

# Optimal transport between determinantal point processes and application to fast simulation

Laurent Decreusefond<sup>a,\*</sup>, Guillaume Moroz<sup>b,1</sup>

<sup>a</sup>*LTCI, Telecom Paris, Institut polytechnique de Paris, Paris, France*

<sup>b</sup>*INRIA Nancy Grand-Est, Nancy, France*

[laurent.decreasefond@mines-telecom.fr](mailto:laurent.decreasefond@mines-telecom.fr) (L. Decreusefond), [guillaume.moroz@inria.fr](mailto:guillaume.moroz@inria.fr) (G. Moroz)

Received: 1 November 2020, Revised: 1 April 2021, Accepted: 10 May 2021,  
Published online: 2 June 2021

**Abstract** Two optimal transport problems between determinantal point processes (DPP for short) are investigated. It is shown how to estimate the Kantorovitch–Rubinstein and Wasserstein-2 distances between distributions of DPP. These results are applied to evaluate the accuracy of a fast but approximate simulation algorithm of the Ginibre point process restricted to a circle. One can now simulate in a reasonable amount of time more than ten thousands points.

## 1 Introduction

Determinantal point processes (DPP) have been introduced in the seventies [24] to model fermionic particles with repulsion like electrons. They recently regained interest since they represent locations of eigenvalues of some random matrices. A determinantal point process is characterized by an integral operator with kernel  $K$  and a reference measure  $m$ . The integral operator is compact and symmetric and thus characterized by its eigenfunctions and its eigenvalues. Following [18], the eigenvalues are not measurable functions of the realizations of the point process so it is difficult to devise how a modification of the eigenfunctions, respectively of the

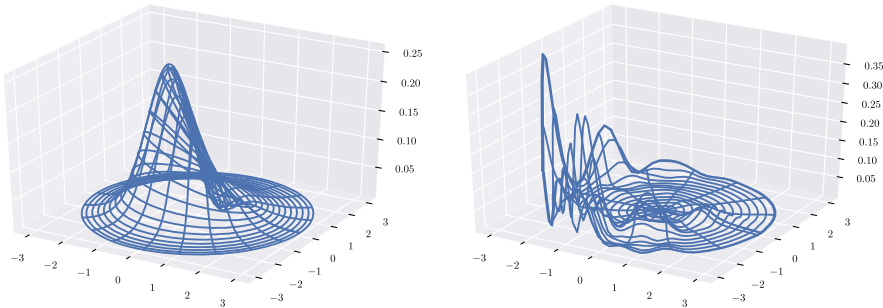
---

\*Corresponding author. This work benefited from fruitful discussions with L. Austruy and F. Maunoury.

<sup>1</sup>Supported by grant ANR-17-CE40-0017 of the French National Research Agency (ANR project AS-PAG). The authors are indebted to the anonymous referees for their very insightful remarks which helped them to improve this paper.

eigenvalues or of the reference measure, modifies the random configurations of a DPP.

A careful analysis of the simulation algorithm given in [18] yields several answers to these questions. For instance, it is clear that the eigenvalues control the distribution of the number of points and the eigenfunctions determine the positions of the atoms once their number is known. The above mentioned algorithm is a beautiful piece of work but requires to draw points according to distributions whose densities are not expressed as combinations of classical functions, hence the necessity to use rejection sampling method. Unfortunately, as the number of drawn points increases, the densities present high peaks and deep valleys which is the most adverse situation for rejection sampling, see Figure 1.



**Fig. 1.** Peaks and valleys of some densities

As a consequence, it is hardly feasible to simulate a DPP with more than one thousand points in less than a few hours. As some DPP arise as locations of eigenvalues of some matrix ensembles, it may seem faster and simpler to draw random matrices and compute their eigenvalues with the optimized libraries to do so. However, there are several drawbacks to this approach: a) we cannot control the domain into which the points fall, and for some applications it may be important to simulate DPP restricted to some compact sets, b) as eigenvalues belong to  $\mathbf{R}$  or  $\mathbf{C}$ , we cannot imagine DPP in higher dimensions with this approach, c) for a stationary DPP, it is often useful to simulate under the Palm measure (see below) which no longer corresponds to a known random matrix ensemble.

Several refinements of Algorithm 1 have been proposed along the years but the most advanced contributions have been made for DPP on lattices, which are of a totally different nature than continuous DPP. We here propose to speed up the simulation of a DPP by reducing the number of eigenvalues considered and approximating the eigenfunctions by functions whose quadrature can be easily inverted to get rid of the rejection part.

We evaluate the impact of these approximations by bounding the distances between the original distribution of the DPP to be simulated and the distribution according to which the points are drawn. The computations of the error margin are specific to the Ginibre DPP but could be done for other processes with radial symmetry like polyanalytic Ginibre ensembles [17] or the Bergman DPP [19].

Actually, there are several notions of distances between the distributions of point processes (see [11] and references therein). We focus here on the total variation distance and on the quadratic Wasserstein distance. The former counts the difference of the number of points in an optimal coupling between two distributions. The latter evaluates the matching distance between two realizations of an optimal coupling provided that it exists.

The paper is organized as follows. We first recall the definition and salient properties of DPP. In Section 3, we briefly introduce the optimal transportation problem in its full generality and give some elements dedicated to point processes. In Section 4, we show how the eigenvalues and eigenfunctions do appear in the evaluation of the distances under scrutiny. In Section 5, we apply these results to the simulation of the Ginibre process.

## 2 Determinantal point processes

Let  $E$  be a Polish space,  $\mathcal{O}(E)$  the family of all nonempty open subsets of  $E$  and  $\mathcal{B}$  denotes the corresponding Borel  $\sigma$ -algebra. In the sequel,  $m$  is a Radon measure on  $(E, \mathcal{B})$ . Let  $\mathfrak{N}_E$  be the space of locally finite subsets of  $E$ , also called the configuration space:

$$\mathfrak{N}_E = \{\xi \subset E : |\Lambda \cap \xi| < \infty \text{ for any compact set } \Lambda \subset E\},$$

equipped with the topology of the vague convergence – the topology generated by the seminorms

$$p_f(\xi) = \left| \sum_{x \in \xi} f(x) \right|$$

for  $f : E \rightarrow \mathbf{R}$  continuous with compact support. This makes  $\mathfrak{N}_E$  a Polish space and we denote by  $\mathcal{F}_{\mathfrak{N}_E}$  its Borelean  $\sigma$ -field (see [20]). We call elements of  $\mathfrak{N}_E$  configurations and identify a locally finite configuration  $\xi$  with the atomic Radon measure  $\sum_{y \in \xi} \delta_y$ , where we have written  $\delta_y$  for the Dirac measure at  $y \in E$ .

Next, let  $\mathfrak{N}_E^f = \{\xi \in \mathfrak{N}_E : \xi(E) < \infty\}$ , the space of all finite configurations on  $E$ . It is naturally equipped with the trace  $\sigma$ -algebra  $\mathcal{F}_{\mathfrak{N}_E^f} = \mathcal{F}|_{\mathfrak{N}_E^f}$ . Note that

$$\mathfrak{N}_E^f = \bigcup_{n=0}^{\infty} \mathfrak{N}_E^{(n)}$$

where

$$\mathfrak{N}_E^{(n)} = \{\xi \in \mathfrak{N}_E^f, \xi(E) = n\}.$$

Since  $\mathfrak{N}_E^{(n)}$  can be identified with  $E^n / \mathfrak{S}_n$  where  $\mathfrak{S}_n$  is the group of permutations over  $n$  elements, every function  $f : \mathfrak{N}_E^f \rightarrow \mathbf{R}$  is in fact equivalent to a family of functions  $(f_n, n \geq 1)$  where  $f_n$  is a symmetric function from  $E^n$  to  $\mathbf{R}$ .

**Definition 1.** A locally finite point process is defined as a probability measure on  $(\mathfrak{N}_E, \mathcal{F}_{\mathfrak{N}_E})$  or as an  $\mathfrak{N}_E$ -valued random variable.

A finite point process is a probability measure on  $(\mathfrak{N}_E^f, \mathcal{F}_{\mathfrak{N}_E^f})$  or an  $\mathfrak{N}_E^f$ -valued random variable.

For  $\Xi_\mu$ , an  $\mathfrak{N}_E^f$ -valued random variable of distribution  $\mu$ , we denote by  $\mu_n$  the distribution of  $\Xi_\mu$  given  $\Xi_\mu(E) = n$ . In view of the previous remark, it can be seen as a probability measure on  $E^n$ , invariant by the action of  $\mathfrak{S}_n$ , which describes the location of the particles given their number.

A random point process  $\mu$  is characterized by its Laplace transform, which is defined for any measurable nonnegative function  $f$  on  $E$  as

$$\mathcal{L}_\mu(f) = \int_{\mathfrak{N}_E} e^{-\sum_{x \in \xi} f(x)} d\mu(\xi).$$

A point process is simple if with probability 1, for any  $x \in E$ ,  $\Xi_\mu(\{x\}) \leq 1$ . In the sequel, we consider only simple point processes so that we remove the word *simple*. Point processes are also often characterized via their correlation functions defined as follows.

**Definition 2.** A point process  $\mu$  is said to have correlation functions  $(\rho_n, n \geq 1)$  if for any  $A_1, \dots, A_n$ , disjoint bounded Borel subsets of  $E$ ,

$$\mathbf{E} \left[ \prod_{i=1}^n \Xi_\mu(A_i) \right] = \int_{A_1 \times \dots \times A_n} \rho_n(x_1, \dots, x_n) dm(x_1) \dots dm(x_n).$$

We see that  $\rho_1$  is the mean density of particles with respect to  $m$ , and

$$\rho_n(x_1, \dots, x_n) dm(x_1) \dots dm(x_n)$$

is the probability of finding at least one particle in the vicinity of each  $x_i$ ,  $i = 1, \dots, n$ .

**Definition 3.** A measure  $\mu$  on  $\mathfrak{N}_E^f$  is regular with respect to the reference measure  $m$  when, for all  $n \geq 1$ , there exist

$$\begin{aligned} j_n &: \mathfrak{N}_E^{(n)} \longrightarrow \mathbf{R}^+, \\ \{x_1, \dots, x_n\} &\longmapsto j_n(x_1, \dots, x_n) \end{aligned}$$

where  $j_n$  is symmetric on  $E^n$  such that for any measurable bounded  $f : \mathfrak{N}_E^f \rightarrow \mathbf{R}$ ,

$$E[f(\Xi_\mu)] = f(\emptyset) + \sum_{n=1}^{\infty} \frac{1}{n!} \int_{E^n} f(x_1, \dots, x_n) j_n(x_1, \dots, x_n) dm(x_1) \dots dm(x_n). \quad (1)$$

The function  $j_n$  is called the  $n$ -th Janossy density. Intuitively, it can be viewed as the probability to have exactly  $n$  points located in the vicinity of  $(x_1, \dots, x_n)$ .

For details about the relationships between correlation functions and Janossy densities, see [7].

### 2.1 Determinantal point processes

For details, we mainly refer to [30]. A (symmetric) determinantal point process on  $X$  is characterized by a kernel  $K$  and a reference measure  $m$ . The map  $K$  is supposed to be a Hilbert–Schmidt operator from  $L^2(E, m)$  into  $L^2(E, m)$  which satisfies the following conditions:

1.  $K$  is a bounded symmetric integral operator on  $L^2(E, m)$ , with kernel  $K(\cdot, \cdot)$ , i.e., for any  $x \in E$ ,

$$Kf(x) = \int_E K(x, y) f(y) dm(y).$$

2. The spectrum of  $K$  is included in  $[0, 1]$ .
3. The map  $K$  is locally trace-class, i.e., for all compact  $\Lambda \subset E$ , the restriction  $K_\Lambda = P_\Lambda K P_\Lambda$  of  $K$  to  $L^2(\Lambda, m)$  is trace-class. Here and hereafter,  $P_\Lambda$  denotes the orthogonal projection from  $L^2(E, m)$  onto  $L^2(\Lambda, m)$ .

