Phase Transitions in Parameter Rich Optimization Problems

Joachim M. Buhmann* Julien Dumazert† Alexey Gronskiy‡ Wojciech Szpankowski§

Abstract

Most real world combinatorial optimization problems are affected by noise in the input data. Search algorithms to identify "good" solutions with low costs behave like the dynamics of large disordered particle systems, e.g., random networks or spin glasses. Such solutions to noise perturbed optimization problems are characterized by *Gibbs distributions* when the optimization algorithm searches for typical solutions by stochastically minimizing costs. The *free energy* that determines the normalization of the Gibbs distribution balances cost minimization relative to entropy maximization.

The problem to analytically compute the free energy of disordered systems has been known as a notoriously difficult mathematical challenge for at least half a century (Talagrand, 2003). We provide rigorous, matching upper and lower bounds on the free energy for two disordered combinatorial optimization problems of random graph instances, the sparse Minimum Bisection Problem (sMBP) and Lawler's Quadratic Assignment Problem (LQAP). These two problems exhibit phase transitions that are equivalent to the statistical behavior of Derrida's Random Energy Model (REM). Both optimization problems can be characterized as *parameter rich* since individual solutions depend on more parameters than the logarithm of the solution space cardinality would suggest for e.g. a coordinate representation.

1 Noisy Combinatorial Optimization

Combinatorial optimization arises in many real world settings and these problems are often notoriously difficult to solve due to data dependent noise in the parameters. Apart from algorithmic questions — like efficient (stochastic) search for solutions with provable guarantees — more theoretical challenges, such as generalization of solutions and their typicality relate to the analytical computation of various macroscopic properties (Frenk et al., 1985) like the *free energy* and these problems remain largely open. Especially, the *free energy* of the corresponding Gibbs distribution is one of those most important macroscopic parameters that often arises in the context of combinatorial optimization. For example, Vannimenus and Mézard (1984) explored the free energy properties of the traveling salesman problem. In this paper we compute the free energy for two optimization problems – sparse Minimum Bisection (sMBP) and Lawler's Quadratic Assignment (LQAP).

Both, sMBP and LQAP belong to a class of optimization problems that can be formulated as follows: Let n be an integer (e.g., number of vertices in a graph, size of a matrix, number of keys in a digital tree, etc.), and $\mathcal{S}_n$ a set of objects (e.g., a set of vertices, elements of a matrix, keys, etc). The data X denote a set of random variables that enter into the definition of an instance (e.g., weights of edges in a weighted graph). One often is interested in asymptotic behavior of the optimal values $R_{\max}$ or $R_{\min}$ defined as

$$R_{\max} = \max_{c\in\mathcal{C}_n}\Big\{ \sum_{i\in\mathcal{S}_n(c)} w_i(X)\Big\} \ , \tag{1.1}$$

$$R_{\min} = \min_{c\in\mathcal{C}_n}\Big\{ \sum_{i\in\mathcal{S}_n(c)} w_i(X)\Big\} \ , \tag{1.2}$$

where $\mathcal{C}_n$ is a set of all feasible solutions, $\mathcal{S}_n(c)$ is a set of objects from $\mathcal{S}_n$ belonging to the c-th feasible solution (e.g., set of edges belonging to a spanning tree), and $w_i(X)$ is the weight assigned to the i-th object.

In this paper the cost function and the optimization task are defined as follows:

$$R(c,X) = \sum_{i\in\mathcal{S}_n(c)} w_i(X) \text{ and} \tag{1.3}$$
$$c_{\text{opt}}(X) = \arg\min_{c\in\mathcal{C}_n} R(c,X).$$

We also denote the cardinality of the feasible set as m (i.e., $m := |\mathcal{C}_n|$) and the cardinality of $\mathcal{S}_n(c)$ as N for all $c \in \mathcal{C}_n$ (i.e., $N := |\mathcal{S}_n(c)|$). Here, we focus on optimization problems in which $\log m = o(N)$ holds true (see Szpankowski, 1995). We call these optimization problems *parameter rich* since the log cardinality of the solution space scales sub-linearly with the number N of objects that belong to a solution c.

*jbuhmann@inf.ethz.ch; this work was supported by SNF Grant # 200021_138117
†julien.dumazert@gmail.com
‡alexey.gronskiy@inf.ethz.ch; this work was supported by SNF Grant # 200021_138117
§spa@cs.purdue.edu; this work was supported by NSF Center for Science of Information (CSoI) Grant CCF-0939370, and in addition by NSF Grants CCF-1524312, and NIH Grant 1U01CA198941-01.

We study these optimization problems in the maximum entropy framework. Therefore, we consider the Gibbs distribution over all configurations (i.e., feasible solutions c). This distribution is parameterized by $\beta = 1/T$ which is the inverse computational temperature T. More precisely, the Gibbs distribution $p_\beta(c|X)$ of $c \in \mathcal{C}_n$ is defined as

$$p_\beta(c|X) = \frac{1}{Z(\beta, X)} \exp(-\beta R(c, X)) \tag{1.4}$$

with partition function

$$Z(\beta, X) = \sum_{c \in \mathcal{C}} \exp(-\beta R(c, X)) \,. \tag{1.5}$$

It is quite revealing to study optimization problems in the maximum entropy framework through the Gibbs distribution. For high temperature when $\beta \to 0$, this distribution selects all configurations uniformly. On the other hand, when $\beta \to \infty$ the Gibbs distribution concentrates on the set of optimal solutions with costs $R_{\min}$. Intermediate values of β define an appropriate resolution of the solution space such that the fluctuations in the input are not overfitted by the optimization algorithm.

