

# MAXIMA AND NEAR-MAXIMA OF A GAUSSIAN RANDOM ASSIGNMENT FIELD

Mordant, G. and J. Segers

LIDAM Discussion Paper ISBA/2021/08

# Maxima and near-maxima of a Gaussian random assignment field

Gilles Mordant\*, Johan Segers

LIDAM/ISBA, UCLouvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium

---

## Abstract

The assumption that the elements of the cost matrix in the classical assignment problem are drawn independently from a standard Gaussian distribution motivates the study of a particular Gaussian field indexed by the symmetric permutation group. The correlation structure of the field is determined by the Hamming distance between two permutations. The expectation of the maximum of the field is shown to go to infinity in the same way as if all variables of the field were independent. However, the variance of the maximum is shown to converge to zero at a rate which is slower than under independence, as the variance cannot be smaller than the one of the cost of the average assignment. Still, the convergence to zero of the variance means that the maximum possesses a property known as superconcentration. Finally, the dimension of the set of near-optimal assignments is shown to converge to zero.

**Keywords:** Extremal field, Gaussian random field, near maximal set, random assignment, superconcentration

---

## 1. Introduction: the Gaussian random assignment field

Optimal assignment is a classical problem appearing in mathematics and computer sciences that has also attracted probabilists' attention. Consider the problem of assigning  $n$  tasks to  $n$  workers such that each task is assigned to one worker and each worker is given a single task. The cost of matching task  $i$  to worker  $j$  is  $c(i, j)$ . An assignment is represented by a permutation  $u$  of  $[n] = \{1, \dots, n\}$ , where  $u(i) = j$  means that task  $i$  is scheduled for worker  $j$ . Let  $S_n$  be the symmetric group of permutations of  $[n]$ . The question is which assignment  $u \in S_n$  has lowest total cost  $\sum_{i \in [n]} c(i, u(i))$ . In the random assignment problem, the costs  $c(i, j)$  are assumed to be drawn from some probability distribution. For space considerations, we refer to Aldous & Steele (2004), Krokmal & Pardalos (2009) or Chatterjee (2019, Section 2.4) for the rich history of the problem, some of the numerous applications and more comprehensive literature reviews.

One of the most important results for the random assignment problem was obtained nearly two decades ago by Aldous (2001). He showed that the expected minimal cost for independent unit exponential entries  $c(i, j)$  converges to  $\zeta(2) = \pi^2/6$  as  $n \rightarrow \infty$ , where  $\zeta$  denotes the Riemann zeta function. This result confirmed a longstanding conjecture by Mézard & Parisi (1987) that could be numerically verified for large cost matrices by Pardalos & Ramakrishnan (1993) thanks to refined interior point methods. Aldous also proved an asymptotical essential uniqueness property, stating that every near-optimal assignment coincides with the optimal one except for a small proportion of pairs  $(i, j)$ . Still for independent unit exponential costs, Wästlund (2005) showed that the variance of the minimal cost is asymptotically equivalent to  $4(\zeta(2) - \zeta(3))/n$  as  $n \rightarrow \infty$ . For uniformly distributed costs, Talagrand (1995) obtained an upper bound on the fluctuations of the minimal cost, whereas Chatterjee (2019) obtained a lower bound for such fluctuations valid for a whole class of cost distributions on the positive half-line. The two-way assignment problem we consider here is also qualified as linear, as opposed to multi-dimensional ones studied in Krokmal et al. (2007). In the quadratic version, for instance, the objective is to find an optimal assignment of  $n$  jobs to  $n$  workers on  $n$  machines.

We consider the (linear) random assignment problem where the costs  $c(i, j)$  are independent and identically distributed Gaussian random variables. Up to the best of our knowledge, we are the first to do so. For convenience, we

---

\*Corresponding author

Email addresses: gilles.mordant@uclouvain.be (Gilles Mordant), johan.segers@uclouvain.be (Johan Segers)

standardize the problem in such a way that the cost of each assignment is standard Gaussian. The collection of costs of all possible assignments is a Gaussian random vector indexed by  $\mathcal{S}_n$ . We use the term Gaussian field because of the metric it induces on  $\mathcal{S}_n$ . This motivates the following definition, which introduces the main object of this paper.

