On the construction of minimax-distance (sub-)optimal designs

Luc Pronzato

Université Côte d'Azur, CNRS, I3S, France

1) Introduction & motivation

Objective:

Approximation/interpolation of a function f: $\mathbf{x} \in \mathscr{X} \subset \mathbb{R}^d \longrightarrow \mathbb{R}$,

(with $\mathscr{X}$ compact: typically, $\mathscr{X} = [0,1]^d$)

- Choose n points $X_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathscr{X}^n$ (the design) where to evaluate f (no repetition)

1) Introduction & motivation

Objective:

Approximation/interpolation of a function f: $\mathbf{x} \in \mathscr{X} \subset \mathbb{R}^d \longrightarrow \mathbb{R}$,
(with $\mathscr{X}$ compact: typically, $\mathscr{X} = [0,1]^d$)

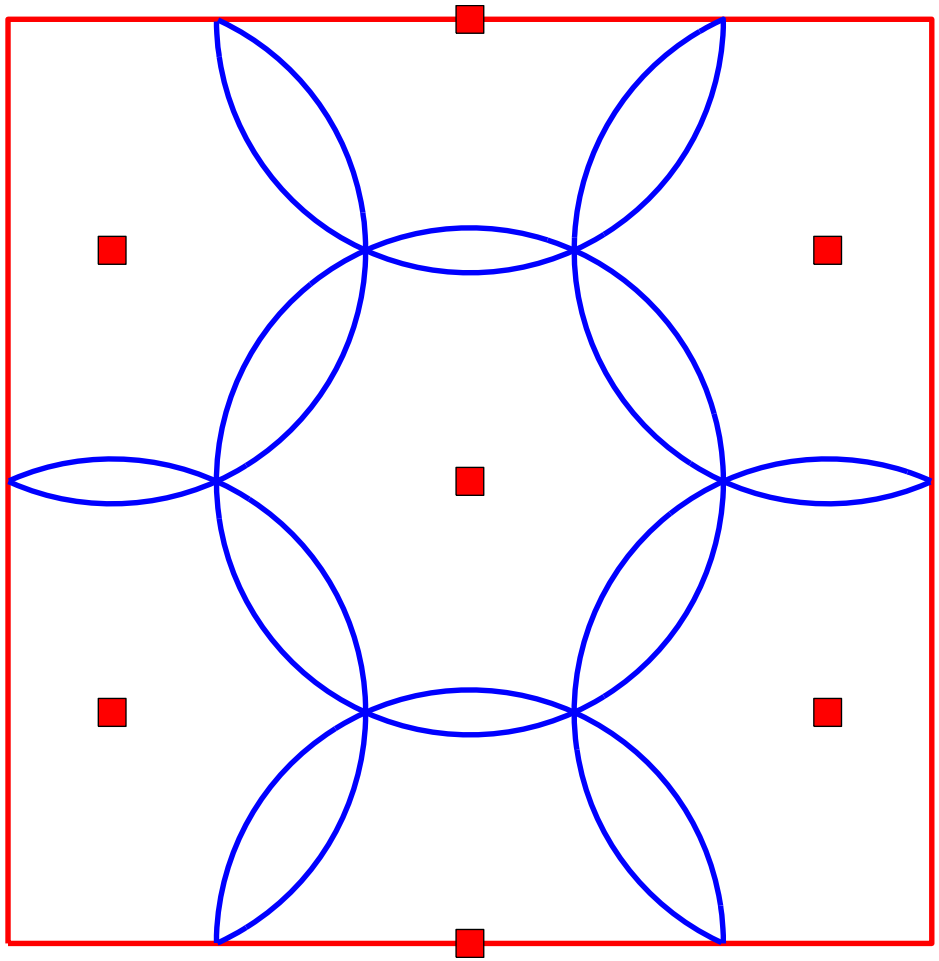
➠ Choose n points $X_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathscr{X}^n$ (the design)
where to evaluate f (no repetition)

Design criterion = minimax distance

➠ minimize $\Phi_{mM}(X_n) = \max_{\mathbf{x}\in\mathscr{X}} \min_{i=1,\ldots,n} \|\mathbf{x} - \mathbf{x}_i\|$ (ℓ_2-distance)
$= \max_{\mathbf{x}\in\mathscr{X}} d(\mathbf{x}, X_n)$
$= d_H(\mathscr{X}, X_n)$ (Hausdorff distance, ℓ_2)
= dispersion of X_n in $\mathscr{X}$ (Niederreiter, 1992, Chap. 6)

X_n^* an optimal n-point design ➔ Φ_{mM}-efficiency $\text{Eff}_{mM}(X_n) = \frac{\Phi^*_{mM,n}}{\Phi_{mM}(X_n)} \in (0,1]$
with $\Phi^*_{mM,n} = \Phi_{mM}(X_n^*)$

$d = 2,\ n = 7$



Why Φ_{mM}? two good reasons (at least) to minimize $\Phi_{mM}(X_n)$:

① Suppose $f \in$ RKHS $\mathcal{H}$ with kernel $K(\mathbf{x}, \mathbf{y}) = C(\|\mathbf{x} - \mathbf{y}\|)$, then

$$\forall \mathbf{x} \in \mathscr{X},\ |f(\mathbf{x}) - \hat{\eta}_n(\mathbf{x})| \leq \|f\|_{\mathcal{H}}\ \rho_n(\mathbf{x})$$ where

- $\hat{\eta}_n(\mathbf{x})$ = BLUP based on the $f(\mathbf{x}_i)$, $i = 1, \ldots, n$
- $\rho_n^2(\mathbf{x})$ = "kriging variance" at $\mathbf{x}$

see, e.g., Vazquez and Bect (2011); Auffray et al. (2012)

Schaback (1995) ➠ $\sup_{\mathbf{x} \in \mathscr{X}} \rho_n(\mathbf{x}) \leq S[\Phi_{mM}(X_n)]$

for some increasing function $S[\cdot]$ (depending on K)

Why Φ_{mM}? **two good reasons (at least) to minimize $\Phi_{mM}(X_n)$:**

① Suppose $f \in$ RKHS $\mathcal{H}$ with kernel $K(\mathbf{x}, \mathbf{y}) = C(\|\mathbf{x} - \mathbf{y}\|)$, then

$$\forall \mathbf{x} \in \mathscr{X},\ |f(\mathbf{x}) - \hat{\eta}_n(\mathbf{x})| \leq \|f\|_{\mathcal{H}}\ \rho_n(\mathbf{x})$$ where

- $\hat{\eta}_n(\mathbf{x})$ = BLUP based on the $f(\mathbf{x}_i)$, $i = 1, \ldots, n$
- $\rho_n^2(\mathbf{x})$ = "kriging variance" at $\mathbf{x}$

see, e.g., Vazquez and Bect (2011); Auffray et al. (2012)

Schaback (1995) ➠ $\sup_{\mathbf{x} \in \mathscr{X}} \rho_n(\mathbf{x}) \leq S[\Phi_{mM}(X_n)]$

for some increasing function $S[\cdot]$ (depending on K)

② X_n^* has no (or few) points on the boundary of $\mathscr{X}$

Evaluation of $\Phi_{mM}(X_n)$? Not considered here!

Evaluation of $\Phi_{mM}(X_n)$? Not considered here!

To evaluate $\Phi_{mM}(X_n) = \max_{\mathbf{x}\in\mathscr{X}} \min_{i=1,\ldots,n} \|\mathbf{x} - \mathbf{x}_i\| = \max_{\mathbf{x}\in\mathscr{X}} d(\mathbf{x}, X_n)$
we need to find $\boxed{\mathbf{x}^* = \arg\max_{\mathbf{x}\in\mathscr{X}} d(\mathbf{x}, X_n)}$

Key idea: replace $\arg\max_{\mathbf{x}\in\mathscr{X}} d(\mathbf{x}, X_n)$ by $\arg\max_{\mathbf{x}\in\mathscr{X}_Q} d(\mathbf{x}, X_n)$
for a suitable finite $\mathscr{X}_Q \in \mathscr{X}^Q$

Replacing $\mathscr{X}_Q$ by a **regular grid**, or first Q points of a **Low Discrepancy Sequence** in $\mathscr{X}$, is not accurate:

➠ $\Phi_{mM}(X_n; \mathscr{X}_Q) \leq \Phi_{mM}(X_n)$ (optimistic result)
$\boxed{\text{requires } Q = \mathcal{O}(1/\epsilon^d) \text{ to have } \Phi_{mM}(X_n) < \Phi_{mM}(X_n; \mathscr{X}_Q) + \epsilon}$

For $d \lesssim 5$, use tools from **algorithmic geometry** (Delaunay triangulation or Voronoï tessellation) ➞ exact result

For larger d, use MCMC with $\mathscr{X}_Q$ = adaptive grid (LP, 2017a)

Bounds on $\Phi^*_{mM,n} = \Phi_{mM}(X^*_n)$ when $\mathscr{X} = [0,1]^d$

Lower bound: the n balls $\mathscr{B}(\mathbf{x}_i, \Phi^*_{mM,n})$ cover $\mathscr{X}$

$\Rightarrow n V_d \, (\Phi^*_{mM,n})^d \geq \text{vol}(\mathscr{X}) \;\; (= 1)$, with $V_d = \text{vol}[\mathscr{B}(0,1)] = \pi^{d/2}/\Gamma(d/2+1)$

$$\underline{R}^*_n = (nV_d)^{-1/d} \leq \Phi^*_{mM,n}$$

Bounds on $\Phi^*_{mM,n} = \Phi_{mM}(X^*_n)$ when $\mathscr{X} = [0,1]^d$

Lower bound: the n balls $\mathscr{B}(\mathbf{x}_i, \Phi^*_{mM,n})$ cover $\mathscr{X}$

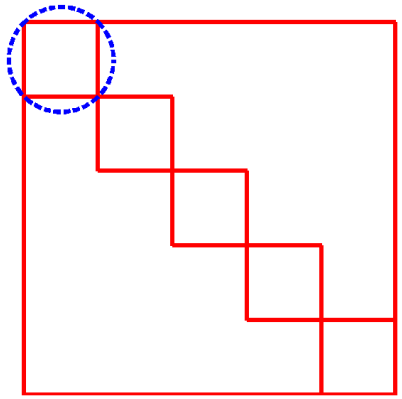
$\Rightarrow n V_d \, (\Phi^*_{mM,n})^d \geq \text{vol}(\mathscr{X}) \;\; (= 1)$, with $V_d = \text{vol}[\mathscr{B}(0,1)] = \pi^{d/2}/\Gamma(d/2+1)$

$$\underline{R}^*_n = (nV_d)^{-1/d} \leq \Phi^*_{mM,n}$$

Upper bound: use any design!

m^d-point regular grid in $\mathscr{X}$:

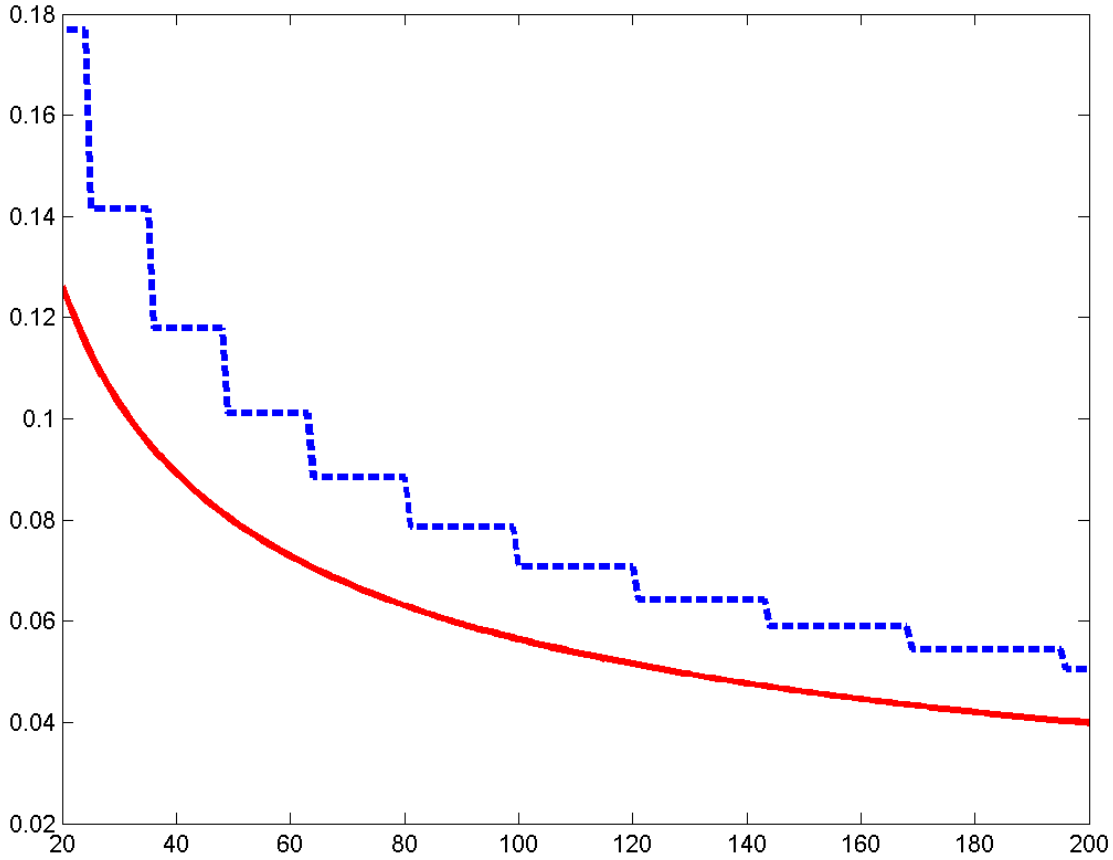
$\Phi^*_{mM,m^d} \leq \frac{\sqrt{d}}{2m}$:



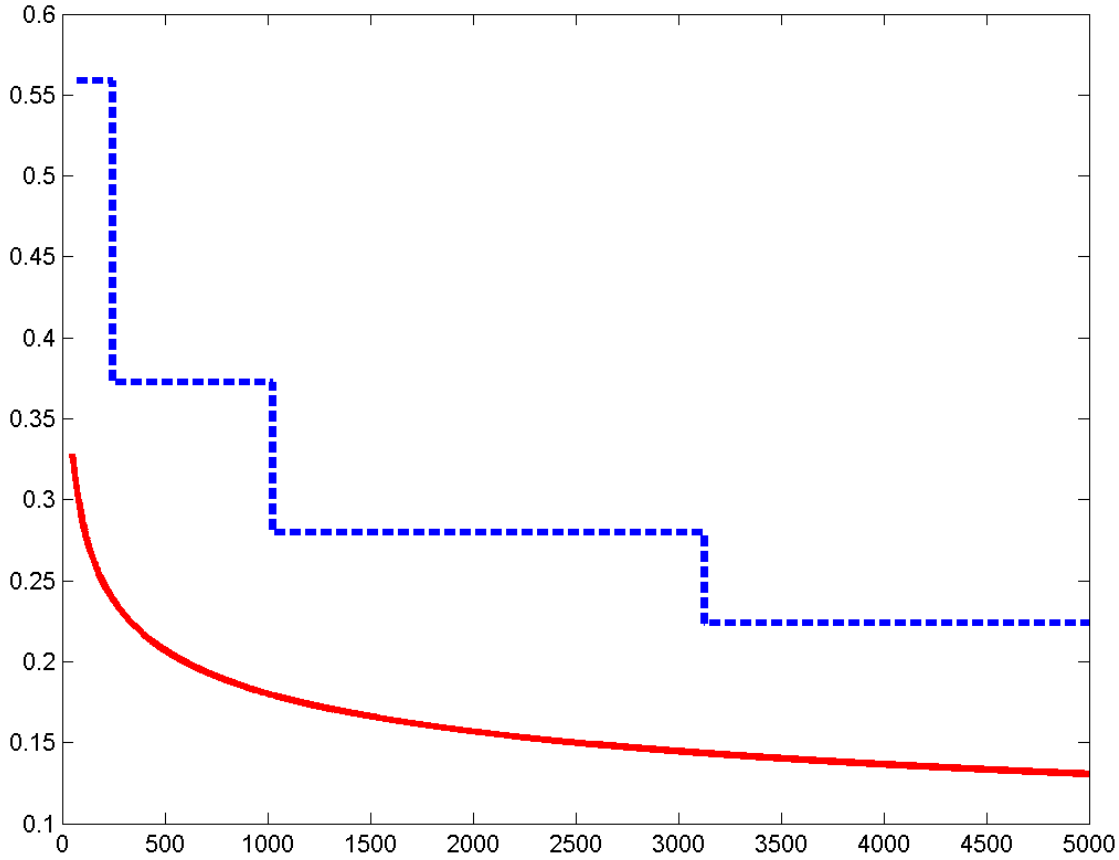
Take $m = \lfloor n^{1/d} \rfloor$, so that $m^d \leq n$ and $\Phi^*_{mM,n} \leq \Phi^*_{mM,m^d}$, therefore

$$\Phi^*_{mM,n} \leq \overline{R}^*_n = \frac{\sqrt{d}}{2\lfloor n^{1/d} \rfloor}$$

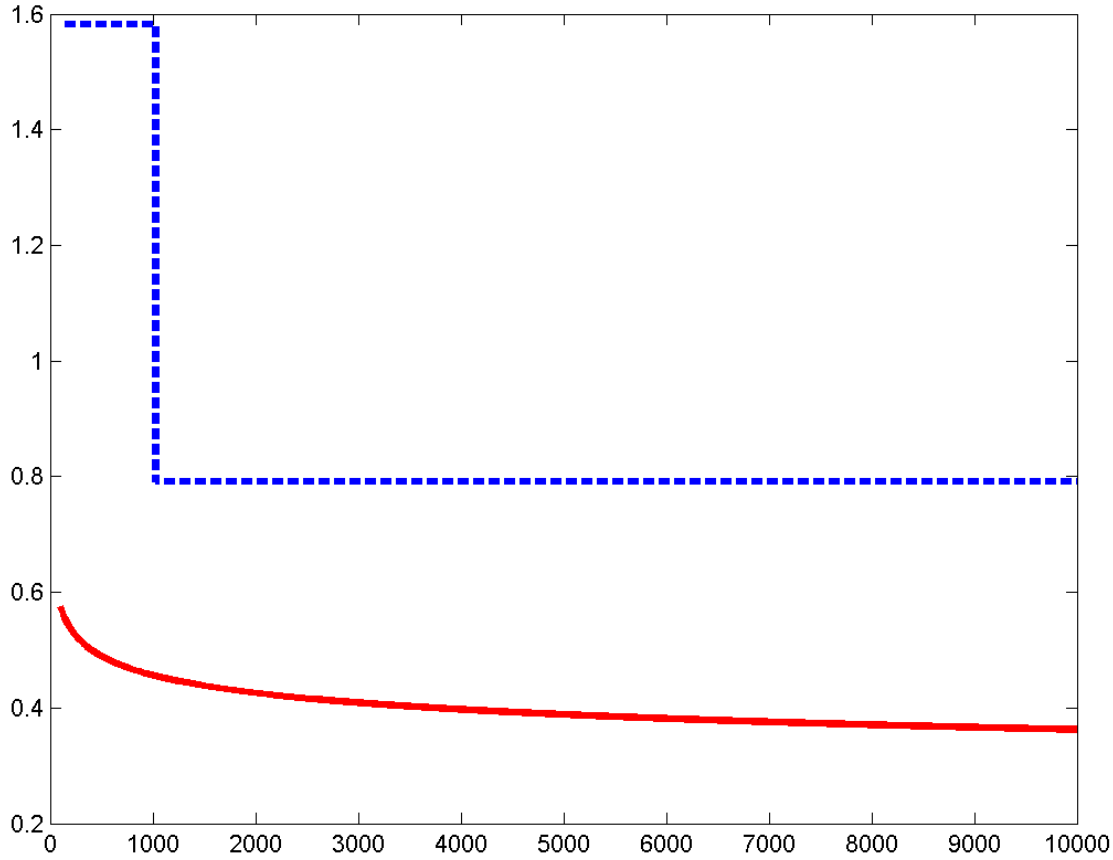
$d = 2$


0.18
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
20
40
60
80
100
120
140
160
180
200

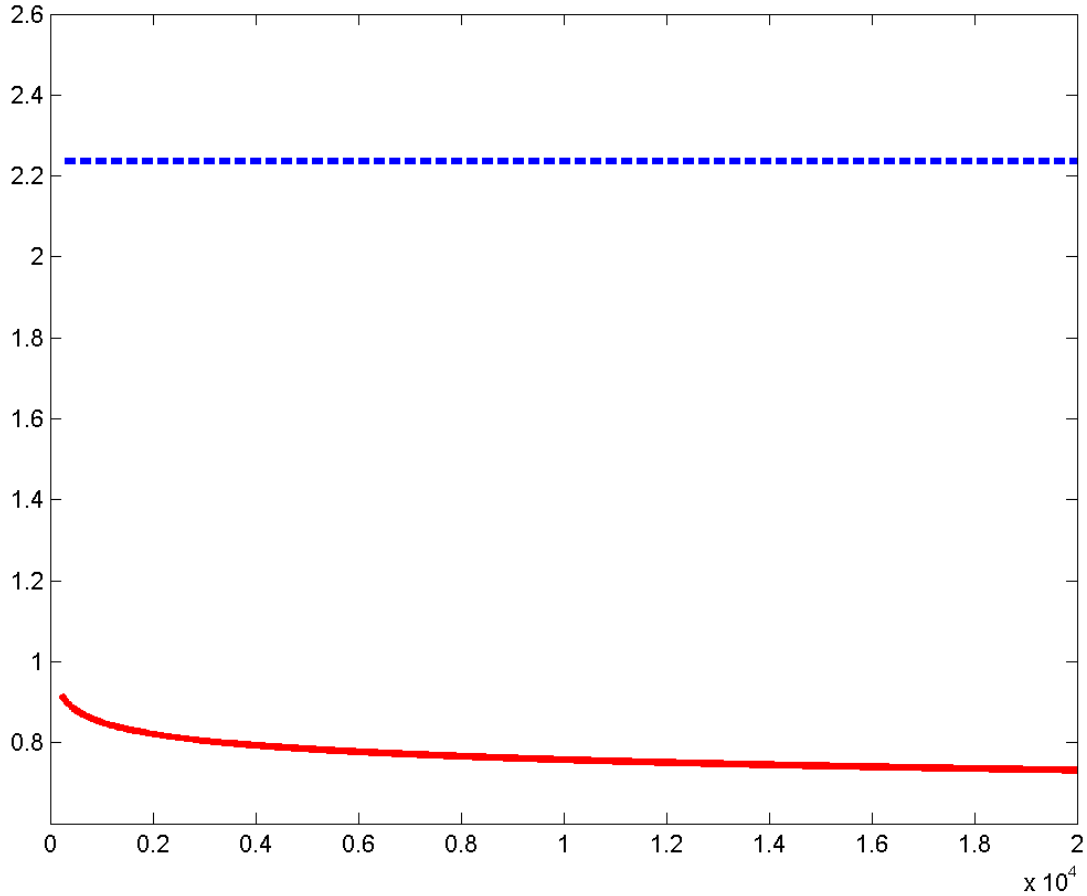
$d = 5$


0.6
0.55
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000

$d = 10$


1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
10000

$d = 20$



- ② Minimization of $\Phi_{mM}(X_n)$ with respect to $X_n \in \mathcal{X}^n$ for a given n

- ② Minimization of $\Phi_{mM}(X_n)$ with respect to $X_n \in \mathcal{X}^n$ for a given n
- ③ n is not fixed ($n_{\min} \leq n \leq n_{\max}$, we may stop before $n_{\max}$ evaluations of f)

