3.36pt

Nested Sampling Designs with Small Covering Radii

Luc Pronzato

(joint work with Anatoly Zhigljavsky, Cardiff Univ.)

Université Côte d'Azur, CNRS, France

March 10, 2022

1 Space-filling design

Objective: approximate $f(\cdot)$ over $\mathscr{X}$ (a compact subset of $\mathbb{R}^d$)
using pairs $(\mathbf{x}_i, f(\mathbf{x}_i))$, $i = 1, 2, \ldots, n$ → observe "everywhere"

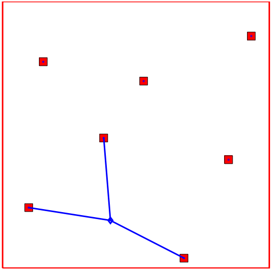
Design $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$

1 Space-filling design

Objective: approximate $f(\cdot)$ over $\mathscr{X}$ (a compact subset of $\mathbb{R}^d$)
using pairs $(\mathbf{x}_i, f(\mathbf{x}_i))$, $i = 1, 2, \ldots, n$ → observe "everywhere"
Design $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$

Covering radius $\mathrm{CR}(\mathbf{X}_n) = \mathrm{CR}_{\mathscr{X}}(\mathbf{X}_n) \triangleq \max_{\mathbf{x}\in\mathscr{X}} \min_{\mathbf{x}_i} \|\mathbf{x} - \mathbf{x}_i\|$

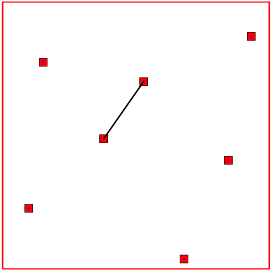
$\mathrm{CR}(\mathbf{X}_n)$ = fill distance = dispersion = miniMax distance criterion



→ we are never far from a design point

Packing radius $\mathrm{PR}(\mathbf{X}_n) \triangleq \frac{1}{2}\min_{i\neq j}\|\mathbf{x}_i - \mathbf{x}_j\|$

$\mathrm{PR}(\mathbf{X}_n)$ = separation radius = $\frac{1}{2}$ Maximin distance criterion

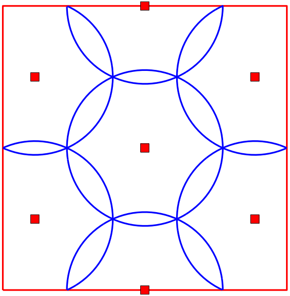


→ easier to compute, but pushes points to the boundary of $\mathscr{X}$

Examples:

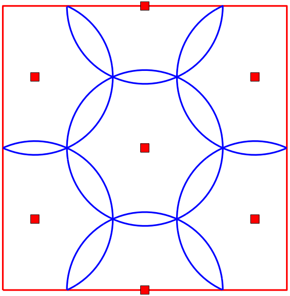
① Covering, miniMax

$d = 2, n = 7$ (radius=CR($\mathbf{X}_n$))

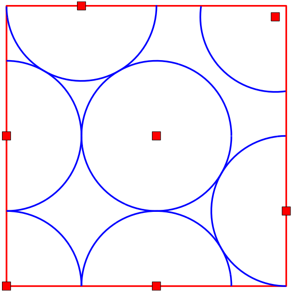


Examples:

① Covering, miniMax
$d = 2, n = 7$ (radius=CR($\mathbf{X}_n$))



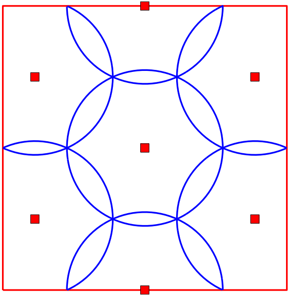
② Packing, Maximin
$d = 2, n = 7$ (radius=PR($\mathbf{X}_n$))



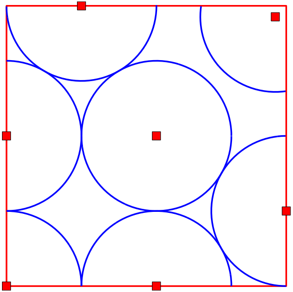
→ Minimise CR and maximise PR; both are difficult

Examples:

① Covering, miniMax
$d = 2, n = 7$ (radius=CR($\mathbf{X}_n$))



② Packing, Maximin
$d = 2, n = 7$ (radius=PR($\mathbf{X}_n$))



→ Minimise CR and maximise PR; both are difficult

We can also minimise the mesh-Ratio $\mathrm{MR}(\mathbf{X}_n) \triangleq \frac{\mathrm{CR}(\mathbf{X}_n)}{\mathrm{PR}(\mathbf{X}_n)}$

Why is the minimisation of $\mathrm{CR}(\mathbf{X}_n)$ important?

$\mathscr{X}$ bounded, with a Lipschitz boundary and satisfying an interior cone condition
K a Sobolev kernel of order α ($\alpha > d/2+1$ for d even, $\alpha > (d+1)/2$ for d odd)
$\eta_n^* =$ RBF interpolator (kriging predictor) for K

Th. (Narcowich et al., 2005): For $f \in W_2^\alpha(\mathscr{X})$, $1 \leq q \leq \infty$, $\exists C_q$ s.t.

$$\|f - \eta_n^*\|_{L_q} = \left(\int_{\mathscr{X}} |f(\mathbf{x}) - \eta_n^*(\mathbf{x})|^q \, \mathrm{d}\mathbf{x} \right)^{1/q} \leq C_q \, \|f\|_{W_2^\alpha(\mathscr{X})} \, \mathrm{CR}(\mathbf{X}_n)^{\alpha - d(1/2-1/q)_+}$$

$\forall$ $\mathbf{X}_n$ such that $\mathrm{CR}(\mathbf{X}_n)$ is small enough (C_q depends on α, d and $\mathscr{X}$)

Why is the minimisation of $\mathrm{CR}(\mathbf{X}_n)$ important?

$\mathscr{X}$ bounded, with a Lipschitz boundary and satisfying an interior cone condition
K a Sobolev kernel of order α ($\alpha > d/2+1$ for d even, $\alpha > (d+1)/2$ for d odd)
$\eta_n^* =$ RBF interpolator (kriging predictor) for K

Th. (Narcowich et al., 2005): For $f \in W_2^\alpha(\mathscr{X})$, $1 \leq q \leq \infty$, $\exists C_q$ s.t.

$$\begin{aligned} \|f - \eta_n^*\|_{L_q} &= \left(\int_{\mathscr{X}} |f(\mathbf{x}) - \eta_n^*(\mathbf{x})|^q \, \mathrm{d}\mathbf{x} \right)^{1/q} \\ &\leq C_q \, \|f\|_{W_2^\alpha(\mathscr{X})} \, \mathrm{CR}(\mathbf{X}_n)^{\alpha - d(1/2-1/q)_+} \end{aligned}$$

$\forall \; \mathbf{X}_n$ such that $\mathrm{CR}(\mathbf{X}_n)$ is small enough (C_q depends on α, d and $\mathscr{X}$)

What if f has lower smoothness than α? ➠ Escape theorem

Th. (Schaback and Wendland, 2006): For $f \in W_2^\beta(\mathscr{X})$, $\beta \leq \alpha$ ($\beta > d/2+1$ for d even and $\beta > (d+1)/2$ for d odd), $\exists C > 0$ s.t.

$$\|f - \eta_n^*\|_{L_2(\mathscr{X})} \leq C \, \|f\|_{W_2^\beta(\mathscr{X})} \, \mathrm{CR}(\mathbf{X}_n)^\beta \, \mathrm{MR}(\mathbf{X}_n)^{(\alpha-\beta)}$$

Objective: construct incremental designs with good-space-filling properties:
$\mathbf{X}_k \subset \mathbf{X}_{k+1} \subset \mathbf{X}_{k+3} \subset \ldots$ with small $\mathrm{CR}(\mathbf{X}_n)$ for $n \in [n_{\min}, n_{\max}]$

Objective: construct <u>incremental</u> designs with good-space-filling properties:
$\mathbf{X}_k \subset \mathbf{X}_{k+1} \subset \mathbf{X}_{k+3} \subset \ldots$ with small $\mathrm{CR}(\mathbf{X}_n)$ for $n \in [n_{\min}, n_{\max}]$

Two types of approaches:

A/ **greedy maximisation** of a set function $f(\mathbf{X}_n)$
If f is non-decreasing and submodular
➠ efficiency bound of (Nemhauser et al., 1978)
(or greedy minimisation of f non-increasing and supermodular)

Objective: construct incremental designs with good-space-filling properties:
$\mathbf{X}_k \subset \mathbf{X}_{k+1} \subset \mathbf{X}_{k+3} \subset \ldots$ with small $\mathrm{CR}(\mathbf{X}_n)$ for $n \in [n_{\min}, n_{\max}]$

Two types of approaches:

A/ **greedy maximisation** of a set function $f(\mathbf{X}_n)$
If f is non-decreasing and submodular
➠ efficiency bound of (Nemhauser et al., 1978)
(or greedy minimisation of f non-increasing and supermodular)

B/ apply a **gradient-type descent** algorithm to a suitable functional $\phi(\xi_n)$
with ξ_n the empirical measure $\xi_n = (1/n)\sum_{i=1}^n \delta_{\mathbf{x}_i}$ for $\mathbf{X}_n$
➠ sometimes simple enough to obtain a convergence rate

A/ Greedy Maximisation $\mathscr{X}_C \subset \mathscr{X}$ = a finite candidate set (C elements)

$f : 2^{\mathscr{X}_C} \to \mathbb{R}$ non-decreasing: $f(\mathcal{A} \cup \{\mathbf{x}\}) \geq f(\mathcal{A}), \forall \mathcal{A} \subset \mathscr{X}_C, \ \mathbf{x} \in \mathscr{X}_C$

A/ Greedy Maximisation $\mathscr{X}_C \subset \mathscr{X}$ = a finite candidate set (C elements)

$f : 2^{\mathscr{X}_C} \to \mathbb{R}$ non-decreasing: $f(\mathcal{A} \cup \{\mathbf{x}\}) \geq f(\mathcal{A})$, $\forall \mathcal{A} \subset \mathscr{X}_C,\ \mathbf{x} \in \mathscr{X}_C$

submodular: $\forall \mathcal{A} \subset \mathscr{B} \in 2^{\mathscr{X}_C},\ \mathbf{x} \in \mathscr{X}_C \setminus \mathscr{B}$,

$f(\mathcal{A} \cup \{\mathbf{x}\}) - f(\mathcal{A}) \geq f(\mathscr{B} \cup \{\mathbf{x}\}) - f(\mathscr{B})$ (diminishing returns property)

A/ Greedy Maximisation $\mathscr{X}_C \subset \mathscr{X}$ = a finite candidate set (C elements)

$f : 2^{\mathscr{X}_C} \to \mathbb{R}$ non-decreasing: $f(\mathcal{A} \cup \{\mathbf{x}\}) \geq f(\mathcal{A})$, $\forall \mathcal{A} \subset \mathscr{X}_C$, $\mathbf{x} \in \mathscr{X}_C$

submodular: $\forall \mathcal{A} \subset \mathscr{B} \in 2^{\mathscr{X}_C}$, $\mathbf{x} \in \mathscr{X}_C \setminus \mathscr{B}$,

$$f(\mathcal{A} \cup \{\mathbf{x}\}) - f(\mathcal{A}) \geq f(\mathscr{B} \cup \{\mathbf{x}\}) - f(\mathscr{B}) \text{ (diminishing returns property)}$$

