PHASE TRANSITIONS FOR THE UNIFORM DISTRIBUTION IN THE PATTERN MAXIMUM LIKELIHOOD PROBLEM AND ITS BETHE APPROXIMATION*

CHUN LAM CHAN†, WINSTON FERNANDES‡, NAVIN KASHYAP§, AND MANJUNATH KRISHNAPUR¶

Abstract. The pattern maximum likelihood (PML) estimate, introduced by Orlitsky et al., is an estimate of the multiset of probabilities in an unknown probability distribution $\mathbf{p}$, the estimate being obtained from n independent and identically distributed samples drawn from $\mathbf{p}$. The PML estimate involves solving a difficult optimization problem over the set of all probability mass functions of finite support. In this paper, we describe an interesting phase transition phenomenon in the PML estimate: at a certain sharp threshold, the uniform distribution goes from being a local maximum to being a local minimum for the optimization problem in the estimate. We go on to consider the question of whether a similar phase transition phenomenon also exists in the Bethe approximation of the PML estimate, the latter being an approximation method with origins in statistical physics. We show that the answer to this question is a qualified "yes." Our analysis involves the computation of the mean and variance of the (i, j)th entry, $a_{i,j}$, in a random $k \times k$ nonnegative integer matrix A with row and column sums all equal to M, drawn according to a distribution that assigns to A a probability proportional to $\Pi_{i,j} \frac{(M - a_{i,j})!}{a_{i,j}!}$.

1. Introduction. Consider the estimation problem in which, given a sequence of n independent and identically distributed (i.i.d.) samples from a fixed but unknown probability distribution $\mathbf{p}$ supported on some discrete alphabet, we are required to estimate the multiset of probabilities in $\mathbf{p}$. In particular, we need not determine the correspondence between the symbols of the underlying alphabet and the probabilities in the multiset. Such a problem arises naturally in the context of universal compression of large-alphabet sources [1] and has several other applications, for example, population estimation from a small number of samples [2]. The multiset of empirical frequencies of the symbols observed in the n samples is a straightforward estimate of the multiset of probabilities in $\mathbf{p}$; this estimate corresponds to the maximum likelihood (ML) estimate of $\mathbf{p}$. However, when the sample size, n, is smaller than the size of the support of the underlying distribution $\mathbf{p}$, the ML estimate may not give a good estimate of the multiset of probabilities in $\mathbf{p}$. An alternative estimate that has been proposed for this regime is the pattern maximum likelihood (PML) estimate, introduced by Orlitsky et al. [1], [2] and described below.

---

†Department of Information Engineering, Chinese University of Hong Kong, Hong Kong (clchan.eric@gmail.com).
‡Dell R&D, Bangalore, India (winfernu@gmail.com).
§Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, India (nkashyap@ece.iisc.ernet.in).
¶Department of Mathematics, Indian Institute of Science, Bangalore, India (manju@math.iisc.ernet.in).

The *pattern* $\boldsymbol{\psi}$ or $\boldsymbol{\psi}(\mathbf{x}^n)$ of a sequence $\mathbf{x}^n = x_1, \ldots, x_n$ is a data structure that keeps track of the order of occurrence and the multiplicities of the distinct symbols in the sequence $\mathbf{x}^n$; for a precise definition, see section 2. The *PML distribution* of a pattern $\boldsymbol{\psi}$ is the multiset of probabilities that maximizes the probability of observing a sequence with pattern $\boldsymbol{\psi}$. It has been argued [2], [3], [4] that for a sequence $\mathbf{x}^n$ sampled from an unknown underlying probability distribution $\mathbf{p}$, the PML distribution of $\boldsymbol{\psi}(\mathbf{x}^n)$ is a good estimate of the multiset of probabilities in $\mathbf{p}$, even in situations where n is much smaller than the support size of $\mathbf{p}$. However, for the purposes of this paper, we view the PML distribution purely as an interesting mathematical object.

The problem of determining the PML distribution (henceforth termed the "PML problem") of a given pattern $\boldsymbol{\psi}$ appears to be computationally hard [4], [5], [6], [7], [8], [9]. It turns out that the PML problem can be very well approximated by its Bethe approximation [10], [11], a technique with roots in statistical physics. The optimization problem in the Bethe approximation can usually be solved highly efficiently using belief propagation algorithms [12].

In this paper, we are concerned with a remarkable phase transition phenomenon[1] observed in the PML problem. For a positive integer k, let U_k denote the uniform distribution on k symbols. Given a pattern $\boldsymbol{\psi}$, we can explicitly compute a quantity $\Upsilon(\boldsymbol{\psi})$ such that for all $k < \Upsilon(\boldsymbol{\psi})$, U_k is a *local maximum*, among all distributions $\mathbf{p}$ with support size k, for the optimization problem within the PML problem, and for all $k > \Upsilon(\boldsymbol{\psi})$, U_k is a *local minimum*. On the basis of this observation, we proposed in [13] a heuristic algorithm for determining whether the uniform distribution is the PML distribution of a given pattern $\boldsymbol{\psi}$.

Given that the Bethe approximation is a very good proxy for the PML distribution, it is natural to ask whether the phase transition phenomenon described above extends to the Bethe approximation as well. We are able to give a qualified affirmative answer to this question. Our answer is given in terms of a sequence of "degree-M optimization problems" such that the degree-1 problem is the original PML problem, and as $M \to \infty$, we obtain the Bethe approximation. We show that for all sufficiently large M, the degree-M optimization problems admit a phase transition phenomenon very similar to that described above for the PML problem. While this falls just short of proving that the Bethe approximation itself admits such a phase transition, it lends strong support in favor of this assertion.

The bulk of our proof of the existence of phase transitions in the degree-M optimization problems involves analyzing a certain probability distribution, denoted by $Q_{k,M}$, on the set of $k \times k$ nonnegative integer matrices with all row and column sums equal to M. This probability distribution and our analysis of it via a discrete Gaussian approximation may be of independent interest.

The remainder of this paper is organized as follows. In section 2, we provide the definitions needed to describe the PML problem, after which we state and prove the corresponding phase transition phenomenon (Theorem 2.1). The Bethe approximation is described in section 3. This section also explains the notion of "degree-M lifted permanents" defined by Vontobel [14], which is used to define our degree-M optimization problems. Section 4 contains a precise statement (Theorem 4.1) of the phase transition phenomenon in the degree-M problems, the proof of which occupies much of the rest of the paper. In particular, section 5 collects together the prop-

---

[1]Our use of the term "phase transition" here is inspired by statistical physics, where the term is often used to describe abrupt changes in behavior of physical (especially, thermodynamical) systems.

erties of the probability distribution $Q_{k,M}$ that are used in the proof. The paper concludes in section 6 with a discussion of the gap remaining in a rigorous proof of the phase transition phenomenon in the Bethe approximation. Some of the more technical proofs from sections 3–5 are presented in appendices.

2. The PML problem and a phase transition phenomenon. Throughout this paper, we use $\mathbb{Z}_+$ to denote the set of nonnegative integers and $\mathbb{R}_+$ to denote the set of nonnegative reals. For a positive integer k, we use $[k]$ to denote the set $\{1, 2, \ldots, k\}$. For any countable set $\mathcal{X}$, we let $\Pi_{\mathcal{X}}$ denote the set of all probability distributions on $\mathcal{X}$:

$$\Pi_{\mathcal{X}} := \left\{ \mathbf{p} = (p(x))_{x \in \mathcal{X}} : \ p(x) \geq 0 \ \forall x \in \mathcal{X}, \ \sum_{x \in \mathcal{X}} p(x) = 1 \right\}.$$

For any positive integer k, we let U_k denote the uniform distribution on $[k]$.

2.1. Patterns and PML. Given a sequence $\mathbf{x}^n = x_1, \ldots, x_n$ over some alphabet, the *pattern* of $\mathbf{x}^n$ is the sequence $\boldsymbol{\psi} = \psi_1, \psi_2, \ldots, \psi_n$ obtained by replacing each x_j by the order of its first occurrence in $\mathbf{x}^n$ [4], [10]. More precisely, for each symbol x occurring in $\mathbf{x}^n$, let $\nu(x)$ denote the number of distinct symbols seen in the *shortest* prefix of $\mathbf{x}^n$ that ends in the symbol x. Then, $\psi_j = \nu(x_j)$ for $j = 1, 2, \ldots, n$. The pattern $\boldsymbol{\psi}(\mathbf{x}^n)$ is defined to have *length* n and *size* m, where m is the number of distinct symbols in $\mathbf{x}^n$. For example, the word "sleepless" has pattern 123342311, which is of length 9 and size 4. We will canonically represent a pattern $\boldsymbol{\psi}$ as $1^{\mu_1} 2^{\mu_2} \ldots m^{\mu_m}$, where μ_j is the *multiplicity* of the symbol j, i.e., the number of times j appears, in $\boldsymbol{\psi}$. Note that $\mu_1 + \cdots + \mu_m = n$. The pattern $\boldsymbol{\psi}$ in our example has canonical form $1^3 2^2 3^3 4$.

Let $\boldsymbol{\psi}$ be a given pattern of length n, and let $\mathbf{p} = (p(x))_{x \in \mathcal{X}}$ be a probability distribution over a discrete (possibly countably infinite) set $\mathcal{X}$. The probability that n i.i.d. samples drawn from the distribution $\mathbf{p}$ form a sequence with pattern $\boldsymbol{\psi}$ is given by

$$(1) \qquad\qquad P(\boldsymbol{\psi}; \mathbf{p}) := \sum_{\mathbf{x}^n : \boldsymbol{\psi}(\mathbf{x}^n) = \boldsymbol{\psi}} \prod_{i=1}^{n} p(x_i).$$

Clearly, all patterns $\boldsymbol{\psi}$ with the same canonical form $1^{\mu_1} 2^{\mu_2} \ldots m^{\mu_m}$ will have the same pattern probability $P(\boldsymbol{\psi}; \mathbf{p})$. Indeed, if $\mathbf{p} = (p_1, \ldots, p_k) \in \Pi_{[k]}$ with $k \geq m$, then we can write

$$(2) \qquad\qquad P(\boldsymbol{\psi}; \mathbf{p}) = \sum_{\sigma} \prod_{i=1}^{m} p_{\sigma(i)}^{\mu_i},$$

where the summation runs over all one-to-one maps $\sigma : [m] \to [k]$.

The right-hand side of (2) can be expressed in an alternative form using the notion of a permanent of a matrix. The *permanent* of a real $k \times k$ matrix $\Theta = (\theta_{i,j})$ is defined as

$$\text{perm}(\Theta) := \sum_{\pi} \prod_{i=1}^{k} \theta_{i, \pi(i)},$$

where the summation is over all permutations $\pi : [k] \to [k]$. With this, it can be verified that (2) can be rewritten as

$$(3) \qquad\qquad P(\boldsymbol{\psi}; \mathbf{p}) = \frac{1}{(k-m)!} \ \text{perm}(\Theta(\boldsymbol{\psi}; \mathbf{p})),$$

where $\Theta(\boldsymbol{\psi}; \mathbf{p})$ is the $k \times k$ matrix $(\theta_{i,j})$ with $\theta_{i,j} = p_i^{\mu_j}$; here, we set[2] $\mu_j = 0$ for $m+1 \leq j \leq k$. The $\frac{1}{(k-m)!}$ term in (3) comes from the fact that each one-to-one map $\sigma : [m] \to [k]$ in the sum (2) can be extended to a permutation $\pi : [k] \to [k]$ in exactly $(k-m)!$ different ways.

The *PML probability* of a pattern $\boldsymbol{\psi}$ of size m is defined as

$$\text{(4)} \qquad P^{\mathrm{PML}}(\boldsymbol{\psi}) := \max_{\mathbf{p}} P(\boldsymbol{\psi}; \mathbf{p}),$$

the maximum being taken over all discrete distributions $\mathbf{p}$ of support size at least m. Any distribution that attains the maximum above is called a *PML distribution* of $\boldsymbol{\psi}$, denoted by $\mathbf{p}^{\mathrm{PML}}(\boldsymbol{\psi})$. For the purposes of this paper, we will assume that the maximum is indeed attained by some discrete distribution $\mathbf{p}$.[3] In this case, there is always a PML distribution with finite support [2]. Hence, we have

$$\text{(5)} \qquad P^{\mathrm{PML}}(\boldsymbol{\psi}) = \max_{k \geq m} \max_{\mathbf{p} \in \Pi_{[k]}} P(\boldsymbol{\psi}; \mathbf{p}).$$

It should be pointed out that, for any $k \geq m$, $P(\boldsymbol{\psi}; \mathbf{p})$ is a continuous function of $\mathbf{p}$ over the compact set $\Pi_{[k]}$, as is evident from (2), and hence, the maximum in $\max_{\mathbf{p} \in \Pi_{[k]}} P(\boldsymbol{\psi}; \mathbf{p})$ is indeed attained.

The problem of determining the PML distribution of a pattern seems to be computationally difficult in general [2], [4], [5], [6], [7], [8], [9]. Algorithms for approximating the PML distribution have been proposed by Orlitsky et al. [4] and Vontobel [10]. Vontobel's algorithm, in particular, uses the Bethe approximation, about which we will have much more to say in section 3.

2.2. Phase transition in the PML problem. Consider now the potentially simpler decision problem of determining whether the PML distribution of a given pattern is a uniform distribution. A natural approach to this problem would be to find a test for whether, for any fixed $k \geq m$, the uniform distribution U_k achieves the maximum in $\max_{\mathbf{p} \in \Pi_{[k]}} P(\boldsymbol{\psi}; \mathbf{p})$. In attempting this approach, we discovered a striking phase transition phenomenon in the PML problem. To describe this, we introduce some notation. For a pattern $\boldsymbol{\psi}$ of size m, and an integer $k \geq m$, let $\beta_k^{\boldsymbol{\psi}} : \Pi_{[k]} \to [0,1]$ be the function defined by the mapping $\mathbf{p} \mapsto P(\boldsymbol{\psi}; \mathbf{p})$. The phase transition phenomenon is made precise in the following theorem.

THEOREM 2.1. *For a pattern $\boldsymbol{\psi}$ of length n with canonical form $1^{\mu_1} 2^{\mu_2} \ldots m^{\mu_m}$, $m \geq 2$, define*

$$\text{(6)} \qquad \Upsilon(\boldsymbol{\psi}) := \frac{n^2 - n}{\sum_{i=1}^m \mu_i^2 - n}.$$

Then, for all integers $k \geq m$, the following holds: when $k < \Upsilon(\boldsymbol{\psi})$, the uniform distribution U_k is a local maximum of the function $\beta_k^{\boldsymbol{\psi}}$, and when $k > \Upsilon(\boldsymbol{\psi})$, U_k is a local minimum.

We clarify a point concerning the statement of the theorem. Note that $\Upsilon(\boldsymbol{\psi})$ is finite iff $\boldsymbol{\psi} \neq 123 \ldots n$. When $\boldsymbol{\psi} = 123 \ldots n$, we take $\Upsilon(\boldsymbol{\psi})$ to be ∞.

---

[2]For consistency, we define $0^0 = 1$. This is also in keeping with the convention used in Definition 3.1 that $0 \log 0 = 0$.

[3]In general, to guarantee that the maximum is always attained, we must allow "mixed" distributions; see [1], [2].

Proof of Theorem 2.1. The proof approach is based on that of Theorem 20 in [14]. Let $\mathbf{p} = U_k$, so that $\mathbf{p}$ is in the interior of the simplex $\Pi_{[k]}$. Pick an arbitrary direction $\boldsymbol{\xi} \in \mathbb{R}^k \setminus \{\mathbf{0}\}$, normalized so that $\|\boldsymbol{\xi}\|_2 = 1$, such that for all t within a sufficiently small interval around 0, the point $\mathbf{p}(t) = \mathbf{p} + t\boldsymbol{\xi}$ continues to lie within $\Pi_{[k]}$. Note that this implies that $\sum_{j=1}^k \xi_j = 0$. Consider the function $g(t) = P(\boldsymbol{\psi}; \mathbf{p}(t))$. We will show that, *independent* of the choice of $\boldsymbol{\xi}$, we have $g'(0) = 0$, $g''(0) < 0$ if $k < \Upsilon(\boldsymbol{\psi})$, and $g''(0) > 0$ if $k > \Upsilon(\boldsymbol{\psi})$. This clearly suffices to prove the theorem.

Now, from (2), $g(t)$ is expressible as $\sum_\sigma g_\sigma(t)$, where $g_\sigma(t) = \prod_{i=1}^m (p_{\sigma(i)} + t\xi_{\sigma(i)})^{\mu_i}$. Differentiation, together with the fact that $p_j = \frac{1}{k}$ for all j, yields $g'_\sigma(0) = \frac{1}{k^{n-1}} \sum_{i=1}^m \mu_i \xi_{\sigma(i)}$. Hence,

$$g'(0) = \sum_\sigma g'_\sigma(0) = \frac{1}{k^{n-1}} \sum_{i=1}^m \mu_i \sum_\sigma \xi_{\sigma(i)}.$$

For any fixed $i \in [m]$, the inner summation $\sum_\sigma \xi_{\sigma(i)}$ can be evaluated as follows. As σ ranges over all one-to-one maps from $[m]$ to $[k]$, for each $j \in [k]$, $\sigma(i)$ takes the value j exactly $\frac{(k-1)!}{(k-m)!}$ times. Hence, $\sum_\sigma \xi_{\sigma(i)} = \frac{(k-1)!}{(k-m)!} \sum_{j=1}^k \xi_j = 0$ by choice of $\boldsymbol{\xi}$. Thus, $g'(0) = 0$.

