Low-degree polynomials and Gaussian Graphical Models

Matteo Russo [*]

August 8, 2021

Abstract

Information-computation gap is the gap between number of samples to make a distinguishing problem information-theoretically possible and number of samples to make a distinguishing problem computationally viable. We study information-computation gaps for the Gaussian Graphical Model (GGM) distinguishing problem employing techniques entailing low degree polynomials and, in particular, multi-sample low degree likelihood ratios (LDLR) and statistical query complexity.

Previously the information-theoretic lower bound has been derived only for the estimation problem of a small set of precision matrices [WWR10]. In this work, we study an associated easier hypothesis testing problem. We prove that below the same information threshold, multivariate Gaussian distributions parameterized by well-conditioned precision matrices has bounded LDLR for arbitrarily large polynomial degree. As complement result, we show for the same distinguishing problem, the LDLR diverges for nearly constant degree. This coincides the fact that efficient distinguishers exist in this case [MVL20]. Combining these two results, we conclude that in this hypothesis testing problem, the low degree model predicts no information-computation gap for any well-conditioned matrix.

For ill-conditioned precision matrices, we obtain similarly divergent LDLR for diagonal dominant matrices, which can be ill-conditioned. This implies either that the hypothesis testing problem we are considering is not hard enough for capturing the information-computation gap or that diagonal dominant matrices are not suitable matrices for predicting the separation in low degree model.

[*] ETH Zürich, D-INFK: mrusso@ethz.ch, Legi number: 19-909-258. This is a master Semester project supervised by Prof.Dr.David Steurer

Contents

1 Introduction 1
1.1 Model . . . 2
1.2 Overview of techniques . . . 2
1.3 Problem and goals . . . 4
1.4 Organization and summary of results . . . 5

2 Techniques 7
2.1 Low Degree Likelihood Ratio Test . . . 8
2.2 Information-Computation Gap . . . 10

3 Related Work 11
3.1 GGM: Information Theoretic Lower Bound . . . 12
3.2 GGM: Efficient algorithms for incoherent precision matrices . . . 13
3.3 GGM: Efficient algorithms for attractive and walk-summable precision matrices 14
3.4 GGM: A first attempt to find an IC Gap via Statistical Dimension . . . 15

4 Proofs of results 17
4.1 LDLR Reformulation for GGMs . . . 17
4.2 Choice of null and alternative distributions . . . 19
4.3 Information threshold from LDLR . . . 24
4.4 Well-conditioned matrices cannot predict an IC Gap via LDLR . . . 25
4.4.1 Low degree bounds for well-conditioned matrices . . . 26
4.4.2 Divergence of LDLR for well-conditioned matrices . . . 28

5 Beyond well-conditioned matrices 29
5.1 Natural generalization from current lower bound . . . 29
5.2 Diagonal dominant precision matrices . . . 29
5.3 Future work: a harder distinguishing problem . . . 31

6 Conclusion 31

References 32

1 Introduction

Nowadays, statistical accuracy in predictive analytics, and several other fields where estimation and hypothesis testing problems are faced on a daily basis, is of the uttermost importance. A plethora of industries uses machine learning tools to support their decision making and high accuracy is a must. Quality of predictions, in fact, heavily depends not only on the amount of data one possesses but also on its quality and dimensionality. The more data one has (and the better it is), the higher accuracy one would obtain.

There is, however, a fundamental barrier in achieving such a high accuracy in statistical problems: the optimal guarantee is often achieved by algorithms such as brute-force search while current known polynomial-time algorithms are sub-optimal. As data streams become gargantuan (together with their dimensionality), exponential-algorithms cannot process all this data in a reasonable amount of time. This gap between the optimal statistical guarantee and the statistical guarantee achievable in polynomial time is often called *information-computation gap*.

In order for us to further understand this notion, let us consider a canonical distinguishing problem known as *planted clique*:

Problem 1.1 (Planted clique)**.** In the hypothesis testing problem of planted clique, we are give a graph with n vertices sampled from one of two following distributions with equal probability $1/2$.

- For null hypothesis, we construct a balanced Erdős-Rényi graph $G(n, 1/2)$, that is a random graph on n nodes such that an edge between two nodes is inserted with probability $1/2$.
- For alternative hypothesis, we take again $G(n, 1/2)$ and plant a clique on k nodes.

One may ask what is the value of *signal-to-noise parameter k* such that (1) it is information-theoretically possible to distinguish between null and alternative hypotheses and (2) there exists a polynomial time distinguisher between the two hypotheses. Put differently, one may ask whether the first and second quantity are near or far-apart from each other, in which case an *information-computation gap* arises. In [AKS98], the authors prove that $k \gtrsim \log n$ in order for the distinguishing to be possible, whereas a long-standing conjecture [AB09] asserts that $k \gtrsim \sqrt{n}$ for the distinguishing to be performed in polynomial time. This would induce an *information-computation gap* insofar as $\log n \ll \sqrt{n}$.

A number of techniques have been employed to predict *information-computation gaps* for a variety of estimation as well as distinguishing problems, among which the most successful (to cite a few) have been Sum-of-Squares proofs [Hop18, HKP+17, BS14] and low-degree polynomials [Hop18, KWB19]. We will make use of the latter together with the notion of *statistical query complexity* and *statistical dimension*, inspired by the work of [BBH+20]. This framework is reminiscent of complexity theory, where classes of problems are classified as easy (polynomial time algorithms exist), hard or even undecidable. If in the realm of traditional algorithmic hardness we are asking which kind of problems fall into which category [AB09], here, we are concerned with statistical hardness, i.e. the two *signal-to-noise parameters* before which the distinguishing problem is impossible, between which it is hard and after which it becomes easy.

This paper will be concerned with exploring the potential existence of *information-computation gaps* in Gaussian Graphical Models [Wai19a, WWR10, MVL20] (described in section 1.1), which have had a multitude of applications ranging from social sciences to computational biology and fMRI imaging [Wai19b].

1.1 Model

Gaussian Graphical Models (GGMs for short) is classical model in high dimensional statistics[Wai19a].

Definition 1.2 (Gaussian graphical model)**.** In Gaussian graphical model, we observe n independent samples $x_1, \dots, x_m \in \mathbb{R}^p$, each of which is drawn from a common underlying normal distribution $\mathcal{N}\left(\mu, \Theta^{-1}\right)$, with $p \in \mathbb{N}$, and the precision matrix $\Theta \in \mathbb{R}^{p \times p}$ is a positive semidefinite matrix.

In this work, we are interested in the Gaussian graphical models, where Θ is d-sparse, i.e no row of the precision matrix can have more than $d \in \mathbb{N}$ non-zero entries. Formally, let $\mathcal{E} := \left\{(i, j) \in [p] \times [p] \mid \Theta_{i,j} \neq 0\right\}$, we have

$$d := \max_{j \in [p]} \left|\left\{j \in [p] \mid (i, j) \in \mathcal{E}\right\}\right| \tag{1.1}$$

If we view the precision matrix as a sparse graph with n vertices, then this means that the degree of vertices are bounded by d.[Wai19a].

1.2 Overview of techniques

Hypothesis Testing. Hypothesis Testing is the problem of distinguishing, with high probability, whether m samples $x_1, \dots, x_m$ have been drawn from a null distribution $D_\emptyset$ or whether from an alternative family of distributions (succinctly denoted as) D_u, where $u \sim \mathcal{S}$. With null hypothesis $H_0 : x_1, \dots, x_m \sim D_\emptyset$ and alternatives $H_u : x_1, \dots, x_m \sim D_u$, we would like both of the following conditions to hold:

$$\textbf{Type I Error.} \qquad \mathbb{P}\left[\exists\, u \in \operatorname{supp}(\mathcal{S}) : H_u \mid x \sim D_\emptyset\right] \leqslant \delta^{(\mathbf{I})} \tag{1.2}$$

$$\textbf{Type II Error.} \qquad \mathbb{P}\left[H_0 \mid \exists\, u \in \operatorname{supp}(\mathcal{S}) : x \sim D_u\right] \leqslant \delta^{(\mathbf{II})} \tag{1.3}$$

Hereby, the probability of an hypothesis really means the probability that the distinguishing test outputs $x \in \operatorname{supp}(D_\emptyset)$ or $x \in \operatorname{supp}(D_u)$, respectively. In words, the above probability statements can simply be seen as the probability of rejecting a true null hypothesis (type I error) and the probability of failing to reject a false null hypothesis (type II error), both of which should be bounded by a small value (regardless of whether it is constant or not).

Since we will be dealing with a computational method that outputs a real value rather than the binary (or k-ary) output of the distinguishing function $\psi(H_0, H_u)$, this naturally inspires the following definition.

Definition 1.3 (β-distinguisher [BBH+20]). Testing function $\psi : \mathbb{R}^{p\times m} \to \mathbb{R}$ on samples $x_1, \dots, x_m \in \mathbb{R}^p$ is called a β-distinguisher for testing problem H_0-vs.-H_u if

$$\left| \mathop{\mathbb{E}}_{x\sim D_\emptyset} \psi(x) - \mathop{\mathbb{E}}_{u\sim\mathcal{S}} \mathop{\mathbb{E}}_{x\sim D_u} \psi(x) \right|^2 \geqslant \beta^2 \cdot \mathop{\mathbf{var}}_{x\sim D_\emptyset} \psi(x)$$

As a side remark, one can note that for $\beta > 1$, Chebyšev's inequality ensures a bounded one sided error.

Depending on the hypothesis distinguishing problem we would like to settle, there are two important quantities that need to be established: the first one being the number of samples below which it is information-theoretically impossible to distinguishing between null and alternative hypotheses, which takes the name of statistical sample complexity and is denoted by m_{STAT}. The second one is the number of samples above which it is the distinguishing between null and alternative hypotheses becomes tractable, that is solvable in polynomial time (in some settings quasi-polynomial or sub-exponential time suffices), which takes the name of computational sample complexity and is denoted by m_{COMP}, as illustrated by the following figure.

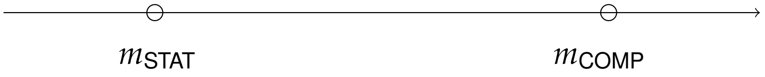

Figure 1: Information-Computation Gap illustration

As shown above, the information-computation gap only arises when $m_{\mathsf{STAT}} \ll m_{\mathsf{COMP}}$ and does not when $m_{\mathsf{STAT}} \simeq m_{\mathsf{COMP}}$ (explained in detail in subsection 2.2).

Low-degree Polynomials. As fully explained in section 2, we have a certain degree of freedom in choosing the testing function $\psi(\cdot)$. In particular, we would like a computational model that closely links the testing function form to the computational complexity of the distinguishing problem: the simplest ψ is a polynomial whose degree is constant or logarithmic in the sample complexity. If there exists polynomial function with degree $O(\log p)$ that performs the distinguishing task with high probability (as per expressions 1.2 and 1.3), then ψ will be called a low-degree polynomial distinguisher.

It is important to note that the low-degree implied lower bounds for distinguishing problems also hold in the estimation version of the same problem, that is they prove hardness of both the distinguishing and the estimation problems. On the other hand, the upper bounds do not imply that tractability of the estimation problem but only of the distinguishing task. In fact, estimation is at least as hard as distinguishing in that we could employ an algorithm for support estimation to solve the distinguishing problem, while the vice versa is not always true [SW20].

The existence of such a distinguisher (in particular, the $\log m$-degree one) is, as a matter of fact, almost interchangeable with computational tractability of the decision problem of distinguishing between H_0 and H_u. This was known as "Hopkins's Conjecture" [Hop18] and, to be precise, it is only true for a wide variety of high-dimensional testing problems such as, to cite a few, Sparse and Tensor PCA, Planted Clique and Stochastic Block Model [BBH+20, KWB19, HL18, BHK+16, ABARS20]. However, the same statement becomes false for explicitly constructed counterexamples such as the Planted 3-XOR [HW20].

