

# Anomalous scaling of the optimal cost in the one-dimensional random assignment problem

**Sergio Caracciolo**

E-mail: [sergio.caracciolo@mi.infn.it](mailto:sergio.caracciolo@mi.infn.it)

Dipartimento di Fisica, University of Milan and INFN, via Celoria 16, I-20133, Milan, Italy

**Matteo D'Achille**

E-mail: [matteo.d-achille@college-de-france.fr](mailto:matteo.d-achille@college-de-france.fr)

Basic Research Division, Centre CEA de Saclay, D36 91191 Gif-sur-Yvette cedex, France

CIRB - Center for Interdisciplinary Research in Biology, Collège de France, 11 Place Marcelin Berthelot, 75231 Paris, France

LI-PaRAD - Laboratoire d'Informatique Parallélisme Réseaux Algorithmes Distribués, 45 avenue des Etats-Unis, 78035 Versailles, France Université de Versailles Saint-Quentin-en-Yvelines - Université Paris Saclay

**Gabriele Sicuro**

E-mail: [gabriele.sicuro@roma1.infn.it](mailto:gabriele.sicuro@roma1.infn.it)

Dipartimento di Fisica, Sapienza Università di Roma, P.le A. Moro 2, I-00185, Rome, Italy

**Abstract.** We consider the random Euclidean assignment problem on the line between two sets of  $N$  random points, independently generated with the same probability density function  $\varrho$ . The cost of the matching is supposed to be dependent on a power  $p > 1$  of the Euclidean distance of the matched pairs. We discuss an integral expression for the average optimal cost for  $N \gg 1$  that generalizes a previous result obtained for  $p = 2$ . We also study the possible divergence of the given expression due to the vanishing of the probability density function. The provided regularization recipe allows us to recover the proper scaling law for the cost in the divergent cases, and possibly some of the involved coefficients. The possibility that the support of  $\varrho$  is a disconnected interval is also analysed. We exemplify the proposed procedure and we compare our predictions with the results of numerical simulations.

## 1. Introduction

Random Euclidean assignment problems (REAPs) are combinatorial optimization problems in which we couple points randomly generated on a given  $d$ -dimensional domain  $\Omega \subseteq \mathbb{R}^d$  in such a way that a certain cost function is minimized. To be more precise, let us call  $\Xi := \{\mathbf{x}_i\}_{i=1,\dots,N} \subset \Omega$  and  $\Upsilon := \{\mathbf{y}_j\}_{j=1,\dots,N} \subset \Omega$  two sets of points, both of cardinality  $N$ , in the  $d$ -dimensional domain  $\Omega$ . All points are supposed to be identically and independently distributed over  $\Omega$  according to some probability

density function (PDF)  $\varrho: \Omega \rightarrow \mathbb{R}^+$ . We also suppose that a certain weight function  $w: \Omega \times \Omega \rightarrow \mathbb{R}^+$  is given, in such a way that  $w(\mathbf{x}_i, \mathbf{y}_j) \equiv w(\|\mathbf{x}_i - \mathbf{y}_j\|)$  is the cost of the coupling  $(\mathbf{x}_i, \mathbf{y}_j)$ ,  $\|\mathbf{x}_i - \mathbf{y}_j\|$  being the Euclidean distance between  $\mathbf{x}_i$  and  $\mathbf{y}_j$ . Then, denoting by  $\mathcal{S}_N$  the set of permutations  $\sigma$  of  $N$  elements, we search for an *optimal assignment*  $\sigma_*$ , such that

$$\varepsilon[\Xi, \Upsilon] := \min_{\sigma \in \mathcal{S}_N} \frac{1}{N} \sum_{i=1}^N w(\mathbf{x}_i, \mathbf{y}_{\sigma(i)}) \equiv \frac{1}{N} \sum_{i=1}^N w(\mathbf{x}_i, \mathbf{y}_{\sigma_*(i)}). \quad (1)$$

Observe that an optimal assignment always exists but it is not unique in general. From the computational point of view, once the sets of points  $\Xi$  and  $\Upsilon$  and the weight function  $w$  are given, finding the minimum cost permutation is an “easy” problem, being the assignment problem in the P complexity class [1]. However, due to the randomness of the point positions, we are typically interested in the *average* properties of the solution, e.g., in the average optimal cost (AOC)

$$\varepsilon_N := \overline{\varepsilon[\Xi, \Upsilon]} = \min_{\sigma \in \mathcal{S}_N} \frac{1}{N} \sum_{i=1}^N w(\mathbf{x}_i, \mathbf{y}_{\sigma(i)}), \quad (2)$$

where we have denoted by  $\bullet$  the average over all possible realizations of the point sets  $\Xi$  and  $\Upsilon$  over  $\Omega$  according to the law  $\varrho$ .

In the available literature on the REAP, the domain  $\Omega$  is typically supposed to be *connected*, and the points to be *uniformly distributed* over it. Many results are available for this specific formulation of the problem, in particular assuming the weight function

$$w_p(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|^p, \quad p \in \mathbb{R}^+. \quad (3)$$

For example, in Ref. [2] the scaling of the AOC in the REAP with weight function  $w_p$  on the unit square has been derived, whereas in Ref. [3, 4] the coefficient to this scaling has been obtained for  $p = 2$  both on the unit square and on the torus, a result later rigorously proved in Ref. [5]. Higher dimensions have been considered in Refs. [3, 4], where the weight function  $w_2$  has been adopted, again assuming uniformly distributed random points on the  $d$ -dimensional hypercube. The problem greatly simplifies in the  $d \rightarrow +\infty$  limit, to be taken before the  $N \rightarrow +\infty$  limit, and indeed the infinite dimensional case has been the first to be considered in the physics literature. The simplification is due to the fact that, in this limit, the correlation among the weights  $w_p(\mathbf{x}_i, \mathbf{y}_j)$  disappears, and the average over the disorder can be easily tackled using the replica trick [6]. This fact has been first shown in a series of seminal works by Orland [7] and Mézard and Parisi [8, 9]. In their treatment of the problem, Mézard and Parisi considered  $p \rightarrow +\infty$  as well, in such a way that  $d/p = r + 1$  for some  $r \geq 0$ . They proved that, in this mean-field version of the original Euclidean problem, the AOC depends on  $r$  only. Their exact results have been later rigorously proved [10, 11, 12], and their techniques have been extended to calculate both finite-size corrections in the mean-field model [9, 13] and finite-dimensional corrections [14, 15, 16]. The results obtained by Mézard and Parisi on random matching problems paved the way to the application of statistical physics techniques, in particular from the theory of disordered systems, to the study of combinatorial optimization problems in presence of randomness [6]. Since their original contributions, the replica trick and the cavity method have been applied to many random optimization problems, and inspired new efficient algorithms for their solution, such as the belief propagation algorithm [17, 18].

The REAP is also naturally related to the Monge-Kantorovich transportation problem between probability measures in the continuum [19]. Indeed, it can be restated as a transportation problem between empirical measures obtained from the same PDF on an Euclidean domain [20, 21]. This correspondence has been fruitfully exploited to obtain some information on the asymptotic behavior of the AOC of the REAP for  $d \geq 1$  [3, 4, 20].

Moving from these recent results, in the present paper, we study the REAP in the lowest possible dimension, i.e.,  $d = 1$ . The one-dimensional REAP is particularly simple to solve when convex weight functions  $w$  are considered. Using the weight function in Eq. (3) with  $p > 1$ , and assuming the points uniformly distributed on the interval, the problem has been discussed and solved in Refs. [22, 23, 24], where a special correspondence between the optimal permutation and a Brownian bridge process on the interval has been found in the thermodynamical limit [22, 25]. The  $p < 0$  case is discussed in Ref. [23]. For results and details on the  $0 < p < 1$  case, see for example Refs. [21, 22]. By means of the aforementioned relation between the REAP and the Monge-Kantorovich theory, the treatment for  $p > 1$  in the uniform distribution case given in Ref. [24] has been later extended to a generic distribution on the line, and a general formula for the AOC has been given for the case of the REAP, assuming a quadratic weight function  $w_2$  [4]. A generalization of this formula, for a generic value of  $p > 1$ , is given in Ref. [20]. As stressed by the authors themselves in Refs. [4, 20], the provided general formula predict  $\varepsilon_N = O(N^{-p/2})$  but it may present some issues, in particular can be divergent depending on the properties of  $\varrho$ . In this case, the expression must be properly regularized, and a different, anomalous asymptotic behavior of  $\varepsilon_N$  may appear [4]. Here we will present a re-derivation of such a general expression for the AOC in the REAP on the line, extending the arguments in Refs. [4, 24] to any value of  $p > 1$  and, moreover, we will give details on the delicate regularization procedure when the scaling of  $\varepsilon_N$  is anomalous.

The paper is organized as follows. In Section 2 we will present the general theory for the REAP on the line. In Section 3 we will discuss the problem of the anomalous scaling of the AOC in the case of gapless support of  $\varrho$ , we will give the recipe for the regularization procedure to be applied to the general formula presented in Section 2, and we will exemplify our results studying different types of PDFs and deriving the scaling of the AOC in all cases. In Section 4 we will consider the effect of the presence of a gap in the support of  $\varrho$ . Finally, in Section 5 we will give our conclusions.