Next, we compute $g''(0) = \sum_\sigma g''_\sigma(0)$. Straightforward computations yield

$$g''_\sigma(0) = \frac{1}{k^{n-2}} \left[\left(\sum_{i=1}^m \mu_i \xi_{\sigma(i)} \right)^2 - \sum_{i=1}^m \mu_i \xi^2_{\sigma(i)} \right].$$

We rewrite the term within square brackets as

$$\sum_{i=1}^m \mu_i(\mu_i - 1)\xi^2_{\sigma(i)} + \sum_{(i,\ell):i\neq\ell} \mu_i \mu_\ell \xi_{\sigma(i)} \xi_{\sigma(\ell)}.$$

Summing over all one-to-one maps $\sigma : [m] \to [k]$, we obtain

$$g''(0) = \frac{1}{k^{n-2}} \left[\sum_{i=1}^m \mu_i(\mu_i - 1) \sum_\sigma \xi^2_{\sigma(i)} + \sum_{(i,\ell):i\neq\ell} \mu_i \mu_\ell \sum_\sigma \xi_{\sigma(i)} \xi_{\sigma(\ell)} \right].$$

As above, $\sum_\sigma \xi^2_{\sigma(i)} = \frac{(k-1)!}{(k-m)!} \sum_{j=1}^k \xi^2_j$ which, since $\|\boldsymbol{\xi}\|_2 = 1$, means that $\sum_\sigma \xi^2_{\sigma(i)} = \frac{(k-1)!}{(k-m)!}$. Similarly, for $i \neq \ell$, $\sum_\sigma \xi_{\sigma(i)} \xi_{\sigma(\ell)} = \frac{(k-2)!}{(k-m)!} \sum_{(s,t)\in[k]^2:s\neq t} \xi_s \xi_t$. We also have $0 = (\sum_{j=1}^k \xi_j)^2$, from which we obtain $\sum_{j=1}^k \xi^2_j = -\sum_{(s,t)\in[k]^2:s\neq t} \xi_s \xi_t$. Hence, again using the fact that $\|\boldsymbol{\xi}\|_2 = 1$, we have $\sum_\sigma \xi_{\sigma(i)} \xi_{\sigma(\ell)} = -\frac{(k-2)!}{(k-m)!}$. Putting it all together, we find that

$$g''(0) = C \left[(k-1) \sum_{i=1}^m \mu_i(\mu_i - 1) - \sum_{(i,\ell)\in[m]^2:i\neq\ell} \mu_i \mu_\ell \right],$$

where $C = \frac{1}{k^{n-2}} \frac{(k-2)!}{(k-m)!}$ is a positive constant independent of $\boldsymbol{\xi}$. Further simplification using the fact that $\sum_{i=1}^m \mu_i = n$ yields

$$g''(0) = C \left[k \left(\sum_{i=1}^m \mu_i^2 - n \right) - (n^2 - n) \right],$$

from which the desired result follows. $\qquad\square$

Theorem 2.1 shows that the uniform distribution U_k is either a local maximum or a local minimum of β_k^ψ for all integers $k \geq m$, except perhaps at the threshold $\Upsilon(\psi)$. Indeed, if $\Upsilon(\psi)$ happens to be an integer, then for $k = \Upsilon(\psi)$, it is possible that U_k is not a local extremum but only a saddle point. For example, for $\psi = 1122$, we have $\Upsilon(\psi) = 3$, and it may be verified that U_3 is not a local extremum for β_3^ψ. This is also a good place to point out that if $\Upsilon(\psi)$ happens to *not* be an integer, so that it lies strictly between two integers v and $v + 1$, then any $\Upsilon'(\psi) \in (v, v + 1)$ would also work as a threshold: indeed, for $m \leq k \leq v$, U_k is a local maximum for β_k^ψ, and for $k \geq v + 1$, U_k is a local minimum.

As a simple corollary of the theorem, we see that a necessary condition for the PML distribution of a pattern ψ to be uniform is that $\Upsilon(\psi) \geq m$. However, this condition is not sufficient in general. In [13], we derive a slightly stronger necessary condition using Theorem 2.1, which is used as the basis for a heuristic algorithm that determines whether a given pattern has a uniform PML distribution.

The intent of this paper, however, is to investigate whether the phase transition phenomenon reported in Theorem 2.1 extends to the Bethe approximation of the PML problem, which we describe in the next section.

3. The Bethe approximation. The Bethe approximation is a method whose origins lie in statistical physics [15], [16]. In the interest of brevity, we describe this approximation only in the context of the PML problem. The motivation and justification behind the definitions in this section are discussed in detail in [14].

3.1. The Bethe PML problem. From (3), we see that computing the pattern probability $P(\psi; \mathbf{p})$ is equivalent to computing the permanent of the matrix $\Theta(\psi; \mathbf{p})$. It is well-known that computing the permanent of a matrix is hard in general; formally, the problem is #P-complete [17]. Many approximation algorithms have been developed for this problem (e.g., [18], [19]), of which the ones based on the Bethe approximation [20], [21], [14] are relevant to us.

Let $\mathcal{D}_k$ denote the set of $k \times k$ doubly stochastic matrices. In the following definition, we use the convention that $0 \log 0 = 0$.

DEFINITION 3.1 (see [14], Corollary 15). *The Bethe permanent of a nonnegative $k \times k$ matrix* $\Theta = (\theta_{i,j})$, *with* $\theta_{i,j} \geq 0$ *for all* i, j, *is defined as*

$$\mathrm{perm}_{\mathrm{B}}(\Theta) := \max_{\Gamma \in \mathcal{D}_k} \exp\left(-F_{\mathrm{B}}(\Gamma, \Theta)\right),$$

where for $\Gamma = (\gamma_{i,j}) \in \mathcal{D}_k$, *we have* $F_{\mathrm{B}}(\Gamma) = U_{\mathrm{B}}(\Gamma, \Theta) - H_{\mathrm{B}}(\Gamma)$, *with*

$$U_{\mathrm{B}}(\Gamma, \Theta) := -\sum_{i,j} \gamma_{i,j} \log(\theta_{i,j}),$$

$$H_{\mathrm{B}}(\Gamma) := -\sum_{i,j} \gamma_{i,j} \log(\gamma_{i,j}) + \sum_{i,j} (1 - \gamma_{i,j}) \log(1 - \gamma_{i,j}).$$

The function $F_{\mathrm{B}}(\Gamma, \Theta)$ in the above definition is called the *Bethe free energy function*. If the pair (Γ, Θ) is such that $\gamma_{i,j} > 0$ but $\theta_{i,j} = 0$ for some (i, j), we define $F_{\mathrm{B}}(\Gamma, \Theta) = \infty$, and correspondingly, $\exp(-F_{\mathrm{B}}(\Gamma, \Theta)) = 0$. With these definitions, $\exp(-F_{\mathrm{B}}(\Gamma, \Theta))$ is a continuous function of (Γ, Θ), so that for any fixed Θ, $\exp(-F_{\mathrm{B}}(\Gamma, \Theta))$ attains a maximum on the compact set $\mathcal{D}_k$. Hence, $\mathrm{perm}_{\mathrm{B}}(\Theta)$ is well-defined.

For positive matrices Θ, we can write

$$\mathrm{perm}_{\mathrm{B}}(\Theta) = \exp\left(-\min_{\Gamma \in \mathcal{D}_k} F_{\mathrm{B}}(\Gamma, \Theta)\right).$$

Vontobel [14, Corollary 23] showed that for any positive matrix Θ, $F_{\mathrm{B}}(\Gamma, \Theta)$ is a convex function of $\Gamma \in \mathcal{D}_k$, so that $\min_{\Gamma \in \mathcal{D}_k} F_{\mathrm{B}}(\Gamma, \Theta)$ is a convex program. Vontobel further proved that the sum-product algorithm (belief propagation) can be used to find this minimum, and hence $\mathrm{perm}_{\mathrm{B}}(\Theta)$, highly efficiently. Since the Bethe permanent is often a very good proxy for the actual permanent [10], [11], [21], [22], having an efficient algorithm to compute it is particularly useful.

For a pattern $\boldsymbol{\psi}$ of size m and a probability distribution $\mathbf{p} \in \Pi_{[k]}$, $k \geq m$, we define, in analogy with (3), the quantity

$$(7) \qquad P_{\mathrm{B}}(\boldsymbol{\psi}; \mathbf{p}) := \frac{1}{(k-m)!} \, \mathrm{perm}_{\mathrm{B}}(\Theta(\boldsymbol{\psi}; \mathbf{p})).$$

We then have $0 \leq P_{\mathrm{B}}(\boldsymbol{\psi}; \mathbf{p}) \leq P(\boldsymbol{\psi}; \mathbf{p}) \leq 1$, the inequalities holding for the following reasons:

- the first inequality is simply a consequence of the nonnegativity of the Bethe permanent;
- the second inequality is because of the fact that $\mathrm{perm}(\Theta) \geq \mathrm{perm}_{\mathrm{B}}(\Theta)$ for any nonnegative matrix Θ, an inequality proved by Gurvits [23], [24];
- the last inequality is a consequence of the fact that $P(\boldsymbol{\psi}; \mathbf{p})$ is a probability.

Thus, $P_{\mathrm{B}}(\boldsymbol{\psi}; \mathbf{p})$ can be viewed as a probability as well.

With this, we define, in analogy with (4) and (5), the *Bethe PML probability* of a pattern $\boldsymbol{\psi}$ to be

$$(8) \qquad P^{\mathrm{BPML}}(\boldsymbol{\psi}) := \sup_{k \geq m} \max_{\mathbf{p} \in \Pi_{[k]}} P_{\mathrm{B}}(\boldsymbol{\psi}; \mathbf{p}).$$

A couple of clarifications on this definition may be needed. One is that for any positive integer k, $\max_{\mathbf{p} \in \Pi_{[k]}} P_{\mathrm{B}}(\boldsymbol{\psi}; \mathbf{p})$ is well-defined. This is because $\chi(\Gamma, \mathbf{p}) := \exp\{-F_{\mathrm{B}}(\Gamma, \Theta(\boldsymbol{\psi}; \mathbf{p}))\}$, as a function of $(\Gamma, \mathbf{p})$, is continuous on the compact set $\mathcal{D}_k \times \Pi_{[k]}$. This implies that $\mathrm{perm}_{\mathrm{B}}(\Theta(\boldsymbol{\psi}; \mathbf{p})) = \max_{\Gamma \in \mathcal{D}_k} \chi(\Gamma, \mathbf{p})$ is a continuous function of $\mathbf{p}$. Hence, $P_{\mathrm{B}}(\boldsymbol{\psi}; \mathbf{p})$, being a continuous function of $\mathbf{p}$, must attain a maximum on the compact set $\Pi_{[k]}$.

A second clarification is that it is not known whether the supremum in (8) is always achieved at some finite k, although empirical evidence suggests that this may indeed be the case [11]. Empirically again, the *Bethe PML distribution*, defined as any distribution $\mathbf{p}$ for which $P_{\mathrm{B}}(\boldsymbol{\psi}; \mathbf{p}) = P^{\mathrm{BPML}}(\boldsymbol{\psi})$, is a very good approximation of the PML distribution of a pattern $\boldsymbol{\psi}$. The "Bethe PML problem" of determining the Bethe PML distribution is also considerably easier to solve numerically than the original PML problem [10], [11].

The question we are interested in addressing is whether the Bethe PML problem exhibits a phase transition analogous to that described for the PML problem in Theorem 2.1. To answer this, we must understand when the uniform distribution U_k is a local maximum or a local minimum in $\Pi_{[k]}$ for the function $\mathbf{p} \mapsto \mathrm{perm}_{\mathrm{B}}(\Theta(\boldsymbol{\psi}; \mathbf{p}))$. A direct approach analogous to that used in the proof of Theorem 2.1 seems difficult as we only have a description of $\mathrm{perm}_{\mathrm{B}}$ as a solution to a convex optimization problem. Instead, we take an indirect approach via the degree-M lifted permanents discussed next.

3.2. Degree-M lifted permanents. As an alternative to defining the Bethe permanent as a solution to an optimization problem, Vontobel gave a combinatorial characterization of this quantity, which we describe here. Let $\Theta = (\theta_{i,j})$ be a given $k \times k$ matrix with nonnegative entries. For a positive integer M, let $\mathcal{P}_M$ denote the

set of all $M \times M$ permutation matrices. Further, let $\mathcal{P}_M^{k \times k}$ be the set of all $kM \times kM$ matrices of the form

$$
(9) \qquad \Lambda = \begin{pmatrix} \Lambda^{(1,1)} & \Lambda^{(1,2)} & \cdots & \Lambda^{(1,k)} \\ \Lambda^{(2,1)} & \Lambda^{(2,2)} & \cdots & \Lambda^{(2,k)} \\ \vdots & \vdots & \ddots & \vdots \\ \Lambda^{(k,1)} & \Lambda^{(k,2)} & \cdots & \Lambda^{(k,k)} \end{pmatrix}
$$

with $\Lambda^{(i,j)} \in \mathcal{P}_M$ for all i, j. For a Λ as above, define

$$
(10) \qquad \Theta \odot \Lambda := \begin{pmatrix} \theta_{1,1}\Lambda^{(1,1)} & \theta_{1,2}\Lambda^{(2,1)} & \cdots & \theta_{1,k}\Lambda^{(1,k)} \\ \theta_{2,1}\Lambda^{(2,1)} & \theta_{2,2}\Lambda^{(2,2)} & \cdots & \theta_{1,k}\Lambda^{(2,k)} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{k,1}\Lambda^{(k,1)} & \theta_{k,2}\Lambda^{(k,2)} & \cdots & \theta_{k,k}\Lambda^{(k,k)} \end{pmatrix}.
$$

DEFINITION 3.2 (see [14], Definition 38). *The degree-M lifted permanent of Θ is defined to be*

$$
\mathrm{perm}_{\mathrm{B},M}(\Theta) := \big\langle \mathrm{perm}(\Theta \odot \Lambda) \big\rangle^{1/M},
$$

where the angular brackets represent the arithmetic average of $\mathrm{perm}(\Theta \odot \Lambda)$ as Λ ranges over the $(M!)^{k^2}$ matrices in $\mathcal{P}_M^{k \times k}$. Equivalently, $\big\langle \mathrm{perm}(\Theta \odot \Lambda) \big\rangle$ is the expected value of $\mathrm{perm}(\Theta \odot \Lambda)$, the expectation being taken over Λ chosen uniformly at random from $\mathcal{P}_M^{k \times k}$.

Note that when $M = 1$, $\mathrm{perm}_{\mathrm{B},M}(\Theta)$ is equal to $\mathrm{perm}(\Theta)$. At the other extreme, as $M \to \infty$, Vontobel [14, Theorem 39] has shown the following identity:

$$
(11) \qquad \limsup_{M \to \infty} \mathrm{perm}_{\mathrm{B},M}(\Theta) = \mathrm{perm}_{\mathrm{B}}(\Theta).
$$

Thus, degree-M lifted permanents interpolate between $\mathrm{perm}(\Theta)$ and $\mathrm{perm}_{\mathrm{B}}(\Theta)$. The advantage of using degree-M lifted permanents as an indirect means of understanding the Bethe permanent is that they can be expressed in a form that is more amenable to analysis. To this end, we introduce a piece of notation that will find much use in this paper. Given positive integers k and M, we let $\mathcal{A}_{k,M}$ denote the set of all $k \times k$ nonnegative integer matrices whose row and column sums are all equal to M.

PROPOSITION 3.3. *For any $k \times k$ matrix $\Theta = (\theta_{i,j})$ and any positive integer M, we have*

$$
\big[\mathrm{perm}_{\mathrm{B},M}(\Theta)\big]^M = (M!)^{2k-k^2} \sum_{(a_{i,j}) \in \mathcal{A}_{k,M}} \prod_{(i,j) \in [k]^2} \theta_{i,j}^{a_{i,j}} \frac{(M - a_{i,j})!}{(a_{i,j})!}.
$$

Using multinomial coefficients, the identity above can be expressed in an alternative, more evocative form:[4]

$$
(12)
$$

$$
\big[\mathrm{perm}_{\mathrm{B},M}(\Theta)\big]^M = \sum_{(a_{i,j}) \in \mathcal{A}_{k,M}} \left[\prod_{(i,j) \in [k]^2} \theta_{i,j}^{a_{i,j}} \right] \frac{\prod_{i=1}^k \binom{M}{a_{i,1},\ldots,a_{i,k}} \prod_{j=1}^k \binom{M}{a_{1,j},\ldots,a_{k,j}}}{\prod_{(i,j) \in [k]^2} \binom{M}{a_{i,j}}}.
$$

Proposition 3.3 is proved in Appendix A.

---

[4]It is also possible to recover this form from Lemma 29 in [25].

4. Phase transition in the Bethe PML problem. As mentioned at the end of section 3.1, we take an indirect approach, via degree-M lifted permanents, to the question of the existence of a phase transition phenomenon in the Bethe PML problem. This approach is based on the intuition that the large-M behavior of these lifted permanents will, by virtue of (11), shed light on the behavior of the Bethe permanent. With this program in mind, we define for a pattern ψ of size $m \geq 2$, and integers $k \geq m$ and $M \geq 1$, a function $\beta_{k,M}^{\psi} : \Pi_{[k]} \to \mathbb{R}_+$ that maps $\mathbf{p} \in \Pi_{[k]}$ to $\mathrm{perm}_{\mathrm{B},M}(\Theta(\psi; \mathbf{p}))$. Recall that U_k denotes the uniform distribution on $[k]$. The aim of this section is to prove the following result.

THEOREM 4.1. *Let ψ be a pattern of length n having canonical form $1^{\mu_1} 2^{\mu_2} \ldots m^{\mu_m}$, $m \geq 2$. There is a threshold $\Upsilon_{\mathrm{B}}(\psi)$ such that for all integers $k \geq m$, the following holds for all sufficiently large M:*

- *when $k < \Upsilon_{\mathrm{B}}(\psi)$, U_k is a local maximum for $\beta_{k,M}^{\psi}$; and*
- *when $k > \Upsilon_{\mathrm{B}}(\psi)$, U_k is a local minimum for $\beta_{k,M}^{\psi}$.*

When $m = 2$, the threshold $\Upsilon_{\mathrm{B}}(\psi)$ may be taken to be the following:

$$\Upsilon_{\mathrm{B}}(\psi) = \begin{cases} \infty & \text{if } \mu_1 = \mu_2 = 1, \\ 2 + \delta & \text{if } \mu_1 = \mu_2 > 1, \\ 1 + \delta & \text{otherwise} \end{cases}$$

for an arbitrary choice of $\delta \in (0, 1)$. When $m \geq 3$, $\Upsilon_{\mathrm{B}}(\psi)$ may be chosen to closely mimic the threshold $\Upsilon(\psi)$ of Theorem 2.1 in the following sense:

- *if $\Upsilon(\psi) = \infty$ (which happens iff $\psi = 123 \ldots n$), then $\Upsilon_{\mathrm{B}}(\psi) = \infty$;*
- *if $\Upsilon(\psi) < \frac{\sqrt{n}+1}{\sqrt{n}-1}$, then we may take $\Upsilon_{\mathrm{B}}(\psi) = \Upsilon(\psi)$;*
- *in all other cases, we may take*

$$\Upsilon_{\mathrm{B}}(\psi) = \frac{\mathcal{U} + n^2 - 2n + \sqrt{(n^2 + 2n - \mathcal{U})^2 - 4n^3}}{2(\mathcal{U} - n)},$$

where $\mathcal{U} = \sum_{i=1}^{m} \mu_i^2$, so that $\Upsilon(\psi) - 1 \leq \Upsilon_{\mathrm{B}}(\psi) < \Upsilon(\psi)$ holds.

The theorem does not explicitly give a comparison between the thresholds $\Upsilon(\psi)$ and $\Upsilon_{\mathrm{B}}(\psi)$ in the case when $m = 2$. This has been done only so that a clean statement of the result could be given. Indeed, there is a close relationship between the two thresholds even in this case: it can be shown using Theorem 2.1 that when $m = 2$,

$$\Upsilon(\psi) = \begin{cases} \infty & \text{if } \mu_1 = \mu_2 = 1, \\ 2 + \frac{1}{\mu_1 - 1} & \text{if } \mu_1 = \mu_2 > 1, \end{cases}$$

and $\Upsilon(\psi)$ lies in the interval $(1, 3)$ otherwise.

In summary, Theorem 4.1 strongly indicates that the Bethe PML problem exhibits a phase transition phenomenon very similar to that proved in Theorem 2.1 for the PML problem. Unfortunately, this does not quite prove that there is indeed such a phase transition in the Bethe PML problem. We make some remarks concerning this in section 6.

The rest of this section is devoted to a proof of Theorem 4.1. The proof proceeds along the same lines as that of Theorem 2.1, except that the calculations are messier. To preserve the flow of this section, we have moved the proofs of some intermediate lemmas, which mainly involve tedious calculations, to the appendices. Also, Proposition 4.4 below, which is also an intermediate step in the proof of Theorem 4.1, but which could be considered an interesting result in its own right, is proved in section 5.

