Interpolation and experimental design with volume sampling

Rémi Bardenet

CNRS & CRIStAL, Univ. Lille, France

Joint work with

Figure: Adrien Hardy, Ayoub Belhadji, Pierre Chainais, Arnaud Poinas

Prologue: numerical integration and DPPs

Tight interpolation rates in RKHSs

Volume sampling for experimental design

Prologue: numerical integration and DPPs

Tight interpolation rates in RKHSs

Volume sampling for experimental design

The goal is to approximate

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

- 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 $w_i = w(x_i, x_{-i})$.
- Typical error is

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

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

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

1. If $\mu = \sum_{x\in\mathcal{X}} \delta_x$, one recovers

$$\mathbb{P}(A \subset X) = \det \mathbf{K}_A.$$

2. $x_1 \sim \frac{1}{N}\mathrm{K}(x,x)\mathrm{d}\mu(x)$ so that $\mathbb{E}\sum_{i=1}^{N} \frac{f(x_i)}{\mathrm{K}(x_i,x_i)} = \int f \mathrm{d}\mu$.
3. A natural choice of $\varphi_k : \mathbb{R}^d \to \mathbb{R}$ is orthogonal polynomials w.r.t. μ.

- 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

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

1. If $\mu = \sum_{x\in\mathcal{X}} \delta_x$, one recovers

$$\mathbb{P}(A \subset X) = \det \mathbf{K}_A.$$

2. $x_1 \sim \frac{1}{N}\mathrm{K}(x,x)\mathrm{d}\mu(x)$ so that $\mathbb{E}\sum_{i=1}^{N} \frac{f(x_i)}{\mathrm{K}(x_i,x_i)} = \int f \mathrm{d}\mu$.
3. A natural choice of $\varphi_k : \mathbb{R}^d \to \mathbb{R}$ is orthogonal polynomials w.r.t. μ.

- 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

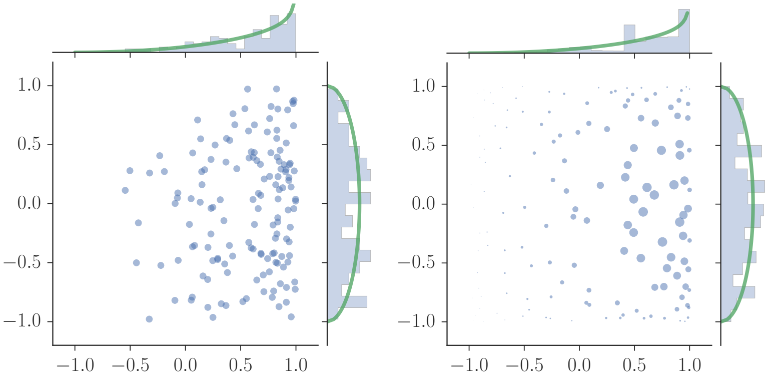
$$x_1, \dots, x_N \sim \frac{1}{N!} \det \left[\mathrm{K}(x_i, x_\ell)\right]_{i,\ell=1}^{N} \mathrm{d}\mu(x_1) \dots \mathrm{d}\mu(x_N).$$

1. If $\mu = \sum_{x \in \mathcal{X}} \delta_x$, one recovers

$$\mathbb{P}(A \subset X) = \det \mathbf{K}_A.$$

2. $x_1 \sim \frac{1}{N}\mathrm{K}(x,x)\mathrm{d}\mu(x)$ so that $\mathbb{E} \sum_{i=1}^{N} \frac{f(x_i)}{\mathrm{K}(x_i,x_i)} = \int f \mathrm{d}\mu$.
3. A natural choice of $\varphi_k : \mathbb{R}^d \to \mathbb{R}$ is orthogonal polynomials w.r.t. μ.

Multivariate orthogonal polynomial ensembles



(a) i.i.d. (b) DPP

Theorem (Bardenet and Hardy 2020)

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

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

where

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

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

- As seen today[1], for $\mu = \mathrm{d}x$, assumptions can be relaxed and K be taken such that $\mathrm{K}(x,x) \propto 1$.

[1] Coeurjolly, Mazoyer, and Amblard 2021.

Plan

Prologue: numerical integration and DPPs

Tight interpolation rates in RKHSs

Volume sampling for experimental design

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

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

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

- 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,

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

- 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 κ, i.e. the completion of

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

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

- 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,

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

- 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 κ, i.e. the completion of

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

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

- 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,

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

- 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

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

where

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

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 an $N \times N$ matrix.

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

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

where

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

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 an $N \times N$ matrix.

A DPP for quadrature in RKHSs: first attempt

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

Algorithm 1: DPP

- Take $\mathrm{K}(x,y) = \sum_{n=1}^{N} e_n(x) e_n(y)$.
- Let $x_1, \dots, x_N \sim 1/N! \det[\mathrm{K}(x_i,x_j)]\, \mathrm{d}\mu(x_1) \dots \mathrm{d}\mu(x_N)$.
- Solve the linear problem for the weights $w_1, \dots, w_N$.