## 2. The random Euclidean assignment problem on the line

Let us start introducing the REAP on the line and discussing its specific properties. As stated in the Introduction, we will consider a PDF  $\varrho(x): \mathbb{R} \rightarrow \mathbb{R}^+$  on the real line,  $\int_{-\infty}^{+\infty} \varrho(x) dx = 1$ . Here and in the following we denote by

$$\Omega := \{x \in \mathbb{R} | \varrho(x) > 0\} \quad (4)$$

the support of  $\varrho$ , and therefore  $\varrho(x) = 0 \ \forall x \in \mathbb{R} \setminus \Omega$ . We will also denote by  $\bar{\Omega}$  the closure of  $\Omega$ , possibly including the points at infinity. The cumulative function  $\Phi(x)$  and the complementary cumulative  $\bar{\Phi}(x) := 1 - \Phi(x)$  are given by

$$\Phi(x) := \int_{-\infty}^x \varrho(\xi) d\xi =: 1 - \bar{\Phi}(x). \quad (5)$$

Let us suppose now that two sets of points both of cardinality  $N$ ,  $\Xi := \{x_i\}_{i=1,\dots,N}$  and  $\Upsilon := \{y_i\}_{i=1,\dots,N}$ , are generated on the line, independently and with the same PDF  $\varrho$ . We will assume that the points in  $\Xi$  and the points in  $\Upsilon$  are labeled in such a way that  $x_i < x_{i+1}$  and  $y_i < y_{i+1}$  for  $i = 1, \dots, N-1$ . We search for the optimal permutation  $\sigma_* \in \mathcal{S}_N$  such that

$$\varepsilon[\Xi, \Upsilon] := \min_{\sigma \in \mathcal{S}_N} \frac{1}{N} \sum_{i=1}^N |y_{\sigma(i)} - x_i|^p \equiv \frac{1}{N} \sum_{i=1}^N |y_{\sigma_*(i)} - x_i|^p, \quad p \in \mathbb{R}. \quad (6)$$

In optimal transport theory, the optimal cost  $\varepsilon[\Xi, \Upsilon]$  is the Kantorovich distance between the two empirical measures  $\varrho_\Xi(x) := \frac{1}{N} \sum_{i=1}^N \delta(x - x_i)$  and  $\varrho_\Upsilon(x) := \frac{1}{N} \sum_{i=1}^N \delta(x - y_i)$ , both having  $\varrho$  as (weak) limit. It turns out that, independently from the PDF  $\varrho$  adopted, the optimal permutation is simply given by  $\sigma_*(i) = i$ , provided that  $p > 1$  [20, 21, 22]. This result greatly simplifies the calculation. If we introduce the transport field

$$\phi_k := y_k - x_k, \quad (7)$$

then we can write

$$\varepsilon_N := \overline{\varepsilon[\Xi, \Upsilon]} = \frac{1}{N} \sum_{k=1}^N \int |\phi|^p \Pr[\phi_k \in d\phi]. \quad (8)$$

where we have used the notation  $z \in dx \Leftrightarrow z \in (x, x + dx)$ . The distribution of the  $k$ th transport field  $\phi_k$  can be obtained observing that

$$\Pr[x_k \in dx] = \binom{N}{k} \bar{\Phi}^{N-k}(x) d\Phi^k(x), \quad (9)$$

and therefore

$$\Pr[\phi_k \in d\phi] = d\phi \binom{N}{k}^2 \iint \delta(\phi - y + x) \bar{\Phi}^{N-k}(x) \bar{\Phi}^{N-k}(y) d\Phi^k(x) d\Phi^k(y). \quad (10)$$

The evaluation of  $\varepsilon_N$  can be performed writing  $\ddagger$

$$\varepsilon_N = \frac{1}{N} \iint_{\Omega \times \Omega} |y - x|^p \sum_{k=1}^N \binom{N}{k}^2 \bar{\Phi}^{N-k}(x) \bar{\Phi}^{N-k}(y) d\Phi^k(x) d\Phi^k(y) \quad (11a)$$

$$= \frac{1}{N} \iint_{\Omega \times \Omega} |y - x|^p {}_2F_1 \left[ 1 - N, 1 - N; 1; \frac{\Phi(x) \Phi(y)}{\bar{\Phi}(x) \bar{\Phi}(y)} \right] d\bar{\Phi}^N(x) d\bar{\Phi}^N(y). \quad (11b)$$

Up to now, no approximation has been performed. Being interested in the case of very large number of points, we observe that a nontrivial large  $N$  limit of Eq. (10) can be

$\ddagger$  To obtain Eq. (11) we have introduced the Gauss hypergeometric function

$${}_2F_1[a, b; c; z] := \sum_{k=0}^{\infty} \frac{(a)_k (b)_k}{(c)_k} \frac{z^k}{k!}, \quad (x)_k := \prod_{n=0}^{k-1} (x + n),$$

and we have used the fact that

$$\sum_{k=1}^N \binom{N}{k}^2 k^2 z^k = N^2 z \sum_{k=0}^{\infty} \frac{(1-N)_k (1-N)_k}{(1)_k} \frac{z^k}{k!} = N^2 z {}_2F_1[1 - N, 1 - N; 1; z].$$

obtained setting  $k = Ns + 1/2$  and introducing the variables  $\xi$  and  $\eta$  such that

$$\Phi(x) = s + \frac{\xi}{\sqrt{N}}, \quad \Phi(y) = s + \frac{\eta}{\sqrt{N}}, \quad (12a)$$

in such a way that  $s$  is kept fixed when  $N \rightarrow +\infty$ . This rescaling has a clear interpretation if we observe that an optimal assignment configuration between  $\Xi$  and  $\Upsilon$  for  $p > 1$  can be mapped, through the cumulative function  $\Phi$ , to an optimal assignment configuration of the same type between points uniformly distributed on  $[0, 1]$ , being  $\Phi$  ordering preserving. As shown in Refs. [23, 24], the optimal assignment between random points uniformly distributed on the unit interval is asymptotically equivalent to a Brownian bridge process after a rescaling of the type in Eq. (12a) is performed. This also implies that the (rescaled) transport field itself can be expressed, in the  $N \rightarrow +\infty$  limit, in terms of the Brownian bridge process composed with the (inverse) cumulative function  $\Phi^{-1}$ . Assuming that  $\Omega = \bar{\Omega}$  and that  $\bar{\Omega}$  is connected — i.e., that  $(\varrho \circ \Phi^{-1})(s) \neq 0$  for all  $s \in [0, 1]$  —, we have

$$\Phi^{-1}\left(s + \frac{\xi}{\sqrt{N}}\right) = \Phi^{-1}(s) + \frac{\xi}{\sqrt{N}\Psi_\varrho(s)} + o\left(\frac{1}{\sqrt{N}}\right), \quad (12b)$$

where we have introduced

$$\Psi_\varrho(s) := (\varrho \circ \Phi^{-1})(s). \quad (12c)$$

A similar equation holds for  $\Phi^{-1}(s + \eta/\sqrt{N})$ . This fact suggests us the following rescaling of  $\phi_k$  to obtain a nontrivial  $N \rightarrow +\infty$  limit,

$$\phi_k = \frac{\varphi(s)}{\sqrt{N}}. \quad (12d)$$

We remark here that

$$\frac{1}{2N} \leq s \leq 1 - \frac{1}{2N}, \quad (13)$$

a fact that will have important consequences in the following. We obtain, at the leading order, the following PDF for  $\varphi$ ,

$$\begin{aligned} \Pr[\varphi(s) \in d\varphi] &= d\varphi \iint \delta\left(\varphi - \frac{\eta - \xi}{\varrho(\Phi^{-1}(s))}\right) \frac{\exp\left(-\frac{\xi^2 + \eta^2}{2s(1-s)}\right)}{2\pi s(1-s)} d\xi d\eta \\ &= d\varphi \frac{\Psi_\varrho(s)}{2\sqrt{\pi s(1-s)}} \exp\left\{-\frac{[\Psi_\varrho(s)]^2}{4s(1-s)}\varphi^2\right\}, \end{aligned} \quad (14)$$

that implies that

$$N^{p/2}\varepsilon_N = \int_0^1 ds \frac{s^{\frac{p}{2}}(1-s)^{\frac{p}{2}}}{[\Psi_\varrho(s)]^p} \int_{-\infty}^{+\infty} |\varphi|^p \frac{e^{-\frac{\varphi^2}{4}}}{2\sqrt{\pi}} d\varphi + o(1) \quad (15a)$$

$$= \frac{2^p}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right) \int_0^1 \left[\frac{\sqrt{s(1-s)}}{\Psi_\varrho(s)}\right]^p ds + o(1) \quad (15b)$$

$$= \frac{2^p}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right) \int_\Omega dx \frac{\Phi^{\frac{p}{2}}(x) \bar{\Phi}^{\frac{p}{2}}(x)}{\varrho^{p-1}(x)} + o(1), \quad (15c)$$

that appears as a generalization of the expression obtained, using a different approach, in Ref. [4] for  $p = 2$ . This result can be stated in a slightly different way saying that, if  $\varrho(x)$  has compact and connected support, then

$$\varepsilon_N = \frac{2^p}{N^{\frac{p}{2}}\sqrt{\pi}}\Gamma\left(\frac{p+1}{2}\right)\int_{\Omega} dx \frac{\Phi^{\frac{p}{2}}(x)\bar{\Phi}^{\frac{p}{2}}(x)}{\varrho^{p-1}(x)} + o\left(\frac{1}{N^{\frac{p}{2}}}\right). \quad (15d)$$

The finiteness of the integral appearing in Eq. (15d) guarantees the typical scaling  $\varepsilon_N = O(N^{-p/2})$  of the AOC for  $N \gg 1$ . We say that the AOC has an *anomalous* scaling whenever the integral diverges. We will show now how information on the anomalous scaling can be extracted from the very same expression in Eqs. (15d) by means of a proper regularization.