Let $\mathbf{p} = U_k$, which lies in the interior of $\Pi_{[k]}$. Pick an arbitrary direction $\boldsymbol{\xi} \in \mathbb{R}^k \setminus \{\mathbf{0}\}$, normalized so that $\|\boldsymbol{\xi}\|_2 = 1$, such that for all t within a sufficiently small interval around 0, the point $\mathbf{p}(t) = \mathbf{p} + t\boldsymbol{\xi}$ continues to lie within $\Pi_{[k]}$. Note that this implies that $\sum_{j=1}^k \xi_j = 0$.

Given a pattern $\boldsymbol{\psi}$ with multiplicities $(\mu_1, \ldots, \mu_m)$, $m \geq 2$, and integers $k \geq m$ and $M \geq 1$, define the function $G_{k,M}(t) = \beta_{k,M}^{\boldsymbol{\psi}}(\mathbf{p}(t))$.[5] As in the proof of Theorem 2.1, the idea is to show that $G'_{k,M}(0) = 0$ and that for a suitable choice of Υ_B independent of $\boldsymbol{\xi}$, the sign of $G''_{k,M}(0)$ depends, for all sufficiently large M, only on whether $k < \Upsilon_\mathrm{B}$ or $k > \Upsilon_\mathrm{B}$. We will in fact prove the statement about the second derivative in the following equivalent form: there exists a threshold Υ_B independent of $\boldsymbol{\xi}$ such that

$$(13) \qquad \lim_{M \to \infty} G''_{k,M}(0) < 0 \ \forall \, k < \Upsilon_\mathrm{B} \quad \text{and} \quad \lim_{M \to \infty} G''_{k,M}(0) > 0 \ \forall \, k > \Upsilon_\mathrm{B}.$$

It is straightforward to show that $G'_{k,M}(0) = 0$; a proof of this will be given in Appendix B as part of the proof of Lemma 4.3 below. To express $G''_{kM}(0)$, we consider once again the set, $\mathcal{A}_{k,M}$, of $k \times k$ nonnegative integer matrices all of whose row and column sums are equal to M. For $A = (a_{i,j}) \in \mathcal{A}_{k,M}$, define

$$(14) \qquad w(A) := \prod_{(i,j)\in[k]^2} \frac{(M - a_{i,j})!}{(a_{i,j})!},$$

and let

$$(15) \qquad Z_{k,M} := \sum_{A \in \mathcal{A}_{k,M}} w(A).$$

Then,

$$Q_{k,M}(A) := \frac{1}{Z_{k,M}} w(A)$$

defines a probability distribution on $\mathcal{A}_{k,M}$. We will study this probability distribution in more detail in section 5. For now, we state and prove a simple fact about the distribution $Q_{k,M}$, which we will need to prove some of the other results in this section.

LEMMA 4.2. *For any entry $a_{i,j}$ of a random matrix $A \in \mathcal{A}_{k,M}$ chosen according to the distribution $Q_{k,M}$, we have*

$$\mathbb{E}[a_{i,j}] = \frac{1}{k} M.$$

Proof. Note that $w(A)$, and, consequently, the probability function $Q_{k,M}(A)$, is invariant to row and column permutations of A. Therefore, the expected value $\mathbb{E}[a_{i,j}]$ must be a constant independent of (i, j). This constant can be explicitly evaluated as follows:

$$\mathbb{E}[a_{i,j}] = \frac{1}{k} \sum_{\ell=1}^k \mathbb{E}[a_{i,\ell}] = \frac{1}{k} \mathbb{E}\left[\sum_{\ell=1}^k a_{i,\ell}\right] = \frac{1}{k} M,$$

the last equality using the fact that $\sum_{\ell=1}^k a_{i,\ell} = M$. □

---

[5]Here, and for the remainder of this section, we will suppress the dependence on $\boldsymbol{\psi}$ in our notation; thus, we write $G_{k,M}(t)$ instead of $G_{k,M}^{\boldsymbol{\psi}}(t)$, Υ_B instead of $\Upsilon_\mathrm{B}(\boldsymbol{\psi})$, and so on.

The above fact is used to derive the expression for $G_{k,M}''(0)$ given next. Let $\mathrm{Var}_{k,M}(a_{1,1})$ denote the variance of the entry $a_{1,1}$ in a random matrix $A \in \mathcal{A}_{k,M}$ chosen according to the distribution $Q_{k,M}$.

LEMMA 4.3. *We have* $G_{k,M}'(0) = 0$ *and*

$$G_{k,M}''(0) = [(M!)^{2k-k^2}]^{\frac{1}{M}}(Z_{k,M})^{\frac{1}{M}}k^{1-n}\left[\frac{k^2}{(k-1)^2}\frac{\mathrm{Var}_{k,M}(a_{1,1})}{M}\left(k\sum_{i=1}^{m}\mu_i^2 - n^2\right) - n\right].$$

The proof of the lemma is deferred to Appendix B. The quantities $Z_{k,M}$ and $\mathrm{Var}_{k,M}(a_{1,1})$ can be determined explicitly for $k = 2$ and asymptotically as $M \to \infty$ for $k \geq 3$.

PROPOSITION 4.4.
(a) $Z_{2,M} = M + 1$ *and* $\mathrm{Var}_{2,M}(a_{1,1}) = \frac{1}{12}M(M+2)$.
(b) *For* $k \geq 3$, *we have*

$$\lim_{M\to\infty}[(M!)^{2k-k^2}]^{\frac{1}{M}}(Z_{k,M})^{\frac{1}{M}} = \frac{(k-1)^{k(k-1)}}{k^{k(k-2)}}$$

and

$$\lim_{M\to\infty}\frac{1}{M}\mathrm{Var}_{k,M}(a_{1,1}) = \frac{(k-1)^3}{k^3(k-2)}.$$

The proof of the proposition will be given in section 5. We digress for a moment from the main line of our argument to record a simple consequence of Proposition 4.4, which may be of some interest. For any integer $k \geq 2$, let $\mathbf{J}_k$ denote the $k \times k$ matrix all of whose entries equal 1. It is clear from Proposition 3.3 that $\mathrm{perm}_{\mathrm{B},M}(\mathbf{J}_k) = [(M!)^{2k-k^2}]^{\frac{1}{M}}(Z_{k,M})^{\frac{1}{M}}$. Proposition 4.4 thus shows that

$$\lim_{M\to\infty}\mathrm{perm}_{\mathrm{B},M}(\mathbf{J}_k) = \frac{(k-1)^{k(k-1)}}{k^{k(k-2)}}$$

for $k \geq 2$. (The fact that this expression holds for $k = 2$ as well is verifiable from $Z_{2,M} = M + 1$.) It should be pointed out that a slightly weaker version of this equality, namely, $\limsup_{M\to\infty}\mathrm{perm}_{\mathrm{B},M}(\mathbf{J}_k) = \frac{(k-1)^{k(k-1)}}{k^{k(k-2)}}$, can be inferred from the work of Vontobel [14]. Indeed, this follows from (11) and the proof of Lemma 48 in [14], which shows that $\mathrm{perm}_{\mathrm{B}}(\mathbf{J}_k)$ equals the expression on the right-hand side above.

Returning to the main line of our argument, as another direct consequence of Lemma 4.3 and Proposition 4.4, we have the following result.

COROLLARY 4.5.
(a) *When* $k = m = 2$,

$$\lim_{M\to\infty}G_{2,M}''(0) = \begin{cases} -n\,2^{1-n} & \text{if } \mu_1 = \mu_2, \\ +\infty & \text{if } \mu_1 \neq \mu_2. \end{cases}$$

(b) *When* $k \geq 3$ *(and* $k \geq m$),

(16) $$\lim_{M\to\infty}G_{k,M}''(0) = \frac{(k-1)^{k(k-1)}}{k^{k(k-2)}}k^{1-n}\left[\frac{k-1}{k(k-2)}\left(k\sum_{i=1}^{m}\mu_i^2 - n^2\right) - n\right].$$

Proof. Only part (a) requires a note of explanation. When $k = m = 2$, the term $k \sum_{i=1}^{m} \mu_i^2 - n^2$ in the expression for $G_{k,M}''(0)$ in Lemma 4.3 reduces to $2(\mu_1^2 + \mu_2^2) - (\mu_1 + \mu_2)^2$, which equals $(\mu_1 - \mu_2)^2$. $\square$

For our purposes, it is only the sign of $\lim_{M\to\infty} G_{k,M}''(0)$ that matters, so we will make much use of the weaker corollary below. As in the statement of Theorem 4.1, we set $\mathcal{U} = \sum_{i=1}^{m} \mu_i^2$.

COROLLARY 4.6.
(a) *When $k = m = 2$, we have $\lim_{M\to\infty} G_{2,M}''(0) < 0$ if $\mu_1 = \mu_2$ and $\lim_{M\to\infty}$ $G_{2,M}''(0) > 0$ if $\mu_1 \neq \mu_2$.*
(b) *When $k \geq 3$, we have $\lim_{M\to\infty} G_{k,M}''(0) \lessgtr 0$ if*

$$k^2(\mathcal{U} - n) - k(\mathcal{U} + n^2 - 2n) + n^2 \lessgtr 0.$$

Proof. It suffices to point out that the condition in part (b) above is equivalent to the term within square brackets in (16) being negative or positive. $\square$

Thus, when $k \geq 3$, the sign of $\lim_{M\to\infty} G_{k,M}''(0)$ depends only on where k lies in relation to the roots of the quadratic polynomial $x^2(\mathcal{U} - n) - x(\mathcal{U} + n^2 - 2n) + n^2$. Note that $\mathcal{U} = \sum_{i=1}^{m} \mu_i^2 \geq \sum_{i=1}^{m} \mu_i = n$, with equality iff $\mu_i = 1$ for all $i \in [m]$, i.e., $\psi = 123\ldots n$. The lemma below summarizes the behavior of the roots of the quadratic equation.

LEMMA 4.7. *For $n \geq 2$, write $q(x) = x^2(\mathcal{U} - n) - x(\mathcal{U} + n^2 - 2n) + n^2$, which has discriminant $D = (n^2 + 2n - \mathcal{U})^2 - 4n^3$. Recall that $\Upsilon = \frac{n^2 - n}{\mathcal{U} - n}$.*
(1) *If $\mathcal{U} = n$ (which happens iff $\psi = 123\ldots n$), then $q(x) = -x(n^2 - n) + n^2$ has $\frac{n}{n-1}$ as its only root. Since $\frac{n}{n-1} \leq 2$, we have $q(k) < 0$ for all $k \geq 3$.*
(2) *If $\mathcal{U} > n$, then we have exactly one of the following two cases:*
  (a) *The discriminant D is strictly negative, which happens iff $\Upsilon < \frac{\sqrt{n}+1}{\sqrt{n}-1}$, so that $q(x)$ has no roots. In this case, $q(k) > 0$ for all k.*
  (b) *The quadratic has two real roots $\rho_1 \leq \rho_2$ given by*

$$\rho_1 = \frac{\mathcal{U} + n^2 - 2n - \sqrt{D}}{2(\mathcal{U} - n)} \quad and \quad \rho_2 = \frac{\mathcal{U} + n^2 - 2n + \sqrt{D}}{2(\mathcal{U} - n)}.$$

  In this case, we have $1 < \rho_1 \leq 2$, so that $q(k) < 0$ for $3 \leq k < \rho_2$, and $q(k) > 0$ for all $k > \rho_2$. Furthermore, $\Upsilon - 1 \leq \rho_2 < \Upsilon$ holds.

The proof of the lemma is given in Appendix C. We now have the tools required to complete the proof of Theorem 4.1.

Proof of Theorem 4.1. Recall that the goal is to show that there is a threshold Υ_B independent of the direction vector ξ such that (13) holds. From Corollary 4.5, we see that $\lim_{M\to\infty} G_{k,M}''(0)$ is independent of ξ. With this, the $m \geq 3$ case of Theorem 4.1 follows directly from Corollary 4.6(b) and Lemma 4.7.

The $m = 2$ case requires a few additional details to be checked. When $\mu_1 = \mu_2 = 1$, Corollary 4.6 and Lemma 4.7(1) show that $\lim_{M\to\infty} G_{k,M}''(0) < 0$ for all $k \geq 2$, and hence, we can take $\Upsilon_B = \infty$.

When $\mu_1 = \mu_2 > 1$, we need to argue that (13) holds for any $\Upsilon_B \in (2,3)$. Corollary 4.6(a) shows that $\lim_{M\to\infty} G_{k,M}''(0) < 0$ for $k = 2$. Therefore, it remains to show that $\lim_{M\to\infty} G_{k,M}''(0) > 0$ for all $k \geq 3$. Similarly, when $\mu_1 \neq \mu_2$, we need to show that $\lim_{M\to\infty} G_{k,M}''(0) > 0$ for all $k \geq 2$, so that we can take $\Upsilon_B \in (1,2)$ in this case. Again, Corollary 4.6(a) takes care of the $k = 2$ case, so we are left with $k \geq 3$.

In summary, we must show that if $(\mu_1, \mu_2) \neq (1,1)$, then $\lim_{M \to \infty} G''_{k,M}(0) > 0$ for all $k \geq 3$. For this, we will appeal to Lemma 4.7(2).

If the discriminant D is negative, then we are done by Lemma 4.7(2)(a). So, we may assume that $D \geq 0$, in which case the situation of Lemma 4.7(2)(b) applies. It suffices to show that when $(\mu_1, \mu_2) \neq (1,1)$, then $\rho_2 \leq 2$. So, let $(\mu_1, \mu_2) = (a, b)$ for some $(a, b) \neq (1,1)$. Note that $D = (2(a+b) + 2ab)^2 - 4(a+b)^3 = 4[(a+b+ab)^2 - (a+b)^3]$. Using the fact that $(a+b)^3 \geq 4ab(a+b)$, we obtain $D \leq 4[(a+b+ab)^2 - 4ab(a+b)] = 4(a+b-ab)^2$. Thus, $\sqrt{D} \leq 2|a+b-ab|$, and hence,

$$(17) \qquad \rho_2 = 1 + \frac{n^2 - \mathcal{U} + \sqrt{D}}{2(U-n)} \leq 1 + \frac{ab + |a+b-ab|}{a^2 + b^2 - (a+b)}.$$

Now, for positive integers a, b, if $ab < a+b$, then $b = 1$. If $(a, b) = (2, 1)$, then observe that $D = -8 < 0$, which cannot happen. So we can have $ab < a+b$ only if $b = 1$ and $a \geq 3$. In this case, (17) becomes $\rho_2 \leq 1 + \frac{a+1}{a^2-a} < 2$, the last inequality holding for any $a > 1 + \sqrt{2}$.

Finally, if $ab \geq a+b$, then (17) reduces to

$$\rho_2 \leq 1 + \frac{2ab - (a+b)}{a^2 + b^2 - (a+b)},$$

which is at most 2. We have thus shown that $\rho_2 \leq 2$ whenever $(\mu_1, \mu_2) \neq (1,1)$, which completes the proof of the $m = 2$ case of the theorem. $\square$

5. The probability distribution $Q_{k,M}$. The main aim of this section is to prove Proposition 4.4. For convenience, we recall the relevant definitions first. We use $\mathcal{A}_{k,M}$ to denote the set of $k \times k$ nonnegative integer matrices whose row and column sums are all equal to M. The probability distribution $Q_{k,M}$ on $\mathcal{A}_{k,M}$ is defined by setting, for $A \in \mathcal{A}_{k,M}$, $Q_{k,M}(A) := \frac{1}{Z_{k,M}} w(A)$, where $w(A)$ and $Z_{k,M}$ are given by (14) and (15). We are especially interested in determining $\mathrm{Var}_{k,M}(a_{1,1})$, the variance of the entry $a_{1,1}$ in a random matrix $A = (a_{i,j}) \in \mathcal{A}_{k,M}$ chosen according to the distribution $Q_{k,M}$.

Recall from Lemma 4 that the expected value $\mathbb{E}[a_{i,j}]$ is a constant independent of (i, j) due to the fact that $Q_{k,M}(A)$ is invariant to row and column permutations of A. This invariance of $Q_{k,M}(A)$ also implies that the variance of $a_{i,j}$ is independent of (i, j). Nonetheless, the explicit computation of $\mathrm{Var}_{k,M}(a_{1,1})$ seems difficult in general, with the notable exception of the case when $k = 2$.

Observe that $\mathcal{A}_{2,M}$ consists precisely of the matrices

$$\begin{bmatrix} a & M-a \\ M-a & a \end{bmatrix}$$

with $a \in \{0, 1, \ldots, M\}$. For any such matrix A, we have $w(A) = 1$, and hence, $Z_{2,M} = |\mathcal{A}_{2,M}| = M + 1$. In particular, $Q_{2,M}$ is just the uniform distribution on $\mathcal{A}_{2,M}$. It follows that $\mathrm{Var}_{2,M}(a_{1,1})$ is the variance of a random variable uniformly distributed on $\{0, 1, \ldots, M\}$. An easy calculation shows this to be $\frac{1}{12} M(M+2)$, thus completing the proof of part (a) of Proposition 4.4.

Henceforth, we consider the case when $k \geq 3$. Our interest is in the regime where k is fixed and M goes to ∞.

5.1. Estimating $Z_{k,M}$. To estimate the normalization constant $Z_{k,M}$, we make use of the fact that

$$w^*_{k,M} \leq Z_{k,M} \leq w^*_{k,M} |\mathcal{A}_{k,M}|,$$

where $w^*_{k,M} := \max_{A \in \mathcal{A}_{k,M}} w(A)$. As we will see below, it is possible to explicitly determine $w^*_{k,M}$. In the regime of interest to us, $w^*_{k,M}$ grows superexponentially in M. Since $|\mathcal{A}_{k,M}| \leq (M+1)^{k^2}$, a quantity that is polynomial in M, we see that the asymptotics of $Z_{k,M}$ are governed by the asymptotics of $w^*_{k,M}$.

Write $M = qk + r$, where $q = \lfloor M/k \rfloor$ and $r = M - kq$. Note that we have $0 \leq r < k$. Let $\mathbf{u} = (u_1, \ldots, u_k) \in \mathbb{Z}^k_+$ be defined by $u_i = q + 1$ for $1 \leq i \leq r$, and $u_i = q$ for $r + 1 \leq i \leq k$. Clearly, $\sum_i u_i = M$. Let U be the circulant matrix having $\mathbf{u}$ as its first row. The fact that U is a circulant matrix assures us that it is in $\mathcal{A}_{k,M}$.

PROPOSITION 5.1. *The matrix U maximizes $w(A)$ among all $A \in \mathcal{A}_{k,M}$, and hence,*

$$w^*_{k,M} = w(U) = \left[\frac{(M - (q+1))!}{(q+1)!}\right]^{kr} \left[\frac{(M - q)!}{q!}\right]^{k(k-r)}.$$

Before giving a proof of the proposition, we briefly discuss its implications. In the regime of fixed k and $M \to \infty$, routine algebraic manipulations using Stirling's approximation yield

(18) $$\lim_{M \to \infty} [(M!)^{2k-k^2}]^{\frac{1}{M}} (w^*_{k,M})^{\frac{1}{M}} = \frac{(k-1)^{k(k-1)}}{k^{k(k-2)}}.$$

Thus, $w^*_{k,M}$ grows superexponentially in M, so that as observed above, its asymptotic behavior governs that of $Z_{k,M}$. Consequently, in the left-hand side of (18), we may replace $w^*_{k,M}$ with $Z_{k,M}$, thereby obtaining the first equality stated in Proposition 4.4(b).

