Optimal transport between determinantal point processes and application to fast simulation

Laurent Decreusefond, Guillaume Moroz

OPTIMAL TRANSPORT BETWEEN DETERMINANTAL POINT PROCESSES AND APPLICATION TO FAST SIMULATION

LAURENT DECREUSEFOND AND GUILLAUME MOROZ

Abstract. We analyze several optimal transportation problems between determinantal point processes. We show how to estimate some of the distances between distributions of DPP they induce. We then apply these results to evaluate the accuracy of a new and fast DPP simulation algorithm. We 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 [21] to model fermionic particles with repulsion like electrons. They recently regained interest since they represent the locations of the eigenvalues of some random matrices. A determinantal point process is characterized by an integral operator of kernel K and a reference measure m. The integral operator is compact and symmetric and is 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 eigenvalues or of the reference measure, modifies the random configurations of a DPP. Conversely, it is also puzzling to know how the usual transformations on point processes like thinning, dilations, displacements, translate onto K and m.

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 quickly present high peaks and deep valleys inducing a high number of rejections, see Figure 1.

As a consequence, it is hardly feasible to simulate a DPP with more than one thousand points in a reasonable amount of time. As a DPP appears as the locations of the eigenvalues of some matrix ensembles, it may seem faster and simpler to draw random matrices and compute the eigenvalues with the optimized libraries to do so. There are several drawbacks to this approach: 1) we cannot control the domain into which the points fall, for some applications it may be important to simulate DPP restricted to some compact sets, 2) as eigenvalues belong to $\mathbf{R}$ or $\mathbf{C}$, we cannot imagine DPP in higher dimensions with this approach, 3) for stationary DPP, it is often useful to simulate under the Palm measure (see below) which is known to correspond to the distribution of the initial DPP with the first eigenvalue removed so no longer corresponds to a random matrix ensemble.

2010 *Mathematics Subject Classification.* 60G55,68W25,68W40.

Key words and phrases. Determinantal point processes, Optimal transport, Simulation.

Supported by grant ANR-17-CE40-0017 of the French National Research Agency (ANR project ASPAG).

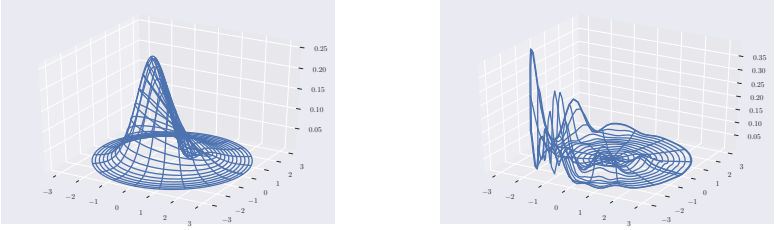

FIGURE 1. Peaks and valleys of some densities.

Several refinements of the 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 DPPs. We here propose to fasten the simulation of a DPP by reducing the number of eigenvalues considered and approximating the eigenfunctions by functions whose quadrature can be easily inversed 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 real distribution according to which the points are drawn.

Actually, there are several notions of distances between the distributions of point processes (see [10] 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 DPPs.

2. Determinantal point processes

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

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

equipped with the topology of the vague convergence. We call elements of $\mathfrak{N}$ configurations and identify a locally finite configuration ξ with the atomic Radon measure $\sum_{y\in\xi} \varepsilon_y$, where we have written ε_y for the Dirac measure at $y \in E$.

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

A random point process is defined as a probability measure on $(\mathfrak{N}, \mathcal{F})$. A random point process μ is characterized by its Laplace transform, which is defined for any measurable non-negative function f on E as

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

Our notations are inspired by those of [14], where the reader can also find a brief summary of many properties of Papangelou intensities.

Definition 1. We define the m-sample measure L on $(\mathfrak{N}^f, \mathcal{F}^f)$ by the identity

$$\int f(\alpha)\,\mathrm{d}L(\alpha) = \sum_{n\geq 0} \frac{1}{n!} \int_{E^n} f(\{x_1,\dots,x_n\})\mathrm{d}m(x_1)\dots\mathrm{d}m(x_n),$$

for any measurable nonnegative function f on $\mathfrak{N}^f$.

Point processes are often characterized via their correlation function defined as:

Definition 2 (Correlation function)**.** A point process μ is said to have a correlation function $\rho : \mathfrak{N}^f \to \mathbf{R}$ if ρ is measurable and

$$\int_{\mathfrak{N}} \sum_{\alpha\subset\xi,\ \alpha\in\mathfrak{N}^f} f(\alpha)\mathrm{d}\mu(\xi) = \int_{\mathfrak{N}^f} f(\alpha)\,\rho(\alpha)\,\mathrm{d}L(\alpha),$$

for all measurable nonnegative functions f on $\mathfrak{N}^f$. For $\xi = \{x_1,\dots,x_n\}$, we will write $\rho(\xi) = \rho_n(x_1,\dots,x_n)$ and call ρ_n the n-th correlation function, where ρ_n is a symmetrical function on E^n.

It can be noted that correlation functions can also be defined by the following property, both characterizations being equivalent in the case of simple point processes.

