

# A Note on the Boltzmann Distribution and the Linear Ordering Problem

Josu Ceberio<sup>1</sup>, Alexander Mendiburu<sup>2</sup>, and Jose A. Lozano<sup>3,4</sup>

<sup>1</sup> Department of Computer Languages and Systems,  
University of the Basque Country UPV/EHU, 48013 Bilbao, Spain  
josu.ceberio@ehu.eus

<sup>2</sup> Department of Computer Architecture and Technology,

<sup>3</sup> Department of Computer Science and Artificial Intelligence,  
University of the Basque Country UPV/EHU, 20018 Donostia, Spain

<sup>4</sup> Basque Center for Applied Mathematics (BCAM), 48009 Bilbao, Spain

**Abstract.** The Boltzmann distribution plays a key role in the field of optimization as it directly connects this field with that of probability. Basically, given a function to optimize, the Boltzmann distribution associated to this function assigns higher probability to the candidate solutions with better quality. Therefore, an efficient sampling of the Boltzmann distribution would turn optimization into an easy task. However, inference tasks on this distribution imply performing operations over an exponential number of terms, which hinders its applicability. As a result, the scientific community has investigated how the structure of objective functions is translated to probabilistic properties in order to simplify the corresponding Boltzmann distribution. In this paper, we elaborate on the properties induced in the Boltzmann distribution associated to permutation-based combinatorial optimization problems. Particularly, we prove that certain characteristics of the linear ordering problem are translated as conditional independence relations to the Boltzmann distribution in the form of  $L - decomposability$ .

**Keywords:** Boltzmann distribution, combinatorial optimization, L-decomposability, linear ordering problem, permutation

## 1 Introduction

In the last decades, the Boltzmann distribution has been a recurrent research topic in the field of combinatorial optimization as it constitutes one of the bridges between this field and that of probability [4]. Basically, given an objective function  $f$ , the associated Boltzmann distribution assigns higher probabilities to those solutions with better objective values. Formally, under this distribution, the probability of a solution  $x$  is defined as

$$P(x) = \frac{\exp(\beta f(x))}{Z_f(\beta)} \quad x \in \Omega \quad (1)$$

where  $\Omega$  denotes the set of all candidate solutions (which is usually of exponential size),  $\beta$  is the Boltzmann constant, and  $Z_f(\beta)$  stands for the partition function. When  $\beta$  equals 0, the distribution becomes uniform. In contrast, when  $\beta$  is large, the probability concentrates around the global optimal solutions.

Due to its ideal properties (the better the solution, the higher its probability), if it were possible to efficiently sample the Boltzmann distribution for any  $\beta$ , optimization would be an easy task [7]. However, in general, there is no closed form expression for  $Z_f(\beta)$ , and therefore, any inference on this distribution usually implies performing a sum over exponentially many terms:

$$Z_f(\beta) = \sum_{y \in \Omega} \exp(\beta f(y)) \quad (2)$$

In spite of these difficulties, the scientific community has researched how the structural properties of the objective function are translated to the probability properties of the corresponding Boltzmann distributions, and how this helps in the optimization process. In this sense, Muhlenbein et al. [7] showed that, in the case of combinatorial problems, it is possible to simplify the representation of the Boltzmann distribution. Particularly, they translated properties of additively decomposable functions into conditional independence assertions between variables that, finally, produced (efficient, in some cases) factorizations of the associated Boltzmann distribution. Motivated by such results, the authors proposed a novel estimation of distribution algorithm based on the Boltzmann distribution.

Nonetheless, the work by Muhlenbein et al. [7] is not applicable to any search space. For instance, when solving permutation-based combinatorial optimization problems, the set of all permutations (of a given size  $n$ ) constitutes the search space of candidate solutions. In this case, due to the *mutual exclusivity constraints* associated with the codification of solutions, the previous conditional independence assertions between variables do not provide information. However, numerous works on permutations probabilistic modelling have defined independence properties that do consider the nature of permutations [3, 2, 8, 5], and therefore, allow to factorize the distribution. The most recurrent of these properties is the *L-decomposability* [2].

According to Csizsar [3], a probability distribution  $P$  over the space of permutations  $\mathbb{S}_n$  is said to be *L-decomposable*, if there are choice probabilities  $P_C(i)$  for all subsets  $C \subseteq \{1, \dots, n\}$  and  $i \in C$ , such that

$$P(\sigma) = \prod_{r=1}^n P_{\{\sigma(r), \dots, \sigma(n)\}}(\sigma(r)) \quad \forall \sigma \in \mathbb{S}_n \quad (3)$$

where  $n$  denotes the size of the permutations and the choice probability  $P_C(i)$  is the probability that item  $i$  is chosen as the best preferred from the subset  $C$  of items. This property implies that the choice probabilities at the  $r$ -th step depend only on the set of items remaining at that step and are independent of the ordering of the previously selected items.

