Free Energy Rates for a Class of Very Noisy Optimization Problems

Joachim M. Buhmann[1], Alexey Gronskiy[1] and Wojciech Szpankowski[2*]

[1] *ETH, Department of Computer Science 8092 Zurich, Switzerland*
[2] *Department of Computer Science, Purdue University, Indiana, U.S.A.*

We study a class of stochastic optimization problems for which the cardinality of the set of feasible solutions (called also configurations) m and the size of every feasible solution N satisfy $\log m = o(N)$. Assuming the data (e.g. weights of a graph, edges in spanning tree problem, elements of matrices in assignment problem, etc.) to be random we adopt the maximum entropy framework by weighting the configurations with a Boltzmann distribution where the inverse computational temperature β controls the cost resolution. For a high noise level in the instances implying low β, we estimate the free energy in the asymptotic limit. This quantity plays a significant role in many applications, including algorithm analysis, robust optimization and so on. In particular, we prove that the free energy exhibits a phase transition in the second order term.

Keywords: Combinatorial optimization, Boltzmann distribution, free energy, partition function

1 Introduction

We consider a class of stochastic 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., set of vertices, elements of a matrix, keys, etc). The data X denote a set of random variables which 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 $R_{\max}(\mathcal{S}_n, X)$ or $R_{\min}(\mathcal{S}_n, X)$ defined as

$$R_{\max}(\mathcal{S}_n, X) = \max_{c\in\mathcal{C}_n}\Big\{ \sum_{i\in\mathcal{S}_n(c)} w_i(c, X)\Big\}\,, \quad R_{\min}(\mathcal{S}_n, X) = \min_{c\in\mathcal{C}_n}\Big\{ \sum_{i\in\mathcal{S}_n(c)} w_i(c, X)\Big\}\,, \tag{1}$$

where $\mathcal{C}_n$ is a set of all feasible solutions (call also *configurations*), $\mathcal{S}_n(c)$ 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 $w_i(c, X)$ is the weight assigned to the i-th object in the c-th feasible solution. Throughout this work, we assume that the weight distribution only depends on the data X but it is invariant on the feasible solution c, if $\mathcal{S}_n(c)$ is given. By this assumption, the data and, consequently, the weights will not change during optimization and we can adopt the notation $w_i(X) := w_i(c, X)$.

* This work was supported by NSF Center for Science of Information (CSoI) Grant CCF-0939370, NSA Grants H98230-11-1-0184 and H98230-11-1-0141, NSF Grants DMS-0800568, and CCF-0830140, the MNSW grant DEC-2013/09/B/ST6/02258, and in addition SNF Grant # 200021_138117. WS is also a Visiting Professor at ETI, Gdańsk University of Technology, Poland.

Combinatorial optimization problems arise in many areas of science and engineering. Among others we mention here: the assignment problem [14], the quadratic assignment problem [5, 8], computation of the minimum spanning tree, the minimum weighted k-clique problem [14], geometric location problems, and so forth. Often, the data entering the problem specification are random and sets of solutions have to be considered as equally likely given the stochastic instance. We analyze this class of random problems in a probabilistic framework which assumes that the weights $w_i(X)$ are Borel functions of X, s.t. $w_i(X)$ are i.i.d. with some distribution $F(\cdot)$. We also assume that the cardinality of the feasible set is m (i.e., $|\mathcal{C}_n| = m$) and the cardinality of $\mathcal{S}_n(c)$ is N for every $c \in \mathcal{C}_n$. Throughout this paper we shall demand that $\log m = o(N)$.

We study these optimization problems in the maximum entropy framework. Therefore, we consider the Boltzmann distribution over all configurations. This distribution is parametrized by $\beta = 1/T$ which is the inverse of temperature T. More precisely, defining the objective function $R(c, X) = \sum_{i \in \mathcal{S}_n(c)} w_i(X)$, the Boltzmann distribution $P(c|X)$ of $c \in \mathcal{C}_n$ is

$$P(c|X) = \frac{1}{Z(\beta, X)} \exp(-\beta R(c, X)), \quad \text{where} \quad Z(\beta, X) = \sum_{c \in \mathcal{C}} \exp(-\beta R(c, X)) \tag{2}$$

is the *partition function*.

It is quite revealing to study optimization problems in the thermodynamic framework through the Boltzmann distribution. In the high temperature when $\beta \to 0$, this distribution selects all configuration uniformly. On the other hand, when $\beta \to \infty$ the Boltzmann distribution concentrates on the set of optimal solutions with costs $R_{\max}$. Furthermore, the partition function $Z(\beta, X)$ can be used to express many interesting properties of optimal or almost optimal solutions [1].

It can also be used to characterize some thermodynamic limits such as entropy and free energy rates [7, 12]. In this paper, we focus on the free energy rates for high temperature when $\beta \to 0$. This limit is most interesting when the instances of optimization problems are affected by strong fluctuations which only support estimation of low 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. The normalization matters! Usually, one defines the free energy rate as

$$\gamma(\beta) = \lim_{n \to \infty} \frac{\mathbb{E}_X[\log Z(\beta, X)]}{\log |\mathcal{C}_n|}. \tag{3}$$

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 [13] the case $\log m \gg O(N)$ was analyzed, while here we focus on a class of optimization problems with $\log m = o(N)$ (e.g., the quadratic assignment problem [5, 8, 9] in which $N = n^2$ and $m = n!$). For this class of optimization problems, Szpankowski [9] proved that any solution is asymptotically optimal with high probability.