### 3. The problem of regularization

The recipe provided by Eqs. (15) for the calculation of the asymptotic AOC might fail due to the presence of divergences, that we have neglected assuming  $\Omega = \bar{\Omega}$  connected. This is indeed the case for some PDFs. To explore this possibility, we will now relax the condition  $\Omega = \bar{\Omega}$ , but not the assumption that the closure  $\bar{\Omega}$  is connected. The set  $\bar{\Omega} \setminus \Omega$  is therefore given at most by isolated points (possibly at infinity). We will consider a disconnected  $\bar{\Omega}$  in Section 4.

The divergence of the expression in Eq. (15) suggests that  $\lim_N N^{p/2}\varepsilon_N = +\infty$ , but gives no hints about the scaling of  $N^{p/2}\varepsilon_N$  with  $N$ . In this case, a regularization must be performed to take into account the discrete nature of the problem, i.e., the finiteness of  $N$ . Such a regularization will allow us to extract information on the anomalous scaling of the AOC and, possibly, on the coefficients appearing in the scaling law. Under the hypothesis  $\Omega \neq \bar{\Omega}$  with  $\bar{\Omega}$  connected, the expression in Eq. (15b) may diverge due to the presence of a point  $x_* \in \partial\Omega$  (possibly at infinity) such that  $\lim_{x \rightarrow x_*} \varrho(x) = 0$ . In particular, denoting by  $s_* = \lim_{x \rightarrow x_*} \Phi(x) \in [0, 1]$ , a non-integrable divergence appears in Eq. (15b) if

$$\Psi_{\varrho}(s) = \begin{cases} O(s^{1/2+1/p}) & \text{if } s_* = 0, \\ O(|s - s_*|^{1/p}) & \text{if } 0 < s_* < 1, \\ O((1 - s)^{1/2+1/p}) & \text{if } s_* = 1. \end{cases} \quad (16)$$

Assuming that  $\bar{\Omega}$  is connected,  $\Omega \neq \bar{\Omega}$  does not automatically imply the presence of an anomalous scaling of the AOC: this is therefore a necessary, but not sufficient, condition.

The divergence can be avoided by means of a cut-off. The correct cut-off to be adopted is suggested by the approximations we have performed to obtain Eqs. (15) from Eq. (11), that is an exact expression.

A first regularization rule is obtained by taking into account Eq. (13) and therefore substituting  $\int_0^1 ds \rightarrow \int_{c_0/N}^{1-c_1/N} ds$  in Eq. (15b), where  $c_0$  and  $c_1$  are two positive regularizing constants that are unspecified at this level. This regularization is required to obtain the proper leading scaling of the asymptotic AOC only if a nonintegrable singularity appears in the integral in Eq. (15b) at  $s_* = 0$  and/or  $s_* = 1$ , and it provides information on the scaling of the  $o(1)$  corrections otherwise.

If a nonintegrable pole  $s_* \in (0, 1)$  is present, Eq. (12a) suggests to incorporate a finite-size correction removing an open ball centered in  $s_*$  having radius  $c_*/\sqrt{N}$  for

some positive regularizing constant  $c_*$  to be determined. Indeed, in Eq. (12a) we have approximated the quantity  $\Phi(x_k)$ , image of the position of the  $k$ th point through the cumulative  $\Phi$ , with its average value  $s = k/N - 1/2N$ , introducing an error that scales as  $O(1/\sqrt{N})$ .

In all cases, it is clear that the coefficients in the AOC scaling obtained after the regularization will depend on the introduced regularizing constants, that have to be determined by means of a fit procedure. We will give now some examples of the approach described above, comparing the obtained predictions with the results of numerical simulations.

### 3.1. Applications

We apply now the considerations above to a series of PDFs  $\varrho$  and we derive the asymptotic scaling of the corresponding AOC, always assuming a weight function in the form in Eq. (3) with  $p > 1$ . As preliminary observation, note that, given a REAP with PDF  $\varrho$  and AOC  $\varepsilon_N$ , introducing the variable  $x \rightarrow \hat{x} = \lambda x + \eta$  with  $\lambda > 0$  and  $\eta \in \mathbb{R}$ , it is easily seen that the old problem is mapped into a new problem with  $\hat{\varrho}(\hat{x}) = \frac{1}{\lambda} \varrho(\frac{\hat{x}-\eta}{\lambda})$  and  $\hat{\varepsilon}_N = \lambda^p \varepsilon_N$ . The results below can be therefore easily extended to more general PDFs involving nontrivial distribution parameters.

*3.1.1. Absence of singularity: flat distribution* Let us start from the simplest case, namely the REAP on the unit interval  $\Omega = [0, 1]$  with points extracted with uniform distribution  $\varrho(x) = \theta(x)\theta(1-x)$ . Here and in the following  $\theta(x)$  is the Heaviside function, such that  $\theta(x) = 1$  if  $x \geq 0$ ,  $\theta(x) = 0$  otherwise. Many results, both numerical and analytical, are available in the literature on this case [5, 22, 23], both with open and with periodic boundary conditions, in particular for  $p = 2$  [3, 24, 26]. The flat PDF case requires no regularization, being  $\Psi_\varrho(s) = 1$  and therefore we will briefly recall the final result only as an application of Eq. (15b) for the sake of completeness and for comparison with other cases studied below. The integral in Eq. (15b) converges for any  $p > -2$  and we obtain

$$\varepsilon_N = \frac{\Gamma(p/2 + 1)}{p + 1} \frac{1}{N^{p/2}} + o\left(\frac{1}{N^{p/2}}\right). \quad (17)$$

The expression above corresponds to the asymptotic AOC for the REAP only for  $p > 1$ . The result for  $p = 2$ , namely  $\lim_N N \varepsilon_N = 1/3$ , can be derived by direct inspection observing that an exact formula is easily obtained for all values of  $N$ ,

$$\varepsilon_N = \frac{1}{N} \sum_{k=1}^N k^2 \binom{N}{k}^2 \int_0^1 \int_0^1 (y-x)^2 (1-x)^{N-k} (1-y)^{N-k} x^{k-1} y^{k-1} dx dy = \frac{1}{3} \frac{1}{N+1}. \quad (18)$$

Numerical evidences for the formulas above can be found in the literature [3, 22, 23], see also Fig. 3a.

*3.1.2. Singularity for  $s_* \in \{0, 1\}$*  Let us now consider a set of examples in which the REAP is given on the line, and such that  $\Psi_\varrho(s_*) = 0$  for a value  $s_* \in \{0, 1\}$ . We consider both exponentially decaying PDFs (the exponential distribution and the Rayleigh distribution) and power-law decaying PDFs (Pareto laws), deriving the proper scaling of the AOC in all discussed cases.

*Exponential distribution* Let us start considering the exponential distribution,

$$\varrho(x) = e^{-x}\theta(x), \quad \Phi(x) = (1 - e^{-x})\theta(x). \quad (19)$$

In this case,  $\Omega = [0, +\infty)$  and a nonintegrable singularity may appear for  $x \rightarrow +\infty$ . For  $1 < p < 2$  Eq. (15b) is convergent and the AOC is

$$N^{p/2}\varepsilon_N = \frac{2^p}{\sqrt{\pi}}\Gamma\left(\frac{p+1}{2}\right) \int_0^1 \left(\frac{s}{1-s}\right)^{\frac{p}{2}} ds + o(1) = \Gamma(p+1)\Gamma\left(1 - \frac{p}{2}\right) + o(1). \quad (20)$$