As carried out in Muhlenbein et al. [7], in this paper we aim to set the basis to also use the Boltzmann distribution to optimize permutation-based optimization problems. To this end, we show that the structure of these problems can be interpreted in the domain of probability as specific properties of the associated Boltzmann distribution. In this manuscript, we take the Linear Ordering Problem (LOP) [6] as a case of study, and demonstrate that its associated Boltzmann distribution is *L-decomposable*.

Nonetheless, it is worth mentioning that this research may be extended to other permutation problems with different structures than that of LOP given rise to other independence properties on the corresponding Boltzmann distribution.

Given a matrix  $B = [b_{k,l}]_{n \times n}$  of numerical entries, the LOP consists of finding a simultaneous permutation  $\sigma$  of the rows and columns of  $B$ , such that the sum of the entries above the main diagonal is maximized (see Fig. 1). The equation below formalizes the LOP function:

$$f(\sigma) = \sum_{k=1}^{n-1} \sum_{l=k+1}^n b_{\sigma(k),\sigma(l)} \quad (4)$$

where  $\sigma(k)$  and  $\sigma(l)$  denote the items ranked at positions  $k$  and  $l$  in the solution  $\sigma$ .

	1	2	3	4	5
1	0	16	11	15	7
2	21	0	14	15	9
3	26	23	0	26	12
4	22	22	11	0	13
5	30	28	25	24	0

Fig. 1: Example of an LOP instance of  $n = 5$ .

In the LOP, the influence of locating a given item at position  $r$  ( $r \in \{1, \dots, n\}$ ) to the quality of the solution  $\sigma$  depends on the distribution of the items in the previous  $\{1, \dots, r-1\}$  and posterior  $\{r+1, \dots, n\}$  positions. However, it does not change for any ordering of the items within the subsets [1]. In the following section, we demonstrate that such a notion of independence between the items is interpreted as *L-decomposability* of the associated Boltzmann distribution.

## 2 Main result

In this section, the main result of the manuscript is introduced: Theorem 1, the *L-decomposability* of the Boltzmann distribution associated to the LOP.

**Theorem 1.** *The Boltzmann distribution associated to the linear ordering problem is L-decomposable.*

*Proof.* In order to verify that a probability distribution  $P(\sigma)$  is *L-decomposable*, it is a necessary and sufficient condition that, for each  $r = 3, \dots, n$  and for each  $i_1, \dots, i_r$ , the conditional probability

$$P(\sigma(r) = i_r | \sigma(1) = i_1, \dots, \sigma(r-1) = i_{r-1}) \quad (5)$$

is a symmetric function of  $i_1, \dots, i_{r-1}$  [2]. Therefore, without loss of generality, by proving the equality below for every permutation  $\pi$  of the items  $\{i_1, \dots, i_{r-1}\}$

$$\begin{aligned} P(\sigma(r) = i_r | \sigma(1) = i_1, \dots, \sigma(r-1) = i_{r-1}) &= \\ P(\sigma(r) = i_r | \sigma(1) = \pi(1), \dots, \sigma(r-1) = \pi(r-1)) & \end{aligned} \quad (6)$$

where  $\forall j, k \in \{1, \dots, r-1\}$ ,  $\pi(j) \in \{i_1, \dots, i_{r-1}\}$  and  $\pi(j) \neq \pi(k)$ , unless  $j = k$ , we prove that  $P(\sigma)$  is *L-decomposable*.

So, from (6), by applying the Bayes' theorem, we obtain

$$\begin{aligned} \frac{P(\sigma(1) = i_1, \dots, \sigma(r-1) = i_{r-1}, \sigma(r) = i_r)}{P(\sigma(1) = i_1, \dots, \sigma(r-1) = i_{r-1})} &= \\ \frac{P(\sigma(1) = \pi(1), \dots, \sigma(r-1) = \pi(r-1), \sigma(r) = i_r)}{P(\sigma(1) = \pi(1), \dots, \sigma(r-1) = \pi(r-1))} & \end{aligned} \quad (7)$$

and, by replacing the terms with the Boltzmann distribution of the LOP (8),

$$P(\sigma) = \frac{\exp(\beta \sum_{k=1}^{n-1} \sum_{l=k+1}^n b_{\sigma(k), \sigma(l)})}{Z_f(\beta)} \quad \sigma \in \mathbb{S}_n \quad (8)$$