Theorem (Belhadji, Bardenet, and Chainais 2019)

Assume $\sum_{n=1}^{N} |\langle g, e_n\rangle|^2 \leqslant 1$. *Let* $r_N = \sum_{m\geqslant N+1} \sigma_m$, *then*

$$\mathbb{E}\left\|\mu_g - \sum_{i=1}^{N} w_i \kappa(x_i,\cdot)\right\|_{\mathcal{F}}^2 \leqslant 2\sigma_{N+1} + 2\left(N r_N + \sum_{\ell=2}^{N} \frac{\sigma_1}{\ell!^2}\left(\frac{N r_N}{\sigma_1}\right)^{\ell}\right).$$

A DPP for quadrature in RKHSs: first attempt

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

Algorithm 1: DPP

- Take $\mathrm{K}(x, y) = \sum_{n=1}^{N} e_n(x)e_n(y)$.
- Let $x_1, \dots, x_N \sim 1/N! \det[\mathrm{K}(x_i, x_j)]\, \mathrm{d}\mu(x_1)\dots\mathrm{d}\mu(x_N)$.
- Solve the linear problem for the weights $w_1, \dots, w_N$.

Theorem (Belhadji, Bardenet, and Chainais 2019)

Assume $\sum_{n=1}^{N} |\langle g, e_n\rangle|^2 \leqslant 1$. *Let* $r_N = \sum_{m\geqslant N+1} \sigma_m$, *then*

$$\mathbb{E}\left\|\mu_g - \sum_{i=1}^{N} w_i\kappa(x_i, \cdot)\right\|_{\mathcal{F}}^2 \leqslant 2\sigma_{N+1} + 2\left(Nr_N + \sum_{\ell=2}^{N}\frac{\sigma_1}{\ell!^2}\left(\frac{Nr_N}{\sigma_1}\right)^{\ell}\right).$$

Algorithm 2: volume sampling

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

Theorem (Belhadji, Bardenet, and Chainais 2020b)

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

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

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

- It is known[2] that $\inf_{\substack{Y \subset \mathcal{F} \\ \dim Y = N}} \sup_{\|g\|_{\mathrm{d}\omega} \leqslant 1} \inf_{y \in Y} \|\mu_g - y\|_{\mathcal{F}}^2 = \sigma_{N+1}$.

[2] Pinkus 2012.

Algorithm 2: volume sampling

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

Theorem (Belhadji, Bardenet, and Chainais 2020b)

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

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

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

- It is known[2] that $\inf_{\substack{Y \subset \mathcal{F} \\ \dim Y = N}} \sup_{\|g\|_{\mathrm{d}\omega} \leqslant 1} \inf_{y \in Y} \|\mu_g - y\|_{\mathcal{F}}^2 = \sigma_{N+1}$.

[2] Pinkus 2012.

Many open problems

- Robustness to RKHS hypothesis / model choice.
- Practical relevance of RKHS hypothesis.
- What should $g \in L^2(\mu)$ be in $\int fg \mathrm{d}\mu$?
- How do we efficiently sample from continuous volume sampling without spectral knowledge? See e.g. Rezaei and Gharan 2019.
- Kernel interpolation is similar to column-subset selection for linear regression[3], where DPPs and VS yield similar bounds[4].

[3] Derezinski and M. Mahoney 2020.
[4] Belhadji, Bardenet, and Chainais 2020a.

Prologue: numerical integration and DPPs

Tight interpolation rates in RKHSs

Volume sampling for experimental design

- Consider $\varphi_1, \cdots, \varphi_p \in L^2(\Omega)$ linearly independent, and

$$Y = \varphi(X)\beta + \varepsilon, \tag{2}$$

  where
  - $X = (x_1, \cdots, x_k)^T$, where $x_1, \ldots, x_k \in \Omega \subset \mathbb{R}^d$,
  - $Y = (y_1, \cdots, y_k)^T \in \mathbb{R}^k$,
  - $\varepsilon \in \mathbb{R}^k$ is $\mathcal{N}(0, \sigma^2 I_k)$.
  - $\varphi(X) = (\varphi_j(x_i)) \in \mathbb{R}^{k \times p}$ is the *design matrix*.
- Assuming $\beta \sim \mathcal{N}(0, \Lambda^{-1})$, the posterior mean for β is

$$\hat{\beta} = (\varphi(X)^T \varphi(X) + \Lambda)^{-1} \varphi(X)^T Y, \tag{3}$$

- The posterior covariance matrix is

$$\sigma^2(\varphi(X)^T \varphi(X) + \Lambda)^{-1}.$$

- Consider $\varphi_1, \cdots, \varphi_p \in L^2(\Omega)$ linearly independent, and

$$Y = \varphi(X)\beta + \varepsilon, \tag{2}$$

where

  - $X = (x_1, \cdots, x_k)^T$, where $x_1, \ldots, x_k \in \Omega \subset \mathbb{R}^d$,
  - $Y = (y_1, \cdots, y_k)^T \in \mathbb{R}^k$,
  - $\varepsilon \in \mathbb{R}^k$ is $\mathcal{N}(0, \sigma^2 I_k)$.
  - $\varphi(X) = (\varphi_j(x_i)) \in \mathbb{R}^{k \times p}$ is the *design matrix*.