**Definition 4.** The determinantal measure on  $\mathfrak{N}_E$  with characteristics  $K$  and  $m$  can be defined through its correlation functions

$$\rho_{n, K}(x_1, \dots, x_n) = \det(K(x_k, x_l))_{1 \leq k, l \leq n},$$

and for  $n = 0$ ,  $\rho_{0, K}(\emptyset) = 1$ .

There is a particular class of DPP which is the building blocks on which general DPP are built upon.

**Definition 5.** A DPP whose spectrum is reduced to the singleton  $\{1\}$  is called a projection DPP. Actually, its kernel is of the form

$$K_\phi(x, y) = \sum_{j=0}^M \phi_j(x) \phi_j(y)$$

where  $M \in \mathbf{N} \cup \{\infty\}$  and  $(\phi_j, j = 0, \dots, M)$  is a family of orthonormal functions of  $L^2(E, m)$ .

If  $M$  is finite, then almost all configurations of such a point process have  $M$  atoms.

Alternatively, when the spectrum of  $K$  does not contain 1, we can define a DPP through its Janossy densities. In this situation, the properties of  $K$  ensure that there exists a sequence  $(\lambda_j, j \geq 0)$  of elements of  $[0, 1)$  with no accumulation point but 0 and a complete orthonormal basis  $(\phi_j, j \geq 0)$  of  $L^2(E, m)$  such that

$$K_\phi(x, y) = \sum_{j \geq 0} \lambda_j \phi_j(x) \phi_j(y).$$

Note that if  $L^2(E, m)$  is a  $\mathbf{C}$ -vector space, we must modify this definition accordingly:

$$K_\phi(x, y) = \sum_{j \geq 0} \lambda_j \phi_j(x) \overline{\phi_j(y)}.$$

For a compact subset  $\Lambda \subset X$ , the map  $J_\Lambda$  is defined by

$$J_\Lambda = (\text{Id} - K_\Lambda)^{-1} K_\Lambda,$$

so that we have

$$(\text{Id} - K_\Lambda) (\text{Id} + J_\Lambda) = \text{Id}.$$

For any compact  $\Lambda \subset E$ , the operator  $J_\Lambda$  is a Hilbert–Schmidt, trace-class operator, whose spectrum is included in  $[0, +\infty)$ . We denote by  $J_\Lambda$  its kernel. For any  $n \in \mathbf{N}$ , any compact  $\Lambda \subset E$ , and any  $(x_1, \dots, x_n) \in \Lambda^n$ , the  $n$ -th Janossy density is given by

$$j_n^\Lambda(x_1, \dots, x_n) = \det (J_\Lambda(x_k, x_l))_{1 \leq k, l \leq n}. \quad (2)$$

We can now state how the characteristics of a DPP are modified by some usual transformations on the configurations.

**Theorem 6.** *Let  $\mu$  be a DPP on  $\mathbf{R}^k$  with kernel  $K$  and reference measure  $m = h \, dx$ . Let  $(\lambda_j, j \geq 0)$  be its eigenvalues counted with multiplicity and  $(\phi_j, j \geq 0)$  the corresponding eigenfunctions.*

1. *A random thinning of probability  $p$  (i.e. we keep each atom with probability  $p$  independently of the others) transforms  $\mu$  into a DPP with kernel  $pK$ .*
2. *A dilation of ratio  $\rho$  transforms  $\mu$  into a DPP of kernel*

$$K_\rho(x, y) = \frac{1}{\rho} K(\rho^{-1/k}x, \rho^{-1/k}y).$$

3. *If  $H$  is a  $C^1$ -diffeomorphism on  $E$ , then*

$$\begin{aligned} \mathcal{H} : \mathfrak{N}_E &\longrightarrow \mathfrak{N}_E, \\ \sum_{x \in \xi} \delta_x &\longmapsto \sum_{x \in \xi} \delta_{H(x)} \end{aligned}$$

*transforms  $\mu$  into a DPP of kernel*

$$K_H(x, y) = K \left( H^{-1}(x), H^{-1}(y) \right)$$

*and reference measure  $m \circ H^{-1}$ , the image measure of  $m$  by  $H$ , see [6].*

4. *If  $K(x, y) = k(x - y)$  then  $\mu$  is stationary [22] and thus admits a stationary Palm measure  $\mu^0$ . From [30, Theorem 6.5], we know that  $\mu^0$  is distributed as the DPP of kernel*

$$K^0(x, y) = K(x, y) - \frac{K(x, 0)K(0, y)}{K(0, 0)}. \quad (3)$$

**Remark 1.** It is straightforward to see that the spectrum of  $K_H$  in  $L^2(E, m \circ H^{-1})$  is the same as the spectrum of  $K$  in  $L^2(E, m)$ . Actually, this transformation will be a particular case of the optimal maps obtained in solving the MKP for the Wassertein-2 distance (see Theorem 17).

**Remark 2.** The Ginibre point process, denoted by  $\mathfrak{G}$ , which will be our toy model, is the stationary DPP with kernel

$$K(x, y) = \frac{1}{\pi} e^{-\frac{1}{2}(|x|^2 + |y|^2)} e^{x\bar{y}}$$

for  $x, y \in \mathbf{C}$ . Expanding the right-most exponential we obtain

$$K(x, y) = \sum_{k=0}^{\infty} \frac{x^k}{\sqrt{\pi} \sqrt{k!}} e^{-|x|^2/2} \frac{\bar{y}^k}{\sqrt{\pi} \sqrt{k!}} e^{-|y|^2/2}. \quad (4)$$

By a change of variables in polar coordinates, we have

$$\begin{aligned} \int_{\mathbf{C}} (x_1 + ix_2)^k (x_1 - ix_2)^m e^{-(x_1^2 + x_2^2)} dx_1 dx_2 &= 2\pi \int_{\mathbf{R}^+} r^{2k+1} e^{-r^2} dr \mathbf{1}_{\{k=m\}} \\ &= \pi k! \mathbf{1}_{\{k=m\}}, \end{aligned}$$

so that (4) gives the orthonormal eigenfunction expansion of  $K$ . If we truncate this expansion to order  $N$ , we obtain the kernel of the point process made by the eigenvalues of complex  $N \times N$  Hermitian matrices with Gaussian entries. This would yield an easy way to simulate an approximation of  $\mathfrak{G}$  limited by the effectiveness of the algorithms which compute the eigenvalues. There are several reasons for which this random matrix based algorithm is not satisfactory, all governed by the practical applications of the Ginibre point process as a model for locations of some items with *repulsiveness*. For instance, in mobile telecommunications systems, antennas cannot be too close to each other if we want to mitigate interference. This proscribes to use Poisson point processes, which are the paragon of processes with no dependency among the particles, as models for their distribution and suggests to use point processes with repulsiveness, like DPP. The first problems with the standard Ginibre point process are that we cannot modify neither the mean number of points per unit of surface, which turns to be  $1/\pi$ , nor *the strength of the repulsiveness*. We then introduce the  $(\lambda, \beta)$ -Ginibre point process, denoted by  $\mathfrak{G}^{\lambda, \beta}$ , whose kernel is given by

$$K_{\lambda, \beta}(x, y) = \frac{\lambda}{\pi} e^{-\frac{\lambda}{2\beta}(|x|^2 + |y|^2)} e^{\frac{\lambda}{\beta} x\bar{y}} \quad (5)$$

where  $\lambda \in \mathbf{R}^+$  and  $\beta \in (0, 1)$ . The mean number per unit of surface of  $\mathfrak{G}^{\lambda, 1}$  is  $\lambda/\pi$  and  $\mathfrak{G}^{\lambda, \beta}$  is obtained by the  $\beta$ -thinning of  $\mathfrak{G}^{\lambda, 1}$  followed by a  $\sqrt{\beta}$  rescaling, so that the mean number of points per unit of surface remains unchanged. We know that  $\mathfrak{G}^{\lambda, \beta}$  converges in distribution to a Poisson point process of intensity  $\lambda/\pi$  when  $\beta$  goes to zero [9]. This means that the more severe the thinning is (i.e.  $\beta$  small), the more the repulsiveness decreases. In [16], it is shown that the locations of antennas in mobile systems of the last generation can be considered as realizations of  $\mathfrak{G}^{\lambda, \beta}$  with  $\lambda$  varying between 1.5 and 3.5 and  $\beta \in [0.6, 1]$ , depending on the operator which deployed the antennas. There is a priori no set of random matrices whose eigenvalues exhibit such a behavior. Furthermore, in practical situations, we are often forced to simulate point processes under their Palm measure. We know from [15], that the Palm measure of  $\mathfrak{G}$  is realized by adding an atom at 0 and removing the atom of  $\mathfrak{G}$  with a

modulus distributed as the absolute value of a Gaussian random variable. On a given realization, this atom is clearly impossible to identify but in view of (3), we see that

$$K^0(x, y) = \frac{1}{\pi} e^{-(|x|^2 + |y|^2)/2} (e^{x\bar{y}} - 1).$$

This amounts to remove the first eigenfunction and to the best of our knowledge, this does not coincide anymore with a DPP arising from a known space of random matrices. All these considerations show how important it is to have a good simulation tool of DPP based solely on the eigenfunction expansion.

### 2.1.1 Simulation of DPP

The simulation algorithm introduced in [18] produces random configurations distributed according to a determinantal point process. It is based on the following lemma.

**Lemma 7.** *Let  $\mu_{K,m}$  be a determinantal point process of a trace-class kernel  $K$  and reference measure  $m$ . Let  $\text{sp}(K; L^2(E, m)) = \{\lambda_j, j \geq 0\}$  and  $(\phi_j, j \geq 0)$  be a CONB of  $L^2(E, m)$  composed of eigenfunctions of  $K$ . Let  $(B(\lambda_j), j \geq 0)$  be a family of independent Bernoulli random variables of respective parameter  $\lambda_j$ . Let*

$$I = \{j \geq 0, B(\lambda_j) = 1\}.$$

Since  $\mathbf{E}[|I|] = \sum_{j=0}^{\infty} \lambda_j = \text{trace}(K) < \infty$ ,  $I$  is a.s. a finite subset of  $\mathbf{N}$ . Consider

$$K_I(x, y) = \sum_{j \in I} \phi_j(x) \phi_j(y)$$

and

$$p_I(x_1, \dots, x_{|I|}) = \frac{1}{|I|!} \det\left(K_I(x_k, x_l), 1 \leq k, l \leq |I|\right).$$

Construct a random configuration  $\xi$  as follows: Given  $I$ , draw points  $(W_1, \dots, W_{|I|})$  with joint density  $p_I$ . Then  $\xi$  is distributed according to  $\mu_{K,m}$ .

In the following, let  $\phi_I(x) = (\phi_j(x), j \in I)$ .