How to obtain good "anytime designs", such that

all nested designs X_n have a high efficiency $\text{Eff}_{mM}(X_n)$, $n_{\min} \leq n \leq n_{\max}$

- ② Minimization of $\Phi_{mM}(X_n)$ with respect to $X_n \in \mathcal{X}^n$ for a given n
- ③ n is not fixed ($n_{\min} \leq n \leq n_{\max}$, we may stop before $n_{\max}$ evaluations of f)

  How to obtain good "anytime designs", such that

  all nested designs X_n have a high efficiency $\text{Eff}_{mM}(X_n)$, $n_{\min} \leq n \leq n_{\max}$
- ④ Design measures that minimize a regularized version of Φ_{mM}

2) Minimization of $\Phi_{mM}(X_n)$, $X_n \in \mathcal{X}^n$, n fixed

2) Minimization of $\Phi_{mM}(X_n)$, $X_n \in \mathscr{X}^n$, n fixed

- General global optimization method (e.g., simulated annealing): not promising
- 2.1) `k-means` and centroids
- 2.2) Stochastic gradient

2.1/ k-means and centroids

Minimize the L_2 energy functional

$$\mathcal{E}_2(\mathcal{T}_n, X_n) = \int_{\mathscr{X}} \left(\sum_{i=1}^{n} I_{\mathcal{C}_i}(\mathbf{x}) \, \|\mathbf{x} - \mathbf{x}_i\|^2 \right) \mathrm{d}\mathbf{x} = \sum_{i=1}^{n} \int_{\mathcal{C}_i} \|\mathbf{x} - \mathbf{x}_i\|^2 \, \mathrm{d}\mathbf{x}$$

where $\mathcal{T}_n = \{\mathcal{C}_i, \, i = 1, \ldots, n\}$ is a tessellation of $\mathscr{X}$

$I_{\mathcal{C}_i}$ = indicator function of $\mathcal{C}_i$

2.1/ k-means and centroids

Minimize the L_2 energy functional

$$\mathcal{E}_2(\mathcal{T}_n, X_n) = \int_{\mathscr{X}} \left(\sum_{i=1}^{n} I_{\mathcal{C}_i}(\mathbf{x}) \, \|\mathbf{x} - \mathbf{x}_i\|^2 \right) \mathrm{d}\mathbf{x} = \sum_{i=1}^{n} \int_{\mathcal{C}_i} \|\mathbf{x} - \mathbf{x}_i\|^2 \, \mathrm{d}\mathbf{x}$$

where $\mathcal{T}_n = \{\mathcal{C}_i, \, i = 1, \ldots, n\}$ is a tessellation of $\mathscr{X}$
$I_{\mathcal{C}_i}$ = indicator function of $\mathcal{C}_i$

Then (Du et al., 1999):

- $\mathcal{C}_i = \mathcal{V}(\mathbf{x}_i)$ = Voronoï region for the site $\mathbf{x}_i$, for all i
  $(\Rightarrow \mathcal{E}_2(\mathcal{T}_n, X_n) = \int_{\mathscr{X}} d^2(\mathbf{x}, X_n) \, \mathrm{d}\mathbf{x})$
- simultaneously $\mathbf{x}_i$ = centroid of $\mathcal{C}_i$ (center of gravity) for all i:
  $\mathbf{x}_i = (\int_{\mathcal{C}_i} \mathbf{x} \, \mathrm{d}\mathbf{x})/\mathrm{vol}(\mathcal{C}_i)$

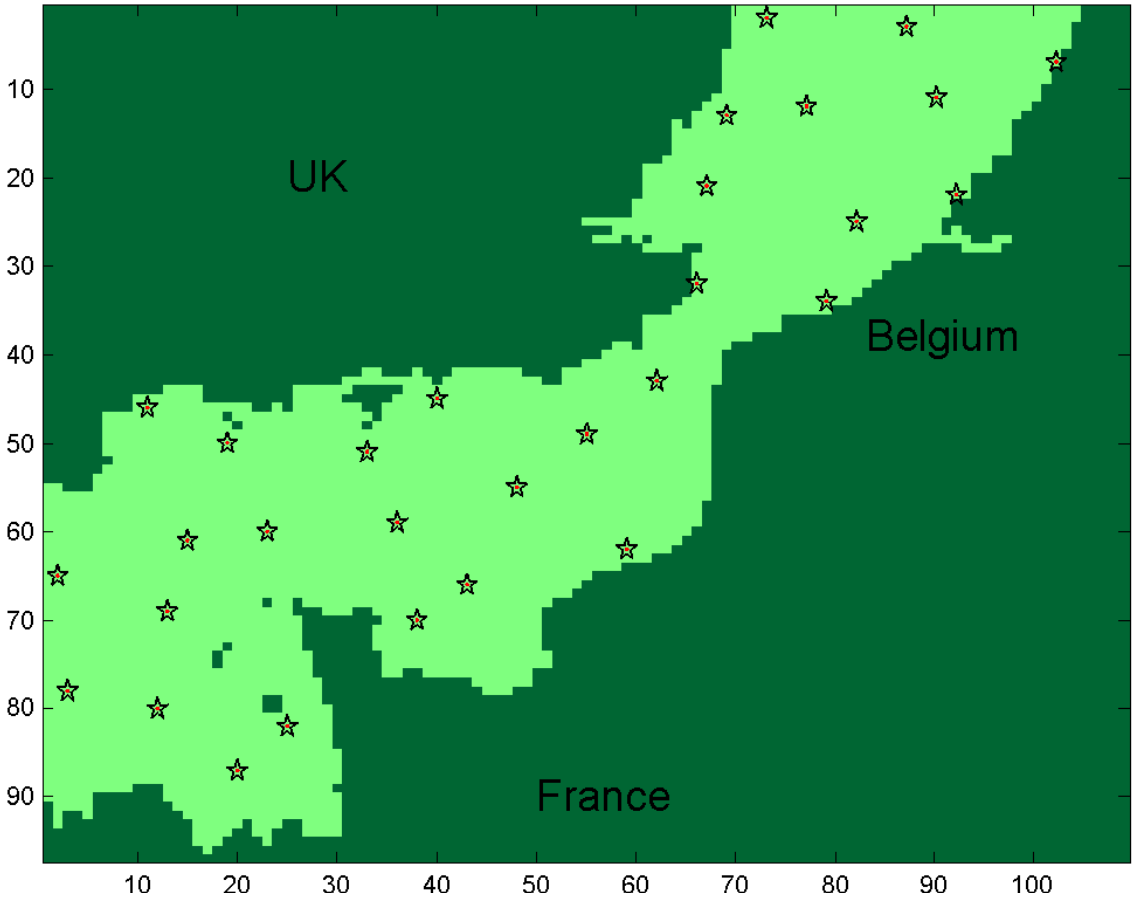
→ such a X_n should thus perform reasonably well in terms of space-filling (Lekivetz and Jones, 2015)

Lloyd's method (1982): (= fixed-point iterations)

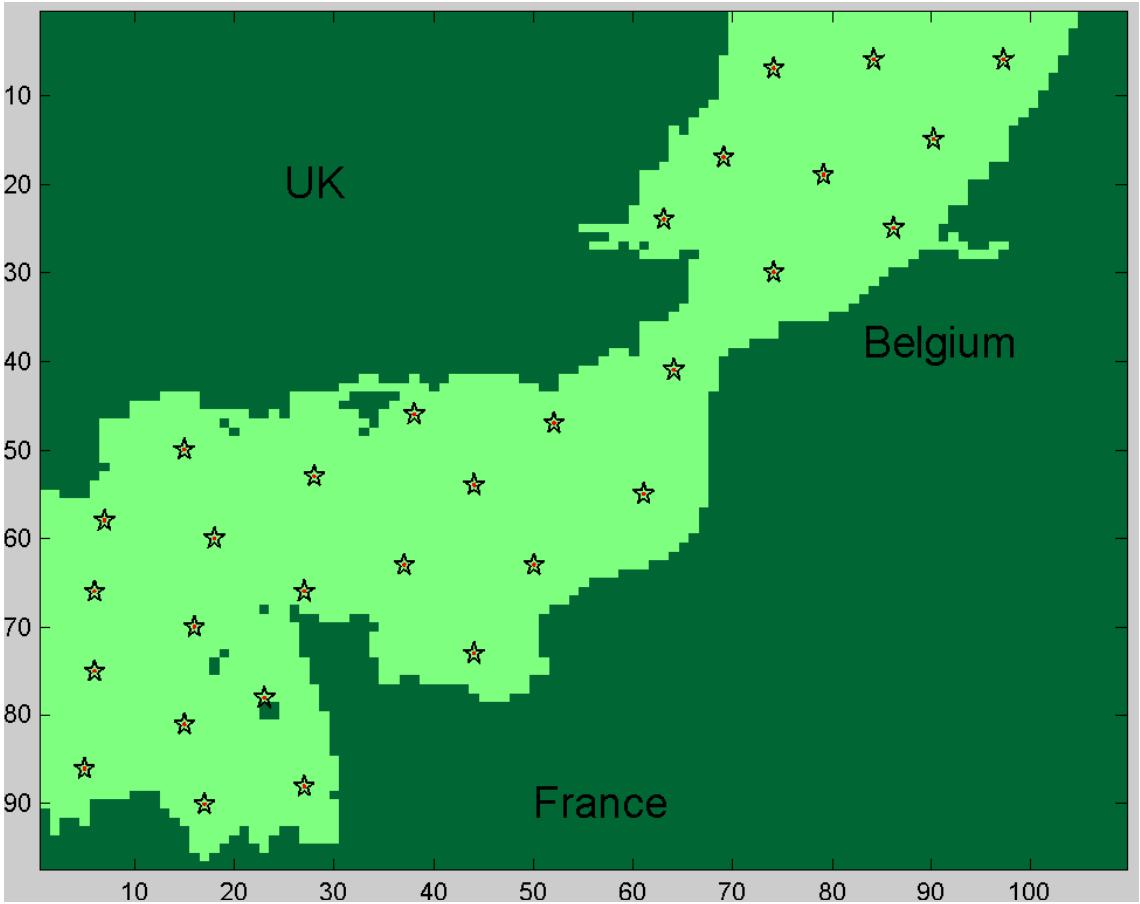
→ Move each $\mathbf{x}_i$ to the centroid of its own Voronoï cell, repeat ...