The partition function $Z(\beta, X)$ can also be used to characterize some thermodynamic limits such as entropy and free energy rates defined in (1.6) below (see Talagrand (2003)). In this paper, we focus on the free energy rates for high temperature when $\beta \to 0$. This high computational temperature limit is most interesting when the instances of optimization problems are affected by strong fluctuations that only support estimation of low cost resolution results.

The *free energy* is related to $\mathbb{E}_X[\log Z(\beta, X)]$ while the free energy *rate* is the *normalized* version of the free energy. Usually, one defines the free energy rate as

$$\mathcal{F}(\beta) = \lim_{n \to \infty} \frac{\mathbb{E}_X[\log Z(\beta, X)]}{\log |\mathcal{C}_n|}. \tag{1.6}$$

However, such a limit may not exist or it may be trivial. The latter refers to the case where either $\log m = \log |\mathcal{C}_n|$ or $N = |\mathcal{S}_n|$ dominates, that is, $\log m \neq \Theta(N)$. In Vannimenus and Mézard (1984) the case $\log m \gg O(N)$ was analyzed, while here we focus on a class of optimization problems with $\log m = o(N)$ or $\log m \ll N$ (e.g., the quadratic assignment problem Frenk et al. (1985); Szpankowski (1995) in which $N = n^2$ and $m = n!$). For this class of optimization problems, Szpankowski (1995) proved that any solution is asymptotically optimal with high probability, and we provide here another interpretation of these results.

It is known that discontinuities of the free energy indicate abrupt changes in the accessibility of solutions and they are closely related to the complexity of the problems (Auffinger and Chen, 2014). We also note that such abrupt changes of macroscopic properties, also known as *phase transitions*, are characteristic features of various large systems (e.g. Łuczak, 1994) and have been generating uninterrupted interest for a long time.

In a parallel field of statistical physics, an interest in large disordered particle systems also aims at finding laws for behavior of macroscopic thermodynamic properties – e.g., the free energy. Many interesting models of such large systems were introduced relatively early, e.g. the *Sherrington-Kirkpatrick (SK) spin glass model* (see Sherrington and Kirkpatrick, 1975). It took, however, some time and effort to develop rigorous techniques for solving them. For example, Derrida (1981) introduced a very simple, but exactly solvable model called the *random energy model (REM)* as the limit of the SK p-spin models family. Later, Aizenman et al. (1987) published an exact solution in the high-temperature phase for the SK model. The question of the exact free energy behavior triggered a new wave of recent research (Bovier et al., 2002; Talagrand, 2003). During the last half century a diverse set of heuristic tools originating in the context of statistical physics was developed, such as the replica method (Parisi, 2009), the cavity method (Mézard and Parisi, 2003) and mean field approximation schemes with belief propagation algorithms.

The main contribution of the paper consists in a mathematically rigorous asymptotic analysis of the free energy for the Sparse Minimum Bisection Problem (sMBP) and for the Lawler Quadratic Assignment problem (LQAP) in the random graph setting. Both problems have small, but yet non-vanishing correlations, between the cost levels of solutions, which render bounding a difficult mathematical problem (Bovier et al., 2002). We discover phase transitions for sMBP and LQAP, which are equivalent to the discontinuities of REM and high-temperature SK (Derrida, 1981; Aizenman et al., 1987). Our results are expected (see Auffinger and Chen, 2014) to foster understanding some fundamental algorithmic complexity properties of these and other optimization problems.

The rest of this paper is organized as follows: In Section 2, we describe our optimization problems in some details. Then we state our main results, namely Theorem 2.2 and Theorem 2.3. Proofs of the main results, constituting the main contribution of the paper, can be found in Sections 3 and 4 (see also (Buhmann et al., 2014)). Details of the proofs can be found in a forthcoming journal version of this research.

2 Main Results

In this paper we focus on the Sparse Minimum Bisection Problem and Lawler's Quadratic Assignment Problem that are formally defined below. Although these two combinatorial optimization problems are specific, many of our results hold for a large class of optimization problems provided that $\log m = o(N)$. In the rest of the paper we will utilize the temperature rescaling $\beta = \widehat{\beta}\sqrt{\log m/N}$ that was theoretically justified in (Buhmann et al., 2014).

Sparse minimum bisection problem (sparse MBP). Consider a complete, undirected, weighted graph $G = (V, E, X)$ of n vertices, where n is an even number. The input data instance X is represented by (random) weights $(W_i)_{i \in E}$ of the graph edges.

A *sparse bisection* is a pair of disjoint subsets of the vertices $c = (U_1, U_2)$: $U_1 \sqcup U_2 \subsetneq V$, whose size is $|U_1| = |U_2| \equiv d$. We shall call the problem sparse if d grows faster than $\log n$ (which we denote as $\log n \ll d$) and slower than $n^{2/7}/\sqrt{\log n}$ (which we write $d \ll n^{2/7}/\sqrt{\log n}$). Now $\mathcal{S}_n = E$ and $\mathcal{C}_n$ is the set of all sparse bisections of graph G, while $\mathcal{S}_n(c)$ is the set of all edges connecting U_1 and U_2. The cost of a bisection c is the sum of the weights of all cut edges $R(c) = \sum_{i \in \mathcal{S}_n(c)} W_i$.

Thus, $|\mathcal{C}_n| = m = \binom{n}{d}\binom{n-d}{d}$, $N = d^2$ and for sparse MBP, the following holds true (we omit here 1/2 constant for the sake of brevity):

$$\log m = \log \binom{n}{d}\binom{n-d}{d} = \log \frac{n!}{(d!)^2(n-2d)!}$$
$$\sim 2d \log n = o(N). \tag{2.7}$$

In summary, the problem falls into the class $\log m = o(N)$ provided $\log n \ll d$, as we assume.