Furthermore, the free energy rate plays an important role in the novel verification approach for cost functions and algorithms proposed recently in [1]. Namely, optimization of $R(c, X')$ and $R(c, X'')$ under the two sample set scenario with *noisy* instances X', X'' requires to be robust w.r.t. noise. A framework of robust optimiztion has been developed in [1, 2] by using information theoretic arguments (see also [3]). Stable recovery of solution sets under two noisy inputs requires to optimize w.r.t. β the quantity

$$\mathcal{I}_\beta = \mathbb{E}_{X', X''} \log \frac{|\mathcal{C}_n| \sum_{c \in \mathcal{C}_n} \exp\bigl(-\beta(R(c, X') + R(c, X''))\bigr)}{\sum_{c \in \mathcal{C}_n} \exp\bigl(-\beta R(c, X')\bigr) \sum_{c \in \mathcal{C}_n} \exp\bigl(-\beta R(c, X'')\bigr)},$$

and then considering the solutions sampled from the Gibbs measure taken at the optimal β. $\mathcal{I}_\beta$ can be interpreted as a generalization capacity for cost functions or algorithms. Maximizing $\mathcal{I}_\beta$ requires to understand how $\mathbb{E}_X \log Z(\beta, X)$ behaves as a function of β.

For the mentioned framework, it is crucial to understand the asymtotics of the terms in the above equation. In the present paper, we establish such asymptotics for certain temperature regimes.

We shall argue that for the case $\log m = o(N)$, the proper normalization requires $\beta = \Theta(\sqrt{\log m/N}) \to 0$. In this *"high noise"* regime we will be able to determine the free energy rate. In fact, we shall prove that the second order term of the free energy rate exhibits a phase transition. We illustrate our findings on the quadratic assignment problem.

2 Main Results and Their Consequences

In this section, we formally introduce the problem, and present our main findings. We aim at understanding the asymptotic behavior of the partition function defined in (2), that is,

$$Z(\beta, X) = \sum_{c \in \mathcal{C}} \exp(-\beta R(c, X)) \tag{4}$$

where $\mathcal{C} := \mathcal{C}_n$ of cardinality $m := |\mathcal{C}_n|$ is the set of configurations or feasible solutions of the objective function $R(c, X)$ defined above.

Since the weights $w_i(X)$ are Borel functions of X, they are random variables. We assume them to be i.i.d. realizations of a random variable $W(X)$ with the probability distribution F that does not depend on i. Furthermore, we postulate that the moment generating function $G(t) = \mathbb{E}_X[\exp(tW(X))] < \infty$ exists for some $t > 0$. In particular, we denote

$$\mu = \mathbb{E}_X[W(X)], \quad \text{and} \quad \sigma^2 = \mathbb{V}_X[W(X)], \tag{5}$$

where $\mathbb{V}_X[W(X)] := \mathbb{E}_X[(W(X) - \mathbb{E}_X[W(X)])^2]$ denotes the variance. To simplify our analysis, we actually shall investigate the centralized weights $\overline{W}(X) := W(X) - \mu$ and denote by $G(\beta)$ the moment generating function of $-\overline{W}(X)$, that is

$$G(\beta) = \mathbb{E}_X[\exp(\beta(-\overline{W}(X)))] < \infty. \tag{6}$$

Our goal is to estimate $\mathbb{E}_X[\log Z(\beta, X)]$ which can be upper bounded by Jensen's inequality as

$$\mathbb{E}_X[\log Z(\beta, X)] \leq \log \mathbb{E}_X[Z(\beta, X)]. \tag{7}$$

Remark. In the following, we will omit X as an argument of $Z(\beta, X)$ and $R(c, X)$ for the sake of simplicity. (The expectation $\mathbb{E}[.]$, the variance $\mathbb{V}[.]$ and other probabilistic operations are still meant to be taken with respect to the randomness of X).

We need to evaluate $\mathbb{E}[Z(\beta)]$, so we proceed as follows

$$\begin{aligned}\mathbb{E}[Z(\beta)] &= \mathbb{E}\Big[\sum_{c \in \mathcal{C}} \exp(-\beta R(c))\Big] = \exp(-\beta N\mu)\mathbb{E}\Big[\sum_{c \in \mathcal{C}} \exp\big(-\beta(R(c) - N\mu)\big)\Big] \\ &= \exp(-\beta N\mu) m G^N(\beta). \end{aligned} \tag{8}$$

Thus

$$\log \mathbb{E}[Z(\beta)] = -\beta N\mu + \log m + N \log G(\beta) \tag{9}$$

since the r.v.s W_i are i.i.d.