Definition 3. A point process μ is said to have correlation functions $(\rho_n,\, n\geq 0)$ if for any $A_1,\dots,A_n$ disjoint bounded Borel subsets of E,

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

Recall that ρ_1 is the mean density of particles with respect to m, and

$$\rho_n(x_1,\dots,x_n)\mathrm{d}m(x_1)\dots\mathrm{d}m(x_n)$$

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

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 $f : \mathfrak{N}_E^f \to \mathbf{R}$ is in fact equivalent to a family of symmetric functions $(f_n,\, n\geq 1)$ where f_n goes from E^n to $\mathbf{R}$. For the sake of notations, we omit the index n of f_n.

Definition 4. A measure μ on $\mathfrak{N}_E^f$ is regular with respect to the reference measure m when there exists

$$\begin{aligned} j\ :\ \bigcup_{n=0}^{\infty}\mathfrak{N}_E^{(n)} &\longrightarrow \mathbf{R}^+ \\ \{x_1,\cdots,x_n\} &\longmapsto j_n(x_1,\cdots,x_n) \end{aligned}$$

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

$$E[f(\xi)] = f(\emptyset) + \sum_{n=1}^{\infty}\frac{1}{n!}\int_{E^n} f(x_1,\dots,x_n)j(x_1,\dots,x_n)\mathrm{d}m(x_1)\dots\mathrm{d}m(x_n). \tag{1}$$

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

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

2.1. **Determinantal point processes.** For details, we mainly refer to [24]. A determinantal point on X is characterized by a kernel K and a reference measure m. The map K is supposed to be an 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 $K(.,.)$, i.e., for any $x \in E$,
$$Kf(x) = \int_E K(x,y) f(y) \mathrm{d}m(y).$$
(2) The spectrum of K is included in $[0,\, 1]$.
(3) The map K is locally of 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 of trace class.

Definition 5. The determinantal measure on $\mathfrak{N}$ with characteristics K and m can be defined through its correlation functions:
$$\rho_{n,\,K}(x_1, \cdots,\, x_n) = \det\big(K\left(x_k,\, x_l\right)\big)_{1\leq k,l\leq n}\,,$$
and for $n = 0$, $\rho_{0,\,K}(\emptyset) = 1$.

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

Definition 6. 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 $(\phi_j,\, j = 0, \cdots, M)$ 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 $(\lambda_i,\, i \geq 1)$ of elements of $[0, 1)$ with no accumulation point but 0 and a complete orthonormal basis $(\phi_i,\, i \geq 1)$ of $L^2(m)$ such that
$$K_\phi(x,y) = \sum_{i\geq 1} \lambda_i \phi_i(x)\phi_i(y).$$
Note that if $L^2(E,\, m)$ is a $\mathbf{C}$-vector space, we must modify this definition accordingly:
$$K_\phi(x,y) = \sum_{i\geq 1} \lambda_i \phi_i(x)\overline{\phi_i(y)}.$$
For a compact subset $\Lambda \subset X$, the map J_Λ is defined by:
$$J_\Lambda = \left(\mathrm{Id} - K_\Lambda\right)^{-1} K_\Lambda,$$
so that we have:
$$\left(\mathrm{Id} - K_\Lambda\right)\left(\mathrm{Id} + J_\Lambda\right) = \mathrm{Id}\,.$$
For any compact $\Lambda \subset E$, the operator J_Λ is an Hilbert-Schmidt, trace class operator, whose spectrum is included in $[0, +\infty)$. We denote by J_Λ its kernel. For

any $n \in \mathbf{N}$, any compact $\Lambda \subset E$, and any $(x_1, \cdots, x_n) \in \Lambda^n$, the n-th Janossy density is given by:

$$j_\Lambda^n(x_1, \cdots, x_n) = \det\left(J_\Lambda(x_k, x_l)\right)_{1 \le k,l \le n}. \tag{2}$$

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

Theorem 1. *Let μ a DPP on $\mathbf{R}^k$ with kernel K and reference measure $m = h dx$. Let $(\lambda_n,\, n \ge 0)$ be its eigenvalues counted with multiplicity and $(\phi_n,\, n \ge 0)$ the corresponds eigenfunctions.*

(1) A random thinning of probability p transforms μ into a DPP of kernel pK.

(2) A dilation of ratio ρ transforms μ 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 $\mathcal{C}^1$-diffeomorphism on E, then

$$\mathcal{H} : \mathfrak{N}_E \longrightarrow \mathfrak{N}_E$$
$$\sum_{x\in\xi} \varepsilon_x \longrightarrow \sum_{x\in\xi} \varepsilon_{H(x)}$$

transforms μ 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 μ is stationary [20] *and thus admits a stationary Palm measure μ^0. From* [24]*, we know that μ^0 is distributed as the DPP of kernel*

$$K^0(x,y) = \sum_{n=1}^{\infty} \lambda_n \phi_n(x)\phi_n(y).$$

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 12).

Remark 2. Recall that for a Poisson process, we can obtain a realization of its Palm measure by just adding an atom at 0 to any of its realization. Unfortunately, we know from [18] that the eigenvalues of a DPP cannot be obtained as measurable functions of the configurations. Hence it is hopeless to construct a realization of μ^0 from a realization of μ.

2.1.1. *Simulation of DPP.* The simulation algorithm introduced [18] is up to now the most efficient to produce random configurations distributed according to a determinantal point process. It is based on the following lemma.