Greedy Algorithm:

- set $\mathbf{X} = \emptyset$

while $|\mathbf{X}| < n$, find $\boxed{\mathbf{x} \in \mathrm{Arg}\max_{\mathbf{x} \in \mathscr{X}_C} f(\mathbf{X} \cup \{\mathbf{x}\})}$, $\mathbf{X} \leftarrow \mathbf{X} \cup \{\mathbf{x}\}$ end while

- **return X** $\boxed{\to \mathbf{X}_n^{GM}}$

→ Complexity $= \mathcal{O}(nC)$

$= \mathcal{O}(\gamma_n nC)$, $\gamma_n \ll 1$, for the lazy-greedy alg. of (Minoux, 1977)

A/ Greedy Maximisation $\mathscr{X}_C \subset \mathscr{X}$ = a finite candidate set (C elements)

$f : 2^{\mathscr{X}_C} \to \mathbb{R}$ non-decreasing: $f(\mathcal{A} \cup \{\mathbf{x}\}) \geq f(\mathcal{A})$, $\forall \mathcal{A} \subset \mathscr{X}_C$, $\mathbf{x} \in \mathscr{X}_C$

submodular: $\forall \mathcal{A} \subset \mathscr{B} \in 2^{\mathscr{X}_C}$, $\mathbf{x} \in \mathscr{X}_C \setminus \mathscr{B}$,

$$f(\mathcal{A} \cup \{\mathbf{x}\}) - f(\mathcal{A}) \geq f(\mathscr{B} \cup \{\mathbf{x}\}) - f(\mathscr{B}) \text{ (diminishing returns property)}$$

Greedy Algorithm:

- set $\mathbf{X} = \emptyset$

while $|\mathbf{X}| < n$, find $\boxed{\mathbf{x} \in \mathrm{Arg}\max_{\mathbf{x} \in \mathscr{X}_C} f(\mathbf{X} \cup \{\mathbf{x}\})}$, $\mathbf{X} \leftarrow \mathbf{X} \cup \{\mathbf{x}\}$ end while

- **return X** $\boxed{\to \mathbf{X}_n^{GM}}$

→ Complexity $= \mathcal{O}(nC)$
$= \mathcal{O}(\gamma_n nC)$, $\gamma_n \ll 1$, for the lazy-greedy alg. of (Minoux, 1977)

Th. (Nemhauser et al., 1978): f non-decreasing and submodular

$$\Rightarrow \forall k \in \{1, \ldots, C\}, \ \frac{f(\mathbf{X}_k) - f(\emptyset)}{f_k^\star - f(\emptyset)} \geq 1 - \frac{1}{e} > 63.2\%$$

where $f_k^\star = \max_{\mathbf{X} \subset \mathscr{X}_C : |\mathbf{X}| \leq k} f(\mathbf{X})$ and $e = \exp(1)$

B/ gradient-type descent = Vertex Direction algorithm

ϕ a convex functional on the set $\mathscr{M}_1^+(\mathscr{X})$ of probability measures on $\mathscr{X}$
$F_\phi(\xi;\nu)$ the directional derivative of $\phi(\cdot)$ at ξ in the direction ν:

$$F_K(\xi;\nu) \triangleq \lim_{\alpha\to 0^+} \frac{\phi[(1-\alpha)\xi+\alpha\nu]-\phi(\xi)}{\alpha}$$

B/ gradient-type descent = Vertex Direction algorithm

ϕ a convex functional on the set $\mathscr{M}_1^+(\mathscr{X})$ of probability measures on $\mathscr{X}$
$F_\phi(\xi;\nu)$ the directional derivative of $\phi(\cdot)$ at ξ in the direction ν:

$$F_K(\xi;\nu) \triangleq \lim_{\alpha \to 0^+} \frac{\phi[(1-\alpha)\xi + \alpha\nu] - \phi(\xi)}{\alpha}$$

Conditional gradient algorithm of (Frank and Wolfe, 1956):

iteration k: $\xi_k \in \mathscr{M}_1^+(\mathscr{X}) \to \xi_{k+1} = (1-\alpha_k)\xi_k + \alpha_k \delta_{\mathbf{x}_{k+1}}, \ \alpha_k \in [0,1]$

with $\boxed{\mathbf{x}_{k+1} \in \mathrm{Arg}\min_{\mathbf{x} \in \mathscr{X}} F_\phi(\xi_k;\delta_{\mathbf{x}})}$ (so that $\xi_{k+1} \in \mathscr{M}_1^+(\mathscr{X})$)

B/ gradient-type descent = Vertex Direction algorithm

ϕ a convex functional on the set $\mathscr{M}_1^+(\mathscr{X})$ of probability measures on $\mathscr{X}$
$F_\phi(\xi;\nu)$ the directional derivative of $\phi(\cdot)$ at ξ in the direction ν:

$$F_K(\xi;\nu) \triangleq \lim_{\alpha\to 0^+} \frac{\phi[(1-\alpha)\xi+\alpha\nu]-\phi(\xi)}{\alpha}$$

Conditional gradient algorithm of (Frank and Wolfe, 1956):

iteration k: $\xi_k \in \mathscr{M}_1^+(\mathscr{X}) \to \xi_{k+1} = (1-\alpha_k)\xi_k + \alpha_k\delta_{\mathbf{x}_{k+1}},\ \alpha_k \in [0,1]$

with $\boxed{\mathbf{x}_{k+1} \in \mathrm{Arg}\min_{\mathbf{x}\in\mathscr{X}} F_\phi(\xi_k;\delta_{\mathbf{x}})}$ (so that $\xi_{k+1} \in \mathscr{M}_1^+(\mathscr{X})$)

Take $\xi_1 = \delta_{\mathbf{x}_1}$ and $\alpha_k = 1/(k+1)$ for all k
= Wynn's **Vertex-Direction** algorithm (1972) for DOE
➞ $\xi_n = (1/n)\sum_{i=1}^n \delta_{\mathbf{x}_i}$ = empirical measure on $\mathbf{X}_n$
$\boxed{\to \mathbf{X}_n^{VD}}$

In practice: replace $\mathscr{X}$ by $\mathscr{X}_C \subset \mathscr{X}$ (with C elements) to choose $\mathbf{x}_{k+1}$
Complexity = $\mathcal{O}(nC)$

Four approaches considered, based on:

1. minimisation of a relaxed version of $\mathrm{CR}(\mathbf{X}_n)$ → A and B
2. minimisation of a Maximum-Mean-Discrepancy (MMD) $=$ distance between ξ_n and μ uniform on $\mathscr{X}$ → A and B
3. maximisation of an integrated covering measure → A
4. geometrical considerations: greedy packing (coffee-house design) → A

2 Minimisation of a relaxed version of CR($\mathbf{X}_n$)

$\mathrm{CR}(\mathbf{X}_n) \triangleq \max_{\mathbf{x}\in\mathscr{X}} \min_{\mathbf{x}_i} \|\mathbf{x}-\mathbf{x}_i\|$ ➠ ℓ_q and L_q relaxations

2 Minimisation of a relaxed version of CR($\mathbf{X}_n$)

$\mathrm{CR}(\mathbf{X}_n) \triangleq \max_{\mathbf{x}\in\mathscr{X}} \min_{\mathbf{x}_i} \|\mathbf{x}-\mathbf{x}_i\|$ ➠ ℓ_q and L_q relaxations

A/ for $\mathbf{X}_n \subset \mathscr{X}$, μ uniform on $\mathscr{X}$ and $q > 0$, denote

$$\Phi_q(\mathbf{X}_n) = \Phi_q(\mathbf{X}_n;\mu) \triangleq \left[\int_{\mathscr{X}} \left(\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{x}-\mathbf{x}_i\|^{-q}\right)^{-1} \mathrm{d}\mu(\mathbf{x})\right]^{1/q}$$

➞ $\forall \mathbf{X}_n \subset \mathscr{X}$, $\Phi_q(\mathbf{X}_n) \to \mathrm{CR}(\mathbf{X}_n)$ as $q \to \infty$

$\mathbf{X}_n \subset \mathscr{X} \to (1/n)\,\Phi_q^q(\mathbf{X}_n,\mu)$ is non-increasing and supermodular

2 Minimisation of a relaxed version of CR($\mathbf{X}_n$)

$\mathrm{CR}(\mathbf{X}_n) \triangleq \max_{\mathbf{x}\in\mathscr{X}} \min_{\mathbf{x}_i} \|\mathbf{x} - \mathbf{x}_i\|$ ➟ ℓ_q and L_q relaxations

A/ for $\mathbf{X}_n \subset \mathscr{X}$, μ uniform on $\mathscr{X}$ and $q > 0$, denote

$$\Phi_q(\mathbf{X}_n) = \Phi_q(\mathbf{X}_n; \mu) \triangleq \left[\int_{\mathscr{X}} \left(\frac{1}{n} \sum_{i=1}^{n} \|\mathbf{x} - \mathbf{x}_i\|^{-q}\right)^{-1} \mathrm{d}\mu(\mathbf{x})\right]^{1/q}$$

➞ $\forall \mathbf{X}_n \subset \mathscr{X}$, $\Phi_q(\mathbf{X}_n) \to \mathrm{CR}(\mathbf{X}_n)$ as $q \to \infty$

$\mathbf{X}_n \subset \mathscr{X} \to (1/n)\, \Phi_q^q(\mathbf{X}_n, \mu)$ is non-increasing and supermodular

➟ Greedy Minimisation, with a candidate set $\mathscr{X}_C \subset \mathscr{X}$ | ➞ $\mathbf{X}_n^{GM}$ |

Replace μ by a Q-point discrete approximation μ_Q ($\to \Phi_q(\mathbf{X}_n; \mu_Q)$)
➞ 2 discrete sets $\mathscr{X}_C$ and $\mathscr{X}_Q$ (compute $C \times Q$ pairwise distances)
Complexity $= \mathcal{O}(nCQ)$ ($\mathcal{O}(\gamma_n nCQ)$ for lazy-greedy version)

B/ for $\xi \in \mathscr{M}_1^+(\mathscr{X})$, μ uniform on $\mathscr{X}$ and $q > 0$, denote

$$\phi_q(\xi) = \phi_q(\xi;\mu) \triangleq \left[\int_{\mathscr{X}} \left(\int_{\mathscr{X}} \|\mathbf{z}-\mathbf{x}\|^{-q}\,\mathrm{d}\xi(\mathbf{z})\right)^{-1}\,\mathrm{d}\mu(\mathbf{x})\right]^{1/q},\ \ q \neq 0\,,$$

➠ $\phi_q(\xi_n) = \Phi_q(\mathbf{X}_n)$ for $\xi_n = (1/n)\sum_{i=1}^n \delta_{\mathbf{x}_i}$

(P & Zhigljavsky, 2019): $\phi_q^q(\cdot)$ is convex for $q > 0$ (strictly if $q \in (0,d)$),
$F_\phi(\xi_k;\delta_{\mathbf{x}})$ known explicitly

B/ for $\xi \in \mathscr{M}_1^+(\mathscr{X})$, μ uniform on $\mathscr{X}$ and $q > 0$, denote

$$\phi_q(\xi) = \phi_q(\xi;\mu) \triangleq \left[\int_{\mathscr{X}} \left(\int_{\mathscr{X}} \|\mathbf{z}-\mathbf{x}\|^{-q}\,\mathrm{d}\xi(\mathbf{z})\right)^{-1} \mathrm{d}\mu(\mathbf{x})\right]^{1/q}, \; q \neq 0\,,$$