**Definition 1.1.** Let  $n$  be a positive integer and let  $(c(i, j) : i, j \in [n])$  be a random matrix of independent standard Gaussian random variables. The Gaussian random assignment field  $g_n = (g_{n,u} : u \in \mathcal{S}_n)$  is defined by

$$g_{n,u} = \frac{1}{\sqrt{n}} \sum_{i=1}^n c(i, u(i)).$$

The normalization by  $1/\sqrt{n}$  in the definition ensures that each variable  $g_{n,u}$  is standard Gaussian. Being a linear transformation of a vector of independent Gaussian random variables,  $g_n$  has a multivariate Gaussian distribution, i.e.,  $g_n$  is a Gaussian random vector. The correlation matrix  $R_n = (r_n(u, v))_{u, v \in \mathcal{S}_n}$  is easily seen to be

$$r_n(u, v) = \mathbb{E}(g_{n,u} g_{n,v}) = \frac{1}{n} |\{i \in [n] : u(i) = v(i)\}|, \quad u, v \in \mathcal{S}_n. \quad (1)$$

The Hamming distance between two permutations  $u, v \in \mathcal{S}_n$  is  $d_H(u, v) = |\{i \in [n] : u(i) \neq v(i)\}|$ . The  $L_2$  metric induced by  $g_n$  on  $\mathcal{S}_n$  is related to the Hamming distance via

$$\sqrt{\mathbb{E}((g_{n,u} - g_{n,v})^2)} = \sqrt{2(1 - r_n(u, v))} = \sqrt{2d_H(u, v)/n}. \quad (2)$$

Two permutations  $u$  and  $v$  are far apart if the variables  $g_{n,u}$  and  $g_{n,v}$  are nearly orthogonal, which happens if  $u(i)$  and  $v(i)$  are different for most  $i \in [n]$ . This metric structure motivates the use of the term random ‘field’.

Already since Pierre Rémond de Montmort’s famous 1713 treatise *Essay d’analyse sur les jeux de hazard*, it is known that, for any permutation  $u \in \mathcal{S}_n$ , the proportion of permutations  $v \in \mathcal{S}_n$  such that the set  $\{i \in [n] : u(i) = v(i)\}$  of fixed points with  $u$  has size  $k \in \{0, 1, \dots, n\}$  is equal to

$$\frac{1}{k!} \sum_{\ell=0}^{n-k} \frac{(-1)^\ell}{\ell!}, \quad (3)$$

a number which quickly converges to  $e^{-1}/k!$  as  $n \rightarrow \infty$ . It follows that for large  $n$ , most assignments  $g_{n,u}$  and  $g_{n,v}$  are nearly uncorrelated—the probability that a unit-Poisson random variable is not larger than 4 is higher than 99%. Also, the average over all  $(n!)^2$  correlations  $r_n(u, v)$  is equal to  $1/n$ ; see the proof of Proposition 2.2 below.

Still, the Gaussian random assignment field  $g_n$  comprises  $n!$  dependent Gaussian variables generated out of only  $n^2$  independent ones. There are thus two conflicting intuitions regarding the dependence within  $g_n$ . On the one hand, it is a high-dimensional random vector supported by a relatively low-dimensional subspace. On the other hand, two arbitrarily chosen components are nearly uncorrelated. As  $n$  grows, each of both effects becomes stronger. The question is then which of the two tendencies dominates. The answer may depend on the functional of  $g_n$  of interest.

In the context of random assignment problems it is natural to study the lowest and highest achievable costs, i.e., the minimum and maximum over all assignments:

$$W_n = \min_{u \in \mathcal{S}_n} g_{n,u} \quad \text{and} \quad M_n = \max_{u \in \mathcal{S}_n} g_{n,u}. \quad (4)$$

By symmetry,  $W_n$  has the same distribution as  $-M_n$ . We therefore only study the maximum  $M_n$ , as is common in the theory of Gaussian fields. Rather than as a cost, we could view  $c(i, j)$  as the gain (when positive) or loss (when negative) of assigning ‘player’  $i$  to ‘game’  $j$ . The highest possible gain over all assignments is then equal to  $M_n$ .

Our first objective is to study the asymptotic behavior of the expectation and the variance of  $M_n$ . Theorem 2.3, our first main result, states that  $\mathbb{E}(M_n)$  is asymptotically equivalent to  $\sqrt{2 \log(n!)}$ . This rate is the same as the one for the maximum of  $n!$  independent standard Gaussian variables, and is in fact the fastest rate possible for a centered Gaussian field with unit variances (Chatterjee, 2014, Eq. (A.3)). In the terminology of Chatterjee (2014, Chapter 8), this means that the Gaussian random assignment field  $g_n$  is *extremal*. Moreover, by Corollary 2.4,  $\text{Var}(M_n)$  converges to zero as  $n \rightarrow \infty$ . Since each individual variable  $g_{n,u}$  has unit variance, this means that, according to the same

reference,  $M_n$  is *superconcentrated*. We refer to Chatterjee (2014) for an in-depth treatment of the implications of the superconcentration phenomenon. According to Theorem 1.18 in that book, for instance, with high probability there are a large number of assignments  $u \in \mathcal{S}_n$  that are mutually distant from each other according to the metric (2) but which are near-maximal, i.e., the relative difference between  $g_{n,u}$  and  $M_n$  is small. Tanguy (2015) illustrates the superconcentration phenomenon for extrema of stationary Gaussian processes and some other Gaussian models.