➠ Algorithmic geometry (Voronoï tessellation) if d very small,
use a finite set $\mathscr{X}_Q$ otherwise

30 points from Sobol' LDS



k-means clustering (30 clusters) of 1,000 point from Sobol' LDS



However... minimax-optimal design is related to the construction of a centroidal tessellation for

$$\mathcal{E}_q(\mathcal{T}_n, X_n) = \int_{\mathscr{X}} \left(\sum_{i=1}^{n} I_{\mathcal{C}_i}(\mathbf{x}) \, \|\mathbf{x} - \mathbf{x}_i\|^q \right) \mathrm{d}\mathbf{x} = \sum_{i=1}^{n} \int_{\mathcal{C}_i} \|\mathbf{x} - \mathbf{x}_i\|^q \, \mathrm{d}\mathbf{x}$$

for $q \to \infty$

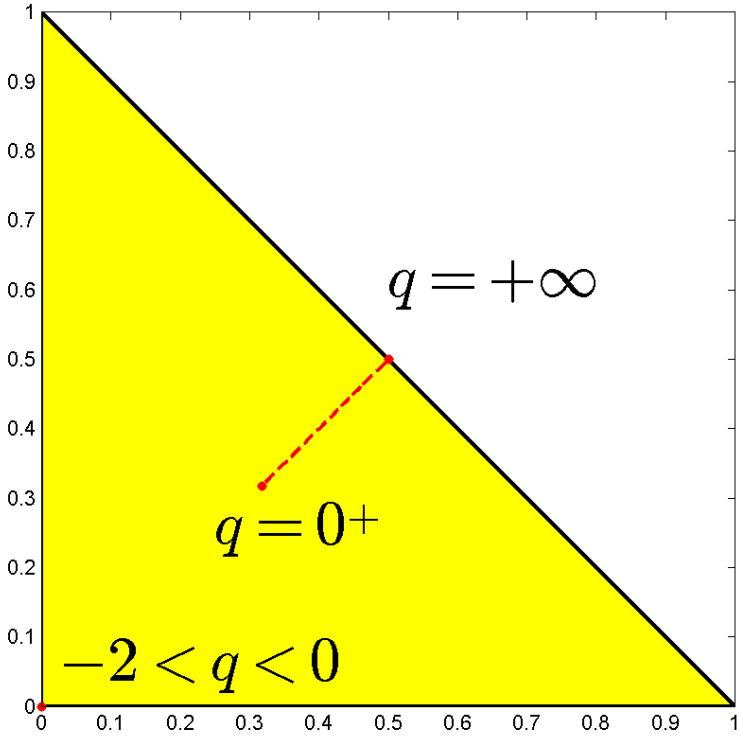
➠ use Chebyshev centers

However... minimax-optimal design is related to the construction of a centroidal tessellation for

$$\mathcal{E}_q(\mathcal{T}_n, X_n) = \int_{\mathcal{X}} \left(\sum_{i=1}^{n} l_{\mathcal{C}_i}(\mathbf{x}) \, \|\mathbf{x} - \mathbf{x}_i\|^q \right) \mathrm{d}\mathbf{x} = \sum_{i=1}^{n} \int_{\mathcal{C}_i} \|\mathbf{x} - \mathbf{x}_i\|^q \, \mathrm{d}\mathbf{x}$$

for $q \to \infty$

➠ use Chebyshev centers

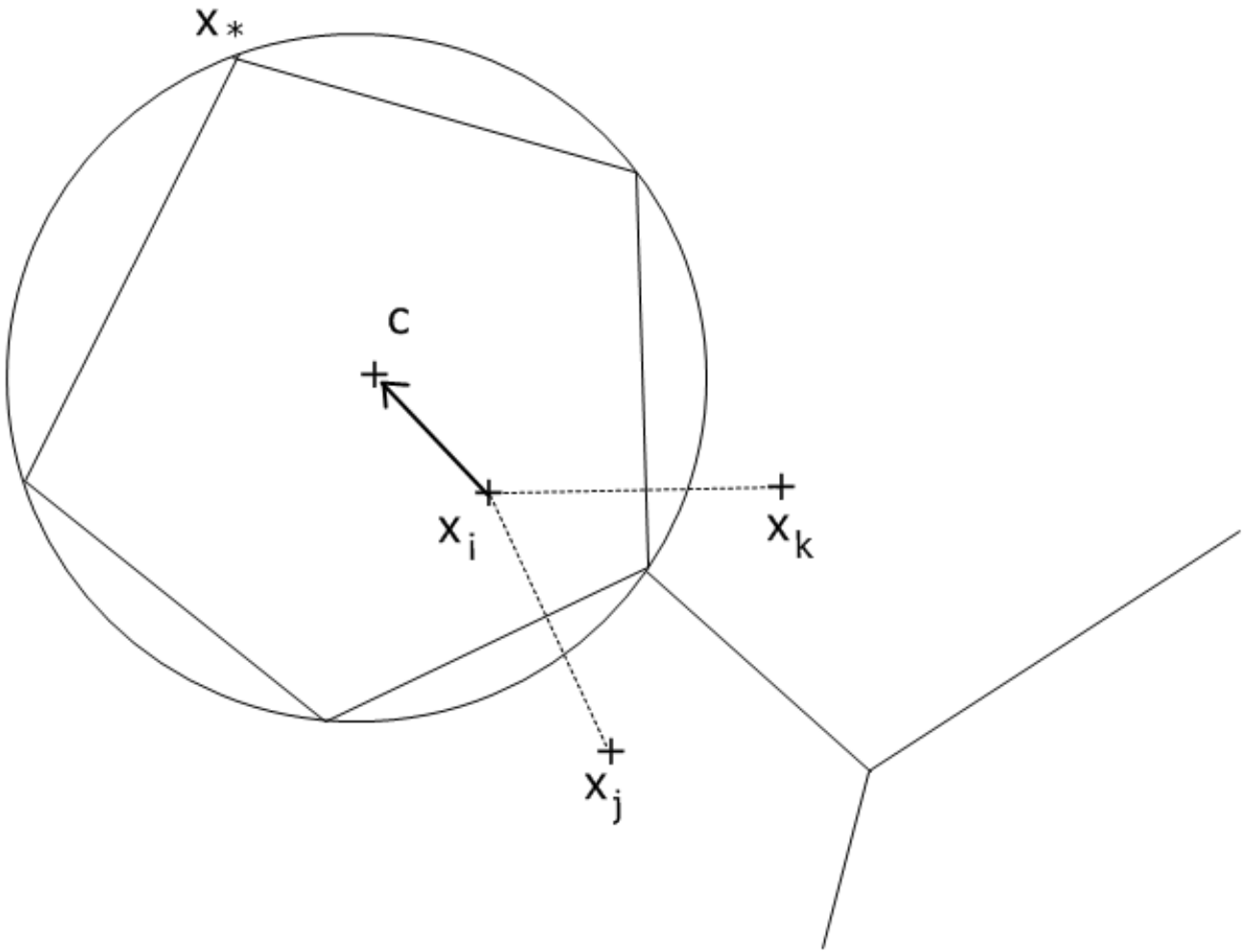


Variant of Lloyd's method:

0) Select $X_n^{(1)}$ and $\epsilon \ll 1$, set $k = 1$
1) Compute the Voronoï tessellation $\{\mathcal{V}_i,\ i = 1, \ldots, n\}$ of $\mathscr{X}$ (or $\mathscr{X}_Q$) based on $X_n^{(k)}$
2) For $i = 1, \ldots, n$
   - ➤ determine the smallest ball $\mathscr{B}(\mathbf{c}_i, r_i)$ enclosing $\mathcal{V}_i$ (= convex QP problem)
   - ➤ replace $\mathbf{x}_i$ by $\mathbf{c}_i$ in $X_n^{(k)}$ (Chebyshev center of $\mathcal{V}_i$)
3) if $\Phi_{mM}(\mathbf{X}_n^{(k)}) - \Phi_{mM}(\mathbf{X}_n^{(k+1)}) < \epsilon$, then stop; otherwise $k \leftarrow k + 1$, return to step 1

→ Move each $\mathbf{x}_i$ to the Chebyshev center of its own Voronoï cell, repeat ...

[$\Phi_{mM}(\mathbf{X}_n^{(k)})$ decreases monotonically, convergence to a local minimum (or a saddle point)]


x*
c
x_i
x_k
x_j

Determination of the smallest enclosing ball containing $\mathcal{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$
(vertices of a Voronoï cell, points of $\mathscr{X}_Q$ closest to $\mathbf{x}_i$):

$$\Leftrightarrow \text{ minimize } f(\mathbf{c}) = \max_{i=1,\ldots,N} \|\mathbf{z}_i - \mathbf{c}\|^2 \text{ with respect to } \mathbf{c} \in \mathbb{R}^d$$

Determination of the smallest enclosing ball containing $\mathcal{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$
(vertices of a Voronoï cell, points of $\mathscr{X}_Q$ closest to $\mathbf{x}_i$):

$$\Leftrightarrow \text{ minimize } f(\mathbf{c}) = \max_{i=1,\ldots,N} \|\mathbf{z}_i - \mathbf{c}\|^2 \text{ with respect to } \mathbf{c} \in \mathbb{R}^d$$

Direct problem = convex QP

Take any $\mathbf{c}_0 \in \mathbb{R}^d$, minimize $\|\mathbf{c} - \mathbf{c}_0\|^2 + t$

with respect to $(\mathbf{c}, t) \in \mathbb{R}^{d+1}$,

subject to $\|\mathbf{z}_i - \mathbf{c}_0\|^2 - 2(\mathbf{z}_i - \mathbf{c}_0)^\top(\mathbf{c} - \mathbf{c}_0) \leq t\,, \;\; i = 1, \ldots, N$

(N linear constraints)

Determination of the smallest enclosing ball containing $\mathcal{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$

Dual problem = similar to an optimal design problem:

maximize $\mathrm{trace}[\mathbf{V}(\xi)]$, with ξ a prob. measure on $\mathcal{Z}$,

$\mathbf{V}(\xi)$ = covariance matrix for ξ

center of the ball = $\mathbf{c}(\xi) = \int_{\mathcal{Z}} \mathbf{z}\, \xi(\mathrm{d}\mathbf{z})$

Determination of the smallest enclosing ball containing $\mathcal{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\}$

Dual problem = similar to an optimal design problem:

maximize $\mathrm{trace}[\mathbf{V}(\xi)]$, with ξ a prob. measure on $\mathcal{Z}$,

$\mathbf{V}(\xi)$ = covariance matrix for ξ

center of the ball = $\mathbf{c}(\xi) = \int_{\mathcal{Z}} \mathbf{z}\, \xi(\mathrm{d}\mathbf{z})$

→ Algorithms of the exchange-type (Yildirim, 2008)
(≈ Fedorov algorithm for D-optimal design: optimal step length is available)

→ One can remove inessential points from $\mathcal{Z}$: (LP, 2017b)

➠ Combine this with the use of a standard QP solver for the direct problem

2.2/ Stochastic gradient

d **is large**: Lloyd's algorithm cannot be used (computational geometry is too complicated, regular grids or LDS are not dense enough)

$$\text{minimize } \mathcal{E}_q{}^*(X_n) = \int_{\mathscr{X}} \left(\sum_{i=1}^{n} I_{\mathcal{V}_i}(\mathbf{x}) \, \|\mathbf{x} - \mathbf{x}_i\|^q \right) \mathrm{d}\mathbf{x}$$

with $\mathcal{V}_i$ = Voronoï region for the site $\mathbf{x}_i$

2.2/ Stochastic gradient

d **is large**: Lloyd's algorithm cannot be used (computational geometry is too complicated, regular grids or LDS are not dense enough)

$$\text{minimize } \mathcal{E}_q{}^*(X_n) = \int_{\mathscr{X}} \left(\sum_{i=1}^{n} I_{\mathcal{V}_i}(\mathbf{x}) \, \|\mathbf{x} - \mathbf{x}_i\|^q \right) \mathrm{d}\mathbf{x}$$

with $\mathcal{V}_i$ = Voronoï region for the site $\mathbf{x}_i$

→ Stochastic gradient algorithm:
(MacQueen, 1967) for $q = 2$, (Cardot et al., 2012) for $q = 1$

0) $k = 1$, $X_n^{(1)}$, set $n_{i,0} = 0$ for all $i = 1, \ldots, n$

1) sample X uniformly distributed in $\mathscr{X}$

2) find $i^* = \arg\min_{i=1,\ldots,n} \|X - \mathbf{x}_i^{(k)}\|$, $n_{i^*,k} \leftarrow n_{i^*,k} + 1$ $\quad [\leftarrow X \in \text{cell } \mathcal{V}_i^*]$