Lawler quadratic assignment problem (Lawler QAP). Lawler (1963) introduced a generalization of the classical quadratic assignment problem (see Beckman and Koopmans, 1957), where the distance and weight matrices are replaced by one large array. Namely, the input data instance X is represented by a 4-dimensional $n \times n \times n \times n$-matrix Q with i.i.d. values. The solution space $\mathcal{C}_n$ is the set of the n-element permutations $\mathbf{S}_n$.

The cost function is $R(\pi, Q) = \sum_{i,j=1}^{n} Q_{i,j,\pi(i),\pi(j)}$ for $\pi \in \mathbf{S}_n$. In our notation, $N = |\mathcal{S}_n(\pi)| = n^2$ and $m = |\mathcal{C}_n| = n!$, and thus

$$\log m = \log n! \sim n \log n = o(N) \tag{2.8}$$

is fulfilled, i.e. the problem falls into the class $\log m = o(N)$.

We are now in the position to present our main results. We start with spelling out common assumptions.

Common Theorem Setting (CTS). *Consider a class of combinatorial optimization problems in which the cardinality of feasible solutions set m and the size N of a feasible solution are related as $\log m = o(N)$. Assume that weights W_i are identically (not necessarily independently) distributed with mean μ and variance σ^2 and that the moment generating function of negative centered weights $(-\overline{W}_i)$ is finite, i.e. $\overline{G}(t) \equiv \mathbb{E}[\exp(-t\overline{W}_i)] < \infty$ exists for some $t > 0$. Further assume that within a given solution, the weights are mutually independent, i.e.*

$$\forall c \in \mathcal{C}_n, \text{ the set } \{W_i \mid i \in \mathcal{S}_n(c)\} \tag{2.9}$$

is a set of mutually independent variables. Define a scaling $\beta = \widehat{\beta}\sqrt{\log m/N}$, where $\widehat{\beta}$ is a constant.

In order to present a full picture, we first cite a fairly tight upper bound on the free energy (for the proof, see Buhmann et al., 2014, Theorem 1).

Theorem 2.1. (Buhmann et al., 2014) *Under the Common Theorem Setting, the following upper bound holds:*

$$\lim_{n\to\infty} \frac{\mathbb{E}[\log Z] + \widehat{\beta}\mu\sqrt{N \log m}}{\log m} \le \begin{cases} 1 + \frac{\widehat{\beta}^2\sigma^2}{2}, & \widehat{\beta} < \frac{\sqrt{2}}{\sigma}, \\ \widehat{\beta}\sigma\sqrt{2}, & \widehat{\beta} \ge \frac{\sqrt{2}}{\sigma} \end{cases} \tag{2.10}$$

provided $\log m = o(N)$.

Interestingly, the above theorem indicates that there exists a phase transition of the upper bound of the free energy (compare to Talagrand, 2003; Mézard and Montanari, 2009). We must stress, however, that the above bound is not tight in general. Consider the (non-sparse) MBP with $d = n/2$ vertices. Under the same general assumptions for the weights, it can be shown that a tighter bound holds for $\widehat{\beta} \le \frac{1}{\sqrt{\log 2\sigma}}$ (see our full paper Buhmann et al. (2017))

$$\lim_{n\to\infty} \frac{\mathbb{E}[\log Z] + \widehat{\beta}\mu\sqrt{N \log m}}{\log m} \le 1 + \frac{\widehat{\beta}^2\sigma^2}{4}, \tag{2.11}$$

which shows that the bound in Theorem 2.1 is not tight.

We proceed now to present our results. For some combinatorial optimization problems, the asymptotical upper bound of Theorem 2.1 turns out to be tight. Below are two main results that give the matching lower bound for the sparse MBP and Lawler QAP. We should point out that for the sparse MBP we developed a novel approach to prove the matching lower bound, since the techniques proposed in (Talagrand, 2003, Chapter 1) seem not to work in our setting.

Theorem 2.2. *Consider sparse MBP (complying with the Common Theorem Setting), whose edge weights have mean μ and variance σ^2. Then the following holds:*

$$\lim_{n\to\infty} \frac{\mathbb{E}[\log Z] + \widehat{\beta}\mu\sqrt{N \log m}}{\log m} = \begin{cases} 1 + \frac{\widehat{\beta}^2\sigma^2}{2}, & \widehat{\beta} < \frac{\sqrt{2}}{\sigma}, \\ \widehat{\beta}\sigma\sqrt{2}, & \widehat{\beta} \ge \frac{\sqrt{2}}{\sigma} \end{cases} \tag{2.12}$$

provided $\log n \ll d \ll n^{2/7}/\sqrt{\log n}$.

For the Lawler QAP, we adapt the proof technique proposed by Talagrand (2003) and present a sketch of proof in Section 4.

Theorem 2.3. *Consider Lawler QAP (complying with the Common Theorem Setting), whose matrix entries have mean μ and variance σ^2. Then the following holds:*

$$\lim_{n\to\infty} \frac{\mathbb{E}[\log Z] + \widehat{\beta}\mu\sqrt{N\log m}}{\log m} = \begin{cases} 1 + \frac{\widehat{\beta}^2\sigma^2}{2}, & \widehat{\beta} < \frac{\sqrt{2}}{\sigma}, \\ \widehat{\beta}\sigma\sqrt{2}, & \widehat{\beta} \geq \frac{\sqrt{2}}{\sigma}. \end{cases} \tag{2.13}$$

that matches the upper bound of Theorem 2.1.