Still, the lower bound  $\text{Var}(M_n) \geq 1/n$  that we derive in Proposition 2.2 implies that the rate at which  $\text{Var}(M_n)$  converges to zero is slower than the one of the maximum of  $n!$  independent standard Gaussian variables, which is  $1/\log(n!) \sim 1/(n \log n)$  instead. The lower bound is also sharper than the bound  $\text{Var}(M_n) \geq c/\log(n!)$ , with  $c$  a universal positive constant, that follows from Ding et al. (2015, Corollary 1.9). The tool to find our variance lower bound is elementary and works by ‘sourcing out’ the average cost  $\bar{g}_n = (n!)^{-1} \sum_{u \in \mathcal{S}_n} g_{n,u}$ .

Our second topic of interest is the number of assignments which are nearly optimal. The following definition comes from (Chatterjee, 2014, Section 8.3).

**Definition 1.2.** *The near maximal set of a Gaussian field  $g = (g_1, \dots, g_N)$  with  $M = \max_{u \in [N]} g_u$  and  $m = \mathbb{E}(M)$  is*

$$A(\epsilon) = \{u \in [N] : g_u > (1 - \epsilon)m\}, \quad \epsilon > 0.$$

The ratio  $\log |A(\epsilon)| / \log N$  can be thought of as the (fractal) dimension of  $A(\epsilon)$ . Let  $A_n(\epsilon)$  be the near-maximal set of the Gaussian random assignment field  $g_n$ . According to our second main result, Theorem 2.5, the expectation of  $\log |A_n(\epsilon)|$ , provided  $A_n(\epsilon)$  is non-empty, is of smaller order than  $\log(n!)$ . In particular, the expected dimension of  $A_n(\epsilon)$  when non-empty converges to zero.

The main results are collected in Section 2 and all proofs in Section 3. A discussion in Section 4 of open questions concludes the paper.

## 2. Main results

Some of our bounds on the expectation and the variance of the maximum  $M_n$  of the Gaussian random assignment field  $g_n$  rely on a simple principle. In a Gaussian random vector of which the covariance matrix has constant row sums, it is possible to write all variables as the sum of a common component and an independent Gaussian random vector. ‘Outsourcing’ this common component produces bounds on the expectation and the variance of the maximum.

**Lemma 2.1.** *Let  $g = (g_1, \dots, g_N)$  be a centered Gaussian random field with  $\text{Var}(g_u) = 1$  for all  $u \in [N]$ . Put  $M = \max_{u \in [N]} g_u$ . Suppose the row means  $N^{-1} \sum_{v \in [N]} \text{Cov}(g_u, g_v)$  of the covariance matrix are the same for all  $u \in [N]$  and let  $s^2$  denote their common value. Then  $0 \leq s^2 \leq 1$  and*

$$\begin{aligned} \mathbb{E}(M) &\leq \sqrt{2(1 - s^2)N}, \\ \text{Var}(M) &\geq s^2. \end{aligned}$$

**Proposition 2.2.** *For all integer  $n \geq 1$ , the maximum  $M_n$  in (4) of the Gaussian random assignment field  $g_n$  in Definition 1.1 satisfies*

$$\begin{aligned} \mathbb{E}(M_n) &\leq \sqrt{2(1 - 1/n) \log(n!)}, \\ \text{Var}(M_n) &\geq 1/n. \end{aligned}$$

Since  $g_n$  is a centered Gaussian field with unit variances, Eq. (A.3) in Chatterjee (2014) implies that  $\mathbb{E}(M_n) \leq \sqrt{2 \log(n!)}$ . Proposition 2.2 thus constitutes a modest improvement. In the introduction, we already noted that the lower bound for the variance improves the inequality  $\text{Var}(M_n) \geq c/(n \log n)$  that follows from Corollary 1.9 in Ding et al. (2015).

The question is at what rate  $\mathbb{E}(M_n)$  goes to infinity. On the one hand, Proposition 2.2 implies that it cannot do so faster than  $\sqrt{2 \log(n!)}$ . On the other hand, Theorem 8.3 in Chatterjee (2014) yields the existence of a universal positive constant  $C$  such that

$$\mathbb{E}(M_n) \geq \sqrt{2 \log(n!)} - C \left( \log \log(n!) + \log \left( \sum_{u,v \in \mathcal{S}_n} (n!)^{-2/(1+r_n(u,v))} \right) \right)^{1/2}$$

with  $r_n(u, v)$  the pairwise correlation in Eq. (1). A careful analysis of the last sum in the lower bound shows that  $\liminf_{n \rightarrow \infty} \mathbb{E}(M_n) / \sqrt{2 \log(n!)} > 0$ . Our first main result, Theorem 2.3 states that the limit exists and is equal to one.