- Assuming $\beta \sim \mathcal{N}(0, \Lambda^{-1})$, the posterior mean for β is

$$\hat{\beta} = (\varphi(X)^T\varphi(X) + \Lambda)^{-1}\varphi(X)^T Y, \tag{3}$$

- The posterior covariance matrix is

$$\sigma^2(\varphi(X)^T\varphi(X) + \Lambda)^{-1}.$$

Experimental design for linear regression

- Consider $\varphi_1, \cdots, \varphi_p \in L^2(\Omega)$ linearly independent, and

$$Y = \varphi(X)\beta + \varepsilon, \tag{2}$$

  where
  - $X = (x_1, \cdots, x_k)^T$, where $x_1, \ldots, x_k \in \Omega \subset \mathbb{R}^d$,
  - $Y = (y_1, \cdots, y_k)^T \in \mathbb{R}^k$,
  - $\varepsilon \in \mathbb{R}^k$ is $\mathcal{N}(0, \sigma^2 I_k)$.
  - $\varphi(X) = (\varphi_j(x_i)) \in \mathbb{R}^{k \times p}$ is the *design matrix*.
- Assuming $\beta \sim \mathcal{N}(0, \Lambda^{-1})$, the posterior mean for β is

$$\hat{\beta} = (\varphi(X)^T \varphi(X) + \Lambda)^{-1} \varphi(X)^T Y, \tag{3}$$

- The posterior covariance matrix is

$$\sigma^2 (\varphi(X)^T \varphi(X) + \Lambda)^{-1}.$$

Experimental design for linear regression

- Consider $\varphi_1, \cdots, \varphi_p \in L^2(\Omega)$ linearly independent, and

$$Y = \varphi(X)\beta + \varepsilon, \tag{2}$$

  where
  - $X = (x_1, \cdots, x_k)^T$, where $x_1, \ldots, x_k \in \Omega \subset \mathbb{R}^d$,
  - $Y = (y_1, \cdots, y_k)^T \in \mathbb{R}^k$,
  - $\varepsilon \in \mathbb{R}^k$ is $\mathcal{N}(0, \sigma^2 I_k)$.
  - $\varphi(X) = (\varphi_j(x_i)) \in \mathbb{R}^{k \times p}$ is the *design matrix*.
- Assuming $\beta \sim \mathcal{N}(0, \Lambda^{-1})$, the posterior mean for β is

$$\hat{\beta} = (\varphi(X)^T \varphi(X) + \Lambda)^{-1} \varphi(X)^T Y, \tag{3}$$

- The posterior covariance matrix is

$$\sigma^2 (\varphi(X)^T \varphi(X) + \Lambda)^{-1}.$$

Experimental design for linear regression

- Consider $\varphi_1, \cdots, \varphi_p \in L^2(\Omega)$ linearly independent, and

$$Y = \varphi(X)\beta + \varepsilon, \tag{2}$$

where

  - $X = (x_1, \cdots, x_k)^T$, where $x_1, \ldots, x_k \in \Omega \subset \mathbb{R}^d$,
  - $Y = (y_1, \cdots, y_k)^T \in \mathbb{R}^k$,
  - $\varepsilon \in \mathbb{R}^k$ is $\mathcal{N}(0, \sigma^2 I_k)$.
  - $\varphi(X) = (\varphi_j(x_i)) \in \mathbb{R}^{k \times p}$ is the *design matrix*.

- Assuming $\beta \sim \mathcal{N}(0, \Lambda^{-1})$, the posterior mean for β is

$$\hat{\beta} = (\varphi(X)^T \varphi(X) + \Lambda)^{-1} \varphi(X)^T Y, \tag{3}$$

- The posterior covariance matrix is

$$\sigma^2 (\varphi(X)^T \varphi(X) + \Lambda)^{-1}.$$

- Consider $\varphi_1, \cdots, \varphi_p \in L^2(\Omega)$ linearly independent, and

$$Y = \varphi(X)\beta + \varepsilon, \tag{2}$$

  where
  - $X = (x_1, \cdots, x_k)^T$, where $x_1, \ldots, x_k \in \Omega \subset \mathbb{R}^d$,
  - $Y = (y_1, \cdots, y_k)^T \in \mathbb{R}^k$,
  - $\varepsilon \in \mathbb{R}^k$ is $\mathcal{N}(0, \sigma^2 I_k)$.
  - $\varphi(X) = (\varphi_j(x_i)) \in \mathbb{R}^{k \times p}$ is the *design matrix*.
- Assuming $\beta \sim \mathcal{N}(0, \Lambda^{-1})$, the posterior mean for β is

$$\hat{\beta} = (\varphi(X)^T \varphi(X) + \Lambda)^{-1} \varphi(X)^T Y, \tag{3}$$