we rewrite (7) as follows<sup>5</sup>:

$$\frac{\sum_{\sigma \in S_r} \exp(\sum_{k=1}^{n-1} \sum_{l=k+1}^n b_{\sigma(k), \sigma(l)})}{\sum_{\sigma \in S_{r-1}} \exp(\sum_{k=1}^{n-1} \sum_{l=k+1}^n b_{\sigma(k), \sigma(l)})} = \frac{\sum_{\sigma \in T_r} \exp(\sum_{k=1}^{n-1} \sum_{l=k+1}^n b_{\sigma(k), \sigma(l)})}{\sum_{\sigma \in T_{r-1}} \exp(\sum_{k=1}^{n-1} \sum_{l=k+1}^n b_{\sigma(k), \sigma(l)})} \quad (9)$$

where  $S_r$  and  $S_{r-1}$  denote the set of permutations that agree with  $i_1, \dots, i_r$  and  $i_1, \dots, i_{r-1}$  orderings in the first  $r$  and  $r-1$  positions.  $T_{r-1}$  denotes the set of permutations that agree with  $\pi$  in the first  $r-1$  positions, and  $T_r$  restricts, even more, the previous set by fixing item  $i_r$  at position  $r$ .

Let us first focus on the left-hand side of the equality in (9). Taking into account the positions in which items are already known, we decompose the exponent terms in the numerator and denominator (10) as illustrated in Fig. 2.

<sup>5</sup> Note that partition functions  $Z(\beta)$  and the Boltzmann constant  $\beta$  are cancelled.

$$\frac{\sum_{\sigma \in S_r} \exp \left( \sum_{k=1}^{r-1} \sum_{l=k+1}^r b_{i_k, i_l} + \underbrace{\sum_{k=1}^r \sum_{l=r+1}^n b_{\sigma(k), \sigma(l)}}_{X_1(\sigma)} + \underbrace{\sum_{k=r+1}^{n-1} \sum_{l=k+1}^n b_{\sigma(k), \sigma(l)}}_{Y_1(\sigma)} \right)}{\sum_{\sigma \in S_{r-1}} \exp \left( \sum_{k=1}^{r-2} \sum_{l=k+1}^{r-1} b_{i_k, i_l} + \underbrace{\sum_{k=1}^{r-1} \sum_{l=r}^n b_{\sigma(k), \sigma(l)}}_{X_2(\sigma)} + \underbrace{\sum_{k=r}^{n-1} \sum_{l=k+1}^n b_{\sigma(k), \sigma(l)}}_{Y_2(\sigma)} \right)} \quad (10)$$

For readability purposes, we replace the terms related to  $\sigma(r), \dots, \sigma(n)$  with  $X_1, X_2, Y_1$  and  $Y_2$ .

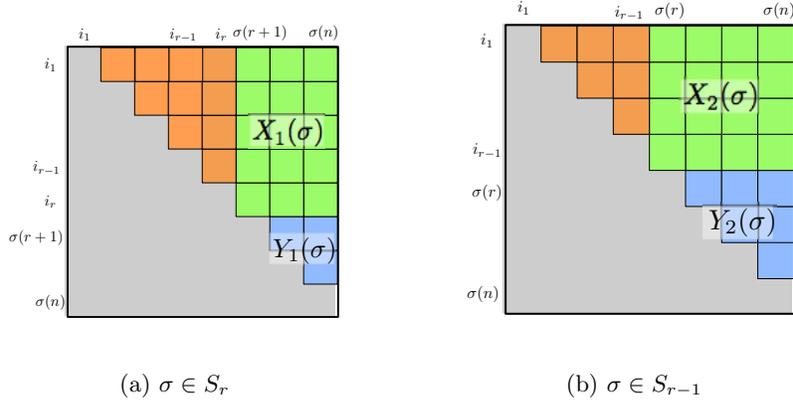


Fig. 2: Decomposition of  $f(\sigma)$  for the solutions in  $S_r$  and  $S_{r-1}$ .