**Theorem 2.3.** *Let  $M_n$  be the maximum of the Gaussian random assignment field  $g_n$ . We have*

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}(M_n)}{\sqrt{2 \log(n!)}} = 1.$$

Theorem 2.3 in combination with (Chatterjee, 2014, Theorem 8.1) implies that  $\text{Var}(M_n)$  tends to zero and thus that  $g_n$  is *superconcentrated*, as discussed in the introduction.

**Corollary 2.4.** *For  $M_n$  and  $g_n$  as in Theorem 2.3, we have*

$$\lim_{n \rightarrow \infty} \text{Var}(M_n) = 0.$$

Next we study the expected size of the near maximal set  $A_n(\epsilon)$  in Definition 1.2 of the Gaussian random assignment field  $g_n$ . Note that the total number of variables is  $|\mathcal{S}_n| = n!$  and that  $\log(n!)$  is asymptotically equivalent to  $n \log(n)$  as  $n \rightarrow \infty$ . According to our second main result, Theorem 2.5, the expected size of  $A_n(\epsilon)$  is substantially smaller. Since  $A_n(\epsilon)$  can be empty with positive probability, we exclude this case from the expectation of its logarithm and we write  $\mathbb{E}[X; A] = \mathbb{E}[X \mathbb{1}_A]$  for a random variable  $X$  and an event  $A$  with indicator  $\mathbb{1}_A$ .

**Theorem 2.5.** *Let  $g_n$  be the Gaussian random assignment field and let  $A_n(\epsilon)$  be its near maximal set. There exist universal constants  $C'$  and  $C''$  such that*

$$\mathbb{E}(\log |A_n(\epsilon)|; A_n(\epsilon) \neq \emptyset) \leq \begin{cases} C' (n \log n)^{3/4} & \text{if } 0 < \epsilon \leq (2n \log n)^{-1/2}, \\ C'' \sqrt{\epsilon} (n \log n) & \text{if } (2n \log n)^{-1/2} < \epsilon < 1. \end{cases}$$

It follows from Theorem 2.5 that the expected dimension  $\log |A_n(\epsilon)| / \log(n!)$  of  $A_n(\epsilon)$  when non-empty converges to zero for  $\epsilon = \epsilon_n \rightarrow 0$ .

### 3. Proofs

**Lemma 3.1.** *Let  $g = (g_1, \dots, g_N)$  be a centred  $N$ -dimensional Gaussian random vector with covariance matrix  $\Gamma = (\gamma_{uv})_{u,v \in [N]}$  such that the row sums  $\sum_{v \in [N]} \gamma_{uv}$  do not depend on  $u \in [N]$ . Let  $s^2 = N^{-1} \sum_{v \in [N]} \gamma_{vv}$  denote the common value of the row means. Then  $0 \leq s^2 \leq \min_{u \in [N]} \gamma_{uu}$  and we have the decomposition*

$$g_u = \bar{g} + h_u, \quad u \in [N], \tag{5}$$

where  $\bar{g} = N^{-1} \sum_{u \in [N]} g_u$  is a centred Gaussian random variable with variance  $s^2$  and  $h = (h_1, \dots, h_N)$  is a centred Gaussian random vector with  $\text{Var}(h_u) = \gamma_{uu} - s^2$ , with  $\bar{g}$  and  $h$  being independent.

*Proof.* The covariance between each  $g_u$  and the arithmetic mean  $\bar{g}$  is

$$\text{Cov}(g_u, \bar{g}) = \frac{1}{N} \sum_{v \in [N]} \text{Cov}(g_u, g_v) = s^2, \quad u \in [N].$$

This common covariance is also equal to the variance of  $\bar{g}$ , since

$$\text{Var}(\bar{g}) = \frac{1}{N} \sum_{u \in [N]} \text{Cov}(g_u, \bar{g}) = s^2.$$

As a consequence,  $s^2$  is non-negative and the variables  $\bar{g}$  and  $g_u - \bar{g}$  are uncorrelated:

$$\text{Cov}(\bar{g}, g_u - \bar{g}) = s^2 - s^2 = 0.$$

It follows that for all  $u \in [N]$  we have

$$\gamma_{uu} = \text{Var}(g_u) = \text{Var}(\bar{g}) + \text{Var}(g_u - \bar{g}) = s^2 + \text{Var}(g_u - \bar{g})$$

and thus

$$\gamma_{uu} - s^2 = \text{Var}(g_u - \bar{g}) \geq 0.$$

Define  $h = (h_1, \dots, h_N)$  by  $h_u = g_u - \bar{g}$  for  $u \in [N]$ . Since the  $(N+1)$ -dimensional random vector  $(\bar{g}, h)$  is jointly Gaussian, the fact that  $\bar{g}$  is uncorrelated with every  $h_u$  implies that  $\bar{g}$  and  $h$  are independent.  $\square$

