Monte Carlo integration with repulsive point processes

Rémi Bardenet

CNRS & CRIStAL, Univ. Lille, France

https://statisfaction-blog.github.io/

Figure imposée

The goal is to approximate

$$\int f \,\mathrm{d}\mu \approx \sum_{i=1}^{N} w_i f(x_i).$$

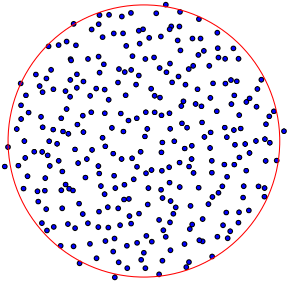
- How to choose the nodes x_i?
- How to choose the weights w_i?

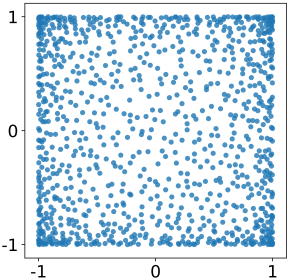
Monte Carlo integration (importance sampling, MCMC, etc.)

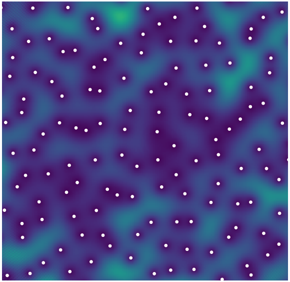
- Choose the nodes randomly, and the weights $w_i = w_i(x_1, \dots, x_N)$.
- Typical error is

$$\sqrt{\mathbb{E}\left[\int f \,\mathrm{d}\mu - \sum_{i=1}^{N} w_i f(x_i)\right]^2} \sim \frac{1}{\sqrt{N}}.$$

Prologue: Repulsive point processes







Go to CLT

Monte Carlo with DPPs

DPPs lead to tight rates in RKHSs

Repelled point processes

Monte Carlo with DPPs

DPPs lead to tight rates in RKHSs

Repelled point processes

Projection DPPs

- Let $(\varphi_k)_{k=0,\dots,N-1}$ be an orthonormal sequence in $L^2(\mu)$.
- Let $\mathrm{K}(x,y) = \sum_{k=0}^{N-1} \varphi_k(x)\varphi_k(y)$.

Definition (Hough, Krishnapur, Peres, and Virág, 2006)

$X = \{x_1, \dots, x_N\}$ is the DPP with kernel K and reference measure μ if

$$x_1, \dots, x_N \sim \frac{1}{N!} \det \Big[\mathrm{K}(x_i, x_\ell)\Big]_{i,\ell=1}^{N} \mathrm{d}\mu(x_1)\dots\mathrm{d}\mu(x_N).$$

Projection DPPs

- Let $(\varphi_k)_{k=0,\dots,N-1}$ be an orthonormal sequence in $L^2(\mu)$.
- Let $\mathrm{K}(x,y) = \sum_{k=0}^{N-1} \varphi_k(x)\varphi_k(y)$.

Definition (Hough, Krishnapur, Peres, and Virág, 2006)

$X = \{x_1, \dots, x_N\}$ is the DPP with kernel K and reference measure μ if

$$x_1, \dots, x_N \sim \frac{1}{N!} \det \left[\mathrm{K}(x_i, x_\ell) \right]_{i,\ell=1}^{N} \mathrm{d}\mu(x_1) \dots \mathrm{d}\mu(x_N).$$