From the above relation (9) one must conclude that in order to get a nontrivial limit of $\log \mathbb{E}[Z(\beta)]/\log m$ we need to choose the limit $\beta \to 0$. Under this assumption, we can expand $G(\beta)$ in the Taylor series to obtain

$$G(\beta) = 1 + \frac{1}{2}\beta^2\sigma^2 + O(\beta^3). \tag{10}$$

We find as long as $\beta \to 0$

$$\begin{aligned}\log \mathbb{E}[Z(\beta)] &= -\beta N\mu + \log m + N \log G(\beta) = -\beta N\mu + \log m + N \log\Big(1 + \frac{1}{2}\beta^2\sigma^2 + O(\beta^3)\Big)\\ &= -\beta N\mu + \log m + \frac{1}{2}N\beta^2\sigma^2(1 + O(\beta)).\end{aligned} \tag{11}$$

This suggests that the right choice for β is

$$\beta = \widehat{\beta}\sqrt{\frac{\log m}{N}} \tag{12}$$

for some constant $\widehat{\beta}$. Thus we arrive at

$$\frac{\log \mathbb{E}[Z(\beta)] + \beta N\mu}{\log m} = 1 + \frac{1}{2}\widehat{\beta}^2\sigma^2(1 + O(\beta)). \tag{13}$$

In terms of $\mathbb{E}[\log Z(\beta)]$ we find

$$\frac{\mathbb{E}[\log Z(\beta)] + \widehat{\beta}\mu\sqrt{N \log m}}{\log m} \le 1 + \frac{1}{2}\widehat{\beta}^2\sigma^2\left(1 + O\left(\sqrt{\frac{\log m}{N}}\right)\right). \tag{14}$$

But there is a surprise! Let us denote

$$\phi(\beta) = \mathbb{E}[\log Z(\beta)] + \beta N\mu =: \mathbb{E}[\log \widehat{Z}(\beta)] \tag{15}$$

where $\widehat{Z}(\beta) = \sum_{c\in\mathcal{C}} \exp(\beta \overline{R}(c))$ with $\overline{R}(c) = -\sum_{i\in\mathcal{S}(c)} \overline{W}_i$. It is easy to observe that

$$\beta \max_{c\in\mathcal{C}} \overline{R}(c) \le \log \widehat{Z}(\beta). \tag{16}$$

Using the upper bound obtained in (14) we find

$$\frac{\mathbb{E}[\max_{c\in\mathcal{C}} \overline{R}(c)]}{\log m} \le \sqrt{\frac{N}{\log m}}\left(\widehat{\beta}^{-1} + \frac{1}{2}\widehat{\beta}\sigma^2\right). \tag{17}$$

Choosing $\widehat{\beta}^* = \sqrt{2}/\sigma$ that minimizes the right-hand side of (17) we arrive at

$$\mathbb{E}[\max_{c\in\mathcal{C}} \overline{R}(c)] \le \sqrt{2\sigma^2 N \log m} \tag{18}$$

Now proceeding as in Talagrand [12, Proposition 1.1.3] we obtain

$$\phi'(\beta) \le \mathbb{E}[\max_{c\in\mathcal{C}} \overline{R}(c)]. \tag{19}$$

But for $\beta > \beta^* := \widehat{\beta}^* \sqrt{\log m/N}$,

$$\phi(\beta) \le \phi(\beta^*) + \phi'(\beta^*)(\beta - \beta^*), \tag{20}$$

since $\phi(\beta)$ is known to be convex. Applying the upper bound for $\phi'(\beta)$ yields

$$\mathbb{E}[\log \widehat{Z}(\beta)] \le \widehat{\beta}\sigma\sqrt{2}\log m \tag{21}$$

and the second upper bound in Theorem 1.

It is now worth proving that the lower bounds for $\mathbb{E}[\log \widehat{Z}(\beta)]$ are asymptotically the same as (14),(21). For that, we will follow the techniques used in Talagrand [12, Proposition 1.1.5, pp. 11–12].

Let Y be cardinality of the solution subset for which the centered negative cost function (see above) $\overline{R}(c)$ is large enough:

$$Y := \mathrm{card}\{c\colon \overline{R}(c) \ge s\log m\} \quad \text{for some} \quad s \ge 0. \tag{22}$$

It is obvious that

$$\mathbb{E}[Y] = ma, \quad \text{where} \quad a := \mathbb{P}(\overline{R}(c) \ge s\log m). \tag{23}$$

It is quite straightforward then to prove that

$$\mathbb{E}[Y^2] = ma + m(m-1)a, \quad \text{thus} \quad \mathbb{V}[Y] = ma - ma^2 \le ma. \tag{24}$$

Let A denote an event $\{Y \le ma/2\}$. Now by Markov inequality (second transition in the following chain)

$$\mathbb{P}(A) \le \mathbb{P}\big((Y - \mathbb{E}[Y])^2 \ge m^2a^2/4\big) \le 4\mathbb{V}[Y]/(m^2a^2) \le 4/(ma). \tag{25}$$