➠ $\phi_q(\xi_n) = \Phi_q(\mathbf{X}_n)$ for $\xi_n = (1/n)\sum_{i=1}^n \delta_{\mathbf{x}_i}$

(P & Zhigljavsky, 2019): $\phi_q^q(\cdot)$ is convex for $q > 0$ (strictly if $q \in (0,d)$),
$F_\phi(\xi_k;\delta_{\mathbf{x}})$ known explicitly

Minimise by conditional gradient, with a candidate set $\mathscr{X}_C \subset \mathscr{X}$

➠ Vertex-Direction algorithm **→ $\mathbf{X}_n^{VD}$**

Replace μ by a Q-point discrete approximation μ_Q ($\rightarrow \phi_q(\xi_n;\mu_Q)$)
→ 2 discrete sets $\mathscr{X}_C$ and $\mathscr{X}_Q$ (compute $C \times Q$ pairwise distances)
Complexity $= \mathcal{O}(nCQ)$

B/ for $\xi \in \mathscr{M}_1^+(\mathscr{X})$, μ uniform on $\mathscr{X}$ and $q > 0$, denote

$$\phi_q(\xi) = \phi_q(\xi;\mu) \triangleq \left[\int_{\mathscr{X}} \left(\int_{\mathscr{X}} \|\mathbf{z}-\mathbf{x}\|^{-q}\, \mathrm{d}\xi(\mathbf{z})\right)^{-1} \mathrm{d}\mu(\mathbf{x})\right]^{1/q},\ q \neq 0\,,$$

➠ $\phi_q(\xi_n) = \Phi_q(\mathbf{X}_n)$ for $\xi_n = (1/n)\sum_{i=1}^n \delta_{\mathbf{x}_i}$

(P & Zhigljavsky, 2019): $\phi_q^q(\cdot)$ is convex for $q > 0$ (strictly if $q \in (0,d)$), $F_\phi(\xi_k;\delta_{\mathbf{x}})$ known explicitly

Minimise by conditional gradient, with a candidate set $\mathscr{X}_C \subset \mathscr{X}$

➠ Vertex-Direction algorithm $\boxed{\rightarrow \mathbf{X}_n^{VD}}$

Replace μ by a Q-point discrete approximation μ_Q ($\rightarrow \phi_q(\xi_n;\mu_Q)$)

→ 2 discrete sets $\mathscr{X}_C$ and $\mathscr{X}_Q$ (compute $C \times Q$ pairwise distances)

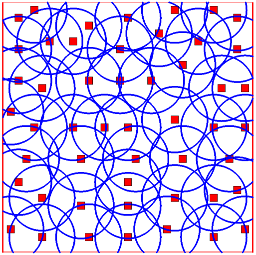
Complexity $= \mathcal{O}(nCQ)$

2 remarks:

- Minimisation of $\phi_q^q(\xi;\mu_Q) \Leftrightarrow A$-optimal design ($\mathrm{trace}[\mathbf{M}^{-1}(\xi)]$ min! for a particular information matrix $\mathbf{M}(\xi)$)
- The optimal measure ξ^* is not uniform on $\mathscr{X}$

Example: $d = 2$, $\mathscr{X} = [0,1]^2$, $n = 50$, $q = 10$
$\mathscr{X}_C = 33 \times 33$ regular grid, $\mathscr{X}_Q = 32 \times 32$ interlaced grid

greedy min. of $\Phi_q(\mathbf{X}_n; \mu_Q)$

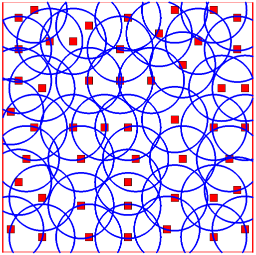


$\mathbf{X}_n^{GM}$

(radius $=$ CR($\mathbf{X}_n$))

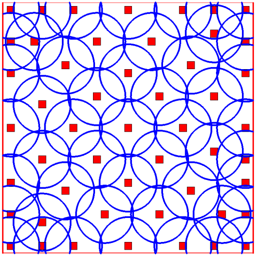
Example: $d = 2$, $\mathscr{X} = [0,1]^2$, $n = 50$, $q = 10$
$\mathscr{X}_C = 33 \times 33$ regular grid, $\mathscr{X}_Q = 32 \times 32$ interlaced grid

greedy min. of $\Phi_q(\mathbf{X}_n; \mu_Q)$



$\mathbf{X}_n^{GM}$

cond. grad. with $\phi_q^q(\xi_n; \mu_Q)$



$\mathbf{X}_n^{VD}$

(radius $=$ CR($\mathbf{X}_n$))

3 Minimisation of a Maximum-Mean-Discrepancy (MMD)

Very much based on:

Sriperumbudur, B., Gretton, A., Fukumizu, K., Schölkopf, B., Lanckriet, G., 2010. Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research* 11 (Apr), 1517–1561.

Sejdinovic, S., Sriperumbudur, B., Gretton, A., Fukumizu, K., 2013. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics* 41 (5), 2263–2291.

Pronzato, L., Zhigljavsky, A., 2020. Bayesian quadrature, energy minimization and space-filling design. SIAM/ASA J. Uncertainty Quantification 8 (3), 959–1011.

3 Minimisation of a Maximum-Mean-Discrepancy (MMD)

Very much based on:

Sriperumbudur, B., Gretton, A., Fukumizu, K., Schölkopf, B., Lanckriet, G., 2010. Hilbert space embeddings and metrics on probability measures. *Journal of Machine Learning Research* 11 (Apr), 1517–1561.

Sejdinovic, S., Sriperumbudur, B., Gretton, A., Fukumizu, K., 2013. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *The Annals of Statistics* 41 (5), 2263–2291.

Pronzato, L., Zhigljavsky, A., 2020. Bayesian quadrature, energy minimization and space-filling design. SIAM/ASA J. Uncertainty Quantification 8 (3), 959–1011.

Let K be a PD kernel on $\mathscr{X} \times \mathscr{X}$, $\mathcal{H}_K$ the associated RKHS
For ν a signed measure on $\mathscr{X}$, define

$$\mathscr{E}_K(\nu) \triangleq \int_{\mathscr{X}^2} K(\mathbf{x},\mathbf{x}')\,\mathrm{d}\nu(\mathbf{x})\mathrm{d}\nu(\mathbf{x}') = \text{energy of } \nu$$

$$P_{K,\nu}(\mathbf{x}) \triangleq \int_{\mathscr{X}} K(\mathbf{x},\mathbf{x}')\,\mathrm{d}\nu(\mathbf{x}') = \text{potential of } \nu \text{ at } \mathbf{x}$$

$[P_{K,\nu}(\cdot) = \text{kernel imbedding of } \nu \text{ into } \mathcal{H}_K]$

For $\underline{f \in \mathcal{H}_K}$, $\mu, \nu \in \mathscr{M}_1^+(\mathscr{X})$ with finite energy

RKHS property $[K_{\mathbf{x}}(\cdot) = K(\mathbf{x}, \cdot)] \Rightarrow$

$$|I_\mu(f) - I_\nu(f)| = \left| \int_{\mathscr{X}} \langle f, K_{\mathbf{x}} \rangle_K \, \mathrm{d}(\mu - \nu)(\mathbf{x}) \right| = |\langle f, P_{K,\mu} - P_{K,\nu} \rangle_K|$$

CS inequality ➞ a Koksma-Hlawka type inequality:

$$\left| \int_{\mathscr{X}} f(\mathbf{x}) \, \mathrm{d}\nu(\mathbf{x}) - \int_{\mathscr{X}} f(\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x}) \right| \leq \|f\|_{\mathcal{H}_K} \mathrm{MMD}_K(\mu, \nu)$$

$$\text{where } \boxed{\mathrm{MMD}_K(\mu, \nu) \triangleq \|P_{K,\mu} - P_{K,\nu}\|_{\mathcal{H}_K} = \mathscr{E}_K^{1/2}(\nu - \mu)}$$

$\mathrm{MMD}_K(\mu, \nu)$ = **Maximum Mean Discrepancy** between μ and ν
(Sriperumbudur et al., 2010; Sejdinovic et al., 2013)

For $\underline{f \in \mathcal{H}_K}$, $\mu, \nu \in \mathscr{M}_1^+(\mathscr{X})$ with finite energy

RKHS property $[K_{\mathbf{x}}(\cdot) = K(\mathbf{x}, \cdot)] \Rightarrow$

$$|I_\mu(f) - I_\nu(f)| = \left| \int_{\mathscr{X}} \langle f, K_{\mathbf{x}} \rangle_K \, \mathrm{d}(\mu - \nu)(\mathbf{x}) \right| = |\langle f, P_{K,\mu} - P_{K,\nu} \rangle_K|$$

CS inequality ➞ a Koksma-Hlawka type inequality:

$$\left| \int_{\mathscr{X}} f(\mathbf{x}) \, \mathrm{d}\nu(\mathbf{x}) - \int_{\mathscr{X}} f(\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x}) \right| \leq \|f\|_{\mathcal{H}_K} \mathrm{MMD}_K(\mu, \nu)$$

$$\text{where } \boxed{\mathrm{MMD}_K(\mu, \nu) \triangleq \|P_{K,\mu} - P_{K,\nu}\|_{\mathcal{H}_K} = \mathscr{E}_K^{1/2}(\nu - \mu)}$$

$\mathrm{MMD}_K(\mu, \nu)$ = **Maximum Mean Discrepancy** between μ and ν
(Sriperumbudur et al., 2010; Sejdinovic et al., 2013)

Space-filling design: take μ uniform on $\mathscr{X}$
➞ find ξ_n (with n support points) minimising $\mathrm{MMD}_K^2(\xi_n, \mu) = \mathscr{E}_K(\xi_n - \mu)$

For $f \in \mathcal{H}_K$, $\mu, \nu \in \mathscr{M}_1^+(\mathscr{X})$ with finite energy

RKHS property $[K_{\mathbf{x}}(\cdot) = K(\mathbf{x}, \cdot)] \Rightarrow$

$$|I_\mu(f) - I_\nu(f)| = \left| \int_{\mathscr{X}} \langle f, K_{\mathbf{x}} \rangle_K \, \mathrm{d}(\mu - \nu)(\mathbf{x}) \right| = |\langle f, P_{K,\mu} - P_{K,\nu} \rangle_K|$$

CS inequality ➞ a Koksma-Hlawka type inequality:

$$\left| \int_{\mathscr{X}} f(\mathbf{x}) \, \mathrm{d}\nu(\mathbf{x}) - \int_{\mathscr{X}} f(\mathbf{x}) \, \mathrm{d}\mu(\mathbf{x}) \right| \leq \|f\|_{\mathcal{H}_K} \mathrm{MMD}_K(\mu, \nu)$$

where $\boxed{\mathrm{MMD}_K(\mu, \nu) \triangleq \|P_{K,\mu} - P_{K,\nu}\|_{\mathcal{H}_K} = \mathscr{E}_K^{1/2}(\nu - \mu)}$

$\mathrm{MMD}_K(\mu, \nu)$ = **Maximum Mean Discrepancy** between μ and ν (Sriperumbudur et al., 2010; Sejdinovic et al., 2013)