What projection DPP samples look like

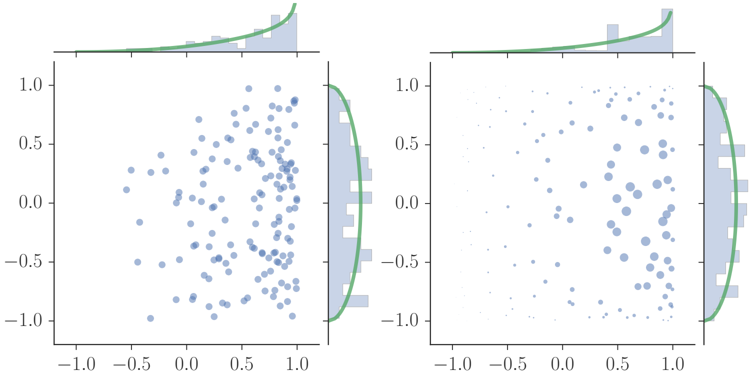


Figure: Left:i.i.d., Right: orthogonal polynomial ensemble (DPP)

Theorem (Bardenet and Hardy, 2020)

Let $\mu(\mathrm{d}x) = \omega(x)\mathrm{d}x$ with ω separable, $\mathscr{C}^1$, positive on the open set $(-1,1)^d$, and satisfying a technical regularity assumption. Let $\varepsilon > 0$. If $x_1, \dots, x_N$ stands for the associated multivariate OP Ensemble, then for every f $\mathscr{C}^1$ vanishing outside $[-1+\varepsilon, 1-\varepsilon]^d$,

$$\sqrt{N^{1+1/d}} \left(\sum_{i=1}^{N} \frac{f(x_i)}{\mathrm{K}(x_i, x_i)} - \int f(x)\mu(\mathrm{d}x) \right) \xrightarrow[N\to\infty]{law} \mathcal{N}(0, \Omega_{f,\omega}^2),$$

where

$$\Omega_{f,\omega}^2 = \frac{1}{2} \sum_{k_1,\dots,k_d=0}^{\infty} (k_1 + \cdots + k_d) \widehat{\left(\frac{f\omega}{\omega_{eq}^{\otimes d}} \right)} (k_1, \dots, k_d)^2,$$

and $\omega_{eq}^{\otimes d}(x) = \pi^{-d}(1-x^2)^{-1/2}$.

- With Thibaut Lemoine, we have a CLT with rate $\sqrt{N^{1+2/d}}$ (soon).

Theorem (Bardenet and Hardy, 2020)

Let $\mu(\mathrm{d}x) = \omega(x)\mathrm{d}x$ with ω $\mathscr{C}^1$ on $(-1,1)^d$. Consider a measure $q(x)\mathrm{d}x$ satisfying the assumptions of the previous theorem, let $K_N(x,y)$ be the corresponding kernel, and $x_1,\dots,x_N$ the associated multivariate OP Ensemble. Then, for every f as before,

$$\sqrt{N^{1+1/d}}\left(\sum_{i=1}^{N}\frac{f(x_i)}{\mathrm{K}(x_i,x_i)}\frac{\omega(x_i)}{q(x_i)} - \int f(x)\mu(\mathrm{d}x)\right) \xrightarrow[N\to\infty]{law} \mathcal{N}\left(0,\Omega_{f,\omega}^2\right),$$

where $\Omega_{f,\omega}^2$ is unchanged.

Plan

Monte Carlo with DPPs

DPPs lead to tight rates in RKHSs

Repelled point processes

RKHSs are spaces of smooth functions with a kernel

- Consider the RKHS $\mathcal{F}$ with kernel k, i.e. the completion of

$$\left\{ \sum_{i=1}^{M} \alpha_i k(x_i, \cdot), M \in \mathbb{N}, \alpha_1, \dots, \alpha_n \in \mathbb{R}, x_1, \dots, x_M \in \mathbb{R}^d \right\}.$$

for the inner product defined by $\langle k(x, \cdot), k(y, \cdot) \rangle_{\mathcal{F}} := k(x, y)$.

- For $f \in \mathcal{F}$ and $x \in \mathcal{X}$, $f(x) = \langle f, k(x, \cdot) \rangle$.