Our proof of Proposition 5.1 is based upon the theory of majorization [26]. For any vector $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$, let $\mathbf{x}_\downarrow = (x_{[1]}, \ldots, x_{[k]})$ denote the permutation of the components of $\mathbf{x}$ such that $x_{[1]} \geq \cdots \geq x_{[k]}$. A vector $\mathbf{x} \in \mathbb{R}^k$ is said to be *majorized* by a vector $\mathbf{y} \in \mathbb{R}^k$, denoted by $\mathbf{x} \preceq \mathbf{y}$, if for $\ell = 1, \ldots, k-1$, we have

$$\sum_{i=1}^{\ell} x_{[i]} \leq \sum_{i=1}^{\ell} y_{[i]},$$

and $\sum_{i=1}^{k} x_{[i]} = \sum_{i=1}^{k} y_{[i]}$ (or equivalently, $\sum_{i=1}^{k} x_i = \sum_{i=1}^{k} y_i$).

The next lemma plays an important role in our proof of Proposition 5.1. Let us define $S_{k,M}$ to be the set of vectors $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{Z}^k_+$ such that $\sum_{i=1}^{k} x_i = M$. Also, recall from above our definition of the vector $\mathbf{u} = (u_1, \ldots, u_k) \in S_{k,M}$ with $u_i = q + 1$ for $1 \leq i \leq r$, and $u_i = q$ for $r + 1 \leq i \leq k$, where $q = \lfloor M/k \rfloor$ and $r = M - kq$.

LEMMA 5.2. *The vector $\mathbf{u}$ is majorized by every $\mathbf{x} \in S_{k,M}$.*

Proof. Consider an arbitrary $\mathbf{x} \in S_{k,M}$. By definition, $\sum_{i=1}^{k} u_i = \sum_{i=1}^{k} x_i = M$, so we must show that $\sum_{i=1}^{\ell} u_{[i]} \leq \sum_{i=1}^{\ell} x_{[i]}$ for $\ell = 1, \ldots, k-1$. Suppose that $\sum_{i=1}^{\ell} u_{[i]} > \sum_{i=1}^{\ell} x_{[i]}$ for some $\ell \in \{1, \ldots, k-1\}$. Note that $\sum_{i=1}^{\ell} u_{[i]} = \ell q + \min(\ell, r)$. Thus,

$$\ell x_{[\ell+1]} \leq \sum_{i=1}^{\ell} x_{[i]} < \sum_{i=1}^{\ell} u_{[i]} = \ell q + \min(\ell, r),$$

from which we infer that $x_{[\ell+1]} < q + \min(1, r/\ell)$. Since $x_{[\ell+1]}$ must be an integer, we deduce that $x_{[\ell+1]} \leq q$.

On the other hand, we also have

$$(k - \ell)x_{[\ell+1]} \geq \sum_{i=\ell+1}^{k} x_{[i]} = M - \sum_{i=1}^{\ell} x_{[i]}$$

$$> M - \sum_{i=1}^{\ell} u_{[i]}$$

$$= qk + r - \big(\ell q + \min(\ell, r)\big)$$

$$\geq q(k - \ell).$$

Hence, $x_{[\ell+1]} > q$, which contradicts the inequality $x_{[\ell+1]} \leq q$ deduced previously. Therefore, our assumption that $\sum_{i=1}^{\ell} u_{[i]} > \sum_{i=1}^{\ell} x_{[i]}$ cannot hold. □

A function $\phi : \mathbb{Z}_+^k \to \mathbb{R}$ is said to be *Schur-concave* if for all $\mathbf{x}, \mathbf{y} \in \mathbb{Z}_+^k$,

$$\mathbf{x} \preceq \mathbf{y} \implies \phi(\mathbf{x}) \geq \phi(\mathbf{y}).$$

Note that if $\mathbf{x}$ and $\mathbf{y}$ are permutations of each other, then $\mathbf{x}_\downarrow = \mathbf{y}_\downarrow$, so that both $\mathbf{x} \preceq \mathbf{y}$ and $\mathbf{y} \preceq \mathbf{x}$ hold. Hence, a Schur-concave function ϕ must be *symmetric*, which means that $\phi(\mathbf{x}) = \phi(\mathbf{y})$ whenever $\mathbf{x}$ and $\mathbf{y}$ are permutations of each other.

A characterization of Schur-concave functions is given in [26, Chapter 3, A.2.b], which we adapt to our context in the proposition below.

PROPOSITION 5.3. *A function $\phi : \mathbb{Z}_+^k \to \mathbb{R}$ is Schur-concave iff it is symmetric, and for each choice of nonnegative integers $s, x_3, \ldots, x_k$, the function $\phi(x, s - x, x_3, \ldots, x_k)$ is monotonically decreasing in x for $x \geq s/2$.*

Now, define the function $\phi : \mathbb{Z}_+^k \to \mathbb{R}$ as follows:

$$(19) \qquad \phi(x_1, \ldots, x_k) = \prod_{j=1}^{k} \frac{\big(\sum_{\ell \neq j} x_\ell\big)!}{x_j!}.$$

An application of Proposition 5.3 shows that this function ϕ is Schur-concave. Indeed, ϕ is obviously symmetric. We claim that, for any choice of nonnegative integers $s, x_3, \ldots, x_k$, the function $\varphi(x) := \phi(x, s - x, x_3, \ldots, x_k)$ is monotonically decreasing in x for integers $x \geq s/2$. To see this, first verify that

$$\frac{\varphi(x+1)}{\varphi(x)} = \frac{(x + 1 + x_3 + \cdots + x_k)(s - x)}{(x + 1)(s - x + x_3 + \cdots + x_k)}.$$

Now, we must show that for $x \geq s/2$, this ratio is at most 1, or equivalently, $(x + 1 + x_3 + \cdots + x_k)(s-x) - (x+1)(s - x + x_3 + \cdots + x_k) \leq 0$. Upon some rearrangement and cancellation of like terms, the left-hand side becomes $(x_3 + \cdots + x_k)(s - x - (x+1))$, which is at most 0 when $x \geq s/2$.

Proof of Proposition 5.1. For any $A \in \mathcal{A}_{k,M}$, let $\mathbf{a}_1, \ldots, \mathbf{a}_k$ denote the rows of A. Note that $w(A) = \prod_{i=1}^{k} \phi(\mathbf{a}_i)$, where ϕ is as defined in (19).

Since all row sums of A are equal to M, the row vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ are all in $S_{k,M}$. By Lemma 5.2 and Schur-concavity of ϕ, we have

$$w(A) = \prod_{i=1}^{k} \phi(\mathbf{a}_i) \leq \prod_{i=1}^{k} \phi(\mathbf{u}) = w(U),$$

which proves the proposition. □

We remark that with a little bit of extra work, it can be shown that when k divides M (and $k \geq 3$), the matrix U, which in this case has all entries equal to M/k, uniquely maximizes $w(A)$ among $A \in \mathcal{A}_{k,M}$. When k does not divide M, any matrix obtained by permuting the rows and columns of U is also a maximizer.

5.2. Estimating $\mathbf{Var}_{k,M}(a_{1,1})$. Recalling that the variance of $a_{i,j}$ is independent of (i,j), we have

$$\text{Var}_{k,M}(a_{1,1}) = \frac{1}{k^2} \sum_{i,j} \text{Var}_{k,M}(a_{i,j}) = \frac{1}{k^2} \mathbb{E}\left[\sum_{i,j}(a_{i,j} - M/k)^2\right].$$

We will recast this in terms of the matrix U defined previously. Recall that this matrix has all its entries $u_{i,j}$ equal to either $\lfloor M/k \rfloor$ or $\lceil M/k \rceil$.

For any matrix $T = (t_{i,j})$, we use $\|T\|^2$ to denote the sum $\sum_{i,j} t_{i,j}^2$. Observe that

$$\mathbb{E}[\|A - U\|^2] = \mathbb{E}\left[\sum_{i,j}(a_{i,j} - u_{i,j})^2\right]$$

$$= \mathbb{E}\left[\sum_{i,j}\left((a_{i,j} - M/k) - (u_{i,j} - M/k)\right)^2\right]$$

$$= \mathbb{E}\left[\sum_{i,j}(a_{i,j} - M/k)^2\right] + \mathbb{E}\left[\sum_{i,j}(u_{i,j} - M/k)^2\right].$$

Hence,

$$(20) \qquad \left|\text{Var}_{k,M}(a_{1,1}) - \frac{1}{k^2}\mathbb{E}[\|A - U\|^2]\right| = \frac{1}{k^2}\mathbb{E}\left[\sum_{i,j}(u_{i,j} - M/k)^2\right] < 1.$$

Thus, for our purposes, it suffices to estimate $\mathbb{E}[\|A - U\|^2]$.

Our strategy to estimate $\mathbb{E}[\|A - U\|^2]$ is to approximate the distribution $Q_{k,M}$ by a suitably chosen discrete Gaussian distribution. The implementation of this strategy rests upon the following lemma, which we prove in Appendix D.

LEMMA 5.4. *Fix $k \geq 3$. Let $\rho = M/k$, and for $A \in \mathcal{A}_{k,M}$, let $T = (t_{i,j}) = A - U$. If $\rho \geq 4$ and $\max_{i,j}|t_{i,j}| \leq \frac{1}{9}\rho$, then*

$$\left|\log \frac{w(A)}{w(U)} - \log \frac{\widetilde{w}(A)}{\widetilde{w}(U)}\right| \leq \frac{4}{\rho^2}\sum_{i,j}(|t_{i,j}| + 1)^3 + \frac{3}{2\rho}\sum_{i,j}|t_{i,j}|,$$

where $\widetilde{w}(A) := \exp\{-\frac{1}{2} \cdot \frac{k-2}{k-1} \cdot \frac{1}{\rho}\sum_{i,j} t_{i,j}^2\}$.

We note in passing that $\widetilde{w}(U) = 1$, so that $\frac{\widetilde{w}(A)}{\widetilde{w}(U)}$ is simply $\widetilde{w}(A)$. Still, we have chosen to write this as $\frac{\widetilde{w}(A)}{\widetilde{w}(U)}$ so as to enable a direct comparison with $\frac{w(A)}{w(U)}$.

Taking cue from Lemma 5.4, we define a discrete Gaussian measure on the set, $\widetilde{\mathcal{A}}_{k,M}$, of all $k \times k$ integer (not necessarily nonnegative) matrices $A = (a_{i,j})$ with row and column sums equal to M. For any such matrix A, define

$$\widetilde{w}(A) := \exp\left\{-\frac{1}{2\sigma_{k,M}^2}\sum_{i,j} t_{i,j}^2\right\},$$

where $(t_{i,j}) = A - U$ and $\sigma_{k,M}^2 := \frac{k-1}{k-2}\left(\frac{M}{k}\right)$. We then define a probability measure $\widetilde{Q}_{k,M}$ on $\widetilde{\mathcal{A}}_{k,M}$ as follows: for $A \in \widetilde{\mathcal{A}}_{k,M}$,

$$\widetilde{Q}_{k,M}(A) := \frac{1}{\widetilde{Z}_{k,M}}\widetilde{w}(A),$$

where $\widetilde{Z}_{k,M} := \sum_{A \in \widetilde{\mathcal{A}}_{k,M}} \widetilde{w}(A)$.

Let $\widetilde{\mathbb{E}}[\cdot]$ denote expectation with respect to the measure $\widetilde{Q}_{k,M}$. To be clear, when we write $\widetilde{\mathbb{E}}[\|A - U\|^2]$, we mean $\sum_{A \in \widetilde{\mathcal{A}}_{k,M}} \|A - U\|^2\, \widetilde{Q}_{k,M}(A)$, while $\mathbb{E}[\|A - U\|^2]$ refers to $\sum_{A \in \mathcal{A}_{k,M}} \|A - U\|^2\, Q_{k,M}(A)$. The following lemma, proved in Appendix E, shows that $\mathbb{E}[\|A - U\|^2]$ is well-approximated by $\widetilde{\mathbb{E}}[\|A - U\|^2]$ as $M \to \infty$.

LEMMA 5.5. $\mathbb{E}[\|A - U\|^2] = \widetilde{\mathbb{E}}[\|A - U\|^2] + o(M)$.

The usefulness of this lemma stems from the fact that $\widetilde{\mathbb{E}}[\|A - U\|^2]$ can be estimated very accurately. To do this, we express $\widetilde{\mathbb{E}}[\|A - U\|^2]$ in a different form. For any $A \in \widetilde{\mathcal{A}}_{k,M}$, note that $A - U$ an integer matrix all of whose row and column sums are equal to 0, i.e., $A - U \in \widetilde{\mathcal{A}}_{k,0}$. For simplicity, let $\mathcal{T}_k := \widetilde{\mathcal{A}}_{k,0}$. We then have $\widetilde{\mathbb{E}}[\|A - U\|^2] = \frac{1}{\widetilde{Z}}\sum_{T \in \mathcal{T}_k} \|T\|^2 \exp\left\{-\frac{1}{2\sigma_{k,M}^2}\|T\|^2\right\}$, where $\widetilde{Z} := \sum_{T \in \mathcal{T}_k} \exp\left\{-\frac{1}{2\sigma_{k,M}^2}\|T\|^2\right\}$.

For any $T = (t_{i,j}) \in \mathcal{T}_k$, each of the entries in the kth row and column of T is linearly dependent on the entries $t_{i,j}$, $1 \le i, j \le k - 1$. To be precise,

$$t_{i,k} = -\sum_{j=1}^{k-1} t_{i,j}, \quad i = 1, \ldots, k - 1;$$

$$t_{k,j} = -\sum_{i=1}^{k-1} t_{i,j}, \quad j = 1, \ldots, k - 1;$$

$$t_{k,k} = -\sum_{i=1}^{k-1} t_{i,k} = \sum_{i=1}^{k-1}\sum_{j=1}^{k-1} t_{i,j}.$$

It follows that $\|T\|^2$ is a positive definite quadratic form in the $(k-1)^2$ variables $t_{i,j}$, $1 \le i, j \le k - 1$. Hence, we can write[6]

$$\|T\|^2 = \mathbf{t}'B\mathbf{t},$$

where $\mathbf{t} \in \mathbb{Z}^{(k-1)^2}$ is a vector with coordinates $t_{i,j}$, $1 \le i, j \le k - 1$ (in some fixed linear order), and B is a symmetric positive definite matrix. With this, we have

$$\widetilde{\mathbb{E}}[\|A - U\|^2] = \frac{1}{\widetilde{Z}} \sum_{\mathbf{t} \in \mathbb{Z}^{(k-1)^2}} \mathbf{t}'B\mathbf{t}\, \exp\left\{-\frac{1}{2\sigma_{k,M}^2}\mathbf{t}'B\mathbf{t}\right\},$$

where $\widetilde{Z}$ can now be written as

$$\widetilde{Z} = \sum_{\mathbf{t} \in \mathbb{Z}^{(k-1)^2}} \exp\left\{-\frac{1}{2\sigma_{k,M}^2}\mathbf{t}'B\mathbf{t}\right\}.$$

---

[6]To avoid notational confusion, we write $\mathbf{t}'$, $\mathbf{x}'$, etc., instead of $\mathbf{t}^T$, x^T, etc., to denote the transpose of $\mathbf{t}$, $\mathbf{x}$, etc.

Thus, we see that $\widetilde{\mathbb{E}}[\|A - U\|^2]$ is equal to the expected value of $\mathbf{t}'B\mathbf{t}$, where $\mathbf{t}$ is a random vector in $\mathbb{Z}^{(k-1)^2}$ distributed according to the discrete Gaussian measure $\frac{1}{Z} \exp\{-\frac{1}{2\sigma_{k,M}^2} \mathbf{t}'B\mathbf{t}\}$. The lemma below is a direct consequence of Proposition F.2 in Appendix F.

LEMMA 5.6. *We have*

$$\widetilde{\mathbb{E}}[\|A - U\|^2] = (k-1)^2\sigma_{k,M}^2 + O\big(\exp(-c_k M)\big),$$

where c_k is a positive constant depending only on k.

Equation (20) and Lemmas 5.5 and 5.6 suffice to prove that

$$\mathrm{Var}_{k,M}(a_{1,1}) = \frac{(k-1)^2}{k^2} \sigma_{k,M}^2 + o(M),$$

from which we obtain the second statement of Proposition 4.4(b).

6. Concluding remarks. The two main results of this paper, namely, Theorems 2.1 and 4.1, encapsulate an abrupt transition in behavior of the uniform distribution U_k in relation to the mapping $\beta_{k,M}^{\boldsymbol{\psi}} : \Pi_{[k]} \to \mathbb{R}_+$ defined by $\mathbf{p} \mapsto \mathrm{perm}_{\mathrm{B},M}(\Theta(\boldsymbol{\psi};\mathbf{p}))$. While the former theorem concerns the $M = 1$ case, the latter theorem holds for all sufficiently large M. Our proofs of these theorems involve picking an arbitrary unit vector $\boldsymbol{\xi} \in \mathbb{R}^k$ such that $\mathbf{p}(t) = U_k + t\boldsymbol{\xi}$ lies within $\Pi_{[k]}$ for all t within a sufficiently small interval around 0, and analyzing the first and second derivatives at $t = 0$ of the function $G_{k,M}(t) = \beta_{k,M}^{\boldsymbol{\psi}}(\mathbf{p}(t))$.

Unfortunately, we have not been able to extend this proof technique to deduce an analogous result for the mapping $\beta_{k,\infty}^{\boldsymbol{\psi}} : \Pi_{[k]} \to [0,1]$ defined by $\mathbf{p} \mapsto \mathrm{perm}_{\mathrm{B}}(\Theta(\boldsymbol{\psi};\mathbf{p}))$. The main technical obstacle here is the fact that the corresponding function $G_{k,\infty}(t) = \beta_{k,\infty}^{\boldsymbol{\psi}}(\mathbf{p}(t))$ may not be differentiable. (We do know that this function is continuous, since $\mathrm{perm}_{\mathrm{B}}(\boldsymbol{\psi};\mathbf{p})$ was noted to be a continuous function of $\mathbf{p}$ in the paragraph following (8).) Note that the identity (11) only allows us to say that $\limsup_{M\to\infty} G_{k,M}(t) = G_{k,\infty}(t)$ pointwise in t. By itself, this is insufficient to claim the differentiability of $G_{k,\infty}$.

It would be possible to prove the desired result if we could show that the functions $G_{k,M}$ satisfy the following two conditions within a sufficiently small interval I about 0: (i) $\lim_{M\to\infty} G_{k,M}(t)$ exists pointwise on I, and (ii) the first three derivatives of $G_{k,M}$ are uniformly bounded on I. Indeed, using the properties of equicontinuous and uniformly convergent functions (for example, Theorems 7.17 and 7.25 in [27]), we can then deduce that $G_{k,\infty}$ is twice-differentiable in the interior of I. Moreover, we would have $\lim_{M\to\infty} G_{k,M}(t) = G_{k,\infty}(t)$, $\lim_{M\to\infty} G'_{k,M}(t) = G'_{k,\infty}(t)$, and $\lim_{M\to\infty} G''_{k,M}(t) = G''_{k,\infty}(t)$ for all t in the interior of I. In particular, these hold at $t = 0$. From this, we could conclude, via Lemma 4.3 and Corollary 4.6, that the mapping $\beta_{k,\infty}^{\boldsymbol{\psi}}$ admits a phase transition at the same threshold as in Theorem 4.1.

