Strong recovery of geometric planted matchings

Dmitriy Kunisky* Jonathan Niles-Weed†

Abstract

We study the problem of efficiently recovering the matching between an unlabelled collection of n points in $\mathbb{R}^d$ and a small random perturbation of those points. We consider a model where the initial points are i.i.d. standard Gaussian vectors, perturbed by adding i.i.d. Gaussian vectors with covariance $\sigma^2 \boldsymbol{I}_d$. In this setting, the maximum likelihood estimator (MLE) can be found in polynomial time as the solution of a linear assignment problem. We establish thresholds on σ^2 for the MLE to perfectly recover the planted matching (making no errors) and to strongly recover the planted matching (making $o(n)$ errors) both for d constant and $d = d(n)$ growing arbitrarily. Between these two thresholds, we show that the MLE makes $n^{\delta+o(1)}$ errors for an explicit $\delta \in (0, 1)$. These results extend a recent line of work on recovering matchings planted in random graphs with independently-weighted edges to the geometric setting. Our proof techniques rely on careful analysis of the combinatorial structure of partial matchings in large, weakly dependent random graphs using the first and second moment methods.

1 Introduction

Consider a set of n unlabelled particles $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ in $\mathbb{R}^d$ undergoing random motion. A short time later, the particles are observed at new locations $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_n\}$. Is it possible to ascertain which particles correspond to which? This problem—known as *multitarget tracking*—was proposed for theoretical analysis by [15], and has a wide range of applications in many scientific contexts where it is useful to infer the trajectories of objects from a succession of still images.

For concreteness, we formalize this question as follows: fix a dimension $d \in \mathbb{Z}_+$, a sample size $n \in \mathbb{Z}_+$, and a noise variance $\sigma^2 \in \mathbb{R}_+$. We first draw $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ independently, then draw noise vectors $\boldsymbol{z}_1, \dots, \boldsymbol{z}_n \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_d)$ independently (of one another and the $\boldsymbol{x}_i$) and set $\boldsymbol{y}_i := \boldsymbol{x}_i + \boldsymbol{z}_i$. We then draw a hidden permutation $\pi^\star \sim \mathsf{Unif}(S_n)$ and observe the tuple $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n, \boldsymbol{y}_{\pi^\star(1)}, \dots, \boldsymbol{y}_{\pi^\star(n)})$. The goal is to estimate the planted permutation $\pi^\star$ from this observation.

While this model is quite natural, rigorously analyzing its statistical and computational properties has proven challenging, chiefly because the pairwise distances $\{\|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2\}_{i,j=1}^n$ are not independent. In the interest of identifying a mathematically tractable alternative, [15] suggested to study a simpler model where independent random variables are substituted for these distances. Under this simplified model, we observe a matrix $\boldsymbol{W} \in \mathbb{R}^{n \times n}$ where, for a random hidden permutation $\pi^\star$, the entries W_{ij} are drawn from a distribution $\mathcal{P}$ when $\pi^\star(i) = j$, and another distribution $\mathcal{Q}$ otherwise, all independently.

Models of this type have attracted significant recent interest in the computer science and statistics communities, and precise results are now known in a number of different settings [18, 36, 42]. Despite this progress, however, the original problem of recovering planted *geometric* matchings to our knowledge has not received any attention since its proposal by [15].

In this work, we make progress on this original question. We precisely characterize the performance of a natural recovery procedure based on the linear assignment problem, and establish thresholds on σ^2 for this procedure to recover the planted matching with various amounts of error. Our results also suggest new conjectures about the performance of a natural online algorithm for multitarget tracking which has been proposed in the signal processing literature [12, 40, 41]. Taken as a whole, our results indicate regimes in which it is possible to recover geometric planted matchings to high accuracy in polynomial time.

[*] Department of Computer Science, Yale University. Email: *dmitriy.kunisky@yale.edu*. Partially supported by NSF grants DMS-1712730 and DMS-1719545. Much of this work was done while DK was at New York University.

[†] Department of Mathematics, Courant Institute of Mathematical Sciences, New York University. Email: *jnw@cims.nyu.edu*. Partially supported by NSF grant DMS-2015291.

We will focus on the *maximum likelihood estimator (MLE)* of $\pi^\star$ from the observations, which is given by

$$\widehat{\pi} := \operatorname*{arg\,max}_{\pi \in S_n} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{y}_{\pi(i)}\|^2 \right) = \operatorname*{arg\,min}_{\pi \in S_n} \sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{y}_{\pi(i)}\|^2. \tag{1.1}$$

One advantage of this estimator is that it does not depend on the variance σ^2, which may not be known in practice. Crucially, despite being given as the solution to an optimization problem over S_n, the estimator can be computed in polynomial time, since it is an instance of the *linear assignment problem*. Solutions may therefore be computed efficiently either by an exact relaxation to a linear program over doubly stochastic matrices, or with specialized combinatorial algorithms such as the Hungarian algorithm [11, 28].

We note that though the MLE is a canonical choice of estimator, it is not the only available polynomial-time approach. Another natural approach is to estimate $\pi^\star$ by greedily matching each point $\boldsymbol{x}_i$ to its nearest neighbor. One can show that this algorithm is competitive with the MLE in some regimes, but is strictly dominated by the MLE when the dimension is large. We discuss this algorithm and a similar greedy algorithm which seeks to maximize the correlation between $\boldsymbol{x}_i$ and its matched point in Appendix A.

We assess the error incurred by the MLE by counting how many indices of $[n]$ it matches incorrectly. We define the (random) set of such errors,

$$\mathcal{E} = \{ i \in [n] : \widehat{\pi}(i) \neq \pi^\star(i) \}. \tag{1.2}$$

We will primarily be concerned with the behavior of the random variable $|\mathcal{E}|$. Its law is unchanged by fixing $\pi^\star$, so we assume without loss of generality that $\pi^\star$ is the identity permutation. We lastly introduce some standard jargon. We say $\widehat{\pi}$ achieves *strong recovery* (of $\pi^\star$) if $|\mathcal{E}| = o(n)$, achieves *perfect recovery* if $|\mathcal{E}| = 0$, and achieves *near-perfect recovery* or *sublinear error* if $0 < |\mathcal{E}| \leq o(n)$. In contrast, we say $\widehat{\pi}$ makes a *macroscopic* number of errors if $|\mathcal{E}| = \Omega(n)$.

Most prior work on planted matching problems has focused on establishing when strong recovery is or is not achieved. We will partly address this question, but we will also study the *polynomial error rate* given by $\frac{\log(1 \vee |\mathcal{E}|)}{\log n}$. As we show below in Section 1.3, this finer control is valuable in applications to multitarget tracking over time. We identify this rate by analyzing the cycle decomposition of $\widehat{\pi}$ and counting the associated *augmenting cycles* of length greater than one; in particular, much of our analysis depends on a precise analysis of the number of augmenting 2-cycles, which we interpret as forming a random graph on $[n]$. We give a further overview of our proof techniques in Section 1.4 below.

The limits of recovering planted matchings under independent weights are increasingly well understood. These models exhibit a phase transition in the recoverability of $\pi^\star$, which was conjectured by [15], proved in a special case by [36], and studied in greater detail and generality by [18, 42]. The approach of [36] in particular may be viewed as an extension to the planted setting of an earlier line of work studying optimal matchings under i.i.d. weights, the so-called *random assignment model* [3, 4, 34, 39]. Despite the sophistication of these results, their techniques rely heavily on the independence assumption, and many of their conclusions remain conjectural in the geometric matching setting.

More broadly, various problems of estimating combinatorial structures from noisy observations have received much attention in recent years. As in our case, the models making strong independence assumptions have been the most amenable to analysis; notable examples include the stochastic block model [1, 17, 37] and the planted clique model [5, 10, 26], both of which may be viewed as models of *community detection* in networks. One of the remarkable phenomena that such models exhibit is the *statistical-to-computational gap*, where in a range of model parameters it is possible to estimate the planted object, but (conjecturally) only with prohibitively costly algorithms (see, e.g., [9]). There is not yet evidence that planted matching problems ever have such gaps, but it is an interesting open question to determine if this in fact ever occurs. We note also that the difference between independent planted matching models and our geometric planted matching model is analogous to the difference between the stochastic block model of network community structure and the stochastic ball model [8, 25] and similar Gaussian mixture models [32, 35] analyzed more recently in the community detection literature.

Finally, the question of optimally matching i.i.d. random points is a classical topic in probability theory and computational geometry [2, 6, 14, 29, 30, 31, 44, 45, 46]. This line of work studies a natural *null model* counterpart to ours, where all $2n$ points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$ are i.i.d. This model is the geometric analogue of the random assignment problem, and it would be interesting to understand whether the optimal transport techniques developed for analyzing matchings of i.i.d. points (such as the PDE approach of [6, 14]) can be imported to the study of geometric planted matching models, in the same way that [36] imported the techniques of [3, 4] related to local weak convergence from the random assignment problem to their independent planted matching model.

The remainder of the paper is organized as follows. In this section we present our main results, a stylized application to multitarget tracking over time, an overview of our proof techniques, and several open questions. In Section 2 we present some preliminary technical tools. In the remaining two sections we prove our main results, which constitute upper and lower bounds with high probability on $|\mathcal{E}|$. In Section 3 we prove upper bounds using the first moment method, and in Section 4 we prove lower bounds using the second moment method.

1.1 Notation Throughout, we focus on the $n \to \infty$ limit and let $d = d(n)$ and $\sigma^2 = \sigma^2(n)$ scale at various rates with n. The asymptotic symbols $o(\cdot), O(\cdot), \omega(\cdot), \Omega(\cdot), \Theta(\cdot), \ll, \sim$, and $\gg$ will have their usual meanings with reference to the limit $n \to \infty$, subscripts such as $O_a(\cdot)$ indicate that the implicit constant depends on the quantity a, and events which occur with probability $1 - o(1)$ are said to hold "with high probability."

We also introduce some further notation for the MLE. We define two *cost matrices* $\boldsymbol{W}^{(0)}, \boldsymbol{W} \in \mathbb{R}^{n\times n}$ with entries

$$W_{ij}^{(0)} := \|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2, \tag{1.3}$$

$$W_{ij} := \langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle, \tag{1.4}$$

and note that, writing $\boldsymbol{P}_\pi$ for the permutation matrix of a permutation π, the MLE is equivalently

$$\widehat{\pi} = \underset{\pi\in S_n}{\arg\min}\langle \boldsymbol{W}^{(0)}, \boldsymbol{P}_\pi\rangle = \underset{\pi\in S_n}{\arg\max}\langle \boldsymbol{W}, \boldsymbol{P}_\pi\rangle, \tag{1.5}$$

since, upon expanding the squared distances, each $\|\boldsymbol{x}_i\|^2$ and $\|\boldsymbol{y}_j\|^2$ occurs exactly once for any π.

For $a, b \in \mathbb{R}$, we write $a \vee b$ for the maximum of a and b and $a \wedge b$ for their minimum. Given $x > 0$, we let $\log_+(x) := 0 \vee \log(x)$.

1.2 Main Results To state our results, we consider three different regimes: the low-dimensional regime where $d = o(\log n)$, the logarithmic regime where $d = \Theta(\log n)$, and the high-dimensional regime where $d = \omega(\log n)$. In each, we identify the behavior of $|\mathcal{E}|$ as a function of σ^2. As our proofs make clear, the difference between these regimes is justified by the fact that the quantity

$$\frac{d}{\log n}\log(1+\sigma^{-2})$$

plays the role of a signal-to-noise ratio for our problem, which suggests that the correct scaling of σ is $\sigma^2 = \Theta(n^{-\xi/d})$ for some $\xi > 0$ in the low-dimensional regime, $\sigma^2 = \Theta(1)$ in the logarithmic regime, and $\sigma^2 = \Theta(\frac{d}{\log n})$ in the high-dimensional regime. Our main results verify these claims.

In the low-dimensional regime, we are able to resolve the thresholds between perfect recovery, strong recovery, and macroscopic error.

THEOREM 1.1. (LOW-DIMENSIONAL REGIME) *Suppose that $d = o(\log n)$.*

1. *(Perfect recovery) If $\sigma^2 = o(n^{-4/d})$, then $|\mathcal{E}| = 0$ with high probability.*
2. *(Small error) If $\sigma^2 = \Theta(n^{-4/d})$, then $\mathbb{E}|\mathcal{E}|$ is bounded; in particular $|\mathcal{E}| \le f(n)$ for any $f(n) = \omega(1)$ with high probability.*
3. *(Sublinear error) If $n^{-4/d} \ll \sigma^2 \ll n^{-2/d}$, then there exists an absolute constant $c > 0$ such that, for any $f(n) = \omega(1)$, with high probability*

$$\frac{c}{\sqrt{d}}\sigma^d n^2 \le |\mathcal{E}| \le f(n)\sigma^d n^2. \tag{1.6}$$

In particular, if $\frac{d}{\log n}\log(1+\sigma^{-2}) \to \xi \in [2, 4]$, then the following convergence in probability holds as $n \to \infty$:

$$\frac{\log(1 \vee |\mathcal{E}|)}{\log n} \to 2 - \frac{\xi}{2}. \tag{1.7}$$

4. *(Linear or nearly-linear error) If $\sigma^2 \geq an^{-2/d}$ for some $a > 0$, then there exists $c = c(a)$ such that $|\mathcal{E}| \geq e^{-cd}n$ with high probability.*

Note that when $\sigma^2 = \Omega(n^{-2/d})$ and d is a constant not depending on n, Theorem 1.1 implies that $|\mathcal{E}| = \Omega(n)$ with high probability; this is the only regime where we are able to show that the MLE actually incurs macroscopic error. When $1 \ll d \ll \log n$ with the same scaling of σ^2, we find the nearly macroscopic $|\mathcal{E}| = \Omega(n^{1-o(1)})$. One way to gain some intuition for this statement is to consider instead a greedy matching algorithm, that matches each $\boldsymbol{x}_i$ (in some sequence) to the nearest unmatched $\boldsymbol{y}_j$. The typical minimum distance between any two of the $\boldsymbol{x}_i$ is typically $\min_{i \neq j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\| = \Theta(n^{-2/d})$, whereby the expected perfect recovery threshold is $\sigma = \Theta(n^{-2/d})$. On the other hand, the minimum distance between any particular $\boldsymbol{x}_i$, say $\boldsymbol{x}_1$, and another $\boldsymbol{x}_j$ is typically $\min_{j \neq 1} \|\boldsymbol{x}_1 - \boldsymbol{x}_j\| = \Theta(n^{-1/d})$. Moreover, this is simultaneously achieved for most of the $\boldsymbol{x}_i$, so the expected strong recovery threshold is $\sigma = \Theta(n^{-1/d})$. We give further comparisons between the MLE and this greedy algorithm as well as another variant thereof in Appendix A.

In the logarithmic regime we obtain similar results, except that the range of σ^2 yielding sublinear errors appears to end at a point when $|\mathcal{E}| = \Theta(n^\delta)$ for some $\delta < 1$. In fact, in Conjecture 1.1 below we predict the existence of a discontinuity in the limiting value of $\frac{\log(1 \vee |\mathcal{E}|)}{\log n}$, where the error rate jumps sharply from $|\mathcal{E}| = \Theta(n^\delta)$ to $|\mathcal{E}| = \Omega(n)$.

THEOREM 1.2. (LOGARITHMIC REGIME) *Suppose that $d \sim a \log n$ for some $a > 0$, and that σ^2 is constant not depending on n.*

1. *(Perfect recovery) If*

$$\sigma^2 < \frac{1}{e^{4/a} - 1}, \tag{1.8}$$

then $|\mathcal{E}| = 0$ with high probability.

2. *(Sublinear error) If*

$$\frac{1}{e^{4/a} - 1} \leq \sigma^2 < \frac{1}{(2e^{1/a} - 1)^2 - 1}, \tag{1.9}$$

then the following convergence in probability holds:

$$\frac{\log(1 \vee |\mathcal{E}|)}{\log n} \to 2 - \frac{a}{2} \log(1 + \sigma^{-2}). \tag{1.10}$$

The quantity on the right side of (1.10) equals zero at the lower limit $\sigma^2 = \frac{1}{e^{4/a}-1}$, and equals $2 - a\log(2e^{1/a} - 1) \in (0,1)$ at the upper limit $\sigma^2 = \frac{1}{(2e^{1/a}-1)^2-1}$ for any $a > 0$. As $a \to \infty$, the width of the sublinear error regime given in (1.9) is $\frac{1}{(2e^{1/a}-1)^2-1} - \frac{1}{e^{4/a}-1} = \frac{1}{8} + o(1)$, so this is indeed a non-trivial range of σ^2 on the critical scale $\sigma^2 = \Theta(1)$.

Next, we treat the remaining high-dimensional regime. Here our results only describe perfect recovery; however, Conjecture 1.1 will again predict that on the scale of σ^2 indicated below, greater noise results in macroscopic error.

THEOREM 1.3. (HIGH-DIMENSIONAL REGIME) *Suppose that $d = \omega(\log n)$. If for some $\epsilon > 0$*

$$\sigma^2 \leq \left(\frac{1}{4} - \epsilon\right) \frac{d}{\log n}, \tag{1.11}$$

then $|\mathcal{E}| = 0$ with high probability.

Finally, we state a supplementary conjecture, which we will discuss in greater detail in Section 1.4, where we show how it is suggested by the first moment combinatorics of augmenting cycles. If true, this conjecture would complete the high-level picture described by our results, in each regime of d showing that for the remaining σ^2 not covered by our results, the MLE makes a macroscopic number of errors.

Conjecture 1.1. *Suppose that any of the following conditions holds:*

1. $1 \ll d \ll \log n$ *and, for some* $\epsilon > 0$, $\sigma^2 \geq n^{-(2-\epsilon)/d}$.
2. $d \sim a \log n$ *and, for some* $\epsilon > 0$, $\sigma^2 \geq \frac{1}{(2e^{1/a}-1)^2-1} + \epsilon$.
3. $d = \omega(\log n)$ *and, for some* $\epsilon > 0$, $\sigma^2 \geq (\frac{1}{4} + \epsilon)\frac{d}{\log n}$.

Then, for some $c = c(\epsilon) > 0$, $|\mathcal{E}| \geq cn$ *with high probability.*

If true, Conjecture 1.1 together with Theorem 1.2 would surprisingly imply a discontinuity in the value of $\frac{\log(1 \vee |\mathcal{E}|)}{\log n}$ as a function of σ^2 when $d = a \log n$ at $\sigma^2 = \frac{1}{(2e^{1/a}-1)^2-1}$: from the left this quantity would tend to a limit $2 - a\log(2e^{1/a} - 1)$ strictly smaller than 1, while from the right it would equal 1. As $a \to 0$, the size of this jump would shrink, recovering in the limit the continuous behavior of the $d \ll \log(n)$ case. We illustrate these error curves and the predicted jump in Figure 4; see also Section 1.4 for discussion of theoretical evidence for this prediction.

1.3 Stylized Application: Online Tracking of Brownian Motions As an application of our results, we consider a stylized motion tracking model, similar to the one proposed by [15]. Suppose that $\boldsymbol{x}_1(t), \dots, \boldsymbol{x}_n(t) \in \mathbb{R}^d$ are independent standard Brownian motions in dimension $d = O(1)$, started from $\boldsymbol{x}_i(0)$ independent standard Gaussian vectors. We view these Brownian motions as the evolution of indistinguishable particles, whose motion we would like to track over time: for some fixed $\delta > 0$, we observe this collection of particles (but not their labels) at times $t = k\delta$ for each integer $k \geq 0$. On the basis of these observations, we would like to track the identities of each particle over some large interval $t \in [0, T]$ as accurately as possible.

A natural approach is an iterative matching algorithm: having observed the point set $X_k = \{\boldsymbol{x}_1(k\delta), \dots, \boldsymbol{x}_n(k\delta)\}$ for each integer $k \geq 0$, repeatedly compute the MLE matching $\widehat{\pi}_k$ between X_{k-1} and X_k for $k \geq 1$. Then, the composition $\widehat{\pi} = \widehat{\pi}_1 \cdots \widehat{\pi}_K$ gives a plausible matching between X_0 and X_K, which attempts to track the Brownian motions up to time $T = K\delta$. In fact, this approach is frequently used in practical engineering applications in concert with various preprocessing and filtering pipelines [12, 40, 41]. We illustrate a small example in Figure 1. How large can we make this T while having the final matching correctly identify at least, say, half of the particles, i.e., having $\widehat{\pi}$ fix at least half of the points of $[n]$?[1] Let us define the expectation of this time,

$$T_{\max} = T_{\max}(\delta, n) := \delta \cdot \mathbb{E}\min\{K : \widehat{\pi}_1 \cdots \widehat{\pi}_K \text{ has fewer than } n/2 \text{ fixed points}\}. \tag{1.12}$$

Clearly we expect decreasing δ—taking snapshots more frequently—to increase $T_{\max}$. We can use our results for d constant to make an informal prediction as to the behavior of this tradeoff. The displacement of a Brownian motion in time δ has law $\mathcal{N}(0, \delta \boldsymbol{I}_d)$, so each time step looks like our earlier setup with $\sigma^2 = \delta$. Thus suppose $n^{-4/d} \ll \delta \ll n^{-2/d}$. Then, we expect the error incurred by $\widehat{\pi}_k$ to be roughly $\delta^{d/2}n^2$ for each k. Supposing that these errors affect different indices in each time step, we then expect to make $\Omega(n)$ errors in total once $K > n/(\delta^{d/2}n^2) = \delta^{-d/2}/n$. Thus, we expect $T_{\max} \sim \delta K = \delta^{1-d/2}/n$.

One case to which this argument certainly does *not* apply is $d = 1$: in this case, the difference between the positions of any two particles is itself a Brownian motion which will eventually cross zero (meaning that the particles will collide), and by a standard argument of time inversion of Brownian motion will in fact cross zero infinitely many times in the vicinity of any such crossing (meaning that the particles will collide infinitely many times immediately following their first collision). Indeed, we illustrate in Figure 2 below that, when $d = 1$, the error of tracking appears to be driven by such collisions and does not depend at all on the sampling interval δ. However, we conjecture that the above heuristic is sound for larger dimension.

Conjecture 1.2. *Suppose that* $d \geq 2$ *and* $\delta = n^{-\xi/d}$ *for some* $\xi \in [2, 4]$. *Then,* $T_{\max} \sim \frac{\delta^{1-d/2}}{n} f(n) = n^{\xi/2 - \xi/d - 1} f(n)$ *for some* $1/\mathsf{polylog}(n) \leq f(n) \leq \mathsf{polylog}(n)$.

[1] All manner of quantities describing the approach of $\widehat{\pi}$ to a uniformly random permutation, such as total variation distance in the style of results on Markov chain mixing times, would be interesting to consider; we restrict our discussion to the number of fixed points for the sake of simplicity.