Lemma 2. *Let $\mu_{K,m}$ a determinantal point process of a trace-class kernel K and reference measure m. Let $\mathrm{sp}(K; L^2(m)) = \{\lambda_n, i \ge 0\}$ and $(\phi_n,\, i \ge 0)$ a CONB of $L^2(E,\, m)$ composed of eigenfunctions of K. Let $(B(\lambda_n), i \ge 0)$ a family of independent Bernoulli random variables of respective parameter λ_n. Let*

$$I = \{n \ge 0,\, B(\lambda_n) = 1\}.$$

Since $\mathbf{E}\left[|I|\right] = \sum_{n=0}^{\infty} \lambda_n < \infty$, I is a.s. a finite subset of $\mathbf{N}$. Consider

$$K_I(x,y) = \sum_{n\in I} \phi_n(x)\phi_n(y)$$

and

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

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

In the following, let $\phi_I(x) = (\phi_n(x), n \in I)$.
Data: R, I
Result: $W_1, \cdots, W_{|I|}$

Draw W_1 from the distribution with density $\|\phi_I(x)\|^2_{\mathbf{C}^{|I|}}/|I|$;
$e_1 \leftarrow \phi_I(W_1)/\|\phi_I(W_1)\|_{\mathbf{C}^{|I|}}$;
for $i \leftarrow 2$ **to** $|I|$ **do**
  Draw W_i from the distribution with density

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

  $u_i \leftarrow \phi_I(W_i) - \sum_{k=1}^{i-1} e_k.\phi_I(W_i)\; e_k$;
  $e_i \leftarrow u_i/\|u_i\|_{\mathbf{C}^{|I|}}$;
end

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

We have two kind of difficulties here: the drawing of W_i according to a density function with no particular feature so we usually have to resort to rejection sampling; 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. 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 [19, 25] but when it comes to simulate a DPP with a large number of points as it is necessary in some applications [5], the best way remains to use MCMC methods [1]. Unfortunately, by its very construction, this last approach is not feasible when the underlying space E is continuous.

3. Distances derived from optimal transport

For details on optimal transport in $\mathbf{R}^d$ and in general Polish spaces, we refer to [27, 26]. For X and Y two Polish spaces, for μ (respectively ν) a probability measure on X (respectively Y), $\Sigma(\mu, \nu)$ is the set of probability measures on $X \times Y$ whose first marginal is μ and second marginal is ν. We also need to consider a lower semi continuous function c from $X \times Y$ to $\mathbf{R}^+$. The Monge-Kantorovitch problem associated to μ, ν and c, denoted by MKP(μ, ν, c) for short, consists in finding

$$\inf_{\gamma \in \Sigma(\mu,\, \nu)} \int_{X \times Y} c(x, y) \mathrm{d}\gamma(x,\, y). \tag{3}$$

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

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

where (F, G) are such that $F \in L^1(\mathrm{d}\mu)$, $G \in L^1(\mathrm{d}\nu)$ and $F(x) + G(y) \leq c(x, y)$. We will denote by $\mathcal{W}_c(\mu, \nu)$ the value of the infimum in (3). In the sequel, we need the following theorem of Brenier:

Theorem 3. *Let $c(x, y) = 2^{-1}\|x - y\|^2$ be the Euclidean distance on $\mathbf{R}^k$ and μ, ν two probability measures with finite second moment. If the measure μ is absolutely continuous with respect to the Lebesgue measure, there exists a unique optimal measure γ_{opt} which realizes the minimum in (3). Moreover, there exists a unique function $\psi : \mathbf{R}^k \to \mathbf{R}$ such that*

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

Then, we have

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

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

For c a distance on $X = Y$, $\mathcal{W}_c$ 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 4 (See [11])**.** *Let c be a distance on the Polish space (X, d_X). For μ and μ two probability measures on X,*

$$\mathcal{W}_c(\mu, \mu) = \sup_{f \in \mathrm{Lip}_1} \left(\int_X f \, d\mu - \int_X f \, d\nu \right)$$

where

$$\mathrm{Lip}_1 = \{ f : X \to \mathbf{R}, \ \forall x, y \in X, |f(x) - f(y)| \leq d_X(x, y) \}.$$

Note that d_X is the distance which defines the topology of the Polish space X, it may not be identical to c.

The next result is found in [26, Chapter 7].

Theorem 5. *The topologies induced by $\mathcal{W}_c$ and $\mathcal{W}_e$ on $\mathfrak{M}_1(\mathbf{R}^k)$ are strictly stronger than the topology of convergence in distribution.*

4. Distances between point processes

There are several ways to define a distance between point processes. 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 $\mathrm{MKP}(\mu, \nu, c)$ for μ and ν two elements of $\mathfrak{M}_1(\mathfrak{N}_E)$.