- The posterior covariance matrix is

$$\sigma^2(\varphi(X)^T \varphi(X) + \Lambda)^{-1}.$$

- Consider $\varphi_1, \cdots, \varphi_p \in L^2(\Omega)$ linearly independent, and

$$Y = \varphi(X)\beta + \varepsilon, \tag{2}$$

where

  - $X = (x_1, \cdots, x_k)^T$, where $x_1, \ldots, x_k \in \Omega \subset \mathbb{R}^d$,
  - $Y = (y_1, \cdots, y_k)^T \in \mathbb{R}^k$,
  - $\varepsilon \in \mathbb{R}^k$ is $\mathcal{N}(0, \sigma^2 I_k)$.
  - $\varphi(X) = (\varphi_j(x_i)) \in \mathbb{R}^{k \times p}$ is the *design matrix*.

- Assuming $\beta \sim \mathcal{N}(0, \Lambda^{-1})$, the posterior mean for β is

$$\hat{\beta} = (\varphi(X)^T \varphi(X) + \Lambda)^{-1} \varphi(X)^T Y, \tag{3}$$

- The posterior covariance matrix is

$$\sigma^2 (\varphi(X)^T \varphi(X) + \Lambda)^{-1}.$$

Finding optimal designs[6][7]

- Minimizing the posterior covariance is often replaced by

$$\min_{x_1,\ldots,x_k\in\Omega} h\left(\varphi(X)^T\varphi(X)+\Lambda\right), \tag{4}$$

where, e. g., $h = h_A \triangleq \mathrm{Tr}(\cdot^{-1})$ or $h = h_D \triangleq \det(\cdot^{-1})$.

- Still a nonconvex problem, requires enumeration when Ω finite.
- A convex relaxation to (4) is

$$\min_{\nu\in\mathcal{M}(\Omega)} h(G_\nu(\varphi)+\Lambda)\ s.t.\ \nu(\Omega)=k, \tag{5}$$

where
  - $\mathcal{M}(\Omega)$ is the space of Borel measures on Ω,
  - $G_\nu(\varphi) = \left(\int_\Omega \varphi_i(x)\varphi_j(x)d\nu(x)\right) \in \mathbb{R}^{p\times p}$.
- One approximate solution to (4) is to sample i.i.d. from $\nu^\star$, and possibly apply a careful rounding procedure[5].

---

[5] Boyd and Vandenberghe 2004, Chapter 7.5.2.
[6] Pukelsheim 1993.
[7] Pronzato and Pázman 2013.

Finding optimal designs[6][7]

- Minimizing the posterior covariance is often replaced by

$$\min_{x_1,\dots,x_k \in \Omega} h\left(\varphi(X)^T \varphi(X) + \Lambda\right), \tag{4}$$

  where, e. g., $h = h_A \triangleq \mathrm{Tr}(\cdot^{-1})$ or $h = h_D \triangleq \det(\cdot^{-1})$.
- Still a nonconvex problem, requires enumeration when Ω finite.
- A convex relaxation to (4) is

$$\min_{\nu \in \mathcal{M}(\Omega)} h(G_\nu(\varphi) + \Lambda) \ s.t.\ \nu(\Omega) = k, \tag{5}$$

  where
  - $\mathcal{M}(\Omega)$ is the space of Borel measures on Ω,
  - $G_\nu(\varphi) = \left(\int_\Omega \varphi_i(x)\varphi_j(x) d\nu(x)\right) \in \mathbb{R}^{p\times p}$.
- One approximate solution to (4) is to sample i.i.d. from $\nu^\star$, and possibly apply a careful rounding procedure[5].

---

[5] Boyd and Vandenberghe 2004, Chapter 7.5.2.

[6] Pukelsheim 1993.

[7] Pronzato and Pázman 2013.

- Minimizing the posterior covariance is often replaced by

$$\min_{x_1,\dots,x_k\in\Omega} h\left(\varphi(X)^T\varphi(X)+\Lambda\right), \tag{4}$$

where, e. g., $h = h_A \triangleq \mathrm{Tr}(\cdot^{-1})$ or $h = h_D \triangleq \det(\cdot^{-1})$.

- Still a nonconvex problem, requires enumeration when Ω finite.
- A convex relaxation to (4) is

$$\min_{\nu\in\mathcal{M}(\Omega)} h(G_\nu(\varphi)+\Lambda) \ s.t.\ \nu(\Omega)=k, \tag{5}$$

where

  - $\mathcal{M}(\Omega)$ is the space of Borel measures on Ω,
  - $G_\nu(\varphi) = \left(\int_\Omega \varphi_i(x)\varphi_j(x)\mathrm{d}\nu(x)\right) \in \mathbb{R}^{p\times p}$.

- One approximate solution to (4) is to sample i.i.d. from $\nu^\star$, and possibly apply a careful rounding procedure[5].

[5] Boyd and Vandenberghe 2004, Chapter 7.5.2.

[6] Pukelsheim 1993.

[7] Pronzato and Pázman 2013.