A surprising consequence of this conjecture would be that, when $d = 2$, there is a large range of δ over which the improvement in $T_{\max}$ gained for decreasing δ is only logarithmic in δ—the situation is hardly better than $d = 1$—while once $d \geq 3$ this improvement becomes polynomial in δ. This criticality of $d = 2$ seems to resemble similar phenomena in the structure of optimal matchings of i.i.d. points in the null model [2, 29, 30, 46]. While it is difficult to make n sufficiently large to overcome finite-size effects and resolve the exponents we are interested in numerically, as alternative evidence we plot the number of errors over time for a fixed small n and various δ and d in Figure 2. We observe something qualitatively similar to the Conjecture: when $d = 2$ the error changes logarithmically over several orders of magnitude of δ, while once $d = 3$ the error changes much more rapidly, plausibly polynomially.

Proving Conjecture 1.2 would require several improvements over our current results, and represents an interesting question for future work. At a minimum, doing so would require better understanding of the concentration properties of $|\mathcal{E}|$ in the low-dimensional regime. Obtaining stronger concentration bounds would also open the door to understanding what happens when $\delta \ll n^{-4/d}$, when each time step is in our "perfect recovery" regime and most time steps do not introduce new errors.

1.4 Proof Techniques We briefly discuss our proof techniques, with the aim of giving a heuristic theoretical justification of Conjecture 1.1 above. The following is the key structural property obeyed by $\mathcal{E}$: because $\mathcal{E}$ is the set of indices not fixed by $\widehat{\pi}$, by the cycle decomposition of $\widehat{\pi}$ the indices of $\mathcal{E}$ belong to a disjoint union of cycles in $\widehat{\pi}$, and each such cycle $(i_1, \ldots, i_t)$ is *augmenting*, meaning that, performing index arithmetic modulo t,

$$\sum_{k=1}^{t} W_{i_k i_{k+1}} \geq \sum_{k=1}^{t} W_{i_k i_k}, \tag{1.13}$$

the reason being simply that the objective value of $\widehat{\pi}$ in (1.5) must not be increased by replacing any cycle of $\widehat{\pi}$ with the identity mapping.[2] Our analysis is based on considering how many augmenting cycles of various sizes on $[n]$ exist.

There are $\binom{n}{t}(t-1)! \approx n^t/t$ possible t-cycles on $[n]$ (the approximation holding for $t \ll n$), so the total "mass" or sum of the lengths of these cycles is $\approx n^t$. We show that the probability that any given cycle is augmenting is related to the Riemann sum of a particular function $f(\sigma^2, x)$, thus obtaining that

$$\mathbb{P}[t\text{-cycle is augmenting}] \leq \exp\left(-\frac{d}{2}\sum_{j=1}^{t-1} f\left(\sigma^2, \frac{j}{t}\right)\right), \tag{1.14}$$

$$\mathbb{E}[\text{mass of augmenting } t\text{-cycles}] \leq \exp\left(t \log n - \frac{d}{2}\sum_{j=1}^{t-1} f\left(\sigma^2, \frac{j}{t}\right)\right) =: n^{c(t)}. \tag{1.15}$$

We will show that these Riemann sums have a *discrete concavity* property (see Section 2.2), and that consequently $c(t)$ is a convex function of t, as we illustrate in Figure 3. The threshold that Conjecture 1.1 predicts for strong recovery is the location where $\lim_{t\to\infty} c(t)/t$ changes sign from negative to positive, i.e. where the limiting slope of the curves in Figure 3 changes from negative to positive.

When this limiting slope is negative, then in fact the entire curve of $c(t)$ is decreasing, so the dominant contribution is made by augmenting 2-cycles. In this case, we may analyze the number of errors the MLE makes by counting augmenting 2-cycles with the first and second moment methods. When the limiting slope is positive, we expect substantial contributions to be made by t-cycles with large t, which our techniques here do not handle. There is a third threshold when there are $\Omega(n)$ augmenting 2-cycles, the rightmost threshold in Figure 3, beyond which in principle our second moment method might be improved to show that the MLE makes $\Omega(n)$ errors. There are technical obstructions due to correlations in the second moment method that prevent us from carrying this out; moreover, as we emphasize in Figure 4 for the case $d = \Theta(\log n)$, we do *not* expect this analysis alone

[2] Often the term "augmenting cycle" instead refers to an even cycle alternating between rows and columns of $\boldsymbol{W}$, a cycle in the weighted bipartite graph on $2n$ vertices whose weights are given by $\boldsymbol{W}$. However, we will find it more intuitive to think of cycles as permutations on $[n]$ instead, as described here.

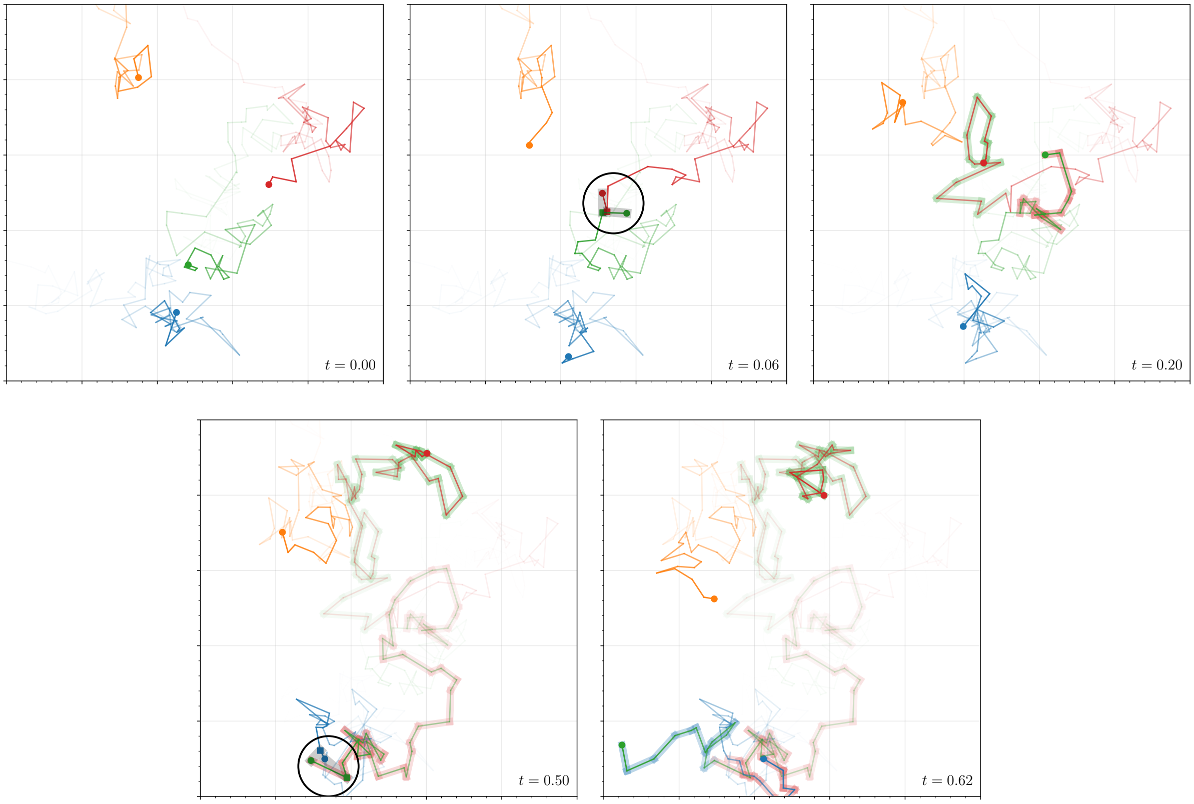

Figure 1: **Online MLE tracking of Brownian motions.** We illustrate how errors accrue in tracking particles by iteratively computing the MLE. We plot the random walks formed by snapshots of four Brownian motions in $\mathbb{R}^2$, and indicate by a circle two times when the permutation produced by the iterated MLE undergoes a transposition from the true labeling. For erroneously labelled points, we show their true label in the thin inner line, and their label by the iterated MLE in the thick outer line. If the points colored orange, red, green, and blue are respectively labelled $1, 2, 3, 4$ at the beginning, then the estimated permutation changes first to $1, 3, 2, 4$, and then to $1, 3, 4, 2$.

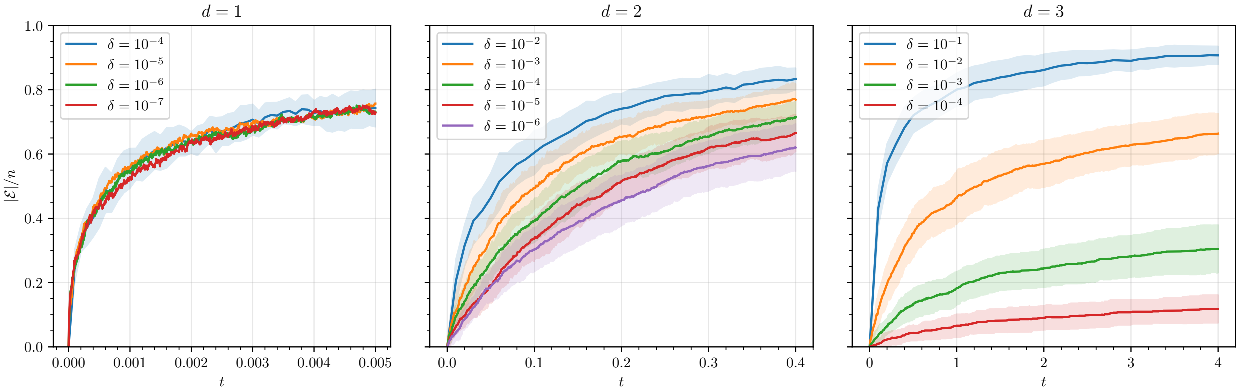

Figure 2: **Dimension-dependent error scaling of MLE tracking.** We plot the error incurred by the iterated MLE estimator over time for tracking $n = 100$ independent Brownian motions in dimensions $d = 1, 2$, and 3, illustrating the differing dependences on the sampling interval δ. Each curve plots an average of 20 independent trials and an error bar of one standard deviation.

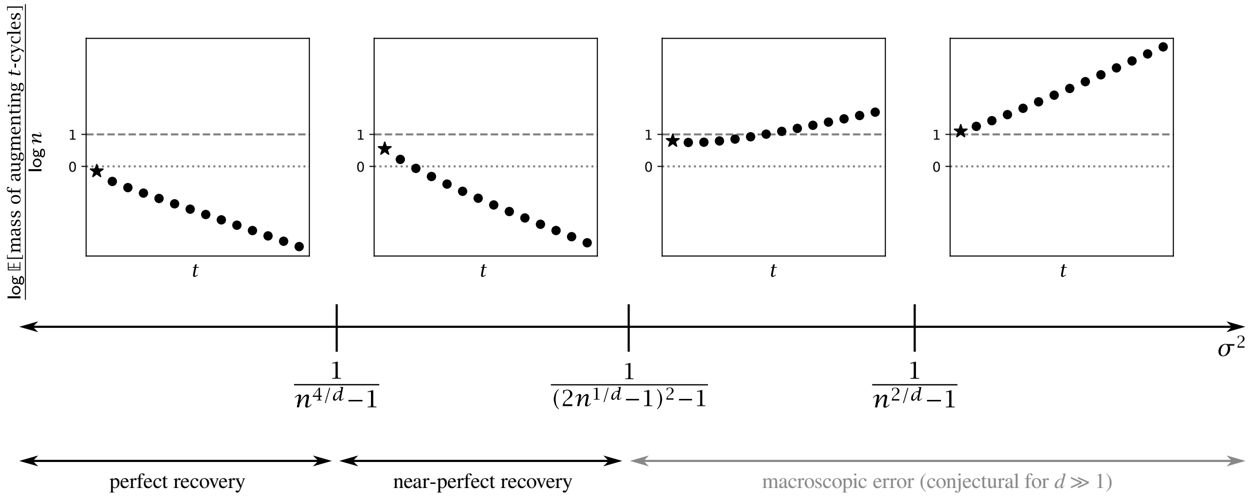
$\frac{\log \mathbb{E}[\text{mass of augmenting } t\text{-cycles}]}{\log n}$
1
0
t
t
t
t
σ^2
$\frac{1}{n^{4/d}-1}$
$\frac{1}{(2n^{1/d}-1)^2-1}$
$\frac{1}{n^{2/d}-1}$
perfect recovery
near-perfect recovery
macroscopic error (conjectural for $d \gg 1$)

Figure 3: **First moments of augmenting cycle counts.** We illustrate our results and the associated thresholds, giving a schematic illustration of the polynomial rate of growth of the total mass of augmenting cycles of various sizes in each regime of the noise parameter σ^2. Regimes marked in black are those described by our results; the one in gray is conjectural. In each plot, a star marks the point plotting the expected mass of augmenting 2-cycles, whose analysis drives our lower bounds on $|\mathcal{E}|$.

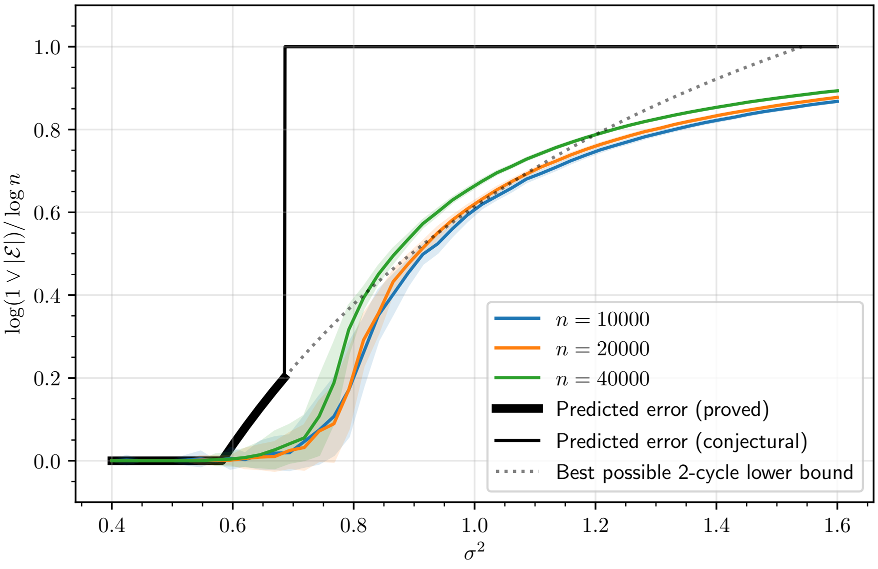

Figure 4: **Discontinuity in polynomial error rate.** We show the predicted jump in the MLE error rate when $d = a \log n$ with $a = 4$ (bold solid line from Theorem 1.2 and thin solid line from Conjecture 1.1) contrasted with the best possible lower bound that could be proved by analyzing only augmenting 2-cycles (dotted line). For increasing n, we also plot the average and one standard deviation error bars for 50 random trials of the MLE at regularly spaced σ^2. Though convergence is very slow with n, the fact that these curves cross the dotted line implies that there is non-trivial contribution to the total error from augmenting cycles of length greater than 2, supporting Conjecture 1.1 in the $d \sim \log n$ regime.

to prove the correct strong recovery threshold—for that, it appears necessary to argue the existence of larger augmenting cycles.

Finally, we remark that this latter threshold is a natural one for *greedy algorithms* that attempt to find a good matching in the matrix $\boldsymbol{W}$ row by row. In Appendix A, we show that the greedy algorithm applied to $\boldsymbol{W}$ in fact achieves strong recovery below this third threshold $\sigma^2 = \frac{1}{n^{2/d}-1}$, which is asymptotically greater than the strong recovery threshold of the MLE $\sigma^2 = \frac{1}{(2n^{1/d}-1)^2-1}$ once $d = \omega(\log n)$ (the former is $\sim \frac{1}{2}\frac{d}{\log n}$, while the latter is $\sim \frac{1}{4}\frac{d}{\log n}$). On the other hand, this algorithm fails completely for $d = o(\log n)$; by contrast, a greedy algorithm applied to $\boldsymbol{W}^{(0)}$ performs similarly to the MLE in that regime but can be worse outside the low-dimensional regime. Across all $d = d(n)$ the three algorithms are generally incomparable. We refer the reader to Appendix A for further discussion of these algorithms.

1.5 Open Questions We conclude with several open questions on the estimation of geometric planted matchings that we find promising for future research.

1. Establish the strong recovery threshold for $d \gg 1$, i.e., prove Conjecture 1.1.

2. Establish the error curve for constant dimension d: what is the function $e(a, d)$ such that, when $\sigma^2 = an^{-2/d}$, then $\mathbb{E}|\mathcal{E}|/n \to e(a, d)$?

3. Prove information-theoretic lower bounds on the $|\mathcal{E}|$ achievable by any computation, and determine in what regimes the MLE is information-theoretically optimal.

4. Are algorithms other than the MLE (including the greedy algorithms we discuss in Appendix A, algorithms computing matchings corresponding to Wasserstein distances W_p with $p \neq 2$, algorithms computing entropy-regularized relaxations of the linear assignment problem [16], and the belief propagation algorithm proposed by [15]) more effective in certain regimes of d and σ^2?

5. Establish the dimension-dependent scaling of the time for which online MLE tracking can consistently track n particles given in Conjecture 1.2, and determine what happens for small time intervals $\delta \ll n^{-4/d}$.

6. More generally, what are effective algorithms for the motion tracking application proposed in Section 1.3? Is there an offline algorithm (processing the entire set of snapshots concurrently) that is superior to the kind of online algorithm we propose?

7. What are the statistics of permutations obtained by computing optimal matchings between a collection of points and their evolution under Brownian motion for some period of time (either just once or with an iterated MLE or greedy algorithm)? How quickly do such permutations converge to the uniform distribution?

2 Preliminaries

2.1 Graph Laplacians and Spectra Given a graph $G = (V, E)$, we write $\boldsymbol{L}^G \in \mathbb{R}^{V \times V}$ for the *graph Laplacian* of G, the symmetric matrix with quadratic form

$$\boldsymbol{x}^\top \boldsymbol{L}^G \boldsymbol{x} = \sum_{\{v,w\} \in E} (x_v - x_w)^2. \tag{2.16}$$

We will particularly be interested in the path and cycle graphs. We write P_t and C_t for the path or cycle, respectively, on t vertices, where we require $t \geq 3$ for C_t to be defined. The following gives the spectra of their respective Laplacians (see, e.g., Example 8.8 for cycles and the discussion following Lemma 10.18 for paths in [38]).

PROPOSITION 2.1. *The eigenvalues of* $\boldsymbol{L}^{P_t}$ *are* $2(1 - \cos(\frac{\pi k}{t}))$ *for* $k = 0, \ldots, t-1$, *and the eigenvalues of* $\boldsymbol{L}^{C_t}$ *are* $2(1 - \cos(\frac{2\pi k}{t})) = 4\sin^2(\frac{\pi k}{t})$ *for* $k = 0, \ldots, t-1$.

2.2 Riemann Sums We have indicated in Section 1.4, and will see more precisely below, that probabilities of cycles being augmenting for the MLE give rise to expressions of the form $\operatorname{Tr}\log(1 + (4\sigma^2)^{-1}\boldsymbol{L}^{C_t})$. Per Proposition 2.1, these may in turn be viewed as Riemann sums of a certain periodic function, and the asymptotic probability of being augmenting for large cycles is therefore related to the integral of this function. Below we set some notation for these objects and present the properties of theirs that we will use.

Definition 2.1. *For any $t \geq 2$, $\sigma^2 > 0$ define*

$$f(\sigma^2, x) := \log\left(1 + \frac{1}{2\sigma^2}(1 - \cos(2\pi x))\right) = \log\left(1 + \frac{1}{\sigma^2}\sin^2(\pi x)\right), \tag{2.17}$$

$$I(\sigma^2) := \int_0^1 f(\sigma^2, x)\, dx, \tag{2.18}$$

$$S(\sigma^2, t) := \sum_{j=1}^{t-1} f\left(\sigma^2, \frac{j}{t}\right). \tag{2.19}$$

In fact, it is possible to evaluate $I(\sigma^2)$ in closed form.

Proposition 2.2. *For all $\sigma^2 > 0$,*

$$I(\sigma^2) = 2\log\left(\frac{1 + \sqrt{1 + \sigma^{-2}}}{2}\right). \tag{2.20}$$

We give the proof in Appendix B by translating the real integral to a complex contour integral.

By elementary real analysis, as $f(\sigma^2, \cdot)$ is continuous on $[0, 1]$, we have the following convergence.

Proposition 2.3. *For any $\sigma^2 > 0$, we have*

$$\lim_{t\to\infty} \frac{S(\sigma^2, t)}{t} = I(\sigma^2). \tag{2.21}$$

We will, however, need to be substantially more precise for our applications. The following are the main technical results that much of our analysis will rely on, a discrete analog of concavity for the Riemann sums of $f(\sigma^2, \cdot)$ as well as a matching opposite bound, which together allow us to formulate linear lower bounds on the $S(\sigma^2, t)$.

Lemma 2.1. (Riemann sum discrete concavity) *For $\sigma^2 > 0$, $S(\sigma^2, t) - S(\sigma^2, t-1)$ is strictly decreasing in $t \geq 3$ and approaches $I(\sigma^2)$ as $t \to \infty$. In particular, $S(\sigma^2, t) - S(\sigma^2, t-1) > I(\sigma^2)$ for all $t \geq 3$.*

Lemma 2.2. (Riemann sum upper bound) *For $t \geq 2$ and $\sigma^2 > 0$, $S(\sigma^2, t) < tI(\sigma^2)$.*

Corollary 2.1. (Riemann sum lower bound) *For all $t_0 \geq 2$ and $t > t_0$, we have*

$$S(\sigma^2, t) > S(\sigma^2, t_0) + (t - t_0)I(\sigma^2) = tI(\sigma^2) - (t_0 I(\sigma^2) - S(\sigma^2, t_0)), \tag{2.22}$$

where the constant term satisfies $t_0 I(\sigma^2) - S(\sigma^2, t_0) > 0$.

The third result follows immediately from the first two. We give the proofs of the first two results in Appendix C. The proofs rely on a combinatorial relationship between the sums $S(\sigma^2, t)$ and the *Lucas polynomials*, which solve a Fibonacci-like recurrence that allows very precise asymptotics via a polynomial-valued analogue of Binet's formula.

3 Upper Bounds and First Moment Method

3.1 Counting Augmenting Cycles To prove upper bounds on $|\mathcal{E}|$, we use the first moment method and bound $\mathbb{E}|\mathcal{E}|$ by counting the numbers of augmenting cycles of various sizes. First, we bound the probability that a cycle of a given size is augmenting.