Note that for the solutions in  $S_r$  and  $S_{r-1}$ , the items in the first  $r - 1$  positions are equal, i.e.  $i_1, \dots, i_{r-1}$ . Thus, we extract the common factor from the numerator and denominator (the terms that correspond to the computation of items  $i_1, \dots, i_{r-1}$ ), and the equation is simplified to

$$\frac{\sum_{\sigma \in S_r} \exp \left( \sum_{k=1}^{r-1} b_{i_k, i_r} + X_1(\sigma) + Y_1(\sigma) \right)}{\sum_{\sigma \in S_{r-1}} \exp (X_2(\sigma) + Y_2(\sigma))} \quad (11)$$

Now, we perform a similar procedure on the right-hand side of the equality in (9) by following the decomposition of the objective function in Fig. 3:

$$\frac{\sum_{\sigma \in T_r} \exp \left( \sum_{k=1}^{r-2} \sum_{l=k+1}^{r-1} b_{\pi(k), \pi(l)} + \sum_{k=1}^{r-1} b_{\pi(k), i_r} + X_1(\sigma) + Y_1(\sigma) \right)}{\sum_{\sigma \in T_{r-1}} \exp \left( \sum_{k=1}^{r-2} \sum_{l=k+1}^{r-1} b_{\pi(k), \pi(l)} + X_2(\sigma) + Y_2(\sigma) \right)} \quad (12)$$

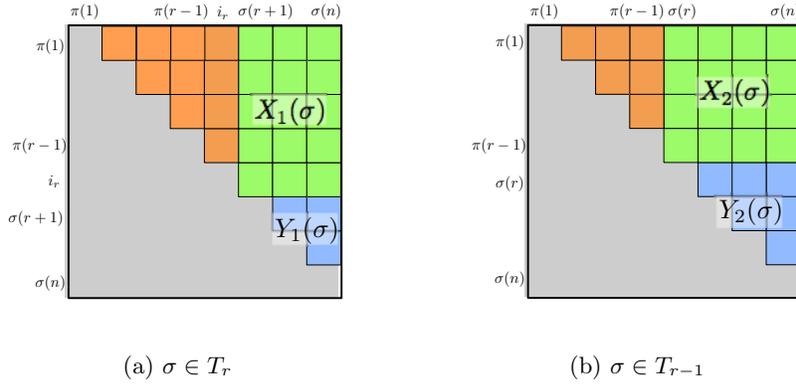


Fig. 3: Decomposition of  $f(\sigma)$  for the solutions in  $T_r$  and  $T_{r-1}$ .

By extracting the common factor corresponding to positions  $1, \dots, r-1$  we obtain

$$\frac{\sum_{\sigma \in T_r} \exp \left( \sum_{k=1}^{r-1} b_{\pi(k), i_r} + X_1(\sigma) + Y_1(\sigma) \right)}{\sum_{\sigma \in T_{r-1}} \exp \left( X_2(\sigma) + Y_2(\sigma) \right)} \quad (13)$$

Since the same set of items is ordered in the first  $r-1$  positions in the permutations in  $S_r$  and  $T_r$ , then  $\sum_{k=1}^{r-1} b_{i_k, i_r}$  equals  $\sum_{k=1}^{r-1} b_{\pi(k), i_r}$ , and thus, (13) equals (11). Therefore, we proved that (6) is true and the Boltzmann distribution associated to the LOP is  $L$ -decomposable.

## Acknowledgements

This work has been partially supported by the Research Groups 2013-2018 (IT-609-13) programs (Basque Government) and TIN2013-41272P (Ministry of Science and Technology). Jose A. Lozano is also supported by BERC 2014-2017 and Elkartek programs (Basque government) and Severo Ochoa Program SEV-2013-0323 (Spanish Ministry of Economy and Competitiveness).

## References

1. Josu Ceberio, Alexander Mendiburu, and Jose A. Lozano. The Linear Ordering Problem Revisited. *European Journal of Operational Research*, 241(3):686–696, 2014.
2. Douglas E Critchlow, Michael A Fligner, and Joseph S Verducci. Probability models on rankings. *Journal of Mathematical Psychology*, 35(3):294 – 318, 1991.
3. Villo Csiszar. On L-decomposability of random orderings. *Journal of Mathematical Psychology*, 53:294–297, 2009.
4. Scott Kirkpatrick, C Daniel Gelatt, Mario P Vecchi, et al. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.
5. Duncan Luce R. *Individual Choice Behavior*. Wiley, New York, 1959.
6. Rafael Martí and Gerhard Reinelt. *The linear ordering problem: exact and heuristic methods in combinatorial optimization*, volume 175. Springer, 2011.
7. Heinz Muhlenbein, Thilo Mahnig, and Fhg Ais. Evolutionary algorithms and the boltzmann distribution. In *FOGA 7*, pages 525–556, 2003.
8. Robin L. Plackett. The Analysis of Permutations. *Journal of the Royal Statistical Society*, 24(10):193–202, 1975.