**Data:**  $R, I$

**Result:**  $W_1, \dots, W_{|I|}$

Draw  $W_1$  from the distribution with density  $\|\phi_I(x)\|_{\mathbf{C}^{|I|}}^2 / |I|$ ;

$e_1 \leftarrow \phi_I(W_1) / \|\phi_I(W_1)\|_{\mathbf{C}^{|I|}}$ ;

**for**  $j \leftarrow 2$  **to**  $|I|$  **do**

    Draw  $W_j$  from the distribution with density

$$p_j(x) = \frac{1}{|I| - j + 1} \left( \|\phi_I(x)\|_{\mathbf{C}^{|I|}}^2 - \sum_{k=1}^{j-1} |e_k \cdot \phi_I(x)|^2 \right)$$

$u_j \leftarrow \phi_I(W_j) - \sum_{k=1}^{j-1} e_k \cdot \phi_I(W_j) e_k$ ;

$e_j \leftarrow u_j / \|u_j\|_{\mathbf{C}^{|I|}}$ ;

**end**

**Algorithm 1:** Sampling of the locations of the points given the set  $I$  of active Bernoulli random variables



We have two kinds of difficulties here: drawing of  $W_i$  according to a density function with no particular feature so we usually have to resort to rejection sampling; furthermore, when  $|I|$  is large the computation of the density may be costly as it contains a sum of  $|I|$  terms. Figure 1 also suggests that when the number of points becomes high, the profile of the conditional density might be very chaotic with high peaks and deep valleys, involving a large number of rejections in the sampling of this density.

To illustrate this second difficulty, we simulate  $\mathfrak{G}^{\lambda, \beta}$  restricted to  $B(0, R)$ , which we denote by  $\mathfrak{G}_R^{\lambda, \beta}$ . A nice feature of radially symmetric DPP, like Ginibre, Bergmann or polyanalytic Ginibre point processes is that the eigenfunctions expansion of their restriction to a ball is easily calculated by just renormalizing the eigenvalues and the eigenfunctions. For  $\mathfrak{G}_R^{\lambda, \beta}$ , we obtain

$$\lambda_j = \frac{\gamma(j+1, R^2)}{j!},$$

$$\phi_j(x) = \sqrt{\frac{\lambda}{\pi \gamma(j+1, R^2)}} \left( \sqrt{\frac{\lambda}{\beta}} x \right)^j e^{-\frac{\lambda}{2\beta}|x|^2}.$$

For a simulation of a realization of  $\mathfrak{G}_R^{\lambda, \beta}$ , we define the overhead due to rejections by the ratio of the total number of points drawn and the number of accepted points.

**Table 1.** Average overhead (over 10 runs) due to rejections in the simulation of  $\mathfrak{G}_{10}^{3, \beta}$

$\beta$	1	0.7	0.5	0.25	0.1
Overhead	3.1	4.5	6.1	13	29

These are the problems we intend to address in the following.

**Remark 3.** Note that this algorithm is fully applicable even if  $E$  is a discrete finite space. It has been improved in several ways [21, 31, 14] but when it comes to simulate a DPP with a large number of points as it is necessary in some applications [4], the best way remains to use MCMC methods [1]. By its very construction, this last approach is not applicable when the underlying space  $E$  is continuous, though there exist some approximate algorithms (see [28] and references therein) based on it.

### 3 Distances derived from optimal transport

For details on optimal transport in  $\mathbf{R}^d$  and in general Polish spaces, we refer to [33, 32]. For two Polish spaces  $X$  and  $Y$ , for  $\mu$  (respectively  $\nu$ ) a probability measure on  $X$  (respectively  $Y$ ),  $\Sigma(\mu, \nu)$  is the set of probability measures on  $X \times Y$  whose first marginal is  $\mu$  and second marginal is  $\nu$ . We also need to consider a lower semicontinuous function  $c$  from  $X \times Y$  to  $\mathbf{R}^+$ . The Monge–Kantorovitch problem associated to  $\mu, \nu$  and  $c$ , denoted by  $\text{MKP}(\mu, \nu, c)$  for short, consists in finding

$$\inf_{\gamma \in \Sigma(\mu, \nu)} \int_{X \times Y} c(x, y) d\gamma(x, y). \quad (6)$$

More precisely, since  $X$  and  $Y$  are Polish and  $c$  is l.s.c., it is known from the general theory of optimal transportation, that there exists an optimal measure  $\gamma \in \Sigma(\mu, \nu)$  and that the minimum coincides with

$$\sup_{(F, G) \in \Phi_c} \left( \int_X F \, d\mu + \int_Y G \, d\nu \right),$$

where  $(F, G) \in \Phi_c$  if and only if  $F \in L^1(d\mu)$ ,  $G \in L^1(d\nu)$  and  $F(x) + G(y) \leq c(x, y)$ . We will denote by  $\mathcal{W}_c(\mu, \nu)$  the value of the infimum in (6). In the sequel, we need the following theorem of Brenier [32, Chapter 2] or [25].

**Theorem 8.** *Let  $c(x, y) = 2^{-1}\|x - y\|^2$  be the Euclidean distance on  $\mathbf{R}^k$  and  $\mu, \nu$  two probability measures with finite second moment. If the measure  $\mu$  is absolutely continuous with respect to the Lebesgue measure, there exists a unique optimal measure  $\gamma_{opt}$  which realizes the minimum in (6). Moreover, there exists a unique function  $\psi : \mathbf{R}^k \rightarrow \mathbf{R}$  such that*

$$y = x - \nabla \psi(x), \quad \gamma_{opt}\text{-a.s.}$$

Then we have

$$\mathcal{W}_e(\mu, \nu) = \frac{1}{2} \int_{\mathbf{R}^k} \|\nabla \psi\|_{\mathbf{R}^k}^2 \, d\mu.$$

The square root of  $\mathcal{W}_e(\mu, \nu)$  (the subscript  $e$  stands for Euclidean) defines a distance on  $\mathfrak{M}_1(\mathbf{R}^k)$ , the set of probability measures on  $\mathbf{R}^k$ , called the Wasserstein-2 distance.

When  $X = Y$  and  $c$  is a distance on  $X$ ,  $\mathcal{W}_c$  also defines a distance on  $\mathfrak{M}_1(\mathbf{R}^k)$ , often called Kantorovitch–Rubinstein or Wasserstein-1 distance. It admits the alternative characterization.

**Theorem 9** (See [12, Chapter 11]). *Let  $(X, c)$  be a Polish space. For  $\mu$  and  $\nu$ , two probability measures on  $X$ ,*

$$\text{dist}_{\text{KR}}(\mu, \nu) := \mathcal{W}_c(\mu, \nu) = \sup_{f \in \text{Lip}_1(X, c)} \left( \int_X f \, d\mu - \int_X f \, d\nu \right),$$

where

$$\text{Lip}_1(X, c) = \{f : X \rightarrow \mathbf{R}, \forall x, y \in X, |f(x) - f(y)| \leq c(x, y)\}.$$

#### 4 Distances between point processes

There are several ways to define a distance between point processes on the same underlying space  $E$ . We here focus on two of them. They are constructed similarly: Choose a cost function  $c$  on  $\mathfrak{N}_E$  and then consider  $\mathcal{W}_c$  defined by the solution of MKP( $\mu, \nu, c$ ) for  $\mu$  and  $\nu$ , two elements of  $\mathfrak{M}_1(\mathfrak{N}_E)$ .

**Definition 10.** Consider the distance in total variation  $\text{dist}_{\text{TV}}$  between two configurations (viewed as discrete measures):

$$\text{dist}_{\text{TV}}(\xi, \zeta) = (\xi \Delta \zeta)(E)$$

where  $\xi \Delta \zeta$  is the symmetric difference between the two sets  $\xi$  and  $\zeta$ , i.e. we count the number of distinct atoms between  $\xi$  and  $\zeta$ . Then, for  $\mu$  and  $\nu$  belonging to  $\mathfrak{M}_1(\mathfrak{N}_E)$ ,

their Kantorovitch–Rubinstein distance is defined by

$$\begin{aligned} \text{dist}_{\text{KR}}(\mu, \nu) &= \inf_{\substack{\text{law}(\Xi_\mu)=\mu \\ \text{law}(\Xi_\nu)=\nu}} \mathbf{E} [(\Xi_\mu \Delta \Xi_\nu)(E)] \\ &= \sup_{f \in \text{Lip}_1(\mathfrak{N}_E, \text{dist}_{\text{TV}})} \left( \int_{\mathfrak{N}_E} f(\xi) \, \text{d}\mu(\xi) - \int_{\mathfrak{N}_E} f(\zeta) \, \text{d}\nu(\zeta) \right). \quad (7) \end{aligned}$$

**Remark 4.** By the definition of the total variation, it is straightforward that for any compact set  $\Lambda \subset E$ , the map

$$\begin{aligned} \Omega_\Lambda : (\mathfrak{N}_E, \text{dist}_{\text{TV}}) &\longrightarrow \mathbf{N}, \\ \xi &\longmapsto \xi(\Lambda) \end{aligned}$$

is Lipschitz. Let  $(\mu_n, n \geq 1)$  be a sequence of point processes and denote by  $\Xi_n$  an  $\mathfrak{N}_E$ -valued random variable whose distribution is  $\mu_n$ . Similarly, for another element  $\nu \in \mathfrak{M}_1(\mathfrak{N}_E)$ , let  $\Upsilon$  be an  $\mathfrak{N}_E$ -valued random variable whose distribution is  $\nu$ . In view of (7) and Theorem 9, if  $\text{dist}_{\text{KR}}(\mu_n, \nu)$  tends to zero then for any compact set  $\Lambda$ , the sequence of random variables  $(\Xi_n(\Lambda), n \geq 1)$  converges in distribution to  $\Upsilon(\Lambda)$ .

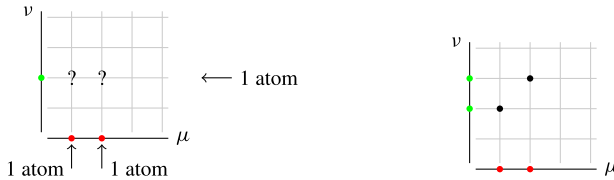
In the sequel, we assume that  $E = \mathbf{R}^k$  and the reference measure  $m$  is the Lebesgue measure on  $\mathbf{R}^k$ . For the quadratic distance, the cost function is the squared Euclidean distance (the  $1/2$  factor is purely cosmetic but traditional)

$$c_e = \frac{1}{2} \|x - y\|^2$$

and we define a cost between configurations (see also [3, 2]) as the lifting of  $c_e$  on  $\mathfrak{N}_E$ : For two configurations  $\xi_1, \xi_2$ , let

$$c(\xi_1, \xi_2) = \inf \left\{ \int c_e(x, y) \, \text{d}\beta(x, y), \beta \in \Sigma(\xi_1, \xi_2) \right\},$$

where  $\Sigma(\xi_1, \xi_2)$  denotes the set of  $\beta \in \mathfrak{N}_{E \times E}$ , i.e. configurations on  $E \times E$ , having marginals  $\xi_1$  and  $\xi_2$ . First remark that when  $\xi_1(E)$  is finite, the cost is finite only if  $\xi_1(E) = \xi_2(E)$ , otherwise  $\Sigma(\xi_1, \xi_2)$  is empty and then, by convention, the cost is infinite.



**Fig. 2.** Cost between two configurations. On the left, there is no feasible coupling as the number of atoms are different. When the two configurations have the same cardinality, there are several possible coupling. One is given on the right

Moreover, the optimum cost is attained at the permutation of  $\{1, \dots, \xi_1(E)\}$  which minimizes the sum of the squared distances

$$c(\xi_1, \xi_2) = \frac{1}{2} \min_{\sigma \in \mathfrak{S}_{\xi_1(E)}} \sum_{j=1}^{\xi_1(E)} \|x_i - y_{\sigma(i)}\|^2$$

where  $\xi_1 = \{x_j, 1 \leq j \leq \xi_1(E)\}$  and  $\xi_2 = \{y_j, 1 \leq j \leq \xi_1(E)\}$ . For infinite configurations, it is not immediate that the cost function so defined has the minimum regularity required to consider an optimal transport problem. According to [29], this is indeed true as  $c$  is lower semicontinuous on  $\mathfrak{N}_E \times \mathfrak{N}_E$ . We can then consider the Monge–Kantorovitch problem MKP( $\mu, \nu, c$ ) on  $\mathfrak{M}_1(\mathfrak{N}_E)$ . The main theorem of [8] is the following (see Definition 1 for the notations).