- Minimizing the posterior covariance is often replaced by

$$\min_{x_1,\dots,x_k \in \Omega} h\left(\varphi(X)^T \varphi(X) + \Lambda\right), \qquad (4)$$

  where, e. g., $h = h_A \triangleq \mathrm{Tr}(\cdot^{-1})$ or $h = h_D \triangleq \det(\cdot^{-1})$.
- Still a nonconvex problem, requires enumeration when Ω finite.
- A convex relaxation to (4) is

$$\min_{\nu \in \mathcal{M}(\Omega)} h(G_\nu(\varphi) + \Lambda) \ s.t. \ \nu(\Omega) = k, \qquad (5)$$

  where
  - $\mathcal{M}(\Omega)$ is the space of Borel measures on Ω,
  - $G_\nu(\varphi) = \left(\int_\Omega \varphi_i(x)\varphi_j(x)\mathrm{d}\nu(x)\right) \in \mathbb{R}^{p\times p}$.
- One approximate solution to (4) is to sample i.i.d. from $\nu^\star$, and possibly apply a careful rounding procedure[5].

[5] Boyd and Vandenberghe 2004, Chapter 7.5.2.
[6] Pukelsheim 1993.
[7] Pronzato and Pázman 2013.

- Minimizing the posterior covariance is often replaced by

$$\min_{x_1,\dots,x_k\in\Omega} h\left(\varphi(X)^T\varphi(X)+\Lambda\right), \tag{4}$$

where, e. g., $h = h_A \triangleq \mathrm{Tr}(\cdot^{-1})$ or $h = h_D \triangleq \det(\cdot^{-1})$.

- Still a nonconvex problem, requires enumeration when Ω finite.
- A convex relaxation to (4) is

$$\min_{\nu\in\mathcal{M}(\Omega)} h(G_\nu(\varphi)+\Lambda)\ s.t.\ \nu(\Omega)=k, \tag{5}$$

where

  - $\mathcal{M}(\Omega)$ is the space of Borel measures on Ω,
  - $G_\nu(\varphi) = \left(\int_\Omega \varphi_i(x)\varphi_j(x)\mathrm{d}\nu(x)\right) \in \mathbb{R}^{p\times p}$.

- One approximate solution to (4) is to sample i.i.d. from $\nu^\star$, and possibly apply a careful rounding procedure[5].

[5] Boyd and Vandenberghe 2004, Chapter 7.5.2.

[6] Pukelsheim 1993.

[7] Pronzato and Pázman 2013.

- Minimizing the posterior covariance is often replaced by

$$\min_{x_1,\dots,x_k\in\Omega} h\left(\varphi(X)^T\varphi(X)+\Lambda\right), \tag{4}$$

  where, e. g., $h = h_A \triangleq \mathrm{Tr}(\cdot^{-1})$ or $h = h_D \triangleq \det(\cdot^{-1})$.
- Still a nonconvex problem, requires enumeration when Ω finite.
- A convex relaxation to (4) is

$$\min_{\nu\in\mathcal{M}(\Omega)} h(G_\nu(\varphi)+\Lambda)\ s.t.\ \nu(\Omega)=k, \tag{5}$$

  where
  - $\mathcal{M}(\Omega)$ is the space of Borel measures on Ω,
  - $G_\nu(\varphi) = \left(\int_\Omega \varphi_i(x)\varphi_j(x)\mathrm{d}\nu(x)\right) \in \mathbb{R}^{p\times p}$.
- One approximate solution to (4) is to sample i.i.d. from $\nu^\star$, and possibly apply a careful rounding procedure[5].

[5] Boyd and Vandenberghe 2004, Chapter 7.5.2.

[6] Pukelsheim 1993.

[7] Pronzato and Pázman 2013.

Definition-Proposition

PVS(ν, φ, Λ) is the point process on Ω with Janossy measures

$$j_n(x_1, \cdots, x_n)\mathrm{d}x_1 \cdots \mathrm{d}x_n = \frac{\det(\varphi(x)^T \varphi(x) + \Lambda)}{\det(G_\nu(\varphi) + \Lambda)\exp(\nu(\Omega))}\mathrm{d}\nu(x_1)\cdots\mathrm{d}\nu(x_n).$$

for all $n \in \mathbb{N}$ and $x \in \Omega^n$, where $\varphi(x) = (\varphi_i(x_j)) \in \mathbb{R}^{n\times p}$.

- Note how the number of points in random so far.
- PVS favorizes diversity over i.i.d. samples from ν.
- PVS generalizes seminal papers[8] focusing on Ω finite or $\Lambda = 0$. All of them condition on cardinality.

[8]Dereziński, Warmuth, and D.J. Hsu 2018; Nikolov, Singh, and Tantipongpipat 2019; Dereziński, Warmuth, and D. Hsu 2019; Dereziński, Liang, and M.W. Mahoney 2020.

Proposition

Let X be the DPP with kernel

$$\mathrm{K}(x, y) = \varphi(x)(G_\nu(\varphi) + \Lambda)^{-1}\varphi(y)^T$$

and reference measure ν, and let Y be an independent Poisson point process with intensity ν. Then $X \cup Y$ follows PVS(ν, φ, Λ).