3) $\mathbf{x}_{i^*}^{(k+1)} = \mathbf{x}_{i^*}^{(k)} - \gamma_{i^*,k} \underbrace{q\|X - \mathbf{x}_{i^*}^{(k)}\|^{q-2} (\mathbf{x}_{i^*}^{(k)} - X)}_{=\text{gradient}}$, $k \leftarrow k + 1$,

return to step 1, stop when $k = K$

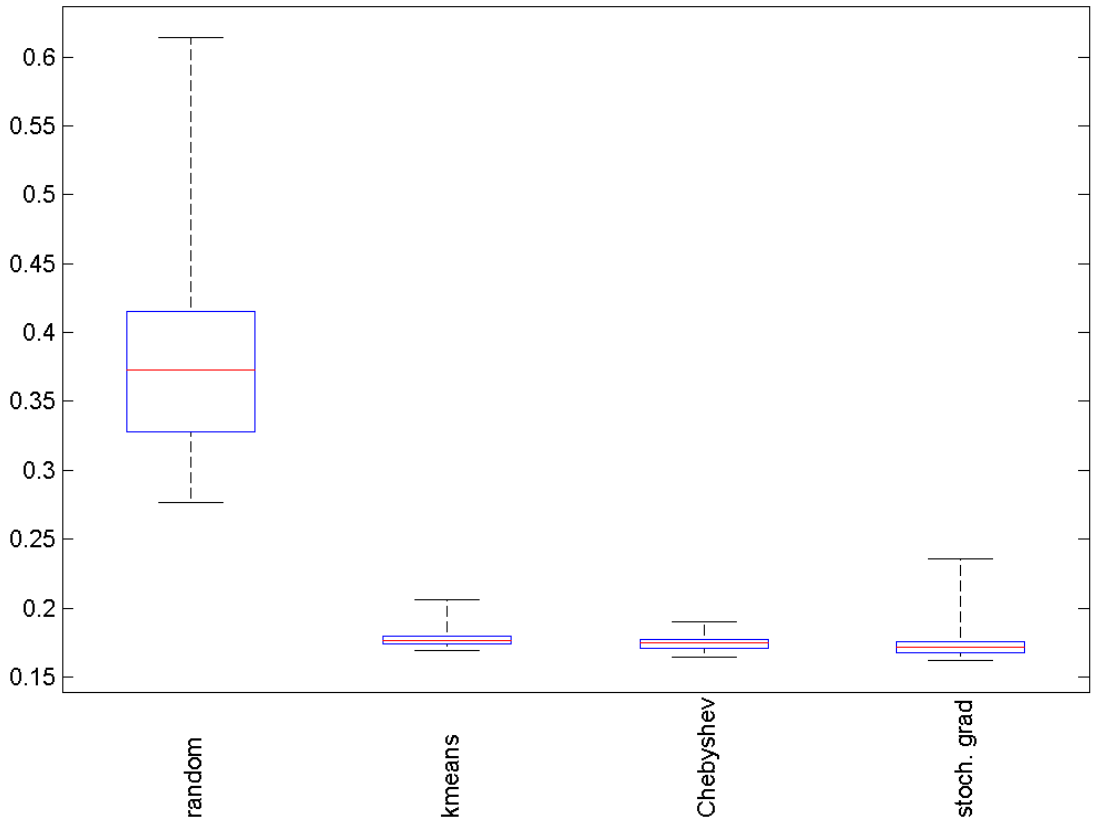
- Typical choice for $\gamma_{i^*,k} = c/n_{i^*,k}^{\alpha}$, with $\alpha \in (1/2, 1]$
  and consider $\widehat{X}_n = \frac{1}{K} \sum_{k=1}^{K} X_n^{(k)}$ when $\alpha < 1$
- Little information to store (no grid or other finite approximation of $\mathscr{X}$)
  $\rightarrow$ can also be used with large d

Example: $n = 10\,d$

all methods are initialized at the same random design, 100 repetitions

`k-means` and Lloyd's method with Chebyshev centers use 2^{d+8} points from a LDS (Sobol')

$$d = 2,\ n = 20\ (\underline{R}_n^* \approx 0.1262,\ \overline{R}_n^* \approx 0.1768)$$

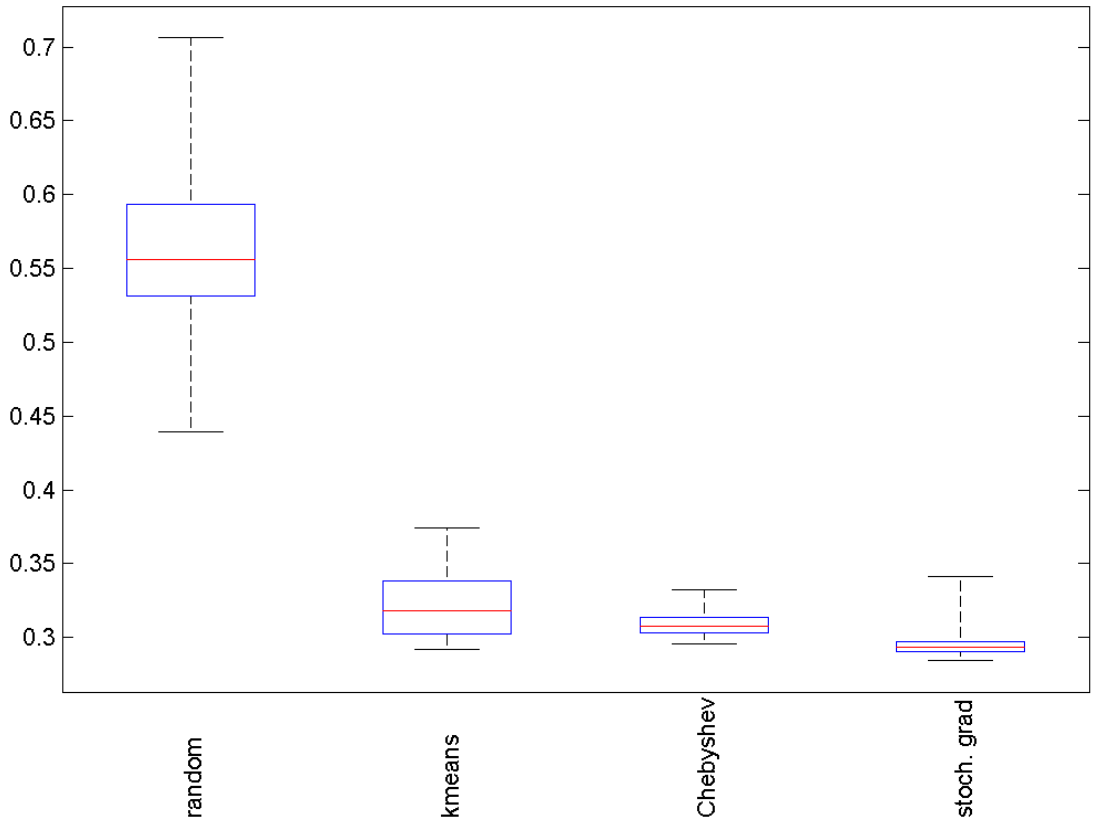


Example: $n = 10\,d$

all methods are initialized at the same random design, 100 repetitions

`k-means` and Lloyd's method with Chebyshev centers use 2^{d+8} points from a LDS (Sobol')

$$d = 3,\ n = 30\ (\underline{R}_n^* \approx 0.1996,\ \overline{R}_n^* \approx 0.2887)$$

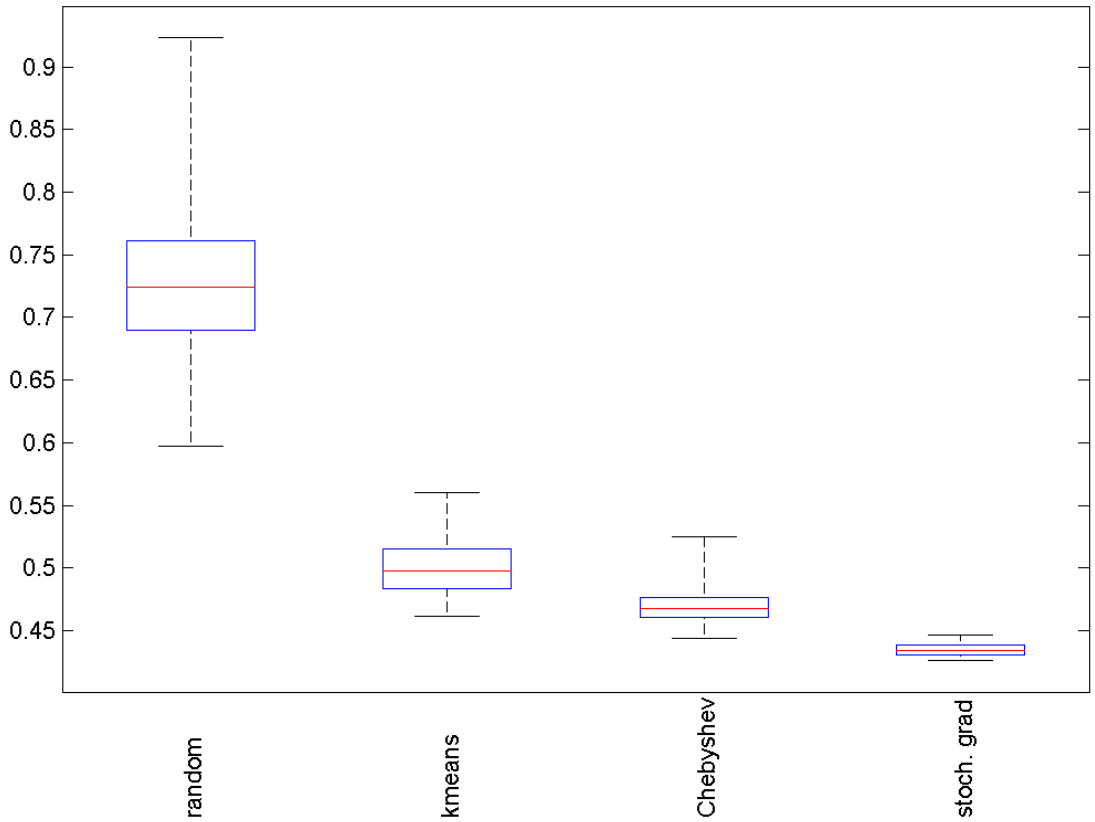


Example: $n = 10\,d$

all methods are initialized at the same random design, 100 repetitions

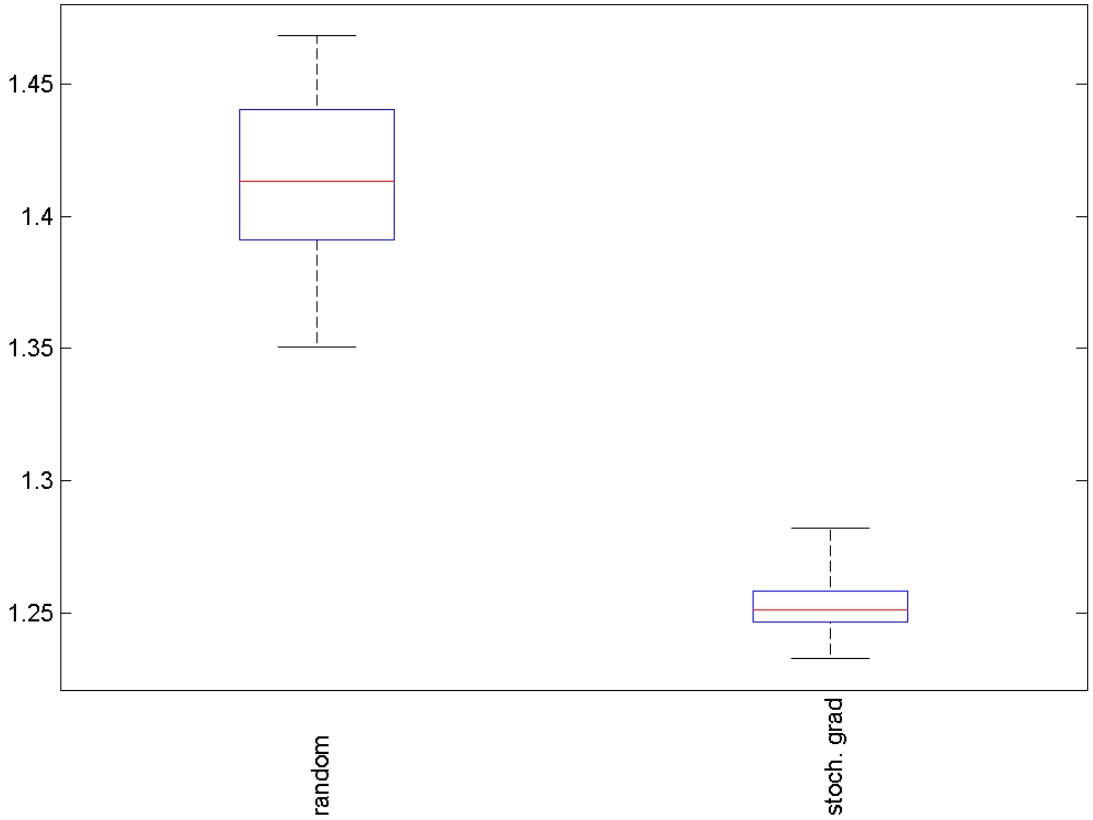
`k-means` and Lloyd's method with Chebyshev centers use 2^{d+8} points from a LDS (Sobol')

$$d = 4,\ n = 40\ (\underline{R}_n^* \approx 0.2668,\ \overline{R}_n^* = 0.5)$$



Example:

$$d = 10,\ n = 100\ (\underline{R}_n^* \approx 0.5746,\ \overline{R}_n^* \approx 1.5811)$$