**Theorem 11.** *Let  $\Lambda \subset E$  be a compact set. Let  $\mu$  be a regular probability measure on  $\mathfrak{N}_\Lambda^f$  and  $\nu$  be a probability measure on  $\mathfrak{N}_E$ . We denote by  $j_n^\mu$  the  $n$ -th order Janossy density of  $\mu$ . Then  $\mu_n$  satisfies*

$$d\mu_n(x_1, \dots, x_n) = \frac{1}{\mu(\zeta(\Lambda) = n)} j_n^\mu(x_1, \dots, x_n) dx_1 \dots dx_n,$$

and is called the normalized Janossy measure. The Monge–Kantorovitch distance, associated to  $c$ , between  $\mu$  and  $\nu$  is finite if and only if the following two conditions hold

1.  $\mu(\zeta(\Lambda) = n) = \nu(\xi(E) = n) := \mathfrak{c}_n$  for any integer  $n \geq 0$ ,
2.  $\sum_{n \geq 1} \mathfrak{c}_n \mathcal{W}_c(\mu_n, \nu_n)$  is finite.

For any  $n \geq 1$ , let  $\varphi_n : \Lambda^{(n)} \rightarrow E^{(n)}$  be the optimal transport map between  $\mu_n$  and  $\nu_n$  given by Theorem 8. Then, for  $\xi \in \mathfrak{N}_\Lambda^{(n)}$ ,

$$\zeta = \varphi_n(\xi), \quad \gamma_{\text{opt-a.s.}}$$

Moreover,

$$\mathcal{W}_c(\mu, \nu) = \sum_{n \geq 1} \mathfrak{c}_n \mathcal{W}_c(\mu_n, \nu_n). \quad (8)$$

We denote by  $\text{dist}_{W_2}$  the Wasserstein-2 distance between  $\mu$  and  $\nu$  defined by

$$\text{dist}_{W_2}(\mu, \nu) = \sqrt{\mathcal{W}_c(\mu, \nu)}.$$

This means that whenever the distance between  $\mu$  and  $\nu$  is finite, there exists a strong coupling which works as follows: 1) draw a discrete random variable with the distribution of  $\xi(\Lambda)$ , and let  $l$  be the obtained value 2) draw the points of  $\xi$  according to  $\mu_l$  and then 3) apply the map  $\varphi_l$  to  $\xi$ . The obtained configuration is distributed according to  $\nu_l$ .

It is shown in [8] that for two Poisson point processes of respective intensities  $\sigma_1$  and  $\sigma_2$ , the distance defined above is finite if and only if  $\sigma_1(E) = \sigma_2(E)$  and

$$\zeta = \sum_{x \in \xi} \delta_{t(x)}, \quad \gamma_{\text{opt-a.s.}}$$

where  $t$  is the optimal transport map between  $\sigma_1/\sigma_1(E)$  and  $\sigma_2/\sigma_2(E)$  for the Euclidean cost as defined in Theorem 8. Note that the optimal map is a transformation which is applied to each atom independently of the others. This amounts to say that

$$\varphi_n(x_1, \dots, x_n) = (t(x_1), \dots, t(x_n)),$$

instead of

$$\varphi_n(x_1, \dots, x_n) = (t_1(x_1, \dots, x_n), \dots, t_n(x_1, \dots, x_n)),$$

in non-Poissonian cases.

#### 4.1 Distances between DPP

For determinantal point processes, we can evaluate the effect of a modification of the eigenvalues with the Kantorovitch–Rubinstein distance and the effect of a modification of the eigenvectors with the Wasserstein-2 distance.

**Lemma 12.** *Let  $\mu$  and  $\nu$  be two determinantal point processes with respective kernels  $K_\mu$  and  $K_\nu$ . Assume that  $K_\mu$  and  $K_\nu$  are two projection kernels in some Hilbert space  $L^2(E, m)$  such that  $K_\mu = K_\nu + L$  where  $L$  is another projection kernel, orthogonal to  $K_\nu$ , i.e.  $\text{im}(L) \subset \ker K_\nu$ . Then*

$$\text{dist}_{\text{KR}}(\mu, \nu) \leq \text{rank}(L). \quad (9)$$

**Proof.** The hypothesis means that there exists  $(\phi_j, j = 0, \dots, n+l)$ , a family of orthonormal functions in  $L^2(E, m)$ , such that

$$K_\nu(x, y) = \sum_{j=0}^n \phi_j(x)\phi_j(y) \text{ and } L(x, y) = \sum_{j=n+1}^{n+l} \phi_j(x)\phi_j(y).$$

Since  $L$  is a positive symmetric operator, this exactly means that  $K_\nu \prec K_\mu$  in the Loewner ordering. According to the Strassen theorem [15, 23], there exists a probability space  $(\Omega', \mathcal{A}', \mathbf{P}')$  on which we can construct three  $\mathfrak{N}_E$ -valued random variables  $\Xi'_\mu$ ,  $\Xi'_\nu$  and  $\Upsilon'$  such that  $\Xi'_\mu$  and  $\Xi'_\nu$  have respective distributions  $\mu, \nu$  and

$$\Xi'_\mu = \Xi'_\nu \cup \Upsilon' \text{ and } \Xi'_\nu \cap \Upsilon' = \emptyset, \mathbf{P}' \text{ a.s.}$$

This implies that

$$\Xi'_\mu \Delta \Xi'_\nu(E) = \Upsilon'(E) = l.$$

According to the first definition of  $\text{dist}_{\text{KR}}$ , see (7), this implies (9).  $\square$

**Theorem 13.** *Let  $\Lambda$  be a compact subset of  $E$  and  $(\phi_j, j \geq 0)$  be a CONB of  $L^2(\Lambda, m)$  and*

$$K_\mu(x, y) = \sum_{j=0}^{\infty} \lambda_n^\mu \phi_j(x)\phi_j(y),$$

$$K_\nu(x, y) = \sum_{j=0}^{\infty} \lambda_n^\nu \phi_j(x)\phi_j(y)$$

where

$$\lambda_j^\nu \leq \lambda_j^\mu, \forall j \geq 0.$$

Let  $\mu$  (respectively  $\nu$ ) be a determinantal point process of characteristics  $K_\mu$  and  $h_\mu$  (respectively  $K_\nu$  and  $h_\nu$ ). Then

$$\text{dist}_{\text{KR}}(\mu, \nu) \leq \sum_{j=0}^{\infty} (\lambda_j^\mu - \lambda_j^\nu). \quad (10)$$

**Proof.** We make a coupling of the Bernoulli random variables which appear in Lemma 7 by using the same sequence of uniform random variables: Let  $(U_j, j \geq 0)$  be a sequence of independent, identically uniformly distributed over  $[0, 1]$ , random variables, and let us consider

$$X_j^\mu = \mathbf{1}_{\{U_j \leq \lambda_j^\mu\}} \text{ and } X_j^\nu = \mathbf{1}_{\{U_j \leq \lambda_j^\nu\}}.$$

Note that

$$\mathbf{P}(X_j^\nu \neq X_j^\mu) = (\lambda_j^\mu - \lambda_j^\nu). \quad (11)$$

Let  $I_\mu = \{j \geq 0, X_j^\mu = 1\}$  and  $I_\nu = \{j \geq 0, X_j^\nu = 1\}$ . In view of the hypothesis,  $X^\nu \leq X^\mu$  hence  $I_\nu \subset I_\mu$ . In other words,  $K_{I_\mu}$  and  $K_{I_\nu}$  are two projection kernels which satisfy the hypothesis of Lemma 12. Hence, there exists a realization  $(\Xi_\mu, \Xi_\nu)$  of  $\Sigma(\mu, \nu)$  given  $I_\mu$  and  $I_\nu$ , such that

$$\text{dist}_{\text{TV}}(\Xi_\mu, \Xi_\nu) = \sum_{j=0}^{\infty} \mathbf{1}_{\{X_j^\nu \neq X_j^\mu\}}.$$

Gluing these realizations together, we get a coupling  $(\Xi_\mu, \Xi_\nu)$  such that

$$\begin{aligned} \mathbf{E}[\text{dist}_{\text{TV}}(\Xi_\mu, \Xi_\nu)] &= \mathbf{E}[\mathbf{E}[\text{dist}_{\text{TV}}(\Xi_\mu, \Xi_\nu) \mid I_\mu, I_\nu]] \\ &= \mathbf{E}\left[\sum_{j=0}^{\infty} \mathbf{1}_{\{X_j^\nu \neq X_j^\mu\}}\right] \\ &= \sum_{j=0}^{\infty} (\lambda_j^\mu - \lambda_j^\nu), \end{aligned}$$

according to (11). Since the Kantorovitch–Rubinstein distance is obtained as the infimum over all couplings of the total variation distance between  $\Xi_\mu$  and  $\Xi_\nu$ , this particular construction shows that (10) holds.  $\square$

The next corollary is an immediate consequence of the alternative definition of the Kantorovitch–Rubinstein distance for point processes, see (7).

**Corollary 14.** *With the hypothesis of Theorem 13, let  $\Xi_\mu$  and  $\Xi_\nu$  be random point processes of respective distributions  $\mu$  and  $\nu$ . Then we have that*

$$\sup_{A \subset \Lambda} \text{dist}_{\text{TV}}(\Xi_\mu(A), \Xi_\nu(A)) \leq \sum_{n=0}^{\infty} (\lambda_n^\mu - \lambda_n^\nu).$$

This means that the Kantorovitch–Rubinstein distance between point processes focuses on the number of atoms in any compact. As we shall see now, the Wasserstein-2 distance evaluates the matching distance between configurations when they have the same cardinality.

**Theorem 15.** *Let  $\mu$  (respectively  $\nu$ ) be a determinantal point process of characteristics  $K_\mu$  and  $h_\mu$  (respectively  $K_\nu$  and  $h_\nu$ ) on a compact set  $\Lambda \subset \mathbf{R}^k$ . The Wasserstein-2 distance between  $\mu$  and  $\nu$  is finite if and only if*

$$sp(K_\mu; L^2(h_\mu dx)) = sp(K_\nu; L^2(h_\nu dx)).$$

**Proof.** If the Wasserstein-2 distance between  $\mu$  and  $\nu$  is finite, then, according to point 1 of Theorem 11, the distributions of  $\Xi_\mu(\Lambda)$  and  $\Xi_\nu(\Lambda)$  coincide. We also know from Lemma 7 that

$$\Phi_\mu(z) = \mathbf{E}_\mu \left[ z^{\Xi_\mu(\Lambda)} \right] = \prod_{\lambda \in sp K_\mu} (1 - \lambda + \lambda z), \quad (12)$$

the infinite product being convergent since  $\text{trace } K_\mu = \sum_{\lambda \in sp K_\mu} \lambda$  is finite. A similar formula holds for  $\Xi_\nu(\Lambda)$ . The zeros of  $\phi_\mu$  are  $(1 - 1/\lambda)$  for  $\lambda \in sp K_\mu$ , counted with multiplicity. Hence the zeros of the holomorphic functions  $\phi_\mu$  and  $\phi_\nu$  appear to be negative and isolated. Let

$$m(\Phi, r) = \text{number of zeros (counted with multiplicity) of } \Phi \text{ in } B(0, r).$$

By the properties of zeros of holomorphic functions we have

$$m(\Phi_\mu, r) = m(\Phi_\nu, r) \text{ for any } r \geq 0.$$

Hence, the zeros of  $\phi_\mu$  and  $\phi_\nu$  coincide and so do the two spectra.