- Under general assumptions, $\mathcal{F} \subset L^2(\mathrm{d}\mu)$, is dense, there is an ON basis (e_n) of $L^2(\mathrm{d}\mu)$ and $\sigma_n \to 0$ such that, pointwise,

$$k(x, y) = \sum_{n \geqslant 1} \sigma_n e_n(x) e_n(y).$$

- In that case, $f \in \mathcal{F}$ if and only if $\sum_n \sigma_n^{-1} |\langle f, e_n \rangle|^2$ converges.

RKHSs are spaces of smooth functions with a kernel

- Consider the RKHS $\mathcal{F}$ with kernel k, i.e. the completion of

$$\left\{ \sum_{i=1}^{M} \alpha_i k(x_i, \cdot), M \in \mathbb{N}, \alpha_1, \ldots, \alpha_n \in \mathbb{R}, x_1, \ldots, x_M \in \mathbb{R}^d \right\}.$$

  for the inner product defined by $\langle k(x, \cdot), k(y, \cdot) \rangle_{\mathcal{F}} := k(x, y)$.

- For $f \in \mathcal{F}$ and $x \in \mathcal{X}$, $f(x) = \langle f, k(x, \cdot) \rangle$.

- Under general assumptions, $\mathcal{F} \subset L^2(\mathrm{d}\mu)$, is dense, there is an ON basis (e_n) of $L^2(\mathrm{d}\mu)$ and $\sigma_n \to 0$ such that, pointwise,

$$k(x, y) = \sum_{n \geqslant 1} \sigma_n e_n(x) e_n(y).$$

- In that case, $f \in \mathcal{F}$ if and only if $\sum_n \sigma_n^{-1} |\langle f, e_n \rangle|^2$ converges.

- Consider the RKHS $\mathcal{F}$ with kernel k, i.e. the completion of

$$\left\{ \sum_{i=1}^{M} \alpha_i k(x_i, \cdot), M \in \mathbb{N}, \alpha_1, \dots, \alpha_n \in \mathbb{R}, x_1, \dots, x_M \in \mathbb{R}^d \right\}.$$

for the inner product defined by $\langle k(x, \cdot), k(y, \cdot) \rangle_{\mathcal{F}} := k(x, y)$.

- For $f \in \mathcal{F}$ and $x \in \mathcal{X}$, $f(x) = \langle f, k(x, \cdot) \rangle$.

- Under general assumptions, $\mathcal{F} \subset L^2(\mathrm{d}\mu)$, is dense, there is an ON basis (e_n) of $L^2(\mathrm{d}\mu)$ and $\sigma_n \to 0$ such that, pointwise,

$$k(x, y) = \sum_{n \geqslant 1} \sigma_n e_n(x) e_n(y).$$

- In that case, $f \in \mathcal{F}$ if and only if $\sum_n \sigma_n^{-1} |\langle f, e_n \rangle|^2$ converges.

RKHSs are spaces of smooth functions with a kernel

- Consider the RKHS $\mathcal{F}$ with kernel k, i.e. the completion of

$$\left\{ \sum_{i=1}^{M} \alpha_i k(x_i, \cdot), M \in \mathbb{N}, \alpha_1, \dots, \alpha_n \in \mathbb{R}, x_1, \dots, x_M \in \mathbb{R}^d \right\}.$$

  for the inner product defined by $\langle k(x, \cdot), k(y, \cdot) \rangle_{\mathcal{F}} := k(x, y)$.

- For $f \in \mathcal{F}$ and $x \in \mathcal{X}$, $f(x) = \langle f, k(x, \cdot) \rangle$.

- Under general assumptions, $\mathcal{F} \subset L^2(\mathrm{d}\mu)$, is dense, there is an ON basis (e_n) of $L^2(\mathrm{d}\mu)$ and $\sigma_n \to 0$ such that, pointwise,

$$k(x, y) = \sum_{n \geqslant 1} \sigma_n e_n(x) e_n(y).$$

- In that case, $f \in \mathcal{F}$ if and only if $\sum_n \sigma_n^{-1} |\langle f, e_n \rangle|^2$ converges.