Space-filling design: take μ uniform on $\mathscr{X}$

➞ find ξ_n (with n support points) minimising $\mathrm{MMD}_K^2(\xi_n, \mu) = \mathscr{E}_K(\xi_n - \mu)$

➞ "classical" L_2-discrepancies (extreme, centered, symmetric, wrap-around...) are obtained for particular kernels (Hickernell, 1998)

$\text{MMD}_K(\cdot,\cdot)$ defines a pseudo-metric on $\mathscr{M}_1^+$
Does it define a metric? $\Leftrightarrow$ K is **characteristic**

Definition

K is Integrally Strictly Positive Definite (ISPD) on $\mathscr{M}$ (set of finite signed Borel measures on $\mathscr{X}$) when $\mathscr{E}_K(\nu) > 0$ for any nonzero $\nu \in \mathscr{M}$

Definition

K is Conditionally Integrally Strictly Positive Definite (CISPD) on $\mathscr{M}$ when it is ISPD on $\mathscr{M}_0$; that is, when $\mathscr{E}_K(\nu) > 0$ for all nonzero $\nu \in \mathscr{M}$ with $\nu(\mathscr{X}) = 0$

$\text{MMD}_K(\cdot,\cdot)$ defines a pseudo-metric on $\mathscr{M}_1^+$
Does it define a metric? $\Leftrightarrow$ K is **characteristic**

Definition

K is Integrally Strictly Positive Definite (ISPD) on $\mathscr{M}$ (set of finite signed Borel measures on $\mathscr{X}$) when $\mathscr{E}_K(\nu) > 0$ for any nonzero $\nu \in \mathscr{M}$

Definition

K is Conditionally Integrally Strictly Positive Definite (CISPD) on $\mathscr{M}$ when it is ISPD on $\mathscr{M}_0$; that is, when $\mathscr{E}_K(\nu) > 0$ for all nonzero $\nu \in \mathscr{M}$ with $\nu(\mathscr{X}) = 0$

Sriperumbudur et al. (2010):

- K bounded & ISPD $\Rightarrow$ K is strictly positive definite
  ($\rightarrow$ defines a RKHS $\mathcal{H}_K$)
- if K uniformly bounded: characteristic $\Leftrightarrow$ CISPD

$\mathrm{MMD}_K(\cdot,\cdot)$ defines a pseudo-metric on $\mathscr{M}_1^+$

Does it define a metric? $\Leftrightarrow$ K is **characteristic**

Definition

K is Integrally Strictly Positive Definite (ISPD) on $\mathscr{M}$ (set of finite signed Borel measures on $\mathscr{X}$) when $\mathscr{E}_K(\nu) > 0$ for any nonzero $\nu \in \mathscr{M}$

Definition

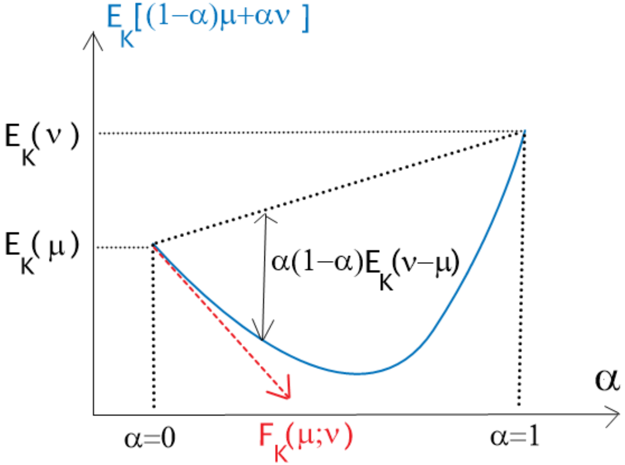
K is Conditionally Integrally Strictly Positive Definite (CISPD) on $\mathscr{M}$ when it is ISPD on $\mathscr{M}_0$; that is, when $\mathscr{E}_K(\nu) > 0$ for all nonzero $\nu \in \mathscr{M}$ with $\nu(\mathscr{X}) = 0$

Sriperumbudur et al. (2010):

- K bounded & ISPD $\Rightarrow$ K is strictly positive definite (→ defines a RKHS $\mathcal{H}_K$)
- if $\underbrace{K \text{ uniformly bounded}}_{\text{assumed in the following}}$: characteristic $\Leftrightarrow$ CISPD

For many kernels K (Gaussian, Matérn, distance-induced kernels of Székely and Rizzo (2013)...):

- $\mathrm{MMD}_K(\cdot,\cdot)$ defines a metric for probability measures
- $\mathscr{E}_K(\cdot)$ is strictly convex



A/ Greedy MMD Minimisation: $\xi_n = (1/n)\sum_{i=1}^n \delta_{\mathbf{x}_i}$

$\mathrm{MMD}_K^2(\mu,\xi_n) = \mathscr{E}_K(\xi_n - \mu) = \mathbf{1}_n^\top \mathbf{K}_n \mathbf{1}_n - 2\,\mathbf{1}_n^\top \mathbf{p}_n(\mu) + \mathscr{E}_K(\mu)$

where $\mathbf{1}_n = (1,\ldots,1)^\top$, $\{\mathbf{K}_n\}_{i,j} = K(\mathbf{x}_i,\mathbf{x}_j)$,

and $\mathbf{p}_n(\mu) = [P_{K,\mu}(\mathbf{x}_1),\ldots,P_{K,\mu}(\mathbf{x}_n)]^\top$

→ $\mathbf{x}_{k+1}$ minimises $\mathrm{MMD}_K^2(\mu,\xi_{n+1})$

➠ $\mathbf{x}_{k+1} \in \mathrm{Arg}\min_{\mathbf{x}\in\mathscr{X}} \sum_{i=1}^k K(\mathbf{x}_i,\mathbf{x}) + \frac{1}{2}\,K(\mathbf{x},\mathbf{x}) - (k+1)\,P_{K,\mu}(\mathbf{x})$

→ $\mathbf{X}_n^{GM}$

A/ Greedy MMD Minimisation: $\xi_n = (1/n)\sum_{i=1}^n \delta_{\mathbf{x}_i}$

$$\mathrm{MMD}_K^2(\mu,\xi_n) = \mathscr{E}_K(\xi_n - \mu) = \mathbf{1}_n^\top \mathbf{K}_n \mathbf{1}_n - 2\,\mathbf{1}_n^\top \mathbf{p}_n(\mu) + \mathscr{E}_K(\mu)$$

where $\mathbf{1}_n = (1,\ldots,1)^\top$, $\{\mathbf{K}_n\}_{i,j} = K(\mathbf{x}_i,\mathbf{x}_j)$,
and $\mathbf{p}_n(\mu) = [P_{K,\mu}(\mathbf{x}_1),\ldots,P_{K,\mu}(\mathbf{x}_n)]^\top$

→ $\mathbf{x}_{k+1}$ minimises $\mathrm{MMD}_K^2(\mu,\xi_{n+1})$

➠ $\mathbf{x}_{k+1} \in \mathrm{Arg}\min_{\mathbf{x}\in\mathscr{X}} \sum_{i=1}^k K(\mathbf{x}_i,\mathbf{x}) + \frac{1}{2}K(\mathbf{x},\mathbf{x}) - (k+1)\,P_{K,\mu}(\mathbf{x})$

→ $\mathbf{X}_n^{GM}$

Remark: Sequential Bayesian Quadrature = greedy MMD minimisation for $\xi_n^* = \sum_{i=1}^n w_i\,\delta_{\mathbf{x}_i}$ with optimal weights $(w_1,\ldots,w_n) = \mathbf{p}_n^\top(\mu)\mathbf{K}_n^{-1}$

→ $\mathrm{MMD}_K^2(\mu,\xi_n^*) = \mathscr{E}_K(\xi_n^* - \mu) = \mathscr{E}_K(\mu) - \mathbf{p}_n^\top(\mu)\mathbf{K}_n^{-1}\mathbf{p}_n(\mu)$

B/ Conditional gradient descent (Vertex-Direction algorithm):

Directional derivative of $\mathscr{E}_K(\cdot)$ at ξ in the direction ν:

$$\begin{aligned} F_K(\xi;\nu) &= \lim_{\alpha\to 0^+} \frac{\mathscr{E}_K[(1-\alpha)\xi+\alpha\nu]-\mathscr{E}_K(xi)}{\alpha} \\ &= 2\left[\int_{\mathscr{X}^2} K(\mathbf{x},\mathbf{x}')\,\mathrm{d}\nu(\mathbf{x})\mathrm{d}\xi(\mathbf{x}') - \mathscr{E}_K(\xi)\right] \end{aligned}$$

$$\Rightarrow \boxed{F_K(\xi;\delta_{\mathbf{x}}) = 2[P_{K,\xi}(\mathbf{x}) - \mathscr{E}_K(\xi)]}$$

B/ Conditional gradient descent (Vertex-Direction algorithm):

Directional derivative of $\mathscr{E}_K(\cdot)$ at ξ in the direction ν:

$$
\begin{aligned}
F_K(\xi;\nu) &= \lim_{\alpha\to 0^+} \frac{\mathscr{E}_K[(1-\alpha)\xi+\alpha\nu]-\mathscr{E}_K(xi)}{\alpha} \\
&= 2\left[\int_{\mathscr{X}^2} K(\mathbf{x},\mathbf{x}')\,\mathrm{d}\nu(\mathbf{x})\mathrm{d}\xi(\mathbf{x}') - \mathscr{E}_K(\xi)\right]
\end{aligned}
$$

$$\Rightarrow \boxed{F_K(\xi;\delta_{\mathbf{x}}) = 2[P_{K,\xi}(\mathbf{x}) - \mathscr{E}_K(\xi)]}$$

We do not want to minimise $\mathscr{E}_K(\xi)$ but $\mathscr{E}_K(\xi-\mu) = \mathrm{MMD}_K^2(\xi,\mu)$ for a given μ

➞ $F_{\mathrm{MMD}_K^2}(\xi,\delta_{\mathbf{x}}) = 2\left[\boxed{P_{K,\xi}(\mathbf{x}) - P_{K,\mu}(\mathbf{x})} + \int_{\mathscr{X}} P_{K,\mu}(\mathbf{x}')\,\mathrm{d}\xi(\mathbf{x}') - \mathscr{E}_K(\xi)\right]$

➠ $\boxed{\mathbf{x}_{k+1} \in \mathrm{Arg\,min}_{\mathbf{x}\in\mathscr{X}} \left[\frac{1}{k}\sum_{i=1}^k K(\mathbf{x},\mathbf{x}_i) - P_\mu(\mathbf{x})\right]}$

This is called *Kernel Herding* in machine learning: $\boxed{\rightarrow \mathbf{X}_n^{KH}}$

3 remarks:

- Greedy MMD minimisation and kernel herding behave similarly, with $\text{MMD}(\xi_n, \mu)$ decreasing like $\log(n)/n$,
- In practice, use a finite candidate set $\mathscr{X}_C$
  ($\rightarrow$ complexity $= \mathcal{O}(nC)$ for n iterations)

3 remarks:

- Greedy MMD minimisation and kernel herding behave similarly, with $\mathrm{MMD}(\xi_n, \mu)$ decreasing like $\log(n)/n$,
- In practice, use a finite candidate set $\mathscr{X}_C$
  ($\rightarrow$ complexity $= \mathcal{O}(nC)$ for n iterations)
- $\mathbf{x}_{k+1}$ is easy to determine when $P_{K,\mu}(\mathbf{x})$ is available
  - replace μ by a discrete measure μ_Q
    $\rightarrow$ the *support points* of (Mak and Joseph, 2018) minimise $\mathrm{MMD}_K(\xi_n, \mu)$ for K the energy-distance kernel Székely and Rizzo (2013)
  - compute $P_{K,\mu}(\mathbf{x})$ explicitly when:

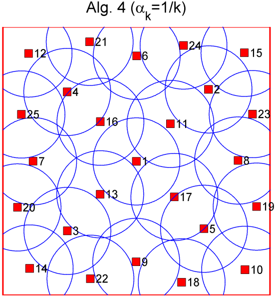
- K is separable on $\mathscr{X} = \times_{i=1}^d \mathscr{X}_i$: $K(\mathbf{x}, \mathbf{x}') = K^{\otimes}(\mathbf{x}, \mathbf{x}') = \prod_{i=1}^d K_i(x_i, x_i')$
- $\mu = \otimes_{i=1}^d \mu_i$ is a product measure on $\mathscr{X} = \times_{i=1}^d \mathscr{X}_i$

$\rightarrow$ $P_{K,\mu}(\mathbf{x}) = \prod_{i=1}^d P_{K_i,\mu_i}(x_i)$

(= product of one dimensional integrals)

Example: $\mathscr{X} = [0,1]^2$, $n = 25$, $\mathscr{X}_C = 2^{17} = 131\,072$ Sobol' points
K = tensor product of Matérn 3/2

$$K_{3/2,\theta}(x,x') = (1+\sqrt{3}\theta|x-x'|)\exp(-\sqrt{3}\theta|x-x'|),\ \theta = 10$$



(radius = $\mathrm{CR}(\mathbf{X}_n^{KH})$)

Minimum-Norm-point algorithm of (Bach et al., 2012):

replace ξ_n (uniform on its support) by $\hat{\xi}_n$ having the same support but optimal weights, positive with sum $= 1$

➠ Simpler version: use optimal weights with sum $= 1$ (explicit form)

(extra comput. cost → $\mathcal{O}(n^2 C)$ for n iterations (P., 2021)) → $\mathbf{X}_n^{MN}$

Minimum-Norm-point algorithm of (Bach et al., 2012):

replace ξ_n (uniform on its support) by $\hat{\xi}_n$ having the same support but optimal weights, positive with sum $= 1$

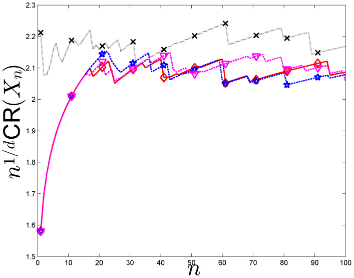
➠ Simpler version: use optimal weights with sum $= 1$ (explicit form)

(extra comput. cost → $\mathcal{O}(n^2 C)$ for n iterations (P., 2021)) → $\mathbf{X}_n^{MN}$

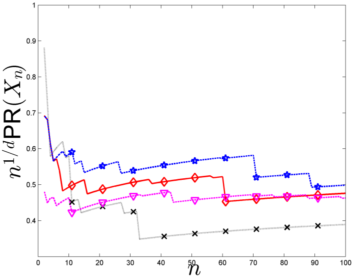
➠ Comparison with Sobol' sequence → $\mathbf{X}_n^{S}$

Example: $d = 10$, $\mathscr{X}_C = 2^{12}$ points of scrambled Sobol' in $\mathscr{X} = [0,1]^{10}$
$n = 100$, $\mathbf{x}_1 = (1/2, \ldots, 1/2)^\top$, $\xi_1 = \delta_{\mathbf{x}_1}$

$\mathbf{X}_n^S$, $\mathbf{X}_n^{KH}$, $\mathbf{X}_n^{KH-\log}$, $\mathbf{X}_n^{MN}$



$\mathbf{X}_n^S$, $\mathbf{X}_n^{KH}$, $\mathbf{X}_n^{KH-\log}$, $\mathbf{X}_n^{MN}$



$\mathbf{X}_n^{KH}$: K = tensor product of Matérn 3/2

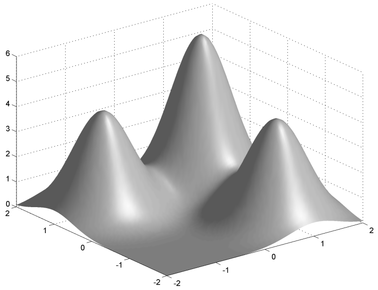
$$K_{3/2,\theta}(x,x') = (1 + \sqrt{3}\theta|x - x'|)\exp(-\sqrt{3}\theta|x - x'|), \ \theta = n^{1/d}$$

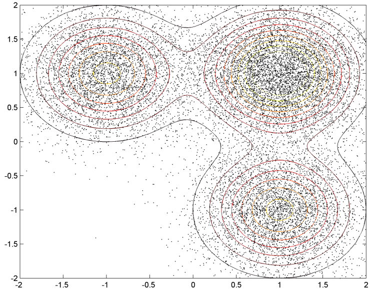
$\mathbf{X}_n^{KH-\log}$: $K(\mathbf{x},\mathbf{x}') = \prod_{i=1}^d \log(1/|x_i - x_i'|)$

MMD minimisation is not restricted to μ being uniform:

Example: Gaussian mixture $\mu = \sum_{j=1}^{3} \beta_j \, \mu_{\mathcal{N}}(\mathbf{a}_j, \sigma_j)$, $C = 2^{14} = 16\,384$

(for $K_\theta(\mathbf{x}, \mathbf{x}') = \exp -(\theta \, \|\mathbf{x} - \mathbf{x}'\|^2)$, we know $P_\mu(\cdot)$ and $\mathscr{E}_K(\mu)$)

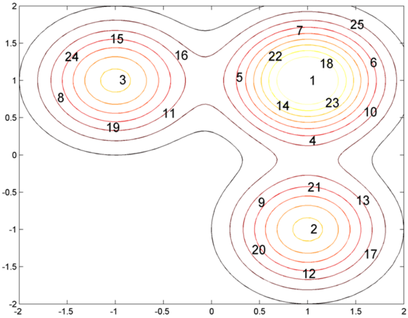




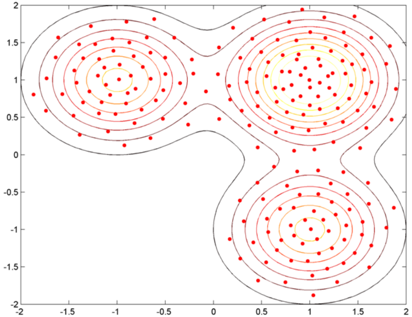
MMD minimisation is not restricted to μ being uniform:

Example: Gaussian mixture $\mu = \sum_{j=1}^{3} \beta_j \, \mu_{\mathcal{N}}(\mathbf{a}_j, \sigma_j)$, $C = 2^{14} = 16\,384$

(for $K_\theta(\mathbf{x}, \mathbf{x}') = \exp -(\theta \, \|\mathbf{x} - \mathbf{x}'\|^2)$, we know $P_\mu(\cdot)$ and $\mathscr{E}_K(\mu)$)



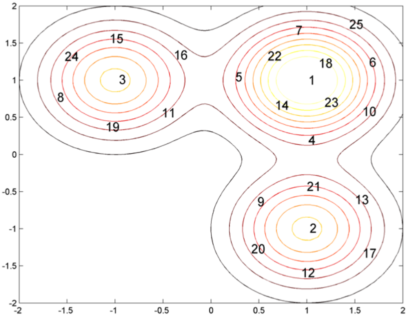
$n = 25$



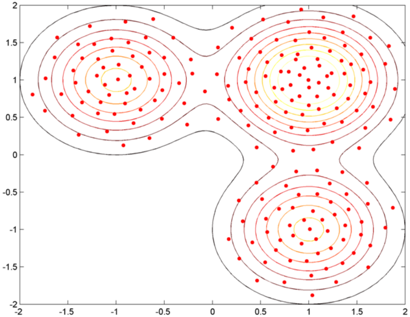
$n = 200$

MMD minimisation is not restricted to μ being uniform:

Example: Gaussian mixture $\mu = \sum_{j=1}^{3} \beta_j \, \mu_{\mathcal{N}}(\mathbf{a}_j, \sigma_j)$, $C = 2^{14} = 16\,384$
(for $K_\theta(\mathbf{x}, \mathbf{x}') = \exp -(\theta \, \|\mathbf{x} - \mathbf{x}'\|^2)$, we know $P_\mu(\cdot)$ and $\mathscr{E}_K(\mu)$)



$n = 25$



$n = 200$

- Comparison between kernel herding, greedy MMD minimisation and Sequential Bayesian Quadrature (P., 2021)
- Extension to Stein discrepancy (Teymur et al., 2021) (K'_μ such that $P_\mu(\cdot) \equiv 0$ and $\mathscr{E}_{K'_\mu}(\mu) = 0$ without knowing the normalising constant in μ)
- Singular kernels (via completely monotone functions) (P. & Zhigljavsky, 2021)

4 Maximisation of an integrated covering measure

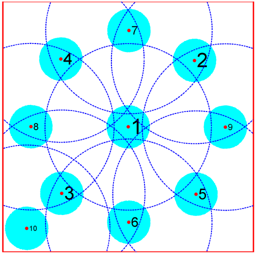
μ uniform on $\mathscr{X}$

$$F_{\mathbf{X}_n}(r) \triangleq \mu\{\mathbf{x} \in \mathscr{X} : d(\mathbf{x}, \mathbf{X}_n) \leq r\}$$
$$= \textit{distance c.d.f.}$$

$\mathbf{X}_n,\ n = 10$

$r = 0.25 \times \mathrm{CR}(\mathbf{X}_n)$

$\rightarrow F_{\mathbf{X}_n}(r) \simeq 0.22$



4 Maximisation of an integrated covering measure

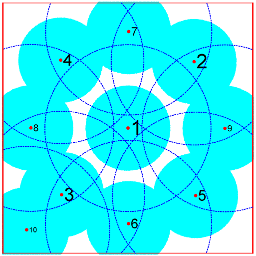
μ uniform on $\mathscr{X}$

$$F_{\mathbf{X}_n}(r) \triangleq \mu\{\mathbf{x} \in \mathscr{X} : d(\mathbf{x}, \mathbf{X}_n) \leq r\}$$
$$= \textit{distance c.d.f.}$$

$\mathbf{X}_n,\ n = 10$

$r = 0.5 \times \mathrm{CR}(\mathbf{X}_n)$

$\rightarrow F_{\mathbf{X}_n}(r) \simeq 0.75$



4 Maximisation of an integrated covering measure

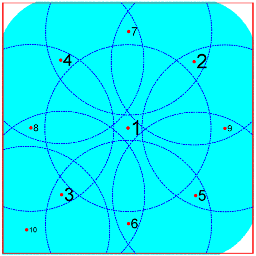
μ uniform on $\mathscr{X}$

$$F_{\mathbf{X}_n}(r) \triangleq \mu\{\mathbf{x} \in \mathscr{X} : d(\mathbf{x}, \mathbf{X}_n) \leq r\}$$
$$= \textit{distance c.d.f.}$$