1.3 Problem and goals

Having introduced GGMs and the general estimation techniques we would like to employ, we are now ready to state formally what our distinguishing problem is. In subsection 3.1, we will make explicit that $m_{\mathsf{STAT}} \in O\left(\log p/\kappa^2\right)$. Then, $m_{\mathsf{COMP}} \in \omega\left(\log p/\kappa^2\right)$ would imply an information-computation gap for the GGM problem. We are interested in the following two questions:

1. Estimating the support of Θ with high probability
2. Estimating Θ with small approximation error

These two problems are different and are not strictly harder than each other [KKMM20]. We anticipate that it is impossible to reconstruct the support if we allow arbitrarily small entries in Θ. Therefore we add a restriction on the minimum normalized edge strength or non-degeneracy parameter κ, which is introduced in [MVL20]: recalling that $\mathcal{E} := \left\{(i,j) \in [p]\times[p] \mid \Theta_{i,j} \neq 0\right\}$,

$$\kappa := \min_{(i,j)\in\mathcal{E}} \frac{|\Theta_{ij}|}{\sqrt{\Theta_{ii}\cdot\Theta_{jj}}} \tag{1.4}$$

Problem 1.4 (GGM Distinguishing Problem)**.** Let $d \ll p$ and κ be defined as above, the κ non-degenerate d-sparse p-dimensional GGM problem $\mathcal{G}(p,d,\kappa)$ is to distinguish the following distributions with high probability:

$$\begin{aligned} H_0 &: \ D_\emptyset = \mathcal{N}\left(\mathbf{0}, \Theta^{-1}\right) \\ H_u &: \ D_u = \mathcal{N}\left(\mathbf{0}, (\Theta+\Delta_u)^{-1}\right) \end{aligned}$$

where Δ_u is randomly sampled such that $\Theta + \Delta_u$ has support different but *close* to Θ.

Note that this problem is easier than the estimating the support of distributions, since the null distribution and alternative distribution has different support.

Goals. In both of the problems 1 and 2 of above, we are interested in polynomial (and even quasi-polynomial or sub-exponential) time algorithms using number of samples $\mathrm{poly}(d, p, \kappa^{-1})$. As will be seen more in detail in section 3, $p^{O(d)}$-time algorithms exist for such estimation problems with polynomial sample complexity. Moreover, if we assume that the matrix Θ is well-conditioned, then $\mathrm{poly}(p, d)$ time algorithms exists. Further if the matrix is *walk-summable* or *attractive*, then $\mathrm{poly}(p, d)$ time algorithms also exists [KKMM20]. However, it remains open to settle the following major question.

Question 1.5. Can we find $p^{o(d)}$-time algorithm using $\mathrm{poly}(d, p, 1/\kappa)$ samples for general precision matrix Θ or can we prove lower bound against it?

Although we aim to answer this question, it seems to be beyond reach currently. In this work, we mainly analyze the well-conditioned matrices, and thus cannot answer this question. However, some of our LDLR analysis applies for general precision matrices and can be helpful for future work. Further we answer the following different question:

Question 1.6. Is it possible to improve the $O(\log p/\kappa^2)$ sample complexity in the GGM distinguishing problem, if the precision matrix of null distribution is any well-conditioned matrix?

The previous information-theoretic lower bound is derived for $\Theta_0 = \mathrm{Id}_p$ and Δ_u containing two non-zero entries uniformly sampled from $[p] \times [p]$. Our result indicates that under similar choices of Δ_u, the $O(\log p/\kappa^2)$ samples are statistically necessary for all well-conditioned precision matrices, if Δ_u has only two non-zero non-diagonal entries and u.a.r sampled. Thus the previous algorithm result should be tight for any well conditioned sparse precision matrix. This means that for the hypothesis testing problem of Gaussian Graphical Model with well-conditioned precision matrix, there is no information-computation gap.

1.4 Organization and summary of results

The paper is organized as follows: in section 2, we describe the techniques we will employ in the proofs of our results. We start from the general description of likelihood ratio to then specialize to low degree likelihood ratio (LDLR) both in the single and multi sample cases, thereby outlining the connection between LDLR and information-computation gap. Section 3 is dedicated to the understanding of how previous literature has faced the problem of finding efficient algorithms for the GGM problem, or attempt to prove lower bounds to show its nonexistence. The crux of our work lies in section 4, where we first reformulate the GGM distinguishing problem in terms of the low degree likelihood ratio. We, then, prove the following results on the use of multi-sample LDLR for the GGM distinguishing problem, which we state as informal theorems below.

Theorem 1.7 (Theorem 4.5, *informal*)**.** *If we consider a null and alternative hypotheses, each parametrized by distinct positive semidefinite matrices Θ, Θ' that are well-conditioned, then,* *whenever the number of samples is $\omega(1/\kappa^2)$ (even much smaller than the information threshold), then the low degree likelihood ratio will always diverge.*

This suggests that LDLR may not give us any insight into the GGM distinguishing problem, in that null and alternative are too far apart from each other and discerning which is which becomes easy. The two distributions need to be close to each other, which indeed justifies next theorem.

Theorem 1.8 (Theorem 4.9, *informal*)**.** *If we consider a null hypothesis parametrized by the identity matrix, i.e. where all samples are independently drawn from a multivariate standard Gaussian, and an alternative that is parametrized by a slightly perturbed identity matrix, then the distinguishing task between the two distributions becomes hard when distinguishers use less than $O(\log p/\kappa^2)$ samples (the information threshold).*

The above theorem should be interpreted as a result that retrieves and confirms that we need at least $O(\log p/\kappa^2)$ samples to perform the distinguishing and excludes all algorithms that have a lower sample complexity.

The next two results are the main ones in this work: the first complements the (informal) theorem above, in that it excludes a large class of matrices from predicting an information-computation gap. As a matter fact, we know there are efficient algorithms for the GGM distinguishing/estimation problem with such matrices; nevertheless, the low degree model detects a bounded LDLR which would imply hardness: a contradiction.

Theorem 1.9 (Theorem 4.10, *informal*)**.** *If we consider a null hypothesis parametrized by a well-conditioned matrix, and an alternative that is parametrized by a slightly perturbed well-conditioned matrix, then the distinguishing task between the two distributions becomes hard when distinguishers use less than $O(\log p/\kappa^2)$ samples (the information threshold).*

If we look at the proof of the theorem a bit more closely, we will notice that the matrix does not need to be well-conditioned but a wider class of matrices cannot predict an information-computation gap via the LDLR model. The other (complementary) result essentially states that, for the same distinguishing problem and same order of sample size, the LDLR diverges for degree $d \in \Theta(\log p)$.

Theorem 1.10 (Theorem 5.1, *informal*)**.** *If we consider a null hypothesis parametrized by a well-conditioned matrix, and an alternative that is parametrized by a slightly perturbed well-conditioned matrix, then when number of samples is $\Omega(\log p/\kappa^2)$, for large enough polynomial degree $\Theta(\log p)$, the LDLR becomes $p^{\Omega(1)}$.*

In conclusion, section 6 summarizes our findings giving an outlook to possible future direction employing LDLR as a main workhorse or Sum-of-Squares proofs as another possible (and more powerful) proof system to use.

2 Techniques

The workhorse to formally understand whether the distinguishing between the null hypothesis H_0 and the alternatives H_u can be done efficiently will be the Low Degree Likelihood Ratio Test (LDLR for short). Before delving into the LDLR description and analysis, it is essential to understand why the plain and famous Neyman-Pearson Likelihood Ratio (LR for short) is not suitable for a large class of testing problems. We start with some useful definitions to then state and prove Neyman-Pearson's Testing Lemma.

Definition 2.1. In the $\mathcal{L}^2$-space of functions, define for distribution D

- $\langle f, g \rangle_D := \mathbb{E}_{x \sim D} f(x) g(x)$ represents the inner product of two functions f, g. [1]
- $\|f\| := \sqrt{\langle f, f \rangle_D}$ represents the norm of function f.
- $f^{\leqslant d}$ represents the projection of f onto the span of functions whose coordinate-degree is at most d. We will canonically refer to such functions as d-simple [Hop18].

Let us consider a null distribution $D_\emptyset$ and a set of alternative distributions (succinctly denoted as) D_u. Neyman-Pearson's Testing Lemma[2] states that there exists no better testing function than the Likelihood Ratio indicator $\psi(x)$ in the $\mathcal{L}^2$-space of functions and if we allow unbounded computation time. Hereby, we denote the likelihood ratio $\bar{D}_u(x) := \frac{D_u(x)}{D_\emptyset(x)}$. Given such optimality, one may wonder whether we can distinguish between null and alternative distribution with high probability by only calculating $\bar{D}_u(x)$. The following definition, proposition and lemma will be the way to do so in an unbounded computation sense.

Definition 2.2 (Definition 1.11 in [KWB19])**.** A sequence of distributions $D_\emptyset$ is contiguous to another sequence of distributions D_u if, whenever $D_\emptyset^{(n)}(A_n) \to 0$, $D_u^{(n)}(A_n) \to 0$, for $n \to \infty$. Hereby, A_n is a sequence of events belonging to the support of the distributions.[3]

Proposition 2.3 (Proposition 1.12 in [KWB19])**.** *If two sequences of distributions $D_\emptyset$, D_u are contiguous, then they are statistically indistinguishable, i.e. there exists no testing function ψ such that both the probability of type I error 1.2 and type II error 1.3 simultaneously to* 0*.*

Lemma 2.4 (Lemma 1.13 in [KWB19])**.** *If, for* $\left\|\bar{D}_u^{(n)}\right\| := \mathbb{E}_{x \sim D_\emptyset^{(n)}} \left(\bar{D}_u^{(n)}(x)\right)^2$,

$$\limsup_{n \to \infty} \left\|\bar{D}_u^{(n)}\right\| < \infty$$

then, $D_\emptyset$ and D_u are contiguous.[4]

[1] In the proceedings, the inner product will be computed with respect to $D_\emptyset$.

[2] Appendix A.1 in [KWB19]

[3] By this notation, we mean that the sequence of probability measures (given by the distributions) for this sequence of events tends to 0 as n increases.

[4] The notation is $\bar{D}_u^{(n)} := D_u^{(n)} / D_\emptyset^{(n)}$.

The above lemma, hence, gives us a way to argue about (in)distinguishability through a Second-Moment Method computation on the likelihood ratio: if the latter remains bounded, then null and alternative distributions remain contiguous. Despite $\bar{D}_u$'s proven optimality, for many testing as well as estimation problems computing $\bar{D}_u(x)$ may be intractable. As a matter of fact, here by bounded contiguity of distributions, we mean that no high probability efficient testing algorithm exists and, hence, Neyman-Pearson is not truly applicable.

2.1 Low Degree Likelihood Ratio Test

We would therefore like to find a suitable testing function $\psi^{\leqslant d}$ of coordinate degree at most d that can distinguish between H_0 and H_u in time $p^{\tilde{O}(d)}$, where p is the dimensionality of the samples, i.e. each sample in the subsequent work will always be $x_i \in \mathbb{R}^p$.

Theorem 2.5 (Single-Sample Low Degree Likelihood Ratio)**.** *Let us consider a null distribution $D_\emptyset$ and a set of alternative distributions (succinctly denoted as) D_u. The testing function minimizing the expected probability of outputting the wrong hypothesis is*

$$\bar{D}_u^{\leqslant d} - 1 \in \arg \max_{\psi \text{ } d\text{-simple}} \mathop{\mathbb{E}}_{x \sim D_\emptyset} \psi(x) \tag{2.1}$$

Therefore,

$$\max_{\psi \text{ } d\text{-simple}} \mathop{\mathbb{E}}_{x \sim D_u} \psi(x) = \left\| \bar{D}_u^{\leqslant d} - 1 \right\|$$

This statement can be viewed as a bounded computation version of lemma 2.4 and its proof also follows essentially the same steps, which we leave to the reader.

On another note, from definition 2.1 above, denoting by $D^{\otimes k}$ the joint distribution of k independent samples from D, we immediately have that,

$$\left\langle f^{\otimes k}, g^{\otimes k} \right\rangle_{D^{\otimes k}} = \left\langle f, g \right\rangle_D^k$$

As we are dealing with joint distributions $D^{\otimes k}$, let us introduce a notion of degree for a testing function that takes more than one sample at a time.

Definition 2.6 (Samplewise degree [BBH+20])**.** Testing function $\psi : (\mathbb{R}^p)^{\otimes m} \to \mathbb{R}$ has samplewise degree (d, k) if $\psi(x_1, \ldots, x_m)$ can be expressed as a linear combination of functions whose coordinate degree is at most d in each x_i and degree nonzero in at most k of the m x_i's.

Following definition 2.1, $\psi^{\leqslant d,k}$ represents the projection of ψ onto the span of functions whose samplewise degree is at most (d, k). We can now extend the (implicit) definition of single-sample LDLR coming from theorem 2.5 to a multi-sample regime.