- Let $f \in \mathcal{F}$, $g \in L^2(\mathrm{d}\mu)$ then

$$\left| \int fg\mathrm{d}\mu - \sum_{i=1}^{N} w_i f(x_i) \right| \leqslant \|f\|_{\mathcal{F}} \left\| \mu_g - \sum_{i=1}^{N} w_i k(x_i, .) \right\|_{\mathcal{F}}, \qquad (1)$$

where

$$\mu_g = \int g(x)k(x,.)\mathrm{d}\mu(x)$$

is the mean element of g.

- Once the nodes $x_1, \dots, x_N$ are known, minimizing the RHS of (1) in w boils down to inverting the $N \times N$ Gram matrix $((k(x_i, x_j)))$.

- Let $f \in \mathcal{F}$, $g \in L^2(\mathrm{d}\mu)$ then

$$\left| \int fg \mathrm{d}\mu - \sum_{i=1}^{N} w_i f(x_i) \right| \leqslant \|f\|_{\mathcal{F}} \left\| \mu_g - \sum_{i=1}^{N} w_i k(x_i, .) \right\|_{\mathcal{F}}, \quad (1)$$

where

$$\mu_g = \int g(x) k(x, .) \mathrm{d}\mu(x)$$

is the mean element of g.

- Once the nodes $x_1, \dots, x_N$ are known, minimizing the RHS of (1) in w boils down to inverting the $N \times N$ Gram matrix $((k(x_i, x_j)))$.

Volume sampling and tight rates

Volume sampling

- Let $x_1, \dots, x_N \sim Z^{-1} \det[k(x_i, x_j)] \, \mathrm{d}\mu(x_1) \dots \mathrm{d}\mu(x_N)$
- Solve the linear program for the weights $w_1, \dots, w_N$.

Theorem (Belhadji, Bardenet, and Chainais, 2020)

Assume again $\sum_{n=1}^{N} |\langle g, e_n \rangle|^2 \leqslant 1$. *Then*

$$\mathbb{E} \left\| \mu_g - \sum_{i=1}^{N} w_i k(x_i, \cdot) \right\|_{\mathcal{F}}^2 \leqslant \sigma_N \left(1 + \beta_N\right),$$

where $\beta_N = \min_{M \in [2:N]} [(N - M + 1)\sigma_N]^{-1} \sum_{m \geqslant M} \sigma_m$.

- Pinkus, 2012 shows that $\inf_{\substack{Y \subset \mathcal{F} \\ \dim Y = N}} \sup_{\|g\|_{L^2(\mu)} \leqslant 1} \inf_{y \in Y} \|\mu_g - y\|_{\mathcal{F}}^2 = \sigma_{N+1}$.

Go to repelled PPs

Volume sampling

- Let $x_1, \dots, x_N \sim Z^{-1} \det[k(x_i, x_j)] \, \mathrm{d}\mu(x_1) \dots \mathrm{d}\mu(x_N)$
- Solve the linear program for the weights $w_1, \dots, w_N$.

Theorem (Belhadji, Bardenet, and Chainais, 2020)

Assume again $\sum_{n=1}^{N} |\langle g, e_n \rangle|^2 \leqslant 1$. *Then*

$$\mathbb{E} \left\| \mu_g - \sum_{i=1}^{N} w_i k(x_i, \cdot) \right\|_{\mathcal{F}}^2 \leqslant \sigma_N \left(1 + \beta_N\right),$$

where $\beta_N = \min_{M \in [2:N]} [(N - M + 1)\sigma_N]^{-1} \sum_{m \geqslant M} \sigma_m$.

- Pinkus, 2012 shows that $\inf_{\substack{Y \subset \mathcal{F} \\ \dim Y = N}} \sup_{\|g\|_{L^2(\mu)} \leqslant 1} \inf_{y \in Y} \|\mu_g - y\|_{\mathcal{F}}^2 = \sigma_{N+1}$.