$\mathbf{X}_n,\ n = 10$

$r = 0.75 \times \mathrm{CR}(\mathbf{X}_n)$

$\rightarrow F_{\mathbf{X}_n}(r) \simeq 0.98$



4 Maximisation of an integrated covering measure

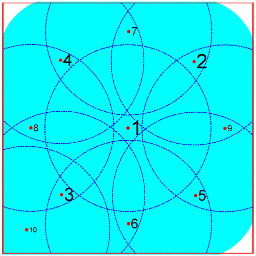
μ uniform on $\mathscr{X}$

$$F_{\mathbf{X}_n}(r) \triangleq \mu\{\mathbf{x} \in \mathscr{X} : d(\mathbf{x}, \mathbf{X}_n) \leq r\}$$
$$= \textit{distance c.d.f.}$$

$$\Phi_r(\mathbf{X}_n) \triangleq F_{\mathbf{X}_n}(r)$$
$= $ *covering measure* of $\mathbf{X}_n$

$$Q_\alpha(\mathbf{X}_n) \triangleq \inf\{t : F_{\mathbf{X}_n}(t) \geq \alpha\}$$
$= \alpha$-quantile of $F_{\mathbf{X}_n}(\cdot)$
(with $Q_1(\mathbf{X}_n) = \mathrm{CR}(\mathbf{X}_n)$)

$\mathbf{X}_n,\ n = 10$
$r = 0.75 \times \mathrm{CR}(\mathbf{X}_n)$
$\rightarrow F_{\mathbf{X}_n}(r) \simeq 0.98$



4 Maximisation of an integrated covering measure

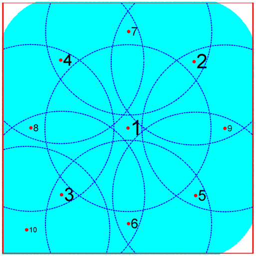
μ uniform on $\mathscr{X}$

$F_{\mathbf{X}_n}(r) \triangleq \mu\{\mathbf{x} \in \mathscr{X} : d(\mathbf{x}, \mathbf{X}_n) \leq r\}$
$= \textit{distance c.d.f.}$

$\Phi_r(\mathbf{X}_n) \triangleq F_{\mathbf{X}_n}(r)$
$=$ *covering measure* of $\mathbf{X}_n$

$Q_\alpha(\mathbf{X}_n) \triangleq \inf\{t : F_{\mathbf{X}_n}(t) \geq \alpha\}$
$= \alpha$-quantile of $F_{\mathbf{X}_n}(\cdot)$
(with $Q_1(\mathbf{X}_n) = \mathrm{CR}(\mathbf{X}_n)$)

$\mathbf{X}_n,\ n = 10$
$r = 0.75 \times \mathrm{CR}(\mathbf{X}_n)$
$\rightarrow F_{\mathbf{X}_n}(r) \simeq 0.98$



- $\Phi_r(\cdot)$ is non-decreasing
- it satisfies $\Phi_r(\emptyset) = 0$
- $\forall \mathbf{x} \in \mathscr{X}$, $\Phi_r(\mathbf{X}_n \cup \{\mathbf{x}\}) - \Phi_r(\mathbf{X}_n)$ is non-increasing with respect to $\mathbf{X}_n$
  $\Rightarrow \Phi_r$ is submodular

For $B > 0$, $q > -1$ and $\mathbf{X}_n \neq \emptyset$, define

$$I_{B,q}(\mathbf{X}_n) \triangleq \boxed{\int_0^B r^q \, F_{\mathbf{X}_n}(r) \, \mathrm{d}r} = \frac{1}{q+1} \left\{ B^{q+1} F_{\mathbf{X}_n}(B) - \int_0^B r^{q+1} \, f_{\mathbf{X}_n}(r) \, \mathrm{d}r \right\}$$
$$= \textit{integrated covering measure}$$

and set $I_{B,q}(\emptyset) = 0$

For $B > 0$, $q > -1$ and $\mathbf{X}_n \neq \emptyset$, define

$$I_{B,q}(\mathbf{X}_n) \triangleq \boxed{\int_0^B r^q \, F_{\mathbf{X}_n}(r)\, \mathrm{d}r} = \frac{1}{q+1}\left\{B^{q+1} F_{\mathbf{X}_n}(B) - \int_0^B r^{q+1}\, f_{\mathbf{X}_n}(r)\, \mathrm{d}r\right\}$$
$$= \textit{integrated covering measure}$$

and set $I_{B,q}(\emptyset) = 0$

The set function $I_{B,q} : \mathbf{X}_n \to I_{B,q}(\mathbf{X}_n)$ is non-decreasing and submodular

→ suitable for greedy maximisation

For $B > 0$, $q > -1$ and $\mathbf{X}_n \neq \emptyset$, define

$$I_{B,q}(\mathbf{X}_n) \triangleq \boxed{\int_0^B r^q\, F_{\mathbf{X}_n}(r)\,\mathrm{d}r} = \frac{1}{q+1}\left\{B^{q+1} F_{\mathbf{X}_n}(B) - \int_0^B r^{q+1}\, f_{\mathbf{X}_n}(r)\,\mathrm{d}r\right\}$$
$$= \textit{integrated covering measure}$$

and set $I_{B,q}(\emptyset) = 0$

The set function $I_{B,q} : \mathbf{X}_n \to I_{B,q}(\mathbf{X}_n)$ is non-decreasing and submodular
→ suitable for greedy maximisation

- $B \geq \mathrm{CR}(\mathbf{X}_n) \Rightarrow F_{\mathbf{X}_n}(B) = 1$

For $B > 0$, $q > -1$ and $\mathbf{X}_n \neq \emptyset$, define

$$I_{B,q}(\mathbf{X}_n) \triangleq \boxed{\int_0^B r^q\, F_{\mathbf{X}_n}(r)\, \mathrm{d}r} = \frac{1}{q+1}\left\{B^{q+1} F_{\mathbf{X}_n}(B) - \int_0^B r^{q+1}\, f_{\mathbf{X}_n}(r)\, \mathrm{d}r\right\}$$

= *integrated covering measure*

and set $I_{B,q}(\emptyset) = 0$

The set function $I_{B,q} : \mathbf{X}_n \to I_{B,q}(\mathbf{X}_n)$ is non-decreasing and submodular

→ suitable for greedy maximisation

- $B \geq \mathrm{CR}(\mathbf{X}_n) \Rightarrow F_{\mathbf{X}_n}(B) = 1$
- $B \geq \mathrm{diam}(\mathscr{X}) \Rightarrow$
  maximising $I_{B,q}(\mathbf{X}_n) \Leftrightarrow$ minimising $\int_0^B r^{q+1}\, f_{\mathbf{X}_n}(r)\, \mathrm{d}r = \mathsf{E}_n\{R^{q+1}\}$, $R \sim f_{\mathbf{X}_n}$

For $B > 0$, $q > -1$ and $\mathbf{X}_n \neq \emptyset$, define

$$I_{B,q}(\mathbf{X}_n) \triangleq \boxed{\int_0^B r^q\, F_{\mathbf{X}_n}(r)\,\mathrm{d}r} = \frac{1}{q+1}\left\{B^{q+1} F_{\mathbf{X}_n}(B) - \int_0^B r^{q+1}\, f_{\mathbf{X}_n}(r)\,\mathrm{d}r\right\}$$
$$= \textit{integrated covering measure}$$

and set $I_{B,q}(\emptyset) = 0$

The set function $I_{B,q} : \mathbf{X}_n \to I_{B,q}(\mathbf{X}_n)$ is non-decreasing and submodular
➞ suitable for greedy maximisation

- $B \geq \mathrm{CR}(\mathbf{X}_n) \Rightarrow F_{\mathbf{X}_n}(B) = 1$
- $B \geq \mathrm{diam}(\mathscr{X}) \Rightarrow$
  maximising $I_{B,q}(\mathbf{X}_n) \Leftrightarrow$ minimising $\int_0^B r^{q+1}\, f_{\mathbf{X}_n}(r)\,\mathrm{d}r = \mathsf{E}_n\{R^{q+1}\}$, $R \sim f_{\mathbf{X}_n}$
- $(\mathsf{E}_n\{R^{q+1}\})^{1/(q+1)} = E_{q+1}(\mathbf{X}_n) = L^{q+1}$-mean quantization error for $\mathbf{X}_n$ (Graf and Luschgy, 2000), with $E_{q+1}(\mathbf{X}_n) \nearrow \mathrm{CR}(\mathbf{X}_n)$ as $q \to \infty$

For $B > 0$, $q > -1$ and $\mathbf{X}_n \neq \emptyset$, define

$$I_{B,q}(\mathbf{X}_n) \triangleq \boxed{\int_0^B r^q\, F_{\mathbf{X}_n}(r)\,\mathrm{d}r} = \frac{1}{q+1}\left\{B^{q+1} F_{\mathbf{X}_n}(B) - \int_0^B r^{q+1}\, f_{\mathbf{X}_n}(r)\,\mathrm{d}r\right\}$$

= *integrated covering measure*

and set $I_{B,q}(\emptyset) = 0$

The set function $I_{B,q} : \mathbf{X}_n \to I_{B,q}(\mathbf{X}_n)$ is non-decreasing and submodular

➞ suitable for greedy maximisation

- $B \geq \mathrm{CR}(\mathbf{X}_n) \Rightarrow F_{\mathbf{X}_n}(B) = 1$
- $B \geq \mathrm{diam}(\mathscr{X}) \Rightarrow$
  maximising $I_{B,q}(\mathbf{X}_n) \Leftrightarrow$ minimising $\int_0^B r^{q+1}\, f_{\mathbf{X}_n}(r)\,\mathrm{d}r = \mathsf{E}_n\{R^{q+1}\}$, $R \sim f_{\mathbf{X}_n}$
- $(\mathsf{E}_n\{R^{q+1}\})^{1/(q+1)} = E_{q+1}(\mathbf{X}_n) = L^{q+1}$-mean quantization error for $\mathbf{X}_n$ (Graf and Luschgy, 2000), with $E_{q+1}(\mathbf{X}_n) \nearrow \mathrm{CR}(\mathbf{X}_n)$ as $q \to \infty$

$\boxed{\to \mathbf{X}_n^{ICM}}$ details in (Nogales Gómez et al., 2021)

Replace μ by a Q-point discrete approximation μ_Q

➞ 2 discrete sets $\mathscr{X}_C$ and $\mathscr{X}_Q$ (compute $C \times Q$ pairwise distances)

➞ complexity $= \mathcal{O}(nCQ)$ ($= \mathcal{O}(\gamma_n nCQ)$ for lazy-greedy version)

5 Greedy packing

- Take any $\mathbf{x}_1 \in \mathscr{X}$ (e.g., at the center)
- For $k = 1, \ldots, n-1$, take $\mathbf{x}_{k+1}$ as far as possible from $\mathbf{X}_k$

(= coffee-house design of (Müller, 2001, 2007))

5 Greedy packing

- Take any $\mathbf{x}_1 \in \mathscr{X}$ (e.g., at the center)
- For $k = 1, \ldots, n-1$, take $\mathbf{x}_{k+1}$ as far as possible from $\mathbf{X}_k$

(= coffee-house design of (Müller, 2001, 2007))
Th. (Gonzalez, 1985):

$$
\begin{aligned}
\mathrm{CR}(\mathbf{X}_k) &\leq 2\, \mathrm{CR}_k^*\,, \ \forall k \geq 1\,, \\
\mathrm{PR}(\mathbf{X}_k) &\geq \frac{1}{2}\, \mathrm{PR}_k^*\,, \ \forall k \geq 2\,, \\
\mathrm{MR}(\mathbf{X}_k) &\leq 2\,, \qquad \forall k \geq 2\,.
\end{aligned}
$$