Definition 2.7 (Multi-Sample Low Degree Likelihood Ratio [BBH+20])**.** Let us consider a null distribution $D_\emptyset$ and a set of alternative distributions (succinctly denoted as) D_u. Define the m-sample (d,k)-low degree likelihood ratio test function as

$$\mathop{\mathbb{E}}_{u\sim S}\left(\bar{D}_u^{\otimes m}-1\right)^{\leqslant d,k}=\mathop{\mathbb{E}}_{u\sim S}\left(\bar{D}_u^{\otimes m}\right)^{\leqslant d,k}-1 \tag{2.2}$$

This can essentially be seen as the projection of the m-sample likelihood ratio $\mathbb{E}_{u\sim S}\bar{D}_u^{\otimes m}$ onto the span of functions whose samplewise degree is at most (d,k).

We now turn to prove a general theorem about the multi-sample LDLR that will turn out to be useful in the (non-)establishment of an information-computation gap, as explained more in depth in subsection 2.2.

Theorem 2.8 (Multi-Sample LDLR from Statistical Dimension [BBH+20], *reformulated*)**.** *Let us consider the testing problem H_0-vs.-H_u, $u\sim\mathcal{S}$, with respective distributions $D_\emptyset$, D_u and relative distributions $\bar{D}_\emptyset$, $\bar{D}_u$. If*

$$\mathop{\mathbb{E}}_{u,v\sim\mathcal{S}}\left(\langle\bar{D}_u,\bar{D}_v\rangle-1\right)^{\Omega(k)}\leqslant\frac{1}{m^{\Omega(k)}}$$

Then,

$$\left\|\mathop{\mathbb{E}}_{u\sim S}\left(\bar{D}_u^{\otimes m}\right)^{\leqslant d,\Omega(k)}-1\right\|\leqslant O(1)$$

Proof. Our proof strategy will be that of finding a constant upper bound on the (∞,k) samplewise degree LDLR, to then conclude that the multi-sample LDLR is also bounded by a constant. In particular, we know by assumption that, for any $t\leqslant k/2$,

$$\mathop{\mathbb{E}}_{u,v\sim\mathcal{S}}\left(\langle\bar{D}_u,\bar{D}_v\rangle-1\right)^{t}\leqslant\frac{1}{m^{\Omega(t)}}$$

Therefore,

$$\begin{aligned}\mathop{\mathbb{E}}_{u,v\sim S}\left\langle\left(\bar{D}_u^{\otimes m}\right)^{\leqslant\infty,k/2},\left(\bar{D}_v^{\otimes m}\right)^{\leqslant\infty,k/2}\right\rangle&=\sum_{t=1}^{k/2}\binom{m}{t}\mathop{\mathbb{E}}_{u,v\sim\mathcal{S}}\left(\langle\bar{D}_u,\bar{D}_v\rangle-1\right)^t\\&\leqslant\sum_{t=1}^{k/2}\left(\frac{me}{t}\right)^t\cdot O\left(\frac{1}{m^t}\right)\\&\leqslant\sum_{t=1}^{k/2}O\left(\frac{e}{t}\right)^t\\&\leqslant O(1)\end{aligned}$$

The first equality derives from a canonical decomposition of the tensor product into components[5], while the first inequality comes from the well-known fact $\binom{m}{t} \leqslant (me/t)^t$ as $t^t \simeq t!$ (by Stirling's formula). We immediately have that, for all $d \in \mathbb{N}$ (hence, $d < \infty$),

$$\left\| \left(\mathop{\mathbb{E}}_{u\sim S} \bar{D}_u^{\otimes m} \right)^{\leqslant d,k/2} - 1 \right\| \leqslant \mathop{\mathbb{E}}_{u,v\sim S} \left\langle \left(\bar{D}_u^{\otimes m} \right)^{\leqslant \infty,k/2}, \left(\bar{D}_v^{\otimes m} \right)^{\leqslant \infty,k/2} \right\rangle \leqslant O(1)$$

This concludes the proof of the theorem. □

2.2 Information-Computation Gap

In Figure 1, we have seen a pictorial representation of the information-computation gap that we, hereby, make more formal.

Definition 2.9 (Information-Computation Gap)**.** For distinguishing problem H_0-vs.-H_u, we say that an information-computation gap arises if

$$m_{\mathsf{STAT}}(H_0, H_u) \ll m_{\mathsf{COMP}}(H_0, H_u)$$

Hereby, $m_{\mathsf{STAT}}(H_0, H_u)$ denotes sample complexity of needed for any statistical testing procedure, and $m_{\mathsf{COMP}}(H_0, H_u)$ denotes the sample complexity needed for polynomial-time testing algorithms. We call $m_{\mathsf{STAT}}(H_0, H_u)$ as *statistical threshold* and $m_{\mathsf{COMP}}(H_0, H_u)$ as *information threshold*.

It turns out that the LDLR defined in the subsection 2.1 is a powerful tool for studying the separation between statistical threshold and information threshold. In essence, we would like to say that if the norm of such a ratio is bounded by some constant, then it becomes intractable to distinguish between null and alternative hypotheses under such sample complexity. On the other hand, if this ratio tends to infinity (i.e. it is superconstant), then it indicates that polynomial time algorithm exists for the distinguishing problem. This is, in words, the formulation of the "Hopkins's conjecture" [Hop18], which we refer to as a "conjecture" for ease of understanding.

Conjecture 2.10 (Hopkins [Hop18, BBH⁺20], *informal*)**.** *Let us consider two distributions* [6] $D_\emptyset$ *and* D_u*, for* $u \sim \mathcal{S}$*. Then, for ,*

1. ***Single sample.*** *If the LDLR* $\left\| \mathbb{E}_{u\sim S} \left(\bar{D}_u\right)^{\leqslant d} - 1 \right\| \leqslant O(1)$ *for any* $d \in \Omega(\log p)$*, then there exist no* $p^{\tilde{O}(d)}$*-time algorithm that can distinguish* H_0 *from* H_u *with* $1 - o(1)$ *probability. Otherwise, if for some* $d \in \Omega(\log p)$*,* $\left\| \mathbb{E}_{u\sim S} \left(\bar{D}_u\right)^{\leqslant d} - 1 \right\| \geqslant \omega(1)$*, then such an algorithm exists.*

[5] Claim 3.3 in [BBH⁺20]

[6] Not all distributions work in the original formulation of the conjecture: please refer to section 4.2.4 in [KWB19] to have a complete understanding of their properties.

2. ***Multi-sample.*** *If the LDLR* $\left\|\mathbb{E}_{u\sim S}\left(\bar{D}_u^{\otimes m}\right)^{\leqslant d,k} - 1\right\| \leqslant O(1)$ *for any* $d \in \Omega(\log p)$*, then there exist no* $p^{\tilde{O}(d)}$*-time algorithm that can distinguish* H_0 *from* H_u *with* $1 - o(1)$ *probability. Otherwise, if for some* $d \in \Omega(\log p)$, $\left\|\mathbb{E}_{u\sim S}\left(\bar{D}_u^{\otimes m}\right)^{\leqslant d,k} - 1\right\| \geqslant \omega(1)$*, then such an algorithm exists.*

Remark 2.11. Note that when $k \to \infty$, the thresholding test done via multi-sample LDLR becomes close to Neyman-Pearson test, and its consequences are, thus, results about the information threshold. On the other hand, when $k \in O(\log p)$ or $k \in O(1)$, then the result indicates a refutation of statistical tests based on thresholding degree $O(k)$ polynomial, and concerns the computational threshold.

Although the conjecture has been refuted, it actually holds for a large class of high-dimensional estimation-distinguishing problems and the quantity it studies is arguably easy to handle and compute.

Remark 2.12. As a remark to the conjecture above, it has been observed that, for some practical applications, $d \in \omega(1)$ suffices.[7]

We only miss the way to link the (potential) sample complexity gap to the (potential) boundedness of the LDLR. The procedure can be outlined as follows:

```
Meta-procedure: Information-Computation Gap General Procedure
  Result: Information-Computation Gap / No Information-Computation Gap
  1. Fix $m$ to be $m_{\mathsf{STAT}}$, the information threshold;
  2. Determine the value of $k$ such that $\left\|\mathbb{E}_{u\sim S}\left(\bar{D}_u^{\otimes m}\right)^{\leqslant d,\Omega(k)} - 1\right\| \leqslant O(1)$;
  3. if $k \in \omega(\log p)$: output Information-Computation Gap;
     else:                output No Information-Computation Gap;
```

We have followed [BBH+20] in our steps: the LDLR we consider in definition 2.7 does not restrict the degree in each sample, it truly restricts the number of samples used in each polynomial (sample degree). In essence, if this sample degree needs to be large, then it also provides evidence for computational hardness.

3 Related Work

This section is dedicated to previous work done in the context of Gaussian Graphical Models. We present the most important results that have been hitherto established, presenting some of the crucial theorems (relative proof sketches and notation) that will be useful in subsequent arguments of this work. In particular, in subsection 3.1, we review how to obtain an information theoretic lower bound on the sample complexity for the GGM problem

[7] Conjecture 4.6 in [KWB19].

[WWR10]. Afterwards, in subsections 3.2 and 3.3, the authors of [Wai19a, KKMM20] show that polynomial time distinguishing/estimation algorithms exist for well-conditioned as well as attractive and walk-summable matrices. In the end, in subsection 3.4, the authors of [BBH$^+$20] demonstrate a first (unsuccessful) attempt to predict an information-computation gap for the GGM problem, which yet sheds light on various possible approaches to prove or disprove the existence of such a gap.

3.1 GGM: Information Theoretic Lower Bound

In order to understand what the attainable point of reference is, regardless of its computational complexity, we need to establish a necessary condition on the minimum number samples needed in order to perform the Gaussian Graphical Model distiguishing task or, similarly, the precision matrix estimation problem. The authors of [WWR10] proved the following theorem.

Theorem 3.1 (GGM Information-theoretic Lower Bound, Theorem 1 in [WWR10])**.** *Let us consider the family of* 0*-mean GGM problems* $\mathcal{G}(p, d, \kappa)$*. Then, the hypotheses distinguishing and the precision matrix estimation tasks as per problem 1.4 require a minimum number of samples (to be successful with high probability)*

$$m_{STAT} \geqslant \left\{ \frac{\log\binom{p-d}{2} - 1}{4\kappa^2}, \frac{\log\binom{p}{d} - 1}{\frac{1}{2}\left(\log\left(1 + \frac{d\kappa}{1-\kappa}\right) - \frac{d\kappa}{1+(d-1)\kappa}\right)} \right\} \tag{3.1}$$

Proof sketch (section IV-B in [WWR10]). For the first bound, let us construct the following restricted ensembles family of graphical models: this defines a graph composed of one clique over a set of 2 vertices $\mathcal{I} \subseteq [p]$, and a second clique over a disjoint set of d vertices $\mathcal{J} \subseteq [p]$.

(Edge set) $\mathcal{E} := \left\{(i, j) \in [p] \times [p] \mid i, j \in \mathcal{I} \text{ or } i, j \in \mathcal{J},\ |\mathcal{I}| = 2,\ |\mathcal{J}| = d\right\}$

(Precision matrix) $\Theta := \mathrm{Id}_p + \vartheta\, \mathbf{1}_{\mathcal{I}}\, \mathbf{1}_{\mathcal{I}}^\top + \vartheta\, \mathbf{1}_{\mathcal{J}}\, \mathbf{1}_{\mathcal{J}}^\top$, for some parameter $\vartheta \geqslant 0$.

For the second bound, let us construct another restricted ensembles family of graphical models: in turn, such a construction defines a graph composed of a clique over a set of d vertices $\mathcal{I} \subseteq [p]$.

(Edge set) $\mathcal{E} := \left\{(i, j) \in [p] \times [p] \mid i, j \in \mathcal{I},\ |\mathcal{I}| = d\right\}$

(Precision matrix) $\Theta := \mathrm{Id}_p + \vartheta\, \mathbf{1}_{\mathcal{I}}\, \mathbf{1}_{\mathcal{I}}^\top$, for some parameter $\vartheta \geqslant 0$.

In both cases, we apply Fano's inequality together with entropy-based bounds (chapter 15 in [Wai19b]) to obtain the sharp minimax (i.e. information-theoretic) lower bound in the theorem statement. □

Observation 3.2. We observe that we could express the bound of 3.1 in Landau order as follows because $d \ll p$ (we are in the sparse setting)

$$m_{\mathsf{STAT}} \in \Omega\left(\frac{\log p}{\kappa^2}\right) \tag{3.2}$$

This is the lower bound we will use throughout the paper.