Definition 7. Consider $\mathrm{dist}_{\mathrm{TV}}$ the distance in total variation between two configurations (viewed as discrete measures):

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

where $\xi\Delta\zeta$ is the symmetric difference between the two sets ξ and ζ, i.e. we count the number of distinct atoms between ξ and ζ. Then, for μ and ν belonging to $\mathfrak{M}_1(\mathfrak{N}_E)$, their Kantorovitch-Rubinstein distance is defined by

$$\begin{aligned} \mathcal{W}_{\mathrm{KR}}(\mu, \nu) &= \inf_{\substack{\mathrm{law}(\xi)=\mu \\ \mathrm{law}(\zeta)=\nu}} \mathbf{E}\left[(\xi \Delta \zeta)(E)\right] \\ &= \sup_{f \in \mathrm{Lip}_1(\mathfrak{N}_E)} \left(\int_{\mathfrak{N}_E} f(\xi) \mathrm{d}\mu(\xi) - \int_{\mathfrak{N}_E} f(\zeta) \mathrm{d}\nu(\zeta) \right). \end{aligned} \tag{4}$$

Remark 4. For any compact set $\Lambda \subset E$, the map

$$\begin{aligned} \Xi_\Lambda \,:\, \mathfrak{N}_E &\longrightarrow \mathbf{N} \\ \xi &\longmapsto \xi(\Lambda) \end{aligned}$$

is Lipschitz. Let $(\mu_n,\, n \geq 1)$ be a sequence of point processes and denote by ξ_n an $\mathfrak{N}_E$-valued random variable whose distribution is μ_n. Similarly, for another element $\nu \in \mathfrak{M}_1(\mathfrak{N}_E)$, let ζ be an $\mathfrak{N}_E$-valued random variable whose distribution is ν. In view of (4) and Theorem 4, if $\mathcal{W}_{\mathrm{KR}}(\mu_n, \nu)$ tends to zero then for any compact set Λ, the sequence of random variables $(\xi_n(\Lambda),\, n \geq 1)$ converges in distribution to $\zeta(\Lambda)$.

For the quadratic distance, we first consider, on $E = \mathbf{R}^k$, the cost function as $\rho(x,y) = 2^{-1}\|x-y\|^2$ and we define a cost between configurations (see also [4, 3, 2]) as the 'lifting' of ρ on $\mathfrak{N}_E$:

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

where $\Sigma(\xi_1,\xi_2)$ denotes the set of $\beta \in \mathfrak{N}_{E\times E}$ having marginals ξ_1 and ξ_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. Moreover, the cost is attained at the permutation of $\{1,\cdots,\xi_1(E)\}$ 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 [23], this is indeed true as c is lower semi continuous on $\mathfrak{N}_E \times \mathfrak{N}_E$. We can then consider the Monge-Kantorovitch problem $\mathrm{MKP}(\mu,\nu,c)$ on $\mathfrak{M}_1(\mathfrak{N}_E)$. The main theorem of [8] is the following. For Λ a compact subset of E, by definition of locally finite point process, the number of points of $\xi_{|\Lambda}$ is finite hence we can write

$$\mathfrak{N}_\Lambda = \bigcup_{n=0}^{\infty} \mathfrak{N}_\Lambda^{(n)}$$

where

$$\mathfrak{N}_\Lambda^{(n)} = \{\xi \in \mathfrak{N}_\Lambda,\ \xi(\Lambda) = n\}\,.$$

Definition 8. A probability measure μ on $\mathfrak{N}_E$ is said to be regular whenever it admits Janossy densities of any order.

Theorem 6. *Let $\Lambda \subset E$ a compact set. Let μ be a regular probability measure on $\mathfrak{N}_\Lambda$ and ν be a probability measure on $\mathfrak{N}_E$. The Monge-Kantorovitch distance, associated to c, between μ and ν 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) $\sum_{n\geq 1} \mathfrak{c}_n\, \mathcal{W}_e(\mu_n,\, \nu_n)$ is finite.

Then, the solution of $\mathrm{MKP}(\mu,\nu,c)$ *is attained at a unique point γ_{opt} and there exists a unique map*

$$\begin{aligned} \varphi \,:\, \bigcup_{n=0}^{\infty} \mathfrak{N}_\Lambda^{(n)} &\longrightarrow \bigcup_{n=0}^{\infty} \mathfrak{N}_E^{(n)} \\ \xi = \{x_1,\cdots,x_n\} &\longmapsto \{\varphi_n\big(y;\,(x_1,\cdots,x_n)\big),\ y \in \xi\} \in E^n \end{aligned}$$

such that for $\xi \in \mathfrak{N}_\Lambda^{(n)}$,

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

Moreover,

$$\mathcal{W}_c(\mu,\,\nu) = \sum_{n\geq 1} \mathfrak{c}_n \mathcal{W}_e(\mu_n,\,\nu_n). \tag{5}$$

If ν *is regular, then for* $y \in \{x_1, \cdots, x_n\}$,

$$\varphi_n(y, (x_1, \cdots, x_n)) = y - \nabla_y \psi_n(x_1, \cdots, x_n)$$

where $\mathrm{Id} - \nabla\psi_n$ *is the optimal transportation map between the n-th Janossy normalized densities* $\mathfrak{c}_n^{-1}\, j_n^\mu$ *and* $\mathfrak{c}_n^{-1}\, j_n^\nu$.

This means that whenever the distance between μ and ν is finite, there exists a strong coupling which works as follows: 1) draw a discrete random variable with the distribution of $\xi(\Lambda)$, let ι the obtained value 2) draw the points of ξ according to μ_ι and then 3) apply the map $\varphi_\iota(.,\,\xi)$ to each point of ξ. The configuration which is obtained is distributed according to ν_ι.

It is shown in [8] that for two Poisson point processes of respective intensity σ_1 and σ_2, the distance defined above is finite if and only if $\sigma_1(E) = \sigma_2$ and

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

where t is the optimal transport map between σ_1 and σ_2 for the Euclidean cost as defined in Theorem 3. Note that the optimal map is a transformation which is applied to each atom irrespectively of the others. In full generality, for non Poisson processes, the amount by which an atom is moved depends on the other locations.