We would unhesitatingly conjecture that condition (i) above is true, but we are less sure about condition (ii). It is actually not difficult to show that $G'_{k,M}$ is uniformly bounded within some small interval I, but the uniform boundedness of the second and third derivatives presents difficulties. It is in fact entirely possible that this approach will not work, as it may be the case that $G_{k,\infty}$ is not differentiable. However, we do believe that the ultimate result is still true: there is sufficient numerical evidence in favor of there being a phase transition at the threshold $\Upsilon_{\mathrm{B}}(\boldsymbol{\psi})$, defined in Theorem 4.1, for the uniform distribution in the Bethe PML problem.

Appendix A. Proof of Proposition 3.3. From Definition 3.2, we see that $\mathrm{perm}_{\mathrm{B},M}(\Theta)$ involves the permanents of $kM \times kM$ matrices $\Theta \odot \Lambda$ as in (10). Each such permanent involves a sum over permutations $\pi : [kM] \to [kM]$. We identify a permutation $\pi : [kM] \to [kM]$ with the permutation matrix in $\mathcal{P}_{kM}$ whose (s,t)th entry is a 1 iff $\pi(s) = t$; in a slight abuse of notation, we let π denote this permutation matrix as well. We will find it convenient to view any matrix $\pi \in \mathcal{P}_{kM}$ as a block matrix of the form

$$
(21) \qquad \pi = \begin{pmatrix} \pi^{(1,1)} & \pi^{(1,2)} & \cdots & \pi^{(1,k)} \\ \pi^{(2,1)} & \pi^{(2,2)} & \cdots & \pi^{(2,k)} \\ \vdots & \vdots & \ddots & \vdots \\ \pi^{(k,1)} & \pi^{(k,2)} & \cdots & \pi^{(k,k)} \end{pmatrix},
$$

where for $1 \le i, j \le k$, the (i,j)th block $\pi^{(i,j)}$ is the $M \times M$ submatrix of π located at the intersection of the rows and columns indexed by $(i-1)M + 1, \ldots, iM$ and $(j-1)M + 1, \ldots, jM$, respectively. Let $a_{i,j}(\pi)$ denote the number of 1s in $\pi^{(i,j)}$.

Given two 0/1-matrices $\mathsf{P} = (p_{i,j})$ and $\mathsf{Q} = (q_{i,j})$ of the same size, we write $\mathsf{P} \le \mathsf{Q}$ if for all i, j, we have $p_{i,j} \le q_{i,j}$ (or equivalently, $p_{i,j} = 1 \implies q_{i,j} = 1$). Using the newly introduced notation, we observe that

$$
(22) \qquad \mathrm{perm}(\Theta \odot \Lambda) = \sum_{\pi \in \mathcal{P}_{kM} : \pi \le \Lambda} \prod_{(i,j) \in [k]^2} \theta_{i,j}^{a_{i,j}(\pi)},
$$

since permutations $\pi \nleq \Lambda$ contribute only 0s to the permanent. Hence,

$$
\langle \mathrm{perm}(\Theta \odot \Lambda) \rangle = (M!)^{-k^2} \sum_{\Lambda \in \mathcal{P}_M^{k \times k}} \sum_{\pi \in \mathcal{P}_{kM} : \pi \le \Lambda} \prod_{(i,j) \in [k]^2} \theta_{i,j}^{a_{i,j}(\pi)}
$$
$$
(23) \qquad = (M!)^{-k^2} \sum_{\pi \in \mathcal{P}_{kM}} \sum_{\Lambda \in \mathcal{P}_M^{k \times k} : \pi \le \Lambda} \prod_{(i,j) \in [k]^2} \theta_{i,j}^{a_{i,j}(\pi)}.
$$

Now, for a given $\pi \in \mathcal{P}_{kM}$, the number of matrices Λ as in (9) such that $\pi \le \Lambda$ is equal to $\prod_{(i,j) \in [k]^2} (M - a_{i,j}(\pi))!$. This is because, for each (i,j), $\pi^{(i,j)}$ determines the positions of $a_{i,j}(\pi)$ 1s in $\Lambda^{(i,j)}$, and the positions of the remaining 1s in $\Lambda^{(i,j)}$ can be chosen in $(M - a_{i,j}(\pi))!$ ways to make $\Lambda^{(i,j)}$ a permutation matrix. Therefore, carrying on from (23), we have

$$
(24) \qquad \langle \mathrm{perm}(\Theta \odot \Lambda) \rangle = (M!)^{-k^2} \sum_{\pi \in \mathcal{P}_{kM}} \prod_{(i,j) \in [k]^2} (M - a_{i,j}(\pi))! \, \theta_{i,j}^{a_{i,j}(\pi)}.
$$

For any $\pi \in \mathcal{P}_{kM}$, the $k \times k$ matrix $A(\pi) := (a_{i,j}(\pi))$ is a nonnegative integer matrix whose row and column sums are all equal to M. Recall that, just prior to the statement of Proposition 3.3, we introduced the notation $\mathcal{A}_{k,M}$ to denote the set of all such $k \times k$ matrices. Thus, we can write (24) as

$$
(25) \qquad \langle \mathrm{perm}(\Theta \odot \Lambda) \rangle = (M!)^{-k^2} \sum_{A = (a_{i,j}) \in \mathcal{A}_{k,M}} |\mathcal{P}_{kM}(A)| \prod_{(i,j) \in [k]^2} (M - a_{i,j})! \, \theta_{i,j}^{a_{i,j}},
$$

where $\mathcal{P}_{kM}(A) := \{\pi \in \mathcal{P}_{kM} : A(\pi) = A\}$. Therefore, the proof of Proposition 3.3 would be complete once we prove the following lemma.

LEMMA A.1. *For $A = (a_{i,j}) \in \mathcal{A}_{k,M}$, we have*

$$|\mathcal{P}_{kM}(A)| = \frac{(M!)^{2k}}{\prod_{(i,j)\in[k]^2}(a_{i,j})!}.$$

Proof. Given a matrix $A = (a_{i,j}) \in \mathcal{A}_{k,M}$, we can construct permutation matrices $\pi \in \mathcal{P}_{kM}$ such that $A(\pi) = A$ by following the three steps described below. Our description views a $kM \times kM$ matrix π as a $k \times k$ block matrix as in (21).

1. Fix an $i \in [k]$. For each $j \in [k]$, pick $a_{i,j}$ rows of $\pi^{(i,j)}$ within which to place 1s. Since π cannot have two 1s in the same row, the number of ways in which these $a_{i,j}$ rows, $j = 1, \ldots, k$, can be picked is the multinomial coefficient $\binom{M}{a_{i,1},\ldots,a_{i,k}}$. Then, letting i range over $[k]$, we see that the number of ways in which rows of π can be so chosen is $\prod_i \binom{M}{a_{i,1},\ldots,a_{i,k}}$.

2. Fix a $j \in [k]$. For each $i \in [k]$, pick $a_{i,j}$ columns of $\pi^{(i,j)}$ within which to place 1s. By a similar argument as above, this can be done in $\binom{M}{a_{1,j},\ldots,a_{k,j}}$ ways. So, letting j range over $[k]$, we see that the number of ways in which columns of π can be so chosen is $\prod_j \binom{M}{a_{1,j},\ldots,a_{k,j}}$.

3. Fix a pair $(i,j) \in [k] \times [k]$, and consider the submatrix of $\pi^{(i,j)}$ formed by the points of intersection of the $a_{i,j}$ rows and columns chosen in the first two steps. For π to be a permutation matrix, this submatrix should be a permutation matrix as well. Hence, there are $(a_{i,j})!$ ways of placing 0s and 1s within this submatrix. All other entries of $\pi^{(i,j)}$ must be 0s. Letting (i,j) range over $[k] \times [k]$, we determine the number of possible choices in this step to be $\prod_{i,j}(a_{i,j})!$.

Thus, putting together the counts in the three steps, we obtain

$$|\mathcal{P}_{kM}(A)| = \left[\prod_i \binom{M}{a_{i,1},\ldots,a_{i,k}}\right]\left[\prod_j \binom{M}{a_{1,j},\ldots,a_{k,j}}\right]\left[\prod_{i,j}(a_{i,j})!\right],$$

which simplifies to the expression in the statement of the lemma. $\qquad\square$

Appendix B. Proof of Lemma 4.3. We introduce some convenient notation to be used in the proof. For $A = (a_{i,j}) \in \mathcal{A}_{k,M}$, define $\gamma_A(t) := \prod_{i,j}(p_i + t\xi_i)^{\mu_j a_{i,j}}$ and $\gamma(t) := \sum_{A \in \mathcal{A}_{k,M}} w(A)\gamma_A(t)$. We will need the values of $\gamma(0)$, $\gamma'(0)$, and $\gamma''(0)$, for which we need to compute $\gamma_A(0)$, $\gamma'_A(0)$, and $\gamma''_A(0)$.

Determining $\gamma_A(0)$ is easy: since $p_i = \frac{1}{k}$ for all $i \in [k]$, we have

$$\gamma_A(0) = \prod_{i,j}\left(\frac{1}{k}\right)^{\mu_j a_{i,j}} = k^{-\sum_{i,j}\mu_j a_{i,j}} = k^{-Mn},$$

the last equality using $\sum_i a_{i,j} = M$, and $\sum_j \mu_j = n$. Hence,

$$(26) \qquad \gamma(0) = \sum_A w(A)\gamma_A(0) = k^{-Mn}Z_{k,M}.$$

Similarly, taking the derivative of $\gamma_A(t)$, it is straightforward to show that

$$\gamma'_A(0) = k^{1-Mn}\sum_{i,j}\xi_i\mu_j a_{i,j},$$

and hence,
$$\gamma'(0) = \sum_A w(A)\gamma'_A(0) = k^{1-Mn} \sum_{i,j} \xi_i \mu_j \sum_A a_{i,j} w(A).$$

Now, $\sum_A a_{i,j} w(A) = Z_{k,M}\mathbb{E}[a_{i,j}]$, where $\mathbb{E}[\cdot]$ denotes expectation taken with respect to a random matrix $A \in \mathcal{A}_{k,M}$ distributed according to $Q_{k,M}$. By Lemma 4, $\mathbb{E}[a_{i,j}]$ is a constant independent of (i,j), and hence,

$$(27) \qquad \gamma'(0) = k^{1-Mn} Z_{k,M}(\text{const.})\left(\sum_i \xi_i\right)\left(\sum_j \mu_j\right) = 0,$$

since $\sum_i \xi_i = 0$ by choice of the direction vector $\boldsymbol{\xi}$.

Calculation of the second derivative $\gamma''(0)$ requires a lot more work. To start with, routine differentiation yields

$$\gamma''_A(0) = k^{2-Mn}\left[\left(\sum_{i,j} \xi_i \mu_j a_{i,j}\right)^2 - \sum_{i,j} \xi_i^2 \mu_j a_{i,j}\right],$$

which we can plug into

$$\gamma''(0) = \sum_A w(A)\gamma''_A(0) = Z_{k,M}\mathbb{E}[\gamma''_A(0)]$$

to get

$$\gamma''(0) = Z_{k,M}k^{2-Mn}\left(\mathbb{E}\left[\left(\sum_{i,j} \xi_i \mu_j a_{i,j}\right)^2\right] - \sum_{i,j} \xi_i^2 \mu_j \mathbb{E}[a_{i,j}]\right)$$

$$(28) \qquad = Z_{k,M}k^{2-Mn}\left(\mathbb{E}\left[\left(\sum_{i,j} \xi_i \mu_j a_{i,j}\right)^2\right] - \frac{Mn}{k}\right).$$

For the second equality above, we used Lemma 4 and $\sum_{i,j} \xi_i^2 \mu_j = (\sum_i \xi_i^2)(\sum_j \mu_j) = n$, since $\|\boldsymbol{\xi}\|_2 = 1$.

To determine $\mathbb{E}\left[\left(\sum_{i,j} \xi_i \mu_j a_{i,j}\right)^2\right]$, we write

$$(29) \qquad \mathbb{E}\left[\left(\sum_{i,j} \xi_i \mu_j a_{i,j}\right)^2\right] = \sum_{i,j} \xi_i^2 \mu_j^2 \mathbb{E}[a_{i,j}^2] + \sum_{\substack{(i,j),(i',j'): \\ (i,j)\neq(i',j')}} \xi_i \xi_{i'} \mu_j \mu_{j'} \mathbb{E}[a_{i,j} a_{i',j'}].$$

As argued for $\mathbb{E}[a_{i,j}]$, the expected value $\mathbb{E}[a_{i,j}^2]$ is also a constant independent of (i,j). With this, the first term on the right-hand side of (29) can be expressed as

$$(30) \qquad \sum_{i,j} \xi_i^2 \mu_j^2 \mathbb{E}[a_{i,j}^2] = \mathbb{E}[a_{1,1}^2] \sum_j \mu_j^2$$

using $\|\boldsymbol{\xi}\|_2 = 1$.

Turning our attention to the second term on the right-hand side of (29), we note that for $(i,j) \neq (i',j')$, by virtue of the invariance of $Q_{k,M}$ with respect to row and column permutations,

$$(31) \qquad \mathbb{E}[a_{i,j} a_{i',j'}] = \begin{cases} \mathbb{E}[a_{1,1}a_{2,2}] & \text{if } i \neq i' \text{ and } j \neq j', \\ \mathbb{E}[a_{1,1}a_{1,2}] & \text{otherwise.} \end{cases}$$

Now,

$$\mathbb{E}[a_{1,1}a_{1,2}] = \frac{1}{k-1}\sum_{j=2}^{k}\mathbb{E}[a_{1,1}a_{1,j}] = \frac{1}{k-1}\mathbb{E}\left[a_{1,1}\sum_{j=2}^{k}a_{1,j}\right]$$

$$(32) \qquad = \frac{1}{k-1}\mathbb{E}[a_{1,1}(M-a_{1,1})] = \frac{1}{k-1}\left[\frac{M^2}{k} - \mathbb{E}[a_{1,1}^2]\right],$$

where we used Lemma 4 to get the last equality. By a similar argument,

$$\mathbb{E}[a_{1,1}a_{2,2}] = \frac{1}{k-1}\mathbb{E}\left[a_{1,1}\sum_{i=2}^{k}a_{i,2}\right] = \frac{1}{k-1}\mathbb{E}[a_{1,1}(M-a_{1,2})]$$

$$(33) \qquad = \frac{1}{k-1}\left[\frac{M^2}{k} - \mathbb{E}[a_{1,1}a_{1,2}]\right] = \frac{1}{(k-1)^2}\left[\frac{k-2}{k}M^2 + \mathbb{E}[a_{1,1}^2]\right],$$

this time using (32) to get the last equality. Plugging (30)–(33) into (29), and then performing some careful bookkeeping, we eventually obtain

$$\mathbb{E}\left[\left(\sum_{i,j}\xi_i\mu_j a_{i,j}\right)^2\right] = \frac{k}{(k-1)^2}\operatorname{Var}_{k,M}(a_{1,1})\left[k\sum_j\mu_j^2 - n^2\right],$$

and hence,

$$(34) \qquad \gamma''(0) = Z_{k,M}k^{2-Mn}\left(\frac{k}{(k-1)^2}\operatorname{Var}_{k,M}(a_{1,1})\left[k\sum_j\mu_j^2 - n^2\right] - \frac{Mn}{k}\right).$$

Finally, by Proposition 3.3, we have

$$G_{k,M}(t) = \operatorname{perm}_{\mathrm{B},M}(\Theta(\boldsymbol{\psi};\mathbf{p}(t))) = [(M!)^{2k-k^2}]^{\frac{1}{M}}\gamma(t)^{\frac{1}{M}}.$$

Taking the derivative with respect to t and setting $t = 0$, we obtain

$$G'_{k,M}(0) = [(M!)^{2k-k^2}]^{\frac{1}{M}}\frac{1}{M}\gamma(0)^{\frac{1}{M}-1}\gamma'(0) = 0,$$

since $\gamma'(0) = 0$—see (27).

Similarly, calculating the second derivative at $t = 0$, we get

$$G''_{k,M}(0) = [(M!)^{2k-k^2}]^{\frac{1}{M}}\frac{1}{M}\gamma(0)^{\frac{1}{M}}\frac{\gamma''(0)}{\gamma(0)}.$$

Plugging in the expressions for $\gamma(0)$ and $\gamma''(0)$ given in (26) and (34), respectively, we obtain the expression for $G''_{k,M}(0)$ recorded in the statement of Lemma 4.3.

Appendix C. Proof of Lemma 4.7. Part (1) of the lemma is obvious, so we concern ourselves with part (2). Here, there are two claims that need proof:

Claim C.1. The discriminant D is strictly negative iff $\Upsilon < \frac{\sqrt{n}+1}{\sqrt{n}-1}$.

Claim C.2. When $D \geq 0$, so that real roots ρ_1 and ρ_2 exist, we have $1 < \rho_1 \leq 2$ and $\Upsilon - 1 \leq \rho_2 < \Upsilon$.

Proof of Claim C.1. Note that $D = (n^2+2n-\mathcal{U})^2 - 4n^3 < 0$ iff $|n^2+2n-\mathcal{U}| < 2n^{3/2}$. We may remove the absolute value in the latter condition since $n^2 > \mathcal{U}$. Thus, $D < 0$

iff $n^2 + 2n - \mathcal{U} < 2n^{3/2}$, which upon some rearrangement becomes $\mathcal{U} - n > (n - \sqrt{n})^2$. Thus, recalling that $\Upsilon = \frac{n^2 - n}{\mathcal{U} - n}$, we see that $D < 0$ is equivalent to

$$\Upsilon < \frac{n^2 - n}{(n - \sqrt{n})^2}.$$

Upon canceling the common factor $\sqrt{n}(n - \sqrt{n})$, the right-hand side simplifies to $\frac{\sqrt{n}+1}{\sqrt{n}-1}$. □

Proof of Claim C.2. We start by writing $\rho_1 = 1 + \frac{n^2 - \mathcal{U} - \sqrt{D}}{2(\mathcal{U} - n)}$. To show that $\rho_1 > 1$, it suffices to show that $n^2 - \mathcal{U} > \sqrt{D}$, or equivalently, $(n^2 - \mathcal{U})^2 > D$. Routine algebra shows that $(n^2 - \mathcal{U})^2 - D = 4n\mathcal{U}$, which is of course positive.

For the rest of this proof, it will be convenient to define $a = \mathcal{U} - n$, $b = \mathcal{U} + n^2 - 2n$ and $c = n^2$, so that $q(x) = ax^2 - bx + c$. With this, $\rho_1 = \frac{b - \sqrt{b^2 - 4ac}}{2a}$.