Next, we derive lower bounds for $\mathbb{E}[\log Z(\beta)]$ on the events A and $\Omega \setminus A$. For the latter, we have:

$$\widehat{Z}(\beta) = \sum_{c\in\mathcal{C}} \exp(\beta\overline{R}) \ge \sum_{c\in\mathcal{C}} \exp(\beta s\log m) \ge \frac{m}{2} a\exp(\beta s\log m), \tag{26}$$

thus

$$\mathbb{E}[\mathbb{1}_{\Omega\setminus A} \log \widehat{Z}(\beta)] \ge (1 - 4/(ma))(\log m - \log 2 + \log a + \beta s\log m). \tag{27}$$

For event A, we derive the lower bound in the following way. Choosing an arbitrary solution c_0, we notice that $Z(\beta) \ge \exp(\beta\overline{R}(c_0))$ and thus

$$\mathbb{E}[\mathbb{1}_A \log \widehat{Z}(\beta)] \ge -\beta\mathbb{E}[-\mathbb{1}_A\overline{R}(c)] \ge -\beta\mathbb{E}[|\overline{R}(c)|] \ge -L\sigma\beta\sqrt{N}, \tag{28}$$

where L is some constant coming from expectation of half-normal distribution, which is the thermodynamic limit distribution for $|\overline{R}(c)|$.

Combining (27) and (28), we obtain

$$\mathbb{E}[\log \widehat{Z}(\beta)] \ge \Big(1 - \frac{4}{ma}\Big)(\log m - \log 2 + \log a + \beta s\log m) - L\sigma\beta\sqrt{N}. \tag{29}$$

From the properties of centered gaussian, which is the limiting distribution of $\overline{R}(c)$ in the thermodynamic limit, we get the following bound on a:

$$\frac{\sigma\sqrt{N}}{Ls\log m}\exp\Big(-\frac{s^2\log^2 m}{2\sigma^2 N}\Big) \le a \le \exp\Big(-\frac{s^2\log^2 m}{2\sigma^2 N}\Big), \tag{30}$$

which means that in the thermodynamic limit ($n \to \infty$) holds true $ma \to \infty$, thus (29) turns into (we also normalize here)

$$\frac{\mathbb{E}[\log \widehat{Z}(\beta)]}{\log m} \ge 1 - s^2\frac{\log m}{2\sigma^2 N} + \beta s - \frac{L\sigma\beta\sqrt{N}}{\log m} + T, \tag{31}$$

where $T = \log\left(\frac{\sigma\sqrt{N}}{Ls\log m}\right)/\log m$. Now for the regime $\beta \le \widehat{\beta}^*\sqrt{\frac{\log m}{N}}$ we choose $s := \beta\frac{\sigma^2 N}{\log m}$, which yields a lower bound

$$\frac{\mathbb{E}[\log \widehat{Z}(\beta)]}{\log m} \ge 1 + \frac{\beta\sigma^2 N}{2\log m} - \frac{L\sigma\beta\sqrt{N}}{\log m} + T = 1 + \frac{\widehat{\beta}^2\sigma^2}{2} + \frac{L\sigma\beta\sqrt{N}}{\log m} + T. \tag{32}$$

And for regime $\beta \ge \widehat{\beta}^*\sqrt{\frac{\log m}{N}}$ we choose $s = \sqrt{\frac{2\sigma^2 N}{\log m}}$, which yields a lower bound

$$\frac{\mathbb{E}[\log \widehat{Z}(\beta)]}{\log m} \ge \beta\sqrt{\frac{2\sigma^2 N}{\log m}} - \frac{L\sigma\beta\sqrt{N}}{\log m} + T = \widehat{\beta}\sqrt{2}\sigma + \frac{L\sigma\beta\sqrt{N}}{\log m} + T. \tag{33}$$

The remaining terms $L\sigma\beta\sqrt{N}/\log m$ and T are small in the termodynamic limit, so we obtain the requested lower bound.

In passing, one should observe that bounds the bounds (both lower and upper) for two regimes are distinctively different. Thus the normalized free energy rate exhibits a phase transition at the second order term.

In summary, we have just proved the following finding.

Theorem 1 *Consider a class of combinatorial optimization problems in which the cardinality of feasible solutions m and the size of feasible solution N are related as $\log m = o(N)$. Assume that weights W_i are i.i.d. distributed with mean μ and variance σ^2 and moment generating function $G(\beta) < \infty$ for some $\beta > 0$. Define*

$$\beta = \widehat{\beta}\sqrt{\frac{\log m}{N}}. \tag{34}$$

Then the function $Z(\beta)$ satisfies

$$\lim_{n\to\infty}\frac{\mathbb{E}[\log Z(\beta)] + \widehat{\beta}\mu\sqrt{N\log m}}{\log m} = \begin{cases} 1 + \frac{\widehat{\beta}^2\sigma^2}{2} + O\left(\sqrt{\frac{\log m}{N}}\right) & \widehat{\beta} < \frac{\sqrt{2}}{\sigma}, \\ \widehat{\beta}\sigma\sqrt{2} & \widehat{\beta} \ge \frac{\sqrt{2}}{\sigma}. \end{cases} \tag{35}$$

In other words, the free energy rate $\gamma(\beta) = \lim_{n\to\infty}\mathbb{E}[\log Z(\beta)]/\log m$ becomes

$$\gamma(\beta) = -\widehat{\beta}\mu\sqrt{N\log m} + O(\log m) \tag{36}$$

with a phase transition at the second order term.