3.2 GGM: Efficient algorithms for incoherent precision matrices

The authors of [FHT08] have designed an algorithm that estimates a Gaussian Graphical Model through an ℓ_1 convex relaxation log-determinant program.

```
Algorithm: Graphical-Lasso [FHT08]
  Result: Reconstructed precision matrix $\hat{\Theta}$
  Input: $x_1, \ldots, x_m \in \mathbb{R}^p$
  1. Let $\hat{\Sigma} = \frac{1}{m} \sum_{i \in [m]} x_i x_i^\top$ be the sample covariance matrix;
  2. Fix the regularization parameter $\lambda_m \in O\left(\alpha^{-1}\sqrt{\frac{\log p}{m}}\right)$, for $\alpha$ defined in 3.3;
  3. Solve the following log-determinant program:
        $$\hat{\Theta} \in \arg\min_{\Theta \succcurlyeq \mathbf{0}} \mathbf{tr}(\Theta\hat{\Sigma}) - \log \mathbf{det}(\Theta) + \lambda_m \cdot \sum_{i \neq j} |\Theta_{ij}|$$
     where we could think of the last term as an $\ell_1$-regularization term applied on the
     off-diagonal entries of the sought precision matrix;
  4. Output $\hat{\Theta}$;
```

In [RWRY08], the authors prove that `Graphical-Lasso` 2 is indeed also an efficient algorithm. Nevertheless, this efficiency (both in terms of sample complexity and computational complexity) heavily depends its incoherence parameter that weighs the influence of irrelevant variables on relevant ones and is defined as follows.

Definition 3.3. The incoherence parameter α of precision matrix Θ for GGM problem $\mathcal{G}(p, d, \kappa)$ is a nonnegative real value satisfying

$$\max_{e \in T^c} \left\| \Gamma^*_{eT} \left(\Gamma^*_{TT}\right)^{-1} \right\|_1 \leqslant 1 - \alpha \tag{3.3}$$

Hereby, $T := E \cup \{(j, j) \mid j \in [p]\}$ is the set of row/column indices associated with edges in the graph as well as self-edges: clearly, $T^c = ([p] \times [p]) \setminus T$. Moreover, $\Gamma^* := \nabla^2 \mathcal{L}_m(\Theta) = \Theta^{-1} \otimes \Theta^{-1} \in \mathbb{R}^{p^2 \times p^2}$, where $\mathcal{L}_m(\Theta) := \mathbf{tr}(\Theta\hat{\Sigma}) - \log \mathbf{det}(\Theta)$ is the loss defined in Graphical Lasso (excluding regularization) and $\otimes$ is the Kronecker product between two matrices.

With this definition at hand, we in fact get the following bound on the number of samples needed to perform the distinguishing/estimation tasks.

Theorem 3.4 (GGMs with Incoherent Precision Matrices, Proposition 11.10 in [Wai19b])**.** *Let us consider the family of* 0*-mean GGM problems* $\mathcal{G}(p, d, \kappa)$ *with an* α*-incoherent precision matrix* Θ*. Then,* `Graphical-Lasso` *2 requires for the hypotheses distinguishing and the precision matrix estimation tasks 1.4 a number of samples (to be successful with high probability)*

$$m \in O\left(\frac{d^2 \log p}{\alpha^2}\right) \simeq m_{STAT} \tag{3.4}$$

Moreover, `Graphical-Lasso` *runs in* $p^{O(d)}$ *time in the worst case.*

This theorem statement should not lead us astray: in fact, we would be tempted to conclude that since Graphical Lasso is an efficient algorithm and theorem 3.4 entails no information-computation gap, then, in general there exists no information-computation gap for GGMs. However, this is only true when the precision matrix Θ satisfies the incoherence condition with parameter α, a strong assumption that generally does not hold.

3.3 GGM: Efficient algorithms for attractive and walk-summable precision matrices

The results presented in [KKMM20] that we are interested in are essentially two: they prove that no information-computation gap arises in the case of attractive or walk-summable precision matrices.

Attractive and walk-summable matrices. Let us first give the definitions of these two families of matrices before explicitly mentioning the theorems that establish the sample complexity for the GGM precision matrix distinguishing/estimation problem.

Definition 3.5 (Attractive matrices, Definition 10 in [KKMM20])**.** A matrix Θ is said to be attractive if, for all $i, j \in [p] \times [p]$, such that $i \neq j$, we have that $\Theta_{ij} \leqslant 0$.

Following this definition, algorithm `GreedyAndPrune` perform uses `OMP` (Orthogonal Matching Pursuit [TG07]) to learn a candidate neighborhood of a given node: this subroutine is essentially a greedy forward selection method employed to minimize conditional variance. It then prunes away all the non-neighbours from this candidate neighbourhood and obtains and estimate of the whole graph representation Θ.

Theorem 3.6 (GGMs with Attractive Precision Matrices, Theorem 7 in [KKMM20])**.** *Let us consider the family of* 0*-mean GGM problems* $\mathcal{G}(p, d, \kappa)$ *with attractive precision matrix* Θ*. Then,* `GreedyAndPrune` *requires for the hypotheses distinguishing and the precision matrix estimation tasks 1.4 a number of samples (to be successful with high probability)*

$$m \in O\left(\frac{\log p}{\kappa^2}\right) \simeq m_{STAT} \tag{3.5}$$

Moreover, `GreedyAndPrune` runs in $O\left(p^{d+1}\right)$ time in the worst case.

The above theorem essentially states that no information-computation gap arises for attractive precision matrices. The authors prove a very similar result for walk-summable precision matrices.

Definition 3.7 (Walk-summable matrices, Definition 4 in [KKMM20])**.** A matrix Θ is said to be walk-summable (symmetric diagonally dominant) if, for all $i \in [p]$, we have that $\Theta_{ii} \geqslant \sum_{j \neq i} \left|\Theta_{ij}\right|$.

We remark that the above truly defines a special case of walk-summable matrices (described more fully in Definition 3 of [KKMM20]), i.e. the SDD ones, but the theorem below holds for all walk-summable models.

Theorem 3.8 (GGMs with Walk-summable Precision Matrices, Theorem 17 in [KKMM20])**.** *Let us consider the family of 0-mean GGM problems $\mathcal{G}(p, d, \kappa)$ with walk-summable precision matrix Θ. Then, `GreedyAndPrune` requires for the hypotheses distinguishing and the precision matrix estimation tasks 1.4 a number of samples (to be successful with high probability)*

$$m \in O\left(\frac{\log p}{\kappa^6}\right) > m_{STAT} \tag{3.6}$$

As before, `GreedyAndPrune` runs in $O\left(p^{d+1}\right)$ time in the worst case.

Note the κ^6 rather than κ^2 factor above: this slightly worse dependence could make us think that an information-computation gap exists, but it is likely that it does not and that such a dependence derives from an artifact of their analysis. Indeed, the authors also design `Hybrid-MB`, an algorithm that achieves a sample complexity of $O\left(\log p / \kappa^4\right)$. In conclusion, let us recall that our goal was that of finding an instance of precision matrix Θ such that an information-computation gap arises (if, at all, existent). From what above, we know for sure that we cannot search for them in the realm of attractive nor walk-summable matrices.

3.4 GGM: A first attempt to find an IC Gap via Statistical Dimension

The authors of [BBH+20] develop a whole new framework to prediction information-computation gaps based on the concept of *statistical query complexity* and its relation to low-degree polynomials. This framework has proved powerful in prediction such gaps for a variety of problems but it has not yet been enough to predict its presence in the case of GGMs. A first unsuccessful attempt uses the connection between the k-th moment of the inner product between null and alternative relative distributions and LDLR. Their self-explanatory choice of null distribution is one parametrized by the identity matrix. The alternative distribution is a planted version of the null, where the planting is the adjacency matrix of a random signed d-regular graph on s vertices which are sampled uniformly at random from the entire set of vertices $[p]$. The rationale of this choice is that the two hypotheses will be close to each other for small enough s, being reminiscent of the *planted clique* problem [AB09].

Theorem 3.9 (GGMs with planted alternative distribution, Lemma 8.31 and Corollary 8.32 in [BBH+20])**.** *Let us consider the family of* 0*-mean GGM problems* $\mathcal{G}(p, d, \kappa)$ *where*

$$\begin{aligned} H_0 &: \ D_\emptyset = \mathcal{N}\left(\mathbf{0}, \mathrm{Id}_p\right) \\ H_u &: \ D_u = \mathcal{N}\left(\mathbf{0}, \left(\mathrm{Id}_p + \kappa\Delta_u\right)^{-1}\right) \end{aligned}$$

where Δ_u *is the adjacency matrix of a random signed d-regular graph on s vertices chosen u.a.r. from* $[p]$*. The hypotheses distinguishing problem 1.4 has*

$$\mathop{\mathbb{E}}_{u,v\sim\mathcal{S}} \left\langle \bar{D}_u, \bar{D}_v \right\rangle^k \leqslant \left(1 + \left(\frac{s^2}{p}\right)^{1/k} \cdot \left(\exp\left(\frac{1}{2} s d\kappa^2\right) - 1\right)\right)^k \tag{3.7}$$

Therefore, by theorem 2.8, we immediately get

$$\left\| \mathop{\mathbb{E}}_{u\sim S} (\bar{D}_u^{\otimes m})^{\leqslant t, k/2} - 1 \right\| \leqslant O(1) \tag{3.8}$$

for some t and so long as $d \ll s \ll p$, $\kappa \leqslant O\left(\frac{1}{\sqrt{d}}\right)$ *and* $s d\kappa^2 \leqslant O\left(k \log m\right)$*.*

We prefer not to give a proof of this theorem as it resembles our proofs in the results section and instead focus on the consequences of such a statement. As a matter of fact, the bounds proved above by the authors are not sharp enough to predict an information-computation gap for the GGM distinguishing problem. Indeed, an easy consequence of the above conclusion is that the number of samples we are allowed to take can be as large as $m \leqslant \frac{1}{2}\left(\frac{p}{s^2}\right)^{1/k} \cdot \frac{1}{\exp\left(\frac{1}{2} s d\kappa^2\right) - 1}$. Let us now think about setting $m = m_{\mathsf{STAT}} \in O\left(\log p/\kappa^2\right)$, $s \in O\left(\log n\right)$ and κ small enough, to see which kind of polynomial degree k (and subsequent $p^{O(k)}$ SDP algorithm) we are able to exclude. By the fact that $s d\kappa^2 \leqslant O\left(k \log m\right)$, we get

$$d\kappa^2 \log p \lesssim k \log\left(\frac{\log p}{\kappa^2}\right) \implies \frac{d\kappa^2 \log p}{\log\log p - 2\log\kappa} \lesssim k$$

Recalling that $\kappa \in O\left(1/\sqrt{d}\right)$, the last expression becomes

$$k \gtrsim \frac{\log p}{\log d}$$

This means that any degree-$O\left(\log p/\log d\right)$ polynomial distinguisher can be directly ruled out, but we would need to exclude all degree-$O\left(d \log p/\log d\right)$ and/or all degree-$O\left(\log p/\kappa \log d\right)$ polynomial distinguishers for an information-computation gap to arise.

Despite the inability of predicting on information-computation gap, the author's techniques are extremely useful and interesting, which is why we will make extensive use of them throughout this paper. In particular, we will generalize the search of an alternative hypothesis by using a general "null" matrix Θ (and not the identity Id_p) and use a more general planting strategy to obtain various families of "alternatives".

4 Proofs of results

We begin this section with susbsection 4.1 where we reformulate the GGM distinguishing problem in terms of the low-degree likelihood ratio, while proving a multitude of lemmas and subsequent corollaries that will be helpful in the proofs of subsections 4.2 and 4.4.

4.1 LDLR Reformulation for GGMs

We are now able to reformulate the GGM problem introduced in section 1.3 in terms of low-degree likelihood ratio. Before going into mathematical detail, let us provide some intuition on the reformulation and review the crucial points we need to address.

Let us fix the sample complexity as $m_{\mathsf{STAT}} \in O\left(\log p/\kappa^2\right)$ (refer to subsection 3.1 for a more detailed derivation), which is the number of samples needed information theoretically to be able to recover the support of precision matrix Θ. Our goal is to give evidence that under such sample complexity, the distinguishing task is hard in the low degree sense. This will be the evidence that information-computation gap exists. Alternatively, we can restrict to polynomial time algorithms under the low degree model, and see whether we can get a strictly larger sample complexity lower bound than the information threshold $\log p/\kappa^2$.

