Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Posterior agreement for large parameter-rich optimization problems

Joachim M. Buhmann[a,1], Julien Dumazert[a], Alexey Gronskiy[a,*,1], Wojciech Szpankowski[b,2]

[a] ETH Zurich, 8092 Zurich, Switzerland
[b] Purdue University, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 22 December 2016
Accepted 10 April 2018
Available online xxxx
Communicated by J. Díaz

Keywords:
Free energy
Uncertain optimization
Gibbs distribution
Random energy model
Partition function asymptotics
Minimum Bisection
Quadratic Assignment
Information Criterion

ABSTRACT

Most real world combinatorial optimization problems are affected by noise in the input data, thus behaving in the high noise limit like large disordered particle systems, e.g. spin glasses or random networks. Due to uncertainty in the input, optimization of such disordered instances should infer stable posterior distributions of solutions conditioned on the noisy input instance. The maximum entropy principle states that the most stable distribution given the noise influence is defined by the *Gibbs distribution* and it is characterized by the *free energy*. In this paper, we first provide rigorous asymptotics of the difficult problem to compute the free energy for two combinatorial optimization problems, namely the sparse Minimum Bisection Problem (sMBP) and Lawler's Quadratic Assignment Problem (LQAP). We prove that both problems exhibit phase transitions equivalent to the discontinuous behavior of Derrida's Random Energy Model (REM). Furthermore, the derived free energy asymptotics lead to a theoretical justification of a recently introduced concept [3] of *Gibbs posterior agreement* that measures stability of the Gibbs distributions when the cost function fluctuates due to randomness in the input. This relatively new stability concept may potentially provide a new method to select robust solutions for a large class of optimization problems.

1. Introduction

1.1. Overview

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 defining such instances. Algorithms that minimize these noisy instances or approximate their global minimum return a solution that is a random variable due to input randomness and that is

* Corresponding author.
E-mail addresses: jbuhmann@inf.ethz.ch (J.M. Buhmann), julien.dumazert@gmail.com (J. Dumazert), alexeygr@inf.ethz.ch (A. Gronskiy), szpan@purdue.edu (W. Szpankowski).

[1] This work was supported by SNF Grant # 200021_138117.
[2] 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.

https://doi.org/10.1016/j.tcs.2018.04.015

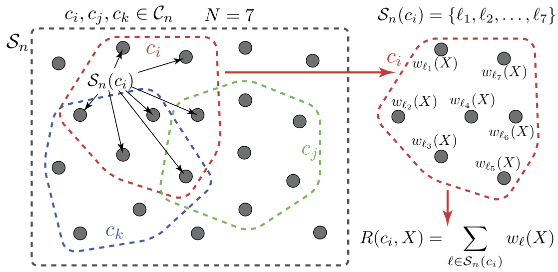

Fig. 1. Illustration of the notation: each of the solutions (examples shown in the figure are c_i, c_j, c_k) includes N (in the figure $N = 7$) objects from the underlying set $\mathcal{S}_n$. The cost function of a solution is the sum of weights assigned to the objects, which belong to that solution.

most often highly unstable. Therefore, we ask the natural questions: What is the distribution of the output returned by the algorithm? Can we stabilize such an output distribution by regularizing the algorithm?

Algorithm design in noise affected real world settings requires both statistical as well as computational considerations: first, we have to ensure that outputs of algorithms are typical in a statistical sense, i.e., they have to occur with high probability. Second, such typical outputs have to be computable in an efficient way with efficient resources. The reader should notice that statistical requirements dominate computational ones in an epistemological sense: A computational result has to be rejected if it is atypical since it lacks predictive power. Computationally, we might require significantly different algorithmic resources (time and space) to calculate typical solutions for typical inputs compared to minimizing the empirical risk.

Due to the statistical nature of inference, we have to efficiently compute posterior distributions of solutions given input data. Open theoretical issues emerge for this strategy, e.g., analytical computation of macroscopic properties like entropy, expected log-partition function or expected costs [12,25]. The expected log-partition function known also as the *free energy*, appeared in the context of combinatorial optimization since the mid 80's; see e.g., Vannimenus and Mézard [26] which explored the free energy properties of the traveling salesman problem. An intriguing property of free energy is the emergence of discontinuities of certain order when changing the concentration of the posterior distribution. Such abrupt changes of macroscopic properties, also known as *phase transitions*, are characteristic features of various large systems and have generated a long-lasting interest in theory of discrete structures (see [7,15]).

The concept "*free energy*" found also applications in theoretical computer science. Recently, in a series of papers on robust learning, Buhmann [3], Busse et al. [5] introduced a robustness score function called the *expected log-posterior agreement* (eLPA) for measuring "goodness" of robust solutions. Although the eLPA arose in a different field, it is tightly connected to computing free energies, as we see later in the paper. Furthermore, estimating the free energy for combinatorial optimization problems allow us to justify theoretically some experimental results obtained for these problems.

For the sake of completeness we should mention here that the statistical physics community developed an equally intensive research interest for finding *theoretical* laws that govern the behavior of macroscopic thermodynamic properties as the free energy. Many interesting models of such large systems were introduced relatively early, e.g. the *Sherrington–Kirkpatrick (SK) spin glass model* (see [22]). It required, however, considerable time and effort to develop rigorous techniques for solving them. For example, Derrida [9] introduced a very simple, but exactly solvable model called *random energy model (REM)* as the limit of the SK models family. Later, Aizenman et al. [1] published an exact solution in the high-temperature phase for the SK model. The general question of the exact free energy behavior became increasingly fascinating: it triggered a new wave of latest research (see [2,25]). The reader should also note that many interesting heuristic tools have been developed in the context of statistical physics over the last several decades, such as the replica method [20], the cavity method [19] and meanfield approximation schemes with belief propagation algorithms.

1.2. Notation and setting

We consider optimization problems that can be formulated as follows (for explanation see Fig. 1): let n be some integer determining the size of the problem (e.g., number of vertices in a graph, size of a matrix, etc.), and $\mathcal{S}_n$ a finite set of objects (e.g., set of edges, elements of a matrix, etc). Let X denote the input to the problem (data).

Define $\mathcal{C}_n$ as a finite set of all feasible solutions (e.g. bisections of a graph), and $\mathcal{S}_n(c) \subseteq \mathcal{S}_n$, $c \in \mathcal{C}_n$, as a finite set of objects belonging to the feasible solution c (e.g., set of edges belonging to a bisection). Let $w_i(X) = W_i$, $i \in \mathcal{S}_n$, be the weight assigned to the i-th object. In this paper we consider optimization problems for which the cost function and optimization task are defined as follows:

$$R(c, X) = \sum_{i \in \mathcal{S}_n(c)} w_i(X) \quad \text{and} \quad c_{\text{opt}}(X) = \arg\min_{c \in \mathcal{C}_n} R(c, X). \tag{1}$$

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., $N := |\mathcal{S}_n(c)|$). In this paper, we focus on optimization problems in which $\log m = o(N)$ holds true (see [23]). We call