PROPOSITION 3.1. *Let C be any fixed t-cycle in $[n]$. Then,*

$$\mathbb{P}[C \textit{ is augmenting}] \le \exp\left(-\frac{d}{2}S(\sigma^2, t)\right). \tag{3.23}$$

Proof. Without loss of generality we may suppose that $C = (1, \dots, t)$. Let us consider the cases $t = 2$ and $t \ge 3$ separately. If $t = 2$, then C is augmenting if and only if

$$W_{1,2} + W_{2,1} \ge W_{1,1} + W_{2,2}, \tag{3.24}$$

which in turn holds if and only if

$$\langle \boldsymbol{z}_1, \boldsymbol{x}_2 - \boldsymbol{x}_1\rangle + \langle \boldsymbol{z}_2, \boldsymbol{x}_1 - \boldsymbol{x}_2\rangle \ge \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2. \tag{3.25}$$

Here, conditional on the $\boldsymbol{x}_i$, the law of the left-hand side is $\mathcal{N}(0, 2\sigma^2\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2)$ since $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are i.i.d. with law $\mathcal{N}(0, \sigma^2\boldsymbol{I}_d)$. Therefore, we compute

$$\begin{aligned}\mathbb{P}[C \text{ augmenting}] &= \mathop{\mathbb{E}}_{\boldsymbol{x}_1,\boldsymbol{x}_2} \mathop{\mathbb{P}}_{g\sim\mathcal{N}(0,2\sigma^2\|\boldsymbol{x}_1-\boldsymbol{x}_2\|^2)}[g \ge \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2] \\ &= \mathop{\mathbb{E}}_{\boldsymbol{x}_1,\boldsymbol{x}_2} \mathop{\mathbb{P}}_{g\sim\mathcal{N}(0,1)}\left[g \ge \sqrt{\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2}{2\sigma^2}}\right] \\ &\le \mathop{\mathbb{E}}_{\boldsymbol{x}_1,\boldsymbol{x}_2} \exp\left(-\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2}{4\sigma^2}\right)\end{aligned}$$

To evaluate the remaining expectation, we must understand the spectrum of the quadratic form involved. Writing $\boldsymbol{x}$ for the concatenation of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, we may write $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^2 = \boldsymbol{x}^\top(\boldsymbol{L}^{P_2} \otimes \boldsymbol{I}_d)\boldsymbol{x}$, where $\boldsymbol{L}^{P_2} \in \mathbb{R}^{2\times 2}$ is the Laplacian of the path graph on two vertices, using the notation of Proposition 2.1. By the Proposition, the eigenvalues of $\boldsymbol{L}^{P_2}$ are 0 and 2. Therefore, continuing by applying an orthogonal change of basis diagonalizing the quadratic form and evaluating the χ^2 moment generating function that appears, we find

$$\begin{aligned} &= \det\left(\boldsymbol{I}_{2d} + \frac{1}{2\sigma^2}(\boldsymbol{L}^{P_2} \otimes \boldsymbol{I}_d)\right)^{-1/2} \\ &= \det\left(\boldsymbol{I}_2 + \frac{1}{2\sigma^2}\boldsymbol{L}^{P_2}\right)^{-d/2} \\ &= \left(1 + \frac{1}{\sigma^2}\right)^{-d/2} \\ &= \exp\left(-\frac{d}{2}\log\left(1 + \frac{1}{\sigma^2}\right)\right) \\ &= \exp\left(-\frac{d}{2}S(\sigma^2, 2)\right), \end{aligned} \tag{3.26}$$

as claimed.

Now, suppose $t \ge 3$. Then C is augmenting if and only if

$$W_{t,1} + \sum_{i=1}^{t-1} W_{i,i+1} \ge \sum_{i=1}^{t} W_{i,i}, \tag{3.27}$$

which in turn holds if and only if

$$\langle \boldsymbol{z}_1, \boldsymbol{x}_t - \boldsymbol{x}_1\rangle + \sum_{i=2}^{t}\langle \boldsymbol{z}_i, \boldsymbol{x}_{i-1} - \boldsymbol{x}_i\rangle \ge \frac{1}{2}\left(\|\boldsymbol{x}_t - \boldsymbol{x}_1\|_2^2 + \sum_{i=2}^{t}\|\boldsymbol{x}_{i-1} - \boldsymbol{x}_i\|_2^2\right). \tag{3.28}$$

Again, let $\boldsymbol{x}$ be the concatenation of the $\boldsymbol{x}_i$. Then, we have

$$\|\boldsymbol{x}_t - \boldsymbol{x}_1\|_2^2 + \sum_{i=2}^{t} \|\boldsymbol{x}_{i-1} - \boldsymbol{x}_i\|_2^2 = \boldsymbol{x}^\top (\boldsymbol{L}^{C_t} \otimes \boldsymbol{I}_d)\boldsymbol{x}, \tag{3.29}$$

where $\boldsymbol{L}^{C_t}$ is the Laplacian of the cycle graph C_t on t vertices. Thus the law of the left-hand side of (3.28) conditional on the $\boldsymbol{x}_i$ is $\mathcal{N}(0, \sigma^2 \boldsymbol{x}^\top (\boldsymbol{L}^{C_t} \otimes \boldsymbol{I}_d)\boldsymbol{x})$, while the right-hand side is $\frac{1}{2}\boldsymbol{x}^\top (\boldsymbol{L}^{C_t} \otimes \boldsymbol{I}_d)\boldsymbol{x}$. (We note the two differences from the case $t = 2$: the path graph is replaced by the cycle graph, and an extra factor of $\frac{1}{2}$ appears on the right-hand side.) An analogous computation to before gives

$$\begin{aligned}
\mathbb{P}[C \text{ augmenting}] &= \mathop{\mathbb{E}}_{\boldsymbol{x}_1,\dots,\boldsymbol{x}_t} \mathop{\mathbb{P}}_{g \sim \mathcal{N}(0, \sigma^2 \boldsymbol{x}^\top (\boldsymbol{L}^{C_t} \otimes \boldsymbol{I}_d)\boldsymbol{x})} \left[g \geq \frac{\boldsymbol{x}^\top (\boldsymbol{L}^{C_t} \otimes \boldsymbol{I}_d)\boldsymbol{x}}{2} \right] \\
&= \mathop{\mathbb{E}}_{\boldsymbol{x}_1,\dots,\boldsymbol{x}_t} \mathop{\mathbb{P}}_{g \sim \mathcal{N}(0,1)} \left[g \geq \sqrt{\frac{\boldsymbol{x}^\top (\boldsymbol{L}^{C_t} \otimes \boldsymbol{I}_d)\boldsymbol{x}}{4\sigma^2}} \right] \\
&\leq \mathop{\mathbb{E}}_{\boldsymbol{x}_1,\dots,\boldsymbol{x}_t} \exp\left(-\frac{\boldsymbol{x}^\top (\boldsymbol{L}^{C_t} \otimes \boldsymbol{I}_d)\boldsymbol{x}}{8\sigma^2} \right) \\
&= \det\left(\boldsymbol{I}_{dt} + \frac{1}{4\sigma^2} \boldsymbol{L}^{C_t} \otimes \boldsymbol{I}_d \right)^{-1/2} \\
&= \det\left(\boldsymbol{I}_t + \frac{1}{4\sigma^2} \boldsymbol{L}^{C_t} \right)^{-d/2}
\end{aligned}$$

and substituting in the eigenvalues of $\boldsymbol{L}$ from Proposition 2.1, we have

$$\begin{aligned}
&= \left(\prod_{j=0}^{t-1} \left\{ 1 + \frac{1}{2\sigma^2} \left(1 - \cos\left(\frac{2\pi j}{t} \right) \right) \right\} \right)^{-d/2} \\
&= \exp\left(-\frac{d}{2} \sum_{j=0}^{t-1} \log\left(1 + \frac{1}{2\sigma^2} \left(1 - \cos\left(\frac{2\pi j}{t} \right) \right) \right) \right) \\
&= \exp\left(-\frac{d}{2} S(\sigma^2, t) \right), \qquad (3.30)
\end{aligned}$$

again giving the result. ∎

COROLLARY 3.1. *For any d, n, σ^2,*

$$\mathbb{E}|\mathcal{E}| \leq \sum_{t=2}^{n} \exp\left(t \log n - \frac{d}{2} S(\sigma^2, t) \right). \tag{3.31}$$

Proof. $\mathcal{E}$ is a disjoint union of augmenting cycles, so $|\mathcal{E}|$ is at most the sum of the lengths of all augmenting cycles. The result then follows from linearity of expectation and applying that the number of t-cycles in $[n]$ is $\leq n^t/t$ and the probability bound of Proposition 3.1. ∎

With these expressions for the expected masses of augmenting cycles of various sizes in hand, we may describe more precisely why the situation presented in Figure 3 arises: the limiting exponent above as $t \to \infty$ is $\sim t \log n (1 - \frac{d}{2 \log n} I(\sigma^2))$, thus the transition around $I(\sigma^2) = 2\log(\frac{1+\sqrt{1+\sigma^{-2}}}{2}) = \frac{2 \log n}{d}$, or $\sigma^2 = \frac{1}{(2n^{1/d}-1)^2-1}$, determines whether the expected mass of large augmenting cycles diverges or not, which we conjecture is the correct strong recovery threshold. Moreover, it will turn out that when strong recovery is possible, then the dominant contribution is by augmenting 2-cycles, whose exponent is $2\log n - \frac{d}{2}S(\sigma^2, 2) = 2\log n - \frac{d}{2}\log(1+\sigma^{-2})$, and this changes sign at $\sigma^2 = \frac{1}{n^{4/d}-1}$, which is the perfect recovery threshold.

3.2 Perfect Recovery In this section we give a sufficient condition for perfect recovery, which proves Part 1 of Theorem 1.1, Part 1 of Theorem 1.2, and Theorem 1.3.

LEMMA 3.1. *Let $s_0 := 2^{1/d}$, and suppose that*

$$\sigma^2 \leq \frac{1}{s_0^{\omega(1)} n^{4/d} - 1}. \tag{3.32}$$

Then, $\mathbb{E}|\mathcal{E}| \to 0$, so, in particular, $|\mathcal{E}| = 0$ with high probability.

Before proceeding with the proof, let us indicate how this implies the claimed results for specific scalings of d. When $d \ll \log n$, then s_0 is bounded and the denominator in the bound above goes to infinity as $n \to \infty$, so the condition is satisfied whenever $\sigma^2 \ll n^{-4/d}$, giving Part 1 of Theorem 1.1.

When $d = a \log n$, then $n^{4/d} = e^{4/a}$, and there exists $f(n) = \omega(1)$ such that $s_0^{f(n)} \to 1$. Thus the condition is satisfied whenever σ^2 is bounded below $\frac{1}{e^{4/a}-1}$, giving Part 1 of Theorem 1.2.

Finally, when $d = \omega(\log n)$, then for any $\epsilon > 0$ again we may choose $f(n) = \omega(1)$ such that $s_0^{f(n)} = 2^{f(n)/d} \leq n^{\epsilon/d}$. Thus the condition is satisfied whenever $\sigma^2 \leq \frac{1}{n^{(4+\epsilon)/d}-1} \sim \frac{1}{4+\epsilon}\frac{d}{\log n}$, giving Theorem 1.3.

Proof. Rearranging the assumption on σ^2, we have

$$2 - \frac{d}{2\log n} S(\sigma^2, 2) = 2 - \frac{d \log(1+\sigma^{-2})}{2 \log n} \leq -\omega\left(\frac{d \log s_0}{\log n}\right) = -\omega\left(\frac{1}{\log n}\right). \tag{3.33}$$

Also, since by Lemma 2.2 we have $S(\sigma^2, 2) < 2I(\sigma^2)$, we further have

$$2 - \frac{d}{2\log n} S(\sigma^2, 2) > 2 - \frac{d}{2 \log n} 2I(\sigma^2) = 2\left(1 - \frac{d}{2\log n} I(\sigma^2)\right). \tag{3.34}$$

Towards bounding the exponents appearing in Corollary 3.1, we manipulate

$$\begin{aligned} t \log n - \frac{d}{2} S(\sigma^2, t) &= \log n \left(t - \frac{d}{2\log n} S(\sigma^2, t)\right) \\ \text{(by Corollary 2.1 with } t_0 = 2) \quad &\leq \log n \left(t - \frac{d}{2 \log n}(S(\sigma^2, 2) + (t-2) I(\sigma^2))\right) \\ &= \log n \left(2 - \frac{d}{2\log n} S(\sigma^2, 2) + (t-2)\left(1 - \frac{d}{2\log n} I(\sigma^2)\right)\right) \end{aligned}$$

and substituting in our bounds from above,

$$\begin{aligned} &\leq \frac{t}{2} \log n \left(2 - \frac{d}{2 \log n} S(\sigma^2, 2)\right) \\ &\leq -\omega(1) \cdot t. \qquad (3.35) \end{aligned}$$

Applying this to Corollary 3.1, we find

$$\mathbb{E}|\mathcal{E}| \leq \sum_{t=2}^{n} (e^{-\omega(1)})^t = o(1), \tag{3.36}$$

and the second result follows by Markov's inequality. □

3.3 Small Error Upper Bound We next prove a similar result to the above that gives Part 2 of Theorem 1.1 and the upper bound for the case of Part 2 of Theorem 1.2 where σ^2 takes its lower bound, $\sigma^2 = \frac{1}{e^{4/a}-1}$.

LEMMA 3.2. *Let $s_0 := 2^{1/d}$, and suppose that*

$$\sigma^2 \leq \frac{1}{s_0^{O(1)} n^{4/d} - 1}. \tag{3.37}$$

Then, $\mathbb{E}|\mathcal{E}| = O(1)$, so, in particular, for any $f(n) = \omega(1)$, we have $|\mathcal{E}| \leq f(n)$ with high probability.

The argument from the previous proof applies verbatim with $\omega(\cdot)$ replaced by $O(\cdot)$ throughout, and shows that $\mathbb{E}|\mathcal{E}| = O(1)$, whereby the result again follows by Markov's inequality.

3.4 Sublinear Error Upper Bound Finally we give an upper bound on $|\mathcal{E}|$ that holds in the sublinear error regime. This implies the upper bound of Part 3 of Theorem 1.1 and the remainder of the upper bound of Part 2 of Theorem 1.2 not covered by the previous proof.

LEMMA 3.3. *Let $s_0 := 2^{1/d}$, and suppose that*

$$\sigma^2 \leq \frac{1}{(2s_0^{\omega(1)} n^{1/d} - 1)^2 - 1}. \tag{3.38}$$

Then,

$$\mathbb{E}|\mathcal{E}| = O\left(\left(1 + \frac{1}{\sigma^2}\right)^{-d/2} n^2\right), \tag{3.39}$$

so in particular for any $f(n) = \omega(1)$ we have, with high probability,

$$|\mathcal{E}| \leq f(n)\left(1 + \frac{1}{\sigma^2}\right)^{-d/2} n^2. \tag{3.40}$$

Proof. Rearranging the assumption on σ^2, we have

$$1 - \frac{d}{2\log n} I(\sigma^2) = 1 - \frac{d}{\log n}\log\left(\frac{1+\sqrt{1+\sigma^{-2}}}{2}\right) \leq -\omega\left(\frac{1}{\log n}\right), \tag{3.41}$$

as before (the difference with the above settings being that such a bound no longer holds for $2 - \frac{d}{2\log n} S(\sigma^2, 2)$). Following the previous argument applied to Corollary 3.1, we find

$$\mathbb{E}|\mathcal{E}| \leq \sum_{t=2}^{n} n^{2 - \frac{d}{2\log n} S(\sigma^2, 2)} (e^{-\omega(1)})^{t-2} = O(n^{2 - \frac{d}{2\log n} S(\sigma^2,2)}) = O\left(\left(1 + \frac{1}{\sigma^2}\right)^{-d/2} n^2\right), \tag{3.42}$$

and the second result again follows by Markov's inequality. □

4 Lower Bounds and Second Moment Method

To prove lower bounds on $|\mathcal{E}|$, we will apply the second moment method to show that there exists a large number of vertex-disjoint augmenting 2-cycles. That is, we will study the random variable

$$M := \text{maximum number of vertex-disjoint augmenting 2-cycles in } [n]. \tag{4.43}$$

The following shows that M being large guarantees a large number of errors in the MLE.

PROPOSITION 4.1. *$|\mathcal{E}| \geq M$.*

Proof. It is impossible for (i, j) to be an augmenting transposition and to have both $\widehat{\pi}(i) = i$ and $\widehat{\pi}(j) = j$, since then π formed by composing the transposition (i, j) with $\widehat{\pi}$ would have a higher likelihood than $\widehat{\pi}$. Thus, for every pair in a maximal collection of M augmenting 2-cycles, at least one of its vertices must be labelled incorrectly by $\widehat{\pi}$, and the result follows. □

Conveniently, this quantity admits a graph-theoretic interpretation. Namely, the set of augmenting 2-cycles may be described by a graph on $[n]$:

$$G^{\mathsf{aug}} := (V = [n], E = \{\{i, j\} : (i, j) \text{ is an augmenting 2-cycle}\}). \tag{4.44}$$

With this notation, M is the size of the largest matching in this graph:

$$M = \text{number of edges in the largest matching in } G^{\mathsf{aug}}. \tag{4.45}$$

Thus our task is to show that a large matching exists in a random graph. In particular, we will want to show that there exists a matching of size $\Omega(|E| \wedge n)$, i.e., a matching of size asymptotically as large as possible subject to the basic constraints that it can exceed neither the number of vertices nor the number of edges.

There is an extensive literature on similar questions for Erdős-Rényi (ER) random graphs; however, most of these results analyze concrete algorithms for finding large matchings rather than using the second moment method [7, 20, 27, 47]. Indeed, to the best of our knowledge no previous work has tried to show the existence of large matchings in random graphs using the second moment method—perhaps thanks to the success of analyzing algorithms and to the "effectiveness" of such results, which provide an algorithm in addition to an existence proof. However, our graph G^{aug} has a more complicated dependence structure, so the second moment method is more convenient, and we draw inspiration from a line of work applying an adjusted second moment method to other extremal problems in ER random graphs, especially the chromatic number and independence number [22, 33, 43].

Remark 4.1. *When the degree of all vertices in G^{aug} is bounded with high probability by some $d_{\max}$, then algorithmic techniques do show that a large matching exists, since a greedy algorithm matching vertices arbitrarily until no more can be matched will produce a matching of at least $|E|/2d_{\max}$ edges. One may control the maximum degree in our case by appealing to the probability bounds of Proposition 4.3 for the star graph. However, this no longer applies in the critical regime where the average degree is constant (when we expect a nearly-linear number of errors in the MLE), in which case in an ER graph the largest degree is of logarithmic order, and we expect a similar behavior for G^{aug}.*

4.1 Statistics of G^{aug} We will think of G^{aug} as being well-approximated by an ER random graph, albeit with some stronger dependencies among various subgraphs. We begin by precisely describing the probability of any particular edge belonging to G^{aug}, which is the edge probability of the analogous ER graph.

Proposition 4.2. (Edge probability in G^{aug}) *Define*

$$p := \mathbb{P}[\{i,j\} \in E(G^{\mathsf{aug}})], \tag{4.46}$$

$$\widehat{p} := \frac{p}{\exp(-\frac{d}{2}S(\sigma^2, 2))}, \tag{4.47}$$

which do not depend on $i, j \in [n]$ distinct. Then, for all n, d, and $\sigma^2 \leq \frac{1}{40}d$,

$$\frac{1}{1000}\sqrt{\frac{1+\sigma^2}{d}} \leq \widehat{p} \leq 1. \tag{4.48}$$

We give the proof, an application of bounds on Gaussian Mills' ratios, in Appendix D.

Next, we control more coarsely the probability that a given graph occurs as a subgraph of G^{aug}. The following is a general parametrized bound, which relates these probabilities to Laplacians with weighted edges.

Proposition 4.3. *Suppose $G = (V, E)$ for some $V \subseteq [n]$. Let $\boldsymbol{\Delta} \in \mathbb{R}^{E \times V}$ be the edge-vertex incidence matrix for G, i.e., the matrix having non-zero entries $\Delta_{\{i,j\},k}$ only when $i = k$ or $j = k$, with one of these equaling 1 and the other equaling -1 (chosen arbitrarily) for each row index $\{i,j\} \in E$. Note that $\boldsymbol{\Delta}^\top \boldsymbol{\Delta} = \boldsymbol{L}$, the graph Laplacian. Then, for any diagonal matrix $\boldsymbol{D} \succeq \boldsymbol{0}$,*

$$\begin{aligned} \mathbb{P}[G \subseteq G^{\mathsf{aug}}] &\leq \det\left(\boldsymbol{I}_V + 2\boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta} - \sigma^2(\boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta})^2\right)^{-d/2} \\ &= \exp\left(-\frac{d}{2}\sum_{i=1}^{|V|} \log\left(1 + 2\lambda_i(\boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta}) - \sigma^2 \lambda_i(\boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta})^2\right)\right), \end{aligned} \tag{4.49}$$

where $\lambda_i(\boldsymbol{A})$ denote the eigenvalues of a symmetric matrix $\boldsymbol{A}$.

