problem is $(\exp(\Theta(n^\varepsilon)), m, \Theta(1))$ -GFP hard with $m = n^{s^*/2 - \Theta(\varepsilon)} / \lambda_{s^*}^2$.

1250 To use the equivalence with SQ, we need to compute the χ^2 -divergence, that is

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

1251 Let us bound

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

1252 where we denoted $M = \|L_u\|_{\mathbb{Q}}^2 - 1 = O(\exp(n^{\varepsilon/2}))$. Thus

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

1253 where we used that $t \log(t) = \tilde{\Theta}(n^{\varepsilon/2})$ by assumption. We can therefore apply Theorem 3 with
1254 $q' = \exp(n^{\varepsilon/2})$ and $t = n^{\varepsilon/2}$ (so that $t \leq \log(q)/\log(m) = \tilde{\Theta}(n^{\varepsilon})$). The model is (q', m') -SQ
1255 hard with

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

1256 which concludes the proof. \square

1257 *Proof of Theorem 9.* The proof proceeds similarly as the proof of Theorem 8. The main difference is
1258 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

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

1259 With this modification, the rest of the proof is identical and we omit it. \square

1260 **Lemma 6** (Tail bound for sparse prior). *Let u, v be independently sampled from the prior $\pi =$*
1261 *$\text{Unif}(\{u \in \pm \frac{1}{\sqrt{k}}\{0, 1\}^n : \|u\|_0 = k\})$. Then for any $t \geq 0$, we have*

$$\pi^2(\langle u, v \rangle \geq t) \leq \exp(-c \min\{nt^2, kt\}), \quad (31)$$

1262 for some universal constant $c > 0$.

1263 C.4 Single-Index Models

1264 We recall the definition of the generative exponent from [12]:

1265 **Definition 14** (Generative exponent). *The generative exponent of the Single-index model in Defini-*
1266 *tion 6 is defined as*

$$s^* := \min\{s \geq 1 : \lambda_s \neq 0\}, \quad \text{with} \quad \lambda_s := \|\zeta_s(Y)\|_{\mu_y}, \quad \zeta_s(y) := \mathbb{E}[h_s(z)|y] \quad (32)$$

1267 where h_s is the degree- s Hermite polynomial.

1268 In particular, the inner-product of likelihood functions is given by

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

1269 which is identical to the inner-product (29) for NGCA models. All our results only depend on
1270 $\langle L_u, L_v \rangle$: thus, all our statements for NGCA hold identically for single-index models, including the
1271 nonnegativity with \mathbb{Z}_2 (Lemma 5), as well as the examples of GFP hardness with uniform and sparse
1272 priors. For completeness, we state separate theorems for single-index models:

1273 **Theorem 10** (GFP hardness of SI models, uniform prior). *Consider a SI model with generative*
1274 *exponent s^* and the uniform prior $\pi = \text{Unif}(\mathcal{S}^{n-1})$. For any $\varepsilon \in (0, 1/2)$, the SI model is*
1275 *$(\exp(\Theta(n^\varepsilon)), m, O(1))$ -GFP hard with*

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

1276 Moreover, via our equivalence theorem, the model is $(\exp(n^{\Theta(\varepsilon)}), m^{1-\Theta(\varepsilon)})$ -SQ hard.

1277 **Theorem 11** (GFP hardness of SI models, sparse prior). *Consider a SI model with generative*
 1278 *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 $\varepsilon \in (0, 1/2)$*
 1279 *so that $k = n^{\Omega(\varepsilon)}$, the SI model is $(\exp(\Theta(n^\varepsilon)), m, O(1))$ -GFP hard with*

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

1280 *Moreover, via our equivalence theorem, the model is $(\exp(n^{\Theta(\varepsilon)}), m^{1 - \Theta(\varepsilon)})$ -SQ hard.*

1281 The SQ lower bounds in Theorem 10 and Theorem 11 were proven in [12] and [9] via direct argument.
 1282 Here, we obtain these bounds via equivalence to GFP hardness.