The prediction above is fully consistent with numerical results, see Fig. 1a. It is evident, however, that a divergence appears when  $p = 2$ . Indeed, with reference to Eq. (16), we have that  $\Psi_\varrho(s) = 1 - s = O((1-s)^{1/2+1/p})$  if  $p \geq 2$ . To better understand the nature of this divergence, it is useful to calculate  $\varepsilon_N$  directly, as in Eq. (11), taking advantage of the fact that the computation can be performed exactly in this case for any value of  $N$ ,

$$\begin{aligned} N\varepsilon_N &= \sum_{k=1}^N k^2 \binom{N}{k}^2 \int_0^1 ds \int_0^1 dt \ln^2 \frac{1-s}{1-t} (st)^{k-1} (1-s)^{N-k} (1-t)^{N-k} \\ &= \sum_{k=1}^N \frac{2}{k} = 2H_N = 2\ln N + 2\gamma_E + \frac{1}{N} - \frac{1}{6N^2} + o\left(\frac{1}{N^2}\right), \end{aligned} \quad (21)$$

where  $H_N$  is the  $N$ th harmonic number and  $\gamma_E$  is the Euler's gamma constant. In Fig. 1b we numerically verified the expression in Eq. (21). The appearance of the divergence in our integral expression in Eq. (15b) is therefore due to an actual (logarithmic) divergence of  $N\varepsilon_N$  for  $N \rightarrow +\infty$ . Following the criterion given in Section 2, Eq. (15b) for  $p \geq 2$  must be regularized as

$$\begin{aligned} N\varepsilon_N &\approx \frac{2^p}{\sqrt{\pi}}\Gamma\left(\frac{p+1}{2}\right) \int_0^{1-c/N} \left(\frac{s}{1-s}\right)^{\frac{p}{2}} ds \\ &= \frac{2^{p+1}}{(p+2)\sqrt{\pi}}\Gamma\left(\frac{p+1}{2}\right) \left(1 - \frac{c}{N}\right)^{1+p/2} {}_2F_1\left[\frac{p}{2}, \frac{p}{2} + 1; \frac{p}{2} + 2; 1 - \frac{c}{N}\right]. \end{aligned} \quad (22)$$

The expression above must be interpreted as a regularization-dependent *asymptotic* formula for the AOC. In particular, the large  $N$  expansion will provide us the scaling properties of the optimal cost, up to some coefficients possibly depending on the regularization. For example, for  $p = 2$  the expression above becomes

$$N\varepsilon_N = 2\ln N - 2\log c - 2 + o(1), \quad (23)$$

that is perfectly compatible with the exact formula in Eq. (21), whereas the finite-size correction depends on the regularization  $c$ . With reference to Eq. (21), we can also infer that  $c = e^{-\gamma_E - 1}$ . For  $p > 2$  we can expand Eq. (22) as

$$N^{p/2}\varepsilon_N = \frac{2^{p+1}c^{1-p/2}}{(p-2)\sqrt{\pi}}\Gamma\left(\frac{p+1}{2}\right) N^{\frac{p}{2}-1} + o\left(N^{\frac{p}{2}-1}\right). \quad (24a)$$



In particular, for  $2 < p < 4$  we have

$$N^{p/2} \varepsilon_N = \frac{2^{p+1} c^{1-p/2}}{(p-2)\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right) N^{\frac{p}{2}-1} + \Gamma\left(1 - \frac{p}{2}\right) \Gamma(p+1) + o(1). \quad (24b)$$

We obtain therefore the scaling  $\varepsilon_N = O(N^{-1})$  for the leading term but we cannot give a prediction for the coefficient in front of it, due to its dependence on the regularization constant  $c$ . We have, instead, a complete analytic prediction for the finite-size correction. For  $p = 4$ , a new logarithmic correction appears. In this case, indeed, our formula in Eq. (22) gives

$$N^2 \epsilon = \frac{12}{c} N - 24 \ln N + O(1). \quad (24c)$$

Once again, the coefficient of the leading term is inaccessible, despite the fact that the correct scaling is recovered, but we obtain a prediction of a logarithmic correction, with its coefficient. We do expect, but it is not obvious *a priori*, that the value of  $c$  appearing in Eqs. (24) is the same that we have obtained for  $p = 2$ . Performing a fit on our numerical results, presented in Fig. 1c, we get  $c = 0.203(2)$  for  $p = 3$ ,  $0.2084(4)$  for  $p = 4$  and  $c = 0.2069(5)$  for  $p = 5$ , that are all very close to the value of  $c = 0.206549 \dots$  analytically obtained for  $p = 2$ .

*Rayleigh distribution* As further example of regularization in the case of a PDF with exponential tail, we consider now the Rayleigh distribution,

$$\varrho(x) = x e^{-\frac{x^2}{2}} \theta(x), \quad \Phi(x) = \left(1 - e^{-\frac{x^2}{2}}\right) \theta(x). \quad (25)$$

In this case  $\Omega = (0, +\infty)$  and we have that  $\Psi_\varrho(s) = (1-s)\sqrt{-2\ln(1-s)}$  that is infinitesimal both in  $s = 0$  and in  $s = 1$ . In particular,  $\Psi_\varrho(s) = \sqrt{2s} + O(s^{3/2})$  for  $s \rightarrow 0$  and therefore, according to Eq. (16), there are no integrability issues for  $s \rightarrow 0$  for any value of  $p > 1$ . On the contrary, for  $s \rightarrow 1$   $\Psi_\varrho(s) = O((1-s)^{1/2+1/p})$  for any  $p > 1$ . The integral is therefore always divergent and a regularization is needed. We proceed in the usual way, restricting ourselves to the  $p = 2$  case,

$$N \varepsilon_N \approx \int_0^{1-\frac{c}{N}} \frac{s}{s-1} \frac{1}{\ln(1-s)} ds = \gamma_E + \ln \ln \frac{N}{c} + \int_{N/c}^{+\infty} \frac{dz}{z^2 \ln z} = \ln \ln N + \gamma_E + o(1). \quad (26)$$

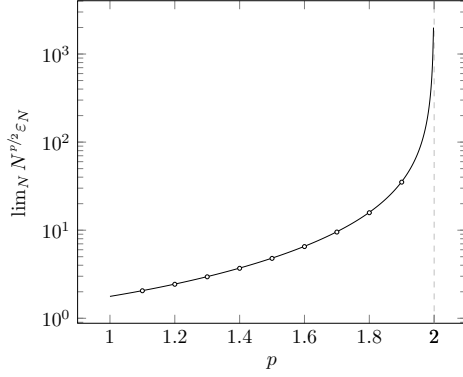
Our results have been numerically verified, as shown in Fig. 1d.

*Pareto distribution* Let us now consider a power-law decaying PDF, e.g., a Pareto distribution,

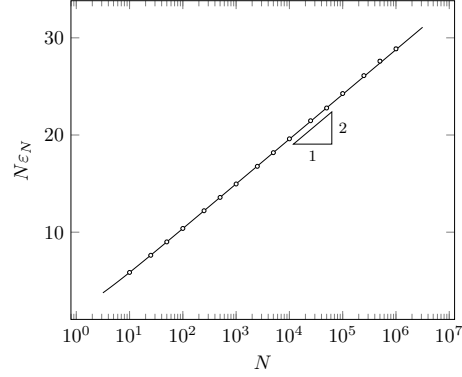
$$\varrho_\alpha(x) = \frac{\alpha}{x^{\alpha+1}} \theta(x-1), \quad \Phi_\alpha(x) = \frac{x^\alpha - 1}{x^\alpha} \theta(x-1), \quad \alpha > 0. \quad (27)$$

Here we have  $\Omega = [1, +\infty)$ . If we consider

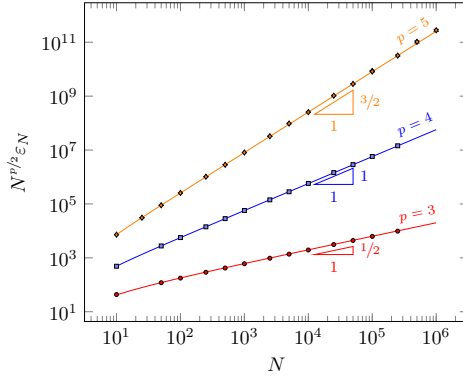
$$p < \frac{2\alpha}{\alpha+2}, \quad (28)$$



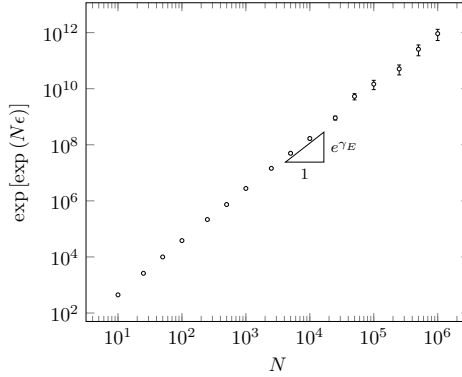
(a) Values of  $\epsilon_N$  in the case of exponentially distributed points, obtained for  $p \in (1, 2)$  compared with the theoretical prediction in Eq. (20) (smooth line). The asymptotic value of  $N^{p/2} \epsilon_N$  for each value of  $p$  has been obtained fitting the numerical results for  $N$  up to  $2.5 \cdot 10^5$  points, assuming the scaling  $f(N) = \epsilon + \epsilon_1 N^{\frac{p}{2}-1}$ .



(b) Numerical results for the AOC for  $p = 2$  in the case of exponentially distributed points, showing a logarithmic divergence in agreement with the prediction in Eq. (23). The smooth line is the prediction in Eq. (21).



(c) Numerical results for the AOC for  $p > 2$  in the case of exponentially distributed points. The smooth lines are fits obtained assuming the scaling behavior in Eqs. (24).