Conversely, if the two spectra coincide, it remains to verify point 2 of Theorem 11. By the very definition of  $\mathcal{W}_e$ ,

$$\mathcal{W}_e(\mu_n, \nu_n) \leq \int_\Lambda \|x\|^2 (d\mu_n + d\nu_n) \leq \sup_{x \in \Lambda} \|x\|^2 (\mu_n(\Lambda) + \nu_n(\Lambda)).$$

Thus, using the notations of Theorem 11, we have

$$\begin{aligned} \sum_{n \geq 1} \mathcal{W}_e(\mu_n, \nu_n) \epsilon_n &\leq \sup_{x \in \Lambda} \|x\|^2 \sum_{n \geq 1} (\mu_n(\Lambda) + \nu_n(\Lambda)) \epsilon_n \\ &\leq (\sup_{x \in \Lambda} \|x\|)^2 (\mathbf{E}[\Xi_\mu(\Lambda)] + \mathbf{E}[\Xi_\nu(\Lambda)]) \\ &= (\sup_{x \in \Lambda} \|x\|)^2 (\text{trace}(K_\mu) + \text{trace}(K_\nu)). \end{aligned}$$

This quantity is finite, hence the Wasserstein-2 distance between  $\mu$  and  $\nu$  is finite as soon as the spectra are equal.  $\square$

The next lemma is a straightforward consequence of Lemma 7.

**Lemma 16.** *Let  $\mu$  be a determinantal point process of characteristics  $K_\mu$  and  $h_\mu$ . For a finite subset  $I$ , let*

$$c_I = \prod_{i \in I} \lambda_i^\mu \prod_{j \in I^c} (1 - \lambda_j^\mu),$$

where  $\lambda_i^\mu$ 's are the eigenvalues of  $K_\mu$ . Then, its  $n$ -th Janossy density is given by

$$j_n^\mu(x_1, \dots, x_n) = \sum_{\substack{I \subset \mathbf{N} \\ |I|=n}} c_I p_I(x_1, \dots, x_n).$$

This means that given  $\Xi_\mu(E) = n$ , the points are distributed according to the probability density

$$\mathfrak{p}_n^\mu : (x_1, \dots, x_n) \mapsto c_n^{-1} \sum_{\substack{I \subset \mathbf{N} \\ |I|=n}} c_I p_I(x_1, \dots, x_n) \text{ where } c_n = \sum_{\substack{I \subset \mathbf{N} \\ |I|=n}} c_I.$$

**Proof.** Consider that  $\Xi_\mu$  is constructed with Algorithm 1 and denote by  $I_{\Xi_\mu}$  the set of indices of the Bernoulli random variables which are equal to 1 for the drawing of  $\Xi_\mu$ . For any bounded  $f : \mathfrak{N}_E^f \rightarrow \mathbf{R}$ ,

$$\begin{aligned} \mathbf{E}[f(\Xi_\mu)] &= f(\emptyset) + \sum_{n=1}^{\infty} \mathbf{E}[f(\Xi_\mu) \mathbf{1}_{\{\Xi_\mu(E)=n\}}] \\ &= f(\emptyset) + \sum_{n=1}^{\infty} \sum_{\substack{J \subset \mathbf{N} \\ |J|=n}} \mathbf{E}[f(\Xi_\mu) | I_{\Xi_\mu} = J] c_J \\ &= f(\emptyset) + \sum_{n=1}^{\infty} \sum_{\substack{J \subset \mathbf{N} \\ |J|=n}} c_J \int_{E^n} f(x_1, \dots, x_n) p_J(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= f(\emptyset) + \sum_{n=1}^{\infty} c_n \int_{E^n} f(x_1, \dots, x_n) \mathfrak{p}_n^\mu(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$

The result follows by identification with (1).  $\square$

Then, Theorem 11 applies as follows.

**Theorem 17.** Suppose that the hypothesis of Theorem 15 holds. Let  $\text{Id} - \nabla \psi_n$  be the optimal transport map between  $\mathfrak{p}_n^\mu$  and  $\mathfrak{p}_n^\nu$ . Then the optimal coupling is given by the following rule: For  $\Xi_\mu$  such that  $\Xi_\mu(E) = n$ , it is coupled with  $\Xi_\nu$ , the configuration with  $n$  atoms described by

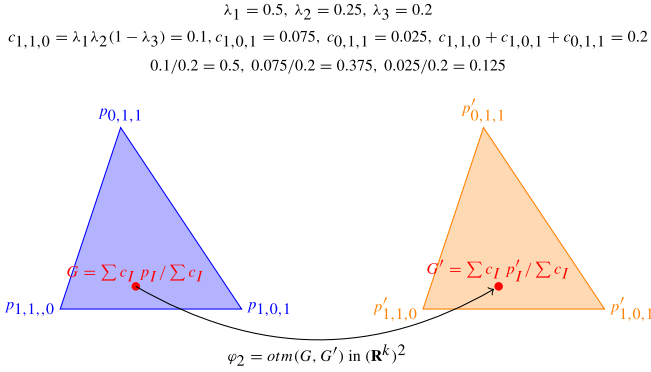
$$\Xi_\nu = \sum_{x \in \Xi_\mu} \delta_{x - \nabla_x \psi_n(\Xi_\mu)}.$$

Furthermore,

$$\begin{aligned} \mathcal{W}_c(\mu, \nu) &= \sum_{n=1}^{\infty} c_n \mathcal{W}_e(\mathfrak{p}_n^\mu, \mathfrak{p}_n^\nu) \\ &= \frac{1}{2} \sum_{n=1}^{\infty} \int_{E^n} \|\nabla \psi_n(x_1, \dots, x_n)\|_{E^n}^2 j_n^\mu(x_1, \dots, x_n) dx_1 \dots dx_n. \end{aligned}$$



Theorems 15 and 17 mean that two determinantal point processes are strongly coupled when and only when their eigenvalues are identical. Moreover, the eigenvalues also control the convex combination of the densities of the projection DPP which appear in the Janossy densities.



**Fig. 3.** For two rank-3 DPP, to compute the optimal map at stage 2 (i.e. for configurations with two points), we compute the coefficient of the convex combination which will be used on both sides to compute the Janossy density. Then we solve the optimal transport problem between these two densities. This gives the map to be applied to configurations with 2 points of the first DPP to obtain the matching configuration in the second DPP

#### 4.2 Determinantal projection processes

Recall from Definition 5 that a projection DPP has a spectrum reduced to  $\{1\}$ . When it is of finite rank  $M$ , almost all its configurations have  $M$  points distributed according to the density

$$p_\phi(x_1, \dots, x_M) = \frac{1}{M!} \det(K_\phi(x_i, x_j), 1 \leq i, j \leq M). \quad (13)$$

Theorem 17 cannot be used as is since a projection DPP do not possess Janossy densities. However, the initial definition of  $\mathcal{W}_c$  can still be used.

**Theorem 18.** Let  $\psi = (\psi_j, 1 \leq j \leq M)$  and  $\phi = (\phi_j, 1 \leq j \leq M)$  be two orthonormal families of  $L^2(E; m)$ . Let  $\mu_\psi$  and  $\mu_\phi$  be the two projection DPP associated to these families. Then

$$\mathcal{W}_c(\mu_\psi, \mu_\phi) \leq \inf_{\sigma \in \mathfrak{S}_M} \sum_{j=1}^M \mathcal{W}_e(|\psi_j|^2 dm, |\phi_{\sigma(j)}|^2 dm).$$

**Proof.** We know that the points of  $\mu_\psi$  (respectively  $\mu_\phi$ ) are distributed according to  $p_\psi$  (respectively  $p_\phi$ ) given by (13). Let  $\gamma$  be a probability measure on  $E^M \times E^M$  whose marginals are  $p_\psi dm$  and  $p_\phi dm$ . We know that

$$\mathcal{W}_c(\mu_\psi, \mu_\phi) = \int_{E^M \times E^M} \inf_{\sigma \in \mathfrak{S}_M} \sum_{j=1}^M |x_j - y_{\sigma(j)}|_E^2 d\gamma(x_1, \dots, x_M, y_1, \dots, y_M)$$

$$\leq \sum_{j=1}^M \int_{E^M \times E^M} |x_j - y_j|_E^2 \, d\gamma(x_1, \dots, x_M, y_1, \dots, y_M).$$

We know from Algorithm 1, that the marginal distribution of a single atom of  $\mu_\psi$  has distribution

$$d\mu_\psi^1(x) = \frac{1}{M} \sum_{j=1}^M |\psi_j(x)|^2 \, dm(x).$$

Since  $p_\psi$  and  $p_\phi$  are both invariant with respect to permutations, we obtain

$$\begin{aligned} \mathcal{W}_c(\mu_\psi, \mu_\phi) &\leq M \int_{E \times E} |x_1 - y_1|_E^2 \, d\gamma(x_1, \dots, x_M, y_1, \dots, y_M) \\ &\leq M \mathcal{W}_e(\mu_\psi^1, \mu_\phi^1). \end{aligned}$$

If  $\gamma_i^1$  is a coupling between  $|\psi_i|^2 \, dm$  and  $|\phi_i|^2 \, dm$  then  $M^{-1} \sum_{i=1}^M \gamma_i^1$  is a coupling between  $\mu_\psi^1$  and  $\mu_\phi^1$ . Hence,

$$\begin{aligned} \mathcal{W}_c(\mu_\psi, \mu_\phi) &\leq \sum_{j=1}^M \int_{E \times E} |x_1 - y_1|^2 \, d\gamma_j^1(x_1, y_1) \\ &\leq \sum_{j=1}^M \mathcal{W}_e(|\psi_j|^2 \, dm, |\phi_j|^2 \, dm). \end{aligned}$$

Since we can arrange the elements of the families  $\psi$  and  $\phi$  in any order, the result follows.  $\square$

## 5 Simulation

In this section, we show how the previous theorems can be applied to give some guarantees when we make an approximate simulation of a Ginibre point process. We will consider Ginibre point processes but our reasoning could be applied to any rotational invariant determinantal process like the polyanalytic ensembles [17, 13] or the Bergman process [19]. For the Ginibre process, which will be our toy model, its restriction to  $\mathcal{B}_R$ , denoted by  $\mathfrak{G}_R^{1,1}$ , has a kernel of the form ([10])

$$K^R(x, y) = \sum_{n=0}^{\infty} \lambda_n^R \phi_n^R(x) \overline{\phi_n^R(y)}$$

where

$$\begin{aligned} \lambda_n^R &= \frac{\gamma(n+1, R^2)}{n!}, \\ \phi_n^R(x) &= \frac{1}{\sqrt{\pi \gamma(n+1, R^2)}} x^n e^{-|x|^2/2}, \end{aligned}$$

with  $\gamma(n, r)$  being the lower incomplete gamma function. We denote by  $\mathfrak{G}_{R,N}^{1,1}$  the truncated Ginibre process whose kernel is the truncation of  $K_R$  to its first  $N$  components:

$$K_N^R(x, y) = \sum_{n=0}^{N-1} \lambda_n^R \phi_n^R(x) \overline{\phi_n^R(y)}.$$

We assume here  $\lambda = 1$  and  $\beta = 1$  only for the sake of simplicity, the computations could be done similarly for any values of  $\lambda$  and  $\beta$ . In the following, we remove the superscript  $(1, 1)$  as  $\lambda$  and  $\beta$  are fixed. The strict application of Algorithm 1 for the simulation of  $\mathfrak{G}_R$ , requires to compute all the quantities of the form

$$\lambda_n^R \prod_{k=n+1}^{\infty} (1 - \lambda_k^R)$$

to determine which Bernoulli random variables are *active*. Strictly speaking, this is unfeasible. However, it is a well-known observation that the number of points of  $\mathfrak{G}_R$  is about  $R^2$ . So it is likely that  $\mathfrak{G}_R$  and  $\mathfrak{G}_{R,N_R}$  should be *close* for  $N_R$  close to  $R^2$ . This is what proves the next theorem.