Our strategy will indeed be that of finding precision matrix such that the sample degree k needs to be of the order of $\log p/\kappa \log d$ or $d \log p/\log d$, both of which are much larger than the information theoretic threshold. For the parameters, if κ is small, then this will increase difficulty on distinguishing the alternative distribution against a Gaussian distribution with "null" covariance matrix Θ. If the number of samples we obtain is smaller, then this will also increase computational hardness.

Problem 4.1 (GGM Distinguishing Problem, *LDLR reformulation*)**.** Let $d \ll p$ and κ be defined as above, the LDLR reformulation of the κ non-degenerate d-sparse p-dimensional GGM problem is finding a precision matrix Θ and a planting matrix Δ_u to prove

$$\left\| \mathop{\mathbb{E}}_{u\sim S} (\bar{D}_u^{\otimes m})^{\leqslant d,k/2} - 1 \right\| \leqslant O(1) \tag{4.1}$$

If no such matrices exist, then prove that for all "null" and "alternative" precision matrices Θ and Θ_u,

$$\left\| \mathop{\mathbb{E}}_{u\sim S} (\bar{D}_u^{\otimes m})^{\leqslant d,k/2} - 1 \right\| \geqslant \omega(1) \tag{4.2}$$

Clearly, as the distributions come from $H_0 : D_\emptyset = \mathcal{N}\big(\mathbf{0}, \Theta^{-1}\big)$ and $H_u : D_u = \mathcal{N}\Big(\mathbf{0}, (\Theta + \Delta_u)^{-1}\Big)$, this formulation is exactly equivalent to the one in 1.4, where the computational model we utilize is the low degree one.

We would now like to rewrite general procedure 1 for Gaussian Graphical Models distinguishing problem. Before doing so however, let us prove the following easy lemma (as well as two direct corollaries) that will come in handy in the rewriting as well as the remainder of the work.

Lemma 4.2. *Let us consider $D_\emptyset = \mathcal{N}\big(\mathbf{0}, \Theta^{-1}\big)$, $D_u = \mathcal{N}\Big(\mathbf{0}, (\Theta + U)^{-1}\Big)$ and $D_v = \mathcal{N}\Big(\mathbf{0}, (\Theta + V)^{-1}\Big)$, for $U, V \in \mathbb{R}^{p \times p}$ symmetric real matrices and $\Theta + U + V \succ \mathbf{0}$, $\Theta + U \succ \mathbf{0}$, $\Theta + V \succeq \mathbf{0}$. We have that, for relative densities $\bar{D}_u, \bar{D}_v$,*

$$\big\langle \bar{D}_u, \bar{D}_v \big\rangle = \frac{1}{\sqrt{\boldsymbol{det}\Big(\mathrm{Id}_p - (\Theta + U)^{-1} U V (\Theta + V)^{-1}\Big)}}$$

Proof. Let us compute the inner product in $\mathbb{R}^p$ for Gaussian densities.

$$\begin{aligned}
\big\langle \bar{D}_u, \bar{D}_v \big\rangle &= \frac{1}{\sqrt{(2\pi)^p \mathbf{det}(\Theta) \cdot \mathbf{det}(\Theta + U)^{-1} \cdot \mathbf{det}(\Theta + V)^{-1}}} \cdot \int_{\mathbb{R}^p} \exp\left(-\frac{1}{2} x^\top (\Theta + U + V) x\right) dx \\
&= \sqrt{\frac{\mathbf{det}(\Theta + U + V)^{-1}}{\mathbf{det}(\Theta) \cdot \mathbf{det}(\Theta + U)^{-1} \cdot \mathbf{det}(\Theta + V)^{-1}}} \\
&= \sqrt{\frac{1}{\mathbf{det}(\Theta) \cdot \mathbf{det}(\Theta + U + V) \cdot \mathbf{det}(\Theta + U)^{-1} \cdot \mathbf{det}(\Theta + V)^{-1}}} \\
&= \sqrt{\frac{1}{\mathbf{det}(\Theta) \cdot \mathbf{det}\Big(\Theta^{-1}((\Theta + U)(\Theta + V) - UV)(\Theta + V)^{-1}(\Theta + U)^{-1}\Big)}} \\
&= \frac{1}{\sqrt{\mathbf{det}(\Theta) \cdot \mathbf{det}\Big(\Theta^{-1}\Big(\mathrm{Id}_p - (\Theta + U)^{-1} U V (\Theta + V)^{-1}\Big)\Big)}} \\
&= \sqrt{\frac{\mathbf{det}(\Theta)}{\mathbf{det}(\Theta) \cdot \mathbf{det}\Big(\mathrm{Id}_p - (\Theta + U)^{-1} U V (\Theta + V)^{-1}\Big)}} \\
&= \frac{1}{\sqrt{\mathbf{det}\Big(\mathrm{Id}_p - (\Theta + U)^{-1} U V (\Theta + V)^{-1}\Big)}}
\end{aligned}$$

The first equality comes from the usual integration of a multivariate Gaussian density with covariance matrix $(\Theta + U + V)^{-1}$. The last steps of this chain of equalities derive from

$\mathbf{det}\, X^{-1} = \mathbf{det}^{-1}\, X$, $\mathbf{det}\, XY = \mathbf{det}\, X \cdot \mathbf{det}\, Y$ and $\Theta + U + V = \Theta^{-1}((\Theta + U)(\Theta + V) - UV)$. This concludes the proof of the lemma. □

Corollary 4.3. *(Claim D.1 in [BBH+20]) In the setting of Lemma 4.2, we get that for* $\Theta = \mathrm{Id}_p$,

$$\langle \bar{D}_u, \bar{D}_v \rangle = \frac{1}{\sqrt{\boldsymbol{det}\left(\mathrm{Id}_p - \left(\mathrm{Id}_p + U\right)^{-1} UV \left(\mathrm{Id}_p + V\right)^{-1}\right)}}$$

For the second corollary of lemma 4.2, we defer its statement and proof to subsection 4.2 (as corollary 4.7), where it will be more natural to introduce it given the choices of null and alternative distributions.

GGM Procedure. We are now able to specialize the meta-procedure in 1 to GGMs. Below, we will need to assume (or impose in our subsequent derivation) not only that $p \gg d$ but also that $\kappa \in \omega\left(\sqrt{\log d / d}\right)$. This assumption is not far-fetched since $\kappa \overset{d\to\infty}{\longrightarrow} 0$, which is exactly one of our requirements for the non-degeneracy parameter.

```
Procedure: GGMs Information-Computation Gap
  Result: Information-Computation Gap / No Information-Computation Gap
  1. Fix $m = \log p / \kappa^2$, the information threshold;
  2. Determine the value of $k$ such that $\left\| \mathbb{E}_{u\sim S}\left(\bar{D}_u^{\otimes m}\right)^{\leqslant d, \Omega(k)} - 1 \right\| \leqslant O(1)$;
  3. if $k \in O\left(\frac{d \log p}{\log d}\right) \gg \frac{\log p}{\kappa^2}$: output Information-Computation Gap;
     else:                                output No Information-Computation Gap;
```

We are now ready to use this reformulation of the GGM distinguishing problem to prove general results about LDLR.

4.2 Choice of null and alternative distributions

The aim of the remaining part of this work will be that of identifying class of matrices such that it is impossible to predict an information-computation gap. We first prove a general fact about the low degree likelihood ratio, which resembles theorem 2.8 and, then, use it for our purposes.

Fact 4.4. *Let us consider two distributions* $D_\emptyset$ *and* D_u*, for* $u \sim \mathcal{S}$*. We have that*

$$\left\| \mathop{\mathbb{E}}_{u\sim S}\left(\bar{D}_u^{\otimes m}\right)^{\leqslant \infty, k/2} - 1 \right\|^2 = \sum_{t=1}^{k/2} \binom{m}{t} \mathop{\mathbb{E}}_{u,v\sim\mathcal{S}} \left(\langle \bar{D}_u, \bar{D}_v \rangle - 1\right)^t$$

In particular, if for any $t > 0$, if

$$\mathop{\mathbb{E}}_{u,v\sim\mathcal{S}}\left(\langle\bar{D}_u,\bar{D}_v\rangle - 1\right)^t \leqslant \beta\cdot\eta^t$$

then,

$$\left\|\mathop{\mathbb{E}}_{u\sim S}\left(\bar{D}_u^{\otimes m}\right)^{\leqslant\infty,k/2} - 1\right\|^2 \leqslant \beta\cdot\sum_{t=1}^{k/2}\binom{m}{t}\eta^t \leqslant \beta\cdot e^{\eta m}$$

Proof. By the proof of theorem 2.8, we know that the first equality below holds:

$$\begin{aligned}\left\|\mathop{\mathbb{E}}_{u\sim S}\left(\bar{D}_u^{\otimes m}\right)^{\leqslant\infty,k/2} - 1\right\|^2 &= \sum_{t=1}^{k/2}\binom{m}{t}\mathop{\mathbb{E}}_{u,v\sim\mathcal{S}}\left(\langle\bar{D}_u,\bar{D}_v\rangle - 1\right)^t \\ &\leqslant \beta\cdot\sum_{t=1}^{k/2}\binom{m}{t}\eta^t \\ &\leqslant \beta\cdot\sum_{t=1}^{k/2}\left(\frac{me}{t}\right)^t\cdot\eta^t \\ &\leqslant \beta\cdot e^{\eta m}\end{aligned}$$

Hereby, the second inequality derives from the assumption that the inner product t^{th} moment is bounded by $\beta\cdot\eta^t$, which concludes the proof of the claim. □

This fact can be interpreted as follows: given distribution-specific parameters β, η, if $\beta\cdot\eta^t$ bounds from above the t^{th} moment of the inner product between relative distributions, then we also obtain a bounded low degree likelihood ratio.

We are now ready to argue what a natural choice for null and alternative distributions is. For what concerns the null distribution, it is naturally to choose a positive semidefinite matrix $\Theta\in\mathbb{R}^{p\times p}$ and sample vectors independently from $\mathcal{N}\left(\mathbf{0},\Theta^{-1}\right)$. For the alternative distribution, a naïve choice would that of $\mathcal{N}\left(\mathbf{0},(\Theta')^{-1}\right)$ for a different PSD matrix Θ'. However, if we consider a fixed precision matrix Θ', then the low degree likelihood ratio diverges for sample complexity much smaller than $O(\log p/\kappa^2)$, as shown by the following theorem.

Theorem 4.5 (LDLR divergence for general alternative)**.** *Consider any two positive semidefinite matrices Θ,Θ' with different support and the same diagonal entries, and suppose the non-zero entries have absolute value larger than κ. Further suppose that condition number[8] of these two matrices are bounded by $\gamma\in\Omega(1)$. Then for the hypothesis testing problem between*

- *Null distribution $D_\emptyset$: $x_1,x_2,\ldots,x_m\sim\mathcal{N}\left(\mathbf{0},\Theta^{-1}\right)$*

[8] We mean condition number of a matrix M in the canonical sense, i.e. the ratio between M's largest and smallest eigenvalue: $\gamma(M) := \lambda_{\max}(M)/\lambda_{\min}(M)$.

- *Alternative distribution D_u: $x_1, x_2, \dots, x_m \sim \mathcal{N}\big(\mathbf{0}, (\Theta')^{-1}\big)$*

we have $\mathbb{E}_{u,v\sim\mathcal{S}}\big\langle \bar{D}_u, \bar{D}_u\big\rangle^2 \in \Omega(\kappa^2/\gamma^2)$, and, thus,

$$\left\| \mathop{\mathbb{E}}_{u\sim\mathcal{S}} \big(\bar{D}_u^{\otimes m}\big)^{\leqslant\infty,\leqslant 1} \right\| \in \omega(1)$$

whenever $m \in \omega(\gamma^2/\kappa^2)$.