The reader should notice that the free energy of sMBP (2.12) and of LQAP (2.13) exhibit a phase transition that is asymptotically equivalent to that of Derrida's *Random Energy Model (REM)* (Derrida, 1981, Sect. V). We like to emphasize that both sMBP and LQAP *introduce weak correlation* between costs of pairs of solutions, while REM defines a mathematically less challenging setting without correlations between costs.

3 Proof of Theorem 2.2

In this section we present the proof of the matching lower bound for Sparse MBP. The proof technique that we propose here is novel to the best of our knowledge, however, see (Magner et al., 2016) for the first use of it.

The proof is broken into several lemmas. Let us start with defining a r.v. D as an elementwise overlap (i.e. number of shared *edges*) between two solutions, where the solutions are sampled *uniformly at random*. We will refer to this uniform distribution of solution pairs as $\mathcal{D}$. A minor difficulty with identifying non-overlapping solutions consists in differentiating it between *edge-non-overlapping* and *vertex-non-overlapping* pairs of solutions.

Lemma 3.1. *The following holds*

$$\frac{\#\{\text{vertex-non-ovrlp}\}}{m^2} = 1 - \Theta(d^2/n). \tag{3.14}$$

Proof. Observe that

$$\begin{aligned} \frac{\#\{\text{vertex-non-ovrlp}\}}{m^2} &= \frac{\binom{n}{d}\binom{n-d}{d}\binom{n-2d}{d}\binom{n-3d}{d}}{\binom{n}{d}^2\binom{n-d}{d}^2} \\ &= \frac{\binom{n-2d}{d}\binom{n-3d}{d}}{\binom{n}{d}\binom{n-d}{d}}. \end{aligned} \tag{3.15}$$

We now do the expansion via Stirling's approximation, for any integer $0 \leq \nu < n/d$:

$$\begin{aligned} \binom{n-\nu d}{d} &\leq \frac{(n-\nu d)^d}{d!} \\ &= \frac{n^d(1-\nu d/n)^d}{d!} \sim \frac{n^d(1-\nu d^2/n)}{d!}. \end{aligned} \tag{3.16}$$

Similarly,

$$\begin{aligned} \binom{n-\nu d}{d} &\geq \frac{(n-(\nu+1)d)^d}{d!} \\ &= \frac{n^d(1-(\nu+1)d/n)^d}{d!} \sim \frac{n^d(1-(\nu+1)d^2/n)}{d!}. \end{aligned} \tag{3.17}$$

Applying these bounds we find

$$\begin{aligned} \frac{\#\{\text{vertex-non-ovrlp}\}}{m^2} &\leq \frac{(1-2d^2/n)(1-3d^2/n)}{(1-d^2/n)(1-2d^2/n)} \\ &\sim 1 - 2d^2/n \end{aligned} \tag{3.18}$$

and

$$\begin{aligned} \frac{\#\{\text{vertex-non-ovrlp}\}}{m^2} &\geq \frac{(1-3d^2/n)(1-4d^2/n)}{(1-d^2/n)} \\ &\sim 1 - 6d^2/n. \end{aligned} \tag{3.19}$$

This completes the proof. □

Lemma 3.2. *The following holds:*

$$\mathbb{P}_{\mathcal{D}}(D = 0) \sim \frac{\#\{\textit{vertex-non-ovrlp}\}}{m^2} \tag{3.20}$$

Proof.

$$\begin{aligned} \mathbb{P}_{\mathcal{D}}(D = 0) = \frac{1}{m^2}\Big(&\#\{\text{vertex-non-ovrlp}\} \\ &+ \#\{\text{edge-non-ovrlp} \mid \text{vertex-ovrlp}\}\Big). \end{aligned} \tag{3.21}$$

Since the following set inclusion holds:

$$\{\text{edge-non-ovrlp}|\text{vertex-ovrlp}\} \subseteq \{\text{vertex-ovrlp}\},$$

we can conclude that

$$\begin{aligned} &\frac{\#\{\text{edge-non-ovrlp} \mid \text{vertex-ovrlp}\}}{m^2} \\ &\leq \frac{\{\text{vertex-ovrlp}\}}{m^2} \\ &= \frac{m^2 - \{\text{vertex-non-ovrlp}\}}{m^2} \\ &= 1 - 1 + \Theta(d^2/n) = o(1) \end{aligned} \tag{3.22}$$

where the last equation comes from Lemma 3.1 and the convergence follows from the initial sparsity assumption $\log n \ll d \ll n^{2/7}/\sqrt{\log n}$. Hence, we arrive at

$$\frac{\mathbb{P}_{\mathcal{D}}(D=0)}{\#\{\text{vertex-non-ovrlp}\}/m^2}$$
$$= 1 + \frac{\#\{\text{edge-non-ovrlp} \mid \text{vertex-ovrlp}\}/m^2}{\#\{\text{vertex-non-ovrlp}\}/m^2}$$
$$= 1 + \frac{o(1)}{1+o(1)} = 1 + o(1), \tag{3.23}$$

which proves the lemma. □

These two lemmas allow us to estimate the expected value of D, as follows.

Lemma 3.3. *We have*

$$\mathbb{E}_{\mathcal{D}} D = \mathcal{O}(d^4/n). \tag{3.24}$$

Proof. To compute $\mathbb{E}_{\mathcal{D}}D$, we observe

$$\begin{aligned}\mathbb{E}_{\mathcal{D}} D &= 0 \cdot \mathbb{P}_{\mathcal{D}}(D=0) + \sum_{k=1}^{N} k \cdot \mathbb{P}_{\mathcal{D}}(D=k)\\ &\leq N \sum_{k=1}^{N} \mathbb{P}_{\mathcal{D}}(D=k)\\ &= d^2 \cdot \mathbb{P}_{\mathcal{D}}(D \neq 0) = d^2\big(1 - \mathbb{P}_{\mathcal{D}}(D=0)\big)\\ &\sim \Theta(d^4/n),\end{aligned} \tag{3.25}$$

where the last asymptotic equivalence follows from Lemmas 3.1 and 3.2. The less-than-equal sign turns Θ into $\mathcal{O}$. The lemma is proven. □

Now we are in the position to prove Theorem 2.2.