**Theorem 19.** *Let  $c > 0$  and  $N_R = (R + c)^2$ . For  $R > c$ , we have*

$$\text{dist}_{\text{KR}}(\mathfrak{G}_R, \mathfrak{G}_{R,N_R}) \leq \sqrt{\frac{2}{\pi}} R e^{-c^2}.$$

Actually, the proof says that with high probability,  $\mathfrak{G}_R$  and  $\mathfrak{G}_{R,N_R}$  do coincide.

**Proof.** First, using the integral expression  $\gamma(j, x) = \int_{t=0}^x t^{j-1} e^{-t} dt$ , observe that  $\sum_{j=1}^{\infty} \frac{\gamma(j, x)}{\Gamma(j)} = x$ . Then, using the formula  $\gamma(n+1, x) = n\gamma(n, x) - x^n e^{-x}$ , we have by induction

$$\sum_{j=n+1}^{\infty} \frac{\gamma(j, x)}{\Gamma(j)} = \frac{x^n e^{-x} - (n-x)\gamma(n, x)}{\Gamma(n)}.$$

For  $n = (R + c)^2$  and  $x = R^2$ , this implies

$$\sum_{j \geq (R+c)^2} \lambda_j^R \leq (R+c)^2 \frac{R^{2(R+c)^2} e^{-R^2}}{(R+c)^2!}.$$

Using the bound  $n! \geq \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$ ,

$$\begin{aligned} \sum_{j \geq (R+c)^2} \lambda_j^R &\leq \frac{R+c}{\sqrt{2\pi}} \frac{R^{2(R+c)^2} e^{(R+c)^2 - R^2}}{(R+c)^{2(R+c)^2}} \\ &\leq \frac{R+c}{\sqrt{2\pi}} e^{(R+c)^2 - R^2 - 2(R+c)^2 \log(1 + \frac{c}{R})} \\ &\leq \frac{R+c}{\sqrt{2\pi}} e^{(R+c)^2 - R^2 - 2(R+c)^2 \frac{\frac{c}{R}}{1 + \frac{c}{R}}} \end{aligned}$$

$$\leq \frac{R+c}{\sqrt{2\pi}} e^{-c^2}.$$

Since  $R > c$ , the proof is complete.  $\square$

As a corollary of the previous proof, we have

$$\mathbf{P}(\exists j \geq (R+c)^2, \mathbf{B}(\lambda_j^R) = 1) \leq \sum_{j \geq (R+c)^2} \lambda_j^R \leq \kappa R e^{-c^2} \quad (14)$$

for  $R$  large enough. This means that the number of *active* Bernoulli random variables in Algorithm 1 is less than  $(R+c)^2$  with high probability. We can also provide a lower bound on the cardinality of  $I$ .

**Lemma 20.** *For any  $R > c > 0$ ,*

$$\mathbf{P}(\text{card}(I) < (R-c)^2) \leq \frac{1}{\sqrt{2\pi}} R e^{-c^2}.$$

**Proof.** As in the previous proof, we will reduce the problem to bounding a sum of reduced incomplete gamma functions.

$$\begin{aligned} \mathbf{P}(\text{card}(I) < (R-c)^2) &= 1 - \mathbf{P}(\text{card}(I) \geq (R-c)^2) \\ &\leq 1 - \prod_{0 \leq j < \lfloor (R-c)^2 \rfloor} \mathbf{P}(\mathbf{B}(\lambda_j^R) = 1) \\ &\leq \sum_{0 \leq j < \lfloor (R-c)^2 \rfloor} (1 - \mathbf{P}(\mathbf{B}(\lambda_j^R) = 1)) \\ &\leq \sum_{1 \leq j \leq \lfloor (R-c)^2 \rfloor} \frac{\Gamma(j, R^2)}{\Gamma(j)}. \end{aligned}$$

Using the formula  $\Gamma(n+1, x) = n\Gamma(n, x) + x^n e^{-x}$ , we have by induction

$$\sum_{j=1}^n \frac{\Gamma(j, x)}{\Gamma(j)} = \frac{x^n e^{-x} - (x-n)\Gamma(n, x)}{\Gamma(n)}.$$

For  $n = \lfloor (R-c)^2 \rfloor$  and  $x = R^2$ , this implies

$$\mathbf{P}(\text{card}(I) < (R-c)^2) \leq (R-c)^2 \frac{R^{2(R-c)^2} e^{-R^2}}{(R-c)^{2!}}.$$

Using Stirling formula,

$$\begin{aligned} \mathbf{P}(\text{card}(I) < (R-c)^2) &\leq \frac{R-c}{\sqrt{2\pi}} \frac{R^{2(R-c)^2} e^{(R-c)^2 - R^2}}{(R-c)^{2(R-c)^2}} \\ &\leq \frac{R-c}{\sqrt{2\pi}} e^{(R-c)^2 - R^2 - 2(R-c)^2 \log(1 - \frac{c}{R})} \end{aligned}$$

$$\begin{aligned}
 &\leq \frac{R-c}{\sqrt{2\pi}} e^{\frac{(R-c)^2 - R^2 + 2(R-c)^2 \frac{c}{1-\frac{c}{R}}}{2}} \\
 &\leq \frac{R-c}{\sqrt{2\pi}} e^{-c^2}.
 \end{aligned}$$

The proof is thus complete.  $\square$

The combination of Lemma 20 and (14) shows that the cardinality of  $I$  is of the order of  $R^2$  with high probability.

### 5.1 Inverse transform sampling

The next step of the algorithm is to draw the points according to a density given by a determinant. Since we do not have explicit expression of the inverse cumulative function of these densities, we have to resort to rejection sampling. Fortunately, the particular form of the eigenfunctions of isotropic point processes, like the Ginibre point process, is prone to the simulation of modulus and arguments by inverting their respective cumulative distribution function. The key remark is that along the iterations which are necessary to draw the  $W_i$ , the densities always have the same form:

$$p_i(z) = \sum_{k \in I} a_{k,i} z^{n_k} f_k(|z|), \text{ for } i = 1, \dots, |I|. \quad (15)$$

Without any approximation,  $f_k$  corresponds to the radial component of  $p_I$  and is given by (19). When we approximate the eigenfunctions,  $f_k$  is replaced by  $\tilde{f}_k$  as in (20). This new approach is summarized in Algorithm 2.

**Lemma 21** (Cumulative distribution of the modulus). *Let*

$$p(z) = \sum_{i \in I} a_i z^{n_i} f_i(|z|) \quad (16)$$

and

$$P(r) = \int_{\rho=0}^r \int_{\theta=0}^{2\pi} |p(\rho e^{j\theta})|^2 \rho \, d\rho \, d\theta.$$

The following equality holds:

$$P(r) = \sum_{i \in I} |a_i|^2 F_i(r^2)$$

where

$$F_i(r^2) = \int_{\rho=0}^{r^2} \pi \rho^{n_i} f_i^2(\sqrt{\rho}) \, d\rho.$$

Given a sequence of complex numbers  $W_\ell$  for  $\ell$  from 1 to  $|I|$ , we denote by  $e_\ell$  the orthonormal vectors obtained by the Gram–Schmidt orthonormalization of the vectors  $\phi_\ell^R(W_\ell)$ . Let also  $M_\ell \in \mathbb{R}^{|I|}$  be the vector  $(|e_{\ell,i}|^2)_{i \in I}$  where  $e_{\ell,i}$  is the coordinate of index  $i$  of  $e_\ell$ . Moreover, let  $U_F(r) = (F_i(r))_{i \in I}$ . Finally, let  $U_i$  be the sequence of vectors defined by induction with  $U_1 = (1)_{i \in I}$  and  $U_{i+1} = U_i - M_i$ . Then, using Lemma 21, we observe that the cumulative distribution associated with the density  $p_i$

in Algorithm 1 is the function  $\frac{1}{|I|-i+1}U_i \cdot U_F(r)$ . Thus drawing the modulus of  $W_i$  in Algorithm 1 is reduced to sampling  $c_i$  uniformly in  $[0, 1]$  and solving the equation

$$c_i = \frac{1}{|I|-i+1}U_i \cdot U_F(r). \quad (17)$$

Knowing  $U_i$ , we can compute  $U_{i+1}$  in  $O(|I|)$  arithmetic operations. By construction Equation (17) has only one solution in the interval  $[0, 1]$ , thus using the bisection method ([5]), it can be solved with precision  $\delta$  using  $O(|I| \log(1/\delta))$  evaluation of  $F_i$ .

Given the modulus, we can now simulate the arguments.

**Lemma 22** (Cumulative distribution of the argument). *Let  $p$  be given by (16) and*

$$Q(r, \alpha) = \int_{\theta=0}^{\alpha} |p(re^{i\theta})|^2 r \, d\theta.$$

Then  $Q$  can be rewritten as a sum of  $|I|^2$  terms:

$$Q(r, \alpha) = \sum_{i,k \in I} a_i \bar{a}_k G_{i,k}(r, \alpha)$$

$$\text{where } G_{i,k}(r, \alpha) = \begin{cases} rg_i(r)g_k(r) \frac{e^{j(n_i-n_k)\alpha} - 1}{j(n_i - n_k)} & \text{if } i \neq j \text{ and } g_i(r) = r^{n_i} f_i(r). \\ rg_i^2(r)\alpha & \text{if } i = k \end{cases}$$

Similarly to the simulation of the modulus, for  $\ell$  from 1 to  $|I|$ , let  $A_\ell \in \mathbb{C}^{|I|^2}$  be the vector  $(e_{\ell,i} \bar{e}_{\ell,k})_{i,k \in I}$ . Let  $V_G(r, \alpha) = (G_{i,k}(r, \alpha))_{i,k \in I}$ . Let  $(V_i)_{i=1 \dots |I|-1}$  be the sequence of vectors defined by recurrence with  $V_1 = (\mathbf{1}_{i=k})_{i,k \in I}$  and  $V_{i+1} = V_i - A_i$ . Drawing the argument of  $W_i$  in Algorithm 1 is now reduced to sampling  $c_i$  uniformly in  $[0, 1]$  and solving the equation (i.e. inverting numerically the cdf):

$$c_i = \frac{1}{V_i \cdot V_G(r, 2\pi)} V_i \cdot V_G(r, \alpha). \quad (18)$$

Computing  $V_{i+1}$  from  $V_i$  requires  $O(|I|^2)$  arithmetic operations. Then, for fixed  $r$ , (18) can be solved up to precision  $\delta$  in  $O(|I|^2 + |I| \log \delta)$  arithmetic operations and evaluations of the  $f_i$ , using a dichotomy approach.

The total cost of sampling the  $W_i$  with this approach is  $O(|I|^3 + |I|^2 \log \delta)$  operations. We will see in the next section how we can reduce this complexity using an approximation of the eigenfunctions.

Gathering the results of this section, we get in Algorithm 2 an efficient method to sample points from a symmetric projection point process.