1283 C.5 Truncated Statistics

1284 Our first result is that all α -convex truncated models satisfy Assumption 1.

1285 **Lemma 7.** *Consider an α -convex truncated model in Definition 7. For any K, K' two symmetric*
 1286 *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*
 1287 *Truncated Model satisfies Assumption 1 for the trivial group.*

1288 *Proof.* For any K , it holds $L_K(x) = 1(x \in K)/\mathbb{Q}(K)$, $x \in \mathbb{R}^n$. Hence,

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

1289 But the so-called Gaussian correlation inequality for symmetric convex bodies in convex geometry
 1290 [37] states exactly that for any symmetric convex bodies K, K' it holds $\mathbb{Q}(K \cap K') \geq \mathbb{Q}(K)\mathbb{Q}(K')$
 1291 yielding the result. \square

1292 C.5.1 A new SQ lower bound for convex truncation

1293 As we mentioned in the main body [15] recently proposed a polynomial-time algorithm that can
 1294 achieve detection for any α -convex truncation setting with $O(n/\alpha^2)$ samples. In [15], it is also proven
 1295 that for some convex bodies K , $\Omega(n/\alpha)$ samples are information-theoretically required, leaving an
 1296 open gap between $\Theta(n/\alpha)$ samples and $\Theta(n/\alpha^2)$ samples. Using the GFP-hardness to SQ-hardness
 1297 framework we prove that for some prior on K , it is SQ-hard to distinguish with $\tilde{o}(n/\alpha^2)$ samples,
 1298 providing evidence that the polynomial-time method from [15] cannot be improved.

1299 To do so, we focus on the following prior on K , a variant of which has been studied in [15] to
 1300 prove their information-theoretic lower bound of $\Omega(n/\alpha)$ samples. To define it we let $K = K_v =$
 1301 $\{x \in \mathbb{R}^d : |\langle x, v \rangle| \leq \kappa\}$ for any $v \in \text{Unif}(\{-1/\sqrt{d}, 1/\sqrt{d}\}^d)$. Here, we choose $\kappa = \kappa(\alpha, d)$ is
 1302 such that the Gaussian measure of each K_v is $1 - \alpha$. Then our prior is uniform among $K_v, v \sim$
 1303 $\text{Unif}(\{-1/\sqrt{d}, 1/\sqrt{d}\}^d)$. We refer to the α -convex truncation setting with this prior as the “ α -Slice
 1304 Convex Truncation” model.

1305 We first point out that for any $m = \omega(n/\alpha)$, detection with m samples is always possible in the
 1306 α -Slice Convex Truncation model from a time-inefficient method. Indeed, one can brute-force search
 1307 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
 1308 \mathbb{P} , there always exists such a vector v and hence the brute force search algorithm will find it with
 1309 probability 1. Under \mathbb{Q} though a direct union bound gives that such a v exists only with probability
 1310 at most $2^d(1 - \alpha)^m = o(1)$ for any $m = \omega(d/\alpha)$. Hence, the algorithm can detect with probability
 1311 $1 - o(1)$. In that context, we prove the following result.

1312 **Theorem 12** (ρ_{Id} -FP- and SQ-hardness of Convex Truncation). *Let $n \in \mathbb{N}$ growing and arbitrary*
 1313 *$\alpha = \alpha_n \in (0, 1)$. There exists a universal constant $C > 0$ and a prior π on the convex bodies K of*
 1314 *Gaussian volume $1 - \alpha$ such that for any $q \in \mathbb{N}$ with $q = e^{o(\alpha n)}$, the α -Convex Truncation model*
 1315 *under π is $(q, \frac{Cn}{\alpha^2 \log(1/\alpha)^{3/2} \log q})$ - ρ_{Id} -FP-hard.*

1316 *In particular, for any constant $T > 0$ if $\alpha = \omega(\frac{(\log n)^T}{n})$ then the α -Convex Truncation model under*
 1317 *π is $(2^{(\log n)^T}, \frac{Cn}{\alpha^2 \log(1/\alpha)^{3/2} (\log n)^{T+1}})$ -SQ hard.*

1318 Satisfyingly the proof of this result is also relatively short.

1319 *Proof of Theorem 12.* Observe that for $L_u := L_{K_u}$ we have via standard Hermite expansion (identi-
1320 cal to the argument in [15, Line (32), proof of Claim 24],

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

1321 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
1322 Φ is the CDF of a standard Gaussian.

1323 Now, conveniently, the authors [27] have already studied the Hermite mass of indicators of symmetric
1324 intervals around 0. Indeed, applying [27, Lemma 27] for $j = 2, \theta = k$ imply that

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

1325 where ϕ is the PDF of a standard Gaussian. But by standard tail bounds $\kappa = O(\sqrt{\log(1/\alpha)})$ and
1326 therefore from the Mill's ratio bound $\phi(\kappa) = \Theta((1 - \Phi(\kappa))\kappa)$ which all together give

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

1327 Parseval's identity gives $\sum_{i>0} f_i^2 = \alpha(1 - \alpha) \leq \alpha$, and hence we conclude that for some constant
1328 $C > 0$

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

1329 Now $\langle L_u, L_v \rangle_{\mathbb{Q}}$ is an increasing function of $\langle u, v \rangle^2$. Hence, for any $q > 0$ there exists $\delta_0(q) > 0$
1330 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
1331 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 $\delta_0(q) \leq \delta = C' \frac{\log q}{n}$.