Proof. Let us now introduce an event A for some ϵ we choose later:

$$A := \{Z \geq \epsilon \mathbb{E}Z\}. \tag{3.26}$$

This implies, by Chebychev inequality,

$$\begin{aligned}1 - \mathbb{P}(A) &\leq \mathbb{P}\big(|Z - \mathbb{E}Z| \geq (1-\epsilon)\mathbb{E}Z\big)\\ &\leq \frac{\mathrm{Var}Z}{(1-\epsilon)^2(\mathbb{E}Z)^2}.\end{aligned} \tag{3.27}$$

From (Buhmann et al., 2014, Lemma 1), we have the following equation:

$$\mathrm{Var}Z = (\mathbb{E}Z)^2\Big(\mathbb{E}_{\mathcal{D}}\Big(\frac{G(2\beta)}{G^2(\beta)}\Big)^D - 1\Big), \tag{3.28}$$

which, in turn, yields asymptotically (see Buhmann et al. (2014)):

$$\mathrm{Var}Z \sim (\mathbb{E}Z)^2\big(\sigma^2\beta^2\mathbb{E}_{\mathcal{D}}D\big). \tag{3.29}$$

Thus the term (3.27) can be further rewritten:

$$\begin{aligned}1 - \mathbb{P}(A) &\leq \frac{\mathrm{Var}Z}{(1-\epsilon)^2(\mathbb{E}Z)^2} \sim \frac{\sigma^2\beta^2\mathbb{E}_{\mathcal{D}}D}{(1-\epsilon)^2}\\ &= \mathcal{O}\Big(\frac{\beta^2\mathbb{E}_{\mathcal{D}}D}{(1-\epsilon)^2}\Big)\\ &= \mathcal{O}\Big(\frac{d^4 \log m}{n(1-\epsilon)^2 N}\Big)\\ &= \mathcal{O}\Big(\frac{d^3 \log n}{n(1-\epsilon)^2}\Big),\end{aligned} \tag{3.30}$$

where we used above Lemma 3.3 for $\mathbb{E}_{\mathcal{D}}D$ asymptotics. We now proceed to compute $\mathbb{E}\log Z$ along the way of (Magner et al., 2016):

$$\begin{aligned}\mathbb{E}\log Z &= \mathbb{E}[\log Z \mid A] \cdot \mathbb{P}(A) + \mathbb{E}[\log Z \mathbb{1}(\bar{A})]\\ &\geq (\log \mathbb{E}Z + \log \epsilon)\mathbb{P}(A) + \mathbb{E}[\log Z \mathbb{1}(\bar{A})].\end{aligned} \tag{3.31}$$

Now remember from (Buhmann et al., 2014, Eq. (12)) that $\log \mathbb{E}Z$ can be written as

$$\log \mathbb{E}Z = -\beta N\mu + \log m + \frac{1}{2}N\beta^2\sigma^2 + o(\beta^2). \tag{3.32}$$

Let the above expression be denoted as $L(\beta, N, m, \sigma)$ for the sake of brevity. So, using (3.30), we rewrite (3.31):

$$\begin{aligned}\mathbb{E}\log Z &\geq \big(L(\beta,N,m,\sigma) + \log\epsilon\big) \cdot \Big(1 - \mathcal{O}\Big(\frac{d^3\log n}{n(1-\epsilon)^2}\Big)\Big)\\ &\qquad + \mathbb{E}[\log Z \mathbb{1}(\bar{A})]\\ &= L(\beta,N,m,\sigma) + \log\epsilon\\ &\qquad - \big(L(\beta,N,m,\sigma) + \log\epsilon\big) \cdot \mathcal{O}\Big(\frac{d^3\log n}{n(1-\epsilon)^2}\Big)\\ &\qquad + \mathbb{E}[\log Z \mathbb{1}(\bar{A})].\end{aligned} \tag{3.33}$$

Thus,

$$\begin{aligned}\frac{\mathbb{E}\log Z + \beta N\mu}{\log m} &\geq 1 + \frac{\widehat{\beta}^2\sigma^2}{2} + \frac{\log\epsilon}{\log m}\\ &\qquad - \big(L(\beta,N,m,\sigma) + \log\epsilon\big) \cdot \mathcal{O}\Big(\frac{d^3\log n}{n\log m(1-\epsilon)^2}\Big)\\ &\qquad + \frac{\mathbb{E}[\log Z\mathbb{1}(\bar{A})]}{\log m}.\end{aligned} \tag{3.34}$$

Now we introduce the below assumption (3.35), which will prove to be true later:

$$\frac{d^3\log n}{n(1-\epsilon)^2} \to 0 \quad (n \to \infty). \tag{3.35}$$

With this assumption, we notice

$$\big(L(\beta,N,m,\sigma) + \log\epsilon\big) \cdot \mathcal{O}\Big(\frac{d^3\log n}{n(1-\epsilon)^2}\Big) = o(1). \tag{3.36}$$

Thus

$$\frac{\mathbb{E}\log Z+\beta N\mu}{\log m}\gtrsim 1+\frac{\widehat{\beta}^2\sigma^2}{2}$$
$$+\frac{\log\epsilon}{\log m}+\frac{\mathbb{E}[\log Z\mathbb{1}(\bar{A})]}{\log m}. \tag{3.37}$$