## 5.2 Compact Ginibre and approximation

Using Theorem 18 with a well-chosen approximation, we will show that we can reduce in Algorithm 2 the complexity of steps A. and B. from  $O(|I|^2)$  to  $O(|I|^{1.5})$  operations with high probability. The idea to approximate the eigenfunctions was already present in [27] but it was not used to its whole power. We here approximate the eigenfunctions so that they contain substantially less monomials than the originals.

**Data:**  $R, I$ **Result:**  $W_1, \dots, W_{|I|}$ Draw  $W_1$  from the distribution with density  $\|\phi_I(x)\|_{\mathbb{C}^{|I|}}^2/|I|$ ; $e_1 \leftarrow \phi_I(W_1)/\|\phi_I(W_1)\|_{\mathbb{C}^{|I|}}$ ; $U_1 = (1)_{i \in I}$ ; $V_1 = (\mathbf{1}_{i=k})_{i,k \in I}$ ;**for**  $i \leftarrow 2$  **to**  $|I|$  **do**    A. Update vectors  $U_i$  and  $V_i$  for next point simulation         $M_i \leftarrow (|e_{i,\ell}|^2)_{\ell \in I}$ ;         $A_i \leftarrow (e_{i,k} \overline{e_{i,\ell}})_{k,\ell \in I}$ ;         $U_i \leftarrow U_{i-1} - M_i$ ;         $V_i \leftarrow V_{i-1} - A_i$ ;    B. Draw point  $W_i$         Draw  $c_i$  from the uniform distribution in the interval  $[0, 1]$ ;         $r_i \leftarrow$  solution of  $c_i = \frac{1}{|I|-i+1} U_i \cdot U_F(r)$ ;        Draw  $d_i$  from the uniform distribution in the interval  $[0, 1]$ ;         $\alpha_i \leftarrow$  solution of  $d_i = \frac{1}{V_i \cdot V_G(r_i, 2\pi)} V_i \cdot V_G(r_i, \alpha)$ ;         $W_i \leftarrow r_i e^{i\alpha_i}$ ;    C. Compute new vector  $e_i$          $u_i \leftarrow \phi_I(W_i) - \sum_{k=1}^{i-1} e_k \cdot \phi_I(W_i) e_k$ ;         $e_i \leftarrow u_i / \|u_i\|_{\mathbb{C}^{|I|}}$ ;**end****Algorithm 2:** Simulation of a compact symmetric projection point process restricted to the disc  $\mathcal{B}_R$ 

For a given constant  $c > 0$  and for an integer  $n$ , let  $R_n$  be the ring between the circles of radii  $u_n = \min(R, \sqrt{n} + c)$  and  $l_n = \max(0, \min(\sqrt{n}, R) - c)$ . Let

$$\mu_n = \int_{R_n} |\phi_n^R(z)|^2 dz = \frac{\gamma(n+1, u_n^2) - \gamma(n+1, l_n^2)}{\gamma(n+1, R^2)},$$

$$f_n(|z|) = \frac{1}{\sqrt{\pi} \gamma(n+1, R^2)} e^{-\frac{|z|^2}{2}}. \quad (19)$$

and define the following approximated functions:

$$\tilde{f}_n(|z|) = \begin{cases} f_n(|z|)/\sqrt{\mu_n} & \text{if } z \in R_n, \\ 0 & \text{otherwise,} \end{cases} \quad (20)$$

and let

$$\tilde{\phi}_n^R(z) = z^n \tilde{f}_n(|z|).$$

**Remark 5** (Complexity simplification). For a complex  $z \in B(0, R)$ ,  $z$  belongs to  $R_n$  for integers  $n$  such that

$$(|z| - c)^2 \leq n \leq (|z| + c)^2.$$

There are  $[4c|z|] \leq [4cR]$  such integers. This means that the modified function  $\tilde{U}_F$  contains at most  $4cR$  nonzero coefficients instead of  $R^2$  coefficients in the original expression. The numerical effort to find  $c_i$  and  $d_i$  in Algorithm 2 is reduced accordingly.

We now show that replacing  $\phi_n^R$  by  $\tilde{\phi}_n^R$  does not cost much in terms of accuracy. In the course of the proof, we need the so-called Bakry–Emery–Blower theorem (see [32, Theorem 9.9 and pages 291–292]).

**Theorem 23** (Bakry–Emery–Blower theorem). *Let  $V$  be a  $C^2$  function on  $\mathbf{R}$  with  $\int e^{-V(x)} dx = 1$  and  $V''(x) \geq \lambda > 0$  then the measure  $\rho_\infty$  of density  $e^{-V}$  with respect to the Lebesgue measure satisfies the Talagrand inequality: For any probability measure  $\rho$ ,*

$$\mathcal{W}_e(\rho, \rho_\infty) \leq H(\rho | \rho_\infty) = \int \rho(r) \log \frac{\rho(r)}{\rho_\infty(r)} dr. \quad (21)$$

**Theorem 24.** *For any  $I \subset \{1, \dots, N_R\}$ ,*

$$\mathcal{W}_c(\mu_\phi, \mu_{\tilde{\phi}}) \leq \sum_{j \in I} \log \left( \frac{1}{\mu_j} \right).$$

**Proof.** According to Theorem 18, it is sufficient to evaluate

$$\mathcal{W}_e \left( |\phi_j^R|^2 dx, |\tilde{\phi}_j^R|^2 dx \right)$$

for any  $j \in I$ . Denote the two measures involved in the previous optimal transport problem by

$$\zeta_j(dx) = |\phi_j^R(x)|^2 dx, \quad \tilde{\zeta}_j(dx) = |\tilde{\phi}_j^R(x)|^2 dx.$$

These are two radially symmetric measures on  $\mathbf{R}^2$ . We still denote by  $\zeta$  and  $\tilde{\zeta}$  the two measures they induce on the polar coordinates  $(r, \theta)$ . Consider

$$\zeta_j(dr | \theta) = c_j r^{2j+1} e^{-r^2} \mathbf{1}_{[0, R]}(r) \text{ where } c_j = \frac{1}{\gamma(j+1, R^2)},$$

the distribution of  $r$  given  $\theta$  under  $\zeta$  and the same quantity for  $\tilde{\zeta}$ . If we have a coupling between these two measures, i.e. if we have a measure

$$\gamma_\theta(dr, dr') \in \Sigma \left( \zeta_j(dr | \theta), \tilde{\zeta}_j(dr' | \theta) \right),$$

then the measure

$$\frac{1}{2\pi} \gamma_\theta(dr, dr') d\theta$$

is a coupling between  $\zeta_j$  and  $\tilde{\zeta}_j$ . It follows that

$$\mathcal{W}_e(\zeta_j, \tilde{\zeta}_j) \leq \mathcal{W}_e \left( c_j r^{2j+1} e^{-r^2} \mathbf{1}_{[0, R]}(r) dr, \tilde{c}_j r^{2j+1} e^{-r^2} \mathbf{1}_{R_j}(r) dr \right)$$

where

$$\tilde{c}_j = \frac{1}{\mu_j}.$$

Furthermore, we have

$$-\frac{d^2}{dr^2} \log(r^{2j+1} e^{-r^2}) = \frac{2j+1}{r^2} + 2 \geq 2.$$



Hence the Bakry–Emery–Blower criterion (Theorem 23) is satisfied and the measure

$$\rho_\infty(\mathrm{d}r) = c_j r^{2j+1} e^{-r^2} \mathbf{1}_{[0, R]}(r) \mathrm{d}r$$

satisfies the Talagrand inequality. Apply (21) to

$$\rho_j(\mathrm{d}r) = \tilde{c}_j r^{2j+1} e^{-r^2} \mathbf{1}_{R_j}(r) \mathrm{d}r$$

to obtain

$$\mathcal{W}_e(\rho_j, \rho_\infty) \leq \log \left( \frac{1}{\rho_\infty(R_j)} \right) = \log \left( \frac{1}{\mu_j} \right).$$

The proof is thus complete.  $\square$

Finally, using the same techniques as above, we bound the sum of  $\log \left( \frac{1}{\mu_i} \right)$  in the following lemma.

**Lemma 25.** *There exists a constant  $\kappa$  such that for  $\sqrt{\log R} \leq c \leq R$ :*

$$\sum_{n=0}^{\infty} \log \left( \frac{1}{\mu_n} \right) \leq \kappa R^2 e^{-c^2}.$$

**Proof.** We split the sum in three parts:

$$\begin{aligned} S_1 &= \sum_{n=0}^{(R-c)^2-1} \log \left( \frac{1}{\mu_n} \right), \\ S_2 &= \sum_{n=(R-c)^2}^{R^2-1} \log \left( \frac{1}{\mu_n} \right), \\ S_3 &= \sum_{n=R^2}^{\infty} \log \left( \frac{1}{\mu_n} \right). \end{aligned}$$

We will first prove that the terms in  $S_1$  and  $S_2$  are  $O(e^{-c^2})$  and the terms  $\log \left( \frac{1}{\mu_{R^2+k}} \right)$  in  $S_3$  are  $O(Re^{-c^2} (1 - \frac{1}{R})^k)$ .

For  $c \geq 1$  and  $n \leq (R-c)^2$ , we show that  $\mu_n^{-1}$  is roughly equal to  $\frac{\gamma(n+1, R^2)}{\Gamma(n+1)}$ :

$$\begin{aligned} &\gamma(n+1, u_n^2) - \gamma(n+1, l_n^2) \\ &= \Gamma(n+1) - \Gamma(n+1, u_n^2) - \gamma(n+1, l_n^2) \\ &\geq \Gamma(n+1) - \frac{u_n^{2(n+1)}}{u_n^2 - n - 1} e^{-u_n^2} - \frac{l_n^{2(n+1)}}{n+1 - l_n^2} e^{-l_n^2} \\ &\geq \Gamma(n+1) - \frac{u_n^2 e^{-c^2}}{(u_n^2 - n - 1)\sqrt{2\pi n}} \Gamma(n+1) - \frac{l_n^2 e^{-c^2}}{(n+1 - l_n^2)\sqrt{2\pi n}} \Gamma(n+1) \\ &\geq \Gamma(n+1)(1 - e^{-c^2}). \end{aligned}$$

This implies that

$$\frac{1}{\mu_n} \leq \frac{\gamma(n+1, R^2)}{\Gamma(n+1)(1-e^{-c^2})}$$

and

$$\log\left(\frac{1}{\mu_n}\right) \leq \frac{1}{1-e^{-c^2}} \left( \frac{\Gamma(n+1, R^2)}{\Gamma(n+1)} + e^{-c^2} \right).$$

Thus, for  $c \geq 1$ , we have

$$S_1 \leq 2Re^{-c^2} + 2R^2e^{-c^2}.$$

For  $S_2$  we have  $u_n = R$  and  $l_n = \sqrt{n} - c$  such that

$$\frac{1}{\mu_n} = \frac{1}{1 - \frac{\gamma(n+1, (\sqrt{n}-c)^2)}{\gamma(n+1, R^2)}}.$$