*Proof of Lemma 2.1.* We apply Lemma 3.1. Since all variables  $g_u$  have unit variance, it follows that  $0 \leq s^2 \leq 1$ . Let  $\bar{g}$  and  $h$  be as in Lemma 3.1. In view of the decomposition (5), the maximum  $M$  over the Gaussian field  $g$  is equal to

$$M = \bar{g} + L \quad \text{where} \quad L = \max_{u \in [n]} h_u.$$

The variables  $\bar{g}$  and  $L$  are independent, whence

$$\text{Var}(M) = \text{Var}(\bar{g}) + \text{Var}(L) \geq \text{Var}(\bar{g}) = s^2.$$

Furthermore,

$$\mathbb{E}(M) = \mathbb{E}(\bar{g}) + \mathbb{E}(L) = \mathbb{E}(L).$$

If  $s^2 = 1$ , then  $h$  is degenerate at the origin and  $L = 0$  almost surely. If  $s^2 < 1$ , then we can write  $h = \sqrt{1-s^2}\tilde{h}$  where  $\tilde{h} = (\tilde{h}_1, \dots, \tilde{h}_N)$  is a centred Gaussian field with  $\text{Var}(\tilde{h}_u) = 1$  for all  $u \in [N]$ . We have  $L = \sqrt{1-s^2}\tilde{L}$  with  $\tilde{L} = \max_{u \in [N]} \tilde{h}_u$ . By Eq. (A.3) in Chatterjee (2014), we have  $\mathbb{E}(\tilde{L}) \leq \sqrt{2 \log(N)}$ . We conclude that

$$\mathbb{E}(M) = \sqrt{1-s^2} \mathbb{E}(\tilde{L}) \leq \sqrt{2(1-s^2) \log(N)}. \quad \square$$

*Proof of Proposition 2.2.* We apply Lemma 2.1. By permutation symmetry, the covariance matrix of  $g_n$  has constant row sums: for  $u, v, w \in \mathcal{S}_n$ , we have  $\text{Cov}(g_{u,n}, g_{v,n}) = \text{Cov}(g_{w \circ u, n}, g_{w \circ v, n})$ , and thus, for  $u_1, u_2 \in \mathcal{S}_n$ ,

$$\sum_{v \in \mathcal{S}_n} \text{Cov}(g_{u_1, n}, g_{v, n}) = \sum_{v \in \mathcal{S}_n} \text{Cov}(g_{u_2, n}, g_{u_2 \circ u_1^{-1} \circ v, n}) = \sum_{v' \in \mathcal{S}_n} \text{Cov}(g_{u_2, n}, g_{v', n}).$$

The common value  $s^2$  of the row means of the covariance matrix of  $g_n$  is equal to the variance of the average  $\bar{g}_n = (n!)^{-1} \sum_{u \in \mathcal{S}_n} g_u$  over all assignments. But the latter is proportional to the average over all individual costs:

$$\begin{aligned} \bar{g}_n &= \frac{1}{n!} \sum_{u \in \mathcal{S}_n} \frac{1}{\sqrt{n}} \sum_{i \in [n]} c(i, u(i)) \\ &= \frac{1}{n! \sqrt{n}} \sum_{i \in [n]} \sum_{u \in \mathcal{S}_n} c(i, u(i)) \\ &= \frac{1}{n! \sqrt{n}} \sum_{i \in [n]} \sum_{j \in [n]} c(i, j) \underbrace{|\{u \in \mathcal{S}_n : u(i) = j\}|}_{=(n-1)!} \\ &= \frac{1}{n \sqrt{n}} \sum_{i \in [n]} \sum_{j \in [n]} c(i, j). \end{aligned}$$

It follows that the common value of the row means of the covariance matrix of  $g_n$  is

$$s^2 = \text{Var}(\bar{g}_n) = \frac{1}{n^3} \cdot n^2 \cdot 1 = \frac{1}{n}.$$

Lemma 2.1 yields the inequalities  $\mathbb{E}(M_n) \leq \sqrt{2(1-n^{-1}) \log(n!)}$  and  $\text{Var}(M_n) \geq n^{-1}$  for all positive integer  $n$ .  $\square$

*Proof of Theorem 2.3.* In view of Theorem 2.2, it suffices to show that

$$\liminf_{n \rightarrow \infty} \frac{\mathbb{E}(M_n)}{\sqrt{2 \log(n!)}} \geq 1. \quad (6)$$

To do so, we construct a lower bound to  $M_n$  by a greedy approach. Define a random permutation  $u_c \in \mathcal{S}_n$  recursively by passing through the rows  $i$  of the cost matrix  $c$  one after the other, each time discarding the columns  $j$  that have already been selected in the previous steps:

- Row  $i = 1$ : let  $u_c(1)$  be the index  $j \in [n]$  that maximizes  $c(1, j)$ .
- Row  $i = 2$ : let  $u_c(2)$  be the index  $j \in [n] \setminus \{u_c(1)\}$  that maximizes  $c(2, j)$ .
- ...
- Row  $i = n$ : let  $u_c(n)$  be the only remaining element in  $[n] \setminus \{u_c(1), \dots, u_c(n-1)\}$ .