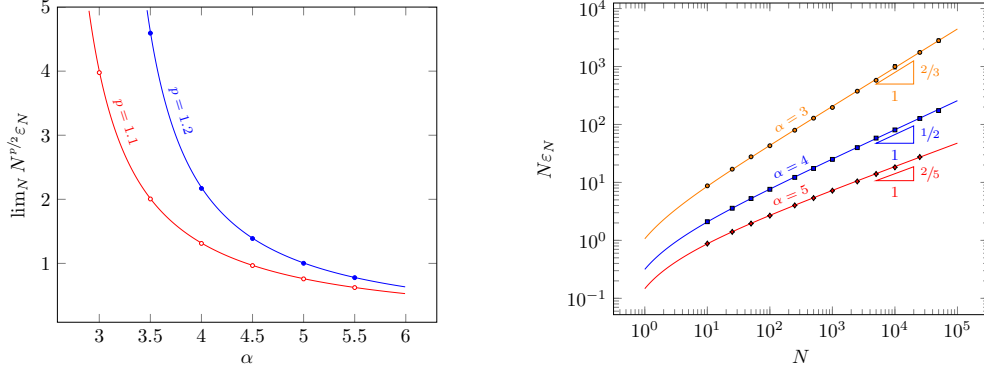
(d) Plot for  $\exp[\exp(N \epsilon_N)] \sim N^{e^{\gamma_E}}$  in the case of Rayleigh distribution and  $p = 2$ , to be compared with the theoretical prediction in Eq. (26).

**Figure 1.** AOC in the case of exponential distribution and Rayleigh distribution. Error bars are represented but smaller than the markers.

Eq. (15b) gives a finite result, namely

$$\begin{aligned} \lim_N N^{p/2} \epsilon_N &= \frac{2^p}{\alpha^p \sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right) \int_0^1 s^{\frac{p}{2}} (1-s)^{\frac{p}{2}-\frac{p}{\alpha}-p} ds \\ &= \frac{2^p}{\alpha^p \sqrt{\pi}} \frac{\Gamma\left(\frac{p}{2}+1\right) \Gamma\left(1-\frac{2+\alpha}{2\alpha}p\right) \Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(2-\frac{p}{\alpha}\right)}. \end{aligned} \quad (29)$$

This formula has been verified, for a set of values of  $p$  and  $\alpha$ , in Fig. 2a. When  $(2-p)\alpha \leq 2p$  the integral does not converge (in particular, does not converge for any



(a) Values of  $\epsilon_N$  obtained for  $p = 1.1$  and  $p = 1.2$  and different values of  $\alpha$ , compared with the theoretical prediction in Eq. (29) (smooth lines). The asymptotic value of  $N^{p/2} \epsilon_N$  for each value of  $p$  has been obtained fitting the numerical results for  $N$  up to  $10^5$  points, assuming the scaling  $f(N) = \epsilon + \epsilon_1 N^{p \frac{\alpha+2}{2\alpha} - 1}$ .

(b) Numerical results for the AOC for  $p = 2$  and different values of  $\alpha$ . The fits are obtained using a fitting function in the form given by Eq. (31); we obtained  $c = 0.0668(5)$  for  $\alpha = 3$ ,  $c = 0.0939(5)$  for  $\alpha = 4$  and  $c = 0.1121(6)$  for  $\alpha = 5$ .

**Figure 2.** AOC in the case of Pareto distribution. Error bars are represented but smaller than the markers.

value of  $\alpha$  when  $p = 2$ ). Indeed, with reference to Eq. (16),  $\Psi_{\ell\alpha}(s) = \alpha(1-s)^{1+\frac{1}{\alpha}}$ , and, therefore, a non-integrable singularity appears for  $s_* = 1$  when  $1 + \frac{1}{\alpha} \geq \frac{1}{2} + \frac{1}{p}$ . We can proceed regularizing the integral for  $p \geq \frac{2\alpha}{\alpha+2}$ ,

$$\begin{aligned}
 N^{p/2} \epsilon_N &\approx \frac{2^p}{\alpha^p \sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right) \int_0^{1-\frac{c}{N}} s^{\frac{p}{2}} (1-s)^{\frac{p}{2}-\frac{p}{\alpha}-p} ds \\
 &= \frac{2^{p+1}}{\alpha^p \sqrt{\pi}} \frac{\Gamma\left(\frac{p+1}{2}\right)}{p+2} \left(1 - \frac{c}{N}\right)^{\frac{p}{2}+1} {}_2F_1\left[\frac{p}{2}+1, \frac{p}{2} \frac{\alpha+2}{\alpha}; \frac{p}{2}+2; 1 - \frac{c}{N}\right] \\
 &= \begin{cases} \frac{2^{p+1}}{\alpha^{p-1} \sqrt{\pi}} \frac{\Gamma\left(\frac{p+1}{2}\right)}{2p+\alpha p-2\alpha} \left(\frac{N}{c}\right)^{p \frac{\alpha+2}{2\alpha} - 1} + o\left(N^{p \frac{\alpha+2}{2\alpha} - 1}\right) & \text{for } p > \frac{2\alpha}{\alpha+2}, \\ \frac{p+2}{2} \ln N - \frac{p+2}{2} (H_{p/2} + \ln c) + o(1) & \text{for } p = \frac{2\alpha}{\alpha+2}. \end{cases} \quad (30)
 \end{aligned}$$

For  $p = 2$ , for example, we find

$$N \epsilon_N = \frac{1}{\alpha} \left(\frac{N}{c}\right)^{2/\alpha} + \begin{cases} \frac{2}{\alpha(\alpha-2)} \left(\frac{N}{c}\right)^{\frac{2}{\alpha}-1} + o(N^{\frac{2}{\alpha}-1}) & \text{for } 0 < \alpha < 2, \\ -\frac{\ln N}{2} + \frac{\ln c - 1}{2} + o(1) & \text{for } \alpha = 2, \\ -\frac{1}{\alpha-2} + o(1) & \text{for } \alpha > 2. \end{cases} \quad (31)$$

In Fig. 2b we numerically verify all the scaling properties above for  $p = 2$  and  $\alpha = 3, 4, 5$ .

**3.1.3. Singularity for  $s_* \in (0, 1)$**  Let us now consider a PDF such that  $\Psi_{\ell}(s_*) = 0$  for  $0 < s_* < 1$  and let us derive the scaling properties of the corresponding AOC in

this case. As an example, we consider

$$\varrho_\alpha(x) = \frac{2 \cos^2(\alpha\pi x)}{1 + \text{sinc}(2\pi\alpha)} \theta(x)\theta(1-x), \quad \alpha \in (0, 1], \quad (32a)$$

$$\Phi_\alpha(x) = x \frac{1 + \text{sinc}(2\pi\alpha)}{1 + \text{sinc}(2\pi\alpha)} \theta(x)\theta(1-x). \quad (32b)$$

Here we have used the definition  $\text{sinc}(x) := \frac{\sin(x)}{x}$ . The distribution above recovers the uniform one for  $\alpha \rightarrow 0$  and it has a (double) zero for  $x = 1/2\alpha \in [1/2, 1]$  if  $\alpha \in [1/2, 1]$ . The support is therefore  $\Omega = [0, 1] \setminus \{1/2\alpha\}$ . In particular, for  $\alpha \in [1/2, 1]$ , we have

$$\Psi_{\varrho_\alpha}(s) = \frac{(6\pi\alpha)^{2/3}}{\sqrt[3]{2 + 2 \text{sinc}(2\pi\alpha)}} (s - s_*)^{2/3} + o\left((s - s_*)^{2/3}\right),$$

with  $s_* := \frac{1}{2\alpha} \frac{1}{1 + \text{sinc}(2\pi\alpha)}$ . (33)

Depending on the value of  $\alpha$ , the general exposition given in the present Section implies therefore three different regimes of the asymptotic AOC.

For  $\alpha \in (0, 1/2)$  the asymptotic AOC is finite for any value of  $p > 1$ . The integral in Eq. (15b) has been evaluated numerically and the prediction has been compared with our numerical results in Fig. 3a, where a perfect agreement is found.