Remarks: First, the reader should note that the critical inverse temperature $\widehat{\beta}^* = \sqrt{2}/\sigma$ relates the low β limit to the high noise regime of optimization problems. Second, it may be useful to provide an heuristic argument behind Theorem 1. Let $Z(\beta) = \exp(-\beta N\mu)\widehat{Z}(\beta)$ where

$$\widehat{Z}(\beta) = \sum_{c\in\mathcal{C}} \exp\left(-\beta\sqrt{N}\frac{\sum_{i\in\mathcal{S}(c)} W_i - N\mu}{\sqrt{N}}\right) = \sum_{c\in\mathcal{C}} \exp\big(\beta\sqrt{N}\cdot\widehat{R}(c)\big). \tag{37}$$

Since W_i are i.i.d. we conclude that (note the minus sign which we put into $\widehat{R}(c)$ for convenience)

$$\widehat{R}(c) = -\frac{\sum_{i\in\mathcal{S}(c)} W_i - N\mu}{\sqrt{N}} \xrightarrow{d} \mathcal{N}(0,\sigma^2) \tag{38}$$

where $\mathcal{N}(0,\sigma^2)$ represents the normal distribution with mean zero and variance σ^2. We now consider two regimes of β. In the first regime, we re-compute $\mathbb{E}[\widehat{Z}(\beta)]$ by noting that its main contribution comes from one large $\widehat{R}(c)$. Indeed, note that (cf. [6])

$$\mathbb{E}\big[\exp\big(\beta\sqrt{N}\widehat{R}(c)\big)\big] \sim \frac{1}{\sqrt{2\pi\sigma}}\int \exp\left(\beta\sqrt{N}t - \frac{t^2}{2\sigma^2}\right)dt. \tag{39}$$

But the above integral achieves at $t_0 = \beta\sigma^2\sqrt{N}$ its maximum value (by the saddle point method)

$$E[\exp(\beta\sqrt{N}\widehat{R}(c))] \sim \exp\left(\frac{\beta^2N\sigma^2}{2}\right). \tag{40}$$

Most of $\mathbb{E}[\widehat{Z}(\beta)]$ comes from this one large $\widehat{R}(c)$. But its probability is bounded by

$$\mathbb{P}\Big(\bigcup_{c\in\mathcal{C}}\widehat{R}(c) > t_0\Big) \le \exp\left(\log m - \frac{\beta^2N\sigma^2}{2}\right) \tag{41}$$

which is very small for

$$\beta > \widehat{\beta}^*\sqrt{\frac{\log m}{N}}, \quad \widehat{\beta}^* = \sqrt{\frac{2}{\sigma^2}}. \tag{42}$$

Thus for $\widehat{\beta} < \widehat{\beta}^*$ we recover the first upper bound in (35).

To find the second bound, we proceed heuristically as follows. For $\beta > \sqrt{\frac{2\log m}{N\sigma^2}}$ we postulate that

$$\widehat{Z}(\beta) \sim \exp\big(\beta\sqrt{N}\max_{c\in\mathcal{C}}\widehat{R}(c)\big) \tag{43}$$

so that

$$\log\widehat{Z}(\beta) \sim \beta\sqrt{N}\max_{c\in\mathcal{C}}\widehat{R}(c). \tag{44}$$

But

$$\mathbb{P}(\max_{c\in\mathcal{C}}\widehat{R}(c) > t) \le \exp\left(\log m - \frac{t^2}{2\sigma^2}\right). \tag{45}$$

Therefore, with high probability we can assume that

$$\max_{c\in\mathcal{C}} \widehat{R}(c) \sim \sqrt{2\sigma^2 \log m}. \tag{46}$$

Combining it with the above lead to the second bound in (35).

In many applications (see [1, 2]) one needs more refined information about $\log Z(\beta)$. In particular, we must know whether $\log Z(\beta)$ is concentrated around $\mathbb{E}[\log Z(\beta)]$. In other words, whether

$$\frac{\log Z(\beta)}{\mathbb{E}[\log Z(\beta)]} \stackrel{pr}{\to} 1 \tag{47}$$

where $\stackrel{pr}{\to}$ represents the convergence in probability.

To address this question, we estimate first variance of $Z(\beta)$ or even better of $\log Z(\beta)$. The next lemma gives us a precise evaluation of the variance $\mathbb{V}[Z(\beta)] = \mathbb{E}[(Z(\beta) - \mathbb{E}[Z(\beta)])^2]$.

Lemma 1 *For any $\beta > 0$ we have*

$$\mathbb{V}[Z(\beta)] = (\mathbb{E}[Z(\beta)])^2 \left(\mathbb{E}_D \left(\frac{G(2\beta)}{(G(\beta))^2}\right)^D - 1\right) \tag{48}$$

where D is a random variable denoting the cardinality of the overlap $\mathcal{S}(c) \cap \mathcal{S}(c')$ for two distinct (chosen uniformly at random) feasible solutions $c, c' \in \mathcal{C}$. As before, $G(\beta)$ is the moment generating function of the weights W.