Proof. Without loss of generality, let us assume that the minimum diagonal entry of Θ is given by 1. Then we denote $\Delta = \Theta' - \Theta$. By the previous computation in lemma 4.2, we have

$$\begin{aligned}\left\| \mathop{\mathbb{E}}_{u\sim\mathcal{S}} \big(\bar{D}_u^{\otimes m}\big)^{\leqslant\infty,\leqslant 1} - 1 \right\|^2 &= \binom{m}{1} \cdot \mathop{\mathbb{E}}_{u\sim\mathcal{S}} \langle \bar{D}_u - 1, \bar{D}_u - 1\rangle \\ &= m \cdot \left(\frac{1}{\sqrt{\mathbf{det}\big(\mathrm{Id}_p - (\Theta')^{-1}\Delta^2\Theta^{-1}\big)}} - 1 \right)\end{aligned}$$

We have $(\Theta')^{-1}\Delta^2\Theta^{-1} = \Phi^\top\Phi$, where $\Phi = \Delta(\Theta')^{-1}$. Now, since Θ^{-1} has conditional number $\gamma \in \Omega(1)$, the minimum eigenvalue of Θ is

$$\lambda_{\min}(\Theta) \geqslant \frac{1}{\gamma}$$

On the other hand, since Δ contains an entry of absolute value at least κ, the maximum singular value of Δ is

$$\sigma_{\max}(\Delta) \geqslant \kappa$$

Therefore the maximum singular value of Φ is at least κ/γ and it follows that $\|\Phi^\top\Phi\| \geqslant \kappa^2/\gamma^2$. Now, by the Matrix-Determinant Lemma

$$\mathbf{det}(\mathrm{Id}_p - \Phi^\top\Phi) \leqslant 1 - \left\|\Phi^\top\Phi\right\| \leqslant 1 - \frac{\kappa^2}{\gamma^2}$$

Thus, we conclude

$$m \cdot \left(\frac{1}{\sqrt{\mathbf{det}\big(\mathrm{Id}_p - (\Theta')^{-1}\Delta^2\Theta^{-1}\big)}} - 1 \right) \geqslant \frac{m\kappa^2}{2\gamma^2}$$

when $m \in \omega(1/\kappa^2)$, the low degree likelihood ratio diverges. □

Remark 4.6. Let us note that if we exchange null and alternative distribution, the theorem still holds. Therefore a corollary is that the low degree thresholding test can truly solve the distinguishing problem with high probability.

This theorem tells us that we cannot take any arbitrary alternative distribution (parametrized by Θ') because, given that the variances of the resulting random variables will be amply different between null and alternative, simple thresholding techniques will make the distinguishing task easy and the LDLR will, thus, be unbounded. In this sense, such a general approach might not be interesting because it does not shed light into the power of LDLR nor does it truly explore how difficult the GGM problem is for null and alternative that are very close to each other.

We have hence established that we need to consider a family of alternative distributions and not just a general one. We are particularly interested in the case where Δ_u is a random symmetric matrix with the only two non-zero entries are sampled uniformly at random from the non-diagonal entries. The rationale of such a choice is that the support of $\Theta + \Delta_u$ and Θ has minimal difference, while the hypothesis testing problem is still strictly harder than estimating the support of Θ.

We indeed introduce a corollary that will be used in the derivation of our contributions. Here, to step back to the notation of lemma 4.2, $U = \Delta_u$, $V = \Delta_v$ are random symmetric matrices with only two non-zero entries, which are sampled uniformly at random from the non-diagonal entries.

Corollary 4.7. *In the setting of Lemma 4.2, we get that for $\Delta_u \in \mathbb{R}^{p\times p}$ being a random matrix obtained from sampling $(i,j) \overset{u.a.r.}{\sim} [p]\times[p] \setminus \{(\ell,\ell) \mid \ell \in [p]\}$, and setting $(\Delta_u)_{ij} = (\Delta_u)_{ji} = \kappa$,*

$$\mathop{\mathbb{E}}_{u,v\sim\mathcal{S}}\left(\left\langle \bar{D}_u, \bar{D}_v\right\rangle - 1\right)^k \simeq \frac{1}{p^2}\cdot \mathop{\mathbb{E}}_{\substack{i,j\sim[p]\times[p]\\ i\neq j}}\left(\frac{1}{\sqrt{\boldsymbol{det}\left(\mathrm{Id}_p - \kappa^2 xx^\top - \kappa^2 yy^\top\right)}} - 1\right)^k$$

$$+\frac{1}{p}\cdot \mathop{\mathbb{E}}_{\substack{i,j\sim[p]\times[p]\\ i\neq j}}\left(\frac{1}{\sqrt{\boldsymbol{det}\left(\mathrm{Id}_p - \kappa^2 wz^\top\right)}} - 1\right)^k$$

where we named $x := (\Theta + \Delta_u)^{-1}_i$ and $y := (\Theta + \Delta_u)^{-1}_j$, respectively the i^{th} and j^{th} column of the planted (or "alternative") covariance matrix. Moreover, w, z are defined in the proof.

Proof. Consider two matrices Δ_u, Δ_v as described in the corollary statement. We get that

- With probability $1 - \frac{2}{p(p-1)}$, we have that $\Delta_u \neq \Delta_v$ and we get various subcases: let us suppose that the indices where the entry κ is planted are i, j for Δ_u, and i', j' for Δ_v.

  1. If $i \notin \{i', j'\}$ and $j \notin \{i', j'\}$, then $\Delta_u\Delta_v = \mathbf{0}$, which happens with probability

  $$1 - \frac{p-1}{p(p-1)} - \frac{2}{p(p-1)} = \frac{p^2 - 2p + 4}{p(p-1)} \simeq 1 - \frac{1}{p}$$

  Hence, $\left\langle \bar{D}_u, \bar{D}_v\right\rangle = 1$.

2. With probability $\frac{p-2}{p(p-1)} \simeq \frac{1}{p}$[9] only one of the following holds

   (a) $i = i'$, then $(\Delta_u \Delta_v)_{j,j'} = \kappa^2$ and all the other entries are 0.

   (b) $i = j'$, then $(\Delta_u \Delta_v)_{j,i'} = \kappa^2$ and all the other entries are 0.

   (c) $j = i'$, then $(\Delta_u \Delta_v)_{i,j'} = \kappa^2$ and all the other entries are 0.

   (d) $j = j'$, then $(\Delta_u \Delta_v)_{i,i'} = \kappa^2$ and all the other entries are 0.

   Note that (a)-(b) and (c)-(d) cannot hold simultaneously as $i \neq j$ and $i' \neq j'$ by the way we sample (off-diagonally), nor can (a)-(c) and (b)-(d) for the same reason. If (a)-(d) or (b)-(c) hold simultaneously then $\Delta_u = \Delta_v$, which is the case described below. Hence,

$$\begin{aligned}\left\langle \bar{D}_u, \bar{D}_v \right\rangle &= \frac{1}{\sqrt{\mathbf{det}\Big(\mathrm{Id}_p - (\Theta + \Delta_u)^{-1} \Delta_u \Delta_v (\Theta + \Delta_v)^{-1}\Big)}} \\ &= \frac{1}{\sqrt{\mathbf{det}\Big(\mathrm{Id}_p - \kappa^2 (\Theta + \Delta_u)_j^{-1} (\Theta + \Delta_v)_{j'}^{-\top}\Big)}} \\ &= \frac{1}{\sqrt{\mathbf{det}\big(\mathrm{Id}_p - \kappa^2 w z^\top\big)}}\end{aligned}$$

   This is case (a), whereby $w := (\Theta + \Delta_u)_j^{-1}$ and $z := (\Theta + \Delta_v)_{j'}^{-1}$. For cases (b), (c), (d), we can very similarly compute the inner product and define w, z.

- With probability $\frac{2}{p(p-1)} \simeq \frac{1}{p^2}$, we have that $\Delta_u = \Delta_v$, Δ_u^2 will have entry κ^2 in positions (i, i) and (j, j), and, by lemma 4.2,

$$\begin{aligned}\left\langle \bar{D}_u, \bar{D}_v \right\rangle =& \frac{1}{\sqrt{\mathbf{det}\Big(\mathrm{Id}_p - (\Theta + \Delta_u)^{-1} \Delta_u^2 (\Theta + \Delta_u)^{-1}\Big)}} \\ =& \frac{1}{\sqrt{\mathbf{det}\Big(\mathrm{Id}_p - \kappa^2 \Big[(\Theta + \Delta_u)_i^{-1} (\Theta + \Delta_u)_i^{-\top} + (\Theta + \Delta_u)_j^{-1} (\Theta + \Delta_u)_j^{-\top}\Big]\Big)}} \\ =& \frac{1}{\sqrt{\mathbf{det}\big(\mathrm{Id}_p - \kappa^2 x x^\top - \kappa^2 y y^\top\big)}}\end{aligned}$$

Subtracting 1 and raising to the power of k concludes the proof as neither of these operations influences the probability. □

[9] The probability that for instance two row indices are equal is $1/p$. Then, we fix this row, say i, and eliminate position i in this row so that we do not sample diagonal entries. Now, the probability that $j = j'$ is $1/(p-1)$ and that $j \neq j'$ is $(p-2)/(p-1)$.

Remark 4.8. We remark that one can expand the determinant expressions in the statement of corollary 4.7 as follows.

$$\begin{aligned}
&\mathbf{det}\big(\mathrm{Id}_p - \kappa^2 xx^\top - \kappa^2 yy^\top\big) = \mathbf{det}\big(\mathrm{Id}_p - \kappa^2 xx^\top\big)\cdot\Big(1-\kappa^2 y^\top\big(\mathrm{Id}_p-\kappa^2 xx^\top\big)^{-1}y\Big)\\
&= \big(1-\kappa^2 x^\top x\big)\cdot\left(1-\kappa^2 y^\top\left(\mathrm{Id}_p+\frac{\kappa^2}{1-\kappa^2 x^\top x}xx^\top\right)y\right)\\
&= \big(1-\kappa^2 x^\top x\big)\cdot\left(1-\kappa^2 y^\top y-\frac{\kappa^4}{1-\kappa^2 x^\top x}y^\top xx^\top y\right)\\
&= 1-\kappa^2 y^\top y-\frac{\kappa^4}{1-\kappa^2 x^\top x}y^\top xx^\top y-\kappa^2 x^\top x+\kappa^4 x^\top xy^\top y+\frac{\kappa^6}{1-\kappa^2 x^\top x}x^\top xy^\top xx^\top y
\end{aligned}$$

The first equality derives from the Matrix-Determinant Lemma applied on matrix $\mathrm{Id}_p - \kappa^2 xx^\top$. The second equality derives from applying the Matrix-Determinant Lemma on the first multiplicand and Sherman-Morrison's formula on the second, and the rest follows by expanding the products. Similarly, one can derive, by simply applying the Matrix-Determinant Lemma,

$$\mathbf{det}\big(\mathrm{Id}_p - \kappa^2 wz^\top\big) = 1-\kappa^2 z^\top w$$

We leave to the reader the simplification of expression in corollary 4.7.

4.3 Information threshold from LDLR

The first result we obtain is recovering the information-theoretic lower bound $m_{\mathsf{STAT}} \in O\big(\log p/\kappa^2\big)$ by taking the sample degree from 1 to p. The following result should not be deceptive: indeed, it does not imply that there exists an information-computation gap. It only indicates that it is impossible to distinguish between the null and alternative distributions with sample complexity $o(\log p/\kappa^2)$.