Proof. The event that $G \subseteq G^{\mathsf{aug}}$ is the same as that, for all $\{i,j\} \in E$, we have $W_{i,j} + W_{j,i} \leq W_{i,i} + W_{j,j}$. Rewriting, this is the event that, for all $\{i,j\} \in E$,

$$-\langle \boldsymbol{z}_i - \boldsymbol{z}_j, \boldsymbol{x}_i - \boldsymbol{x}_j \rangle \geq \|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2. \tag{4.50}$$

Let $\boldsymbol{X}, \boldsymbol{Z} \in \mathbb{R}^{V \times d}$ have the $\boldsymbol{x}_i$ and the $-\boldsymbol{z}_i$ as their rows, respectively. Then, the system above may be rewritten with the help of $\boldsymbol{\Delta}$ as

$$\mathsf{diag}(\boldsymbol{\Delta}\boldsymbol{Z}(\boldsymbol{\Delta}\boldsymbol{X})^\top) \geq \mathsf{diag}(\boldsymbol{\Delta}\boldsymbol{X}(\boldsymbol{\Delta}\boldsymbol{X})^\top). \tag{4.51}$$

Whenever this is true, then we also have

$$\langle \boldsymbol{D}, \boldsymbol{\Delta}\boldsymbol{Z}(\boldsymbol{\Delta}\boldsymbol{X})^\top \rangle \geq \langle \boldsymbol{D}, \boldsymbol{\Delta}\boldsymbol{X}(\boldsymbol{\Delta}\boldsymbol{X})^\top \rangle, \tag{4.52}$$

or, rewriting to isolate $\boldsymbol{Z}$,

$$\langle \boldsymbol{Z}, \boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta}\boldsymbol{X} \rangle \geq \langle \boldsymbol{X}\boldsymbol{X}^\top, \boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta} \rangle. \tag{4.53}$$

Since the entries of $\boldsymbol{Z}$ are i.i.d. with law $\mathcal{N}(0, \sigma^2)$, taking a Chernoff bound and evaluating the Gaussian moment generating function yields

$$\begin{aligned}\mathbb{P}[G \subseteq G^{\mathsf{aug}}] &\leq \mathbb{E}_{\boldsymbol{X}} \frac{\mathbb{E}_{\boldsymbol{Z}} \exp\left(\langle \boldsymbol{Z}, \boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta}\boldsymbol{X}\rangle\right)}{\exp\left(\langle \boldsymbol{X}\boldsymbol{X}^\top, \boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta}\rangle\right)} \\ &= \mathbb{E}_{\boldsymbol{X}} \exp\left(\frac{\sigma^2}{2}\|\boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta}\boldsymbol{X}\|_F^2 - \langle \boldsymbol{X}\boldsymbol{X}^\top, \boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta}\rangle\right)\end{aligned}$$

and, noting that $\|\boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta}\boldsymbol{X}\|_F^2 = \mathrm{Tr}(\boldsymbol{X}^\top(\boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta})^2\boldsymbol{X}) = \langle \boldsymbol{X}\boldsymbol{X}^\top, (\boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta})^2\rangle$, we find

$$= \mathbb{E}_{\boldsymbol{X}} \exp\left(\left\langle \boldsymbol{X}\boldsymbol{X}^\top, \frac{\sigma^2}{2}(\boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta})^2 - \boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta}\right\rangle\right)$$

and evaluating this as a χ^2 moment generating function after an orthogonal change of basis diagonalizing the matrix on the right, we obtain

$$= \det\left(\boldsymbol{I}_V + 2\boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta} - \sigma^2(\boldsymbol{\Delta}^\top \boldsymbol{D}\boldsymbol{\Delta})^2\right)^{-d/2}, \tag{4.54}$$

as claimed. ☐

It is an interesting question to optimize the choice of $\boldsymbol{D}$ in this bound. For our purposes, it suffices to use a simple version for G a path or cycle.

PROPOSITION 4.4. *For any $G = P_t$ with $t \geq 2$ or $G = C_t$ with $t \geq 3$,*

$$\mathbb{P}[G \subseteq G^{\mathsf{aug}}] \leq \exp\left(-\frac{d}{2}S(\sigma^2, t)\right). \tag{4.55}$$

In words, this shows that the probability that a path or cycle in G^{aug} on t vertices has augmenting 2-cycles for all of its edges is at most our bound (Proposition 3.1) on the probability that a cycle on t vertices is augmenting.

Proof. For $G = P_2$ the result follows from Proposition 3.1. We first note that, since P_t is a subgraph of C_t, $\mathbb{P}[C_t \subseteq G^{\mathsf{aug}}] \leq \mathbb{P}[P_t \subseteq G^{\mathsf{aug}}]$ for all $t \geq 3$ (since the event that $P_t \subseteq G^{\mathsf{aug}}$ contains the event that $C_t \subseteq G^{\mathsf{aug}}$ for suitable labellings of the two graphs), so it suffices to consider $G = P_t$. For this case, we choose $\boldsymbol{D} = \frac{1}{2\sigma^2}\boldsymbol{I}_{t-1}$ in Proposition 4.3. That gives

$$\begin{aligned}\mathbb{P}[P_t \subseteq G^{\mathsf{aug}}] &\leq \exp\left(-\frac{d}{2}\sum_{i=1}^{t}\log\left(1 + \frac{1}{\sigma^2}\lambda_i(\boldsymbol{L}^{P_t}) - \frac{1}{4\sigma^2}\lambda_i(\boldsymbol{L}^{P_t})^2\right)\right) \\ &= \exp\left(-\frac{d}{2}\sum_{k=1}^{t-1}\log\left(1 + \frac{2}{\sigma^2}\left(1 - \cos\left(\frac{\pi k}{t}\right)\right) - \frac{1}{\sigma^2}\left(1 - \cos\left(\frac{\pi k}{t}\right)\right)^2\right)\right) \\ &= \exp\left(-\frac{d}{2}\sum_{k=1}^{t-1}\log\left(1 + \frac{1}{\sigma^2}\sin^2\left(\frac{\pi k}{t}\right)\right)\right) \\ &= \exp\left(-\frac{d}{2}S(\sigma^2, t)\right), \end{aligned} \tag{4.56}$$

completing the proof. ☐

REMARK 4.2. *While this approach to bounding $\mathbb{P}[G \subseteq G^{\mathsf{aug}}]$ may seem rather naive, there is reason to believe it is close to optimal up to constant factors in σ^2: we know from the proof of Proposition 3.1 for $t = 2$ that $\mathbb{P}[\{i,j\} \in E(G^{\mathsf{aug}}) \mid \boldsymbol{x}_i, \boldsymbol{x}_j] \approx \exp(-\frac{1}{4\sigma^2}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2)$, so if we heuristically suppose that the edges of G^{aug} occur independently conditional on $\boldsymbol{x}$, then we find*

$$\begin{aligned}\mathbb{P}[G \subseteq G^{\mathsf{aug}}] &\approx \mathop{\mathbb{E}}_{\boldsymbol{x}} \prod_{\{i,j\}\in E(G)} \mathbb{P}[\{i,j\} \in E(G^{\mathsf{aug}}) \mid \boldsymbol{x}_i, \boldsymbol{x}_j] \\ &\approx \mathop{\mathbb{E}}_{\boldsymbol{x}} \exp\left(-\frac{1}{4\sigma^2}\boldsymbol{x}^\top(\boldsymbol{L}^G \otimes \boldsymbol{I}_d)\boldsymbol{x}\right) \\ &= \det\left(\boldsymbol{I}_V + \frac{1}{2\sigma^2}\boldsymbol{L}^G\right)^{-d/2}\end{aligned}$$

and if, for instance, $G = C_t$ then following the computations in Proposition 3.1 for $t \geq 3$ we would find

$$= \exp\left(-\frac{d}{2}S\left(\frac{\sigma^2}{2}, t\right)\right), \tag{4.57}$$

differing only by a factor of 2 in σ^2 from the bound of Proposition 4.4.

4.2 Concentration-Enhanced Second Moment Method We next review a version of the second moment method that can sometimes improve a weak result of the ordinary method—showing an object exists with quite low probability—to a strong result with high probability by combining it with a concentration inequality. Below, Part (b) is the typical result of a second moment method that has not succeeded in showing that a random variable is positive with high probability, instead only giving a lower bound of exponentially small probability. Part (a) is a concentration inequality, which in our case will come from a martingale argument, showing that the random variable also enjoys concentration around its mean with Gaussian tails. Exploiting the interplay of these two inequalities, we may in fact "repair" the ineffective second moment, as follows.

LEMMA 4.1. *Suppose $X \geq 0$ is a random variable and $m > 0$ are such that the following two statements hold, for some constants $0 < \beta < \alpha$:*

(a) $\mathbb{P}[X - \mathbb{E}X \leq -t] \vee \mathbb{P}[X - \mathbb{E}X \geq t] \leq \exp(-\alpha t^2/m)$ *for all* $t > 0$.

(b) $\mathbb{P}[X \geq m] \geq \exp(-\beta m)$.

Then, for any $0 < \gamma < 1 - \sqrt{\beta/\alpha}$,

$$\mathbb{P}[X > \gamma m] \geq 1 - \exp\left(-\alpha\left(1 - \sqrt{\frac{\beta}{\alpha}} - \gamma\right)^2 m\right). \tag{4.58}$$

Proof. Suppose $\delta \in (0,1)$. Then, whenever $\mathbb{E}X \leq (1-\delta)m$, we have

$$\begin{aligned}\exp(-\beta m) &\leq \mathbb{P}[X \geq m] \\ &\leq \mathbb{P}[X \geq \mathbb{E}X + \delta m] \\ &\leq \exp\left(-\frac{\alpha(\delta m)^2}{m}\right) \\ &= \exp(-\alpha\delta^2 m), \end{aligned} \tag{4.59}$$

whereby $\delta \leq \sqrt{\beta/\alpha}$. Thus, by contrapositive, $\mathbb{E}X > (1-\delta)m$ for all $\delta > \sqrt{\beta/\alpha}$, so $\mathbb{E}X \geq (1 - \sqrt{\beta/\alpha})m$.

Now, for all $0 < \gamma < 1 - \sqrt{\beta/\alpha}$, we have

$$\begin{aligned}\mathbb{P}[X \leq \gamma m] &\leq \mathbb{P}\left[X \leq \mathbb{E}X - \left(1 - \sqrt{\frac{\beta}{\alpha}} - \gamma\right)m\right] \\ &\leq \exp\left(-\alpha\left(1 - \sqrt{\frac{\beta}{\alpha}} - \gamma\right)^2 m\right),\end{aligned} \tag{4.60}$$

as claimed. □

Our formulation here is very similar to that of Frieze in [22], who treats the largest independent set in an ER graph; a similar idea also appeared earlier in [43] for the chromatic number of an ER graph. See also [33] for a survey of related methods.

4.3 Type (a) and (b) Inequalities We now proceed to the main computations for using the concentration-enhanced second moment method, which we state as general claims for all dimensions d. In the following sections we will derive specific consequences for different scalings of d.

Unfortunately, applying our method directly to the random variable M does not afford us sufficient flexibility to adjust the constants α and β such that the condition $\beta < \alpha$ is satisfied. Instead, we will proceed by applying Lemma 4.1 to the following adjustment of M, which is also directly analogous to the approach of Frieze in [22], there credited to Luczak, to the existence of independent sets. Given $r \in \mathbb{Z}_+$, let $n' := \lfloor n/r \rfloor$, and let $A_k = \{(k-1)r+1, \dots, kr\}$ for $k \in [n']$. Then, we call a matching *r-good* if all of its vertices belong to $A_1 \cup \cdots \cup A_{n'}$, and it contains at most one vertex in each A_k. We then work with the random variable

$$M^{(r)} := \text{number of edges in the largest } r\text{-good matching in } G^{\mathsf{aug}}. \tag{4.61}$$

Clearly, $M \geq M^{(r)}$.

Lemma 4.2. (Type (a) inequality) *For all $t > 0$,*

$$\mathbb{P}[M^{(r)} - \mathbb{E}M^{(r)} \leq -t] \vee \mathbb{P}[M^{(r)} - \mathbb{E}M^{(r)} \geq t] \leq \exp\left(-\frac{t^2}{2n'}\right). \tag{4.62}$$

That is, inequality (a) of Lemma 4.1 holds for $M^{(r)}$ for any $m > 0$ with

$$\alpha = \frac{m}{2n'}. \tag{4.63}$$

Proof. For an arbitrary graph G on vertex set $[n]$, let $M^{(r)}(G)$ denote the number of edges in the largest r-good matching in G.

We first claim that, if there exists some $k \in [n']$ such that G and G' differ only on edges incident with A_k, then $|M^{(r)}(G) - M^{(r)}(G')| \leq 1$. Indeed, if the largest matching in G' contains no edge incident with A_k, then the same matching exists in $M^{(r)}(G)$, so $M^{(r)}(G) \geq M^{(r)}(G')$. If the largest matching in G' does contain an edge incident with A_k, then the matching formed by removing that edge exists in $M^{(r)}(G)$, so $M^{(r)}(G) \geq M^{(r)}(G') - 1$. Thus $M^{(r)}(G') - M^{(r)}(G) \leq 1$, and symmetrically $M^{(r)}(G) - M^{(r)}(G') \leq 1$.

Now, view $M^{(r)} = M^{(r)}(G^{\mathsf{aug}})$ as a function of $\boldsymbol{x}_1, \boldsymbol{z}_1, \dots, \boldsymbol{x}_n, \boldsymbol{z}_n$. Form the Doob's martingale $M_k^{(r)} := \mathbb{E}[M^{(r)} \mid \{\boldsymbol{x}_i\}_{i \in A_1 \cup \cdots \cup A_k} \cup \{\boldsymbol{z}_i\}_{i \in A_1 \cup \cdots \cup A_k}]$ for $k = 0, 1, \dots, n'$, for which $M_0^{(r)} = \mathbb{E}M^{(r)}$ and $M_{n'}^{(r)} = M^{(r)}$. By the above claim, $|M_k^{(r)} - M_{k-1}^{(r)}| \leq 1$ for all k, and the result then follows from the Azuma-Hoeffding inequality (see Lemma 1.2 of [33]). □

Our type (b) inequality involves the multinomial entropy function H, defined for $x_1, \dots, x_k \geq 0$ satisfying $x_1 + \cdots + x_k \leq 1$ as

$$H(x_1, \dots, x_k) := -\sum_{i=1}^{k} x_i \log x_i - \left(1 - \sum_{i=1}^{k} x_i\right) \log\left(1 - \sum_{i=1}^{k} x_i\right). \tag{4.64}$$

We use the slightly non-standard notation of omitting what is usually the last argument $1 - \sum_{i=1}^{k} x_i$ to shorten the expressions that arise below; this is, however, in agreement with the standard notation $H(x) = -x\log x - (1-x)\log(1-x)$ for the binomial entropy.

We give a coarsely-bounded exponential rate function below; this will suffice for our purposes and we make no efforts to optimize our analysis at the level of constants on the exponential scale in m. More precise expressions are mentioned in our proof to follow.

Lemma 4.3. (Type (b) inequality) *Suppose $n' \geq 4m$ and $\sigma^2 \leq \frac{d}{40}$. Define as before $p := \mathbb{P}[\{i,j\} \in E(G^{\mathsf{aug}})]$. Then,*

$$\mathbb{P}[M^{(r)} \geq m] \geq \exp\left(-m \sup_{\boldsymbol{x} \in \mathcal{A}} F(\boldsymbol{x}) - O(\log n')\right), \tag{4.65}$$

where, for an absolute positive constant K (e.g., one may take $K = 50$),

$$\mathcal{A} := \left\{(\bar{a}, \bar{b}, \bar{c}, \bar{j}, \bar{k}, \bar{\ell}) \in [0,1]^6 : 2\bar{a} + \bar{b} + 2\bar{c} + \bar{j} + \bar{k} + \bar{\ell} \leq 1\right\}, \tag{4.66}$$

$$R_1 := K + \log\left(\frac{n'^2}{pn^2 m}\right), \tag{4.67}$$

$$R_2 := K + d(S(\sigma^2, 2) - I(\sigma^2)) + 4\log_+\left(\frac{d}{1+\sigma^2}\right) - 2\log r, \tag{4.68}$$

$$F(\bar{a}, \bar{b}, \bar{c}, \bar{j}, \bar{k}, \bar{\ell}) := 7H\left(\bar{a}, \bar{a}, \bar{b}, \bar{c}, \bar{c}, \bar{j}, \bar{k}, \bar{\ell}\right) + \left(\bar{a} + \bar{b} + \bar{c} + \bar{j} + \bar{k} + \bar{\ell}\right)(R_1 \vee R_2). \tag{4.69}$$

That is, if n' and m are functions of $n \to \infty$ with $n' = e^{o(m)}$ then, for any $\epsilon > 0$, for all sufficiently large n, inequality (b) of Lemma 4.1 holds with

$$\beta = \sup_{\boldsymbol{x} \in \mathcal{A}} F(\boldsymbol{x}) + \epsilon. \tag{4.70}$$

The basic idea of the remaining analysis will be to choose r and m to ensure that R_1 and R_2 are very negative, forcing $\bar{a}, \bar{b}, \bar{c}, \bar{j}, \bar{k}, \bar{\ell}$ to be small at the maximizing point. In R_1, we will accomplish this by taking $n' = Cm$ for some fixed C, and $m = cpn^2$ for some sufficiently small c. Then, the first term of F is also small, so $\sup F$ and therefore β may be made arbitrarily small by lowering c. On the other hand, $\alpha = m/2n' = 1/2C$, so we may ensure $\beta < \alpha$ and apply Lemma 4.1, finding that with high probability $M \geq M^{(r)} \geq c'pn^2$ for some $0 < c' < c$.

The following is the main technical preliminary to our proof, which bounds the moment generating function of the number of connected components in the union of two random edge-disjoint perfect matchings.

Proposition 4.5. (Cycle moment generating function) *Let ℓ be even and let K_ℓ be the complete graph on vertex set $[\ell]$. Let Q_1 be any perfect matching in K_ℓ, and let Q_2 be a uniformly random perfect matching in K_ℓ with the edges of Q_1 removed. Write X_ℓ for the random variable giving the number of connected components in $Q_1 \cup Q_2$. Then, for all $\ell \geq 4$ and all $a \geq \ell$,*

$$\mathbb{E}a^{X_\ell} \leq \frac{(\phi^2 a)^{\ell/4}}{(\ell/2)!!} \leq \left(\frac{e^3 a}{\ell}\right)^{\ell/4}, \tag{4.71}$$

where $\phi = (1+\sqrt{5})/2$ denotes the golden ratio.

Proof. We prove our bound inductively. For any fixed a, write $m_\ell := \mathbb{E}a^{X_\ell}$ for each even ℓ, where we take $m_0 = 1$ and $m_2 = 0$. We will prove that

$$m_\ell = \frac{a}{\ell - 3} m_{\ell-4} + \left(1 - \frac{1}{\ell-3}\right) m_{\ell-2}\,. \tag{4.72}$$

Let us assume that (4.72) holds for now and show how to derive the claim. Clearly the first inequality in (4.71) holds for $\ell = 0, 2, 4$. For the inductive step, suppose the bound holds for all values smaller than a given $\ell \geq 6$. The induction hypothesis then implies

$$\begin{aligned} m_\ell &\leq \frac{a}{\ell-3} \cdot \frac{(\phi^2 a)^{\ell/4-1}}{(\ell/2-2)!!} + \left(1 - \frac{1}{\ell-3}\right)\frac{(\phi^2 a)^{\ell/4-1/2}}{(\ell/2-1)!!} \\ &\leq \frac{a}{\ell/2} \cdot \frac{(\phi^2 a)^{\ell/4-1}}{(\ell/2-2)!!} + \frac{(\phi^2 a)^{\ell/4-1/2} a^{1/2}}{(\ell/2)!!} \\ &= \frac{(\phi^2 a^{\ell/4})}{(\ell/2)!}(\phi^{-2} + \phi^{-1})\,, \end{aligned}$$

where we have used that $\ell - 3 \geq \ell/2$ for all $\ell \geq 6$ and that $(\ell/2-1)!! \geq (\ell/2)!!/\sqrt{\ell} \geq (\ell/2)!!/\sqrt{a}$. Since $\phi^{-2} + \phi^{-1} = 1$, this completes the induction and proves the first inequality in (4.71), and the second is an immediate consequence.

All that is left is to establish the promised recurrence (4.72). Note that (Q_1, Q_2) as described are two uniformly random perfect matchings on $[\ell]$ conditioned to be edge disjoint. Each connected component of $Q_1 \cup Q_2$ is a cycle whose edges alternate between Q_1 and Q_2. Let us condition on the size of the component containing the vertex 1. Write i for the neighbor of 1 in Q_1, and j and k for the neighbors of 1 and i, respectively, in Q_2. Since Q_1 and Q_2 are edge-disjoint perfect matchings, 1, i, j, and k are distinct.

If 1 lies in a 4-cycle, then $\{j,k\} \in Q_1$, and removing the vertices $\{1,i,j,k\}$, and corresponding edges from Q_1 and Q_2 yields two uniformly random, edge-disjoint perfect matchings on $\ell - 4$ vertices, with one fewer connected component than $Q_1 \cup Q_2$. Since k is a uniform random vertex from $[\ell] \setminus \{1,i,j\}$, this situation occurs with probability $\frac{1}{\ell-3}$. This gives the first term of (4.72).

On the other hand, if 1 lies in a cycle of length greater than 4, then $\{j,k\} \notin Q_1$. In this case, removing the vertices 1 and i as well as the edges $\{1,i\}$ from Q_1 and $\{1,j\}$ from Q_2 and replacing the edge $\{i,k\}$ by $\{j,k\}$ in Q_2 yields two uniformly random, edge-disjoint perfect matchings on $\ell - 2$ vertices, whose union has the same number of connected components as $Q_1 \cup Q_2$. Since this occurs with probability $1 - \frac{1}{\ell-3}$, this yields the second term of (4.72). □

We will also use the following inequalities among the various functions of σ^2, whose proofs we defer to Appendix C.3.

PROPOSITION 4.6. *For $\sigma^2 > 0$, define*

$$\eta_1 = \eta_1(\sigma^2) := \frac{3}{4}S(\sigma^2, 2) - \frac{1}{4}S(\sigma^2, 4), \tag{4.73}$$

$$\eta_2 = \eta_2(\sigma^2) := S(\sigma^2, 2) - \frac{1}{2}S(\sigma^2, 3), \tag{4.74}$$

$$\eta_3 = \eta_3(\sigma^2) := \frac{1}{2}S(\sigma^2, 2) - \frac{1}{2}I(\sigma^2). \tag{4.75}$$

Then, we have

$$\eta_i \leq \eta_3 \leq \frac{3}{2+8\sigma^2} \text{ for each } i \in \{1,2,3\}. \tag{4.76}$$

We remark that these results are qualitatively sharp, in that the given quantities indeed approach positive constants as $\sigma^2 \to 0$, and decay as $O(\sigma^{-2})$ as $\sigma^2 \to \infty$; proofs of matching opposite bounds follow from similar elementary manipulations to those we give in the proof.

Finally, we will use the following standard properties of the multinomial entropy function H. We note that we adopt the same convention for multinomial coefficients of omitting the last argument as we do for H:

$$\binom{m}{a_1, \dots, a_k} := \frac{m!}{a_1! \cdots a_k! (m - a_1 - \cdots - a_k)!}. \tag{4.77}$$

PROPOSITION 4.7. *The function H satisfies the following properties:*

1. *$H(x_1, \dots, x_k) \leq \log(k+1)$.*
2. *For any $x \in (0,1)$, $tH(x/t)$ is a strictly increasing function of t.*
3. *For any $x_1, \dots, x_k \geq 0$ with $x_1 + \cdots + x_k \leq 1$, $H(x_1 + x_2, x_3, \dots, x_k) \leq H(x_1, x_2, x_3, \dots, x_k)$, and for any $k' < k$, $H(x_1, \dots, x_{k'}) \leq H(x_1, \dots, x_k)$.*
4. *A multinomial coefficient is bounded by the entropy as*

$$\exp\left(mH\left(\frac{a_1}{m}, \cdots, \frac{a_k}{m}\right) - O_k(\log m)\right) \leq \binom{m}{a_1, \dots, a_k} \leq \exp\left(mH\left(\frac{a_1}{m}, \cdots, \frac{a_k}{m}\right)\right). \tag{4.78}$$

Proof. [Proof of Lemma 4.3] Define the random variable

$$N := \#\{r\text{-good matchings on } 2m \text{ vertices of } G^{\mathsf{aug}}\}. \tag{4.79}$$

We then have

$$\mathbb{P}[M^{(r)} \geq m] = \mathbb{P}[N > 0], \tag{4.80}$$

and we will bound the latter from below by the second moment method.