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 7. *Let* μ *and* ν *two determinantal point processes with respective kernels* K_μ *and* K_ν. *Assume that* K_μ *and* K_ν *are two projection kernels in some Hilbert space* $L^2(m)$ *such that* $K_\mu = K_\nu + L$ *where* L *is another projection kernel and* L *is orthogonal to* K_ν. *Then,*

$$\mathcal{W}_{KR}(\mu,\,\nu) \leq \mathit{range}(L). \tag{6}$$

Proof. The hypothesis means that there exists $(\phi_j,\, j = 1, \cdots, l+n)$ a family of orthonormal functions in $L^2(m)$ such that

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

Since L is a positive symmetric operator, this exactly means that $K_\nu \prec K_\mu$ in the Loewner sense. According to [15], there exists ξ', ζ' of respective distribution μ, ν and a point process ω' such that

$$\xi' = \zeta' + \omega' \text{ and } \zeta' \cap \omega' = \emptyset.$$

This implies that

$$\xi' \Delta \zeta'(E) = \omega'(E) = l.$$

According to the first definition of $\mathcal{W}_{\mathrm{KR}}$, see (4), this implies (6). □

Theorem 8. *Let μ (respectively ν) be a determinantal point process of characteristics K_μ and h_μ (respectively K_ν and h_ν) on a compact set $\Lambda \subset \mathbf{R}^k$. Denote by $(\lambda_n^\mu,\, n \geq 0)$ (respectively $(\lambda_n^\nu,\, n \geq 0)$) the eigenvalues of K_μ in $L^2(E,\, h_\mu dx)$ (respectively of K_ν in $L^2(E,\, h_\nu dx)$) counted with multiplicity and ranked in decreasing order. Assume that*

$$\lambda_n^\nu \leq \lambda_n^\mu,\ \forall n \geq 0.$$

Then,

$$\mathcal{W}_{KR}(\mu,\, \nu) \leq \sum_{n=0}^{\infty} |\lambda_n^\mu - \lambda_n^\nu|. \tag{7}$$

Proof. We make a coupling of $(B(\lambda_n^\mu),\, n \geq 0)$ and $(B(\lambda_n^\mu),\, n \geq 0)$ by using the same sequence of uniform random variables: Let $(U_n,\, n \geq 0)$ be a sequence of independent, identically uniformly distributed over $[0, 1]$, random variables, consider

$$X_n^\mu = \mathbf{1}_{\{U_n \leq \lambda_n^\mu\}} \text{ and } X_n^\nu = \mathbf{1}_{\{U_n \leq \lambda_n^\nu\}}.$$

Note that

$$\mathbf{P}(X_n^\nu \neq X_n^\mu) = |\lambda_n^\mu - \lambda_n^\nu|. \tag{8}$$

Let $I_\mu = \{n \geq 0, X_n^\mu = 1\}$ and $I_\nu = \{n \geq 0, X_n^\nu = 1\}$. In view of the hypothesis, $X^\nu \leq X^\mu$ hence $I_\nu \subset I_\mu$. Otherwise stated, K_{I_μ} and K_{I_ν} are two projection kernels which satisfy the hypothesis of Lemma 7. Hence, there exists a realization (ξ, ζ) of $\Sigma(\mu, \nu)$ given I_μ and I_ν, such that

$$\mathrm{dist}_{\mathrm{TV}}(\xi,\, \zeta) = \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n^\nu \neq X_n^\mu\}}.$$

Gluing these realizations together, we get a coupling (ξ, ζ) such that

$$\begin{aligned}
\mathbf{E}\left[\mathrm{dist}_{\mathrm{TV}}(\xi,\, \zeta)\right] &= \mathbf{E}\left[\mathbf{E}\left[\mathrm{dist}_{\mathrm{TV}}(\xi,\, \zeta) \,|\, I_\mu, I_\nu\right]\right] \\
&= \mathbf{E}\left[\sum_{n=0}^{\infty} \mathbf{1}_{\{X_n^\nu \neq X_n^\mu\}}\right] \\
&= \sum_{n=0}^{\infty} |\lambda_n^\mu - \lambda_n^\nu|,
\end{aligned}$$

according to (8). Since the Kantorovitch-Rubinstein distance is obtained as the infimum over all couplings of the total variation distance between ξ and ζ, this particular construction shows that (7) holds. □

The next corollary is an immediate consequence of the alternative definition of the KR distance on point processes, see Eqn. (4).

Corollary 9. *With the hypothesis of Theorem 8, let ξ and ζ be random point process of respective distribution μ and ν. Then, we have that*

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

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 10. *Let μ (respectively ν) be a determinantal point process of characteristics K_μ and h_μ (respectively K_ν and h_ν) on a compact set $\Lambda \subset \mathbf{R}^k$. The Wasserstein-2 distance between μ and ν is finite if and only if*

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

Proof of Theorem 10. According to point 1 of Theorem 6, we must first prove the equality of the spectra. We already know that the eigenvalues of both kernels are between 0 and 1, with no other accumulation point than 0. Furthermore, the distribution of $\zeta(\Lambda)$ is that of the sum of independent Bernoulli random variables of parameters given by the eigenvalues, hence

$$\Phi_\mu(z) = \mathbf{E}_\mu\left[z^{\zeta(\Lambda)}\right] = \prod_{\lambda\in\operatorname{sp} K_\mu}(1-\lambda+\lambda z). \tag{9}$$

The infinite product is convergent since $\operatorname{trace} K_\mu = \sum_{\lambda\in\operatorname{sp} K_\mu}$ is finite.