- In particular, the average cardinality of the underlying DPP is

$$\mathrm{Tr}(G_\nu(\varphi)(G_\nu(\varphi) + \Lambda)^{-1}).$$

- When Λ is large, we shouldn't expect much repulsion.

Theoretical results[9]

- Recall the convex relaxation of optimal design

$$\min_{\nu \in \mathcal{M}(\Omega)} h(G_\nu(\varphi) + \Lambda) \ s.t. \ \nu(\Omega) = k, \tag{6}$$

- Almost by definition

$$\mathbb{E}\left[\det(\varphi(X)^T \varphi(X) + \Lambda)^{-1}\right] = \det(G_\nu(\varphi) + \Lambda)^{-1}. \tag{7}$$

- More subtly,

$$\begin{aligned} &\mathbb{E}\Big[\det(\varphi(X)^T \varphi(X) + \Lambda)^{-1} \Big| |X| = k\Big] \\ &\leqslant \frac{k^p (k-p)!}{k!} \frac{\det(G_\nu(\varphi) + \Lambda)^{-1}}{1 + \frac{p-1}{k-p+1}\left[1 - \det(G_\nu(\varphi)(G_\nu(\varphi) + \Lambda)^{-1})\right]} \end{aligned} \tag{8}$$

with equality when $\Lambda = 0$.

- This implies that

$$\mathbb{E}\left[\left(\frac{\det(\varphi(X)^T \varphi(X) + \Lambda)}{\det(\varphi(X_\star)^T \varphi(X_\star) + \Lambda)}\right)^{1/p} \Bigg| |X| = k\right] \geqslant 1 - \frac{p-1}{k}.$$

[9] Poinas and Bardenet 2021, In revision.

Theoretical results[9]

- Recall the convex relaxation of optimal design

$$\min_{\nu \in \mathcal{M}(\Omega)} h(G_\nu(\varphi) + \Lambda) \ s.t.\ \nu(\Omega) = k, \tag{6}$$

- Almost by definition

$$\mathbb{E}\left[\det(\varphi(X)^T\varphi(X) + \Lambda)^{-1}\right] = \det(G_\nu(\varphi) + \Lambda)^{-1}. \tag{7}$$

- More subtly,

$$\mathbb{E}\Big[\det(\varphi(X)^T\varphi(X) + \Lambda)^{-1}\Big||X| = k\Big] \leqslant \frac{k^p(k-p)!}{k!}\frac{\det(G_\nu(\varphi) + \Lambda)^{-1}}{1 + \frac{p-1}{k-p+1}\left[1 - \det(G_\nu(\varphi)(G_\nu(\varphi) + \Lambda)^{-1})\right]} \tag{8}$$

with equality when $\Lambda = 0$.

- This implies that

$$\mathbb{E}\left[\left(\frac{\det(\varphi(X)^T\varphi(X) + \Lambda)}{\det(\varphi(X_\star)^T\varphi(X_\star) + \Lambda)}\right)^{1/p}\middle||X| = k\right] \geqslant 1 - \frac{p-1}{k}.$$

[9] Poinas and Bardenet 2021, In revision.

Theoretical results[9]

- Recall the convex relaxation of optimal design

$$\min_{\nu \in \mathcal{M}(\Omega)} h(G_\nu(\varphi) + \Lambda) \; s.t. \; \nu(\Omega) = k, \tag{6}$$

- Almost by definition

$$\mathbb{E}\left[\det(\varphi(X)^T \varphi(X) + \Lambda)^{-1}\right] = \det(G_\nu(\varphi) + \Lambda)^{-1}. \tag{7}$$

- More subtly,

$$\mathbb{E}\Big[\det(\varphi(X)^T\varphi(X) + \Lambda)^{-1} \Big| |X| = k\Big] \leqslant \frac{k^p (k-p)!}{k!} \frac{\det(G_\nu(\varphi) + \Lambda)^{-1}}{1 + \frac{p-1}{k-p+1}\left[1 - \det(G_\nu(\varphi)(G_\nu(\varphi) + \Lambda)^{-1})\right]} \tag{8}$$

with equality when $\Lambda = 0$.

- This implies that

$$\mathbb{E}\left[\left(\frac{\det(\varphi(X)^T\varphi(X) + \Lambda)}{\det(\varphi(X_\star)^T\varphi(X_\star) + \Lambda)}\right)^{1/p} \middle| |X| = k\right] \geqslant 1 - \frac{p-1}{k}.$$

[9] Poinas and Bardenet 2021, In revision.