We now estimate the term $\mathbb{E}[\log Z\mathbb{1}(\bar{A})]$. For some solution c, it holds true:

$$\begin{aligned}\mathbb{E}[\log Z\mathbb{1}(\bar{A})] &\geq \mathbb{E}[\log e^{-\beta R(c)}\mathbb{1}(\bar{A})]\\ &= \mathbb{E}[-\beta R(c)\cdot\mathbb{1}(\bar{A})]\\ &= \mathbb{E}[-\beta(\bar{R}(c)+\mathbb{E}R)\cdot\mathbb{1}(\bar{A})]\\ &= \mathbb{E}[-\beta\bar{R}(c)\mathbb{1}(\bar{A})]-\beta\mathbb{E}R\cdot\mathbb{P}(\bar{A})\\ &\geq -\beta\mathbb{E}[|\bar{R}(c)|]-\beta\mathcal{O}(N)(1-\mathbb{P}(A))\\ &\geq -\beta\mathcal{O}(\sqrt{N})-\beta\mathcal{O}\left(N\frac{d^3\log n}{n(1-\epsilon)^2}\right),\end{aligned} \tag{3.38}$$

thus, essentially,

$$\frac{\mathbb{E}[\log Z\mathbb{1}(\bar{A})]}{\log m}\geq-\beta\mathcal{O}\left(N\frac{d^3\log n}{n\log m(1-\epsilon)^2}\right)$$
$$\sim-\mathcal{O}\left(\frac{d^{7/2}\sqrt{\log n}}{n(1-\epsilon)^2}\right). \tag{3.39}$$

Consequently, we can rewrite (3.37) as

$$\frac{\mathbb{E}\log Z+\beta N\mu}{\log m}\gtrsim 1+\frac{\widehat{\beta}^2\sigma^2}{2}$$
$$+\frac{\log\epsilon}{\log m}-\mathcal{O}\left(\frac{d^{7/2}\sqrt{\log n}}{n(1-\epsilon)^2}\right). \tag{3.40}$$

This suggests that d should be $d=o(n^{2/7}/\sqrt{\log n})$ to make the error term negligible.

We will now choose ϵ in order to produce the lower bounds, and then check that assumption (3.35) is fulfilled. For $\widehat{\beta}>\widehat{\beta}^*:=\sqrt{2}/\sigma$ we choose

$$\epsilon=m^{-(1-\widehat{\beta}\sigma\sqrt{2}+\frac{\widehat{\beta}^2\sigma^2}{2})},$$

which gives

$$\frac{\mathbb{E}\log Z+\beta N\mu}{\log m}\gtrsim\widehat{\beta}\sigma\sqrt{2}-o(1). \tag{3.41}$$

For this choice $\epsilon=o(1)$, and

$$\mathcal{O}\left(\frac{d^{7/2}\sqrt{\log n}}{n(1-\epsilon)^2}\right)=o(1). \tag{3.42}$$

so that (3.35) holds.

For $\widehat{\beta}\leq\widehat{\beta}^*:=\sqrt{2}/\sigma$ we choose $\epsilon=1/2$, yielding

$$\frac{\mathbb{E}\log Z+\beta N\mu}{\log m}\gtrsim 1+\frac{\widehat{\beta}^2\sigma^2}{2}+o(1), \tag{3.43}$$

since

$$\frac{\log\epsilon}{\log m}=o(1),\quad\mathcal{O}\left(\frac{d^{7/2}\sqrt{\log n}}{n(1-\epsilon)^2}\right)=o(1) \tag{3.44}$$

and assumption (3.35) holds. This completes the proof of Theorem 2.2. □

4 Proof of Theorem 2.3

In this section we present a sketch of the proof for the matching lower bound for the Lawler QAP.

Proof. Theorem 2.1 gives us a general upper bound. To find the matching lower bound we follow Talagrand (2003) that we briefly review. Let Y be the cardinality of the solution subset for which the centered negative cost function $\bar{R}(c)$ is large enough, that is,

$$Y:=\mathrm{card}\{c\colon\bar{R}(c)\geq u_n\}\quad\text{for some}\quad u_n\geq 0.$$

It is obvious that

$$\mathbb{E}[Y]=ma,\quad\text{where}\quad a:=\mathbb{P}(\bar{R}(c)\geq u_n). \tag{4.45}$$

Let now A define the event: $\{Y\leq ma/2\}$. By Markov inequality we have

$$\begin{aligned}\mathbb{P}(A)&\leq\mathbb{P}\big((Y-\mathbb{E}[Y])^2\geq m^2a^2/4\big)\\ &\leq\frac{4\mathrm{Var}[Y]}{m^2a^2}\leq\frac{4\mathbb{E}[Y^2]}{m^2a^2}-1,\end{aligned} \tag{4.46}$$

and our goal is to prove that $\mathbb{E}[Y^2]/(ma)^2\to 1$.

In our setting we define $a_n=\mathbb{P}(\bar{R}(c)\geq u_n(\widehat{\beta}))$ where

$$u_n(\widehat{\beta})=\begin{cases}\widehat{\beta}\sigma^2\sqrt{N\log m}, & \widehat{\beta}<\widehat{\beta}^*\\ \widehat{\beta}^*\sigma^2\sqrt{N\log m}, & \widehat{\beta}\geq\widehat{\beta}^*.\end{cases} \tag{4.47}$$

We set as above $A=\{Y\leq ma_n/2\}$, and prove that $\mathbb{P}(A)\to 0$ for $\widehat{\beta}^*<\sqrt{2}/\sigma$. Then by Talagrand's approach this will lead to a proof of Theorem 2.3. Details of the proof can be found in our journal version (Buhmann et al., 2017), see also (Buhmann et al., 2014).