Theorem 4.9 (Information threshold from LDLR)**.** *Let us consider $\Theta = \mathrm{Id}_p$ and Δ_u a random symmetric matrix with only two non-zero entries such that one of two non-zero entries in Δ_u is sampled uniformly at random from the non-diagonal entries. Then, there exists constant $c = 2$ such that for $m \leqslant c\log p/\kappa^2$, $\kappa \in o(1)$, and any $k > 0$,*

$$\left\|\mathop{\mathbb{E}}_{u\sim S}\left(\bar{D}_u^{\otimes m}\right)^{\leqslant\infty,k/2}-1\right\| \leqslant O(1)$$

Proof. By fact 4.4 above, it is sufficient to show that

$$\mathop{\mathbb{E}}_{u,v\in\mathcal{S}}\big(\langle\bar{D}_u,\bar{D}_v\rangle-1\big)^t \leqslant \beta\cdot\eta^t$$

In particular, it is sufficient to show what above for $\beta = 1/p$ and $\eta = c'\kappa^2$ (for some constant c'), since in this case we will have

$$\beta\cdot e^{\eta m} \in O(1)$$

Corollary 4.7 together with remark 4.8 implies that, for $u \neq v$,

$$\left\langle \bar{D}_u, \bar{D}_v \right\rangle - 1 = 0$$

On the other hand, for $u = v$, we have

$$\begin{aligned}\left\langle \bar{D}_u, \bar{D}_v \right\rangle - 1 &= \frac{1}{\sqrt{(1 - \kappa^2 x^\top x) \cdot \left(1 - \kappa^2 y^\top y - \frac{\kappa^4}{1 - \kappa^2 x^\top x} y^\top x x^\top y\right)}} - 1 \\ &\lesssim \frac{1}{\sqrt{1 - \kappa^2}} - 1 \\ &\simeq \kappa^2\end{aligned}$$

Here, following the usual notation, $x := (\mathrm{Id}_p + \Delta_u)_i^{-1}$ is the i^{th} column vector of matrix $(\mathrm{Id}_p + \Delta_u)^{-1}$, while $y := (\mathrm{Id}_p + \Delta_u)_j^{-1}$ is the j^{th} column vector of matrix $(\mathrm{Id}_p + \Delta_u)^{-1}$. The first inequality derives from the fact that vectors $x,\ y$ are in norm very close to 1: the maximum eigenvalue of the perturbed identity matrix inverse is close to the maximum eigenvalue of the identity matrix

$$\lambda_{\min}(\mathrm{Id}_p + \Delta_u) \simeq 1 - \kappa \Longrightarrow \lambda_{\max}\left((\mathrm{Id}_p + \Delta_u)^{-1}\right) \simeq \frac{1}{1 - \kappa}$$

The second equality (up to constant factors) holds because $\kappa \in o(1)$. Therefore we have

$$\mathbb{E}\left(\langle \bar{D}_u, \bar{D}_v \rangle - 1\right)^t \leqslant O\left(\frac{\kappa^{2t}}{p}\right)$$

Corollary 4.7 implies

$$\left\| \mathop{\mathbb{E}}_{u \sim S} \left(\bar{D}_u^{\otimes m}\right)^{\leqslant \infty, k/2} - 1 \right\|^2 \leqslant \frac{1}{p} \cdot e^{O\left(\kappa^2 m\right)} \leqslant \frac{1}{p} \cdot e^{O\left(\log p\right)} \leqslant O(1)$$

This finishes the proof. □

We know that a lower bound on the distinguishing problem has to also be a lower bound for the estimation version of the same problem. Therefore, the above result is much stronger than saying that the support of the matrix cannot be recovered exactly. In turn, such a lower bound implies the information lower bound (threshold).

Although this lower bound rules out algorithms using only $o(\log p/\kappa^2)$ samples for all sparse precision matrices, it might happen that such sample complexity is only required for small set of special matrices. Next section will argue why this is, in fact, not the case.

4.4 Well-conditioned matrices cannot predict an IC Gap via LDLR

We have ended last section by showing that no algorithm with $o(\log p/\kappa^2)$ sample complexity can perform the distinguishing task with high probability for all sparse precision matrices with null matrix being the identity.

4.4.1 Low degree bounds for well-conditioned matrices

By the merit of simple computation, it can be shown that such lower bound does hold for a large class of sparse precision matrices, including well-conditioned matrices.

Theorem 4.10. *Let us consider a PSD precision matrix Θ, with condition number γ, and Δ_u* [10] *a random symmetric matrix with only two non-zero entries such that one of two non-zero entries in Δ_u is sampled uniformly at random from the non-diagonal entries. Then, there exists constant c such that for $m \leqslant \frac{c \log p}{\gamma^2 \kappa^2}$, $\kappa \in o(1)$, for arbitrarily large $k > 0$,*

$$\left\| \mathop{\mathbb{E}}_{u \sim S} \left(\bar{D}_u^{\otimes m} \right)^{\leqslant \infty, k/2} - 1 \right\| \leqslant O(1)$$

This complements previous results, which state that there are polynomial time algorithms for well-conditioned matrices above the information threshold.

Proof. Similar to the case of Θ as identity matrix, by corollary 4.7, we have

$$\mathop{\mathbb{E}}_{u,v \sim \mathcal{S}} \left(\langle \bar{D}_u, \bar{D}_v \rangle - 1 \right)^k \simeq \frac{1}{p^2} \cdot \mathop{\mathbb{E}}_{\substack{i,j \sim [p] \times [p] \\ i \neq j}} \left(\frac{1}{\sqrt{\mathbf{det}\left(\mathrm{Id}_p - \kappa^2 x x^\top - \kappa^2 y y^\top\right)}} - 1 \right)^k$$

$$+ \frac{1}{p} \cdot \mathop{\mathbb{E}}_{\substack{i,j \sim [p] \times [p] \\ i \neq j}} \left(\frac{1}{\sqrt{\mathbf{det}\left(\mathrm{Id}_p - \kappa^2 w z^\top\right)}} - 1 \right)^k$$

Thus, we only need to bound the following two terms (using the notation on $x,\ y,\ w,\ z$)

$$\frac{1}{\sqrt{\mathbf{det}\left(\mathrm{Id}_p - \kappa^2 \Theta_{ii} \Theta_{jj} \left(\left\| (\Theta + \Delta_u)_i^{-1} \right\|^2 + \left\| (\Theta + \Delta_u)_j^{-1} \right\|^2 \right) \right)}} - 1$$

$$\frac{1}{\sqrt{\mathbf{det}\left(\mathrm{Id}_p - \kappa^2 \Theta_{ii} \Theta_{jj} (\Theta + \Delta_u)_i^{-1} ((\Theta + \Delta_u)_j^{-1})^\top \right)}} - 1$$

For the first term, we have that

$$\Theta_{ii} \Theta_{jj} \left\| (\Theta + \Delta_u)_i^{-1} \right\|^2 \leqslant \gamma^4$$

[10] We can safely assume that both the spectral norm and the condition number of matrix $\Theta + \Delta_u$ are bounded by γ insofar as if at least one of the two is not small, then the problem will become easily solvable by thresholding as argued in theorem 4.5.

Henceforth,

$$\frac{1}{\sqrt{\mathbf{det}\left(\mathrm{Id}_p - \kappa^2\Theta_{ii}\Theta_{jj}\left(\left\|(\Theta+\Delta_u)_i^{-1}\right\|^2 + \left\|(\Theta+\Delta_u)_j^{-1}\right\|^2\right)\right)}} - 1$$

$$= \frac{1}{\sqrt{1 - \kappa^2\Theta_{ii}\Theta_{jj}\left(\left\|(\Theta+\Delta_u)_i^{-1}\right\|^2 + \left\|(\Theta+\Delta_u)_j^{-1}\right\|^2\right)}} - 1 \leqslant \frac{1}{\sqrt{1-\kappa^2\gamma^4}} - 1$$

$$\leqslant O(\gamma^4\kappa^2)$$

For the second term we have

$$\Theta_{ii}\Theta_{jj}(\Theta+\Delta_u)_j^{-\top}(\Theta+\Delta_u)_i^{-1} \leqslant \tilde{\Theta}_j^{-\top}\tilde{\Theta}_i^{-1}$$

where $\tilde{\Theta} = \mathbf{diag}^{-1}\{\Theta_{ii}\}_{i\in[p]} \cdot \Theta$. By Cauchy-Schwartz inequality, $\tilde{\Theta}_j^{-\top}\tilde{\Theta}_i^{-1} \leqslant \gamma^2$ and, thus,

$$\frac{1}{\sqrt{\mathbf{det}\left(\mathrm{Id}_p - \kappa^2\Theta_{ii}\Theta_{jj}\Theta_i^{-1}(\Theta_j^{-1})^\top\right)}} - 1 = \frac{1}{\sqrt{1-\kappa^2\Theta_{ii}\Theta_{jj}(\Theta+\Delta_u)_j^{-\top}(\Theta+\Delta_u)_i^{-1}}} - 1$$

$$\leqslant O(\gamma^2\kappa^2)$$

For what concerns matrices with constant condition number γ, we can perform similar calculations as for the diagonal matrix, the only difference being that rather than the identity matrix, we use null matrix Θ. By the above,

$$\frac{1}{\sqrt{\mathbf{det}\left(\mathrm{Id}_p - \kappa^2(\Theta+\Delta_u)_i^{-1}(\Theta+\Delta_u)_i^{-\top} - \kappa^2(\Theta+\Delta_u)_j^{-1}(\Theta+\Delta_u)_j^{-\top}\right)}} - 1 \leqslant O\left(\gamma^4\kappa^2\right)$$

$$\frac{1}{\sqrt{\mathbf{det}\left(\mathrm{Id}_p - \kappa^2(\Theta+\Delta_u)_{\mathtt{idx}}^{-1}(\Theta+\Delta_v)_{\mathtt{idx}'}^{-\top}\right)}} - 1 \leqslant O\left(\gamma^2\kappa^2\right)$$

Hereby, $\mathtt{idx}$, $\mathtt{idx}'$ are indices as per cases (a)-(d) in the proof of corollary 4.7, each happening with probability 1/4. Now, let us use corollary 4.7 to obtain

$$\mathop{\mathbb{E}}_{u,v\sim\mathcal{S}}\left(\left\langle \bar{D}_u, \bar{D}_v\right\rangle - 1\right)^t \leqslant \frac{1}{p^2}\cdot O\left(\gamma^2\kappa\right)^{2t} + \frac{1}{p}\cdot O(\gamma\kappa)^{2t} \leqslant O\left(\frac{(\gamma\kappa)^{2t}}{p}\right)$$

The last inequality follows immediately from a bounded condition number $\gamma \in O(1)$ and $\kappa \in o(1)$. To conclude the proof, notice that by fact 4.4 (with $\beta = 1/p$, $\eta = (\gamma\kappa)^2$) and $m \in O\left(\log p/\kappa^2\right)$,

$$\left\|\mathop{\mathbb{E}}_{u\sim S}\left(\bar{D}_u^{\otimes m}\right)^{\leqslant\infty,k/2} - 1\right\|^2 \leqslant \frac{1}{p}\cdot e^{O\left((\gamma\kappa)^2 m\right)} \leqslant \frac{1}{p}\cdot e^{O\left(\log p\right)} \leqslant O(1)$$

where the second to last step follows from γ being constant. □

After following closely the above proof, we observe that it is also possible to relax the assumption of being well-conditioned: as a matter of fact, the proof only deals with precision matrix diagonal entries and covariance matrix column space, thereby, only requiring assumptions on them rather than on the condition number, which is much more restrictive of a condition. Hence, the above theorem rules out the existence of distinguishers with $o\big(\log p/\kappa^2\big)$ sample complexity for distributions parameterized by a large class of sparse "well-behaved" precision matrices.

4.4.2 Divergence of LDLR for well-conditioned matrices

Since the distinguishing problem we consider is weaker than the support estimation problem, the divergence of low degree likelihood ratio does not provide evidence that a polynomial time algorithm exists for well conditioned matrices. However, showing the divergence of LDLR can rule out the possibility that we can derive an unconditional lower bound using such approach.

Theorem 4.11. *Let us consider a PSD precision matrix Θ, with condition number γ, and Δ_u a random symmetric matrix with only two non-zero entries such that one of two non-zero entries in Δ_u is sampled uniformly at random from the non-diagonal entries. Then, there exists constant c such that for $m \geqslant c\gamma^2 \log p/\kappa^2$, $\kappa \in o(1)$,*

$$\left\| \mathop{\mathbb{E}}_{u\sim\mathcal{S}} \big(\bar{D}_u^{\otimes m}\big)^{\leqslant\infty,\leqslant O(\log p)} - 1 \right\| \in \omega(1)$$

Remark 4.12. Note that degree $O(\log p)$ is truly needed, since for $k \in O(1)$, the LDLR becomes bounded for any constant c.