Theoretical results[9]

- Recall the convex relaxation of optimal design

$$\min_{\nu \in \mathcal{M}(\Omega)} h(G_\nu(\varphi) + \Lambda) \; s.t. \; \nu(\Omega) = k, \tag{6}$$

- Almost by definition

$$\mathbb{E}\left[\det(\varphi(X)^T\varphi(X) + \Lambda)^{-1}\right] = \det(G_\nu(\varphi) + \Lambda)^{-1}. \tag{7}$$

- More subtly,

$$\mathbb{E}\Big[\det(\varphi(X)^T\varphi(X) + \Lambda)^{-1}\Big| |X| = k\Big] \leqslant \frac{k^p (k-p)!}{k!} \frac{\det(G_\nu(\varphi) + \Lambda)^{-1}}{1 + \frac{p-1}{k-p+1}\left[1 - \det(G_\nu(\varphi)(G_\nu(\varphi) + \Lambda)^{-1})\right]} \tag{8}$$

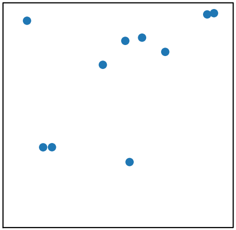
with equality when $\Lambda = 0$.

- This implies that

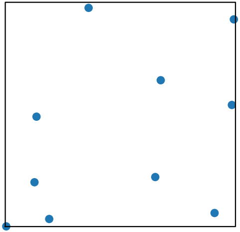
$$\mathbb{E}\left[\left(\frac{\det(\varphi(X)^T\varphi(X) + \Lambda)}{\det(\varphi(X_\star)^T\varphi(X_\star) + \Lambda)}\right)^{1/p} \Big| |X| = k\right] \geqslant 1 - \frac{p-1}{k}.$$

[9] Poinas and Bardenet 2021, In revision.

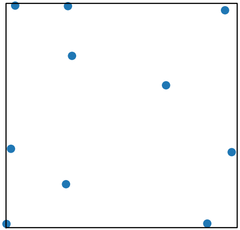
An example with $\Omega = [0,1]^2$, $\varphi_1, \ldots, \varphi_p$ all bivariate polynomials



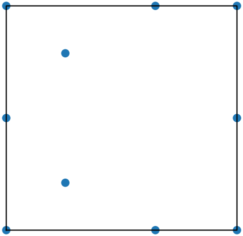
(c) Uniform distribution



(d) PVS (unif.)

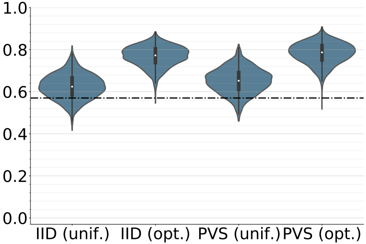


(e) PVS (opt.)

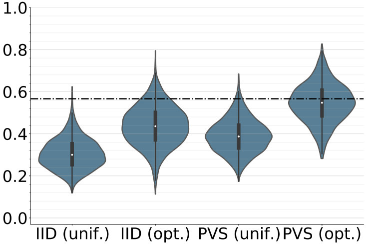


(f) D-Optimal design

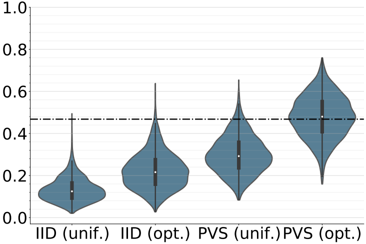
Plotting the D-efficiency



(g) $\Lambda = I_{10}$



(h) $\Lambda = 10^{-2} I_{10}$



(i) $\Lambda = 10^{-4} I_{10}$

Wrapping up

- VS links the repulsiveness of the nodes with the smoothness of the target.
- ☺ Volume sampling gives tight rates for interpolation in RKHSs.[10][11]
- ☺ If we manage to sample it, VS could be a powerful integration tool. See Yoann Jayer's PhD.
- VS yields elegant properties for optimal design in general design spaces.
- ☹ However, VS alone is not competitive with standard OD heuristics.
- VS can still be a useful component for stochastic search heuristics.[12]

PhD and postdoc applications welcome!

- remi.bardenet@gmail.com

[10] Belhadji, Bardenet, and Chainais 2019.
[11] Belhadji, Bardenet, and Chainais 2020a.
[12] Poinas and Bardenet 2021.

Wrapping up

- VS links the repulsiveness of the nodes with the smoothness of the target.
- ☺ Volume sampling gives tight rates for interpolation in RKHSs.[10][11]
- ☺ If we manage to sample it, VS could be a powerful integration tool. See Yoann Jayer's PhD.
- VS yields elegant properties for optimal design in general design spaces.
- ☹ However, VS alone is not competitive with standard OD heuristics.
- VS can still be a useful component for stochastic search heuristics.[12]

PhD and postdoc applications welcome!

- remi.bardenet@gmail.com

[10] Belhadji, Bardenet, and Chainais 2019.
[11] Belhadji, Bardenet, and Chainais 2020a.
[12] Poinas and Bardenet 2021.

Wrapping up

- VS links the repulsiveness of the nodes with the smoothness of the target.
- ☺ Volume sampling gives tight rates for interpolation in RKHSs.[10][11]
- ☺ If we manage to sample it, VS could be a powerful integration tool. See Yoann Jayer's PhD.
- VS yields elegant properties for optimal design in general design spaces.
- ☹ However, VS alone is not competitive with standard OD heuristics.
- VS can still be a useful component for stochastic search heuristics.[12]

PhD and postdoc applications welcome!