To illustrate an application of the above lemma and the behavior of the overlap D we discuss below the quadratic assignment problem [5, 8].

Example 1. *Quadratic Assignment Problem*
In the Quadratic Assignment Problem (further referred to as QAP), we consider two $n \times n$ matrices, namely the weight matrix V and the distance matrix H. The solution space is the set of all the n-element permutations S_n. The objective function is

$$R_{\text{QAP}}(\pi) = \sum_{i,j=1}^{n} V(i,j) \cdot H(\pi(i), \pi(j)), \quad \pi \in S_n.$$

Thus, in our notation, $\mathcal{C} = S_n$, $N = n^2$ and $m = n!$.

Recall that D is a random variable denoting the overlap of two arbitrary feasible solutions. We claim that

$$\mathbb{E}_D[D] = O(1). \tag{49}$$

Indeed, let π, π' be two random permutations and Y_{ij} be an indicator random variable of the event

$$\text{ovr}(i,j) = \{V(i,j) \cdot H(\pi(i), \pi(j)) = V(i,j) \cdot H(\pi'(i), \pi'(j))\},$$

thus

$$\mathbb{E}_D[D] = \mathbb{E}_D\Big[\sum_{i,j=1}^{n} Y_{ij}\Big] = \sum_{i,j=1}^{n} \mathbb{E}_D[Y_{ij}] = \sum_{i,j=1}^{n} \mathbb{P}(\text{ovr}(i,j)).$$

Now note that for each i, j the event $\mathrm{ovr}(i, j)$ occurs, if and only if $\pi(i) = \pi'(i)$ and $\pi(j) = \pi'(j)$. Thus $\mathbb{P}(\mathrm{ovr}(i, j)) = 1/n(n-1)$ and

$$\mathbb{E}_D[D] = \sum_{i,j=1}^{n} \mathbb{P}(\mathrm{ovr}(i, j)) = n^2 \frac{1}{n(n-1)} = O(1), \tag{50}$$

which proves (49).

Now are ready to formulate our second main finding.

Theorem 2 *If* $\beta \mathbb{E}_D[D] \to 0$, *then*

$$\frac{Z(\beta)}{\mathbb{E}[Z(\beta)]} \overset{pr}{\to} 1. \tag{51}$$

More precisely,

$$\mathbb{P}(|Z(\beta) - \mathbb{E}[Z(\beta)]| \geq \epsilon \mathbb{E}[Z(\beta)]) \leq \frac{\mathbb{V}[Z(\beta)]}{\epsilon^2 (\mathbb{E}[Z(\beta)])^2} = O(\beta^2 \mathbb{E}_D[D]) \to 0. \tag{52}$$

Proof: Recall that $G(\beta) = \mathbb{E} \exp(\beta(-W))$, where W is a random variable with expectation μ and variance σ^2. The Taylor expansion of $G(\beta)$ around 0 is

$$G(\beta) = 1 - \beta\mu + \frac{\beta^2 \mathbb{E}[W^2]}{2} + O(\beta^3). \tag{53}$$

Thus,

$$\begin{aligned}\left(\frac{G(2\beta)}{(G(\beta))^2}\right)^D &= \left(\frac{1 - 2\beta\mu + 2\beta^2\mathbb{E}[W^2] + O(\beta^3)}{[1 - \beta\mu + \beta^2\mathbb{E}[W^2]/2 + O(\beta^3)]^2}\right)^D = \left(\frac{1 - 2\beta\mu + 2\beta^2\mathbb{E}[W^2] + O(\beta^3)}{1 - 2\beta\mu + (\mu^2 + \mathbb{E}[W^2])\beta^2 + O(\beta^3)}\right)^D \\ &= (1 + (\mathbb{E}[W^2] - \mu^2)\beta^2 + O(\beta^3))^D = 1 + D\sigma^2\beta^2 + O(\beta^3)\end{aligned} \tag{54}$$

leading to

$$\mathbb{E}_D\left(\frac{G(2\beta)}{(G(\beta))^2}\right)^D = 1 + \sigma^2\beta^2\mathbb{E}_D[D] + O(\beta^3). \tag{55}$$

From the above we obtain the β-asymptotics of $\mathbb{V}[Z(\beta)]$.

$$\mathbb{V}[Z(\beta)] = (\mathbb{E}[Z(\beta)])^2 \left(\mathbb{E}_D\left(\frac{G(2\beta)}{(G(\beta))^2}\right)^D - 1\right) = (\mathbb{E}[Z(\beta)])^2(\sigma^2\beta^2\mathbb{E}_D[D] + O(\beta^3)), \tag{56}$$

and the theorem is proved. □

The last theorem implies that

$$\log Z(\beta) - \mathbb{E}[\log Z(\beta)] \overset{pr}{\to} 0. \tag{57}$$

But we would like to establish a stronger statement, say

$$\log Z(\beta) - \log \mathbb{E}[Z(\beta)] \overset{pr}{\to} 0 \tag{58}$$

or even $\log Z(\beta)/\mathbb{E}[\log Z(\beta)] \xrightarrow{pr} 1$. The following considerations should imply this. Observe that by expanding $\log Z(\beta)$ in the Taylor series around $\mathbb{E}[Z(\beta)]$ we have

$$\log Z(\beta) = \log \mathbb{E}[Z(\beta)] + \frac{Z(\beta) - \mathbb{E}[Z(\beta)]}{\mathbb{E}[Z(\beta)]} - \frac{1}{2} \frac{(Z(\beta) - \mathbb{E}[Z(\beta)])^2}{(\mathbb{E}[Z(\beta)])^2} + \sum_{k=3}^{\infty} \frac{(-1)^{k+1}}{k!} \frac{(Z(\beta) - \mathbb{E}[Z(\beta)])^k}{(\mathbb{E}[Z(\beta)])^k}.$$