Go to repelled PPs

- For $f \in \mathcal{F} \subset L^2(\mu)$, we investigate guarantees on

$$\mathbb{E}\|f - \hat{f}\|^2_{L^2(\mu)}$$

in (Belhadji, Bardenet, and Chainais, 2023, *preprint*).

- In (Rouault, Bardenet, and Maida, 2024), we investigate the Coulomb gas with interaction potential k and confining potential V,

$$\mathrm{d}\mathbb{P}^V_{n,\beta_n}(X_n) = \frac{1}{Z^V_{n,\beta_n}} \mathrm{e}^{-\frac{\beta_n}{2n^2}\sum_{i\neq j} K(x_i,x_j) - \frac{\beta_n}{n}\sum_{i=1}^n V(x_i)} \mathrm{d}x_1 \ldots \mathrm{d}x_n,$$

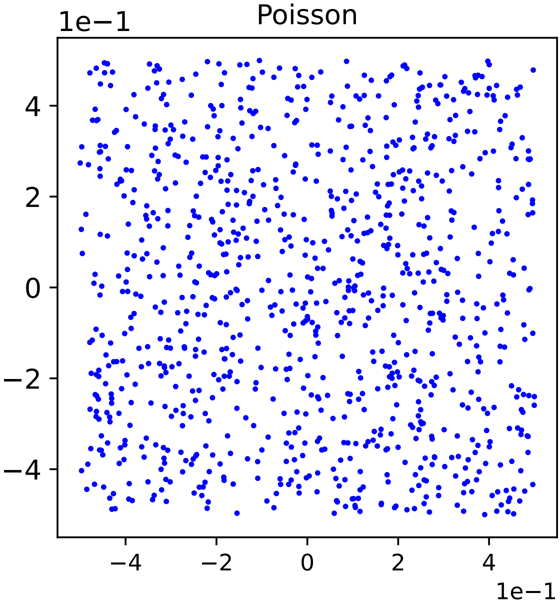
In particular, we prove that for $\beta_n = n^2$ and $r \leqslant 1/\sqrt{n}$,

$$\mathbb{P}^V_{n,\beta_n}\left(\sup_{f\in B_{\mathcal{F}}} \left| \int f \mathrm{d}\mu_n - \int f \mathrm{d}\mu_V \right|^2 > r^2 \right) \leqslant \exp\left(-u_1 \beta_n r^2\right). \quad (2)$$

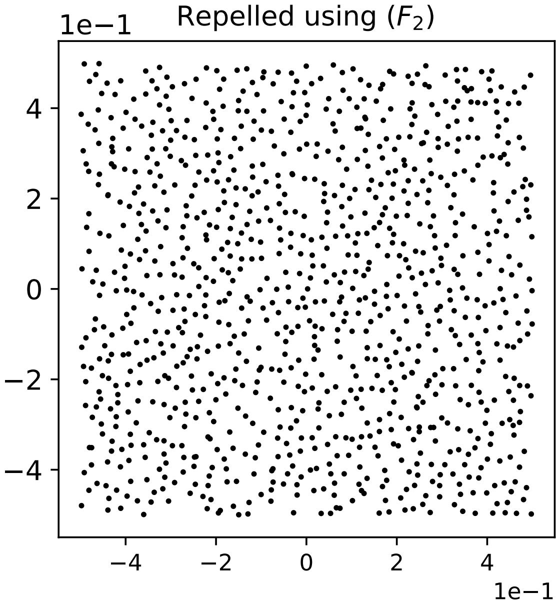
Monte Carlo with DPPs

DPPs lead to tight rates in RKHSs

Repelled point processes


Poisson
1e−1
4
2
0
−2
−4
−4
−2
0
2
4
1e−1

The repelled Poisson point process



Coulomb repulsion leads to variance reduction

- Let $\mathcal{P} \subset \mathbb{R}^d$ be a homogeneous Poisson point process.
- For x in $\mathbb{R}^d$ and a collection C of points in $\mathbb{R}^d$, consider