Formally, define  $u_c \in \mathcal{S}_n$  by

$$\begin{aligned} u_c(1) &= \arg \max_{j \in [n]} c(1, j), \\ u_c(i) &= \arg \max_{j \in [n] \setminus \{u_c(1), \dots, u_c(i-1)\}} c(i, j), \quad i \in \{2, \dots, n\}. \end{aligned}$$

In case of ties, always choose the smallest index available; another tie-splitting rule would work too. Clearly,

$$\begin{aligned} c(1, u_c(1)) &= \max_{j \in [n]} c(1, j), \\ c(i, u_c(i)) &= \max_{j \in [n] \setminus \{u_c(1), \dots, u_c(i-1)\}} c(i, j), \quad i \in \{2, \dots, n\}. \end{aligned}$$

Then  $c(1, u_c(1))$  is the maximum of  $n$  independent standard Gaussian variables;  $c(2, u_c(2))$  is the maximum of  $(n-1)$  independent standard Gaussian variables; etc. Indeed, let  $i \geq 2$ . Since the indices  $u_c(1), \dots, u_c(i-1)$  are a function of the first  $(i-1)$  rows of the cost matrix  $c$ , the conditional distribution of the maximum  $c(i, u_c(i))$  over the remaining  $n-i+1$  variables  $c(i, j)$  with indices  $j$  in  $[n] \setminus \{u_c(1), \dots, u_c(i-1)\}$  in the  $i$ th row given the earlier chosen indices  $u_c(1), \dots, u_c(i-1)$  is independent of the values of those indices.

The random assignment  $u_c$  constructed by this greedy algorithm yields a lower bound to the maximum:

$$M_n \geq \frac{1}{\sqrt{n}} \sum_{i=1}^n c(i, u_c(i)).$$

Let  $\mu_n$  be the expectation of the maximum of  $n$  independent standard Gaussian variables, i.e.,  $\mu_n = \mathbb{E}[\max_{j \in [n]} c(1, j)]$ . It follows that

$$\mathbb{E}(M_n) \geq \frac{1}{\sqrt{n}} \sum_{i \in [n]} \mathbb{E}[c(i, u_c(i))] = \frac{1}{\sqrt{n}} \sum_{i \in [n]} \mu_{n-i+1}.$$

From classical extreme value theory, it is known that<sup>1</sup>

$$\lim_{n \rightarrow \infty} \frac{\mu_n}{\sqrt{2 \log n}} = 1. \quad (7)$$

Let  $\delta \in (0, 1)$ . Since  $\mu_n$  is non-decreasing in  $n$ , we find

$$\mathbb{E}(M_n) \geq \frac{1}{\sqrt{n}} \sum_{i \in [n]} \mu_i \geq \frac{1}{\sqrt{n}} \sum_{i \in [n], i > \delta n} \mu_i \geq \frac{1}{\sqrt{n}} (n - \lceil \delta n \rceil + 1) \mu_{\lceil \delta n \rceil},$$

<sup>1</sup>In fact, more accurate expansions are available. In Resnick (1987), for instance, combine Example 2 on page 71 with Proposition 2.1(iii) on page 77 and Exercise 2.1.2 on page 84.

where  $\lceil x \rceil$  is the smallest integer not less than the scalar  $x$ . By (7), we have

$$\frac{\mu_{\lceil \delta n \rceil}}{\sqrt{2 \log n}} = \frac{\mu_{\lceil \delta n \rceil}}{\sqrt{2 \log \lceil \delta n \rceil}} \sqrt{\frac{\log \lceil \delta n \rceil}{\log n}} \rightarrow 1, \quad n \rightarrow \infty.$$

We arrive at

$$\frac{\mathbb{E}(M_n)}{\sqrt{2n \log n}} \geq \frac{n - \lceil \delta n \rceil + 1}{n} \frac{\mu_{\lceil \delta n \rceil}}{\sqrt{2 \log n}} \rightarrow 1 - \delta, \quad n \rightarrow \infty.$$

As  $\delta \in (0, 1)$  was arbitrary, we find

$$\liminf_{n \rightarrow \infty} \frac{\mathbb{E}(M_n)}{\sqrt{2n \log n}} \geq 1.$$

Since  $\log(n!)/(n \log n) \rightarrow 1$  as  $n \rightarrow \infty$ , the claim (6) follows.  $\square$

*Proof of Corollary 2.4.* Combine Theorem 2.3 with Theorem 8.1 in Chatterjee (2014), stating an upper bound for  $\text{Var}(M_n)$  which tends to zero since  $\mathbb{E}(M_n)/\sqrt{2 \log(n!)}$  tends to one.  $\square$

As a preparation to the proof of Theorem 2.5, we state Theorem 12.4 in Chatterjee (2014). The theorem gives a bound on the size of the near maximal set  $A(\epsilon)$  in Definition 1.2 in terms of two ingredients: the expectation  $m$  of the maximum and the number  $V(\delta)$  of variables that have a correlation larger than  $1 - \delta$  with a given variable.