Taking the expectation, we obtain

$$\mathbb{E}[\log Z(\beta)] = \log \mathbb{E}[Z(\beta)] - \frac{1}{2} \frac{\mathbb{V}[Z(\beta)]}{(\mathbb{E}[Z(\beta)])^2} + \sum_{k=3}^{\infty} \frac{(-1)^{k+1}}{k!} \frac{\mathbb{E}[(Z(\beta) - \mathbb{E}[Z(\beta)])^k]}{(\mathbb{E}[Z(\beta)])^k}. \tag{59}$$

Now we apply Theorem 2 in a stronger form. From the proof of Lemma 1 presented in the next section, we actually can conclude a stronger form of (52), namely, for any $k \geq 2$

$$\mathbb{P}\big(|(Z(\beta) - \mathbb{E}[Z(\beta)])^k| \geq \epsilon(\mathbb{E}[Z(\beta)])^k\big) \leq \frac{\mathbb{E}[(Z(\beta) - \mathbb{E}[Z(\beta)])^k]}{\epsilon(\mathbb{E}[Z(\beta)])^k} = O(\beta^2 \mathbb{E}_D[D_2(k)]) \tag{60}$$

where $D_2(k)$ is a random variable representing the cardinality of the joint pairwise overlaps $\bigcup_{s<t}\big(\mathcal{S}(c_s) \cap \mathcal{S}(c_k)\big)$ among k selected configurations $c_1, \ldots, c_k \in \mathcal{C}$. In many combinatorial optimization problems $E[D_2(k)]$ is constant or grows very slowly (e.g., in the quadratic assignment problem $E[D_2(k)] = O(1)$). Then (58) will follow.

3 Proof of Lemma 1

We prove now Lemma 1. Let

$$Z(\beta) = \exp(-\beta N \mu) \sum_{c \in \mathcal{C}} T(c), \quad T(c) = \exp\Big(\beta\Big(-\sum_{i=1}^{N} \overline{W}_i\Big)\Big). \tag{61}$$

Now define $\widehat{Z}(\beta) = \sum_{c \in \mathcal{C}} T(c)$. To compute $\mathbb{V}[\widehat{Z}(\beta)]$, we proceed as follows

$$\mathbb{E}[(\widehat{Z}(\beta))^2] = \mathbb{E}\Big[\sum_{c \in \mathcal{C}} T(c) \cdot \sum_{c' \in C} T(c')\Big] = \sum_{c, c' \in \mathcal{C}} \mathbb{E} \exp\Big(-\beta\Big(\sum_{i \in \mathcal{S}(c)} \overline{W}_i + \beta \sum_{j \in \mathcal{S}(c')} \overline{W}_j\Big)\Big). \tag{62}$$

Now let the solutions c and c' have an overlap $\mathcal{S}(c, c') := \mathcal{S}(c) \cap \mathcal{S}(c')$ of cardinality $d = d(c, c') := |\mathcal{S}(c, c')|$ (which we will call a *summand overlap*). We also define the symmetric difference $\bar{\mathcal{S}}(c, c') := \mathcal{S}(c) \triangle \mathcal{S}(c')$ and continue the chain of equalities:

$$\mathbb{E}[(\widehat{Z}(\beta))^2] = \sum_{c, c' \in \mathcal{C}} \mathbb{E} \exp\Big(-\beta\Big(2 \sum_{i \in \mathcal{S}(c,c')} \overline{W}_i + \sum_{j \in \bar{\mathcal{S}}(c,c')} \overline{W}_j\Big)\Big). \tag{63}$$

Here the sets of $\mathcal{S}(c,c')$ and $\bar{\mathcal{S}}(c,c')$ are independent, allowing us to decompose the expectation into the product:

$$\mathbb{E}[(\widehat{Z}(\beta))^2] = \sum_{c,c'\in\mathcal{C}} \mathbb{E}\exp\Big(-\beta\Big(2\sum_{i\in\mathcal{S}(c,c')}\overline{W}_i\Big)\Big)\cdot\mathbb{E}\exp\Big(-\beta\Big(\sum_{j\in\bar{\mathcal{S}}(c,c')}\overline{W}'_j\Big)\Big)$$
$$= \sum_{c,c'\in\mathcal{C}}(G(2\beta))^d(G(\beta))^{2(N-d)} = (G(\beta))^{2N}\sum_{c,c'\in\mathcal{C}}\left(\frac{G(2\beta)}{(G(\beta))^2}\right)^d. \tag{64}$$