3) Nested designs

➠ obtain a high Φ_{mM}-efficiency $\text{Eff}_{mM}(X_n) = \frac{\Phi^*_{mM,n}}{\Phi_{mM}(X_n)}$ for **all** X_n, $n_{\min} \leq n \leq n_{\max}$

$[\text{Eff}_{mM}(X_n) \in (0,1]]$

3.1/ Coffee-house design

$\mathbf{x}_1$ at the centre of $\mathscr{X}$, then $\mathbf{x}_{n+1}$ furthest point from X_n for all $n \geq 1$ (called coffee-house design (Müller, 2007, Chap. 4))

3.1/ Coffee-house design

$\mathbf{x}_1$ at the centre of $\mathscr{X}$, then $\mathbf{x}_{n+1}$ furthest point from X_n for all $n \geq 1$
(called coffee-house design (Müller, 2007, Chap. 4))

Guarantees $\boxed{\text{Eff}_{mM}(X_n) = \frac{\Phi^*_{mM,n}}{\Phi_{mM}(X_n)} \geq \frac{1}{2} \text{ and } \text{Eff}_{Mm}(X_n) = \frac{\Phi_{Mm}(X_n)}{\Phi^*_{Mm,n}} \geq \frac{1}{2} \text{ for all } n}$

with $\Phi_{Mm}(X_n) = \min_{i \neq j \in \{1,\ldots,n\}} \|\mathbf{x}_i - \mathbf{x}_j\|$ the maximin-distance criterion,
and $\Phi^*_{Mm,n}$ its optimal (maximum) value

3.1/ Coffee-house design

$\mathbf{x}_1$ at the centre of $\mathscr{X}$, then $\mathbf{x}_{n+1}$ furthest point from X_n for all $n \geq 1$
(called coffee-house design (Müller, 2007, Chap. 4))

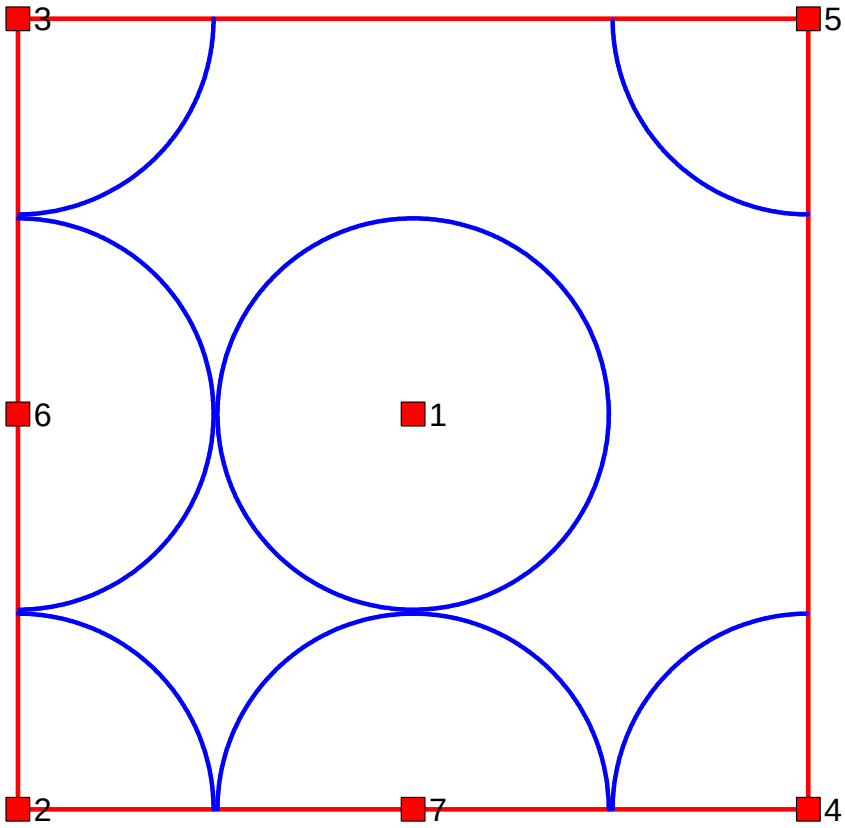
Guarantees $\boxed{\text{Eff}_{mM}(X_n) = \frac{\Phi^*_{mM,n}}{\Phi_{mM}(X_n)} \geq \frac{1}{2} \text{ and } \text{Eff}_{Mm}(X_n) = \frac{\Phi_{Mm}(X_n)}{\Phi^*_{Mm,n}} \geq \frac{1}{2} \text{ for all } n}$

with $\Phi_{Mm}(X_n) = \min_{i \neq j \in \{1,\ldots,n\}} \|\mathbf{x}_i - \mathbf{x}_j\|$ the maximin-distance criterion,
and $\Phi^*_{Mm,n}$ its optimal (maximum) value

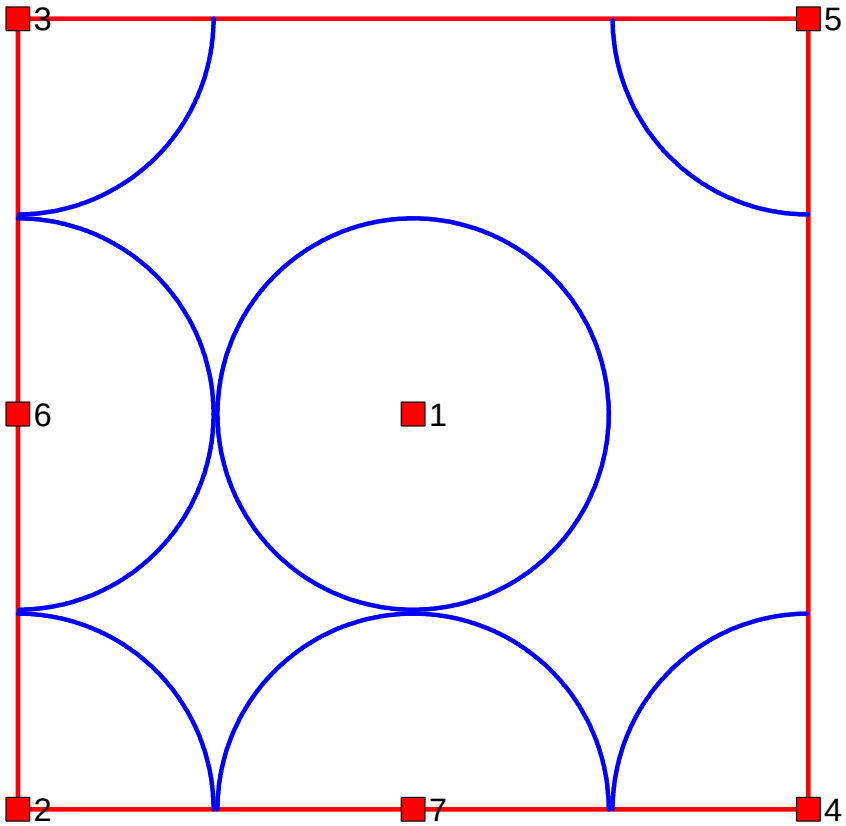
Proof. (Gonzalez, 1985)

- by construction:
  $\Phi_{Mm}(X_{n+1}) \triangleq \min_{\mathbf{x}_i \neq \mathbf{x}_j \in X_{n+1}} \|\mathbf{x}_i - \mathbf{x}_j\| = d(\mathbf{x}_{n+1}, X_n) = \Phi_{mM}(X_n)$
- X^*_n a Φ_{mM}-optimal design: the n balls $\mathscr{B}(\mathbf{x}^*_i, \Phi_{mM}(X^*_n))$, $\mathbf{x}^*_i \in X^*_n$, cover $\mathscr{X}$
  $\Rightarrow$ one of them contains 2 points $\mathbf{x}_i$, $\mathbf{x}_j$ in X_{n+1} *for any* X_{n+1} *($n+1$ points)*
  $\Rightarrow \Phi_{Mm}(X_{n+1}) \leq \|\mathbf{x}_i - \mathbf{x}_j\| \leq 2\Phi_{mM}(X^*_n)$
  $\Rightarrow \Phi^*_{Mm,n+1} \leq 2\Phi_{mM}(X^*_n) \leq 2\Phi_{mM}(X_n) = \Phi_{Mm}(X_{n+1})$

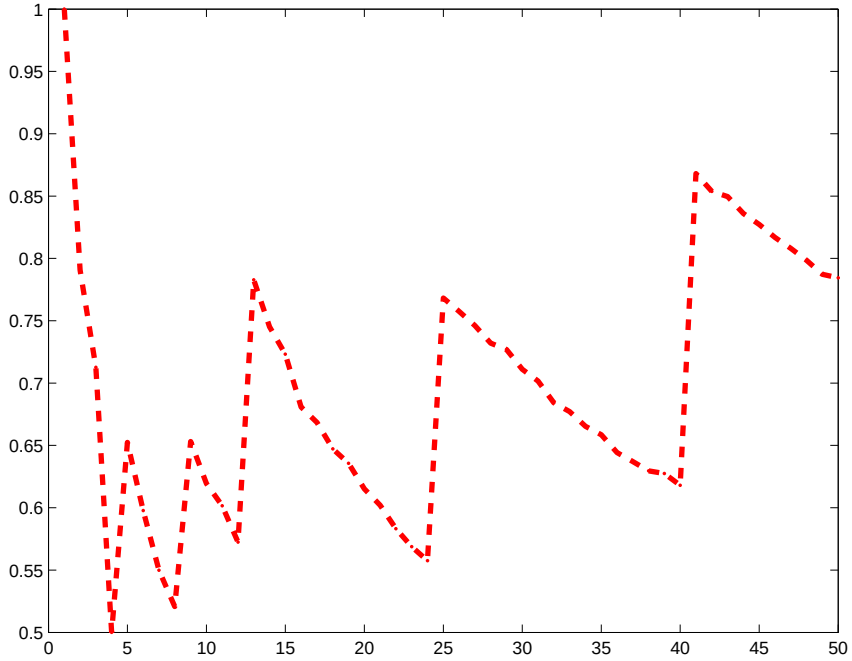
$\mathscr{X} = [0,1]^2$, $n = 7$



$\mathscr{X} = [0,1]^2$, $n = 7$



$\text{Eff}_{mM}(X_n)$, $n = 1 \ldots, 50$



Regular construction ➠ large fluctuations of $\text{Eff}_{mM}(X_n)$

3.2/ Submodularity and greedy algorithms

$\mathscr{X}_Q = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(Q)}\}$ a finite set with Q points in $\mathscr{X}$

(regular grid, first Q points of a LDS — Halton, Sobol' ...)