5 Greedy packing

- Take any $\mathbf{x}_1 \in \mathscr{X}$ (e.g., at the center)
- For $k = 1, \ldots, n-1$, take $\mathbf{x}_{k+1}$ as far as possible from $\mathbf{X}_k$

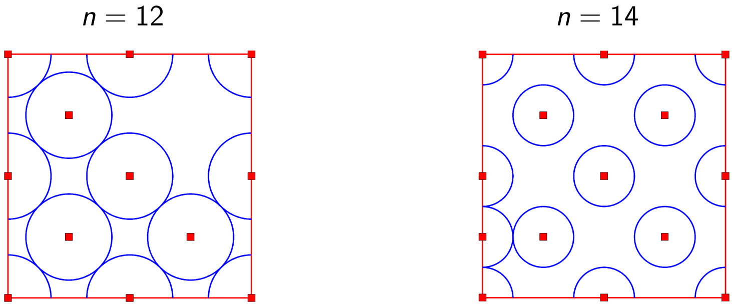
(= coffee-house design of (Müller, 2001, 2007))
Th. (Gonzalez, 1985):

$$\begin{aligned} \mathrm{CR}(\mathbf{X}_k) &\leq 2\,\mathrm{CR}_k^*\,, \ \forall k \geq 1\,, \\ \mathrm{PR}(\mathbf{X}_k) &\geq \frac{1}{2}\,\mathrm{PR}_k^*\,, \ \forall k \geq 2\,, \\ \mathrm{MR}(\mathbf{X}_k) &\leq 2\,, \qquad \forall k \geq 2\,. \end{aligned}$$

Greedy packing is asymptotically optimal for MR:
$\limsup_{n\to\infty} \mathrm{MR}(\mathbf{X}_n) \geq 2$ for any sequence of nested designs $\mathbf{X}_n$ in $\mathscr{X}$ bounded (P. & Zhigljavsky, 2022)

Easy to implement: use a finite candidate set $\mathscr{X}_C \subset \mathscr{X}$
$\rightarrow$ complexity $= \mathcal{O}(nC)$

How does it perform?
Exact behaviour known in some cases ($\mathscr{X} = [0,1]^d$, $d = 2, 4$, maybe 8?)



$\mathrm{PR}(\mathbf{X}_n)$

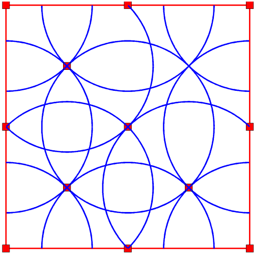
Easy to implement: use a finite candidate set $\mathscr{X}_C \subset \mathscr{X}$

→ complexity $= \mathcal{O}(nC)$

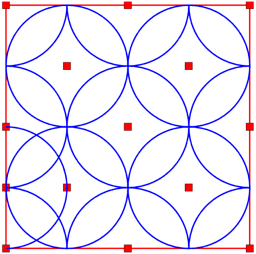
How does it perform?

Exact behaviour known in some cases ($\mathscr{X} = [0,1]^d$, $d = 2, 4$, maybe 8?)

$n = 12$

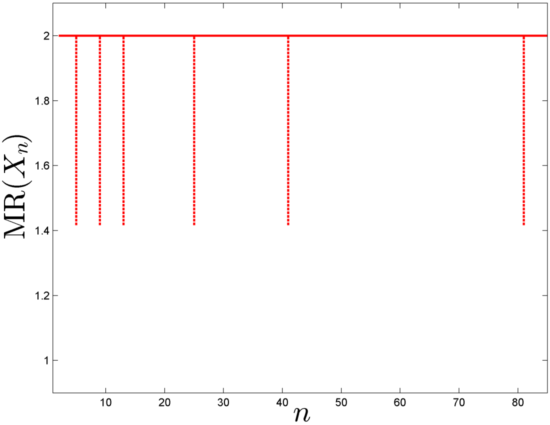


$n = 14$



$\mathrm{CR}(\mathbf{X}_n)$

...but there is no competitor in terms of $\mathrm{MR}(\mathbf{X}_n)$!



$\mathscr{X} = [0,1]^2$, $\mathrm{MR}(\mathbf{X}_n)$, $n = 2, \ldots, 85$

To reduce $\mathrm{CR}(\mathbf{X}_n)$: force points to stay away from the boundary $\partial\mathscr{X}$:

$\rightarrow$ take $\mathbf{x}_{k+1} \in \mathrm{Arg}\max_{\mathbf{x}\in\mathscr{X}} D_\beta(\mathbf{x}, \mathbf{X}_k, \mathscr{X})$ with

where $D_\beta(\mathbf{x}, \mathbf{X}_k, \mathscr{X}) = \min\left\{\min_{\mathbf{x}_i\in\mathbf{X}_k} \|\mathbf{x} - \mathbf{x}_i\|,\ \beta\, d(\mathbf{x}, \partial\mathscr{X})\right\}$, $\beta > 0$

To reduce $\mathrm{CR}(\mathbf{X}_n)$: force points to stay away from the boundary $\partial\mathscr{X}$:

$\rightarrow$ take $\mathbf{x}_{k+1} \in \mathrm{Arg}\max_{\mathbf{x}\in\mathscr{X}} D_\beta(\mathbf{x}, \mathbf{X}_k, \mathscr{X})$ with

where $D_\beta(\mathbf{x}, \mathbf{X}_k, \mathscr{X}) = \min\{\min_{\mathbf{x}_i\in\mathbf{X}_k} \|\mathbf{x}-\mathbf{x}_i\|, \beta\, d(\mathbf{x}, \partial\mathscr{X})\}$, $\beta > 0$

- easy to implement when $\mathscr{X} = [0,1]^d$ ($\rightarrow$ complexity $= \mathcal{O}(nC)$)
- $\beta = \infty$ ➠ greedy packing
- $\beta = 2$ ➠ traditional packing: n non-intersecting balls fully in $\mathscr{X}$
- (Shang and Apley, 2021) $\rightarrow \beta = 2\sqrt{2d}$
  (Nogales Gómez et al., 2021) $\rightarrow \beta = \beta(n,d) = \frac{d}{2\,(n_{\max} V_d)^{-1/d}} - \sqrt{d}$,
  with $V_d = \mathrm{vol}\,(\mathscr{B}(\mathbf{0},1))$

Performance of *boundary-phobic greedy packing*:

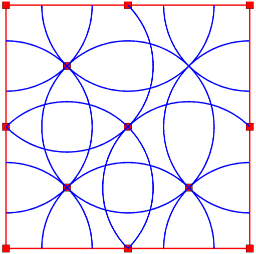
$$\begin{aligned}
\mathrm{CR}(\mathbf{X}_k) &\leq \frac{2}{a}\,\mathrm{CR}_k^*\,,\ \forall k \geq 1\,,\\
\mathrm{PR}(\mathbf{X}_k) &\geq \frac{a}{2}\,\mathrm{PR}_k^*\,,\ \forall k \geq 2\,,\\
\mathrm{MR}(\mathbf{X}_k) &\leq \frac{2}{a}\,, \qquad \forall k \geq 2\,,
\end{aligned}$$

with $a = 1/(1+\sqrt{d}/\beta)$ (P. & Zhigljavsky, 2022)

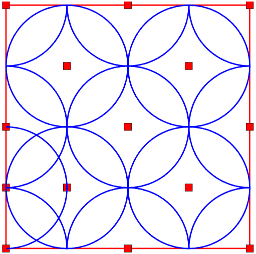
$\mathscr{X} = [0,1]^2$, $\mathrm{CR}(\mathbf{X}_n)$

$\beta = \infty$ → greedy packing

$n = 12$



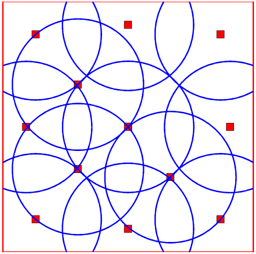
$n = 14$



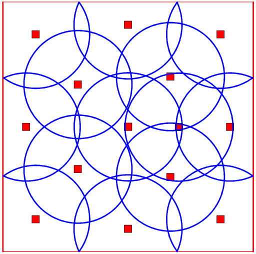
$\mathscr{X} = [0,1]^2$, $\mathrm{CR}(\mathbf{X}_n)$

$\beta = 4$ → boundary-phobic greedy packing

$n = 12$



$n = 14$

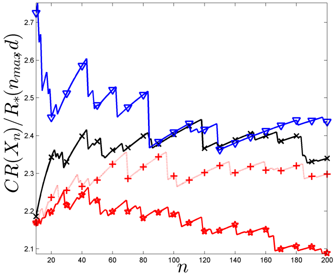


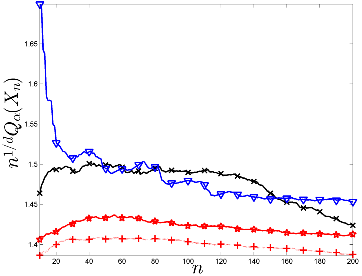
Example: $d = 10$, $n = 200$

$\mathscr{X}_C = 2^{13}$ Sobol' points in $\mathscr{X} = [0,1]^{10}$

$\mathscr{X}_{Q'} = 2^{14}$ Sobol' points, $\mathscr{X}_Q = \mathscr{X}_{Q'} \cup 2^{10}$ vertices

Comparison of $\mathbf{X}_n^{ICM}$ (★ for $\mathscr{X}_Q$, + for $\mathscr{X}_{Q'}$, $q = 10$) with Halton ($\triangledown$) and Sobol' ($\times$)





$R_\star(n,d) = (nV_d)^{-1/d} \leq \mathrm{CR}_n^*$, $\alpha = 0.99$ in $Q_\alpha(\mathbf{X}_n)$

evaluation of $\mathrm{CR}(\mathbf{X}_n)$ and $Q_\alpha(\mathbf{X}_n)$ on 2^{18} Sobol' points + 2^{10} vertices

Example: $d = 10$, $n = 200$

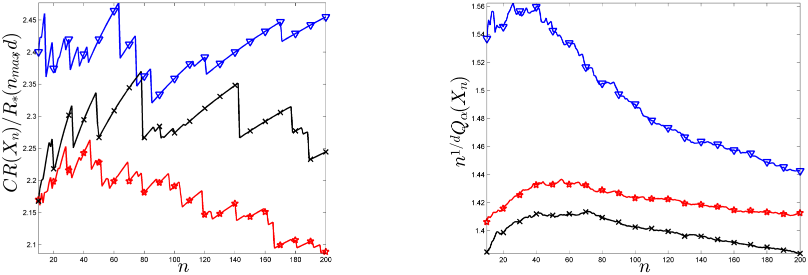
$\mathscr{X}_C = 2^{13}$ Sobol' points in $\mathscr{X} = [0,1]^{10}$

$\mathscr{X}_{Q'} = 2^{14}$ Sobol' points, $\mathscr{X}_Q = \mathscr{X}_{Q'} \cup 2^{10}$ vertices

Comparison of $\mathbf{X}_n^{ICM}$ ($\star$, $q = 10$) with minimisation of (ℓ_q, L_q) relaxed $\mathrm{CR}(\mathbf{X}_n)$:

Greedy Minimisation $\mathbf{X}_n^{GM}$ ($\times$) and Vertex Direction $\mathbf{X}_n^{VD}$ ($\triangledown$), $q = 10$

($\approx$ 7 and 2 times slower)



$R_\star(n,d) = (nV_d)^{-1/d} \leq \mathrm{CR}_n^*$, $\alpha = 0.99$ in $Q_\alpha(\mathbf{X}_n)$

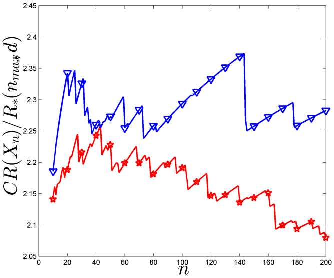
evaluation of $\mathrm{CR}(\mathbf{X}_n)$ and $Q_\alpha(\mathbf{X}_n)$ on 2^{18} Sobol' points + 2^{10} vertices

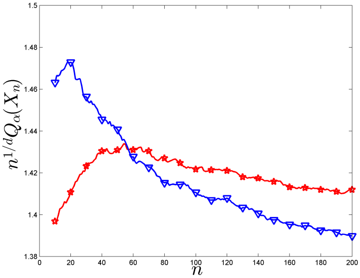
Example: $d = 10$, $n = 200$

$\mathscr{X}_C = 2^{13}$ Sobol' points in $\mathscr{X} = [0,1]^{10}$

$\mathscr{X}_{Q'} = 2^{14}$ Sobol' points, $\mathscr{X}_Q = \mathscr{X}_{Q'} \cup 2^{10}$ vertices

Comparison of $\mathbf{X}_n^{ICM}$ (★, $q = 10$) with Kernel Herding $\mathbf{X}_n^{KH}$ (▽):
(≈ 2 times faster)





$R_\star(n,d) = (nV_d)^{-1/d} \leq \mathrm{CR}_n^*$, $\alpha = 0.99$ in $Q_\alpha(\mathbf{X}_n)$

evaluation of $\mathrm{CR}(\mathbf{X}_n)$ and $Q_\alpha(\mathbf{X}_n)$ on 2^{18} Sobol' points + 2^{10} vertices

Example: $d = 10$, $n = 200$

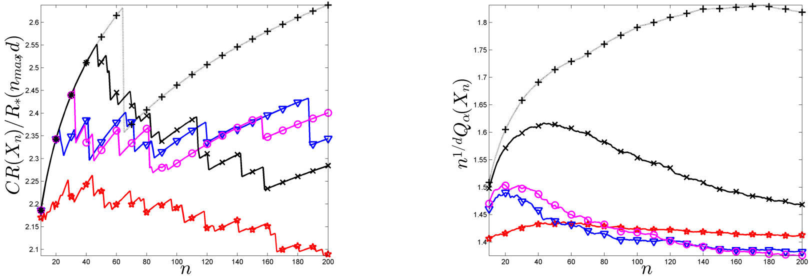
$\mathscr{X}_C = 2^{13}$ Sobol' points in $\mathscr{X} = [0,1]^{10}$

$\mathscr{X}_{Q'} = 2^{14}$ Sobol' points, $\mathscr{X}_Q = \mathscr{X}_{Q'} \cup 2^{10}$ vertices

Comparison of $\mathbf{X}_n^{ICM}$ (★, $q = 10$) with greedy packing ($\approx$ 20 times faster):

$\beta = \infty$ and $\mathscr{X}_C$ ($\times$), $\beta = \infty$ and $\mathscr{X}_C \cup 2^{10}$ vertices ($+$),

$\beta = 2\sqrt{2d}$ ($\circ$), $\beta = \beta(n, d)$ ($\triangledown$)



$R_\star(n, d) = (nV_d)^{-1/d} \leq \mathrm{CR}_n^*$, $\alpha = 0.99$ in $Q_\alpha(\mathbf{X}_n)$

evaluation of $\mathrm{CR}(\mathbf{X}_n)$ and $Q_\alpha(\mathbf{X}_n)$ on 2^{18} Sobol' points $+ 2^{10}$ vertices

6 Conclusions

- Several space-filling criteria:
  CR($\mathbf{X}_n$) is important, but $Q_\alpha(\mathbf{X}_n)$ may be more relevant:
  → it may provide a smaller error $\|f - \eta_n^*\|_{L_q}$, $q < \infty$

6 Conclusions

- Several space-filling criteria:
  CR($\mathbf{X}_n$) is important, but $Q_\alpha(\mathbf{X}_n)$ may be more relevant:
  → it may provide a smaller error $\|f - \eta_n^*\|_{L_q}$, $q < \infty$
- Many methods (some based on heuristics):
  - those using two finite sets $\mathscr{X}_C$ and $\mathscr{X}_Q$ cannot have C, Q very large
    → the choices of the two sets are important
  - those using $\mathscr{X}_C$ only (MMD, greedy packing) are linear in C and n
    → fast and usable for design with large size n and dimension d
    → valuable alternatives to low-discrepancy sequences (Sobol')
  - Minimising the integrated covering measure gives the smallest CR($\mathbf{X}_n$)

6 Conclusions

- Several space-filling criteria:
  $\mathrm{CR}(\mathbf{X}_n)$ is important, but $Q_\alpha(\mathbf{X}_n)$ may be more relevant:
  → it may provide a smaller error $\|f - \eta_n^*\|_{L_q}$, $q < \infty$
- Many methods (some based on heuristics):
  - those using two finite sets $\mathscr{X}_C$ and $\mathscr{X}_Q$ cannot have C, Q very large
    → the choices of the two sets are important
  - those using $\mathscr{X}_C$ only (MMD, greedy packing) are linear in C and n
    → fast and usable for design with large size n and dimension d
    → valuable alternatives to low-discrepancy sequences (Sobol')
  - Minimising the integrated covering measure gives the smallest $\mathrm{CR}(\mathbf{X}_n)$
- No clear winner between greedy minimisation and gradient-type descent when both are possible
- Batch design: optimize one of the criteria considered for a fixed n, as Mak and Joseph (2017, 2018) do with MMD for the energy-distance kernel

6 Conclusions

- Several space-filling criteria:
  $\mathrm{CR}(\mathbf{X}_n)$ is important, but $Q_\alpha(\mathbf{X}_n)$ may be more relevant:
  ➞ it may provide a smaller error $\|f - \eta_n^*\|_{L_q}$, $q < \infty$
- Many methods (some based on heuristics):
  - those using two finite sets $\mathscr{X}_C$ and $\mathscr{X}_Q$ cannot have C, Q very large
    ➞ the choices of the two sets are important
  - those using $\mathscr{X}_C$ only (MMD, greedy packing) are linear in C and n
    ➞ fast and usable for design with large size n and dimension d
    ➞ valuable alternatives to low-discrepancy sequences (Sobol')
  - Minimising the integrated covering measure gives the smallest $\mathrm{CR}(\mathbf{X}_n)$
- No clear winner between greedy minimisation and gradient-type descent when both are possible
- Batch design: optimize one of the criteria considered for a fixed n, as Mak and Joseph (2017, 2018) do with MMD for the energy-distance kernel

Thank you for your attention !

References I

Bach, F., Lacoste-Julien, S., Obozinski, G., 2012. On the equivalence between herding and conditional gradient algorithms. In: Proc. 29th Annual International Conference on Machine Learning. pp. 1355–1362.

Frank, M., Wolfe, P., 1956. An algorithm for quadratic programming. Naval Res. Logist. Quart. 3, 95–110.

Gonzalez, T., 1985. Clustering to minimize the maximum intercluster distance. Theoretical Computer Science 38, 293–306.

Graf, S., Luschgy, H., 2000. Foundations of Quantization for Probability Distributions. Springer, Berlin.

Hickernell, F., 1998. A generalized discrepancy and quadrature error bound. Mathematics of Computation 67 (221), 299–322.

Mak, S., Joseph, V., 2017. Projected support points, with application to optimal MCMC reduction. arXiv preprint arXiv:1708.06897.

Mak, S., Joseph, V., 2018. Support points. Annals of Statistics 46 (6A), 2562–2592.

Minoux, M., 1977. Accelerated greedy algorithms for maximizing submodular set functions. In: Proc. 8th IFIP Conference, Wurzburg (Part 2). Springer, New-York, pp. 234–243.

Müller, W., 2001. Coffee-house designs. In: Atkinson, A., Bogacka, B., Zhigljavsky, A. (Eds.), Optimum Design 2000. Kluwer, Dordrecht, Ch. 21, pp. 241–248.

References II

Müller, W., 2007. Collecting Spatial Data. Springer, Berlin, [3rd ed.].

Narcowich, F., Ward, J., Wendland, H., 2005. Sobolev bounds on functions with scattered zeros, with applications to radial basis function surface fitting. Mathematics of Computation 74 (250), 743–763.

Nemhauser, G., Wolsey, L., Fisher, M., 1978. An analysis of approximations for maximizing submodular set functions–I. Mathematical Programming 14 (1), 265–294.

Nogales Gómez, A., Pronzato, L., Rendas, M.-J., 2021. Incremental space-filling design based on coverings and spacings: improving upon low discrepancy sequences. Journal of Statistical Theory and Practice (to appear, HAL preprint hal-02987983).

Pronzato, L., 2021. Performance analysis of greedy algorithms for minimising a maximum mean discrepancy. hal-03114891, arXiv:2101.07564.

Pronzato, L., Zhigljavsky, A., 2019. Measures minimizing regularized dispersion. Journal of Scientific Computing 78 (3), 1550–1570.

Pronzato, L., Zhigljavsky, A., 2020. Bayesian quadrature, energy minimization and space-filling design. SIAM/ASA J. Uncertainty Quantification 8 (3), 959–1011.

Pronzato, L., Zhigljavsky, A., 2021. Minimum-energy measures for singular kernels. Journal of Computational and Applied Mathematics 382, (113089, 16 pages) hal-02495643.

References III

Pronzato, L., Zhigljavsky, A., 2022. Quasi-uniform designs with asymptotically optimal and near-optimal uniformity constant. hal-03494864, arXiv:2112.10401.

Schaback, R., Wendland, H., 2006. Kernel techniques: from machine learning to meshless methods. Acta Numerica 15, 543–639.

Sejdinovic, S., Sriperumbudur, B., Gretton, A., Fukumizu, K., 2013. Equivalence of distance-based and RKHS-based statistics in hypothesis testing. The Annals of Statistics 41 (5), 2263–2291.

Shang, B., Apley, D., 2021. Full-sequential space-filling design algorithms for computer experiments. Journal of Quality Technology 53 (2), 173–196.

Sriperumbudur, B., Gretton, A., Fukumizu, K., Schölkopf, B., Lanckriet, G., 2010. Hilbert space embeddings and metrics on probability measures. Journal of Machine Learning Research 11, 1517–1561.

Székely, G., Rizzo, M., 2013. Energy statistics: A class of statistics based on distances. Journal of Statistical Planning and Inference 143 (8), 1249–1272.

Teymur, O., Gorham, J., Riabiz, M., Oates, C., 2021. Optimal quantisation of probability measures using maximum mean discrepancy. In: International Conference on Artificial Intelligence and Statistics. pp. 1027–1035, arXiv preprint arXiv:2010.07064v1.

Wynn, H., 1970. The sequential generation of D-optimum experimental designs. Annals of Math. Stat. 41, 1655–1664.