If the Wasserstein-2 distance between μ and ν is finite then $\Phi_\mu = \Phi_\nu$. The zeros of these two holomorphic functions are all greater than 1 and are 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,

$$\left\{\frac{1-\lambda}{\lambda},\ \lambda\in\operatorname{sp} K_\mu\right\} = \left\{\frac{1-\lambda}{\lambda},\ \lambda\in\operatorname{sp} K_\nu\right\}$$

and the two spectra must coincide. Now then, by the very definition of $\mathcal{W}_e$,

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

Thus, we have

$$\begin{aligned}\sum_{n\geq 1}\mathcal{W}_e(\mu_n,\,\nu_n)\ \mu(\zeta(\Lambda)=n) &\leq \sup_{x\in\Lambda}\|x\|^2\sum_{n\geq 1}\Big(\mu_n(\Lambda)+\nu_n(\Lambda)\Big)\mu(\zeta(\Lambda)=n)\\ &= 2\sup_{x\in\Lambda}\|x\|^2\operatorname{trace} K_\mu.\end{aligned}$$

This quantity is finite hence the Wasserstein-2 distance between μ and ν as soon as the spectra are equal. □

The next lemma is a straightforward consequence of Lemma 2.

Lemma 11. *Let μ be a determinantal point process of characteristics K_μ and h_μ. For I a finite subset, let*

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

where the λ_i^μ's are the eigenvalues of K_μ. Then, its n-th Janossy density is given by

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

This means that given $\zeta(E)=n$, the points are distributed according to the probability measure:

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

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

$$\begin{aligned}\mathbf{E}\left[f(\xi)\right] &= f(\emptyset) + \sum_{n=1}^{\infty} \mathbf{E}\left[f(\xi)\mathbf{1}_{\{\xi(E)=n\}}\right] \\ &= f(\emptyset) + \sum_{n=1}^{\infty} \sum_{\substack{J\subset\mathbf{N}\\|J|=n}} \mathbf{E}\left[f(\xi)\,|\,I_\xi = J\right]\, c_J \\ &= f(\emptyset) + \sum_{n=1}^{\infty} \sum_{\substack{J\subset\mathbf{N}\\|J|=n}} c_J \int_{E^n} f(x_1,\cdots,x_n)\, p_J(x_1,\cdots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n \\ &= f(\emptyset) + \sum_{n=1}^{\infty} \mathfrak{c}_n \int_{E^n} f(x_1,\cdots,x_n)\, \mathfrak{p}_n^\mu(x_1,\cdots,x_n)\mathrm{d}x_1\ldots\mathrm{d}x_n.\end{aligned}$$

The result follows by identification with (1). □

Then, Theorem 6 applies as follows.

Theorem 12. *Suppose that the hypothesis of Theorem 10 hold. Let* $\mathrm{Id} - \nabla\psi_n$ *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* ξ *such that* $\xi(E) = n$*, it is coupled with* ζ *the configuration with* n *atoms described by*

$$\zeta = \sum_{x\in\xi} \varepsilon_{x-\nabla_x\psi_n(\xi)}.$$

Furthermore,

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

Theorems 10 and 12 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.

4.2. **Determinantal projection processes.** Recall from Definition 6 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

$$p_\phi(x_1,\cdots,x_M) = \frac{1}{M!}\det\Big(K_\phi(x_i,\, x_j),\, 1\le i,j\le M\Big). \tag{10}$$

Theorem 12 cannot be used as is since projection DPPs do not possess Janossy densities. However, the initial definition of $\mathcal{W}_c$ can still be used.

Theorem 13. *Let* $\psi = (\psi_j,\, 1\le j\le M)$ *and* $\psi = (\psi_j,\, 1\le j\le M)$ *two orthonormal families of* $L^2(E;m)$*. Let* μ_ψ *and* μ_ϕ *the two projection DPP associated to these families. Then,*

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

Proof. We know that the points of μ_ψ (respectively μ_ϕ) are distributed according to p_ψ (respectively p_ϕ) given by (10). Let γ be a probability measure on $E^M \times E^M$ whose marginals are $p_\psi \mathrm{d}m$ and $p_\phi \mathrm{d}m$. We know that

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

We know from Algorithm 1, that the marginal distribution a single atom of μ_ψ has distribution

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

Since p_ψ and p_ϕ are both invariant with respect to permutations, we obtain

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

If γ_i^1 is a coupling between $|\psi_i|^2 \mathrm{d}m$ and $|\phi_i|^2 \mathrm{d}m$ then $M^{-1} \sum_{i=1}^M \gamma_i^1$ is a coupling between μ_ψ^1 and μ_ϕ^1. Hence,

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

Since we can order the elements of the family ψ and ϕ in any order, the result follows. □

5. Simulation

As mentioned previously, implementing Algorithm 1 using rejection sampling involves too many rejections which prevents the algorithm to work for more than 1 000 points. In this section, we will show that we can generate 10 000 points for the Ginibre point process on a compact disc using inverse transform sampling and approximation of the kernel.