Proof. Given the following chain of inequalities,

$$\binom{m}{k} \cdot \mathop{\mathbb{E}}_{u,v\sim\mathcal{S}} \langle \bar{D}_u, \bar{D}_v\rangle^k \geqslant \binom{m}{k} \cdot \mathbb{P}[u=v] \cdot \mathop{\mathbb{E}}_{u,v\sim\mathcal{S}} \big[\langle \bar{D}_u, \bar{D}_v\rangle^k \mid u=v\big] = \binom{m}{k} \cdot \mathop{\mathbb{E}}_{u,v\sim\mathcal{S}} \langle \bar{D}_u, \bar{D}_u\rangle^k$$

it is sufficient to show that for some even number $k \in O(\log p)$,

$$\binom{m}{k} \cdot \mathop{\mathbb{E}}_{u,v\sim\mathcal{S}} \langle \bar{D}_u, \bar{D}_u\rangle^k \in \omega(1)$$

Similar to previous computation (theorem 4.5 for reference), we have

$$\mathop{\mathbb{E}}_{u\sim\mathcal{S}} \langle \bar{D}_u, \bar{D}_u\rangle^k = \frac{1}{p^2} \cdot \left(\frac{1}{\sqrt{\mathbf{det}\big(\mathrm{Id}_p - (\Theta+\Delta_u)^{-1}\Delta_u^2(\Theta+\Delta_u)^{-1}\big)}} - 1 \right)$$

We denote the diagonal matrix of Θ as M. Then since the matrix $\Theta+\Delta_u$ has condition number γ, we have $\left\|M^{-1/2}(\Theta+\Delta_u)M^{-1/2}\right\| \leqslant \gamma$. Further by assumption there exist entry (i,j) in

$M^{-1/2}\Delta_u M^{-1/2}$ which is at least as large as κ, we immediately get $\|M^{-1/2}\Delta_u M^{-1/2}\| \geqslant \kappa$. It follows that

$$\begin{aligned}\|(\Theta+\Delta_u)^{-1}\Delta_u\| &= \left\| M^{-1/2}\Big(M^{-1/2}(\Theta+\Delta_u)M^{-1/2}\Big)^{-1}\Big(M^{-1/2}\Delta_u M^{-1/2}\Big)M^{1/2}\right\| \\ &\geqslant \frac{1}{\gamma}\left\|\Big(M^{-1/2}(\Theta+\Delta_u)M^{-1/2}\Big)^{-1}\right\| \geqslant \frac{\kappa}{\gamma^2}\end{aligned}$$

Since for any positive semidefinite matrix A with spectral norm bounded by 1, the determinant $\mathbf{det}(\mathrm{Id}_p - A) \leqslant 1 - \|A\|$, we have

$$\frac{1}{\sqrt{\mathbf{det}\big(\mathrm{Id}_p - (\Theta+\Delta_u)^{-1}\Delta_u^2(\Theta+\Delta_u)^{-1}\big)}} - 1 \geqslant \frac{\kappa^2}{4\gamma^2}$$

Substituting back, all in all we have

$$\binom{m}{k}\cdot \mathbb{E}\langle \bar{D}_u, \bar{D}_v\rangle^k \geqslant \binom{m}{k}\cdot\frac{1}{p^2}\cdot\left(\frac{\kappa^2}{4\gamma^2}\right)^k$$

Taking $k \in O(\log p)$ and $m \in \Omega(\log p/\kappa^2)$, we obtain

$$\binom{m}{k}\cdot \mathbb{E}\langle \bar{D}_u, \bar{D}_v\rangle^k \in \omega(1)$$

This concludes the proof of the theorem. □

5 Beyond well-conditioned matrices

5.1 Natural generalization from current lower bound

Although we only show that the sample complexity needed by the current polynomial time algorithms is information-theoretically optimal for well-conditioned matrices, our lower bound $O\Big(\frac{\log p}{\gamma^2\kappa^2}\Big)$ directly extends to the case where the matrix is ill-conditioned. This lower bound is worse than $O(\log p/\kappa^2)$, but is also not trivial.

5.2 Diagonal dominant precision matrices

A very typical class of ill-conditioned matrices is diagonal dominant precision matrices with small minimum eigenvalue. For such matrix, we can show that the low degree likelihood ratio diverges under the similar condition as well-conditioned matrices.

Theorem 5.1. *For $\kappa = o(1)$, let us consider a PSD precision matrix Θ with diagonals given by 1, and Δ_u a random symmetric matrix sampled from a distribution supported on $p^{\Omega(1)}$ matrices with* *non-zero entries $|\Delta_u(i,j)| \geqslant \kappa$. Further suppose that $\|\Delta_u\| \leqslant 2$. Then, there exists constant c such that for $m \geqslant c \log p/\kappa^2$, ,*

$$\left\| \mathop{\mathbb{E}}_{u\sim S} \left(\bar{D}_u^{\otimes m}\right)^{\leqslant \infty, \leqslant O(\log p)} - 1 \right\| \in \omega(1)$$

Proof. Given the following chain of inequalities,

$$\binom{m}{k} \cdot \mathop{\mathbb{E}}_{u,v\sim \mathcal{S}} \langle \bar{D}_u, \bar{D}_v \rangle^k \geqslant \binom{m}{k} \cdot \mathbb{P}[u=v] \cdot \mathop{\mathbb{E}}_{u,v\sim \mathcal{S}} \left[\langle \bar{D}_u, \bar{D}_v \rangle^k \mid u = v \right] = \binom{m}{k} \cdot \mathop{\mathbb{E}}_{u,v\sim \mathcal{S}} \langle \bar{D}_u, \bar{D}_u \rangle^k$$

it is sufficient to show that for some even number $k \in O(\log p)$,

$$\binom{m}{k} \cdot \mathop{\mathbb{E}}_{u,v\sim \mathcal{S}} \langle \bar{D}_u, \bar{D}_u \rangle^k \in \omega(1)$$

Similar to previous computation (theorem 4.5 for reference), we have

$$\mathop{\mathbb{E}}_{u\sim \mathcal{S}} \langle \bar{D}_u, \bar{D}_u \rangle^k = \frac{1}{p^{\Omega(1)}} \cdot \left(\frac{1}{\sqrt{\mathbf{det}\left(\mathrm{Id}_p - (\Theta + \Delta_u)^{-1} \Delta_u^2 (\Theta + \Delta_u)^{-1}\right)}} - 1 \right)$$

Now since Θ is diagonally dominant, we have $\|\Theta\| \leqslant 2$. For $\|\Delta_u\| \leqslant 2$, we have $\|\Theta\| \leqslant 2$ $\|\Theta + \Delta_u\| \leqslant 4$. Further since there exist entries in Δ_u are at least as large as κ, we immediately get $\|\Delta_u\| \geqslant \kappa$. It follows that

$$\|(\Theta + \Delta_u)^{-1} \Delta_u\| \geqslant \Omega(\kappa)$$

Since for any positive semidefinite matrix A with spectral norm bounded by 1, the determinant $\mathbf{det}(\mathrm{Id}_p - A) \leqslant 1 - \|A\|$, we have

$$\frac{1}{\sqrt{\mathbf{det}\left(\mathrm{Id}_p - (\Theta + \Delta_u)^{-1} \Delta_u^2 (\Theta + \Delta_u)^{-1}\right)}} - 1 \geqslant \Omega(\kappa^2)$$

Substituting back, all in all we have

$$\binom{m}{k} \cdot \mathbb{E} \langle \bar{D}_u, \bar{D}_v \rangle^k \geqslant \binom{m}{k} \cdot \frac{1}{p^2} \cdot \left(c' \kappa^2\right)^k$$

where c' is a constant. Taking $k \in O(\log p)$ and $m \in \Omega(\log p/\kappa^2)$, we obtain

$$\binom{m}{k} \cdot \mathbb{E} \langle \bar{D}_u, \bar{D}_v \rangle^k \in \omega(1)$$

This concludes the proof of the theorem. □

5.3 Future work: a harder distinguishing problem

It is likely that we cannot prove information-computation gap using the current hypothesis problem where the precision matrix is a fixed matrix. Such hypothesis testing problem can be much easier than the corresponding support estimation problem, since for rejecting the null hypothesis we only need to tell whether a single entry not in the support of Θ is in the support of precision matrix.

A possible way of strengthening the lower bound is considering null distribution where Θ also follows a random distribution. However, for such null distribution, the form of likelihood ratio will be drastically more complicated and technically hard to analyze.

6 Conclusion

In summary, we have shown that a wide family of "well-behaved" matrices for which we know efficient distinguishing/estimation algorithms exist, low degree method cannot predict an information-computation gap for the GGM problem under the multi-sample low degree polynomial model. Our results also generalize to some ill-conditioned matrices, and could be truly helpful in subsequent work.

One important direction that can be continued in future work is that of strengthening the impossibility to predict an information-computation gap to other renowned class of matrices, and provide algorithms for Gaussian Graphical Model with such precision matrices. One possible approach is to frame the GGM estimation and distinguishing problem in the language of *Sum-of-Squares proof system*, try to explore the degree of such proofs and, thereby, understanding the complexity of Gaussian Graphical Models.

Another future direction is to show that for a random distribution of ill-conditioned null matrices and our planting perturbation for the alternative, the LDLR is bounded and, thus, an information-computation gap arises.

References

[AB09] Sanjeev Arora and Boaz Barak, *Computational complexity: A modern approach*, 1st ed., Cambridge University Press, USA, 2009. 1, 2, 16

[ABARS20] Emmanuel Abbe, Enric Boix-Adserà, Peter Ralli, and Colin Sandon, *Graph powering and spectral robustness*, SIAM Journal on Mathematics of Data Science **2** (2020), no. 1, 132–157. 4

[AKS98] Noga Alon, Michael Krivelevich, and Benny Sudakov, *Finding a large hidden clique in a random graph*, Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (USA), SODA '98, Society for Industrial and Applied Mathematics, 1998, p. 594–598. 1

[BBH$^+$20] Matthew Brennan, Guy Bresler, Samuel B. Hopkins, Jerry Li, and Tselil Schramm, *Statistical query algorithms and low-degree tests are almost equivalent*, CoRR (2020). 1, 3, 4, 8, 9, 10, 11, 12, 15, 16, 19

[BHK$^+$16] Boaz Barak, Samuel B. Hopkins, Jonathan A. Kelner, Pravesh Kothari, Ankur Moitra, and Aaron Potechin, *A nearly tight sum-of-squares lower bound for the planted clique problem*, CoRR **abs/1604.03084** (2016). 4

[BS14] Boaz Barak and David Steurer, *Sum-of-squares proofs and the quest toward optimal algorithms*, CoRR **abs/1404.5236** (2014). 1

[FHT08] Jerome Friedman, Trevor Hastie, and Robert Tibshirani, *Sparse inverse covariance estimation with the graphical lasso*, Biostatistics **9** (2008), no. 3, 432–441 (en). 13

[HKP$^+$17] Samuel B. Hopkins, Pravesh K. Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer, *The power of sum-of-squares for detecting hidden structures*, CoRR **abs/1710.05017** (2017). 1

[HL18] Samuel B. Hopkins and Jerry Li, *Mixture models, robustness, and sum of squares proofs*, Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018, 2018, pp. 1021–1034. 4

[Hop18] Samuel Hopkins, *Statistical inference and the sum of squares method*, PhD thesis, Cornell University, 2018. 1, 4, 7, 10

[HW20] Justin Holmgren and Alexander S. Wein, *Counterexamples to the low-degree conjecture*, 2020. 4

[KKMM20] Jonathan A. Kelner, Frederic Koehler, Raghu Meka, and Ankur Moitra, *Learning some popular gaussian graphical models without condition number bounds*, Annual

Conference on Neural Information Processing Systems 2020, NeurIPS 2020, 2020. 4, 5, 12, 14, 15

[KWB19] Dmitriy Kunisky, Alexander S. Wein, and Afonso S. Bandeira, *Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio*, 2019. 1, 4, 7, 10, 11

[MVL20] Sidhant Misra, Marc Vuffray, and Andrey Y. Lokhov, *Information theoretic optimal learning of gaussian graphical models*, Conference on Learning Theory, 2020. 1, 2, 4

[RWRY08] Pradeep Ravikumar, Martin J. Wainwright, Garvesh Raskutti, and Bin Yu, *High-dimensional covariance estimation by minimizing ℓ_1-penalized log-determinant divergence*, 2008. 13

[SW20] Tselil Schramm and Alexander S. Wein, *Computational barriers to estimation from low-degree polynomials*, 2020. 4

[TG07] Joel A. Tropp and Anna C. Gilbert, *Signal recovery from random measurements via orthogonal matching pursuit*, IEEE Transactions on Information Theory **53** (2007), no. 12, 4655–4666. 14

[Wai19a] Martin J. Wainwright, *Graphical models for high-dimensional data*, Cambridge Series in Statistical and Probabilistic Mathematics, p. 347–382, Cambridge University Press, 2019. 2, 12

[Wai19b] ______, *High-dimensional statistics: A non-asymptotic viewpoint*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2019. 2, 12, 14

[WWR10] Wei Wang, Martin J. Wainwright, and Kannan Ramchandran, *Information-theoretic bounds on model selection for gaussian markov random fields*, 2010 IEEE International Symposium on Information Theory, 2010, pp. 1373–1377. 1, 2, 12

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Low degree polynomials and Gaussian Graphical Models

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Russo	Matteo

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
Zürich, 08.08.2021	

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.