**Theorem 3.2** (Chatterjee). *Let  $g = (g_1, \dots, g_N)$  be a centered Gaussian random vector such that  $\text{Var}(g_u) = 1$  for all  $u \in [N]$ . Let  $M$ ,  $m$  and  $A(\epsilon)$  be as in Definition 1.2. For  $u \in [N]$  and  $\delta \in (0, 1)$ , define*

$$B(u, \delta) = \{v \in [N] : \text{Cov}(g_u, g_v) > 1 - \delta\} \quad \text{and} \quad V(\delta) = \max_{u \in [N]} |B(u, \delta)|$$

*Then there exists a universal constant  $C$  such that for any  $\epsilon \in (0, 1)$  we have the bound*

$$\mathbb{E}(\log |A(\epsilon)|; A(\epsilon) \neq \emptyset) \leq \inf_{\delta \in (0, 1)} \left( \log V(\delta) + \frac{C \max\{\epsilon m^2, m\}}{\delta} \right).$$

*Proof of Theorem 2.5.* We apply Theorem 12.4 in Chatterjee (2014), stated for the reader's convenience as Theorem 3.2 above, to the Gaussian random assignment field  $g_n$ . Recall the discussion around Eq. (3). For  $\delta \in (0, 1)$ , the size of the set

$$B_n(u, \delta) = \{v \in \mathcal{S}_n : \text{Cov}(g_u, g_v) \geq 1 - \delta\} = \{v \in \mathcal{S}_n : |\{i \in [n] : u(i) = v(i)\}| > (1 - \delta)n\}$$

does not depend on  $u \in \mathcal{S}_n$  and is equal to

$$V_n(\delta) = |B_n(u, \delta)| = n! \sum_{k \in [n], k > (1 - \delta)n} \frac{1}{k!} \sum_{\ell=0}^{n-k} \frac{(-1)^\ell}{\ell!}.$$

Let  $\lceil x \rceil$  denote the smallest integer not smaller than the scalar  $x$ . Then  $V_n(\delta)$  is bounded by

$$V_n(\delta) \leq \frac{n!}{\lceil (1 - \delta)n \rceil!} \sum_{k \in [n]} \frac{1}{k!} \sum_{\ell=0}^{n-k} \frac{(-1)^\ell}{\ell!} \leq \frac{n!}{\lceil (1 - \delta)n \rceil!} \leq n^{\delta n}.$$

Indeed, the third inequality follows from the fact that the number of integers  $k$  such that  $\lceil (1 - \delta)n \rceil < k \leq n$  is bounded by  $n - (1 - \delta)n = \delta n$ , while the second inequality follows from

$$\begin{aligned} \sum_{k \in [n]} \frac{1}{k!} \sum_{\ell=0}^{n-k} \frac{(-1)^\ell}{\ell!} &= \sum_{k \in [n]} \sum_{\ell=0}^{n-k} \frac{1}{(k + \ell)!} \binom{k + \ell}{\ell} (-1)^\ell \\ &= \sum_{s \in [n]} \frac{1}{s!} \sum_{\ell=0}^{s-1} \binom{s}{\ell} (-1)^\ell \\ &= \sum_{s \in [n]} \frac{1}{s!} ((1 - 1)^s - (-1)^s) = \sum_{s \in [n]} \frac{1}{s!} (-1)^{s-1} \leq 1, \end{aligned}$$

in view of the binomial theorem. By Theorem 3.2 and the bound on  $V_n(\delta)$ , it follows that

$$\mathbb{E}(\log A_n(\epsilon); A_n(\epsilon) \neq \emptyset) \leq \inf_{\delta \in (0,1)} \left( \delta n \log(n) + \frac{C \max\{\epsilon m_n^2, m_n\}}{\delta} \right)$$

where  $m_n = \mathbb{E}(M_n)$ . Since  $m_n \leq \sqrt{2 \log(n!)} \leq \sqrt{2n \log n}$  by Theorem 2.2, we obtain

$$\mathbb{E}(\log A_n(\epsilon); A_n(\epsilon) \neq \emptyset) \leq \inf_{\delta \in (0,1)} \left( \delta n \log(n) + \frac{C \max\{2\epsilon n \log n, \sqrt{2n \log n}\}}{\delta} \right).$$

It remains to distinguish between the following two cases:

- (i) If  $0 < \epsilon \leq 1/\sqrt{2n \log n}$ , then put  $\delta = (n \log n)^{-1/4}$ . Both terms in the sum are of the order  $(n \log n)^{3/4}$ .
- (ii) If  $1/\sqrt{2n \log n} < \epsilon < 1$ , then put  $\delta = \sqrt{\epsilon}$ . Both terms in the sum are of the order  $\sqrt{\epsilon} n \log n$ .