ψ: $2^{\mathscr{X}_Q} \longrightarrow \mathbb{R}$ a set function (to be maximized)

non-decreasing: $\psi(\mathscr{A} \cup \{\mathbf{x}\}) \geq \psi(\mathscr{A})$ for all $\mathscr{A} \subset \mathscr{X}_Q$ and $\mathbf{x} \in \mathscr{X}_Q$

Definition 1:

ψ is submodular iff $\psi(\mathscr{A}) + \psi(\mathscr{B}) \geq \psi(\mathscr{A} \cup \mathscr{B}) + \psi(\mathscr{A} \cap \mathscr{B})$ for all $\mathscr{A}, \mathscr{B} \subset \mathscr{X}_Q$

3.2/ Submodularity and greedy algorithms

$\mathscr{X}_Q = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(Q)}\}$ a finite set with Q points in $\mathscr{X}$
(regular grid, first Q points of a LDS — Halton, Sobol' ...)
ψ: $2^{\mathscr{X}_Q} \longrightarrow \mathbb{R}$ a set function (to be maximized)
non-decreasing: $\psi(\mathscr{A} \cup \{\mathbf{x}\}) \geq \psi(\mathscr{A})$ for all $\mathscr{A} \subset \mathscr{X}_Q$ and $\mathbf{x} \in \mathscr{X}_Q$

Definition 1:
ψ is submodular iff $\psi(\mathscr{A}) + \psi(\mathscr{B}) \geq \psi(\mathscr{A} \cup \mathscr{B}) + \psi(\mathscr{A} \cap \mathscr{B})$ for all $\mathscr{A}, \mathscr{B} \subset \mathscr{X}_Q$

Equivalently, **Definition 1'** (diminishing return property):
ψ is submodular iff $\psi(\mathscr{A} \cup \{\mathbf{x}\}) - \psi(\mathscr{A}) \geq \psi(\mathscr{B} \cup \{\mathbf{x}\}) - \psi(\mathscr{B})$ for all $\mathscr{A} \subset \mathscr{B} \subset \mathscr{X}_Q$ and $\mathbf{x} \in \mathscr{X}_Q \setminus \mathscr{B}$

(a sort of concavity property for set functions)

Greedy Algorithm:

1. set $\mathscr{A} = \emptyset$
2. while $|\mathscr{A}| < k$
   - find $\mathbf{x}$ in $\mathscr{X}_Q$ such that $\psi(\mathscr{A} \cup \{\mathbf{x}\})$ is maximal
   - $\mathscr{A} \leftarrow \mathscr{A} \cup \{\mathbf{x}\}$
3. end while
4. return $\mathscr{A}_k = \mathscr{A}$

Greedy Algorithm:

1. set $\mathscr{A} = \emptyset$
2. while $|\mathscr{A}| < k$
   - find $\mathbf{x}$ in $\mathscr{X}_Q$ such that $\psi(\mathscr{A} \cup \{\mathbf{x}\})$ is maximal
   - $\mathscr{A} \leftarrow \mathscr{A} \cup \{\mathbf{x}\}$
3. end while
4. return $\mathscr{A}_k = \mathscr{A}$

Denote $\psi_k^* = \max_{\mathscr{B} \subset \mathscr{X}_Q, |\mathscr{B}| \leq k} \psi(\mathscr{B})$

Theorem (Nemhauser, Wolsey & Fisher, 1978): When ψ is non-decreasing and submodular, then for all $k \in \{1, \ldots, Q\}$ the algorithm returns a set $\mathscr{A}_k$ such that

$$\frac{\psi(\mathscr{A}_k) - \psi(\emptyset)}{\psi_k^* - \psi(\emptyset)} \geq 1 - (1 - 1/k)^k \geq 1 - 1/\mathrm{e} > 0.6321$$

Greedy Algorithm:

1. set $\mathscr{A} = \emptyset$
2. while $|\mathscr{A}| < k$
   - find $\mathbf{x}$ in $\mathscr{X}_Q$ such that $\psi(\mathscr{A} \cup \{\mathbf{x}\})$ is maximal
   - $\mathscr{A} \leftarrow \mathscr{A} \cup \{\mathbf{x}\}$
3. end while
4. return $\mathscr{A}_k = \mathscr{A}$

Denote $\psi_k^* = \max_{\mathscr{B} \subset \mathscr{X}_Q, |\mathscr{B}| \leq k} \psi(\mathscr{B})$

Theorem (Nemhauser, Wolsey & Fisher, 1978): When ψ is non-decreasing and submodular, then for all $k \in \{1, \ldots, Q\}$ the algorithm returns a set $\mathscr{A}_k$ such that

$$\frac{\psi(\mathscr{A}_k) - \psi(\emptyset)}{\psi_k^* - \psi(\emptyset)} \geq 1 - (1 - 1/k)^k \geq 1 - 1/\mathrm{e} > 0.6321$$

Bad news: we maximize $-\Phi_{mM}$ which is non-decreasing but not submodular

➠ no guaranteed efficiency for sequential optimization

3.3/ Submodular alternatives to minimax

A) Covering measure, c.d.f. and dispersion *[SIAM UQ, Lausanne, 2016]*

For any $r \geq 0$, any $X_n \in \mathscr{X}^n$, define the covering measure of X_n by

$$\boxed{\psi_r(X_n) = \text{vol}\{\mathscr{X} \cap [\cup_{i=1}^n \mathscr{B}(\mathbf{x}_i, r)]\}}$$ ➠ non-decreasing and submodular

3.3/ Submodular alternatives to minimax

A) Covering measure, c.d.f. and dispersion *[SIAM UQ, Lausanne, 2016]*

For any $r \geq 0$, any $X_n \in \mathscr{X}^n$, define the covering measure of X_n by

$$\boxed{\psi_r(X_n) = \text{vol}\{\mathscr{X} \cap [\cup_{i=1}^n \mathscr{B}(\mathbf{x}_i, r)]\}}$$ ➠ non-decreasing and submodular

Maximizing $\psi_r(X_n)$ is equivalent to maximizing

$$F_{X_n}(r) = \psi_r(X_n)/\text{vol}(\mathscr{X}) = \frac{\mu_L\{\mathscr{X} \cap [\cup_{i=1}^n \mathscr{B}(\mathbf{x}_i, r)]\}}{\mu_L(\mathscr{X})}$$

which can be considered as a c.d.f., with $F_{X_n}(r) \in [0,1]$, increasing in r, and $F_{X_n}(0) = 0$, $F_{X_n}(r) = 1$ for any $r \geq \Phi_{mM}(X_n)$

3.3/ Submodular alternatives to minimax

A) Covering measure, c.d.f. and dispersion *[SIAM UQ, Lausanne, 2016]*

For any $r \geq 0$, any $X_n \in \mathscr{X}^n$, define the covering measure of X_n by

$$\boxed{\psi_r(X_n) = \text{vol}\{\mathscr{X} \cap [\cup_{i=1}^n \mathscr{B}(\mathbf{x}_i, r)]\}}$$ ➠ non-decreasing and submodular

Maximizing $\psi_r(X_n)$ is equivalent to maximizing

$$F_{X_n}(r) = \psi_r(X_n)/\text{vol}(\mathscr{X}) = \frac{\mu_L\{\mathscr{X} \cap [\cup_{i=1}^n \mathscr{B}(\mathbf{x}_i, r)]\}}{\mu_L(\mathscr{X})}$$

which can be considered as a c.d.f., with $F_{X_n}(r) \in [0,1]$, increasing in r, and $F_{X_n}(0) = 0$, $F_{X_n}(r) = 1$ for any $r \geq \Phi_{mM}(X_n)$

Take any probability measure μ on $\mathscr{X}$ (e.g., with finite support $\mathscr{X}_Q$)

➠ define $F_{X_n}(r) = \mu\{\mathscr{X} \cap [\cup_{i=1}^n \mathscr{B}(\mathbf{x}_i, r)]\}$

as a function of r ➔ forms a c.d.f.,

as a function of X_n ➔ non-decreasing and submodular

Which r should we take in $F_{X_n}(r)$?

A positive linear combination of non-decreasing submodular functions is non-decreasing and submodular

- Consider $\Psi_{b,B,q}(X_n) = \int_b^B r^q \, F_{X_n}(r) \, \mathrm{d}r$, for $B > b \geq 0$, $q > 0$

➔ guaranteed efficiency bounds when maximizing with a greedy algorithm

Which r should we take in $F_{X_n}(r)$?

A positive linear combination of non-decreasing submodular functions is non-decreasing and submodular

➠ Consider $\Psi_{b,B,q}(X_n) = \int_b^B r^q \, F_{X_n}(r) \, \mathrm{d}r$, for $B > b \geq 0$, $q > 0$

➔ guaranteed efficiency bounds when maximizing with a greedy algorithm

Justification:

$$\Psi_{0,B,q}(X_n) = \frac{B^{q+1}}{q+1} F_{X_n}(B) - \frac{1}{q+1} \int_0^B r^{q+1} \, F_{X_n}(\mathrm{d}r)$$

Take any $B \geq \Phi_{mM}(X_n)$ ➔ $F_{X_n}(B) = 1$

Maximizing $\Psi_{0,B,q}(X_n)$ for B large enough $\Leftrightarrow$ minimizing $\int_0^B r^{q+1} \, F_{X_n}(\mathrm{d}r)$

$$\Leftrightarrow \text{ minimizing } \left[\int_0^B r^{q+1} \, F_{X_n}(\mathrm{d}r) \right]^{1/(q+1)}$$

and $\left[\int_0^B r^{q+1} \, F_{X_n}(\mathrm{d}r) \right]^{1/(q+1)} \to \Phi_{mM}(X_n)$ as $q \to \infty$

Implementation

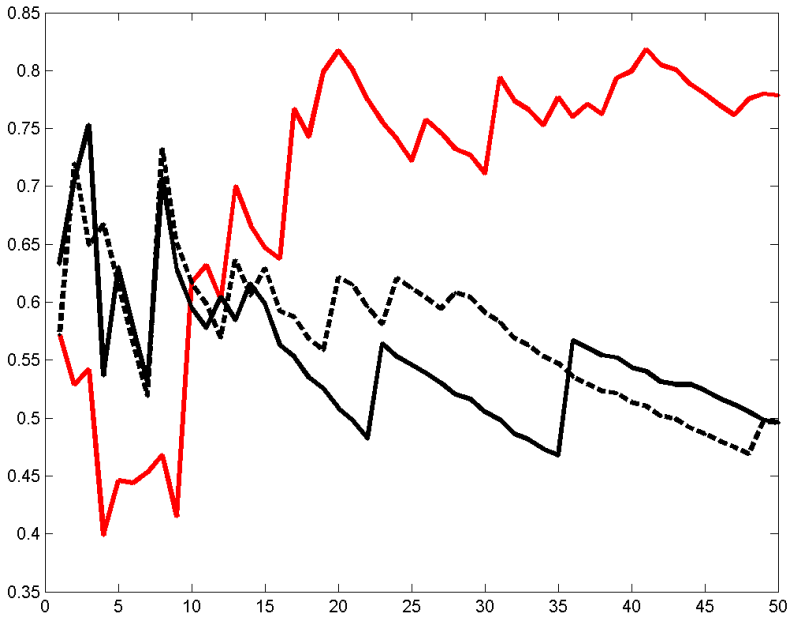
Easy when

- $\mathscr{X}$ approximated by $\mathscr{X}_Q = \{\mathbf{s}_1, \ldots, \mathbf{s}_Q\} \in \mathscr{X}^Q$, $\mu = \frac{1}{Q} \sum_{j=1}^{Q} \delta_{\mathbf{s}_j}$
- $X_n \in \mathscr{X}_Q{}^n$
  (inter-distances $\|\mathbf{s}_i - \mathbf{s}_j\|$ are only computed once)

Ex: $\mathscr{X} = [0,1]^2$, $\mathscr{X}_Q =$ grid with $Q = 33 \times 33 = 1089$ points
$n_{\min} = 15$, $n_{\max} = 50$, $q = 2$ in $\Psi_{b,B,q}(\cdot)$