Now assume that the probability of the two solutions c and c', chosen uniformly at random, to have a d-element overlap is $P_{\text{ovr}}(d)$ and rewrite the above as follows:

$$\mathbb{E}[(\widehat{Z}(\beta))^2] = (G(\beta))^{2N}\sum_{d=0}^{N} m^2 P_{\text{ovr}}(d)\left(\frac{G(2\beta)}{(G(\beta))^2}\right)^d = m^2(G(\beta))^{2N}\sum_{d=0}^{N}P_{\text{ovr}}(d)\left(\frac{G(2\beta)}{(G(\beta))^2}\right)^d$$
$$= (\mathbb{E}[\widehat{Z}(\beta)])^2\sum_{d=0}^{N}P_{\text{ovr}}(d)\left(\frac{G(2\beta)}{(G(\beta))^2}\right)^d. \tag{65}$$

We conclude that

$$\mathbb{V}[\widehat{Z}(\beta)] = \mathbb{E}[(\widehat{Z}(\beta))^2] - (\mathbb{E}[\widehat{Z}(\beta)])^2 = (\mathbb{E}[\widehat{Z}(\beta)])^2\left(\mathbb{E}_D\left(\frac{G(2\beta)}{(G(\beta))^2}\right)^D - 1\right),$$

where D is a random variable denoting the summand overlap in two randomly chosen solutions. Recalling that $Z(\beta) = \exp(-\beta N\mu)\widehat{Z}(\beta)$, we obtain the version without hats:

$$\mathbb{V}[Z(\beta)] = \mathbb{E}[(Z(\beta))^2] - (\mathbb{E}[Z(\beta)])^2 = (\mathbb{E}[Z(\beta)])^2\left(\mathbb{E}_D\left(\frac{G(2\beta)}{(G(\beta))^2}\right)^D - 1\right).$$

This proves Lemma 1.

4 Conclusion and Outlook

This paper discusses the low β asymptotics of the free energy for a class of optimization functions that show a bounded growth of the configuration space w.r.t. the solution complexity. In the concrete example of the quadratic assignment problem, the configuration space is the symmetric group with $n!$ configurations and assignment solutions of complexity n^2 for $n \times n$ matrices. Theorem 1 establishes a proportionality between the critical temperature and the noise level in the optimization problem. Consequently, the low β limit is justified for very noisy optimization problems as they arise in a variety of modern high throughput experimental designs. There, the precision of the individual experiments is traded off against the number of experiments that can be performed. This situations appears to us being prevalent in many *big data* scenarios today.

In addition, we like to emphasize that free energies evaluated on two different data instances determine a model validation criterion for cost functions and algorithms. The results of theorem 1 will enable us to determine the optimal resolution of optimization problems and the optimal precision of algorithms in the high noise limit.

References

[1] J.M. Buhmann, Information theoretic model validation for clustering. *ISIT*, 1398-1402, 2010

[2] J. M. Buhmann, SIMBAD: Emergence of pattern similarity. In Marcello Pelillo, editor, *Similarity-Based Pattern Analysis and Recognition*, Advances in Vision and Pattern Recognition, pages 45–64. Springer Berlin / Heidelberg, 2013.

[3] Joachim M. Buhmann, Matúš Mihal'ák, Rastislav Šrámek, and Peter Widmayer. Robust optimization in the presence of uncertainty. Innovations in Theoretical Computer Science (ITCS), 2013, pp. 505–514.

[4] R. Burkard and U. Fincke, Probabilistic Asymptotic properties of some Combinatorial Optimization Problems, *Discrete Applied Mathematics*, 12, 21-29, 1985.

[5] Frenk, J., van Houweninge, M. and Rinnooy Kan, A., Asymptotic Properties of the Quadratic Assignment Problem, *Mathematics of Operations Research*, 10, 100– 116, 1985.

[6] A. Magner, D. Kihara and W. Szpankowski. Phase Transition in Protein Folding Channel. *ISIT*, 2014 (submitted)

[7] Marc Mézard and Andrea Montanari. *Information, Physics, and Computation*. Oxford University Press, Inc., New York, NY, USA, 2009.

[8] W.T. Rhee, A Note on Asymptotic Properties of the Quadratic Assignment Problem, *Operations Research Letters*, 7, 197-200, 1988.

[9] W. Szpankowski, Combinatorial optimization problems for which almost every algorithm is asymptotically optimal, *Optimization*, 33, 359-368, 1995.

[10] W. Szpankowski. *Average Case Analysis of Algorithms on Sequences*. John Wiley & Sons, Inc., New York, NY, USA, 2001.

[11] Hiroshi Takahata. On the rates in the central limit theorem for weakly dependent random fields. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 64, 1983.

[12] M. Talagrand. *Spin Glasses: A Challenge for Mathematicians*. Springer, New York, NY, USA, 2003.

[13] J. Vannimenus and M. Mézard, On the Statistical Mechanics of Optimization Problems of the traveling Salesman Type, *J. de Physique Lettres*, 45, 1145-1153, 1984.

[14] Weide, B., Random Graphs and Graph Optimization Problems, *SIAM J. Computing*, 9, 552–557, 1980.