Let $\mathcal{M}$ denote the set of r-good matchings of $2m$ vertices of the complete graph on $[n]$, whose cardinality is

$$|\mathcal{M}| = \binom{n'}{2m} r^{2m}(2m-1)!!. \tag{4.81}$$

We then have by linearity of expectation that

$$\mathbb{E}N = p^m|\mathcal{M}|. \tag{4.82}$$

Let Q_0 be a fixed r-good matching of m elements in the complete graph on $[n]$ (say, the graph with edges $\{1,2\},\{3,4\},\dots,\{2m-1,2m\}$). By symmetry, we have

$$\mathbb{E}N^2 = |\mathcal{M}| \sum_{Q \in \mathcal{M}} \mathbb{P}[Q_0 \cup Q \subseteq G^{\mathsf{aug}}], \tag{4.83}$$

and therefore the moment ratio may be written as an average,

$$\frac{\mathbb{E}N^2}{(\mathbb{E}N)^2} = \frac{1}{|\mathcal{M}|} \sum_{Q \in \mathcal{M}} \frac{\mathbb{P}[Q_0 \cup Q \subseteq G^{\mathsf{aug}}]}{p^{2m}}. \tag{4.84}$$

Given a graph G, write $\mathsf{cc}(G)$ for the set of its connected components, $\mathsf{cc}_2(G)$ for the set of its connected components isomorphic to the path on two vertices, $\mathsf{cc}_3(G)$ for the set of those isomorphic to the path on three vertices, and $\mathsf{cc}_{\geq 4}(G)$ for the set of the remaining connected components. Let us abbreviate $G := Q_0 \cup Q$. Note that all components of G are then either cycles of even length at least 4 or paths. Then, by Proposition 4.4, for any connected component $H \in \mathsf{cc}_{\geq 4}(G)$ we have

$$\mathbb{P}[H \subseteq G^{\mathsf{aug}}] \leq \exp\left(-\frac{d}{2}S(\sigma^2, |V(H)|)\right) \tag{4.85}$$

and, applying Corollary 2.1 with $t_0 = 4$, we have

$$\leq \exp\left(-\frac{d}{2}(|V(H)|I - J)\right), \tag{4.86}$$

where $I = I(\sigma^2)$ and $J = J(\sigma^2) = 4I(\sigma^2) - S(\sigma^2, 4) > 0$. For the remainder of this proof, let us follow the above convention abbreviating $I = I(\sigma^2)$ and $J = J(\sigma^2)$, and also writing $S_t = S(\sigma^2, t)$. Using this bound for connected components on at least four vertices and Proposition 4.4 for connected components on three vertices, we then have

$$\begin{aligned}
&\mathbb{P}[G \subseteq G^{\mathsf{aug}}] \\
&= \prod_{H \in \mathsf{cc}(G)} \mathbb{P}[H \subseteq G^{\mathsf{aug}}] \\
&\leq \exp\left(|\mathsf{cc}_2(G)|\log p - |\mathsf{cc}_3(G)|\frac{d}{2}S_3 - \frac{d}{2}\sum_{H \in \mathsf{cc}_{\geq 4}(G)} (|V(H)|I - J)\right) \\
&= \exp\left(|\mathsf{cc}_2(G)|\log p - |\mathsf{cc}_3(G)|\frac{d}{2}S_3 - (|V(G)| - 2|\mathsf{cc}_2(G)| - 3|\mathsf{cc}_3(G)|)\frac{d}{2}I + |\mathsf{cc}_{\geq 4}(G)|\frac{d}{2}J\right) \\
&= \exp\left(-|V(G)|\frac{d}{2}I + |\mathsf{cc}_2(G)|\,(dI + \log p) + |\mathsf{cc}_3(G)|\left(\frac{3d}{2}I - \frac{d}{2}S_3\right) + |\mathsf{cc}_{\geq 4}(G)|\frac{d}{2}J\right).
\end{aligned} \tag{4.87}$$

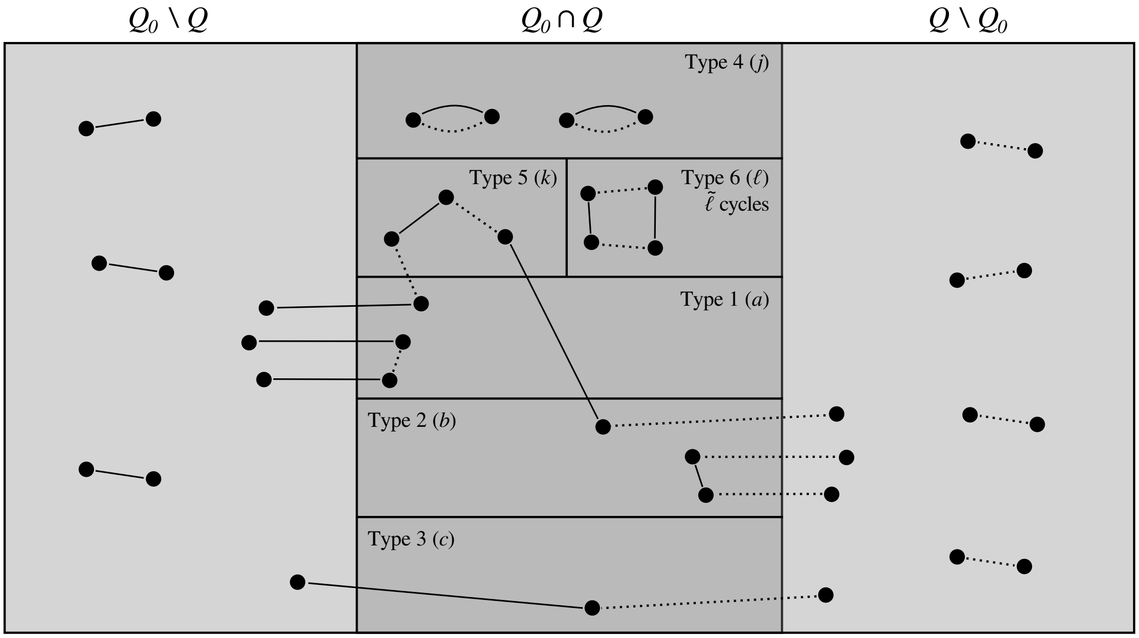

Figure 5: **Union of two matchings.** We illustrate the decomposition of two partial matchings of $[n]$ from the proof of Lemma 4.3. The center region contains all vertices of $V(Q_0) \cap V(Q)$, solid lines indicate edges of $E(Q_0)$, and dotted lines indicate edges of $E(Q)$.

We now divide the sum over $Q \in \mathcal{M}$ into portions over which we may uniformly bound this probability. To do this, for any given Q we introduce the following classification of the vertices of $Q_0 \cap Q$, into "types" 1, 2, 3, 4, 5, and 6. Next to each type, we give the letter that will denote the number of vertices of this type, a, b, c, j, k, and ℓ, respectively:

1. a vertices whose neighbor in Q_0 lies in $Q_0 \setminus Q$ and whose neighbor in Q lies in $Q_0 \cap Q$.

2. b vertices whose neighbor in Q_0 lies in $Q_0 \cap Q$ and whose neighbor in Q lies in $Q \setminus Q_0$.

3. c vertices whose neighbor in Q_0 lies in $Q_0 \setminus Q$ and whose neighbor in Q lies in $Q \setminus Q_0$.

4. j vertices whose neighbors in both Q_0 and Q are equal.

5. k vertices which belong to a path connected component and whose neighbors in Q_0 and Q are different but both lie in $Q_0 \cap Q$.

6. ℓ vertices which belong to a cycle connected component and whose neighbors in Q_0 and Q are different but both lie in $Q_0 \cap Q$.

We also denote by $\widetilde{\ell}$ the number of cycle connected components in $Q_0 \cup Q$ (which all consist of Type 6 vertices). See Figure 5 for an illustration of this decomposition. With these notations, recalling that $|V(Q_0)| = |V(Q)| = 2m$, we have

$$|V(Q_0) \cap V(Q)| = a + b + c + j + k + \ell, \tag{4.88}$$

$$\begin{aligned} |V(Q_0 \cup Q)| &= |V(Q_0)| + |V(Q)| - |V(Q_0) \cap V(Q)| \\ &= 4m - a - b - c - j - k - \ell, \end{aligned} \tag{4.89}$$

$$\begin{aligned} |\mathsf{cc}_2(Q_0 \cup Q)| &= \frac{j}{2} + \frac{2m - 2a - b - 2c - j - k - \ell}{2} + \frac{2m - a - 2b - 2c - j - k - \ell}{2} \\ &= 2m - \frac{3}{2}a - \frac{3}{2}b - 2c - \frac{1}{2}j - k - \ell, \end{aligned} \tag{4.90}$$

$$|\mathsf{cc}_3(Q_0 \cup Q)| = c, \tag{4.91}$$

$$|\mathsf{cc}_{\geq 4}(Q_0 \cup Q)| = \frac{1}{2}a + \frac{1}{2}b + \widetilde{\ell}, \tag{4.92}$$

the final claim following because every path component of length 4 or greater contains exactly two internal vertices of Type 1 or Type 2.

Let $\mathcal{M}_{a,b,c,j,k,\ell,\widetilde{\ell}}$ be the set of Q such that $Q_0 \cup Q$ has the specified number of vertices of each type, and $\widetilde{\ell}$ cycle connected components. We note that this set is empty unless j, $k + a$, and $k + b$, and ℓ are all even, since these sets of vertices must admit perfect matchings (from restrictions of both Q_0 and Q, Q, Q_0, and both Q_0 and Q, respectively). For $Q \in \mathcal{M}_{a,b,c,j,k,\ell,\widetilde{\ell}}$, we have

$$\begin{aligned} &\frac{\mathbb{P}[Q_0 \cup Q \subseteq G^{\mathsf{aug}}]}{p^{2m}} \\ &\leq \exp\Bigg(- (4m - a - b - c - j - k - \ell)\frac{d}{2}I \\ &\qquad + \left(2m - \frac{3}{2}a - \frac{3}{2}b - 2c - \frac{1}{2}j - k - \ell\right)(dI + \log p) \\ &\qquad + c\left(\frac{3d}{2}I - \frac{d}{2}S_3\right) \\ &\qquad + \left(\frac{1}{2}a + \frac{1}{2}b + \widetilde{\ell}\right)\frac{d}{2}J \\ &\qquad - 2m \log p\Bigg) \\ &= \exp\left(a\left(-dI + \frac{d}{4}J - \frac{3}{2}\log p\right) + b\left(-dI + \frac{d}{4}J - \frac{3}{2}\log p\right) + c\left(-\frac{d}{2}S_3 - 2\log p\right)\right. \\ &\qquad \left. + j\left(-\frac{1}{2}\log p\right) + k\left(-\frac{d}{2}I - \log p\right) + \ell\left(-\frac{d}{2}I - \log p\right) + \widetilde{\ell}\left(\frac{d}{2}J\right)\right) \end{aligned}$$

and, using that $\log p = -\frac{d}{2}S_2 + \log \widehat{p}$,

$$\begin{aligned} &= \exp\left((a + b)\left(\eta_1 d - \frac{3}{2}\log \widehat{p}\right) + c\,(\eta_2 d - 2\log \widehat{p}) + (k + \ell)\,(\eta_3 d - \log \widehat{p})\right. \\ &\qquad \left. + j\left(-\frac{1}{2}\log p\right) + \widetilde{\ell}\left(\frac{d}{2}J\right)\right), \end{aligned} \tag{4.93}$$

with η_i as defined in Proposition 4.6,

$$\eta_1 = -I + \frac{1}{4}J + \frac{3}{4}S_2 = \frac{3}{4}S_2 - \frac{1}{4}S_4, \tag{4.94}$$

$$\eta_2 = S_2 - \frac{1}{2}S_3, \tag{4.95}$$

$$\eta_3 = \frac{1}{2}S_2 - \frac{1}{2}I. \tag{4.96}$$

By Proposition 4.6 we have $\eta_i \leq \eta_3$, and by Proposition 4.2 we have $\widehat{p} \geq \frac{1}{1000}\sqrt{\frac{1+\sigma^2}{d}}$. Here and in the remainder of the proof, let K be a large constant that may vary line to line. Substituting these bounds,

$$\begin{aligned}
&\frac{\mathbb{P}[Q_0 \cup Q \subseteq G^{\mathsf{aug}}]}{p^{2m}} \\
&\leq \exp\Bigg((a+b+c+k)\left(K + \eta_3 d + \log_+\left(\frac{d}{1+\sigma^2}\right)\right) \\
&\qquad + j\left(-\frac{1}{2}\log p\right) + \ell\left(K + \eta_3 d + \log_+\left(\frac{d}{1+\sigma^2}\right)\right) + \widetilde{\ell}\left(\frac{d}{2}J\right)\Bigg).
\end{aligned} \tag{4.97}$$

We must also control the size of the subsets $\mathcal{M}_{a,b,c,j,k,\ell,\widetilde{\ell}}$, which is the content of the following technical lemma, whose proof we defer to the conclusion of this section.

Lemma 4.4. *Given ℓ and $\widetilde{\ell}$, let R_0 be a perfect matching of $[\ell]$, and write $\mathsf{Cyc}(\ell, \widetilde{\ell})$ for the number of perfect matchings R of $[\ell]$ edge-disjoint from R_0 such that $R_0 \cup R$ contains exactly $\widetilde{\ell}$ cycles. Define normalizations $\bar{a} := a/2m$ and likewise $\bar{b}, \bar{c}, \bar{j}, \bar{k}$, and $\bar{\ell}$. Then $\mathcal{M}_{a,b,c,j,k,\ell,\widetilde{\ell}}$ satisfies*

$$\begin{aligned}
\frac{\left|\mathcal{M}_{a,b,c,j,k,\ell,\widetilde{\ell}}\right|}{|\mathcal{M}|} &\leq \exp\left(m\left[5H\left(\bar{a}, \bar{a}, \bar{b}, \bar{c}, \bar{c}, \bar{j}, \bar{k}, \bar{\ell}\right) + (\bar{a} + \bar{c})\log 4\right] + O(\log n')\right) \cdot \\
&\qquad r^{-a-b-c-j-k-\ell}\frac{1}{(j-1)!!}\frac{\mathsf{Cyc}(\ell, \widetilde{\ell})}{(\ell-1)!!}.
\end{aligned} \tag{4.98}$$

Putting the bounds (4.97) and (4.98) together,

$$\begin{aligned}
\frac{\mathbb{E}N^2}{(\mathbb{E}N)^2} &\leq \sum_{a,b,c,j,k,\ell,\widetilde{\ell}} \frac{\left|\mathcal{M}_{a,b,c,j,k,\ell,\widetilde{\ell}}\right|}{|\mathcal{M}|} \max_{Q \in \mathcal{M}_{a,b,c,j,k,\ell,\widetilde{\ell}}} \frac{\mathbb{P}[Q_0 \cup Q \subseteq G^{\mathsf{aug}}]}{p^{2m}} \\
&\leq \sum_{a,b,c,j,k,\ell} \exp\Bigg(m\Bigg[5H\left(\bar{a}, \bar{a}, \bar{b}, \bar{c}, \bar{c}, \bar{j}, \bar{k}, \bar{\ell}\right) \\
&\qquad + (\bar{a} + \bar{b} + \bar{c} + \bar{k})\left(K + 2\eta_3 d + 2\log_+\left(\frac{d}{1+\sigma^2}\right) - 2\log r\right) \\
&\qquad + \bar{j}\left(-\log p - 2\log r\right) \\
&\qquad + \bar{\ell}\left(K + 2\eta_3 d + 2\log_+\left(\frac{d}{1+\sigma^2}\right) - 2\log r\right)\Bigg] + O(\log n')\Bigg) \\
&\qquad \frac{1}{(j-1)!!}\sum_{\widetilde{\ell}} \frac{\mathsf{Cyc}(\ell, \widetilde{\ell})}{(\ell-1)!!}(e^{\frac{d}{2}J})^{\widetilde{\ell}}
\end{aligned} \tag{4.99}$$

For the remaining sum over $\widetilde{\ell}$, we use Proposition 4.5. We note first that, since $(\ell-1)!!$ is the total number of matchings on $[\ell]$ and $\mathsf{Cyc}(\ell, \widetilde{\ell})$ is the number of such matchings that are disjoint from a fixed matching and whose union with that matching contains $\widetilde{\ell}$ cycles, we may generally bound $\sum_{\widetilde{\ell}} \frac{\mathsf{Cyc}(\ell,\widetilde{\ell})}{(\ell-1)!!} f(\widetilde{\ell}) \leq \mathbb{E}f(X_\ell)$, where X_ℓ is the random variable from Proposition 4.5. If $e^{\frac{d}{2}J} \leq \ell$, then we may bound

$$\sum_{\widetilde{\ell}} \frac{\mathsf{Cyc}(\ell, \widetilde{\ell})}{(\ell-1)!!}(e^{\frac{d}{2}J})^{\widetilde{\ell}} \leq \sum_{\widetilde{\ell}} \frac{\mathsf{Cyc}(\ell, \widetilde{\ell})}{(\ell-1)!!}(\ell)^{\widetilde{\ell}} \leq (e^3)^{\ell/4}. \tag{4.100}$$

If $e^{\frac{d}{2}J} \geq \ell$, then we have

$$\sum_{\widetilde{\ell}} \frac{\mathsf{Cyc}(\ell, \widetilde{\ell})}{(\ell-1)!!}(e^{\frac{d}{2}J})^{\widetilde{\ell}} \leq \left(\frac{e^{3+\frac{d}{2}J}}{\ell}\right)^{\ell/4}. \tag{4.101}$$

For the remaining term involving j, we bound $(j-1)!! \geq (j/e)^{j/2}$. Combining these estimates,we may incorporate everything under the exponential as

$$\begin{aligned}\frac{\mathbb{E}N^2}{(\mathbb{E}N)^2} \leq \sum_{a,b,c,j,k,\ell} \exp\Bigg(m \Bigg[& 5H\left(\bar{a},\bar{a},\bar{b},\bar{c},\bar{c},\bar{j},\bar{k},\bar{\ell}\right) \\ & + (\bar{a}+\bar{b}+\bar{c}+\bar{k})R_2 \\ & + \bar{j}\left(1 - \log p - 2\log r + \log\left(\frac{1}{j}\right)\right) \\ & + \bar{\ell}\left(K + 2\eta_3 d + 2\log_+\left(\frac{d}{1+\sigma^2}\right) - 2\log r + 0 \vee \left(\frac{d}{4}J + \frac{1}{2}\log\left(\frac{1}{\ell}\right)\right)\right)\Bigg] \\ & + O(\log n')\Bigg)\end{aligned} \tag{4.102}$$

To bound the remaining rates, we first consider the $\bar{j}$ term. Recall that $n' = \lfloor n/r \rfloor \geq n/2r$, so $r \geq n/2n'$. Thus we have

$$\begin{aligned} -\log p - 2\log r + \log\left(\frac{1}{j}\right) &= \log\frac{1}{pr^2 j} \\ &\leq \log\frac{4n'^2}{pn^2 j} \\ &= \log\frac{2n'^2}{pn^2 m} - \log\bar{j}. \end{aligned} \tag{4.103}$$

Extracting a similar expression by adding and subtracting $\frac{1}{2}\log p$ in the $\bar{\ell}$ term when the second term of the maximum is greater than zero,

$$\begin{aligned} & K + 2\eta_3 d + 2\log_+\left(\frac{d}{1+\sigma^2}\right) - 2\log r + \frac{d}{4}J + \frac{1}{2}\log\left(\frac{1}{\ell}\right) \\ & \leq K + \left(2\eta_3 + \frac{1}{4}J\right)d + 2\log_+\left(\frac{d}{1+\sigma^2}\right) + \frac{1}{2}\log p - \log r + \frac{1}{2}\log\frac{2n'^2}{pn^2 m} - \log\bar{\ell} \\ & \leq K + \left(2\eta_3 + \frac{1}{4}J - \frac{1}{4}S_2\right)d + 2\log_+\left(\frac{d}{1+\sigma^2}\right) - \log r + \frac{1}{2}\log\frac{2n'^2}{pn^2 m} - \log\bar{\ell} \end{aligned}$$

and we notice $2\eta_3 + \frac{1}{4}J - \frac{1}{4}S_2 = S_2 - I + I - \frac{1}{4}S_4 - \frac{1}{4}S_2 = \frac{3}{4}S_2 - \frac{1}{4}S_4 = \eta_1$, so, using Proposition 4.6 to bound $\eta_1 \leq \eta_3$,

$$\begin{aligned} & \leq K + \eta_3 d + 2\log_+\left(\frac{d}{1+\sigma^2}\right) - \log r + \frac{1}{2}\log\frac{2n'^2}{pn^2 m} - \log\bar{\ell} \\ & \leq \frac{1}{2}(R_1 + R_2) - \log\bar{\ell} \end{aligned} \tag{4.104}$$

We note also that $-\bar{j}\log\bar{j} \leq H(\bar{j})$ and likewise $-\bar{\ell}\log\bar{\ell} \leq H(\bar{\ell})$, and both of these are bounded by $H(\bar{a},\bar{a},\bar{b},\bar{c},\bar{c},\bar{j},\bar{k},\bar{\ell})$ by Proposition 4.7.

Applying these observations,

$$\begin{aligned}\frac{\mathbb{E}N^2}{(\mathbb{E}N)^2} \leq \sum_{a,b,c,j,k,\ell} \exp\Bigg(m \Bigg[& 7H\left(\bar{a},\bar{a},\bar{b},\bar{c},\bar{c},\bar{j},\bar{k},\bar{\ell}\right) \\ & + (\bar{a}+\bar{b}+\bar{c}+\bar{k})R_2 + \bar{j}R_1 \\ & + \bar{\ell}\max\left\{R_2, \frac{1}{2}(R_1+R_2)\right\}\Bigg] \\ & + O(\log n')\Bigg), \end{aligned} \tag{4.105}$$

which implies the result. ∎

To complete the proof, it remains to justify Lemma 4.4.