So from now on we focus on the proof of $\mathbb{P}(A)\to 0$ or equivalently that $\mathbb{E}[Y^2]/(ma_n)^2\to 1$. First, note that if solutions (permutations) π and π' have k common points, then they share k^2 entries of the 4-dimensional matrix Q (out of the $N=n^2$ entries appearing in $R(\pi,X)$). Besides, since the solution space $\mathcal{C}_n$ (the set of permutations of size n) is a group, there exists a permutation π'' such that $\pi'=\pi\circ\pi''$. Thus, counting the common points between π and π' is equivalent to counting the fixed points of π'', which is a well-studied problem.

The number of permutations with k fixed points is the rencontre number (see e.g., Szpankowski (2001))

$$D_{n,k}=\frac{n!}{k!}\sum_{j=0}^{n-k}\frac{(-1)^j}{j!}. \tag{4.48}$$

Therefore, the number of ordered pairs of permutations sharing k fixed points is

$$B_{n,k} = n! D_{n,k} = \frac{n!^2}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}. \tag{4.49}$$

We now evaluate $\mathbb{E}[Y^2]$ defined above. We have

$$\begin{aligned}
\mathbb{E}[Y^2] &= \sum_{\pi,\pi' \in C_n} \mathbb{P}\big(\bar{R}(\pi, X) \geq u_n(\widehat{\beta}) \\
&\qquad \text{and } \bar{R}(\pi', X) \geq u_n(\widehat{\beta})\big) \\
&= \sum_{k=0}^{N} B_{n,k} \mathbb{P}\big(O_{n,k} + I_{n,k} \geq u_n(\widehat{\beta}) \\
&\qquad \text{and } O'_{n,k} + I_{n,k} \geq u_n(\widehat{\beta})\big),
\end{aligned} \tag{4.50}$$

where $O_{n,k}, O'_{n,k} \sim \mathcal{N}(0, N-k^2)$, $I_{n,k} \sim \mathcal{N}(0,k^2)$ are independent, $I_{n,k}$ represents the sum of the entries shared by the two solutions, and $O_{n,k}, O'_{n,k}$ the entries exclusive to one of the two. Above $\mathcal{N}(0,\sigma^2)$ means the normal distribution with mean 0 and variance σ^2.

Let us now bound the probability

$$\begin{aligned}
p_{n,k}(\widehat{\beta}) &= \mathbb{P}\big(O_{n,k} + I_{n,k} \geq u_n(\widehat{\beta}) \\
&\qquad \text{and } O'_{n,k} + I_{n,k} \geq u_n(\widehat{\beta})\big)
\end{aligned} \tag{4.51}$$

that two solutions with k^2 shared entries exceed the threshold $u_n(\widehat{\beta})$. This is exactly the probability that the two coordinates of a multivariate centered normal vector with covariance matrix $\left(\begin{smallmatrix} n^2 & k^2 \\ k^2 & n^2 \end{smallmatrix}\right)\sigma^2$ exceed $u_n(\widehat{\beta})$. Applying the results of Savage (1962) on multivariate Gaussian bounds, we get for $k < n$:

$$\begin{aligned}
p_{n,k} &\leq \frac{\sigma^2}{2\pi u_n(\widehat{\beta})^2} \sqrt{\frac{(n^2+k^2)^3}{n^2-k^2}} \\
&\qquad \times \exp\left(-\frac{u_n(\widehat{\beta})^2}{(n^2+k^2)\sigma^2}\right) \\
&= \frac{1}{2\pi\widehat{\beta}^2\sigma^2} \frac{1}{n^2 \log n!} \sqrt{\frac{(n^2+k^2)^3}{n^2-k^2}} \\
&\qquad \times \exp\left(-\widehat{\beta}^2\sigma^2 \frac{n^2 \log n!}{(n^2+k^2)\sigma^2}\right).
\end{aligned} \tag{4.52}$$

For $k = n$, we have $p_{n,n} = a_n$ and we know that

$$\begin{aligned}
a_n &\sim \frac{n\sigma}{\sqrt{2\pi} u_n(\widehat{\beta})} \exp\left(-\frac{u_n(\widehat{\beta})^2}{2n^2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi \log n!}\widehat{\beta}\sigma} \exp\left(-\frac{\widehat{\beta}^2\sigma^2}{2} \log n!\right).
\end{aligned} \tag{4.53}$$

Combining equations (4.49), (4.50), (4.52) and (4.53) yields

$$\begin{aligned}
\frac{\mathbb{E}[Y^2]}{m^2 a_n^2} \lesssim S_n &= \sum_{k=0}^{n-1} \frac{1}{k!} \left(\sum_{j=0}^{n-k} \frac{(-1)^j}{j!}\right) \frac{1}{n^2} \sqrt{\frac{(n^2+k^2)^3}{n^2-k^2}} \\
&\qquad \times \exp\left(\left(1 - \frac{n^2}{n^2+k^2}\right)\widehat{\beta}^2\sigma^2 \log n!\right) \\
&\qquad + \frac{\sqrt{2\pi \log n!}\widehat{\beta}\sigma}{n!} \exp\left(\frac{\widehat{\beta}^2\sigma^2}{2} \log n!\right).
\end{aligned} \tag{4.54}$$