When  $\alpha = 1/2$  the singularity  $s_*$  moves to 1. We obtain the regularized integral

$$N^{p/2} \varepsilon_N = \frac{2^p}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right) \int_0^{1 - \frac{c_1/2}{\sqrt[3]{N}}} dx \frac{\Phi_\alpha^{\frac{p}{2}}(x) \bar{\Phi}_\alpha^{\frac{p}{2}}(x)}{\varrho_\alpha^{p-1}(x)} + o(1) \propto \begin{cases} N^{\frac{p}{6}-1} & \text{for } p > 6, \\ \ln N & \text{for } p = 6, \\ \text{constant} & \text{for } 1 < p < 6. \end{cases} \quad (34)$$

We verified the scaling above in Fig. 3c. In the  $p = 6$  case, in particular, we have

$$N^3 \varepsilon_N = \frac{160}{27\pi^4} \ln N + O(1). \quad (35)$$

For  $\alpha \in (1/2, 1]$ , instead, there is a singularity in  $s_* \in [1/2, 1)$  and the regularization procedure has to be modified. In this case we have to exclude from the integration domain a ball centered in  $s_*$  and radius  $O(1/\sqrt{N})$ . Observing that

$$\begin{aligned} \Phi_\alpha^{-1}\left(\frac{1}{2\alpha(1 + \text{sinc}(2\pi\alpha))} \pm \frac{c}{\sqrt{N}}\right) &= \frac{1}{2\alpha} \pm \frac{1}{\alpha} \sqrt[3]{\frac{3c(1 + \text{sinc}(2\pi\alpha))}{2\pi^2\sqrt{N}}} + o\left(\frac{1}{\sqrt[6]{N}}\right), \\ &\equiv \frac{1}{2\alpha} \pm \frac{\hat{c}_\alpha}{\sqrt[6]{N}} + o\left(\frac{1}{\sqrt[6]{N}}\right) \end{aligned} \quad (36)$$

and denoting the “regularized” domain by

$$\Omega_\alpha := [0, 1] \setminus \left(\frac{1}{2\alpha} - \frac{\hat{c}_\alpha}{\sqrt[6]{N}}, \frac{1}{2\alpha} + \frac{\hat{c}_\alpha}{\sqrt[6]{N}}\right), \quad (37)$$

we can write the regularized integral as

$$N^{p/2} \varepsilon_N = \frac{2^p}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right) \int_{\Omega_\alpha} dx \frac{\Phi_\alpha^{\frac{p}{2}}(x) \bar{\Phi}_\alpha^{\frac{p}{2}}(x)}{\varrho_\alpha^{p-1}(x)} + o(1) \propto \begin{cases} N^{\frac{2p-3}{6}} & \text{for } p > 3/2, \\ \ln N & \text{for } p = 3/2, \\ \text{constant} & \text{for } 1 < p < 3/2, \end{cases} \quad (38)$$

where we limited ourselves to the scaling behavior of the AOC. In Fig. 3b we present our numerical results for  $p = 2$  and  $p = 3$ . In both cases, the scaling predicted in Eq. (38) is confirmed. For  $p = 3/2$ , in particular, we find

$$N^{3/4} \varepsilon_N = \frac{\Gamma(1/4)}{6\alpha(1 + \operatorname{sinc}(2\pi\alpha))} \left( \frac{1 + \operatorname{sinc}(2\pi\alpha) - \frac{1}{2\alpha}}{2\alpha\pi^2} \right)^{3/4} \ln N + O(1). \quad (39)$$

In Fig. 3d we show our numerical results for this case, once again in agreement with the prediction. Observe that the different regularization applied in this case implies a completely different scaling of the asymptotic AOC respect to the one obtained for  $s_* = 1$ .

#### 4. Assignment on disjoint intervals: an example

In the examples above, and in the general remarks in Section 3, we have always assumed that the domain  $\Omega$  is such that  $\bar{\Omega}$  is a connected interval, and therefore  $\Phi$  is an invertible function on  $\bar{\Omega}$ . This is not the case if the domain  $\Omega$  has a “gap”. In this Section we will study the effects of such a gap on the asymptotic AOC. We will limit ourselves to the case  $\Omega = A \cup B$  with  $\bar{A}, \bar{B}$  connected intervals such that  $\bar{A} \cap \bar{B} = \emptyset$ . In the following we will assume that  $\forall x \in A$  and  $\forall y \in B$ ,  $x < y$ . To avoid complications due to the presence of singularities in the integrals, we will also assume that  $\Omega = \bar{\Omega}$ . The lack of invertibility of  $\Phi$  is due in this case to the fact that  $\lim_{x \rightarrow \sup A} \Phi(x) = \lim_{x \rightarrow \inf B} \Phi(x) = \hat{s}$ , despite the fact that  $a = \sup A \neq \inf B = b$ . We expect that our approach proposed in Section 2 fails in this situation, because the transport field  $\phi_k$  in Eq. (7) is not infinitesimal in general for  $N \rightarrow +\infty$ .

In the simple case mentioned here,  $\Omega = A \cup B$  with  $\bar{A} \cap \bar{B} = \emptyset$ , the exact formula in Eq. (11) can be written as

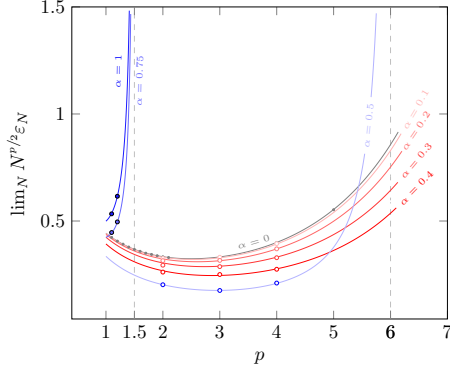
$$\varepsilon_N = \frac{1}{N} \int d\phi |\phi|^p \sum_{k=1}^{\infty} \left[ p_k^{(AA)}(\phi) + p_k^{(BB)}(\phi) + p_k^{(AB)}(\phi) + p_k^{(BA)}(\phi) \right]. \quad (40)$$

In the expression above, the quantity

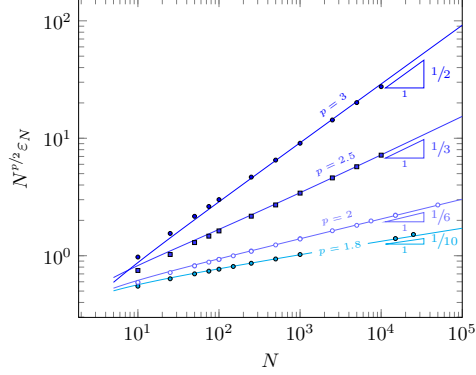
$$\begin{aligned} p_k^{(XY)}(\phi) d\phi &:= \Pr[\phi_k \in d\phi, x_k \in X, y_k \in Y] \\ &= d\phi \binom{N}{k}^2 \iint_{X \times Y} \delta(\phi - y + x) \bar{\Phi}^{N-k}(x) \bar{\Phi}^{N-k}(y) d\bar{\Phi}^k(x) d\bar{\Phi}^k(y) \end{aligned} \quad (41)$$

is the joint probability that the  $k$ th transport field  $\phi_k = y_k - x_k$  takes value in the interval  $(\phi, \phi + d\phi)$ ,  $x_k \in X$  and  $y_k \in Y$ . We expect that, to obtain a nontrivial  $N \rightarrow +\infty$  limit from  $p_k^{(AA)}$  and  $p_k^{(BB)}$ , we have to rescale  $\phi_k$  following Eq. (12), due to the fact that matched points in the same interval can be arbitrarily close in the thermodynamical limit. Indeed, we can repeat the same calculations in Section 2 performing the rescaling in Eqs. (12) and recovering, with the same caveat, a limiting distribution exactly in the form given in Eq. (14),

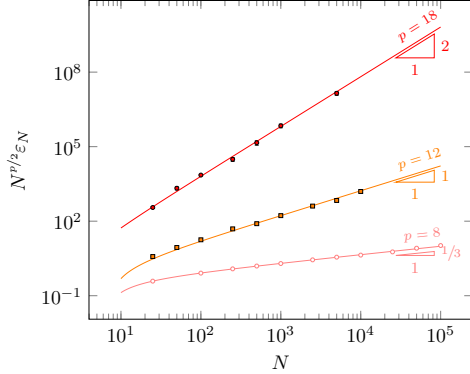
$$\frac{1}{N} \sum_{k=1}^N \Pr[\phi_k \in d\phi, x_k \in A, y_k \in A] + \frac{1}{N} \sum_{k=1}^N \Pr[\phi_k \in d\phi, x_k \in A, y_k \in B]$$



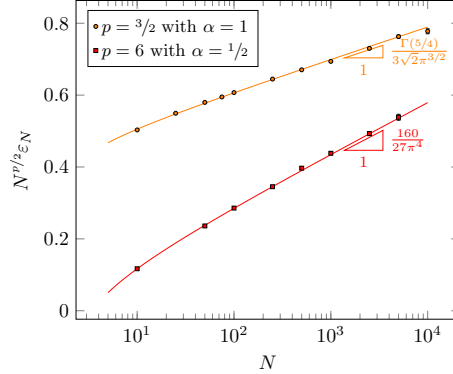
(a) Values of  $\epsilon_N$  in the case of points distributed with PDF given in Eq. (32), obtained for different values of  $p$  and  $\alpha$  for which  $\lim_N N^{p/2} \epsilon_N$  is finite, compared with the theoretical prediction obtained using Eq. (15b) (smooth lines). The limit curve for  $\alpha = 0$ , given by Eq. (17), is also represented (gray), along with the numerical results for the asymptotic AOC in the uniform distribution case.



(b) Numerical results for the AOC in the case of points distributed with PDF given in Eq. (32) with  $\alpha = 1$ . The smooth lines are fits obtained a scaling of the type  $f(N) = N^{p/3-1/2} (\epsilon_1 + \epsilon_2/N) + \epsilon_0$ . The obtained scaling are in agreement with the prediction in Eq. (38).