Proof. [Proof of Lemma 4.4] We begin with a combinatorial bound. Below, line by line, the factors count the number of ways to choose the vertices of $V(Q) \cap V(Q_0)$, the number of ways to choose the vertices of $V(Q) \setminus V(Q_0)$, the number of ways to draw the edges of $E(Q)$ incident with $V(Q) \cap V(Q_0)$, and the number of ways to draw the edges of $E(Q)$ between pairs of $V(Q) \setminus V(Q_0)$:

$$\begin{aligned}
&|\mathcal{M}_{a,b,c,j,k,\ell,\widetilde{\ell}}| \\
&\leq \binom{m}{a,\, c,\, \frac{b+k}{2},\, \frac{j}{2},\, \frac{\ell}{2}} 2^{a+c} \cdot \\
&\binom{n' - a - \frac{b}{2} - c - \frac{j}{2} - \frac{k}{2} - \frac{\ell}{2}}{2m - a - b - c - j - k - \ell} r^{2m-a-b-c-j-k-\ell} \cdot \\
&\binom{b+k}{k} (k+a-1)!! \binom{2m - a - b - c - j - k - \ell}{b+c} (b+c)! \mathsf{Cyc}(\ell, \widetilde{\ell}) \cdot \\
&(2m - a - 2b - 2c - j - k - \ell - 1)!!
\end{aligned}$$

Let us introduce $C := n'/2m$, which satisfies $C \geq 2$ by assumption. Then, applying the entropy bound for multinomial coefficients wherever possible,

$$\begin{aligned}
&\leq \exp\Bigg(m \Bigg[H\left(2\bar{a}, 2\bar{c}, \bar{b} + \bar{k}, \bar{j}, \bar{\ell}\right) \\
&\qquad + 2\left(C - \left(\bar{a} + \frac{1}{2}\bar{b} + \bar{c} + \frac{1}{2}\bar{j} + \frac{1}{2}\bar{k} + \frac{1}{2}\bar{\ell} \right) \right) H\left(\frac{1 - \bar{a} - \bar{b} - \bar{c} - \bar{j} - \bar{k} - \bar{\ell}}{C - (\bar{a} + \frac{1}{2}\bar{b} + \bar{c} + \frac{1}{2}\bar{j} + \frac{1}{2}\bar{k} + \frac{1}{2}\bar{\ell})} \right) \\
&\qquad + 2(1 - \bar{a} - \bar{b} - \bar{c} - \bar{j} - \bar{k} - \bar{\ell}) H\left(\frac{\bar{b} + \bar{c}}{1 - \bar{a} - \bar{b} - \bar{c} - \bar{j} - \bar{k} - \bar{\ell}} \right) \\
&\qquad + 2(\bar{b} + \bar{k}) H\left(\frac{\bar{k}}{\bar{b} + \bar{k}} \right) + 2(\bar{a} + \bar{c}) \log 2 \Bigg] \Bigg) \cdot \\
&r^{2m-a-b-c-j-k-\ell} (k+a-1)!!(b+c)!(2m - a - 2b - 2c - j - k - \ell - 1)!!(j-1)!!(\ell-1)!! \cdot \\
&\frac{1}{(j-1)!!} \frac{\mathsf{Cyc}(\ell, \widetilde{\ell})}{(\ell-1)!!}
\end{aligned} \tag{4.106}$$

We will in particular need to bound the fraction of $\mathcal{M}$ occupied by each of these subsets. To that end, we note that

$$|\mathcal{M}| = \binom{n'}{2m} r^{2m} (2m-1)!! \geq \exp\left(m \left[2CH\left(\frac{1}{C} \right) \right] - O(\log n') \right) r^{2m} (2m-1)!!. \tag{4.107}$$

Considering the quotient of factorials and double factorials that will remain, an entropy bound again yields

$$\begin{aligned}
&\frac{(k+a-1)!!(b+c)!(2m - a - 2b - 2c - j - k - \ell - 1)!!(j-1)!!(\ell-1)!!}{(2m-1)!!} \\
&\leq \exp\left(-mH\left(\bar{a} + \bar{k}, \bar{b} + \bar{c}, \bar{b} + \bar{c}, \bar{j}, \bar{\ell} \right) + O(\log n') \right),
\end{aligned} \tag{4.108}$$

where we have used Proposition 4.7, that $A!!$ obeys the same exponential asymptotics as $\sqrt{A!}$, and that $m = O(n')$

so that we may replace the $\log m$ term with $\log n'$. Thus we find

$$\begin{aligned}
&\frac{|\mathcal{M}_{a,b,c,j,k,\ell,\widetilde{\ell}}|}{|\mathcal{M}|} \\
&\leq \exp\Bigg(m\Bigg[H\left(2\bar{a}, 2\bar{c}, \bar{b}+\bar{k}, \bar{j}, \bar{\ell}\right) \\
&\qquad\qquad + 2\left(C - \left(\bar{a} + \frac{1}{2}\bar{b} + \bar{c} + \frac{1}{2}\bar{j} + \frac{1}{2}\bar{k} + \frac{1}{2}\bar{\ell}\right)\right) H\left(\frac{1 - \bar{a} - \bar{b} - \bar{c} - \bar{j} - \bar{k} - \bar{\ell}}{C - (\bar{a} + \frac{1}{2}\bar{b} + \bar{c} + \frac{1}{2}\bar{j} + \frac{1}{2}\bar{k} + \frac{1}{2}\bar{\ell})}\right) \\
&\qquad\qquad + 2(1 - \bar{a} - \bar{b} - \bar{c} - \bar{j} - \bar{k} - \bar{\ell}) H\left(\frac{\bar{b}+\bar{c}}{1 - \bar{a} - \bar{b} - \bar{c} - \bar{j} - \bar{k} - \bar{\ell}}\right) \\
&\qquad\qquad - H\left(\bar{a}+\bar{k}, \bar{b}+\bar{c}, \bar{b}+\bar{c}, \bar{j}, \bar{\ell}\right) - 2CH\left(\frac{1}{C}\right) \\
&\qquad\qquad + 2(\bar{b}+\bar{k})H\left(\frac{\bar{k}}{\bar{b}+\bar{k}}\right) + 2(\bar{a}+\bar{c})\log 2\Bigg] + O(\log n')\Bigg) \\
&\qquad r^{-a-b-c-j-k-\ell} \frac{1}{(j-1)!!} \frac{\mathsf{Cyc}(\ell, \widetilde{\ell})}{(\ell-1)!!}
\end{aligned}$$

and repeatedly use Proposition 4.7 to bound the entropies,

$$\begin{aligned}
&\leq \exp\left(m\left[5H\left(\bar{a}, \bar{a}, \bar{b}, \bar{c}, \bar{c}, \bar{j}, \bar{k}, \bar{\ell}\right) + (\bar{a}+\bar{c})\log 4\right] + O(\log n')\right) \\
&\qquad r^{-a-b-c-j-k-\ell} \frac{1}{(j-1)!!} \frac{\mathsf{Cyc}(\ell, \widetilde{\ell})}{(\ell-1)!!}, \qquad (4.109)
\end{aligned}$$

where we have used that $C \geq 2$ ensures that the sum of the two terms involving C is at most zero. ◻

4.4 Sublinear Error Lower Bound We now prove the following application of the above results, which implies the lower bound of Part 3 of Theorem 1.1 and of Part 2 of Theorem 1.2.

LEMMA 4.5. *Define* $\widehat{d} := 1 + d \wedge \log n \geq 2$ *and* $s := \widehat{d}^{1/d}$. *Suppose that*

$$\frac{1}{s^{-\omega(1)}n^{4/d} - 1} \leq \sigma^2 \leq \frac{1}{(2s^{\omega(1)}n^{1/d} - 1)^2 - 1}. \qquad (4.110)$$

Then, there exists an absolute constant $c > 0$ *such that, with high probability,*

$$|\mathcal{E}| \geq \frac{c}{\sqrt{\widehat{d}}}\left(1 + \frac{1}{\sigma^2}\right)^{-d/2} n^2. \qquad (4.111)$$

Proof. Let us bound $\widehat{p}$ from below under these assumptions, which amounts to bounding $\frac{d}{1+\sigma^2}$ from above. We always have $\frac{d}{1+\sigma^2} \leq d$, and, using the lower bound above along with $1 - e^{-x} \leq x$, we have

$$\frac{d}{1+\sigma^2} \leq d(1 - s^{\omega(1)}n^{-4/d}) \leq d(1 - n^{-4/d}) \leq 4\log n. \qquad (4.112)$$

Thus,

$$\frac{d}{1+\sigma^2} \leq d \wedge 4\log n \leq 4\widehat{d}, \qquad (4.113)$$

and so, by Proposition 4.2

$$\widehat{p} \geq \frac{1}{1000}\sqrt{\frac{1+\sigma^2}{d}} \geq \frac{1}{2000}\widehat{d}^{-1/2}. \qquad (4.114)$$

Next, we bound the other term of p, $\exp(-\frac{d}{2}S(\sigma^2,2))$, from above and below. Since $S(\sigma^2,2)$ is a decreasing function of σ^2, by the lower bound on σ^2 we have

$$S(\sigma^2,2) \le S\left(\frac{1}{s^{-\omega(1)}n^{4/d}-1},2\right) = \log(s^{-\omega(1)}n^{4/d}) \le \frac{4\log n - \omega(\log \widehat{d})}{d}, \tag{4.115}$$

and thus

$$\exp\left(-\frac{d}{2}S(\sigma^2,2)\right) \ge \frac{1}{n^2}\exp(\omega(\log \widehat{d})), \tag{4.116}$$

whereby $pn^2 = \widehat{p}\exp(-\frac{d}{2}S(\sigma^2,2))n^2 \to \infty$ as $n \to \infty$. On the other hand, by the upper bound on σ^2 we have

$$S(\sigma^2,2) \ge S\left(\frac{1}{(2s^{\omega(1)}n^{1/d}-1)^2-1},2\right) = 2\log(2s^{\omega(1)}n^{1/d}-1) \ge \frac{2\log n}{d}, \tag{4.117}$$

whereby

$$\exp\left(-\frac{d}{2}S(\sigma^2,2)\right) \le \frac{1}{n}, \tag{4.118}$$

so, since $\widehat{p} \le 1$, $pn \le 1$ as well.

With these properties in mind, let us set up an application of Lemma 4.1 via Lemmata 4.2 and 4.3, with which we will seek to show that $M \gtrsim pn^2$ with high probability. Fix $c > 0$ a small constant, and take

$$r := \left\lfloor \frac{2}{cpn} \right\rfloor \in \left[\frac{1}{cpn}, \frac{4}{cpn}\right], \tag{4.119}$$

$$n' := \left\lfloor \frac{n}{r} \right\rfloor \in \left[\frac{1}{4}cpn^2, cpn^2\right], \tag{4.120}$$

$$m := \left\lfloor \frac{1}{32}cpn^2 \right\rfloor \in \left[\frac{1}{64}cpn^2, \frac{1}{16}cpn^2\right]. \tag{4.121}$$

Then by Lemma 4.2, the type (a) inequality, holds with $\alpha = m/2n' \ge \frac{1}{128}$.

For the type (b) inequality, we have $n' \ge 4m$ by our choice, and by the upper bound on σ^2,

$$\sigma^2 \le \frac{1}{4n^{1/d}(n^{1/d}-1)} \le \frac{1}{4(n^{1/d}-1)} \le \frac{d}{4\log n}, \tag{4.122}$$

so, for sufficiently large n, the conditions of Lemma 4.3 are satisfied. It remains to control the rates R_1 and R_2 appearing in the Lemma, and thus to bound β.

We note in advance that, by the upper bound on σ^2 and since $I(\sigma^2)$ is a decreasing function,

$$I(\sigma^2) \ge I\left(\frac{1}{(2s^{\omega(1)}n^{1/d}-1)^2-1}\right) = \frac{2\log n}{d} + \omega(\log s) = \frac{2\log n + \omega(\widehat{d})}{d}. \tag{4.123}$$

The quantities appearing in these rates satisfy

$$\frac{n'^2}{pn^2m} \le 64c \tag{4.124}$$

$$\begin{aligned} d(S(\sigma^2,2) - I(\sigma^2)) - 2\log r &\le -dI(\sigma^2) + 2\log n + 2\log c \\ &\le -\log n\left(\frac{d}{\log n}I(\sigma^2) - 2\right) \\ &\le -\omega(\log \widehat{d}). \end{aligned} \tag{4.125}$$

We thus have

$$R_1 = K + \log\left(\frac{n'^2}{pn^2 m}\right) \le K + \log 64c, \tag{4.126}$$

$$R_2 = K + d(S(\sigma^2,2) - I(\sigma^2)) + 4\log_+\left(\frac{d}{1+\sigma^2}\right) - 2\log r = -\omega(\log \widehat{d}), \tag{4.127}$$

using in the latter our earlier result that $\frac{d}{1+\sigma^2} = O(\widehat{d})$.

For any $D > 0$, we may therefore choose c small enough that $R_i \le -D$ for $i = 1, 2$, so the whole rate function F in Lemma 4.3 satisfies

$$F(\bar{a}, \bar{b}, \bar{c}, \bar{j}, \bar{k}, \bar{\ell}) \le 7H(\bar{a}, \bar{a}, \bar{b}, \bar{c}, \bar{c}, \bar{j}, \bar{k}, \bar{\ell}) - D(\bar{a} + \bar{b} + \bar{c} + \bar{j} + \bar{k} + \bar{\ell})$$

The first term is bounded uniformly by $7\log 9$, so sufficiently large D we may ensure that F is negative if any of $\bar{a}, \bar{b}, \bar{c}, \bar{j}, \bar{k}$, or $\bar{\ell}$ is at least ϵ. On the other hand if all of the parameters are at most ϵ, then the first term is at most $-7(8\epsilon\log\epsilon + (1-\epsilon)\log(1-\epsilon))$, which tends to zero as $\epsilon \to 0$. Thus for sufficiently small c we may make the supremum $\sup_{\boldsymbol{x}\in\mathcal{A}} F(\boldsymbol{x})$ bounded by any arbitrarily small positive number. In particular, for any $\epsilon > 0$ there exists $c > 0$ such that the type (b) inequality holds with $\beta < \epsilon$. For c sufficiently small we may thus ensure, e.g., $\beta \le \frac{1}{512} \le \frac{\alpha}{4}$. Thus we may take $\gamma = \frac{1}{4}$ in Lemma 4.1, which gives that, with high probability, $M \ge \frac{1}{2048} cpn^2$. Substituting our lower bound on $\widehat{p}$ then gives the result as stated. □

4.5 Linear or Nearly-Linear Error Lower Bound Finally, we prove the following result, which yields Part 4 of Theorem 1.1.

LEMMA 4.6. *Let $s := d^{1/d}$. Suppose that $1 \le d \ll \log n$, and that for some $a \in \mathbb{R}$,*

$$\sigma^2 \ge \frac{1}{(2s^a n^{1/d} - 1)^2 - 1}. \tag{4.128}$$

Then, there exists $c = c(a) > 0$ such that

$$|\mathcal{E}| \ge e^{-cd} n \tag{4.129}$$

with high probability.

Proof. We first produce similar preliminary bounds to before. As before we have $\frac{d}{1+\sigma^2} \le d$, and thus

$$\widehat{p} \ge \frac{1}{1000} d^{-1/2}. \tag{4.130}$$

For the bounds on the other, exponential factor in p, we have, again assuming n is sufficiently large,

$$S(\sigma^2, 2) \le S\left(\frac{1}{(2s^a n^{1/d} - 1)^2 - 1}, 2\right) = 2\log(2s^a n^{1/d} - 1) \le 2\log(2s^a n^{1/d}), \tag{4.131}$$

whereby

$$\exp\left(-\frac{d}{2} S(\sigma^2, 2)\right) \ge \frac{1}{2^d s^{ad} n} = \frac{d^{-a}}{2^d n}. \tag{4.132}$$

Thus we may bound

$$pn = \widehat{p}\exp\left(-\frac{d}{2} S(\sigma^2, 2)\right) n \ge \frac{1}{2000} d^{-a-1/2} 2^{-d}. \tag{4.133}$$

We take a similar choice of parameters to the previous proof, only now taking r a constant not depending on n. Let $c = c(a) > 0$ be a constant to be fixed later, and take

$$r := \lfloor 2e^{cd} \rfloor \in \left[e^{cd}, 4e^{cd}\right], \tag{4.134}$$

$$n' := \left\lfloor \frac{n}{r} \right\rfloor \in \left[\frac{1}{4}e^{-cd}n, e^{-cd}n\right], \tag{4.135}$$

$$m := \left\lfloor \frac{1}{32}e^{-cd}n \right\rfloor \in \left[\frac{1}{64}e^{-cd}n, \frac{1}{16}e^{-cd}n\right]. \tag{4.136}$$

Then by Lemma 4.2, the type (a) inequality, holds as before with $\alpha = m/2n' \geq \frac{1}{128}$. The conditions of Lemma 4.3 are again satisfied. To control the rates appearing there, we again have

$$\frac{{n'}^2}{pn^2m} \leq 64\frac{e^{-cd}}{pn} = O((2e^{-c})^d d^{a+1/2}), \tag{4.137}$$

and for the other rate we use that, by Proposition 4.6, for all σ^2 we have $S(\sigma^2, 2) - I(\sigma^2) \leq \frac{3}{2}$, so

$$d(S(\sigma^2, 2) - I(\sigma^2)) - 2\log r \leq d\left(\frac{3}{2} - 2c\right). \tag{4.138}$$

Thus, for c sufficiently large the rates appearing in F are

$$R_1 = K + \log\left(\frac{{n'}^2}{pn^2m}\right) \leq K - \frac{c}{2}d, \tag{4.139}$$

$$R_2 = K + d(S(\sigma^2, 2) - I(\sigma^2)) + 4\log_+\left(\frac{d}{1+\sigma^2}\right) - 2\log r \leq K - \frac{c}{2}d, \tag{4.140}$$

thus choosing c large enough we may again make both rates arbitrarily negative, and the remainder of the proof goes through as before. □

Acknowledgments

We thank Cristopher Moore for helpful discussions about the second moment method, and the anonymous reviewers for many valuable suggestions.

A Greedy Algorithms and a Gaussian Limit

To supplement our discussion of the MLE, we describe two natural greedy algorithms for estimating the planted permutation $\pi^\star$ and discuss their performance heuristically. We believe the computations presented here are accurate, but for the sake of brevity we will not give detailed proofs. The algorithms we analyze here are also *improper* in the sense that they do not return a permutation; rather, they output an assignment of each $\boldsymbol{x}_i$ to an element of $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_n\}$ with no restriction that each element is matched exactly once. We leave more careful analysis of greedy algorithms which output a permutation to future work.

To summarize before presenting the details, the first greedy algorithm, where distance is measured as ordinary ℓ^2 distance, will match the performance of the MLE when $d = o(\log n)$ but will make $n - o(n)$ errors once $d = \omega(\log n)$. The second, which greedily selects the point with largest inner product with $\boldsymbol{x}_i$, will make $n - o(n)$ errors when $d = o(\log n)$ but will sometimes (though not always) improve on the MLE in the $d = \omega(\log n)$ regime. It is unclear what simplifying assumptions are reasonable when $d = \Theta(\log n)$, so we leave this case aside here; numerical evidence suggests that all three algorithms are competitive and none strictly dominates another in this regime.

A.1 Algorithm 1: Greedy Distance The first algorithm we consider is perhaps the most immediately appealing greedy algorithm, which attempts to match each point to its nearest neighbor. This may be viewed as greedily matching rows to columns in the matrix $\boldsymbol{W}^{(0)}$ of pairwise squared distances between the $\boldsymbol{x}_i$ and $\boldsymbol{y}_j$ formed as an intermediate step in our derivation of the MLE. As a proxy for the error incurred by such an algorithm, we consider the number of $\boldsymbol{x}_i$ whose nearest neighbor among the $\{\boldsymbol{y}_j\}_{j=1}^n$ is not equal to $\boldsymbol{y}_i$, the set of which we denote

$$\mathcal{E}^{\mathsf{dist}} := \{i \in [n] : \|\boldsymbol{x}_i - \boldsymbol{y}_i\|^2 > \|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2 \text{ for some } j \neq i\}. \tag{A.1}$$

As another, simpler variant, we may also consider

$$\widetilde{\mathcal{E}}^{\mathsf{dist}} := \{(i, j) \in [n]^2 : \|\boldsymbol{x}_i - \boldsymbol{y}_i\|^2 > \|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2\}, \tag{A.2}$$

which satisfies

$$\frac{1}{n-1}|\widetilde{\mathcal{E}}^{\mathsf{dist}}| \leq |\mathcal{E}^{\mathsf{dist}}| \leq |\widetilde{\mathcal{E}}^{\mathsf{dist}}|. \tag{A.3}$$

By linearity of expectation, we have

$$\mathbb{E}|\widetilde{\mathcal{E}}^{\mathsf{dist}}| = n(n-1) \cdot \mathbb{P}[\|\boldsymbol{x}_1 - \boldsymbol{y}_1\|^2 > \|\boldsymbol{x}_1 - \boldsymbol{y}_2\|^2]$$

Here, we note that $\boldsymbol{x}_1 - \boldsymbol{y}_1 = \boldsymbol{z}_1$ whose squared norm has law $\chi^2(d)$ scaled by σ^2, and is independent from $\boldsymbol{x}_1 - \boldsymbol{y}_2 = \boldsymbol{x}_1 - \boldsymbol{x}_2 - \boldsymbol{z}_2$, whose squared norm has law $\chi^2(d)$ scaled by $2 + \sigma^2$. Thus, we may rewrite this probability in terms of two independent $A, B \sim \chi^2(d)$ as

$$= n(n-1) \cdot \mathbb{P}\left[\frac{A}{B} < \frac{\sigma^2}{2 + \sigma^2}\right]$$

The ratio A/B has the F distribution $F(\frac{d}{2}, \frac{d}{2})$, whose density is $\frac{\Gamma(d)}{\Gamma(d/2)^2} x^{d/2-1}(1+x)^{-d}dx$. Thus, so long as $\sigma^2 = o(1)$, we will have from integrating an initial segment of this density that

$$\lesssim \sigma^d n^2, \tag{A.4}$$

which, up to lower-order terms, is the same as the expected number of augmenting 2-cycles for the MLE. In particular, $\mathbb{E}|\mathcal{E}^{\mathsf{dist}}|$ is bounded by the same quantity and it is reasonable to believe, so long as this is $O(n)$, that this bound is tight. We also expect this first moment computation to be an accurate estimate of the typical size of $\mathcal{E}^{\mathsf{dist}}$. Thus, we find that the error rate of a greedy distance algorithm asymptotically that of the MLE so long as $d = o(\log n)$.