$$F_C(x) = \sum_{y \in C, y \neq x} \frac{x - y}{\|x - y\|^d},$$

- Consider the repelled point process

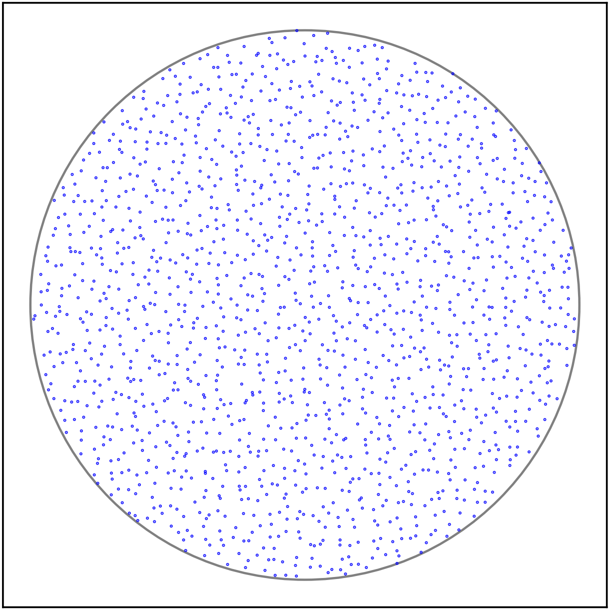
$$\Pi_\varepsilon \mathcal{P} \triangleq \{x + \varepsilon F_{\mathcal{P}}(x), \quad x \in \mathcal{P}\}.$$

Theorem (Hawat, Bardenet, and Lachièze-Rey, 2023)

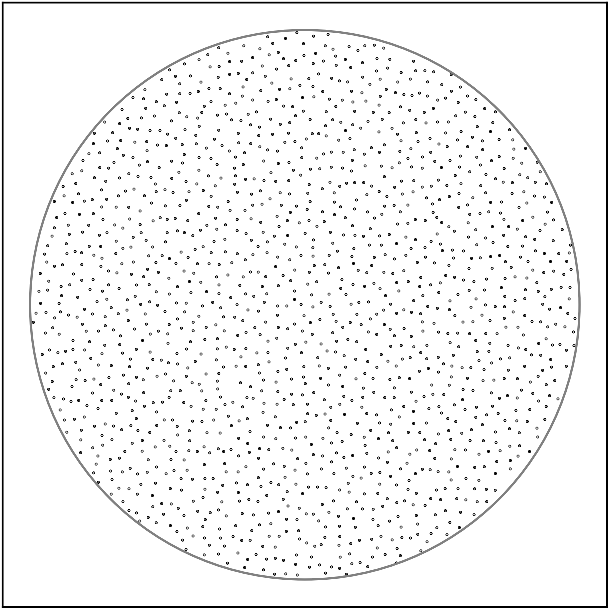
- $\Pi_\varepsilon \mathcal{P}$ is well-defined and has the same intensity λ as $\mathcal{P}$.
- For f compactly supported in $S \subset \mathbb{R}^d$, there is an $\varepsilon > 0$ such that

$$\mathrm{Var}\left[\frac{1}{\lambda} \sum_{x \in S \cap \Pi_\varepsilon \mathcal{P}} f(x)\right] < \mathrm{Var}\left[\frac{1}{\lambda} \sum_{x \in S \cap \mathcal{P}} f(x)\right].$$

The repelled Ginibre point process[1]



The repelled Ginibre point process[1]