1332 Combining the above we have that for any $q = e^{o(\alpha n)}$,

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

1333 as long as $m = O(d/(\alpha^2 \log(1/\alpha)^{3/2} \log q))$. So we conclude the $(q, O(d/(\alpha^2 \log(1/\alpha)^{3/2} \log q)))$ -
1334 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 $T > 0$,

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

1335 By our equivalence Theorem 3 for $m_{\text{IT}} = (\log n)^T$ we derive for $\log q = (\log n)^{T+1}$, $t = (\log n)^T$
1336 the $(2^{(\log n)^T}, O(n/(\alpha^2 \log(1/\alpha)^{3/2} (\log n)^{T+1})))$ -SQ hardness of the model. \square

1337 D Counterexample

1338 Below, we provide details on the counterexample described in Section 4. We first prove the formula
1339 in Lemma 1 for the inner-product of likelihood ratios.

1340 *Proof of Lemma 1.* By definition, for any $u \in \{0, 1\}^{n+1}$,

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

1341 For any $u, v \in \{0, 1\}^{n+1}$, the inner product $\langle L_u, L_v \rangle$ satisfies

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

1342 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 $a_i = 1 + r^2$;
1343 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
1344 $a_i = 1 + r^2 \cdot \alpha$. We deduce that $a_i = 1 + r^2 \cdot \alpha^{u_i+v_i}$ and the lemma follows. \square

1345 Let us consider the m -sample version of the hypothesis testing problem. The null hypothesis is then
1346 $\mathbb{Q}^{\otimes m}$ and the alternative hypothesis is $\mathbb{E}_{u \sim \pi} \mathbb{P}_u^{\otimes m}$, where u is sampled from the following two-point
1347 prior π :

$$u = \begin{cases} (1, 0, \dots, 0), & \text{w.p. } \rho, \\ (0, 1, \dots, 1), & \text{w.p. } 1 - \rho. \end{cases} \quad (33)$$

1348 We abbreviate these vectors as $u_1 = (1, 0, \dots, 0)$ and $u_2 = (0, 1, \dots, 1)$ for convenience. Using
1349 Lemma 1, it holds that

$$\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle = \prod_{i=0}^n (1 + r^2 \cdot \alpha^{u_i+v_i})^m. \quad (34)$$

1350 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
1351 therefore the model verifies Assumption 1 for the trivial group. For convenience, we restate the
1352 theorem from the main text below:

1353 **Theorem 13.** *For the two-point prior π in (34) with $\rho = \exp(-n^\varepsilon/2)$, and for $r = n^{-1/2}$, $\alpha =$
1354 $n^{-1+2\varepsilon}$, $m = n^{1-\varepsilon}$ and $D = n^\varepsilon$, where $\varepsilon > 0$ is any small constant, the following hold. The
1355 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
1356 but not $(n^{-1}, m, \exp(\Theta(n^\varepsilon)))$ -FP hard. Moreover, via our equivalence theorem the model is
1357 $(e^{n^{\Theta(\varepsilon)}}, n^{1-\Theta(\varepsilon)})$ -SQ hard.*

1358 *Proof.* Let us first show it is FP easy. Define $\delta := \delta(n^{-1/2})$ the supremum over δ such that
1359 $\pi^2(\langle u, v \rangle \geq \delta) \geq 1/n$. We observe when $u \neq v$, then we must have $\langle u, v \rangle = 0 < \delta$ by the choice of
1360 the two points prior with $\langle u_1, u_2 \rangle = 0$. Therefore,

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

1361 We deduce the following lower bound

$$\begin{aligned}\mathbb{E}_{u,v}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(\langle u, v \rangle < \delta)] &\geq \mathbb{E}_{u,v}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(u \neq v)] \\ &= \pi^2[u \neq v] \cdot \mathbb{E}_{u,v}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle | u \neq v]\end{aligned}$$

1362 When conditioned on $u \neq v$, we get

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

1363 by applying Eq. (34), with $u_i + v_i = 1$ for all $0 \leq i \leq n$. Inserting the parameters stated in the
1364 lemma, we obtain