The claim follows. □

#### 4. Discussion

Proposition 2.2 in combination with Corollary 2.4 yields  $1/n \leq \text{Var}(M_n) \rightarrow 0$  as  $n \rightarrow \infty$ . The correct rate at which  $\text{Var}(M_n)$  converges to zero remains an open problem. For unit exponential costs, Wästlund (2005) showed that the variance of the minimal costs is asymptotically equivalent to a constant times  $1/n$  as  $n \rightarrow \infty$ . It is however unclear if this also occurs for the Gaussian random assignment field.

In Lemma 2.1, we have shown that the variance of the maximum of a Gaussian array could be bounded from below by the variance of the sample mean under the fixed-row-sum assumption on the covariance matrix. What are (easily verifiable) necessary and sufficient conditions on the covariance matrix for this lower bound to provide the correct order of the variance of the maximum?

Finally, limit theorems for the maxima of Gaussian random variables are well-covered in the literature—see for instance Leadbetter et al. (1983) and the references therein—in particular in the weakly dependent case and for stationary sequences. In the latter case, assuming some regularity on the correlation function, Mittal & Ylvisaker (1975) proved that Gumbel limits, mixtures of Gumbel and Gaussian distributions and Gaussian distributions could arise depending on the speed at which correlations tend to zero. The maximum of the Gaussian random assignment field possesses characteristics both of a maximum (over all possible assignments) and of a sum (over all jobs). How does the limit distribution look like? Although the setting is different, the answer could give more insight into the limit distribution of the empirical Wasserstein distance (Panaretos & Zemel, 2019), a much coveted result, which corresponds to a matching problem with a matrix of random but dependent costs.

#### Acknowledgments

The first author thanks Sourav Chatterjee for some insightful remarks. The comments by Stefka Kirilova and Vincent Plassier are also gratefully acknowledged. The authors would also like to thank the Associate Editor and an anonymous Referee for comments that helped improve the presentation of the paper and for pointing out particularly relevant references. The second author gratefully acknowledges funding by FNRS-F.R.S. grant CDR J.0146.19.

#### References

- Aldous, D. (2001). The  $\zeta(2)$  limit in the random assignment problem. *Random Structures & Algorithms*, 18, 381–418.
- Aldous, D., & Steele, M. J. (2004). The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures* (pp. 1–72). Springer.
- Chatterjee, S. (2014). *Superconcentration and Related Topics*. Springer Monographs in Mathematics. Cham: Springer.
- Chatterjee, S. (2019). A general method for lower bounds on fluctuations of random variables. *The Annals of Probability*, 47, 2140–2171.
- Ding, J., Eldan, R., Zhai, A. et al. (2015). On multiple peaks and moderate deviations for the supremum of a Gaussian field. *The Annals of Probability*, 43, 3468–3493.

- Krokhmal, P. A., Grundel, D. A., & Pardalos, P. M. (2007). Asymptotic behavior of the expected optimal value of the multidimensional assignment problem. *Mathematical Programming*, 109, 525–551.
- Krokhmal, P. A., & Pardalos, P. M. (2009). Random assignment problems. *European Journal of Operational Research*, 194, 1–17.
- Leadbetter, M. R., Lindgren, G., & Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. New York: Springer-Verlag.
- Mézard, M., & Parisi, G. (1987). On the solution of the random link matching problems. *Journal de Physique*, 48, 1451–1459.
- Mittal, Y., & Ylvisaker, D. (1975). Limit distributions for the maxima of stationary Gaussian processes. *Stochastic Processes and their Applications*, 3, 1–18.
- Panaretos, V. M., & Zemel, Y. (2019). Statistical aspects of Wasserstein distances. *Annual Review of Statistics and its Application*, 6, 405–431.
- Pardalos, P. M., & Ramakrishnan, K. (1993). On the expected optimal value of random assignment problems: Experimental results and open questions. *Computational Optimization and Applications*, 2, 261–271.
- Resnick, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes* volume 4 of *Applied Probability. A Series of the Applied Probability Trust*. New York: Springer-Verlag.
- Talagrand, M. (1995). Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, 81, 73–205.
- Tanguy, K. (2015). Some superconcentration inequalities for extrema of stationary Gaussian processes. *Statistics & Probability Letters*, 106, 239–246.
- Wästlund, J. (2005). *The variance and higher moments in the random assignment problem*. Linköping University Electronic Press.