(c) Numerical results for the AOC in the case of points distributed with PDF given in Eq. (32) with  $\alpha = 1/2$ . The represented fits have been obtained assuming a fitting function  $f(N) = N^{p/6-1} (\epsilon_1 + \epsilon_2/N) + \epsilon_0$ . The obtained scaling laws are in agreement with the prediction in Eq. (34).



(d) Numerical results for the AOC in the case of points distributed with PDF given in Eq. (32) in the cases in which a logarithmic divergence appears. The smooth lines are fits obtained assuming a scaling  $f(N) = \epsilon \ln N + \epsilon_0 + \epsilon_1/\ln N$ , where  $\epsilon$  has been provided by the predictions in Eq. (39) and Eq. (35).

**Figure 3.** AOC in the case of points distributed with PDF given in Eq. (32). Error bars are represented but smaller than the markers.

$$= \mathrm{d}\varphi \int_0^1 \mathrm{d}s \frac{\Psi_\varrho(s)}{2\sqrt{\pi s(1-s)}} \exp \left\{ -\frac{[\Psi_\varrho(s)]^2}{4s(1-s)} \varphi^2 \right\} + o(1). \quad (42)$$

This formula is exactly the expression we would have obtained if  $\bar{\Omega}$  were connected. If convergent, as it will happen under the hypotheses adopted here, this contribution will give a  $O(N^{-p/2})$  term in the expression of  $\varepsilon_N$  for  $N \gg 1$ .

On the other hand, the last two contributions in Eq. (40) corresponds to the matching transport between the two components of  $\Omega$ , i.e.,  $A \rightarrow B$  or  $B \rightarrow A$ , and therefore the transport field in this case is of the order of the distance between  $A$  and  $B$ , namely  $\inf B - \sup A$ . The asymptotic rescaling given in Eqs. (12), therefore, cannot be applied to this term. However, from the fact that two matched points  $x_k$  and  $y_k$  have  $|\Phi(x_k) - \Phi(y_k)| = O(1/\sqrt{N})$ , if  $x_k \in A$  and  $y_k \in B$  this implies that

$$\Phi(x_k) = \hat{s} + \frac{\xi_k}{\sqrt{N}}, \quad \Phi(y_k) = \hat{s} + \frac{\eta_k}{\sqrt{N}} \quad (43)$$

with  $\xi_k < 0$  and  $\eta_k > 0$ , and therefore

$$x_k = a + \frac{\xi_k}{\sqrt{N}\varrho(a)} + o\left(\frac{1}{\sqrt{N}}\right), \quad y_k = b + \frac{\eta_k}{\sqrt{N}\varrho(b)} + o\left(\frac{1}{\sqrt{N}}\right), \quad (44)$$

where  $a = \sup A$  and  $b = \inf B$  and, under our hypotheses,  $\varrho(a) \neq 0$  and  $\varrho(b) \neq 0$ . The relations above suggest the rescaling  $\phi_k \rightarrow b - a + \hat{\varphi}_k/\sqrt{N}$  for  $k = Ns + 1/2$ . A nontrivial distribution for  $\hat{\varphi}$  is obtained assuming  $s = \hat{s} + \sigma/\sqrt{N}$ . Indeed

$$\begin{aligned} & \frac{1}{N} \sum_{k=1}^N \Pr[\phi_k \in \mathrm{d}\phi, x_k \in A, y_k \in B] \\ &= \frac{\mathrm{d}\phi}{N} \sum_{k=1}^N \binom{N}{k}^2 \int_0^{\hat{s}} \mathrm{d}u^k \int_{\hat{s}}^1 \mathrm{d}v^k \delta(\phi - \Phi^{-1}(v) + \Phi^{-1}(u)) (1-u)^{N-k} (1-v)^{N-k} \\ &\approx \mathrm{d}\hat{\varphi} \int_{-\infty}^{+\infty} \frac{\mathrm{d}\sigma}{\sqrt{N}} \int_{-\infty}^0 \mathrm{d}\xi \int_0^{+\infty} \mathrm{d}\eta \delta\left(\hat{\varphi} - \frac{\eta}{\varrho(b)} + \frac{\xi}{\varrho(a)}\right) \frac{\exp\left(-\frac{(\xi-\sigma)^2 + (\eta-\sigma)^2}{2\hat{s}(1-\hat{s})}\right)}{2\pi\hat{s}(1-\hat{s})} \\ &= \frac{\mathrm{d}\hat{\varphi}}{\sqrt{N}} \begin{cases} \frac{\varrho(a)\varrho(b)}{2(\varrho(a)-\varrho(b))} \left[ \operatorname{Erf}\left(\frac{\varrho(a)\hat{\varphi}}{\sqrt{2\hat{s}(1-\hat{s})}}\right) - \operatorname{Erf}\left(\frac{\varrho(b)\hat{\varphi}}{\sqrt{2\hat{s}(1-\hat{s})}}\right) \right] \theta(\hat{\varphi}) & \varrho(a) \neq \varrho(b), \\ \varrho^2(a)\hat{\varphi} \frac{e^{-\frac{\varrho^2(a)\hat{\varphi}^2}{4\hat{s}(1-\hat{s})}}}{2\sqrt{\pi\hat{s}(1-\hat{s})}} \theta(\hat{\varphi}) & \varrho(a) = \varrho(b), \end{cases} \\ &:= \frac{1}{\sqrt{N}} \Pr[\hat{\varphi} \in \mathrm{d}\hat{\varphi}, A \rightarrow B]. \quad (45) \end{aligned}$$

In a similar manner the expression for  $\Pr[\hat{\varphi} \in \mathrm{d}\hat{\varphi}, B \rightarrow A]$  can be obtained. Collecting our results, we can write down the contribution to the asymptotic probability for  $\hat{\varphi}$  given by the matching between points of different subintervals as

$$\Pr[\hat{\varphi}(s) \in \mathrm{d}\hat{\varphi}, A \leftrightarrow B] = \frac{\mathrm{d}\hat{\varphi}}{\sqrt{N}} \begin{cases} \varrho(a)\varrho(b) \frac{\operatorname{Erf}\left(\frac{\varrho(b)|\hat{\varphi}|}{\sqrt{2\hat{s}(1-\hat{s})}}\right) - \operatorname{Erf}\left(\frac{\varrho(a)|\hat{\varphi}|}{\sqrt{2\hat{s}(1-\hat{s})}}\right)}{2(\varrho(b)-\varrho(a))} & \varrho(a) \neq \varrho(b), \\ \varrho^2(a)|\hat{\varphi}| \frac{e^{-\frac{\varrho^2(a)\hat{\varphi}^2}{4\hat{s}(1-\hat{s})}}}{2\sqrt{\pi\hat{s}(1-\hat{s})}} & \varrho(a) = \varrho(b). \end{cases} \quad (46)$$

Observe that the previous contributions are not normalized in  $\hat{\varphi}$ . This is due to the fact that they appear as  $O(1/\sqrt{N})$  corrections to the distribution  $\Pr[\varphi(s) \in d\varphi]$  that has Eq. (42) as leading term: higher order corrections to  $\Pr[\phi_k \in d\phi, x_k \in A, y_k \in A]$  and  $\Pr[\phi_k \in d\phi, x_k \in B, y_k \in B]$ , that would guarantee for  $N \gg 1$  the total integral of the corrections to  $\Pr[\varphi(s) \in d\varphi]$  to be zero, have not been computed. This will be irrelevant for our final computation, because the matching field is  $O(1/\sqrt{N})$  when matching points in the same interval, but  $O(1)$  when matching points in different intervals. The final result is

$$\begin{aligned} \varepsilon_N &= \frac{|b-a|^p}{\sqrt{N}} \int_{-\infty}^{+\infty} \Pr[\hat{\varphi}(s) \in d\hat{\varphi}, A \leftrightarrow B] + o\left(\frac{1}{\sqrt{N}}\right) \\ &= \frac{2|b-a|^p}{\sqrt{N}} \sqrt{\frac{\hat{s}(1-\hat{s})}{\pi}} + o\left(\frac{1}{\sqrt{N}}\right), \end{aligned} \quad (47)$$

irrespectively from the fact that  $\varrho(a) = \varrho(b)$  or not. Remarkably, the coefficient in front of the leading term does not depend on  $\varrho(x)$  but only on the average fraction of points that are in each of the two subintervals, i.e., on  $\hat{s}$ . Moreover, the obtained scaling can be intuitively justified observing that the number of points of  $\Xi$  that are expected to fall, e.g., in  $A$  are  $N\hat{s}$ , but the fluctuations to this number scale as  $\sqrt{N}$ . The same reasoning applies to  $\Upsilon$ . This means that  $O(\sqrt{N})$  points in  $A$  have necessarily to be matched with points in  $B$  with a matching cost that is  $O(|b-a|^p)$ , giving a final  $O(1/\sqrt{N})$  contribution to  $\varepsilon_N$ .