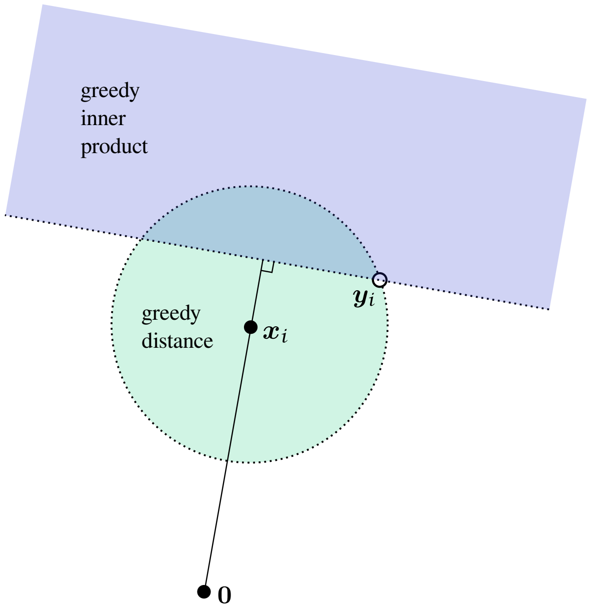

Figure 6: **Greedy algorithm proximal regions.** We illustrate the regions which must not contain any other $\boldsymbol{y}_j$ in order for $\boldsymbol{x}_i$ to be matched with $\boldsymbol{y}_i$ in the two greedy algorithms we consider in Appendix A.

However, once $d = \omega(\log n)$ this algorithm is much less effective than the MLE. In that case, in the critical scaling for the MLE we have $\sigma^2 = \Theta(\frac{d}{\log n}) = \omega(1)$, so $\frac{\sigma^2}{2+\sigma^2} \to 1$ as $n \to \infty$. By evaluating the probability as a Laplace integral, we thus find

$$\begin{aligned} == \mathbb{P}\left[\frac{A}{B} < \frac{\sigma^2}{2+\sigma^2}\right] &\approx \frac{\Gamma(d)}{\Gamma(\frac{d}{2})^2} \exp\left(d\left[\frac{1}{2}\log\left(\frac{\sigma^2}{2+\sigma^2}\right) - \log\left(1 + \frac{\sigma^2}{2+\sigma^2}\right)\right]\right) \\ &\approx \exp\left(d\left[\frac{1}{2}\log\left(\frac{\sigma^2}{2+\sigma^2}\right) - \log\left(\frac{2+2\sigma^2}{2+\sigma^2}\right) + \log 2\right]\right) \\ &= \exp\left(d \log \frac{\sqrt{\sigma^2(2+\sigma^2)}}{1+\sigma^2}\right), \end{aligned} \tag{A.5}$$

whereby

$$\frac{\log(1 \vee \mathbb{E}|\widetilde{\mathcal{E}}^{\mathsf{dist}}|)}{\log n} \approx 2 - \frac{d}{2\log n}\log\frac{(1+\sigma^2)^2}{\sigma^2(2+\sigma^2)} \approx 2 - \frac{d}{2\log n}\log\left(1 + \frac{1}{\sigma^4}\right) \approx 2 - \frac{d}{2\sigma^4 \log n}. \tag{A.6}$$

So, while the MLE achieves perfect recovery for some $\sigma^2 = \Theta(d/\log n)$, the greedy distance algorithm only achieves perfect recovery for the asymptotically smaller $\sigma^2 = O(\sqrt{d/\log n})$.

A more informal way to make the same prediction is to first observe that, for large d, we have the distributional approximation $\chi^2(d) \approx \mathcal{N}(d, 2d)$. Then, when $\sigma^2 \gg 1$ the distances $\|\boldsymbol{x}_i - \boldsymbol{y}_j\|^2$ are distributed approximately as $\mathcal{N}((2+\sigma^2)d, 2(2+\sigma^2)^2 d)$. Likewise the $\|\boldsymbol{x}_i - \boldsymbol{y}_i\|^2 = \|\boldsymbol{z}_i\|^2$ are distributed approximately as $\mathcal{N}(\sigma^2 d, 2\sigma^4 d)$. Moreover, we may make the simplifying assumption of thinking of these distances as independent. Then, we expect strong recovery to only be possible when $\min_{j \neq i} \|\boldsymbol{z}_j\|^2 \approx (2+\sigma^2)d - \sqrt{4(2+\sigma^2)^2 d \log n}$ is at least the typical $\|\boldsymbol{z}_i\|^2 \approx \sigma^2 d$. This gives $(2+\sigma^2)\sqrt{d\log n} \lesssim d$, or $\sigma^2 \lesssim \sqrt{d/\log n}$, as claimed above.

A.2 Algorithm 2: Greedy Inner Product The second algorithm we consider applies the same greedy matching approach to the cost matrix $\boldsymbol{W}$ formed for the MLE by subtracting out the norm terms when the squared distances are expanded. The analogous error set is then

$$\mathcal{E}^{\mathsf{prod}} := \{i \in [n] : \langle \boldsymbol{x}_i, \boldsymbol{y}_i \rangle < \langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle \text{ for some } j \neq i\}. \tag{A.7}$$

Here a useful shortcut allows us to dispense with the low-dimensional case easily: as is apparent from Figure 6, if $i \notin \mathcal{E}^{\mathsf{prod}}$ then $\boldsymbol{y}_i$ is a vertex of the convex hull of $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$. In particular then, this algorithm will only achieve even weak recovery when the convex hull of n i.i.d. standard Gaussian vectors in $\mathbb{R}^d$ has $\Omega(n)$ vertices with high probability. As has been shown in the literature on this so-called *Gaussian polytope* (e.g., [24]), the expected number of vertices is $(\log n)^{O(d)}$, whereby whenever $d = o(\log n / \log\log n)$ the greedy inner product algorithm will not achieve even weak recovery, having $|\mathcal{E}^{\mathsf{prod}}| = n - o(n)$.

On the other hand, when $d = \omega(\log n)$ we believe that the instance $\boldsymbol{W}$ should, loosely speaking, behave in law like a matrix with independent entries (we will say more about how our computations here relate to prior work on such models below). In this case, the "planted" or diagonal and "null" or off-diagonal distributions should be approximately

$$W_{ii} = \langle \boldsymbol{x}_i, \boldsymbol{y}_i \rangle = \|\boldsymbol{x}_i\|^2 + \langle \boldsymbol{x}_i, \boldsymbol{z}_i \rangle \overset{(d)}{\approx} \mathcal{N}(d, \sigma^2 d) =: \mathcal{P}, \tag{A.8}$$

$$W_{ij} = \langle \boldsymbol{x}_i, \boldsymbol{y}_j \rangle = \langle \boldsymbol{x}_i, \boldsymbol{x}_j + \boldsymbol{z}_j \rangle \overset{(d)}{\approx} \mathcal{N}(0, \sigma^2 d) =: \mathcal{Q}, \tag{A.9}$$

where we use that, because $\sigma^2 \gg 1$, we may neglect the fluctuations coming from terms not involving any $\boldsymbol{z}_i$.

If this approximation is sound, then we expect $\max_{j \neq i} W_{ij} \approx \sqrt{2\sigma^2 d \log n}$. On the other hand, $n - o(n)$ of the diagonal terms are of size $\Theta(d)$. Therefore we expect the strong recovery regime to be when $\sqrt{2\sigma^2 d \log n} < d$, or $\sigma^2 < \frac{1}{2}\frac{d}{\log n}$. This is strictly larger than the strong recovery regime $\sigma^2 < \frac{1}{4}\frac{d}{\log n}$ of the MLE. We note that the former threshold $\sigma^2 = \frac{1}{2}\frac{d}{\log n}$ as the threshold we illustrated in Figure 3 when the number of augmenting 2-cycles for the MLE becomes macroscopic.

On the other hand, we also expect $\min_i W_{ii} \approx d - \sqrt{2\sigma^2 d \log n}$, so we expect the *perfect* recovery regime for the greedy inner product algorithm to be when $d - \sqrt{2\sigma^2 d \log n} < \sqrt{2\sigma^2 d \log n}$, or $\sigma^2 < \frac{1}{8}\frac{d}{\log n}$. This is strictly *smaller* than the perfect recovery regime $\sigma^2 < \frac{1}{4}\frac{d}{\log n}$ of the MLE (which is the same as the strong recovery regime of the MLE). Thus the final picture that emerges for the greedy inner product algorithm when $d = \omega(\log n)$ is that it achieves strong recovery for a greater range of σ^2, but has a region of sublinear error $\frac{1}{8}\frac{d}{\log n} < \sigma^2 < \frac{1}{2}\frac{d}{\log n}$, while the MLE has no region of sublinear error on this scale, instead achieving perfect recovery when $\sigma^2 < \frac{1}{4}\frac{d}{\log n}$; from the point of view of the polynomial error rate, the two algorithms are thus incomparable.

A.3 Gaussian Limit in High Dimension The independent Gaussian limit discussed above falls in the range of models treated by previous works [18, 36, 42]. In particular, it was predicted in [36, 42] and proved for certain models (not including the Gaussian model of $\mathcal{P}$ and $\mathcal{Q}$ above) in [18] that the strong recovery threshold in such a model should correspond to $\sqrt{n}B(\mathcal{P}, \mathcal{Q}) = 1$, where $B(\mathcal{P}, \mathcal{Q})$ is the Bhattacharyya coefficient. This may be computed in closed form for Gaussian distributions, which gives that the critical σ^2 should satisfy

$$n = \frac{3 + 2\sigma^2}{2\sqrt{(2+\sigma^2)(1+\sigma^2)}} \exp\left(\frac{d}{2(3+2\sigma^2)}\right). \tag{A.10}$$

As $n \to \infty$ the prefactor and the constant term in the exponent denominator are irrelevant, so this predicts a critical transition at $n = \exp(d/4\sigma^2)$, or $\sigma^2 = \frac{1}{4}\frac{d}{\log n}$.

Per our discussion above, this is the correct strong recovery threshold for the MLE; indeed, the proof of the positive results in [18] goes by analyzing the MLE, so this is not surprising. However, the greedy algorithm applied to this model (to agree with the setting of [18], we should think of the input as the matrix $\boldsymbol{W}$ with entries distributed roughly according to $\mathcal{P}$ and $\mathcal{Q}$, rather than the "raw" point sets $\{\boldsymbol{x}_i\}$ and $\{\boldsymbol{y}_i\}$) achieves a better strong recovery threshold of $\sigma^2 = \frac{1}{2}\frac{d}{\log n}$. Essentially the same is noted in Remark 1 of [18], where the authors bring up a similar independent Gaussian model as an instance where the Bhattacharyya coefficient does *not* give a correct prediction.

Our discussion above, however, gives some further nuance to this point if one is interested in sublinear error rates in addition to just strong recovery. Namely, both in our model for high dimension and in the independent Gaussian model, the greedy algorithm achieves an inferior perfect recovery threshold, and, more generally, an inferior sublinear error rate to the MLE whenever $\sigma^2 < \frac{1}{4}\frac{d}{\log n}$, but a superior rate whenever $\frac{1}{4}\frac{d}{\log n} < \sigma^2 < \frac{1}{2}\frac{d}{\log n}$.

B Evaluation of Integral: Proof of Proposition 2.2

Recall our claim,

$$I(\sigma^2) := \int_0^1 \log\left(1 + \frac{1}{2\sigma^2}(1-\cos(2\pi x))\right) dx = 2\log\left(\frac{1+\sqrt{1+\sigma^{-2}}}{2}\right). \tag{B.11}$$

To lighten the notation, let us set $\lambda = \sigma^2$. Differentiating under the integral sign, we have

$$I'(\lambda) = -\frac{1}{\lambda}\int_0^1 \frac{1-\cos(2\pi x)}{(2\lambda+1)-\cos(2\pi x)} dx$$

which we may write as a contour integral over C the complex unit circle

$$\begin{aligned} &= -\frac{1}{2\pi\lambda}\oint_C \frac{1-\frac{z+z^{-1}}{2}}{(2\lambda+1)-\frac{z+z^{-1}}{2}}\frac{dz}{iz} \\ &= -\frac{1}{2\pi i\lambda}\oint_C \frac{(z-1)^2}{z(z^2-(4\lambda+2)z+1)} dz \end{aligned}$$

where the integrand has poles at $z = 0, \rho_-, \rho_+$ for $\rho_\pm = 2\lambda + 1 \pm 2\sqrt{\lambda^2+\lambda}$. Only $z = 0, \rho_-$ lie inside C, so by the residue theorem we have

$$= -\frac{1}{\lambda}\left(\frac{1}{\rho_+\rho_-} + \frac{(\rho_- - 1)^2}{\rho_-(\rho_- - \rho_+)}\right)$$

which after some algebra reduces to

$$= -\frac{1}{\lambda}\left(1 - \sqrt{\frac{\lambda}{\lambda+1}}\right). \tag{B.12}$$

Since $\lim_{\lambda\to\infty} I(\lambda) = 0$, we then have

$$I(\lambda) = \int_\lambda^\infty \frac{1}{t}\left(1 - \sqrt{\frac{t}{t+1}}\right) dt$$

where the integrand has the explicit antiderivative $\log(t) - \log(1+\sqrt{\frac{t}{t+1}}) + \log(1-\sqrt{\frac{t}{t+1}})$ whose limit as $t \to \infty$ is $-\log(4)$, whereby we finish

$$\begin{aligned} &= -\log(4) - \log(\lambda) + \log\left(1+\sqrt{\frac{\lambda}{\lambda+1}}\right) - \log\left(1-\sqrt{\frac{\lambda}{\lambda+1}}\right) \\ &= \log\left(\frac{1+\sqrt{\frac{\lambda}{1+\lambda}}}{4\lambda(1-\sqrt{\frac{\lambda}{1+\lambda}})}\right) \end{aligned}$$

which after some algebra reduces to

$$= \log\left(\left(\frac{1+\sqrt{1+\lambda^{-1}}}{2}\right)^2\right), \tag{B.13}$$

as claimed.

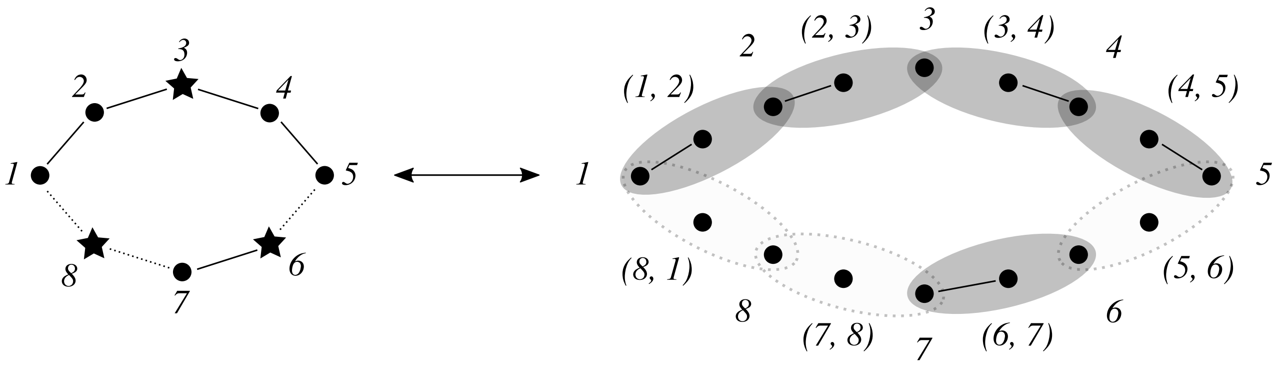

Figure 7: **Spanning forests and matchings.** We illustrate the bijection between rooted spanning forests on the t-cycle and matchings on the $2t$-cycle used in the proof of Lemma C.1.

C Riemann Sum Analysis

In this appendix we prove our bounds on the Riemann sums $S(\sigma^2, t)$. We will proceed by relating $S(\sigma^2, t)$ to the following well-known family of polynomials.

Definition C.1. (Lucas polynomials) *The* Lucas polynomials $L_k(x) \in \mathbb{R}[x]$ *for* $k \geq 0$ *are defined by the recursion*

$$L_0(x) = 2, \tag{C.14}$$

$$L_1(x) = x, \tag{C.15}$$

$$L_k(x) = xL_{k-1}(x) + L_{k-2}(x) \text{ for } k \geq 2. \tag{C.16}$$

The recursion may be solved as follows, a version of the usual approach for a second-order recurrence, only now parametrized by x (see, e.g., [23]).

Proposition C.1. (Binet's formula) *Let* $\alpha(x), \beta(x)$ *be the roots of* $t^2 - tx - 1$, *i.e.,*

$$\alpha(x) = \frac{x + \sqrt{x^2+4}}{2}, \tag{C.17}$$

$$\beta(x) = \frac{x - \sqrt{x^2+4}}{2}. \tag{C.18}$$

Then, $L_n(x) = \alpha(x)^n + \beta(x)^n$.

The following is then the key statement relating the Lucas polynomials to our Riemann sums.

Lemma C.1. *For any* $t \geq 3$, $\exp(S(\sigma^2, t)) = (4\sigma^2)^{-t}(L_{2t}(2\sigma) - 2)$.

We note that the same formula does not hold for $t = 2$: the left-hand side is $1 + \sigma^{-2}$, while the right-hand side is $1 + \sigma^{-2}/4$. As we will see in the course of the proof, that is because the formula depends on the eigenvalues of the t-cycle graph C_3 appearing in the summation in $S(\sigma^2, t)$.

Proof. [Proof of Lemma C.1] Recall that we denote by C_t the cycle on t vertices and by $\boldsymbol{L}^{C_t}$ its graph Laplacian. Let $\lambda_{t,1}, \dots, \lambda_{t,t}$ denote the eigenvalues of $\boldsymbol{L}^{C_t}$. Then, we have

$$\exp(S(\sigma^2, t)) = \prod_{j=1}^{t}\left(1 + \frac{1}{4\sigma^2}\lambda_{t,j}\right) = \sum_{k=0}^{t} E_k(\lambda_{t,1}, \dots, \lambda_{t,t})(4\sigma^2)^{-k}, \tag{C.19}$$

where

$$E_k(a_1, \dots, a_t) := \sum_{1 \leq i_1 < \dots < i_k \leq t} a_{i_1} \cdots a_{i_k} \tag{C.20}$$

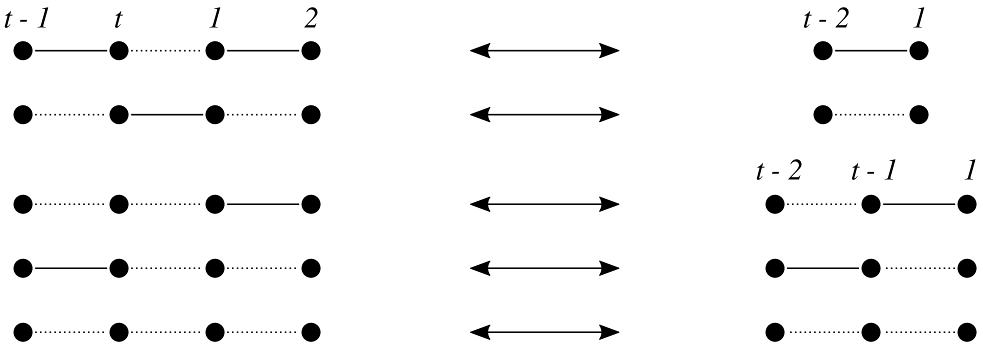

Figure 8: **Contracting matchings in cycles.** We illustrate the bijection between matchings in a t-cycle and those in either a $(t-1)$- or $(t-2)$-cycle used in the proof of Lemma C.1.

are the elementary symmetric polynomials.

By a generalization of the matrix-tree theorem (see Theorem 7.5 of [13]), $E_k(\lambda_{t,1}, \dots, \lambda_{t,t})$ is equal to the number of spanning forests of C_t containing $t-k$ connected components and with each connected component having an assigned root vertex. The condition of a spanning forest having $t-k$ connected components is, for the specific case of the graph C_t, also equivalent to the forest containing k edges.

When $k < t$, then these forests are in bijection with the matchings on C_{2t} also containing k edges. An explicit bijection is as follows. Suppose the vertices of C_{2t} are labelled $0, \dots, 2t-1$ and the vertices of C_t labelled $0, \dots, t-1$. Suppose M is a matching in C_{2t}. We build a spanning forest F of C_t by including the edge $\{i, i+1\}$ whenever either $\{2i, 2i+1\}$ or $\{2i+1, 2i+2\}$ is included in M, and by declaring i a root vertex in C_t if $2i$ is not adjacent to any edges of M in C_{2t}. We illustrate this mapping in Figure 7, which shows that every connected component of F formed this way indeed has a unique root vertex (located where the pairs of consecutive edges containing edges of M switch from "leaning" counterclockwise to clockwise), and that the knowledge of the connected components and the root vertices of F uniquely determines the preimage M. This holds so long as $k < t$; however, when $k = t$, then there are two matchings on C^{2t} with $k = t$ edges, while $E_t(\lambda_{t,1}, \dots, \lambda_{t,t}) = \det(\boldsymbol{L}^{C_t}) = 0$.

Let $M_{t,k}$ denote the number of matchings of k edges in C_t. The result then follows from showing that $L_t(x)$ is the *matching polynomial* of C_t for $t \geq 3$:

$$L_t(x) = \sum_{k=0}^{\lfloor t/2 \rfloor} M_{t,k} x^{t-2k}. \tag{C.21}$$

This fact is known, though often phrased differently (e.g., Section 6 of [21]), but we give the simple proof here for the sake of completeness. The statement is easily verified for $t = 3, 4$, and then it suffices to show the coefficient recursion for $t \geq 5$ that

$$M_{t,k} = M_{t-1,k} + M_{t-2,k-1}. \tag{C.22}$$

We present a bijective proof of this recursion in Figure 8: fixing a sequence of three consecutive edges in C_t, we map any matching in C_t to a matching in either C_{t-1} or C_{t-2} by a suitable replacement of this sequence by either two edges or one edge, and this mapping is visibly a bijection. □

C.1 Discrete Concavity: Proof of Lemma 2.1 It suffices to show that $S(\sigma^2, t) - S(\sigma^2, t-1)$ is decreasing in t. (This establishes that the limit $\lim_{t\to\infty}(S(\sigma^2, t) - S(\sigma^2, t-1))$ exists, and since $S(\sigma^2, t)/t \to I(\sigma^2)$, the former limit must then equal $I(\sigma^2)$.) Equivalently, it suffices to show that, for all $t \geq 3$,

$$2S(\sigma^2, t) \overset{?}{>} S(\sigma^2, t-1) + S(\sigma^2, t+1). \tag{C.23}$$

We consider separately the case $t = 3$. Introducing for the sake of convenience a new variable $y = \sigma^{-2}$, we have the polynomials

$$\exp(S(y^{-1},2)) = 1 + y\sin^2\left(\frac{1}{2}\pi\right)$$

$$= 1 + y, \tag{C.24}$$

$$\exp(S(y^{-1},3)) = \left(1 + y\sin^2\left(\frac{1}{3}\pi\right)\right)\left(1 + y\sin^2\left(\frac{2}{3}\pi\right)\right)$$

$$= \left(1 + \frac{3}{4}y\right)^2, \tag{C.25}$$

$$\exp(S(y^{-1},4)) = \left(1 + y\sin^2\left(\frac{1}{4}\pi\right)\right)\left(1 + y\sin^2\left(\frac{2}{4}\pi\right)\right)\left(1 + y\sin^2\left(\frac{3}{4}\pi\right)\right)$$

$$= \left(1 + \frac{1}{2}y\right)^2(1+y) \tag{C.26}$$

Thus it suffices to show that, for all $y > 0$,

$$\left(1 + \frac{3}{4}y\right)^4 \overset{?}{>} \left(1 + \frac{1}{2}y\right)^2(1+y)^2, \tag{C.27}$$

which follows by the arithmetic-geometric mean inequality which gives $(1 + \frac{1}{2}y)(1+y) < (\frac{1}{2}(1 + \frac{1}{2}y + 1 + y))^2 = (1 + \frac{3}{4}y)^2$.