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

1365 This shows that under our parameter choice, the task is $(n^{-1/2}, m, \exp(\Theta(n^\varepsilon)))$ -FP easy.

1366 Let us now show that this model is GFP hard. We will prove that the model is ρ_{Id} -FP hard and conclude
1367 using the implication Theorem 2.1. Under the trivial group, we have $\rho_{\text{Id}}(u, v) = \langle L_u, L_v \rangle_{\mathbb{Q}} - 1$.
1368 From Eq. (34), the m -sample inner product of likelihood ratio is given for $u = v = u_1$ by

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

1369 and for $u = v = u_2$ by

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

1370 Because $\alpha \ll 1$, it is not hard to notice

$$\langle L_{u_1}^{\otimes m}, L_{u_2}^{\otimes m} \rangle < \langle L_{u_2}^{\otimes m}, L_{u_2}^{\otimes m} \rangle < \langle L_{u_1}^{\otimes m}, L_{u_1}^{\otimes m} \rangle. \quad (35)$$

1371 From the definition of π , it holds that

$$\pi^2(\{u = u_2, v = u_2\} \cup \{u \neq v\}) = 1 - e^{-n^\varepsilon} \geq 1 - q^{-2}, \quad (36)$$

1372 using that we set $q = \exp(D/2)$. Combining with Eq. (35), we conclude that the event $\{\rho(u, v) \leq$
1373 $r(q)\} \subset \{u = u_2, v = u_2\} \cup \{u \neq v\}$. This allows us to estimate the upper bound as

$$\begin{aligned}\mathbb{E}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(r \leq r(q))] &\leq \mathbb{E}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(\{u = u_2, v = u_2\} \cup \{u \neq v\})] \\ &\leq (1 + \alpha^2 r^2)^{nm} \cdot (1 + r^2)^m.\end{aligned} \quad (37)$$

1374 Inserting our choice of parameters, we obtain

$$\begin{aligned}\mathbb{E}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle \cdot \mathbf{1}(r \leq r(q))] &\leq (1 + n^{-3+4\varepsilon})^{n^{2-\varepsilon}} \cdot (1 + n^{-1})^{n^{1-\varepsilon}} \\ &\leq \exp(n^{-1+3\varepsilon} + n^{-\varepsilon}) \leq 1 + 2n^{-\varepsilon}.\end{aligned} \quad (38)$$

1375 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.

1376 Finally, let us use the SQ-GFP equivalence in Theorem 3 to show that the model is also SQ hard,
1377 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
1378 theorem, we need to compute the χ^2 -divergence. Denoting $X = \langle L_u^{\otimes 4t}, L_v^{\otimes 4t} \rangle$ with $t = n^{\varepsilon/2}$,

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

1379 Thus, we obtain

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

1380 and we deduce the model is $(e^{D/2}, m^{1-\Theta(\varepsilon)})$ -SQ hard. \square

1381 E Proofs of equivalence

1382 E.1 Proof of Theorem 2

1383 *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
1384 event

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

1385 it clearly holds $\pi^{\otimes 2}(A) \geq 1 - q^{-2}$ and, since G is a group, $G^{\otimes 2}(A) = A$. Hence,

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

1386 implying the desired result.

1387 We now focus on the other direction. By decomposing the likelihood ratio inner product, we obtain

$$\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} = (\langle L_u, L_v \rangle_{\mathbb{Q}} - 1 + 1)^m = \sum_{t=0}^m \binom{m}{t} \cdot (\langle L_u, L_v \rangle_{\mathbb{Q}} - 1)^t. \quad (39)$$

1388 Taking expectation over the prior $\pi^{\otimes 2}$ conditioned on *any* event A satisfying $G^{\otimes 2}(A) = A$ and
1389 $\pi^2(A) = 1 - q^{-2}$, we have

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

1390 where in the second equality, we used that G is a π -preserving transformation, and for the inequality
1391 we use Assumption 1.