*Uniform distribution with a gap* To better exemplify this situation, let us consider the following PDF on  $\Omega = [0, 1/2 - \alpha/2] \cup [1/2 + \alpha/2, 1]$  with  $\alpha \in [0, 1)$  and  $q \in (0, 1)$ ,

$$\varrho_{\alpha,q}(x) = \begin{cases} \frac{2q}{1-\alpha} & \text{if } x \in [0, 1/2 - \alpha/2], \\ \frac{2-2q}{1-\alpha} & \text{if } x \in [1/2 + \alpha/2, 1], \\ 0 & \text{otherwise.} \end{cases} \quad (48)$$

A gap of width  $\alpha$  is present in  $\Omega = \bar{\Omega}$  when  $\alpha \neq 0$ . With reference to the notation adopted in this Section, in this case  $\hat{s} = q$  for any value  $\alpha \in (0, 1)$  and therefore Eq. (47) applies immediately, giving us

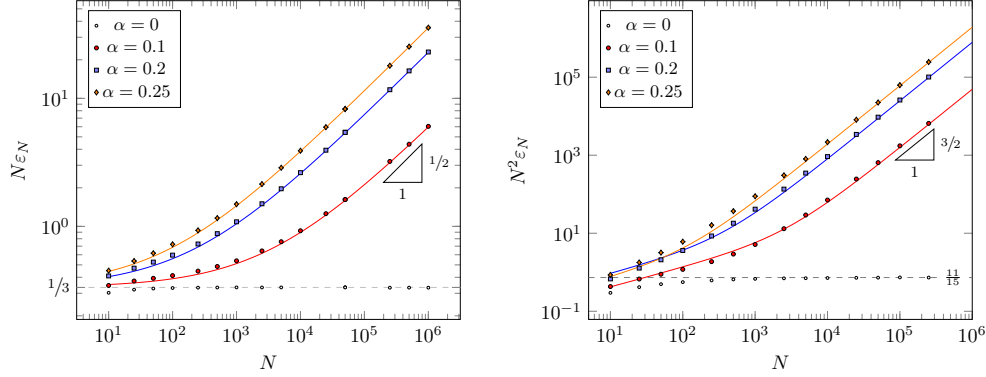
$$\varepsilon_N = 2\alpha^p \sqrt{\frac{q(1-q)}{\pi N}} + o\left(\frac{1}{\sqrt{N}}\right). \quad (49)$$

This scaling law has been numerically verified. In Fig. 4a we consider  $q = 1/2$  and  $p = 2$ , whereas in Fig. 4b we assume  $q = 3/4$  and  $p = 4$ , in both cases with different values of  $\alpha$ . The prediction in Eq. (49) is in perfect agreement with the numerical results.

## 5. Conclusions

We have discussed the REAP on the line with convex weight cost  $w(x, y) = |x - y|^p$  for  $p > 1$ . We have assumed the points to be independently and randomly distributed on the line, according with a PDF  $\varrho(x): \mathbb{R} \rightarrow \mathbb{R}^+$ . We have given a general expression





(a) Numerical results for the AOC in the case of points distributed with PDF given in Eq. (48) with  $q = 1/2$  and  $p = 2$ . The fits are obtained assuming a scaling  $f(N) = \alpha^2 \sqrt{N/\pi} + \epsilon$ , where  $\epsilon$  is a fitting parameter. The dashed line represents the asymptotic limit for  $q = 0$  predicted by Eq. (15).

(b) Numerical results for the AOC in the case of points distributed with PDF given in Eq. (48) with  $q = 3/4$  and  $p = 4$ . The fits are obtained assuming a scaling  $f(N) = 1/4 \sqrt{3/\pi} \alpha^4 N^{3/2} + \epsilon \sqrt{N}$ , where  $\epsilon$  is a fitting parameter. The dashed line represents the asymptotic limit for  $q = 0$  predicted by Eq. (15).

**Figure 4.** AOC in the case of points distributed with the “gapped” PDF given in Eq. (48). Error bars are represented but smaller than the markers.

for the asymptotic AOC  $N^{p/2}\epsilon_N$ . We have shown that this general expression is possibly divergent, due to regions of very low density of points, i.e., to the zeros of  $\varrho(x)$ . We have given a recipe to take into account the effects of the discreteness of the problem when, denoting by  $\Omega = \{x \in \mathbb{R}: \varrho(x) > 0\}$ , the set  $\bar{\Omega} \setminus \Omega$  is made up by isolated points (possibly including the point at infinity). We have then exemplified our approach, applying our recipe to a set of examples, extracting the exact scaling of the asymptotic AOC and, if possible, the coefficients appearing in it. Finally, we have considered the case in which the set  $\Omega$  has gap, i.e., is composed by two disjoint intervals, showing that, in this situation, the effect of fluctuations in the number of points falling in each sub-interval dominates the asymptotic AOC.

The present work extends and completes the analysis given in Refs. [4, 24] for the quadratic REAP. The importance of the discussed regularization is however not restricted to the one-dimensional random assignment. Indeed, the same formulas for the AOC discussed here appear in other one-dimensional random optimization problems, such as the random Euclidean 2-matching [27] and the Traveling Salesman Problem [28] in the bipartite case. The understanding of the proper regularization to be adopted, when the simpler expression cannot be used and an anomalous scaling appears, is therefore relevant for a larger class of optimization problems, to which the analysis presented here can be applied. Finally, in Ref. [4] an integral expression for the asymptotic AOC is given for  $d > 1$ . As the authors stress therein, the higher dimensional case might also require a regularization, depending on the properties of the domain and of the PDF adopted to extract the points. The criteria for such a regularization remain an open problem for future investigations.

## Acknowledgments

The authors are grateful to Luigi Ambrosio for useful discussions and correspondence, and for pointing us to Ref. [20]. The authors also thank Giorgio Parisi, for many discussions on the assignment problem in Euclidean spaces. GS acknowledges the financial support of the Simons foundation (Grant No. 454949, Giorgio Parisi). The work presented in this article was supported by the project “Meccanica statistica e complessità”, a research grant funded by PRIN 2015 (Agreement no. 2015K7KK8L).

## References

- [1] Papadimitriou C and Steiglitz K 1998 *Combinatorial Optimization: Algorithms and Complexity* Dover Books on Computer Science Series (Dover Publications)
- [2] Ajtai M, Komlós J and Tusnády G 1984 *Combinatorica* **4** 259–264
- [3] Caracciolo S, Lucibello C, Parisi G and Sicuro G 2014 *Phys. Rev. E* **90** 012118
- [4] Caracciolo S and Sicuro G 2015 *Phys. Rev. Lett.* **115** 230601
- [5] Ambrosio L, Stra F and Trevisan D 2018 *Probab. Theory Relat. Fields* (Preprint [arXiv:1611.04960](https://arxiv.org/abs/1611.04960))
- [6] Mézard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* Lecture Notes in Physics Series (World Scientific Publishing Company, Inc.)
- [7] Orland H 1985 *J. Phys. Lett.* **46** 763–770
- [8] Mézard M and Parisi G 1985 *J. Phys. Lett.* **46** 771–778
- [9] Mézard M and Parisi G 1987 *J. Phys.* **48** 1451–1459
- [10] Aldous D J 1992 *Probab. Theory Relat. Fields* **93** 507–534
- [11] Linusson S and Wästlund J 2004 *Probab. Theory Relat. Fields* **128** 419–440
- [12] Nair C, Prabhakar B and Sharma M 2005 *Random Struct. Algorithms* **27** 413–444
- [13] Parisi G and Ratiéville M 2002 *Eur. Phys. J. B* **29** 457–468
- [14] Mézard M and Parisi G 1988 *J. Phys.* **49** 2019–2025
- [15] Houdayer J, Boutet de Monvel J H and Martin O C 1998 *Eur. Phys. J. B* **6** 383–393
- [16] Lucibello C, Parisi G and Sicuro G 2017 *Phys. Rev. E* **95**(1) 012302
- [17] Bayati M, Shah D and Sharma M 2008 *IEEE Trans. Inf. Theory* **54** 1241–1251
- [18] Mézard M and Montanari A 2009 *Information, Physics, and Computation* Oxford Graduate Texts (OUP Oxford)
- [19] Villani C 2009 *Optimal transport. Old and new.* (Springer–Verlag)
- [20] Bobkov S and Ledoux M 2016 preprint URL <https://perso.math.univ-toulouse.fr/ledoux/files/2016/12/MEMO.pdf>
- [21] McCann R J 1999 *Proc. R. Soc. A: Math., Phys. Eng. Sci.* **455** 1341–1380
- [22] Boniolo E, Caracciolo S and Sportiello A 2014 *J. Stat. Mech.* **2014** P11023
- [23] Caracciolo S, D’Achille M and Sicuro G 2017 *Phys. Rev. E* **96**(4) 042102
- [24] Caracciolo S and Sicuro G 2014 *Phys. Rev. E* **90** 042112
- [25] Del Barrio E, Giné E and Utzet F 2005 *Bernoulli* **11** 131–189
- [26] Caracciolo S and Sicuro G 2015 *Phys. Rev. E* **91** 062125
- [27] Caracciolo S, Di Gioacchino A and Malatesta E M 2018 Solution for a Bipartite Euclidean 2-matching in one dimension. In preparation.
- [28] Caracciolo S, Di Gioacchino A, Gherardi M and Malatesta E M 2018 (Preprint [arXiv:1802.01545](https://arxiv.org/abs/1802.01545))