In this section, we will consider Ginibre point processes but our reasoning could be applied to any rotational invariant determinantal process like the polyanalytic ensembles [16, 13] or the Bergman process [17]. For these processes, it is relatively easy to compute the eigenvalues and the eigenfunctions of the kernel of their restriction to a ball centered at the origin. For the Ginibre process, which will be our toy model, its restriction to $\mathcal{B}_R$, denoted by $\mathfrak{G}^R$, has a kernel of the form

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

where [9]

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

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

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

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 $\mathfrak{G}_{N_R}^R$ should be *close* for N_R close to R^2. This is what proves the next theorem.

Theorem 14. *Let $c>0$ and $N_R=(R+c)^2$. For $R>c$, we have*

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

Actually, the proof says that with high probability, $\mathfrak{G}^R$ and $\mathfrak{G}_{N_R}^R$ 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 $n! \geq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$

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

Since $R>c$, the proof is complete. □

As a corollary of the previous proof, we have

$$\mathbf{P}(\exists n \geq (R+c)^2, \mathrm{Ber}(\lambda_n^R)=1) \leq \sum_{n\geq (R+c)^2} \lambda_n^R \leq \kappa R e^{-c^2} \tag{11}$$

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 15. *For any $R > c > 0$,*

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

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

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

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:

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

Using Stirling formula

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

The proof is thus complete. □

The combination of Lemma 15 and (11) 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, even it has not been noticed to the best of our knowledge, the particular form of the eigenfunctions of the Ginibre like processes is prone to the simulation of modulus and arguments by inverting their respective cumulative distribution function. This new approach is summarized in Algorithm 2.

Lemma 16 (Simulation of the modules)**.** *Let $p(z) = \sum_{i \in I} a_i z^{n_i} f_i(|z|)$ 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_ℓ for ℓ from 1 to $|I|$, we denote by e_ℓ the orthonormal vectors obtained by Gram-Schmidt orthonormalization of the vectors $\phi_I^R(W_\ell)$. Let also $M_\ell \subset \mathbb{R}^{|I|}$ be the vector $(|e_{\ell,i}|^2)_{i \in I}$ where $e_{\ell,i}$ is the coordinate of index i of e_ℓ. 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 drawing the module of W_i in Algorithm 1 is reduced to sampling uniformly c_i in $[0, 1]$ and solving the equation:

$$c_i = \frac{1}{|I| - i + 1} U_i \cdot U_F(r). \tag{12}$$

Knowing U_i, we can compute U_{i+1} in $O(|I|)$ arithmetic operations. Using a dichotomy approach, Equation (12) can be solved with precision ε using $O(|I| \log \varepsilon)$ evaluation of the F_i.

Given the moduli, we can now simulate the arguments.

Lemma 17 (Simulation of the arguments)**.** *Let $p = \sum_{i \in I} a_i z^{n_i} f_i(|z|)$ 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:

$$Q(r, \alpha) = \sum_{i,k \in I} a_i \overline{a_k} G_{i,k}(r, \alpha)$$

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

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

$$c_i = \frac{1}{V_i \cdot V_G(r, 2\pi)} V_i \cdot V_G(r, \alpha). \tag{13}$$

Computing V_{i+1} from V_i requires $O(|I|^2)$ arithmetic operations. Then, for fixed r, Equation (13) can be solved up to precision ε in $O(|I|^2 + |I| \log \varepsilon)$ arithmetic operations and evaluations of the f_i, using a dichotomy approach.

The total cost of sampling the W_i with this approach is $O(|I|^3 + |I|^2 \log \varepsilon)$ 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.

Data: R, I
Result: $W_1, \cdots, W_{|I|}$

Draw W_1 from the distribution with density $\|\phi_I(x)\|^2_{\mathbf{C}^{|I|}}/|I|$;
$e_1 \leftarrow \phi_I(W_1)/\|\phi_I(W_1)\|_{\mathbf{C}^{|I|}}$;
$U_1 = (1)_{i\in I}$;
$V_1 = (\mathbb{1}_{i=k})_{i,k\in I}$;
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
  $u_i \leftarrow \phi_I(W_i) - \sum_{k=1}^{i-1} e_k.\phi_I(W_i)\ e_k$;
  $e_i \leftarrow u_i/\|u_i\|_{\mathbf{C}^{|I|}}$;
end

Algorithm 2: Simulation of a compact symmetric projection point process restricted to the disc $\mathcal{B}_R$

5.2. **Compact Ginibre and approximation.** Using Theorem 13 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.

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

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

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

and define the following approximated functions :

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

and let

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

We now show that replacing ϕ_n^R by $\tilde{\phi}_n^R$ does not cost much in terms of accuracy.

Theorem 18. *For any* $I \subset \{1, \cdots, N_R\}$,

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

Proof. According to Theorem 13, it is sufficient to evaluate

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

for any $j \in I$. Denote the two measures involved in the previous OTP by

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

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

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

the distribution of r given θ under ζ and the same quantity for $\tilde{\zeta}$. If we have a coupling Σ_θ between these two measures, then

$$(r,\theta) \longmapsto (\Sigma_\theta(r),\theta)$$

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

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

where

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

We have

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

Hence the Bakry-Emery criterion [26] entails that the measure

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

satisfies the Talagrand inequality: For any probability measure ρ

$$\mathcal{W}_e(\rho,\,\rho_\infty) \le H(\rho\,|\,\rho_\infty) = \int \rho(r)\log\frac{\rho(r)}{\rho_\infty(r)}\mathrm{d}r$$

Apply this identity to

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

yields

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

□

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

Lemma 19. *There exists a constant κ such that for $\sqrt{\log R} \le c \le R$:*

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

Proof. We split the sum in three parts:

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

We will first prove that the terms in S_1 and S_2 are $O(e^{-c^2})$ and the terms $\log(\frac{1}{\mu_{R^2+k}})$ 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 $\dfrac{\gamma(n+1,R^2)}{\Gamma(n+1)}$:

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

This implies that

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

and

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

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

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

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 :

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

Thus, summing from R^2 to ∞ 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.