We now give a proof for $\rho_1 \leq 2$. Suppose that $\rho_1 \geq 2$. We would then have $b - 4a \geq \sqrt{b^2 - 4ac}$. This yields two inequalities that must necessarily be satisfied: $b - 4a \geq 0$ and $(b-4a)^2 \geq \sqrt{b^2 - 4ac}$. The latter inequality simplifies to $4a - 2b + c \geq 0$, so that the two inequalities that must be satisfied are

$$(35) \qquad b - 4a \geq 0 \quad \text{and} \quad 4a - 2b + c \geq 0.$$

Plugging in the expressions for a, b, and c, we obtain $n^2 + 2n \geq 3\mathcal{U}$ and $2\mathcal{U} \geq n^2$. Combining these inequalities, we get

$$(36) \qquad \frac{1}{2}n^2 \leq \mathcal{U} \leq \frac{1}{3}(n^2 + 2n),$$

and hence, $\frac{1}{2}n^2 \leq \frac{1}{3}(n^2 + 2n)$. Upon rearrangement, this becomes $n^2 - 4n \leq 0$, which yields $0 \leq n \leq 4$.

As we do not consider pattern lengths $n < 2$, we must deal with $n = 2, 3, 4$. Plugging these values of n into (36), we obtain $(n, \mathcal{U}) = (2, 2)$, $(3, 5)$, and $(4, 8)$ as the only valid solutions. The assumption of part (2) of the lemma is that $\mathcal{U} > n$, so we cannot have $(n, \mathcal{U}) = (2, 2)$. Also, $(n, \mathcal{U}) = (3, 5)$ is not possible as this yields a negative discriminant. We are thus forced to conclude that if $\rho_1 \geq 2$, then $(n, \mathcal{U}) = (4, 8)$. Indeed, in this case, we have $q(x) = 4x^2 - 16x + 16$, so that $\rho_1 = \rho_2 = 2$. We have thus proved that $\rho_1 \leq 2$ always holds, and in fact, it holds with equality iff $(n, \mathcal{U}) = (4, 8)$, i.e., $\boldsymbol{\psi} = 1^2 2^2$.

To prove that $\Upsilon - 1 \leq \rho_2 < \Upsilon$, consider the difference $\rho_2 - \Upsilon$. It may be verified that this difference can be expressed as

$$\rho_2 - \Upsilon = \frac{1}{2(\mathcal{U} - n)} \left[-(n^2 - \mathcal{U}) + \sqrt{(n^2 - \mathcal{U})^2 - 4n(\mathcal{U} - n)} \right],$$

which is obviously strictly negative, given that $\mathcal{U} > n$ and $n^2 > \mathcal{U}$. Thus, $\rho_2 < \Upsilon$.

Now, suppose $\rho_2 - \Upsilon \leq -1$. Then, using the expression given above for $\rho_2 - \Upsilon$, we must have $\sqrt{(n^2 - \mathcal{U})^2 - 4n(\mathcal{U} - n)} \leq n^2 + 2n - 3\mathcal{U}$. This yields two inequalities to be satisfied: $n^2 + 2n - 3\mathcal{U} \geq 0$ and $(n^2 - \mathcal{U})^2 - 4n(\mathcal{U} - n) \leq (n^2 + 2n - 3\mathcal{U})^2$. The latter inequality can be manipulated into the following equivalent form: $4(\mathcal{U} - n)(2\mathcal{U} - n^2) \geq 0$. Since $\mathcal{U} > n$, this inequality is satisfied iff $2\mathcal{U} \geq n^2$. Thus, $\rho_2 - \Upsilon \leq -1$ only if the inequalities in (36) hold. As argued earlier, these inequalities are satisfied only if $(n, \mathcal{U}) = (4, 8)$, in which case it may be verified that $\rho_2 - \Upsilon = -1$. This proves

that $\rho_2 \geq \Upsilon - 1$ always holds, and again, it holds with equality iff $(n, \mathcal{U}) = (4, 8)$, i.e., $\psi = 1^2 2^2$.   □

This completes the proof of Lemma 4.7.

Appendix D. Proof of Lemma 5.4. Throughout the proof, we fix $k \geq 3$. We introduce some convenient notation: $\rho := M/k$, $\varepsilon_{i,j} := u_{i,j} - \rho$, and $f(x) := x(M + 1 - x)$. Now, consider the ratio $w(A)/w(U) = w(U + T)/w(U)$:

$$\frac{w(U + T)}{w(U)}$$

$$= \prod_{(i,j):t_{i,j} \neq 0} \frac{(M - (u_{i,j} + t_{i,j}))!}{(M - u_{i,j})!} \frac{u_{i,j}!}{(u_{i,j} + t_{i,j})!}$$

$$= \prod_{(i,j):t_{i,j} < 0} f(u_{i,j} - |t_{i,j}| + 1) \cdots f(u_{i,j}) \prod_{(i,j):t_{i,j} > 0} \frac{1}{f(u_{i,j} + 1) \cdots f(u_{i,j} + t_{i,j})}$$

$$= \prod_{(i,j):t_{i,j} < 0} \frac{f(u_{i,j} - |t_{i,j}| + 1) \cdots f(u_{i,j})}{f(\rho)^{|t_{i,j}|}} \prod_{(i,j):t_{i,j} > 0} \frac{f(\rho)^{t_{i,j}}}{f(u_{i,j} + 1) \cdots f(u_{i,j} + t_{i,j})}.$$

The last equality above holds because $\sum_{i,j} t_{i,j} = 0$. Thus,
(37)
$$\log \frac{w(U + T)}{w(U)} = \sum_{(i,j):t_{i,j} < 0} \sum_{\ell = -|t_{i,j}| + 1}^{0} \log \frac{f(u_{i,j} + \ell)}{f(\rho)} - \sum_{(i,j):t_{i,j} > 0} \sum_{\ell = 1}^{t_{i,j}} \log \frac{f(u_{i,j} + \ell)}{f(\rho)}.$$

Now, we need estimates for the summands $\log \frac{f(u_{i,j} + \ell)}{f(\rho)}$. Observe that, since $u_{i,j} + t_{i,j} = a_{i,j} \geq 0$, the integers ℓ that appear in (37) all satisfy $u_{i,j} + \ell \geq 1$. We will make use of this observation a little later.

We first derive useful estimates for the ratios $\frac{f(u_{i,j} + \ell)}{f(\rho)}$. Note that $f(x_2) - f(x_1) = (x_2 - x_1)(M + 1 - (x_1 + x_2))$. Using this and the fact that $M = k\rho$, we can write, for $\ell \in \mathbb{Z}$,

$$\frac{f(u_{i,j} + \ell)}{f(\rho)} = 1 + \frac{f(u_{i,j} + \ell) - f(\rho)}{f(\rho)}$$

$$= 1 + \frac{(\ell + \varepsilon_{i,j})((k - 2)\rho - \ell - \varepsilon_{i,j} + 1)}{\rho((k - 1)\rho + 1)}$$

(38)
$$= 1 + \left(\frac{\ell + \varepsilon_{i,j}}{\rho} \right) \left(\frac{k - 2}{k - 1} \right) (1 + \gamma(\ell)),$$

where $1 + \gamma(\ell) = \left(1 - \frac{\ell + \varepsilon_{i,j} - 1}{(k-2)\rho} \right) \left(1 + \frac{1}{(k-1)\rho} \right)^{-1}$. Observe that

(39)
$$1 + \gamma(\ell) \leq 1 - \frac{\ell + \varepsilon_{i,j} - 1}{(k-2)\rho} \leq 1 + \frac{|\ell| + 2}{(k-2)\rho},$$

using $\varepsilon_{i,j} \in (-1, 1)$. On the other hand, using $\left(1 + \frac{1}{(k-1)\rho} \right)^{-1} \geq 1 - \frac{1}{(k-1)\rho}$, we also have

$$1 + \gamma(\ell) \geq \left(1 - \frac{\ell + \varepsilon_{i,j} - 1}{(k-2)\rho} \right) \left(1 - \frac{1}{(k-1)\rho} \right)$$

$$= 1 - \frac{\ell + \varepsilon_{i,j}}{(k-2)\rho} + \frac{\rho + \ell + \varepsilon_{i,j} - 1}{(k-1)(k-2)\rho^2}$$

$$= 1 - \frac{\ell + \varepsilon_{i,j}}{(k-2)\rho} + \frac{\ell + u_{i,j} - 1}{(k-1)(k-2)\rho^2}$$

$$(40) \qquad \geq 1 - \frac{\ell + \varepsilon_{i,j}}{(k-2)\rho}$$

since, as observed earlier, $\ell + u_{i,j} \geq 1$ for the integers ℓ that appear in (37). It should also be pointed out that the first inequality above requires $1 - \frac{\ell+\varepsilon_{i,j}-1}{(k-2)\rho}$ to be non-negative. If $1 - \frac{\ell+\varepsilon_{i,j}-1}{(k-2)\rho} < 0$, then $1 + \gamma(\ell)$ is still lower bounded by (40), since we now have $1 + \gamma(\ell) = \left(1 - \frac{\ell+\varepsilon_{i,j}-1}{(k-2)\rho}\right)\left(1 + \frac{1}{(k-1)\rho}\right)^{-1} \geq 1 - \frac{\ell+\varepsilon_{i,j}-1}{(k-2)\rho}$.

Thus, from (38)–(40), we obtain for any integer ℓ occurring in (37),

$$(41) \qquad \frac{f(u_{i,j} + \ell)}{f(\rho)} = 1 + \zeta(\ell)(1 + \gamma(\ell)),$$

where $\zeta(\ell) = (\frac{k-2}{k-1})(\frac{\ell+\varepsilon_{i,j}}{\rho})$, and $|\gamma(\ell)| \leq \frac{|\ell|+2}{(k-2)\rho}$. Observe that for $|\ell| \geq 1$, $|\zeta(\ell)(1 + \gamma(\ell))| \leq \frac{2|\ell|}{\rho}(1 + \frac{3|\ell|}{\rho})$, which is at most $\frac{1}{2}$ for $\frac{|\ell|}{\rho} \leq \frac{1}{9}$. (Indeed, it is easy to check that $2x(1+3x) \leq \frac{1}{2}$ for $|x| \leq \frac{1}{12}(\sqrt{15}-2) = 0.1560\ldots$.) Also, verify that $|\zeta(0)(1+\gamma(0))| \leq \frac{1}{\rho}(1 + \frac{2}{(k-2)\rho}) \leq \frac{1}{\rho}(1 + \frac{2}{\rho})$, which is at most $\frac{3}{8}$ for $\rho \geq 4$. Henceforth, we assume $\rho \geq 4$ and $\frac{|\ell|}{\rho} \leq \frac{1}{9}$, in keeping with the hypotheses of the lemma.

From (41), via the inequality $x - x^2 \leq \log(1+x) \leq x$, valid for $|x| \leq \frac{1}{2}$, we obtain

$$-\left[\zeta(\ell)(1 + \gamma(\ell))\right]^2 \leq \log \frac{f(u_{i,j} + \ell)}{f(\rho)} - \zeta(\ell)(1 + \gamma(\ell)) \leq 0,$$

and hence,

$$\zeta(\ell)\gamma(\ell) - \left[\zeta(\ell)(1 + \gamma(\ell))\right]^2 \leq \log \frac{f(u_{i,j} + \ell)}{f(\rho)} - \zeta(\ell) \leq \zeta(\ell)\gamma(\ell).$$

It follows that

$$(42) \qquad \left| \log \frac{f(u_{i,j} + \ell)}{f(\rho)} - \zeta(\ell) \right| \leq |\zeta(\ell)\gamma(\ell)| + \left[\zeta(\ell)(1 + \gamma(\ell))\right]^2.$$

Now, $|\zeta(\ell)\gamma(\ell)| \leq \frac{1}{k-1}\frac{(|\ell|+1)(|\ell|+2)}{\rho^2} \leq \frac{2}{k-1}(\frac{|\ell|+1}{\rho})^2 \leq (\frac{|\ell|+1}{\rho})^2$. Furthermore, $|1+\gamma(\ell)| \leq 1 + |\gamma(\ell)| \leq 1 + \frac{|\ell|+2}{\rho} \leq 1 + \frac{1}{9} + \frac{1}{4} = \frac{29}{18}$, as we have assumed $\frac{|\ell|}{\rho} \leq \frac{1}{9}$ and $\rho \geq 4$. From this, we get $|(\zeta(\ell)(1 + \gamma(\ell))| \leq \frac{29}{18}\frac{|\ell|+1}{\rho}$. Plugging these estimates into (42), we obtain

$$(43) \qquad \left| \log \frac{f(u_{i,j} + \ell)}{f(\rho)} - \zeta(\ell) \right| \leq \left(1 + \frac{29^2}{18^2}\right)\left(\frac{|\ell|+1}{\rho}\right)^2 \leq 4\left(\frac{|\ell|+1}{\rho}\right)^2.$$

From (43), we can deduce estimates for sums of the form $\sum_\ell \log \frac{f(u_{i,j}+\ell)}{f(\rho)}$, which we can use in (37). Indeed, for an integer $t < 0$, with $|t| \leq \frac{1}{9}\rho$, we have

$$\left| \sum_{\ell=-|t|+1}^{0} \log \frac{f(u_{i,j} + \ell)}{f(\rho)} - \sum_{\ell=-|t|+1}^{0} \zeta(\ell) \right| \leq \frac{4}{\rho^2} \sum_{\ell=-|t|+1}^{0} (|\ell| + 1)^2,$$

which yields

$$\left| \sum_{\ell=-|t|+1}^{0} \log \frac{f(u_{i,j}+\ell)}{f(\rho)} + \frac{1}{2}\left(\frac{k-2}{k-1}\right)\left(\frac{1}{\rho}\right)\left[|t|^2 + |t|(2\varepsilon_{i,j}-1)\right]\right|$$

$$\leq \frac{2}{3\rho^2}\left[|t|(|t|+1)(2|t|+1)\right].$$

The above bound can be brought into the following looser but simpler form:

$$(44) \qquad \left| \sum_{\ell=-|t|+1}^{0} \log \frac{f(u_{i,j}+\ell)}{f(\rho)} + \frac{1}{2}\left(\frac{k-2}{k-1}\right)\left(\frac{t^2}{\rho}\right)\right| \leq \frac{4|t|^3}{\rho^2} + \frac{3|t|}{2\rho}.$$

Similarly, for an integer $0 < t \leq \frac{1}{9}\rho$, we can obtain

$$(45) \qquad \left| \sum_{\ell=1}^{t} \log \frac{f(u_{i,j}+\ell)}{f(\rho)} - \frac{1}{2}\left(\frac{k-2}{k-1}\right)\left(\frac{t^2}{\rho}\right)\right| \leq \frac{4(t+1)^3}{\rho^2} + \frac{3t}{2\rho}.$$

Lemma 5.4 readily follows from (37), (44), and (45).

Appendix E. Proof of Lemma 5.5. As in Lemma 5.4 and Appendix D, we fix $k \geq 3$ and set $\rho = \frac{M}{k}$. We assume $\rho \geq 4$ throughout.

We will use the notation introduced after the statement of Lemma 5.5 in section 5.2. In particular, the measure $\widetilde{Q}_{k,M}$ defined on the set $\widetilde{\mathcal{A}}_{k,M}$ is equivalent to a discrete Gaussian measure on $\mathbb{Z}^{(k-1)^2}$, which assigns to each $\mathbf{t} \in \mathbb{Z}^{(k-1)^2}$ the mass $\frac{1}{\widetilde{Z}} \exp\left(-\frac{1}{2\sigma_{k,M}^2}\mathbf{t}'B\mathbf{t}\right)$, where $\sigma_{k,M}^2 = \frac{k-1}{k-2}\rho$ and B is a symmetric positive definite matrix.

Let $\delta \in (0, \frac{1}{6})$ be fixed, and define

$$\mathcal{A}_{k,M}(\delta) := \left\{A \in \mathcal{A}_{k,M} : \max_{i,j}|t_{i,j}| \leq \rho^{\frac{1}{2}+\delta}\right\},$$

where $T = (t_{i,j}) = A - U$. By Lemma 5.4, it follows that for any $A \in \mathcal{A}_{k,M}(\delta)$, we have

$$\left|\log \frac{w(A)}{w(U)} - \log \frac{\widetilde{w}(A)}{\widetilde{w}(U)}\right| \leq \hat{c}_k \rho^{-\frac{1}{2}+3\delta},$$

where $\hat{c}_k$ is a positive constant depending only on k. Thus, for $A \in \mathcal{A}_{k,M}(\delta)$, we have

$$(46) \qquad \widetilde{w}(A)\exp(-\hat{c}_k\rho^{-\frac{1}{2}+3\delta}) \leq \frac{w(A)}{w(U)} \leq \widetilde{w}(A)\exp(\hat{c}_k\rho^{-\frac{1}{2}+3\delta}),$$

noting that $\widetilde{w}(U) = 1$. Since $-\frac{1}{2} + 3\delta < 0$, this shows that, as $M \to \infty$ (so that $\rho \to \infty$), the ratio $w(A)/w(U)$ is well-approximated by $\widetilde{w}(A)$ for all $A \in \mathcal{A}_{k,M}(\delta)$. From this, we will be able to deduce that, as $M \to \infty$, the contributions made to $\mathbb{E}[\|A-U\|^2]$ and $\widetilde{\mathbb{E}}[\|A-U\|^2]$ by matrices in $A \in \mathcal{A}_{k,M}(\delta)$ are nearly the same. The next two lemmas show that the matrices outside $\mathcal{A}_{k,M}(\delta)$ make vanishingly small contributions to both the expected values.

LEMMA E.1. *Let* $\widetilde{\mathcal{B}}_{k,M}(\delta) = \widetilde{\mathcal{A}}_{k,M} \setminus \mathcal{A}_{k,M}(\delta)$. *We have*

$$\sum_{A \in \widetilde{\mathcal{B}}_{k,M}(\delta)} \|A-U\|^2 \widetilde{Q}_{k,M}(A) \leq \kappa_k\, \rho^2 \exp\left(-\frac{k-2}{2k}\rho^{2\delta}\right),$$

where κ_k *is a constant depending only on* k.

Proof. For any $A \in \widetilde{\mathcal{A}}_{k,M}$, if the entries of $T = A - U$ are bounded above in magnitude by $\rho^{\frac{1}{2}+\delta}$, then A must have nonnegative entries, and hence, $A \in \mathcal{A}_{k,M}(\delta)$. Therefore, for any $A \in \widetilde{\mathcal{B}}_{k,M}(\delta)$, we have $\|A - U\|^2 = \sum_{i,j} t_{i,j}^2 \geq \rho^{1+2\delta}$. Hence,

$$\sum_{A \in \widetilde{\mathcal{B}}_{k,M}(\delta)} \|A - U\|^2 \, \widetilde{Q}_{k,M}(A) \leq \sum_{A \in \widetilde{\mathcal{A}}_{k,M} : \|A-U\|^2 \geq \rho^{1+2\delta}} \|A - U\|^2 \, \widetilde{Q}_{k,M}(A),$$

and the lemma follows by applying Proposition F.3 in Appendix F with $R = \rho^{1+2\delta}$ and $\tau = \frac{1}{k}$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

LEMMA E.2. *Let* $\mathcal{B}_{k,M}(\delta) = \mathcal{A}_{k,M} \setminus \mathcal{A}_{k,M}(\delta)$. *There is a positive constant* c_k' *depending only on* k *such that for all* $A \in \mathcal{B}_{k,M}(\delta)$, *the bound*

$$(47) \qquad\qquad\qquad\qquad \frac{w(A)}{w(U)} \leq \exp(-c_k' \rho^{2\delta})$$

holds for all sufficiently large M. *Consequently, there is a positive constant* c_k'' *depending only on* k *such that*

$$(48) \qquad\qquad \sum_{A \in \mathcal{B}_{k,M}(\delta)} \|A - U\|^2 \, \frac{w(A)}{w(U)} \leq \exp(-c_k'' \rho^{2\delta})$$

holds for all sufficiently large M.