Now, suppose $t \geq 4$. Then, exponentiating both sides and applying Lemma C.1, it suffices to show that

$$(L_{2t}(x) - 2)^2 \overset{?}{>} (L_{2t-2}(x) - 2)(L_{2t+2}(x) - 2) \tag{C.28}$$

for all $x > 0$. Letting $a = \alpha(x)^2 > 1$, we have $\beta(x)^2 = a^{-1}$. In terms of a, we may then expand either side as

$$(L_{2t}(x) - 2)^2 = 6 + a^{2t} + a^{-2t} - 4a^t - 4a^{-t}, \tag{C.29}$$

$$(L_{2t-2}(x) - 2)(L_{2t+2}(x) - 2) = 4 + a^{2t} + a^{-2t} + a^2 + a^{-2} - 2a^{t-1} - 2a^{-t+1} - 2a^{t+1} - 2a^{-t-1}. \tag{C.30}$$

Therefore, it suffices to show that, for all $a > 0$,

$$0 \overset{?}{<} 2 - 4a^t - 4a^{-t} - a^2 - a^{-2} + 2a^{t-1} + 2a^{-t+1} + 2a^{t+1} + 2a^{-t-1}$$

$$= 2 + 2a^{t-1}(1-a)^2 + 2a^{-t+1}(1-a^{-1})^2 - a^2 - a^{-2}. \tag{C.31}$$

Viewing t for a moment as a continuous parameter, we note that the derivative of the above expression with respect to t is $2\log a(a^{t-1}(1-a)^2 - a^{-t+1}(1-a^{-1})^2) > 0$ for any $t > 1$ and $a > 1$, since $a^{t-1} > a^{-t+1}$ and $(1-a)^2 > (1-a^{-1})^2 = (\frac{a-1}{a})^2$. Thus this expression is increasing in t, so it suffices to consider $t = 4$. In that case, we have the factorization

$$2 + 2a^3(1-a)^2 + 2a^{-3}(1-a^{-1})^2 - a^2 - a^{-2} = \frac{(a-1)^4(2a^6 + 4a^5 + 6a^4 + 7a^3 + 6a^2 + 4a + 2)}{a^5}, \tag{C.32}$$

which shows strict positivity for any $a > 1$.

C.2 Upper Bound: Proof of Lemma 2.2 We want to show $S(\sigma^2, t) < tI(\sigma^2)$. Exponentiating either side, we observe that

$$\exp(tI(\sigma^2)) = \left(\frac{1+\sqrt{1+\sigma^{-2}}}{2}\right)^{2t}, \tag{C.33}$$

$$\begin{aligned}
\exp(S(\sigma^2, t)) &= (4\sigma^2)^{-t}\left(\alpha(\sqrt{4\sigma^2})^{2t} + \beta(\sqrt{4\sigma^2})^{2t} - 2\right) \\
&= (4\sigma^2)^{-t}\left(\left(\frac{\sqrt{4\sigma^2}+\sqrt{4\sigma^2+4}}{2}\right)^{2t} + \left(\frac{\sqrt{4\sigma^2}-\sqrt{4\sigma^2+4}}{2}\right)^{2t} - 2\right) \\
&= \left(\frac{1+\sqrt{1+\sigma^{-2}}}{2}\right)^{2t} + \left(\frac{1-\sqrt{1+\sigma^{-2}}}{2}\right)^{2t} - 2(4\sigma^2)^{-t}
\end{aligned}$$

$$= \exp(tI(\sigma^2)) + \left(\frac{1-\sqrt{1+\sigma^{-2}}}{2}\right)^{2t} - \left(\frac{2^{1/t}}{4\sigma^2}\right)^{t}. \tag{C.34}$$

Thus it suffices to show that

$$\frac{2^{1/t}}{4\sigma^2} \overset{?}{>} \left(\frac{\sqrt{1+\sigma^{-2}}-1}{2}\right)^2, \tag{C.35}$$

which we verify as

$$\frac{2^{1/t}}{4\sigma^2} - \left(\frac{\sqrt{1+\sigma^{-2}}-1}{2}\right)^2 > \frac{1}{4\sigma^2} - \left(\frac{\sqrt{1+\sigma^{-2}}-1}{2}\right)^2 = \frac{1}{2}\left(\sqrt{1+\sigma^{-2}}-1\right) > 0. \tag{C.36}$$

C.3 Miscellaneous Rate Functions: Proof of Proposition 4.6 Again introducing $y = \sigma^{-2}$ and using our expressions for S_t for $t = 2, 3$, and 4 from (C.24), (C.25), and (C.26), respectively, as well as the expression

$$\exp(I(y^{-1})) = \left(\frac{1+\sqrt{1+y}}{2}\right)^2 \in \left[1+\frac{y}{4}, 1+\frac{y}{2}\right], \tag{C.37}$$

we may compute as follows:

$$\exp(\eta_1) = \frac{(1+y)^{3/4}}{(1+\frac{1}{2}y)^{1/2}(1+y)^{1/4}}$$

$$= \sqrt{\frac{1+y}{1+\frac{1}{2}y}}, \tag{C.38}$$

$$\exp(\eta_2) = \frac{1+y}{1+\frac{3}{4}y}, \tag{C.39}$$

$$\exp(\eta_3) = \frac{2\sqrt{1+y}}{1+\sqrt{1+y}}. \tag{C.40}$$

Thus we find $\eta_3 \geq \eta_1$, since by concavity of the square root $\frac{1}{2}(1+\sqrt{1+y}) \leq \sqrt{1+\frac{1}{2}y}$. We also have, by the arithmetic-geometric mean inequality,

$$\frac{\exp(\eta_1)}{\exp(\eta_2)} = \frac{1+\frac{3}{4}y}{\sqrt{(1+y)(1+\frac{1}{2}y)}} \geq 1, \tag{C.41}$$

so $\eta_3 \geq \eta_1 \geq \eta_2$. For the upper bound on η_3, we have

$$\begin{aligned}
\exp(\eta_3) &\leq \sqrt{\frac{1+y}{1+\frac{1}{4}y}} \\
&= \sqrt{1+\frac{\frac{3}{4}y}{1+\frac{1}{4}y}} \\
&\leq \exp\left(\frac{\frac{3}{4}y}{2+\frac{1}{2}y}\right) \\
&= \exp\left(\frac{1}{\frac{2}{3}+\frac{8}{3}\sigma^2}\right). \qquad \text{(C.42)}
\end{aligned}$$

D Edge Probability Prefactor: Proof of Proposition 4.2

We begin by producing the following formula for p: let $g \sim \mathcal{N}(0,\sigma^2)$ and $u \sim \chi^2(d)$ be independent. Then,

$$p = \mathbb{P}\left[g \geq \sqrt{u}\right]. \tag{D.43}$$

We work directly from the earlier expression, in the special case $t = 2$:

$$\begin{aligned}
\mathbb{P}[(1,2) \text{ is augmenting}] &= \mathbb{P}\left[\langle \boldsymbol{z}_1 - \boldsymbol{z}_2, \boldsymbol{x}_2 - \boldsymbol{x}_1 \rangle \geq \|\boldsymbol{x}_2 - \boldsymbol{x}_1\|^2\right] \\
&= \mathbb{P}\left[\left\langle \frac{\boldsymbol{z}_1 - \boldsymbol{z}_2}{\sqrt{2}}, \frac{\boldsymbol{x}_2 - \boldsymbol{x}_1}{\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|} \right\rangle \geq \left\|\frac{\boldsymbol{x}_2 - \boldsymbol{x}_1}{\sqrt{2}}\right\|\right], \qquad \text{(D.44)}
\end{aligned}$$

where we observe now that $(\boldsymbol{x}_2 - \boldsymbol{x}_1)/\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|$ has the law of a uniformly distributed unit vector, and is independent from $\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|/\sqrt{2}$, which has the law of the norm of a standard gaussian vector, which is that of $\sqrt{u}$. Moreover, $(\boldsymbol{z}_1 - \boldsymbol{z}_2)/\sqrt{2}$ has the law $\mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_d)$, so the left-hand side of the probability has law $\mathcal{N}(0,\sigma^2)$, giving the claim.

Note that the upper bound on $\hat{p}$ follows from Proposition 3.1. For the lower bounds, we first give two quantitative lower bounds, a looser one that holds for all $d \geq 1$ and a tighter one that holds for all $d \geq 4$. We will use the following "Mills' ratio" lower bounds on Gaussian tails (see, e.g., [19]): for all $t > 0$,

$$\mathop{\mathbb{P}}_{g\sim\mathcal{N}(0,1)}[g \geq t] \geq \frac{t}{1+t^2}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right) \geq \left(\frac{1}{t} - \frac{1}{t^3}\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{t^2}{2}\right) \tag{D.45}$$

For our first lower bound, we use the first lower bound of (D.45):

$$\begin{aligned}
p &= \mathop{\mathbb{E}}_{u\sim\chi^2(d)} \mathop{\mathbb{P}}_{g\sim\mathcal{N}(0,1)}\left[g \geq \sqrt{\frac{u}{\sigma^2}}\right] \\
&\geq \sqrt{\frac{\sigma^2}{2\pi}} \mathop{\mathbb{E}}_{u} \frac{\sqrt{u}}{\sigma^2 + u} \exp\left(-\frac{u}{2\sigma^2}\right) \\
&= \sqrt{\frac{\sigma^2}{2\pi}} \frac{1}{2^{d/2}\Gamma(\frac{d}{2})} \int_0^\infty \frac{u^{\frac{d-1}{2}}}{\sigma^2 + u} \exp\left(-\frac{1+\sigma^{-2}}{2}u\right) du \\
&= \sqrt{\frac{\sigma^2}{2\pi}} \frac{1}{2^{d/2}\Gamma(\frac{d}{2})} \left(\frac{1+\sigma^{-2}}{2}\right)^{-\frac{d-1}{2}} \int_0^\infty \frac{v^{\frac{d-1}{2}}}{\frac{1+\sigma^2}{2} + v} e^{-v} dv \\
&= \exp\left(-\frac{d}{2}S(\sigma^2, 2)\right) \frac{\sqrt{1+\sigma^2}}{2\sqrt{\pi}} \frac{1}{\Gamma(\frac{d}{2})} \int_0^\infty \frac{v^{\frac{d-1}{2}}}{\frac{1+\sigma^2}{2} + v} e^{-v} dv. \qquad \text{(D.46)}
\end{aligned}$$

Working now with the remaining integral,

$$\begin{aligned}\int_0^\infty \frac{v^{\frac{d-1}{2}}}{\frac{1+\sigma^2}{2}+v}e^{-v}dv &\geq \frac{1}{1+\sigma^2}\int_0^\infty \frac{v^{\frac{d-1}{2}}}{1+v}e^{-v}dv \\ &\geq \frac{1}{1+\sigma^2}\int_1^\infty \frac{v^{\frac{d-1}{2}}}{2v}e^{-v}dv \\ &= \frac{1}{2(1+\sigma^2)}\Gamma\left(\frac{d-1}{2},1\right),\end{aligned} \tag{D.47}$$

and we find that, for all $d \geq 1$,

$$\widehat{p} \geq \frac{1}{4\sqrt{\pi(1+\sigma^2)}}\frac{\Gamma(\frac{d-1}{2},1)}{\Gamma(\frac{d}{2})}$$

and, bounding the Γ functions,

$$\geq \frac{1}{40\sqrt{\pi d(1+\sigma^2)}}. \tag{D.48}$$

For our second lower bound, we suppose that $d \geq 4$ and use the second lower bound of (D.45):

$$\begin{aligned}p &\geq \frac{1}{\sqrt{2\pi}}\mathop{\mathbb{E}}_{u}\left(\left(\left(\frac{\sigma^2}{u}\right)^{1/2}-\left(\frac{\sigma^2}{u}\right)^{3/2}\right)\exp\left(-\frac{u}{2\sigma^2}\right)\right) \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{2^{d/2}\Gamma(\frac{d}{2})}\left(\sigma\int_0^\infty u^{\frac{d-3}{2}}\exp\left(-\frac{1+\sigma^{-2}}{2}u\right)du-\sigma^3\int_0^\infty u^{\frac{d-5}{2}}\exp\left(-\frac{1+\sigma^{-2}}{2}u\right)du\right)\end{aligned}$$

where both integrals converge due to our assumption that $d \geq 4$. Performing the same change of various as before, we find

$$= \frac{1}{\sqrt{2\pi}}\frac{1}{\Gamma(\frac{d}{2})}\left(\frac{\Gamma(\frac{d-1}{2})}{\sqrt{2}}(1+\sigma^2)^{1/2}-\frac{\Gamma(\frac{d-3}{2})}{2\sqrt{2}}(1+\sigma^2)^{3/2}\right)\exp\left(-\frac{d}{2}S(\sigma^2,2)\right), \tag{D.49}$$

and thus, rearranging, we find that

$$\widehat{p} \geq \frac{1}{2\sqrt{\pi}}\cdot\frac{\Gamma(\frac{d-1}{2})}{\Gamma(\frac{d}{2})}(\sigma^2+1)^{\frac{1}{2}}-\frac{1}{4\sqrt{\pi}}\cdot\frac{\Gamma(\frac{d-3}{2})}{\Gamma(\frac{d}{2})}(\sigma^2+1)^{3/2}$$

and bounding the Γ function ratios from above and below,

$$\geq \frac{1}{\sqrt{2\pi}}\left(\frac{1+\sigma^2}{d}\right)^{\frac{1}{2}}-2\left(\frac{1+\sigma^2}{d}\right)^{3/2} \tag{D.50}$$

For $1 \leq d \leq 40$, by assumption we have $\sigma^2 \leq 1$, so by our first bound we find

$$\widehat{p} \geq \frac{1}{40\sqrt{80\pi}} \geq \frac{1}{1000}\sqrt{\frac{1+\sigma^2}{d}}, \tag{D.51}$$

using that $1+\sigma^2 \leq 2$ and $d \geq 1$. For $d \geq 40$, we have $1 \leq \frac{d}{40}$, so $1+\sigma^2 \leq \frac{d}{20}$. On the interval $x \in [0, \frac{1}{20}]$ we have $\frac{1}{\sqrt{2\pi}}x^{1/2}-2x^{3/2} \geq \frac{1}{4}x^{1/2}$, so by our second bound we have

$$\widehat{p} \geq \frac{1}{4}\sqrt{\frac{1+\sigma^2}{d}}. \tag{D.52}$$

Combining the two cases gives the result.

References

[1] E. Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

[2] M. Ajtai, J. Komlós, and G. Tusnády. On optimal matchings. *Combinatorica*, 4(4):259–264, 1984.

[3] D. Aldous. Asymptotics in the random assignment problem. *Probability Theory and Related Fields*, 93(4): 507–534, 1992.

[4] D. J. Aldous. The $\zeta(2)$ limit in the random assignment problem. *Random Structures & Algorithms*, 18(4): 381–418, 2001.

[5] N. Alon, M. Krivelevich, and B. Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3-4):457–466, 1998.

[6] L. Ambrosio, F. Stra, and D. Trevisan. A PDE approach to a 2-dimensional matching problem. *Probability Theory and Related Fields*, 173(1-2):433–477, 2019.

[7] J. Aronson, A. Frieze, and B. G. Pittel. Maximum matchings in sparse random graphs: Karp-Sipser revisited. *Random Structures & Algorithms*, 12(2):111–177, 1998.

[8] P. Awasthi, A. S. Bandeira, M. Charikar, R. Krishnaswamy, S. Villar, and R. Ward. Relax, no need to round: Integrality of clustering formulations. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pages 191–200, 2015.

[9] A. S. Bandeira, A. Perry, and A. S. Wein. Notes on computational-to-statistical gaps: predictions using statistical physics. *arXiv preprint arXiv:1803.11132*, 2018.

[10] B. Barak, S. Hopkins, J. Kelner, P. K. Kothari, A. Moitra, and A. Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.

[11] D. P. Bertsekas. The auction algorithm for assignment and other network flow problems: A tutorial. *Interfaces*, 20(4):133–149, 1990.

[12] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft. Simple online and realtime tracking. In *2016 IEEE international conference on image processing (ICIP)*, pages 3464–3468. IEEE, 2016.

[13] N. Biggs. *Algebraic graph theory*. Cambridge University Press, 1993.

[14] S. Caracciolo, C. Lucibello, G. Parisi, and G. Sicuro. Scaling hypothesis for the Euclidean bipartite matching problem. *Physical Review E*, 90(1):012118, 2014.

[15] M. Chertkov, L. Kroc, F. Krzakala, M. Vergassola, and L. Zdeborová. Inference in particle tracking experiments by passing messages between images. *Proceedings of the National Academy of Sciences*, 107(17): 7663–7668, 2010.

[16] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in Neural Information Processing Systems*, 26:2292–2300, 2013.

[17] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová. Inference and phase transitions in the detection of modules in sparse networks. *Physical Review Letters*, 107(6):065701, 2011.

[18] J. Ding, Y. Wu, J. Xu, and D. Yang. The planted matching problem: Sharp threshold and infinite-order phase transition. *arXiv preprint arXiv:2103.09383*, 2021.

[19] L. Duembgen. Bounding standard gaussian tail probabilities. *arXiv preprint arXiv:1012.2063*, 2010.

[20] M. Dyer, A. Frieze, and B. Pittel. The average performance of the greedy matching algorithm. *The Annals of Applied Probability*, pages 526–552, 1993.

[21] E. J. Farrell. An introduction to matching polynomials. *Journal of Combinatorial Theory, Series B*, 27(1): 75–86, 1979.

[22] A. M. Frieze. On the independence number of random graphs. *Discrete Mathematics*, 81(2):171–175, 1990.

[23] A. Horadam and J. Mahon. Pell and Pell-Lucas polynomials. *The Fibonacci Quarterly*, 23(1):7–20, 1985.

[24] D. Hug, G. O. Munsonius, and M. Reitzner. Asymptotic mean values of Gaussian polytopes. *Beiträge Algebra Geometry*, 45(2):531–548, 2004.

[25] T. Iguchi, D. G. Mixon, J. Peterson, and S. Villar. Probably certifiably correct k-means clustering. *Mathematical Programming*, 165(2):605–642, 2017.

[26] M. Jerrum. Large cliques elude the Metropolis process. *Random Structures & Algorithms*, 3(4):347–359, 1992.

[27] R. M. Karp and M. Sipser. Maximum matching in sparse random graphs. In *22nd Annual Symposium on Foundations of Computer Science (SFCS 1981)*, pages 364–375. IEEE, 1981.

[28] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2 (1-2):83–97, 1955.

[29] M. Ledoux. On optimal matching of Gaussian samples II, 2018.

[30] M. Ledoux. On optimal matching of Gaussian samples. *Journal of Mathematical Sciences*, 238(4):495–522, Apr 2019.

[31] T. Leighton and P. Shor. Tight bounds for minimax grid matching with applications to the average case analysis of algorithms. *Combinatorica*, 9(2):161–187, 1989.

[32] X. Li, Y. Li, S. Ling, T. Strohmer, and K. Wei. When do birds of a feather flock together? k-means, proximity, and conic programming. *Mathematical Programming*, 179(1):295–341, 2020.

[33] C. McDiarmid. On the method of bounded differences. *Surveys in Combinatorics*, 141(1):148–188, 1989.

[34] M. Mézard and G. Parisi. On the solution of the random link matching problems. *Journal de Physique*, 48 (9):1451–1459, 1987.

[35] D. G. Mixon, S. Villar, and R. Ward. Clustering subgaussian mixtures by semidefinite programming. *Information and Inference: A Journal of the IMA*, 6(4):389–415, 2017.

[36] M. Moharrami, C. Moore, and J. Xu. The planted matching problem: phase transitions and exact results. *arXiv preprint arXiv:1912.08880*, 2019.

[37] C. Moore. The computer science and physics of community detection: landscapes, phase transitions, and hardness. *arXiv preprint arXiv:1702.00467*, 2017.

[38] B. Nica. A brief introduction to spectral graph theory. *arXiv preprint arXiv:1609.08072*, 2016.

[39] G. Parisi. A conjecture on random bipartite matching. *arXiv preprint cond-mat/9801176*, 1998.

[40] A. A. Perera, C. Srinivas, A. Hoogs, G. Brooksby, and W. Hu. Multi-object tracking through simultaneous long occlusions and split-merge conditions. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006)*, volume 1, pages 666–673. IEEE, 2006.

[41] B. Sahbani and W. Adiprawita. Kalman filter and iterative-Hungarian algorithm implementation for low complexity point tracking as part of fast multiple object tracking system. In *6th International Conference on System Engineering and Technology (ICSET 2016)*, pages 109–115. IEEE, 2016.

[42] G. Semerjian, G. Sicuro, and L. Zdeborová. Recovery thresholds in the sparse planted matching problem. *Physical Review E*, 102(2):022304, 2020.

[43] E. Shamir and J. Spencer. Sharp concentration of the chromatic number on random graphs $G_{n,p}$. *Combinatorica*, 7(1):121–129, 1987.

[44] P. W. Shor and J. E. Yukich. Minimax grid matching and empirical measures. *The Annals of Probability*, 19(3):1338–1348, 1991.

[45] M. Talagrand. *Upper and lower bounds for stochastic processes: modern methods and classical problems*, volume 60 of *Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge (A Series of Modern Surveys in Mathematics)*. Springer, Heidelberg, 2014.

[46] M. Talagrand. Scaling and non-standard matching theorems. *Comptes Rendus Mathematique*, 356(6):692–695, 2018.

[47] L. Zdeborová and M. Mézard. The number of matchings in random graphs. *Journal of Statistical Mechanics: Theory and Experiment*, 2006(05):P05003, 2006.