Moreover, for  $n+1 \leq R^2$ , we know that

$$\gamma(n+1, R^2) \geq \frac{\Gamma(n+1)}{2}$$

and

$$\gamma(n+1, (\sqrt{n}-c)^2) \leq e^{-c^2} \Gamma(n+1).$$

Combine these identities with the well-known fact

$$\sum_i \log\left(\frac{1}{1-\epsilon_i}\right) \leq \frac{\sum_i \epsilon_i}{1 - \max \epsilon_i}$$

to obtain

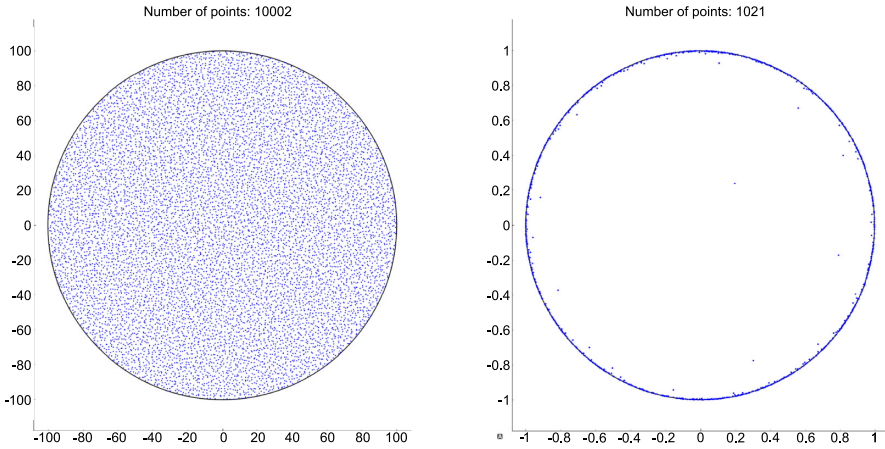
$$S_2 \leq \frac{2cRe^{-c^2}}{1-e^{-c^2}}.$$

Finally for  $S_3$ , we have  $u_n = R$  and  $l_n = R - c$ , so that we have

$$\frac{1}{\mu_n} = \frac{1}{1 - \frac{\gamma(n+1, (R-c)^2)}{\gamma(n+1, R^2)}}.$$

Then remark that

$$\begin{aligned} \frac{\gamma(n, (R-c)^2)}{\gamma(n, R^2)} &\leq \frac{(R-c)^{2n} e^{-(R-c)^2} / (n - (R-c)^2)}{R^{2n} e^{-R^2} / n} \\ &\leq \frac{\left(1 - \frac{c}{R}\right)^{2n} e^{R^2 - (R-c)^2}}{1 - \frac{(R-c)^2}{n}} \\ &\leq \frac{\left(1 - \frac{c}{R}\right)^{2(n-R^2)} e^{-2R^2 \frac{c}{R} - R^2 \frac{c^2}{R^2} + 2cR - c^2}}{1 - \frac{(R-c)^2}{n}} \end{aligned}$$



**Fig. 4.** Simulation of determinantal point processes. Left: kernel  $K(z, w) = \frac{1}{\pi} e^{-\frac{1}{2}(|z|^2 + |w|^2) + z\bar{w}}$ , restricted to the disk of radius 100. Right: kernel  $K(z, w) = \frac{1}{\pi(1-z\bar{w})}$ , restricted to the disk of radius 0.9995

$$\begin{aligned} &\leq \frac{(1 - \frac{c}{R})^{2(n-R^2)} e^{-2c^2}}{1 - \frac{(R-c)^2}{R^2}} \\ &\leq \frac{R}{c} \left(1 - \frac{c}{R}\right)^{2(n-R^2)} e^{-2c^2}. \end{aligned}$$

Thus, summing from  $R^2$  to  $\infty$  we get

$$S_3 \leq \frac{\frac{R^2}{c^2} e^{-2c^2}}{1 - \frac{R}{c} e^{-2c^2}}.$$

The proof is thus complete.  $\square$

### 5.3 Experimental results

An implementation in Python of this algorithm publicly available in [26] allowed us to sample 10 002 points in 2 128 seconds on a 8 core 3Ghz CPU for the Ginibre kernel restricted to a disk of radius 100, shown in Figure 4. In the case of a projection determinantal process, the last matrix  $V$  should be 0 if the vectors  $e_j$  are orthonormal. In our simulation, the norm of  $V$  is  $9.42 \times 10^{-11}$  which is an indicator on the small numerical error that we had with our approximation. For the simulation in a disk of radius 30, leading to roughly 900 points, we ran 10 simulations that all took between 3.2 and 3.4 seconds. The same approach can be used for the DPP with the so-called Bergmann kernel [19]. In this case, the simulation in a disk of radius 0.9995 leads to roughly 1000 points. We ran 10 simulations that all took between 5.9 and 7.3 seconds.

## References

- [1] Anari, N., Gharan, S.O., Rezaei, A.: Monte Carlo Markov chain algorithms for sampling strongly Rayleigh distributions and determinantal point processes. In: 29th Annual Conference on Learning Theory, vol. 49, pp. 103–115 (2016). <http://proceedings.mlr.press/v49/anari16.html>
- [2] Barbour, A.D., Brown, T.C.: The Stein-Chen method, point processes and compensators. *Ann. Probab.* **20**(3), 1504–1527 (1992). [MR1175275](#)
- [3] Barbour, A.D., Maansson, M.: Compound Poisson process approximation. *Ann. Probab.* **30**(3), 1492–1537 (2002). 00027. [MR1920275](#). <https://doi.org/10.1214/aop/1029867135>
- [4] Bardenet, R., Hardy, A.: Monte Carlo with determinantal point processes. *Ann. Appl. Probab.* **30**(1), 368–417 (2020). [MR4068314](#). <https://doi.org/10.1214/19-AAP1504>
- [5] Burden, R.L., Faires, J.D.: Numerical Analysis. Cengage Learning, Boston, MA (2016)
- [6] Camilier, I., Decreusefond, L.: Quasi-invariance and integration by parts for determinantal and permanental point processes. *J. Funct. Anal.* **259**, 268–300 (2010). [MR2610387](#). <https://doi.org/10.1016/j.jfa.2010.01.007>
- [7] Daley, D.J., Vere-Jones, D.: An Introduction to the Theory of Point Processes. Vol. I: Elementary theory and methods, 2nd edn. Probability and its Applications (New York), p. 469. Springer (2003). [MR1950431](#)
- [8] Decreusefond, L.: Wasserstein distance on configurations space. *Potential Anal.* **28**(3), 283–300 (2008). [MR2386101](#). <https://doi.org/10.1007/s11118-008-9077-5>
- [9] Decreusefond, L., Vasseur, A.: Stein’s method and Papangelou intensity for Poisson or Cox process approximation. Working paper or preprint (2018). <https://hal.archives-ouvertes.fr/hal-01832212>
- [10] Decreusefond, L., Flint, I., Vergne, A.: A note on the simulation of the Ginibre point process. *J. Appl. Probab.* **52**(04), 1003–1012 (2015). [1310.0800v2](#). [MR3439168](#). <https://doi.org/10.1239/jap/1450802749>
- [11] Decreusefond, L., Joulin, A., Savy, N.: Upper bounds on Rubinstein distances on configuration spaces and applications. *Commun. Stoch. Anal.* **4**(3), 377–399 (2010). [MR2677197](#). <https://doi.org/10.31390/cosa.4.3.05>
- [12] Dudley, R.M.: Real Analysis and Probability vol. 74. Cambridge University Press (2002). [MR1932358](#). <https://doi.org/10.1017/CBO9780511755347>
- [13] Fenzl, M., Lambert, G.: Precise Deviations for Disk Counting Statistics of Invariant Determinantal Processes. *Int. Math. Res. Not.* (2021). <https://doi.org/10.1093/imrn/rnaa341>
- [14] Gillenwater, J.: Approximate inference for determinantal point processes. PhD thesis, University of Pennsylvania (2014). [MR3321917](#)
- [15] Goldman, A.: The Palm measure and the Voronoi tessellation for the Ginibre process. *Ann. Appl. Probab.* **20**(1), 90–128 (2010). [math/0610243](#). [MR2582643](#). <https://doi.org/10.1214/09-AAP620>
- [16] Gomez, J.-S., Vasseur, A., Vergne, A., Martins, P., Decreusefond, L., Wei, C.: A Case Study on Regularity in Cellular Network Deployment. *Wirel. Commun. Lett.* (2015)
- [17] Haimi, A., Hedenmalm, H.: The polyanalytic Ginibre ensembles. *J. Stat. Phys.* **153**(1), 10–47 (2013). [MR3100813](#). <https://doi.org/10.1007/s10955-013-0813-x>
- [18] Hough, J.B., Krishnapur, M., Peres, Y., Virág, B.: Determinantal processes and independence. *Probab. Surv.* **3**, 206–229 (2006). [MR2216966](#). <https://doi.org/10.1214/154957806000000078>
- [19] Hough, J.B., Krishnapur, M., Peres, Y., Virág, B.: Zeros of Gaussian Analytic Functions and Determinantal Point Processes. University Lecture Series, vol. 51, p. 154. Ameri-

- can Mathematical Society, Providence, RI (2009). [MR2552864](#). <https://doi.org/10.1090/ullect/051>
- [20] Kallenberg, O.: Random Measures, 3rd edn. Academic Press (1983). [MR0818219](#)
- [21] Kulesza, A., Taskar, B.: Determinantal point processes for machine learning. *Found. Trends Mach. Learn.* **5**(2–3), 123–286 (2012). <https://doi.org/10.1561/22000000044>
- [22] Lavancier, F., Møller, J., Rubak, E.: Determinantal point process models and statistical inference. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* **77**(4), 853–877 (2015). [MR3382600](#). <https://doi.org/10.1111/rssb.12096>
- [23] Lindvall, T.: On Strassen’s Theorem on Stochastic Domination. *Electron. Commun. Probab.* **4**, 51–59 (1999). [MR1711599](#). <https://doi.org/10.1214/ECP.v4-1005>
- [24] Macchi, O.: The coincidence approach to stochastic point processes. *Adv. Appl. Probab.* **7**, 83–122 (1975). [MR0380979](#). <https://doi.org/10.2307/1425855>
- [25] McCann, R., Guillen, N.: Five Lectures on Optimal Transportation: Gemetry, Regularity and Applications. CRM Proceedings and Lecture Notes, vol. 56. American Mathematical Society, Providence, Rhode Island (2013). [MR3060502](#). <https://doi.org/10.1090/crmp/056/06>
- [26] Moroz, G.: Determinantal Point Process. Zenodo (2020). <https://doi.org/10.5281/zenodo.4088585>
- [27] Olver, S., Nadakuditi, R.R., Trogdon, T.: Sampling unitary ensembles. *Random Matrices: Theory Appl.* **4**(1), 1550002 (2015). [MR3334666](#). <https://doi.org/10.1142/S2010326315500021>
- [28] Rezaei, A., Gharan, S.O.: A polynomial time MCMC method for sampling from continuous determinantal point processes. In: *Proceedings of the 36th International Conference on Machine Learning*, vol. 97, pp. 5438–5447 (2019). <http://proceedings.mlr.press/v97/rezaei19a.html>.
- [29] Röckner, M., Schied, A.: Rademacher’s theorem on configuration spaces and applications. *J. Funct. Anal.* **169**(2), 325–356 (1999). [MR1730565](#). <https://doi.org/10.1006/jfan.1999.3474>
- [30] Shirai, T., Takahashi, Y.: Random point fields associated with certain Fredholm determinants i: fermion, Poisson and boson point processes. *J. Funct. Anal.* **205**(2), 414–463 (2003). [MR2018415](#). [https://doi.org/10.1016/S0022-1236\(03\)00171-X](https://doi.org/10.1016/S0022-1236(03)00171-X)
- [31] Tremblay, N., Barthelme, P.-O., Amblard, S.: Optimized Algorithms to Sample Determinantal Point Processes. *arXiv e-prints*, 1802–08471 (2018). [1802.08471](#).
- [32] Villani, C.: Topics in Optimal Transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence, RI (2003). [MR1964483](#). <https://doi.org/10.1090/gsm/058>
- [33] Villani, C.: Optimal Transport, Old and New. Lectures Notes in Mathematics. Springer Verlag, New York (2007). [MR2459454](#). <https://doi.org/10.1007/978-3-540-71050-9>