Proof. Given the bound in (47), the bound in (48) follows readily. Indeed, from the definition of U given in section 5.1, it is straightforward to verify that for any $A \in \mathcal{A}_{k,M}$, we have

$$\|A - U\|^2 = \|A\|^2 - k^2 q^2 + kr - 2\sum_{i=1}^{k} \sum_{j:u_{i,j}=q+1} a_{i,j} \leq \|A\|^2 - k^2 q^2 + kr.$$

Hence, for all $M \geq k$ (so that $q \geq 1$), we have $\|A - U\|^2 \leq \|A\|^2 \leq k^2 M^2$. Also, note that $|\mathcal{B}_{k,M}(\delta)| \leq |\mathcal{A}_{k,M}| \leq (M+1)^{k^2}$. Therefore,

$$\sum_{A \in \mathcal{B}_{k,M}(\delta)} \|A - U\|^2 \, \frac{w(A)}{w(U)} \leq k^2 M^2 |\mathcal{B}_{k,M}(\delta)| \exp(-c_k' \rho^{2\delta})$$

$$\leq \exp\big(-c_k' \rho^{2\delta} + O(\log M)\big),$$

from which (48) follows.

The proof of (47) builds on (37). Recall that $A - U = T$, and note that $f(x)/f(\rho) \geq 1$ iff $\rho \leq x \leq M + 1 - \rho$. For all (i,j), define $\hat{t}_{i,j} = \min\{|t_{i,j}|, \rho^{\frac{1}{2}+\delta}\}$. Then, for $t_{i,j} < 0$, we have

$$(49) \qquad\qquad \sum_{\ell=-|t_{i,j}|+1}^{0} \log \frac{f(u_{i,j}+\ell)}{f(\rho)} \leq \sum_{\ell=-\hat{t}_{i,j}+1}^{0} \log \frac{f(u_{i,j}+\ell)}{f(\rho)}.$$

Also, for any $t_{i,j} > 0$ such that $a_{i,j} = u_{i,j} + t_{i,j} \leq M + 1 - \rho$, we have

$$(50) \qquad\qquad \sum_{\ell=1}^{t_{i,j}} \log \frac{f(u_{i,j}+\ell)}{f(\rho)} \geq \sum_{\ell=1}^{\hat{t}_{i,j}} \log \frac{f(u_{i,j}+\ell)}{f(\rho)}.$$

Now, consider $A \in \mathcal{B}_{k,M}(\delta)$. Suppose first that $a_{i,j} \leq M + 1 - \rho$ for all (i,j). Then, from (37), (49), and (50), we have

$$(51) \quad \log \frac{w(A)}{w(U)} \leq \sum_{(i,j):t_{i,j}<0} \sum_{\ell=-\hat{t}_{i,j}+1}^{0} \log \frac{f(u_{i,j}+\ell)}{f(\rho)} - \sum_{(i,j):t_{i,j}>0} \sum_{\ell=1}^{\hat{t}_{i,j}} \log \frac{f(u_{i,j}+\ell)}{f(\rho)}.$$

Now arguing as in the proof of Lemma 5.4 in Appendix D (in particular, using (44) and (45)), we can bound the right-hand side above by

$$(52) \qquad -\frac{1}{2}\left(\frac{k-2}{k-1}\right)\frac{1}{\rho}\sum_{i,j}(\hat{t}_{i,j})^2 + \frac{4}{\rho^2}\sum_{i,j}(\hat{t}_{i,j}+1)^3 + \frac{3}{2\rho}\sum_{i,j}\hat{t}_{i,j}.$$

Now, since $A \in \mathcal{B}_{k,M}(\delta)$, we have $\sum_{i,j}(\hat{t}_{i,j})^2 \geq \rho^{1+2\delta}$, so that the first term above is upper bounded by $-\frac{1}{2}(\frac{k-2}{k-1})\rho^{2\delta}$. Using $\hat{t}_{i,j} \leq \rho^{\frac{1}{2}+\delta}$, the remaining two terms are upper bounded by $\hat{c}_k\rho^{-\frac{1}{2}+3\delta}$ for some positive constant $\hat{c}_k$ that depends only on k. Hence, we have

$$\log \frac{w(A)}{w(U)} \leq -\frac{1}{2}\left(\frac{k-2}{k-1}\right)\rho^{2\delta} + \hat{c}_k\rho^{-\frac{1}{2}+3\delta}.$$

This proves (47) for $A \in \mathcal{B}_{k,M}(\delta)$ with $\max_{i,j} a_{i,j} \leq M + 1 - \rho$.

It remains to consider the case of $A \in \mathcal{B}_{k,M}(\delta)$ with $\max_{i,j} a_{i,j} > M + 1 - \rho$. Now, the reason the approach used above to get to (47) no longer works is that if $a_{i,j} > M + 1 - \rho$, then (50) may not hold. So, what we do instead is to show that for each such A, there exists an $\breve{A} \in \mathcal{B}_{k,M}(\delta)$ with $\max_{i,j} \breve{a}_{i,j} \leq M + 1 - \rho$ such that $w(A) \leq w(\breve{A})$. As argued above, (47) holds for $\breve{A}$; therefore, it holds for A as well.

We now prove the existence of an $\breve{A}$ as required. Let (i,j) be such that $a_{i,j} > M + 1 - \rho$. Then, since $A \in \mathcal{A}_{k,M}$, the following must hold: (i) for all $i' \neq i$ and $j' \neq j$, we have $a_{i',j} < \rho - 1$ and $a_{i,j'} < \rho - 1$, and (ii) there exists some $i' \neq i$ and $j' \neq j$ such that $a_{i',j'} > \rho$. Now, consider the matrix $A^{\pm}$ which has the same entries as A, except for the following: $a_{i,j}^{\pm} = a_{i,j} - 1$, $a_{i',j}^{\pm} = a_{i',j} + 1$, $a_{i,j'}^{\pm} = a_{i,j'} + 1$, and $a_{i',j'}^{\pm} = a_{i',j'} - 1$. Note that $A^{\pm}$ is also in $\mathcal{B}_{k,M}(\delta)$, since $a_{i,j}^{\pm} - u_{i,j} > M - \rho - u_{i,j} \geq M - 2\rho - 1 = (k-2)\rho - 1 \geq \rho - 1 > \rho^{\frac{1}{2}+\delta}$ for $\delta \in (0, \frac{1}{6})$ and $\rho \geq 4$ (as assumed at the beginning of this appendix).

Let $\mathbf{a}_1, \ldots, \mathbf{a}_k$ and $\mathbf{a}_1^{\pm}, \ldots, \mathbf{a}_k^{\pm}$ denote the rows of A and $A^{\pm}$, respectively. Clearly, $\mathbf{a}_\ell = \mathbf{a}_\ell^{\pm}$ for all $\ell \notin \{i, i'\}$. Moreover, it can be directly verified using the definition of the function ϕ in (19) that $\phi(\mathbf{a}_\ell) \leq \phi(\mathbf{a}_\ell^{\pm})$ for $\ell \in \{i, i'\}$. With this, we have

$$w(A) = \prod_{\ell=1}^{k} \phi(\mathbf{a}_\ell) \leq \prod_{\ell=1}^{k} \phi(\mathbf{a}_\ell^{\pm}) = w(A^{\pm}).$$

Note that the procedure of obtaining $A^{\pm}$ from A strictly reduces the (i,j)th entry of A, and does not create any new entries larger than $M + 1 - \rho$. If $A^{\pm}$ still contains an entry larger than $M + 1 - \rho$, we apply the procedure to $A^{\pm}$ to produce a matrix $(A^{\pm})^{\pm}$, and so on. Carrying on in this manner, after finitely many steps, we will obtain the desired matrix $\breve{A}$. □

Lemmas E.1 and E.2 allow us to obtain some useful bounds on the ratio $\frac{\mathbb{E}[\|A-U\|^2]}{\mathbb{E}[\|A-U\|^2]}$.

LEMMA E.3. *There exist positive constants $c_{1,k}$ and $c_{2,k}$ depending only on k, such that for all sufficiently large M, we have*

$$(53) \qquad \exp(-c_{2,k}\,\rho^{-\frac{1}{2}+3\delta}) \le \frac{\sum_{A\in\mathcal{A}_{k,M}} \|A-U\|^2 \frac{w(A)}{w(U)}}{\sum_{A\in\widetilde{\mathcal{A}}_{k,M}} \|A-U\|^2 \widetilde{w}(A)} \le \exp(c_{1,k}\,\rho^{-\frac{1}{2}+3\delta})$$

and

$$(54) \qquad \exp(-c_{2,k}\,\rho^{-\frac{1}{2}+3\delta}) \le \frac{\sum_{A\in\mathcal{A}_{k,M}} \frac{w(A)}{w(U)}}{\sum_{A\in\widetilde{\mathcal{A}}_{k,M}} \widetilde{w}(A)} \le \exp(c_{1,k}\,\rho^{-\frac{1}{2}+3\delta}).$$

Hence, with $c_{3,k} = c_{1,k} + c_{2,k}$, we have

$$(55) \qquad \exp(-c_{3,k}\,\rho^{-\frac{1}{2}+3\delta}) \le \frac{\mathbb{E}[\|A-U\|^2]}{\widetilde{\mathbb{E}}[\|A-U\|^2]} \le \exp(c_{3,k}\,\rho^{-\frac{1}{2}+3\delta}).$$

Proof. Note that (55) follows directly from (53) and (54), since by definition, we have

$$\mathbb{E}[\|A-U\|^2] = \frac{\sum_{A\in\mathcal{A}_{k,M}} \|A-U\|^2 \frac{w(A)}{w(U)}}{\sum_{A\in\mathcal{A}_{k,M}} \frac{w(A)}{w(U)}}$$

and

$$\widetilde{\mathbb{E}}[\|A-U\|^2] = \frac{\sum_{A\in\widetilde{\mathcal{A}}_{k,M}} \|A-U\|^2\,\widetilde{w}(A)}{\sum_{A\in\widetilde{\mathcal{A}}_{k,M}} \widetilde{w}(A)}.$$

We will prove (53) only, as (54) is proved along similar lines. We first write

$$\sum_{A\in\mathcal{A}_{k,M}} \|A-U\|^2 \frac{w(A)}{w(U)} = \sum_{A\in\mathcal{A}_{k,M}(\delta)} \|A-U\|^2 \frac{w(A)}{w(U)} + \sum_{A\in\mathcal{B}_{k,M}(\delta)} \|A-U\|^2 \frac{w(A)}{w(U)},$$

where $\mathcal{B}_{k,M}(\delta)$ is as defined in Lemma E.2. Now, from (46) and (48), we have positive constants $\hat{c}_k$ and c_k'' depending only on k, such that for all sufficiently large M,

$$(56)$$
$$\sum_{A\in\mathcal{A}_{k,M}} \|A-U\|^2 \frac{w(A)}{w(U)} \le \exp(\hat{c}_k\,\rho^{-\frac{1}{2}+3\delta}) \sum_{A\in\mathcal{A}_{k,M}(\delta)} \|A-U\|^2 \widetilde{w}(A) + \exp(-c_k''\rho^{2\delta}).$$

It is easily verified that as M increases, the term $\exp(-c_k''\rho^{2\delta})$ goes to zero fast enough that it can eventually be "absorbed" into the term $\exp(\hat{c}_k\,\rho^{-\frac{1}{2}+3\delta}) \sum_{A\in\mathcal{A}_{k,M}(\delta)} \|A - U\|^2 \widetilde{w}(A)$ by increasing the constant $\hat{c}_k$ slightly. To be precise, for any constant $c_{1,k} > \hat{c}_k$, we have for all sufficiently large M,

$$\sum_{A\in\mathcal{A}_{k,M}} \|A-U\|^2 \frac{w(A)}{w(U)} \le \exp(c_{1,k}\,\rho^{-\frac{1}{2}+3\delta}) \sum_{A\in\mathcal{A}_{k,M}(\delta)} \|A-U\|^2 \widetilde{w}(A)$$
$$\le \exp(c_{1,k}\,\rho^{-\frac{1}{2}+3\delta}) \sum_{A\in\widetilde{\mathcal{A}}_{k,M}} \|A-U\|^2 \widetilde{w}(A),$$

which yields the second inequality in (53).

To prove the first inequality in (53), we begin by noting that

$$\sum_{A \in \mathcal{A}_{k,M}} \|A - U\|^2 \, \frac{w(A)}{w(U)} \geq \sum_{A \in \mathcal{A}_{k,M}(\delta)} \|A - U\|^2 \, \frac{w(A)}{w(U)}$$

(57)
$$\geq \exp(-\hat{c}_k \, \rho^{-\frac{1}{2}+3\delta}) \sum_{A \in \mathcal{A}_{k,M}(\delta)} \|A - U\|^2 \, \widetilde{w}(A)$$

for all sufficiently large M, by (46). Now, as in Lemma E.1, let $\widetilde{\mathcal{B}}_{k,M}(\delta) = \widetilde{\mathcal{A}}_{k,M} \setminus \mathcal{A}_{k,M}(\delta)$, and consider

$$\sum_{A \in \widetilde{\mathcal{A}}_{k,M}} \|A - U\|^2 \, \widetilde{w}(A)$$

$$= \sum_{A \in \mathcal{A}_{k,M}(\delta)} \|A - U\|^2 \, \widetilde{w}(A) + \sum_{A \in \widetilde{\mathcal{B}}_{k,M}(\delta)} \|A - U\|^2 \, \widetilde{w}(A)$$

$$= \sum_{A \in \mathcal{A}_{k,M}(\delta)} \|A - U\|^2 \, \widetilde{w}(A) \; + \; \widetilde{Z}_{k,M} \sum_{A \in \widetilde{\mathcal{B}}_{k,M}(\delta)} \|A - U\|^2 \, \widetilde{Q}_{k,M}(A)$$

$$\leq \sum_{A \in \mathcal{A}_{k,M}(\delta)} \|A - U\|^2 \, \widetilde{w}(A) \; + \; \exp(-c_k''' \rho^{2\delta})$$

for some positive constant c_k''' depending only on k, and for all sufficiently large M. The last inequality above is obtained via Lemma E.1 and Proposition F.1. Thus, we have

$$\sum_{A \in \mathcal{A}_{k,M}(\delta)} \|A - U\|^2 \, \widetilde{w}(A) \geq \sum_{A \in \widetilde{\mathcal{A}}_{k,M}} \|A - U\|^2 \, \widetilde{w}(A) \; - \; \exp(-c_k''' \rho^{2\delta})$$

for all sufficiently large M. Plugging this into (57), and noting as before that $\exp(-c_k''' \rho^{2\delta})$ goes to zero rapidly as M increases, we can readily deduce that for any constant $c_{2,k} > \hat{c}_k$, we have

$$\sum_{A \in \mathcal{A}_{k,M}} \|A - U\|^2 \, \frac{w(A)}{w(U)} \geq \exp(-c_{2,k} \, \rho^{-\frac{1}{2}+3\delta}) \sum_{A \in \widetilde{\mathcal{A}}_{k,M}} \|A - U\|^2 \widetilde{w}(A)$$

for all sufficiently large M. The first inequality in (53) is thus proved. $\qquad\square$

We are now in a position to complete the proof of Lemma 5.5. From (55), it follows that

$$\left| \mathbb{E}[\|A - U\|^2] - \widetilde{\mathbb{E}}[\|A - U\|^2] \right| \; = \; \widetilde{\mathbb{E}}[\|A - U\|^2] \, O\!\left(\rho^{-\frac{1}{2}+3\delta}\right).$$

Since $\widetilde{\mathbb{E}}[\|A - U\|^2] = O(\rho)$ by Lemma 5.6,[7] we conclude that

$$\left| \mathbb{E}[\|A - U\|^2] - \widetilde{\mathbb{E}}[\|A - U\|^2] \right| = O\!\left(\rho^{\frac{1}{2}+3\delta}\right),$$

which proves Lemma 5.5.

Appendix F. Some properties of discrete Gaussian measures. In this appendix, we consider a discrete Gaussian measure defined on $\mathbb{Z}^d$ by $\mu(\mathbf{x}) := \frac{1}{Z} v(\mathbf{x})$,

---

[7]Lemma 5.6 is proved independently of Lemma 5.5.

where $v(\mathbf{x}) = \exp\{-\frac{1}{2\beta}\mathbf{x}'V\mathbf{x}\}$ for $\mathbf{x} \in \mathbb{Z}^d$, $\beta > 0$ and V a symmetric positive definite matrix, and $Z = \sum_{\mathbf{x} \in \mathbb{Z}^d} v(\mathbf{x})$. Let $\mathbf{X}$ be a random variable distributed according to the measure μ. We collect here some results on the measure μ that are used in this paper. These results are valid in the regime where V is fixed and $\beta \to \infty$.

The main tool used in the proofs in this appendix is the Poisson summation formula (see, e.g., [28, Chapter VII, Corollary 2.6] or [29, section 17]). This formula applies to functions $f : \mathbb{R}^d \to \mathbb{C}$, with Fourier transform $\hat{f}$ defined for all $\boldsymbol{\xi} \in \mathbb{R}^d$ as $\hat{f}(\boldsymbol{\xi}) = \int_{\mathbb{R}^d} f(\mathbf{x})e^{i\langle\boldsymbol{\xi},\mathbf{x}\rangle}d\mathbf{x}$, such that

$$|f(\mathbf{x})|, |\hat{f}(\mathbf{x})| \leq \frac{C}{1 + \|\mathbf{x}\|^{d+\delta}} \quad \forall\, \mathbf{x} \in \mathbb{R}^d$$

for some constants $C > 0$ and $\delta > 0$. For such functions f, the Poisson summation formula states that

$$\tag{58} \sum_{\mathbf{x} \in \mathbb{Z}^d} f(\mathbf{x}) = \sum_{\boldsymbol{\xi} \in \mathbb{Z}^d} \hat{f}(2\pi\boldsymbol{\xi}).$$

By a basic fact about the Gaussian density, the function $v(\mathbf{x}) = \exp\{-\frac{1}{2\beta}\mathbf{x}^T V\mathbf{x}\}$ has Fourier transform $\hat{v}(\boldsymbol{\xi}) = \frac{(2\pi)^{d/2}\beta^{1/2}}{\sqrt{\det V}} \exp(-\frac{1}{2}\beta\,\boldsymbol{\xi}'V^{-1}\boldsymbol{\xi})$. Hence, the Poisson summation formula applies, and we have

$$\tag{59} Z = \sum_{\mathbf{x} \in \mathbb{Z}^d} v(\mathbf{x}) = \sum_{\boldsymbol{\xi} \in \mathbb{Z}^d} \hat{v}(2\pi\boldsymbol{\xi}) = \frac{(2\pi)^{d/2}\beta^{1/2}}{\sqrt{\det V}}\, Z^*,$$

where we define $Z^* := \sum_{\boldsymbol{\xi} \in \mathbb{Z}^d} \exp(-\frac{1}{2}4\pi^2\beta\,\boldsymbol{\xi}'V^{-1}\boldsymbol{\xi})$.