□

5.3. **Experimental results.** An implementation in Python of this algorithm publicly available [22] allowed us to sample 10 000 points in 2 128 seconds on a 8 core 3Ghz CPU. The same approach can be used for the DPP with the so-called Bergmann kernel which represents the zeros of some Gaussian analytic functions [17].

References

[1] N. Anari, S. O. Gharan, and A. Rezaei. Monte-Carlo Markov chain algorithms for sampling strongly Rayleigh distributions and determinantal point processes. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 103–115, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR.

[2] A. D. Barbour and T. C. Brown. The Stein-Chen method, point processes and compensators. *Ann. Probab.*, 20(3):1504–1527, 1992.

[3] A. D. Barbour and M. Maansson. Compound Poisson approximation and the clustering of random points. *Adv. in Appl. Probab.*, 32(1):19–38, 2000.

[4] A. D. Barbour and M. Maansson. Compound Poisson process approximation. *Ann. Probab.*, 30(3):1492–1537, 2002.

[5] R. Bardenet and A. Hardy. Monte-Carlo with determinantal point processes. *Annals of Applied Probability*, 30(1):368-417, 2020.

[6] I. Camilier and L. Decreusefond. Quasi-invariance and integration by parts for determinantal and permanental point processes. *Journal of Functional Analysis*, 259:268–300, 2010.

[7] D. J. Daley and D. Vere-Jones. *An introduction to the theory of point processes. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, 2003.

[8] L. Decreusefond. Wasserstein distance on configurations space. *Potential Anal.*, 28(3):283–300, 2008.

[9] L. Decreusefond, I. Flint, and A. Vergne. A note on the simulation of the ginibre point process. *Journal of Applied Probability*, 52(04):1003–1012.
[10] L. Decreusefond, A. Joulin, and N. Savy. Upper bounds on Rubinstein distances on configuration spaces and applications. *Communications on stochastic analysis*, 4(3):377–399, 2010.
[11] R. M. Dudley. *Real Analysis and Probability*, volume 74. Cambridge University Press, 2002.
[12] N. Dunford and J.T. Schwartz. *Linear Operators. Part II*. Wiley Classics Library, 1988.
[13] M. Fenzel and G. Lambert. Precise deviations for disk counting statistics of invariant determinantal processes. arXiv:2003.07776.
[14] H.-O. Georgii and H.J. Yoo. Conditional intensity and Gibbsianness of determinantal point processes. *Journal of Statistical Physics*, 118(1-2):55–84, 2005.
[15] A. Goldman. The Palm measure and the Voronoi tessellation for the Ginibre process. *The Annals of Applied Probability*, 20(1):90–128, February 2010.
[16] A. Haimi and H. Hedenmalm. The polyanalytic Ginibre ensembles. *Journal of Statistical Physics*, 153(1):10–47,2013.
[17] B. Hough, M. Krishnapur, Y. Peres, and B. Virag. *Zeros of Gaussian Analytic Functions and Determinantal Point Processes*. Microsoft Research.
[18] J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág. Determinantal processes and independence. *Probability Surveys*, 3:206–229 (electronic), 2006.
[19] A. Kulesza and B. Taskar. Determinantal point processes for machine learning. *Foundations and Trends® in Machine Learning*, 5(2–3):123–286, 2012.
[20] F. Lavancier, J. Møller, and E. Rubak. Determinantal point process models and statistical inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(4):853–877, 2015.
[21] O. Macchi. The coincidence approach to stochastic point processes. *Advances in Applied Probability*, 7:83–122, 1975.
[22] G. Moroz. Determinantal point process. *Software*, doi:10.5281/zenodo.4088585, `https://gitlab.inria.fr/gmoro/point_process`, October 2020.
[23] M. Röckner and A. Schied. Rademacher's theorem on configuration spaces and applications. *Journal of Functional Analysis*, 169(2):325–356, 1999. 00000.
[24] T. Shirai and Y. Takahashi. Random point fields associated with certain Fredholm determinants i: fermion, Poisson and boson point processes. *Journal of Functional Analysis*, 205(2):414–463, dec 2003.
[25] N. Tremblay and P.-O. Barthelme, S. Amblard. Optimized Algorithms to Sample Determinantal Point Processes. *arXiv e-prints*, page arXiv:1802.08471, February 2018.
[26] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
[27] C. Villani. *Optimal Transport, Old and New*. Lectures Notes in Mathematics. Springer Verlag, New York, 2007.

LTCI, Telecom Paris, Institut polytechnique de Paris, France
Email address: `laurent.decreusefond@mines-telecom.fr`

INRIA Nancy Grand-Est, Nancy, France
Email address: `guillaume.moroz@inria.fr`