It is obvious that the term outside of the sum will tend to 0 as long as $\widehat{\beta} < \sqrt{2}/\sigma$. Let us now address the asymptotics of

$$\begin{aligned}
S_n &= \sum_{k=0}^{n-1} \frac{1}{k!} \left(\sum_{j=0}^{n-k} \frac{(-1)^j}{j!}\right) \frac{1}{n^2} \sqrt{\frac{(n^2+k^2)^3}{n^2-k^2}} \\
&\qquad \times \exp\left(\left(1 - \frac{n^2}{n^2+k^2}\right)\widehat{\beta}^2\sigma^2 \log n!\right).
\end{aligned} \tag{4.55}$$

For that, set $k = o\left(\frac{n}{\log n}\right)$ and consider the following approximations of the above terms. First,

$$\sum_{j=0}^{n-k} \frac{(-1)^j}{j!} = \frac{1}{\mathrm{e}} + \mathcal{O}\left(\frac{1}{(n-k)!}\right). \tag{4.56}$$

Second,

$$\frac{1}{n^2} \sqrt{\frac{(n^2+k^2)^3}{n^2-k^2}} = 1 + \mathcal{O}\left(\frac{k^2}{n^2}\right). \tag{4.57}$$

Eventually,

$$\mathrm{e}^{\left(1-\frac{n^2}{n^2+k^2}\right)\widehat{\beta}^2\sigma^2 \log n!} \sim \mathrm{e}^{\frac{k^2}{n^2} n \log n} = 1 + \mathcal{O}\left(\frac{k \log n}{n}\right). \tag{4.58}$$

Thus S_n becomes

$$\begin{aligned}
S_n &= \sum_{k=0}^{\infty} \frac{1}{k!} \cdot \frac{1}{\mathrm{e}} \cdot \left(1 + \mathcal{O}\left(\frac{k^2}{n^2}\right)\right) \cdot \left(\mathcal{O}\left(\frac{k \log n}{n}\right)\right) \\
&= 1 + \mathcal{O}\left(\frac{1}{n^\epsilon}\right),
\end{aligned} \tag{4.59}$$

provided that $k = \frac{n^{1-\epsilon}}{\log n}$.

In summary,

$$\frac{\mathbb{E}[Y^2]}{m^2 a_n^2} \to 1 \tag{4.60}$$

and

$$\mathbb{P}(A) \leq \frac{4\mathrm{Var}[Y]}{m^2 a_n^2} = \frac{\mathbb{E}[Y^2] - m^2 a_n^2}{m^2 a_n^2} \to 0, \tag{4.61}$$

as needed. □

References

M. Aizenman, J. L. Lebowitz, and D. Ruelle. Some rigorous results on the Sherrington-Kirkpatrick spin glass model. *Communications in Mathematical Physics*, 112(1):3–20, 1987.

A. Auffinger and W.-K. Chen. Free energy and complexity of spherical bipartite models. *Journal of Statistical Physics*, 157(1):40–59, 2014.

M. Beckman and T. C. Koopmans. Assignment problems and the location of economic activities. *Econometrica*, 25:53–76, 1957.

A. Bovier, I. Kurkova, and M. Löwe. Fluctuations of the free energy in the rem and the p-spin sk models. *The Annals of Probability*, 30(2):605–651, 2002.

J. M. Buhmann, A. Gronskiy, and W. Szpankowski. Free energy rates for a class of very noisy optimization problems. In *Analysis of Algorithms (AofA)*, pages 67–78, France, 2014.

J. M. Buhmann, J. Dumazert, A. Gronskiy, and W. Szpankowski. Posterior Agreement for Parameter-Rich Optimization Problems in the Asymptotic Limit, submitted, 2017.

A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38. Springer Science & Business Media, 2009.

B. Derrida. Random-energy model: An exactly solvable model of disordered systems. *Phys. Rev. B*, 24:2613–2626, 1981.

W. Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. Wiley, NY, 2nd edition, 1971.

J. B. G. Frenk, M. van Houweninge, and A. H. G. R. Kan. Asymptotic properties of the quadratic assignment problem. *Mathematics of Operations Research*, 10(1):100–116, 1985.

E. L. Lawler. The quadratic assignment problem. *Management science*, 9(4):586–599, 1963.

T. Łuczak. Phase transition phenomena in random discrete structures. *Discrete Mathematics*, 136(1):225 – 242, 1994.

A. Magner, D. Kihara, and W. Szpankowski. A study of the Boltzmann sequence-structure channel. *Proceedings of the IEEE*, 2016; also In *2016 International Symposium on Information Theory (ISIT)*, Barcelona, Spain, 2016.

M. Mézard and A. Montanari. *Information, Physics, and Computation*. Oxford University Press, 2009.

M. Mézard and G. Parisi. The cavity method at zero temperature. *Journal of Statistical Physics*, 111(1): 1–34, 2003.

G. Parisi. The mean field theory of spin glasses: The heuristic replica approach and recent rigorous results. *Letters in Mathematical Physics*, 88(1):255–269, 2009.

I. R. Savage. Mills' ratio for multivariate normal distributions. *J. Res. Nat. Bur. Standards Sect. B*, 66:93–96, 1962.

D. Sherrington and S. Kirkpatrick. Solvable Model of a Spin-Glass. *Physical Review Letters*, 35:1792–1796, 1975.

W. Szpankowski. Combinatorial optimization problems for which almost every algorithm is asymptotically optimal. *Optimization*, 33:359–368, 1995.

W. Szpankowski. *Avarage Case Analysis of Algorithms on Sequences*, Wiley, New York, 2001.

M. Talagrand. *Spin Glasses: A Challenge for Mathematicians: Cavity and Mean Field Models*. Springer Verlag, 2003.

J. Vannimenus and M. Mézard. On the statistical mechanics of optimization problems of the travelling salesman type. *Phys. Rev. Lett.*, 35:1792–1796, 1984.