1392 Clearly for all $t \geq 0$,

$$\mathbb{E}_{g \sim \text{Unif}(G)} (\langle L_{g(u)}, L_{g(v)} \rangle_{\mathbb{Q}} - 1)^{2t} \geq |G|^{-1} \rho_G(u, v)^{2t}. \quad (40)$$

1393 Therefore, we further conclude that

$$\mathbb{E}[\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} - 1 | A] \geq |G|^{-1} \sum_{t=1}^{\lfloor m/2 \rfloor} \binom{m}{2t} \cdot \mathbb{E}[\rho_G(u, v)^{2t} | A]. \quad (41)$$

1394 Recall that $r(q)$ satisfies

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

1395 Hence, by definition of $r(q)$ we have

$$|G|^{-1} \sum_{t=1}^{\lfloor m/2 \rfloor} \binom{m}{2t} \cdot \mathbb{E}[\rho_G(u, v)^{2t} | A] \geq |G|^{-1} \sum_{t=1}^{\lfloor m/2 \rfloor} \binom{m}{2t} \cdot \mathbb{E}[\rho_G(u, v)^{2t} | \rho_G(u, v) \leq r(q)] \quad (42)$$

$$\geq |G|^{-1} \sum_{t=1}^{\lfloor m/2 \rfloor} \binom{m}{2t} \cdot \mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^{2t} | \rho_G(u, v) \leq r(q)]. \quad (43)$$

1396 In addition, we notice that for each even order $2t+1$ with $t = 1, \dots, \lfloor m/2 \rfloor - 1$, it holds by Lemma 8
 1397 that

$$\frac{\binom{m}{2t+1} \cdot |\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^{2t+1}}{\sqrt{\binom{m}{2t} \cdot |\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^{2t} \cdot \binom{m}{2t+2} \cdot |\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^{2t+2}}} = \frac{\binom{m}{2t+1}}{\sqrt{\binom{m}{2t} \cdot \binom{m}{2t+2}}} \leq 2. \quad (44)$$

1398 Therefore, using the inequality $2\sqrt{ab} \leq a + b$ for $a, b \geq 0$, we obtain

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

1399 Consequently, the right-hand-side of (43) can be further lower bounded by

$$|G|^{-1} \sum_{t=1}^{\lfloor m/2 \rfloor} \binom{m}{2t} \cdot \mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^{2t} \mid \rho_G(u, v) \leq r(q)] \quad (45)$$

$$\geq \frac{|G|^{-1}}{3} \sum_{t=2}^m \binom{m}{t} \cdot \mathbb{E}[|\langle L_u, L_v \rangle_{\mathbb{Q}} - 1|^t \mid \rho_G(u, v) \leq r(q)] \quad (46)$$

$$\geq \frac{|G|^{-1}}{3} \sum_{t=2}^m \binom{m}{t} \cdot \mathbb{E}[(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1)^t \mid \rho_G(u, v) \leq r(q)]. \quad (47)$$

1400 Combining (41), (43), and (47), with the condition of (q, m, ε) -GFP $_{\mathcal{T}}$ hardness, we obtain

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

1401 Again, by the definition of $r(A)$ it follows that

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

1402 and therefore by (39)

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

1403 Next, we aim to upper bound the first order term, namely $m \cdot \mathbb{E}[(\langle L_u, L_v \rangle - 1) \cdot \mathbf{1}(\rho_G(u, v) \leq$
 1404 $r(q))]$. Note that $A' := \{(u, v) : \rho_G(u, v) \leq r(q)\}$ is also $G^{\otimes 2}$ -invariant. Hence, employing also

1405 Assumption 1 we also have

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

1406 Therefore,

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

1407 from which the result follows. \square

1408 **Lemma 8.** For any $t \in \{1, 2, \dots, n-1\}$ and $n \geq 3$, we have

$$\frac{\binom{n}{t}^2}{\binom{n}{t-1} \cdot \binom{n}{t+1}} \leq 4. \quad (48)$$

1409 *Proof.* Note that by the successive ratio between binomial coefficients, we have

$$\frac{\binom{n}{t}^2}{\binom{n}{t-1} \cdot \binom{n}{t+1}} = 1 + \frac{1+n}{t(n-t)} \leq 1 + \frac{1+n}{n-1} = 2 + \frac{2}{n-1} \leq 4. \quad (49)$$