Clearly, $Z^* \geq 1$ since the $\boldsymbol{\xi} = \mathbf{0}$ term in the sum evaluates to 1. In fact, as $\beta \to \infty$, $Z^* \to 1$. This is because $\boldsymbol{\xi}'V^{-1}\boldsymbol{\xi}$ is a positive definite quadratic form, so that $\lim_{\beta\to\infty} \exp(-\frac{1}{2}4\pi^2\beta\,\boldsymbol{\xi}'V^{-1}\boldsymbol{\xi})$ equals 0 if $\boldsymbol{\xi} \neq \mathbf{0}$ and equals 1 if $\boldsymbol{\xi} = \mathbf{0}$. Thus, $Z \to \frac{(2\pi)^{d/2}\beta^{1/2}}{\sqrt{\det V}}$. A rigorous statement of this convergence, along with an estimate of the rate of convergence, is given in Proposition F.1 below. To obtain the convergence rate estimate, we make use of some bounds on the quadratic form $\boldsymbol{\xi}'V^{-1}\boldsymbol{\xi}$.

The matrix V^{-1} can be diagonalized as $U'\Lambda^{-1}U$, where U is an orthogonal matrix and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$ is a diagonal matrix composed of the eigenvalues, $\lambda_1, \ldots, \lambda_d$, of V. Since V is symmetric and positive definite, its eigenvalues are all real and positive. Thus, with $\boldsymbol{\eta} = U\boldsymbol{\xi}$, we have $\boldsymbol{\xi}'V^{-1}\boldsymbol{\xi} = \boldsymbol{\eta}'\Lambda^{-1}\boldsymbol{\eta} = \sum_{i=1}^{d} \frac{1}{\lambda_i}\eta_i^2$. Hence, letting $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalues, respectively, of V, we have

$$\tag{60} \frac{1}{\lambda_{\max}}\|\boldsymbol{\xi}\|^2 = \frac{1}{\lambda_{\max}}\|\boldsymbol{\eta}\|^2 \leq \boldsymbol{\xi}'V^{-1}\boldsymbol{\xi} \leq \frac{1}{\lambda_{\min}}\|\boldsymbol{\eta}\|^2 = \frac{1}{\lambda_{\min}}\|\boldsymbol{\xi}\|^2,$$

the equalities on either side being due to the fact that orthogonal transformations preserve ℓ_2 norms.

PROPOSITION F.1. *In the regime where $\beta \to \infty$, we have*

$$Z = \frac{(2\pi)^{d/2}\beta^{1/2}}{\sqrt{\det V}}\left[1 + O\left(\exp\left(-\frac{2\pi^2}{\lambda_{\max}}\beta\right)\right)\right].$$

Proof. Let $C_0 = \frac{2\pi^2}{\lambda_{\max}}$. It is enough to show that $Z^* = 1 + O\big(\exp(-C_0\beta)\big)$. Since $Z^* \geq 1$, we need a corresponding upper bound. This is done using (60) as follows:

$$
\begin{aligned}
Z^* &\leq \sum_{\boldsymbol{\xi} \in \mathbb{Z}^d} \exp\Big(-C_0\,\beta \sum_{i=1}^d \xi_i^2\Big) = \prod_{i=1}^d \sum_{\xi_i \in \mathbb{Z}} \exp\Big(-C_0\beta\,\xi_i^2\Big) \\
&= \Big[\sum_{\xi \in \mathbb{Z}} \exp\Big(-C_0\beta\,\xi^2\Big)\Big]^d \leq \Big[\sum_{\xi \in \mathbb{Z}} \exp\Big(-C_0\beta\,|\xi|\Big)\Big]^d.
\end{aligned}
$$

The upper bound $1 + O\big(\exp(-C_0\beta)\big)$ now follows from the geometric series summation formula. $\qquad \square$

The Poisson summation formula also applies to the function $\vartheta(\mathbf{x}) = \mathbf{x}'\mathrm{V}\mathbf{x}\,v(\mathbf{x})$. Recall that if f has a twice-differentiable Fourier transform, then the function $g(\mathbf{x}) = \|\mathbf{x}\|^2 f(\mathbf{x})$ has Fourier transform $\hat{g}(\boldsymbol{\xi}) = -\Delta\hat{f}(\boldsymbol{\xi})$, where $\Delta\hat{f}$ is the Laplacian of $\hat{f}$. With $f(\mathbf{x}) = \exp(-\frac{1}{2}\|\mathbf{x}\|^2)$, we have $g(\mathbf{x}) = \|\mathbf{x}\|^2 \exp(-\frac{1}{2}\|\mathbf{x}\|^2)$, and $\frac{1}{\beta}\vartheta(\mathbf{x}) = g(\beta^{-1/2}\mathrm{V}^{1/2}\mathbf{x})$, where $\mathrm{V}^{1/2}$ is the symmetric positive definite square root of V. Now, $\hat{f}(\boldsymbol{\xi}) = (2\pi)^{d/2}\exp(-\frac{1}{2}\|\boldsymbol{\xi}\|^2)$, from which straightforward computations yield $\hat{g}(\boldsymbol{\xi}) = (2\pi)^{d/2}(d - \|\boldsymbol{\xi}\|^2)\exp(-\frac{1}{2}\|\boldsymbol{\xi}\|^2)$. Thus, via a change of variable, we have

$$
\begin{aligned}
\frac{1}{\beta}\hat{\vartheta}(\boldsymbol{\xi}) &= \frac{(2\pi)^{d/2}\beta^{1/2}}{\sqrt{\det \mathrm{V}}}(d - \beta\boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi})\exp\left(-\frac{1}{2}\beta\,\boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi}\right) \\
&= \frac{Z}{Z^*}(d - \beta\,\boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi})\exp\left(-\frac{1}{2}\beta\,\boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi}\right),
\end{aligned}
\tag{61}
$$

where we have used the identity in (59).

PROPOSITION F.2. *In the regime where $\beta \to \infty$, we have*

$$
\mathbb{E}_\mu[\mathbf{X}'\mathrm{V}\mathbf{X}] = \frac{1}{Z}\sum_{\mathbf{x} \in \mathbb{Z}^d} \vartheta(\mathbf{x}) = \beta\,d + O\left(\beta^2 \exp\left(-\frac{2\pi^2}{\lambda_{\max}}\beta\right)\right).
$$

Proof. Plugging (61) into the Poisson summation formula (58), we obtain

$$
\begin{aligned}
\mathbb{E}_\mu[\mathbf{X}'\mathrm{V}\mathbf{X}] &= \frac{1}{Z}\sum_{\mathbf{x} \in \mathbb{Z}^d} \vartheta(\mathbf{x}) = \frac{1}{Z}\sum_{\boldsymbol{\xi} \in \mathbb{Z}^d} \hat{\vartheta}(2\pi\boldsymbol{\xi}) \\
&= \frac{\beta}{Z^*}\sum_{\boldsymbol{\xi} \in \mathbb{Z}^d}(d - 4\pi^2\beta\,\boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi})\exp\left(-\frac{1}{2}4\pi^2\beta\,\boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi}\right) \\
&= \beta\,d - \frac{4\pi^2\beta^2}{Z^*}\sum_{\boldsymbol{\xi} \in \mathbb{Z}^d} \boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi}\,\exp\left(-\frac{1}{2}4\pi^2\beta\,\boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi}\right).
\end{aligned}
\tag{62}
$$

It only remains to show that the final summation in (62) is $O\big(\exp(-\frac{2\pi^2}{\lambda_{\max}}\beta)\big)$. Using (60), the summand can be bounded as

$$
\boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi}\,\exp\left(-\frac{1}{2}4\pi^2\beta\,\boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi}\right) \leq \frac{1}{\lambda_{\min}}\|\boldsymbol{\xi}\|^2 \exp\left(\frac{-2\pi^2}{\lambda_{\max}}\beta\|\boldsymbol{\xi}\|^2\right).
$$

We then have

$$\sum_{\boldsymbol{\xi}\in\mathbb{Z}^d} \boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi}\,\exp\left(-\frac{1}{2}4\pi^2\beta\,\boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi}\right)$$

$$\leq \frac{1}{\lambda_{\min}}\sum_{\boldsymbol{\xi}\in\mathbb{Z}^d}\sum_{i=1}^d \xi_i^2\exp\left(\frac{-2\pi^2}{\lambda_{\max}}\beta\sum_{i=1}^d \xi_i^2\right)$$

$$= \frac{1}{\lambda_{\min}}\sum_{i=1}^d\sum_{\boldsymbol{\xi}\in\mathbb{Z}^d} \xi_i^2\exp\left(\frac{-2\pi^2}{\lambda_{\max}}\beta\sum_{i=1}^d \xi_i^2\right)$$

$$= \frac{d}{\lambda_{\min}}\sum_{\boldsymbol{\xi}\in\mathbb{Z}^d} \xi_1^2\exp\left(\frac{-2\pi^2}{\lambda_{\max}}\beta\sum_{i=1}^d \xi_i^2\right)$$

$$= \frac{d}{\lambda_{\min}}\left[\sum_{\xi_1\in\mathbb{Z}} \xi_1^2\exp\left(\frac{-2\pi^2}{\lambda_{\max}}\beta\,\xi_1^2\right)\right]\prod_{i=2}^d\left[\sum_{\xi_i\in\mathbb{Z}}\exp\left(\frac{-2\pi^2}{\lambda_{\max}}\beta\,\xi_i^2\right)\right]$$

$$= \frac{d}{\lambda_{\min}}\left[\sum_{\xi\in\mathbb{Z}} \xi^2\exp\left(\frac{-2\pi^2}{\lambda_{\max}}\beta\,\xi^2\right)\right]\left[\sum_{\xi\in\mathbb{Z}}\exp\left(\frac{-2\pi^2}{\lambda_{\max}}\beta\,\xi^2\right)\right]^{d-1}$$

$$\leq \frac{d}{\lambda_{\min}}\left[\sum_{\xi\in\mathbb{Z}} \xi^2\exp\left(\frac{-2\pi^2}{\lambda_{\max}}\beta\,|\xi|\right)\right]\left[\sum_{\xi\in\mathbb{Z}}\exp\left(\frac{-2\pi^2}{\lambda_{\max}}\beta\,|\xi|\right)\right]^{d-1}$$

$$(63)\qquad = \frac{d}{\lambda_{\min}}O\left(\exp\left(-\frac{2\pi^2}{\lambda_{\max}}\beta\right)\right)\left[1+O\left(\exp\left(-\frac{2\pi^2}{\lambda_{\max}}\beta\right)\right)\right]^{d-1}$$

$$= O\left(\exp\left(-\frac{2\pi^2}{\lambda_{\max}}\beta\right)\right),$$

the equality in (63) being a consequence of standard geometric series summation formulas. □

Our final result estimates the contribution to $\mathbb{E}_\mu[\mathbf{X}'V\mathbf{X}] = \frac{1}{Z}\sum_{\mathbf{x}}\vartheta(\mathbf{x})$ made by vectors $\mathbf{x}\in\mathbb{Z}^d$ with $\mathbf{x}'V\mathbf{x}\geq R$ for some (large) $R>0$. To this end, define $\mathfrak{Z}(R):=\{\mathbf{x}\in\mathbb{Z}^d:\mathbf{x}'V\mathbf{x}\geq R\}$.

PROPOSITION F.3. *For any $R>0$ and $0<\tau<1$, we have*

$$\frac{1}{Z}\sum_{\mathbf{x}\in\mathfrak{Z}(R)}\vartheta(\mathbf{x})\leq \beta d\,\tau^{-(\frac{d}{2}+1)}\exp\left(-\frac{(1-\tau)R}{2\beta}\right).$$

Proof. For $0<\tau<1$, we write

$$\frac{1}{Z}\sum_{\mathbf{x}\in\mathfrak{Z}(R)}\vartheta(\mathbf{x}) = \frac{1}{Z}\sum_{\mathbf{x}\in\mathfrak{Z}(R)}\mathbf{x}'V\mathbf{x}\,\exp\left(-\frac{1-\tau}{2\beta}\mathbf{x}'V\mathbf{x}\right)\exp\left(-\frac{\tau}{2\beta}\mathbf{x}'V\mathbf{x}\right)$$

$$\leq \exp\left(-\frac{1-\tau}{2\beta}R\right)\frac{1}{Z}\sum_{\mathbf{x}\in\mathbb{Z}^d}\mathbf{x}'V\mathbf{x}\,\exp\left(-\frac{\tau}{2\beta}\mathbf{x}'V\mathbf{x}\right)$$

$$= \exp\left(-\frac{1-\tau}{2\beta}R\right)\tau^{-1}\frac{1}{Z}\sum_{\mathbf{x}\in\mathbb{Z}^d}\vartheta(\tau^{\frac{1}{2}}\mathbf{x})$$

$$(64) \qquad = \exp\left(-\frac{1-\tau}{2\beta}R\right)\tau^{-(\frac{d}{2}+1)}\frac{1}{Z}\sum_{\boldsymbol{\xi}\in\mathbb{Z}^d}\hat{\vartheta}(2\pi\tau^{-\frac{1}{2}}\boldsymbol{\xi})$$

$$(65) \qquad \leq \exp\left(-\frac{1-\tau}{2\beta}R\right)\tau^{-(\frac{d}{2}+1)}\frac{\beta}{Z^*}\sum_{\boldsymbol{\xi}\in\mathbb{Z}^d}d\exp(-\frac{1}{2}4\pi^2\tau^{-1}\beta\,\boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi})$$

$$\leq \exp\left(-\frac{1-\tau}{2\beta}R\right)\tau^{-(\frac{d}{2}+1)}\frac{\beta}{Z^*}\sum_{\boldsymbol{\xi}\in\mathbb{Z}^d}d\exp(-\frac{1}{2}4\pi^2\beta\,\boldsymbol{\xi}'\mathrm{V}^{-1}\boldsymbol{\xi})$$

$$= \exp\left(-\frac{1-\tau}{2\beta}R\right)\tau^{-(\frac{d}{2}+1)}\beta d.$$

In (64) above, we used the Poisson summation formula (58), and (65) follows from (61).   ⬜

REFERENCES

[1] A. ORLITSKY, N. P. SANTHANAM, AND J. ZHANG, *Universal compression of memoryless sources over unknown alphabets*, IEEE Trans. Inform. Theory, 50 (2004), pp. 1469–1481.

[2] A. ORLITSKY, N. P. SANTHANAM, K. VISWANATHAN, AND J. ZHANG, *On Estimating the Probability Multiset*, manuscript, https://www-ee.eng.hawaii.edu/~prasadsn/skelnew59.pdf (2011).

[3] A. ORLITSKY, N. P. SANTHANAM, K. VISWANATHAN, AND J. ZHANG, *On modeling profiles instead of values*, in Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence, 2004.

[4] A. ORLITSKY, S. SAJAMA, N. P. SANTHANAM, K. VISWANATHAN, AND J. ZHANG, *Algorithms for modeling distributions over large alphabets*, in Proceedings of ISIT'04, Chicago, 2004, p. 304.

[5] J. ACHARYA, A. ORLITSKY, AND S. PAN, *Recent results on pattern maximum likelihood*, in Proceedings of ITW'09, Volos, Greece, 2009, pp. 251–255.

[6] A. ORLITSKY AND S. PAN, *The maximum likelihood probability of skewed patterns*, in Proceedings of ISIT'09, Seoul, South Korea, 2009, pp. 1130–1134.

[7] A. ORLITSKY AND S. PAN, *The maximum likelihood probability of unique-singleton, ternary, and length-7 patterns*, Proceedings of ISIT'09, Seoul, South Korea, 2009, pp. 1135–1139.

[8] J. ACHARYA, H. DAS, H. MOHIMANI, A. ORLITSKY, AND S. PAN, *Exact calculation of pattern probabilities*, in Proceedings of ISIT'10, Austin, TX, 2010, pp. 1498–1502.

[9] J. ACHARYA, H. DAS, A. ORLITSKY, AND S. PAN, *Algebraic computation of pattern maximum likelihood*, in Proceedings of ISIT'11, Saint Petersburg, Russia, 2011, pp. 400–404.

[10] P. O. VONTOBEL, *The Bethe approximation of the pattern maximum likelihood distribution*, in Proceedings of ISIT'12, Cambridge, MA, 2012, pp. 2012–2016.

[11] W. FERNANDES, *Calculation of Approximate Pattern Maximum Likelihood*, M.E. Project Report, Department of Electrical Communication Engineering, Indian Institute of Science, 2013.

[12] J. S. YEDIDIA, W. T. FREEMAN, AND Y. WEISS, *Constructing free-energy approximations and generalized belief propagation algorithms*, IEEE Trans. Inform. Theory, 51 (2005), pp. 2282–2312.

[13] W. FERNANDES AND N. KASHYAP, *A phase transition for the uniform distribution in the pattern maximum likelihood problem*, in Proceedings of ITW 2013, Seville, Spain, 2013, pp. 318–322.

[14] P. O. VONTOBEL, *The Bethe permanent of a nonnegative matrix*, IEEE Trans. Inform. Theory, 59 (2013), pp. 1866–1901.

[15] H. BETHE, *Statistical theory of superlattices*, Proc. R. Soc. Lond. A, 150 (1935), pp. 552–575.

[16] R. Peierls, *On Ising's model of ferromagnetism*, Math. Proc. Cambridge Philos. Soc., 32 (1936), pp. 477–481.

[17] L. Valiant, *The complexity of computing the permanent*, Theoret. Comput. Sci., 8 (1979), pp. 189–201.

[18] M. Jerrum, A. Sinclair, and E. Vigoda, *A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries*, J. ACM, 51 (2004), pp. 671–697.

[19] M. Huber and J. Law, *Fast approximation of the permanent for very dense problems*, in Proceedings of the 19th Annual ACM-SIAM Symposium on Discrete Algorithms, San Francisco, CA, 2008.

[20] M. Chertkov, L. Kroc, and M. Vergassola, *Belief Propagation and Beyond for Particle Tracking*, arXiv:0806.1199, 2008.

[21] B. Huang and T. Jebara, *Approximating the Permanent with Belief Propagation*, arXiv:0908.1769, 2009.

[22] L. Gurvits and A. Samorodnitsky, *Bounds on the permanent and some applications*, in Proceedings of FOCS'14, Philadelphia, 2014, pp. 90–99.

[23] L. Gurvits, *Unharnessing the Power of Schrijver's Permanental Inequality*, arXiv:1106.2844, 2011.

[24] L. Gurvits, *Unleashing the power of Schrijver's permanental inequality with the help of the Bethe approximation*, in Proc. Electron. Colloq. Comput. Complexity, TR 2011, pp. 11–169.

[25] P. O. Vontobel, *Counting in graph covers: A combinatorial characterization of the Bethe entropy function*, IEEE Trans. Inform. Theory, 59 (2013), pp. 6018–6048.

[26] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York, 1979.

[27] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed., McGraw-Hill, New York, 1976.

[28] E. M. Stein and G. L. Weiss, *Introduction to Fourier Analysis on Euclidean Spaces*, Princeton University Press, Princeton, NJ, 1971.

[29] A. Barvinok, *Math 669: Combinatorics, Geometry and Complexity of Integer Points*, https://www.math.lsa.umich.edu/~barvinok/latticenotes669.pdf.