Wrapping up

- Volume sampling yields tight MSE bounds in RKHSs.[2][3][4]
- DPP sampling is an active research topic. Check out our Python toolbox DPPy.[5]
- Coulomb repulsion yields variance reduction at a lower cost, and is potentially widely applicable.[6]

Take-home message

- Repulsive point processes yield fast Monte Carlo integration.
- DPPs tie analytic assumptions with node design.
- PhD and postdoc applications welcome! remi.bardenet@gmail.com

[2] Belhadji, Bardenet, and Chainais, 2019.
[3] Belhadji, Bardenet, and Chainais, 2020.
[4] Belhadji, Bardenet, and Chainais, 2023.
[5] Gautier, Bardenet, Polito, and Valko, 2019, github.com/guilgautier/DPPy.
[6] Hawat, Bardenet, and Lachièze-Rey, 2023.

- Volume sampling yields tight MSE bounds in RKHSs.[2][3][4]
- DPP sampling is an active research topic. Check out our Python toolbox `DPPy`.[5]
- Coulomb repulsion yields variance reduction at a lower cost, and is potentially widely applicable.[6]

Take-home message

- Repulsive point processes yield fast Monte Carlo integration.
- DPPs tie analytic assumptions with node design.
- PhD and postdoc applications welcome! remi.bardenet@gmail.com

[2] Belhadji, Bardenet, and Chainais, 2019.
[3] Belhadji, Bardenet, and Chainais, 2020.
[4] Belhadji, Bardenet, and Chainais, 2023.
[5] Gautier, Bardenet, Polito, and Valko, 2019, `github.com/guilgautier/DPPy`.
[6] Hawat, Bardenet, and Lachièze-Rey, 2023.

- Volume sampling yields tight MSE bounds in RKHSs.[2][3][4]
- DPP sampling is an active research topic. Check out our Python toolbox `DPPy`.[5]
- Coulomb repulsion yields variance reduction at a lower cost, and is potentially widely applicable.[6]

Take-home message

- Repulsive point processes yield fast Monte Carlo integration.
- DPPs tie analytic assumptions with node design.
- PhD and postdoc applications welcome! remi.bardenet@gmail.com

[2] Belhadji, Bardenet, and Chainais, 2019.
[3] Belhadji, Bardenet, and Chainais, 2020.
[4] Belhadji, Bardenet, and Chainais, 2023.
[5] Gautier, Bardenet, Polito, and Valko, 2019, `github.com/guilgautier/DPPy`.
[6] Hawat, Bardenet, and Lachièze-Rey, 2023.

- Bardenet, R. and A. Hardy (2020). "Monte Carlo with Determinantal Point Processes". In: *Annals of Applied Probability.*
- Belhadji, A., R. Bardenet, and P. Chainais (2019). "Kernel quadrature with determinantal point processes". In: *Advances in Neural Information Processing Systems (NeurIPS).*
- — (2020). "Kernel interpolation with continuous volume sampling". In: *International Conference on Machine Learning (ICML).*
- — (2023). "Signal reconstruction using determinantal sampling". In: *arXiv preprint arXiv:2310.09437.*
- Gautier, G., R. Bardenet, G. Polito, and M. Valko (2019). "DPPy: Sampling Determinantal Point Processes with Python". In: *Journal of Machine Learning Research; Open Source Software (JMLR MLOSS).*
- Hawat, D., R. Bardenet, and R. Lachièze-Rey (2023). "Repelled point processes with application to numerical integration". In: *arXiv preprint arXiv:2308.04825.*
- Hough, J. B., M. Krishnapur, Y. Peres, and B. Virág (2006). "Determinantal processes and independence". In: *Probability surveys.*
- Pinkus, A. (2012). *N-widths in Approximation Theory*. Vol. 7. Springer Science & Business Media.

Rouault, M., R. Bardenet, and Mylène Maida (2024). "Monte Carlo with kernel-based Gibbs measures: Guarantees for probabilistic herding". In: *arXiv preprint arXiv:2402.11736*.