1410 This completes the proof. \square

1411 **E.2 Proof of Theorem 3**

1412 *Proof of Theorem 3. SQ implies ρ_G -FP.* We have that

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

1413 Now as G is π -preserving that easily implies that for any A such that $G^{\otimes 2}(A) = A$, that

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

1414 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
 1415 $\pi^2(A_r) \geq q^{-2}$ implies $r \leq |G|/m$. Recall that $r(q) > 0$ satisfies $\pi^2(A_{r(q)}) \geq q^{-2}$. In particular,
 1416 $r(q) \leq |G|/m$, and therefore for any $m' \leq m/2$,

$$\begin{aligned} \mathbb{E} [\langle L_u^{\otimes m'}, L_v^{\otimes m'} \rangle_{\mathbb{Q}} \cdot \mathbf{1}(\rho_G(u, v) \leq r(q))] &= \mathbb{E} [(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1 + 1)^{m'} \cdot \mathbf{1}(\rho_G(u, v) \leq r(q))] \\ &\leq \mathbb{E} [(\rho_G(u, v) + 1)^{m'} \cdot \mathbf{1}(\rho_G(u, v) \leq r(q))] \quad (50) \\ &\leq (r(q) + 1)^{m'} \leq (|G|/m + 1)^{m'} \leq 1 + e|G|m'/m. \quad (51) \end{aligned}$$

1417 This concludes the $(q, m', e|G|m'/m)$ - ρ_G -FP hardness.

1418 **ρ_G -FP hardness implies SQ-hardness.** Suppose we have (q, m, ε) - ρ_G -FP hardness

$$\mathbb{E} [\langle L_u^{\otimes m}, L_v^{\otimes m} \rangle_{\mathbb{Q}} \cdot \mathbf{1}(\rho_G(u, v) < r(q))] \leq 1 + \varepsilon, \quad \text{where} \quad \pi^2(\rho_G(u, v) \geq r(q)) = q^2. \quad (52)$$

1419 By definition 4, we have that

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

1420 where the first inequality holds by the definition of the ρ_G -FP hardness and the second equality holds
 1421 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
 1422 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
 1423 $t \geq 0$, we have

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

1424 Hence, combining the two for all even t , with $1 \leq t \leq m$,

$$\max_{\substack{1 \leq t \leq m, \\ t \text{ even}}} \mathbb{E} [(\langle L_u, L_v \rangle_{\mathbb{Q}} - 1)^t \cdot \mathbf{1}(\rho_G(u, v) < r(q))] \leq \frac{1 + \varepsilon}{\binom{m}{t}}. \quad (53)$$

1425 Therefore, we have for any even t with $t \leq m$ that

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

1426 where in the first inequality, we use the Cauchy-Schwarz inequality for the second term and the fact
1427 that $\pi^2(\rho_G(u, v) \geq r(q)) \leq q^2$. In the second term, we use the elementary $\binom{m}{t} \geq (m/t)^t$.

1428 Now focusing on $t \leq \log q / \log m$ we further have

$$\mathbb{E}[(\langle L_u, L_v \rangle - 1)^t] \leq \left(\frac{t(1 + \varepsilon)^{1/t} + \chi^2(\mathbb{P}^{\otimes 4t} \parallel \mathbb{Q}^{\otimes 4t})}{m} \right)^t. \quad (54)$$

1429 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

1430 $q' > 0$ 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

1431 $t = (\log m)^s$ and $q' = e^{\delta(\log m)^{s+1}}$. □

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.

- 1520 [32] Rafał Łatała and Dariusz Matlak. Royen’s proof of the gaussian correlation inequality. In
1521 *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2014–2016*, pages 265–275.
1522 Springer, 2017.
- 1523 [33] Peter McCullagh. *Generalized linear models*. Routledge, 2019.
- 1524 [34] Andrea Montanari, Daniel Reichman, and Ofer Zeitouni. On the limitation of spectral meth-
1525 ods: From the gaussian hidden clique problem to rank-one perturbations of gaussian tensors.
1526 *Advances in Neural Information Processing Systems*, 28, 2015.
- 1527 [35] Karl Pearson. On the systematic fitting of curves to observations and measurements. *Biometrika*,
1528 1(3):265–303, 1902.
- 1529 [36] Emile Richard and Andrea Montanari. A statistical model for tensor pca. *Advances in neural*
1530 *information processing systems*, 27, 2014.
- 1531 [37] Thomas Royen. A simple proof of the gaussian correlation conjecture extended to multivariate
1532 gamma distributions. *arXiv preprint arXiv:1408.1028*, 2014.
- 1533 [38] Tselil Schramm and Alexander S Wein. Computational barriers to estimation from low-degree
1534 polynomials. *The Annals of Statistics*, 50(3):1833–1858, 2022.
- 1535 [39] Ilias Zadik, Min Jae Song, Alexander S Wein, and Joan Bruna. Lattice-based methods surpass
1536 sum-of-squares in clustering. In *Conference on Learning Theory*, pages 1247–1248. PMLR,
1537 2022.
- 1538 [40] Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and
1539 algorithms. *Advances in Physics*, 65(5):453–552, 2016.