- remi.bardenet@gmail.com

[10] Belhadji, Bardenet, and Chainais 2019.
[11] Belhadji, Bardenet, and Chainais 2020a.
[12] Poinas and Bardenet 2021.

Wrapping up

- ▶ VS links the repulsiveness of the nodes with the smoothness of the target.
- ☺ Volume sampling gives tight rates for interpolation in RKHSs.[10][11]
- ☺ If we manage to sample it, VS could be a powerful integration tool. See Yoann Jayer's PhD.
- ▶ VS yields elegant properties for optimal design in general design spaces.
- ☹ However, VS alone is not competitive with standard OD heuristics.
- ▶ VS can still be a useful component for stochastic search heuristics.[12]

PhD and postdoc applications welcome!

- remi.bardenet@gmail.com

[10] Belhadji, Bardenet, and Chainais 2019.
[11] Belhadji, Bardenet, and Chainais 2020a.
[12] Poinas and Bardenet 2021.

Wrapping up

- VS links the repulsiveness of the nodes with the smoothness of the target.
- ☺ Volume sampling gives tight rates for interpolation in RKHSs.[10][11]
- ☺ If we manage to sample it, VS could be a powerful integration tool. See Yoann Jayer's PhD.
- VS yields elegant properties for optimal design in general design spaces.
- ☹ However, VS alone is not competitive with standard OD heuristics.
- VS can still be a useful component for stochastic search heuristics.[12]

PhD and postdoc applications welcome!

- remi.bardenet@gmail.com

[10] Belhadji, Bardenet, and Chainais 2019.
[11] Belhadji, Bardenet, and Chainais 2020a.
[12] Poinas and Bardenet 2021.

Wrapping up

- VS links the repulsiveness of the nodes with the smoothness of the target.
- ☺ Volume sampling gives tight rates for interpolation in RKHSs.[10][11]
- ☺ If we manage to sample it, VS could be a powerful integration tool. See Yoann Jayer's PhD.
- VS yields elegant properties for optimal design in general design spaces.
- ☹ However, VS alone is not competitive with standard OD heuristics.
- VS can still be a useful component for stochastic search heuristics.[12]

PhD and postdoc applications welcome!

- remi.bardenet@gmail.com

[10] Belhadji, Bardenet, and Chainais 2019.
[11] Belhadji, Bardenet, and Chainais 2020a.
[12] Poinas and Bardenet 2021.

References I

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)*.

– (2020a). "A determinantal point process for column subset selection". In: *Journal of Machine Learning Research (JMLR)*.

– (2020b). "Kernel interpolation with continuous volume sampling". In: *International Conference on Machine Learning (ICML)*.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. USA: Cambridge University Press.

Coeurjolly, J.-F., A. Mazoyer, and P.-O. Amblard (2021). "Monte Carlo integration of non-differentiable functions on $[0,1]^{\iota}$, $\iota = 1, \dots, d$, using a single determinantal point pattern defined on $[0,1]^{d}$". In: *Electronic Journal of Statistics*.

Dereziński, M., F. Liang, and M.W. Mahoney (26–28 Aug 2020). "Bayesian experimental design using regularized determinantal point processes". In: *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*. Ed. by Silvia Chiappa and Roberto Calandra. Vol. 108. Proceedings of Machine Learning Research. Online: PMLR, pp. 3197–3207.

Derezinski, M. and M. Mahoney (2020). "Determinantal Point Processes in Randomized Numerical Linear Algebra". In: *arXiv preprint arXiv:2005.03185*.

Dereziński, M., M.K. Warmuth, and D. Hsu (2019). *Unbiased estimators for random design regression*. arXiv pre-print.

Dereziński, M., M.K. Warmuth, and D.J. Hsu (2018). "Leveraged volume sampling for linear regression". In: *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems*, pp. 2510–2519.

Hough, J. B., M. Krishnapur, Y. Peres, and B. Virág (2006). "Determinantal processes and independence". In: *Probability surveys*.

Nikolov, A., M. Singh, and U. T. Tantipongpipat (2019). "Proportional Volume Sampling and Approximation Algorithms for A-Optimal Design". In: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '19. Society for Industrial and Applied Mathematics, pp. 1369–1386.

Pinkus, A. (2012). *N-widths in Approximation Theory*. Vol. 7. Springer Science & Business Media.

Poinas, A. and R. Bardenet (2021). "On proportional volume sampling for experimental design in general spaces". In: *In revision*.

Pronzato, L. and A. Pázman (2013). *Design of Experiments in Nonlinear Models: Asymptotic Normality, Optimality Criteria and Small-Sample Properties*. Lecture Notes in Statistics 212. Springer-Verlag New York.

Pukelsheim, F. (1993). *Optimal design of experiments*. Vol. 50. siam.

Rezaei, A. and S. O. Gharan (2019). "A Polynomial Time MCMC Method for Sampling from Continuous Determinantal Point Processes". In: *International Conference on Machine Learning*, pp. 5438–5447.