➠ $\text{Eff}_{mM}(X_n)$ as a function of n

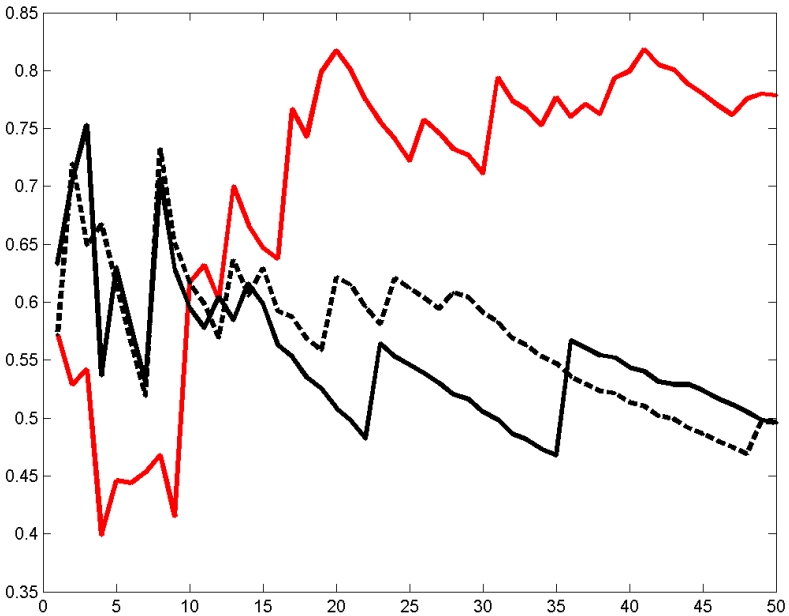
$\text{Eff}_{mM}(X_n)$: $\Psi_{b,B,q}(\cdot)$ —,
Halton LDS —, Sobol' LDS - -



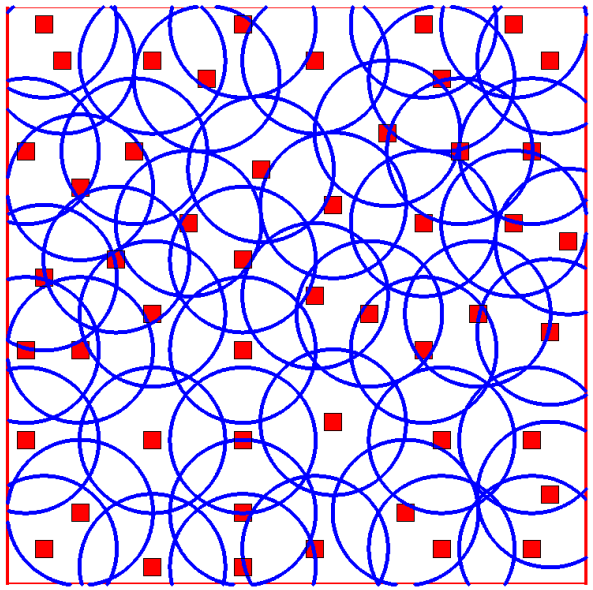
Ex: $\mathscr{X} = [0,1]^2$, $\mathscr{X}_Q =$ grid with $Q = 33 \times 33 = 1089$ points
$n_{\min} = 15$, $n_{\max} = 50$, $q = 2$ in $\Psi_{b,B,q}(\cdot)$

➠ $\text{Eff}_{mM}(X_n)$ as a function of n

$\text{Eff}_{mM}(X_n)$: $\Psi_{b,B,q}(\cdot)$ —,
Halton LDS —, Sobol' LDS - -



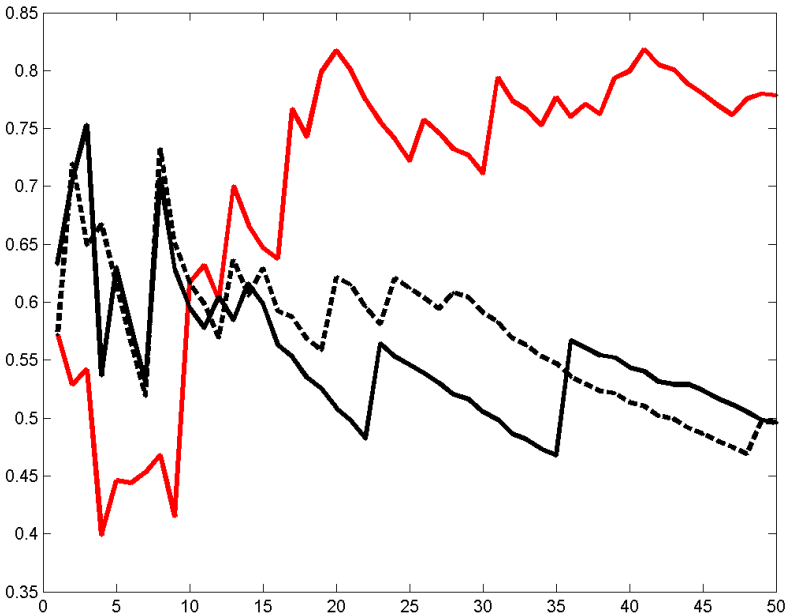
$X_{n_{\max}}$ with $\Psi_{b,B,q}(\cdot)$



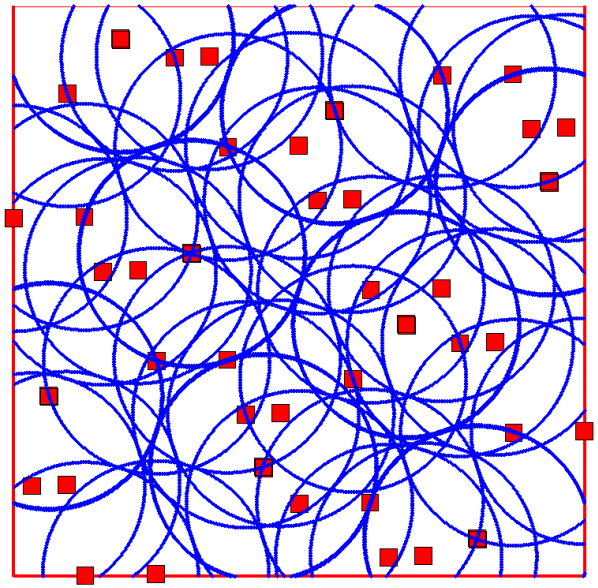
Ex: $\mathscr{X} = [0,1]^2$, $\mathscr{X}_Q =$ grid with $Q = 33 \times 33 = 1089$ points
$n_{\min} = 15$, $n_{\max} = 50$, $q = 2$ in $\Psi_{b,B,q}(\cdot)$

➠ $\text{Eff}_{mM}(X_n)$ as a function of n

$\text{Eff}_{mM}(X_n)$: $\Psi_{b,B,q}(\cdot)$ —,
Halton LDS —, Sobol' LDS - -



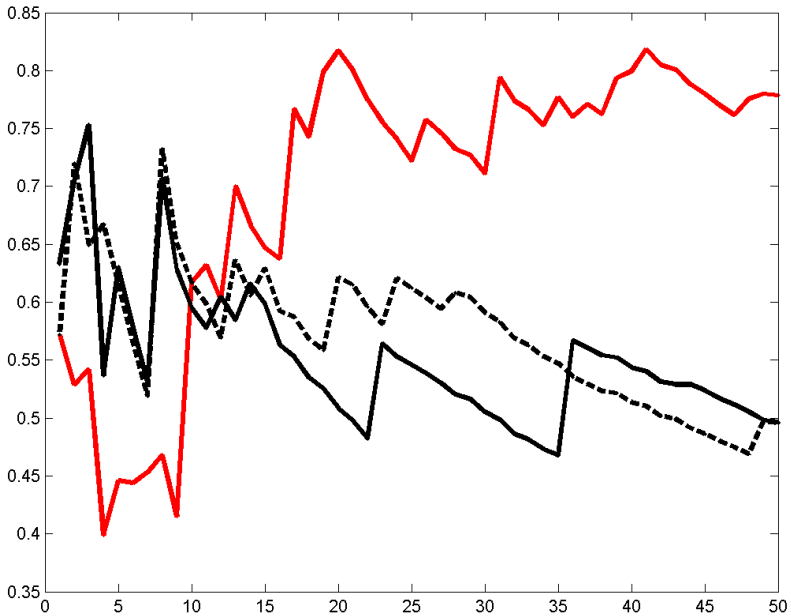
First $n_{\max}$ points of Sobol' LDS



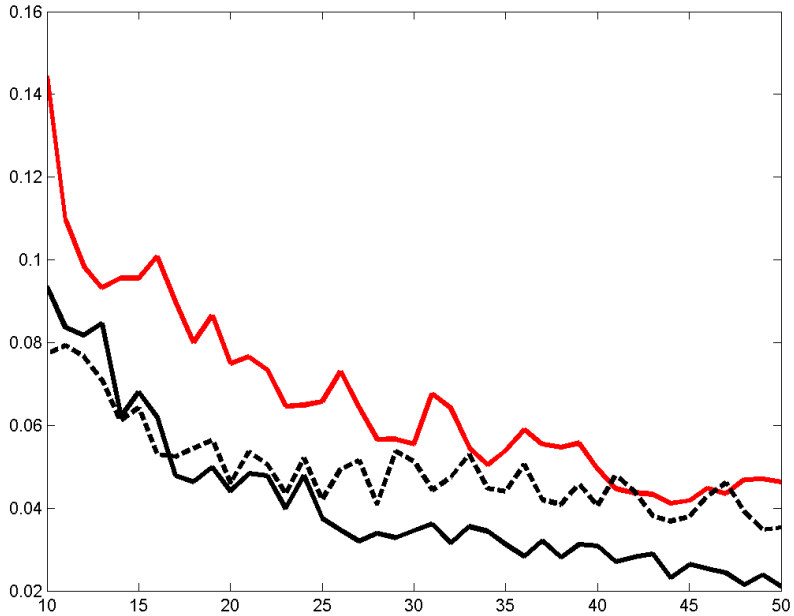
Ex: $\mathscr{X} = [0,1]^2$, $\mathscr{X}_Q =$ grid with $Q = 33 \times 33 = 1089$ points
$n_{\min} = 15$, $n_{\max} = 50$, $q = 2$ in $\Psi_{b,B,q}(\cdot)$

➠ $\text{Eff}_{mM}(X_n)$ as a function of n

$\text{Eff}_{mM}(X_n)$: $\Psi_{b,B,q}(\cdot)$ —,
Halton LDS —, Sobol' LDS - -



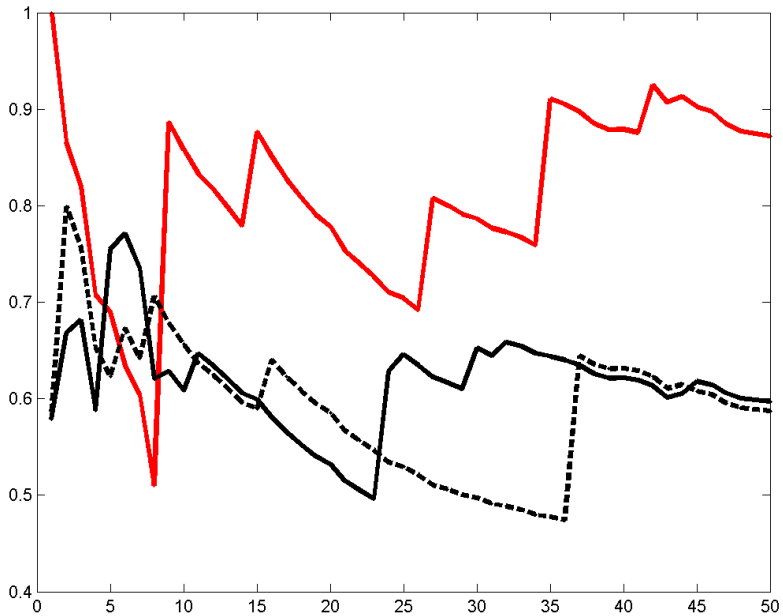
Centered L_2 discrepancies



Ex: $\mathscr{X} = [0,1]^3$, $\mathscr{X}_Q =$ grid with $Q = 11^3 = 1331$ points
$n_{\min} = 15$, $n_{\max} = 50$, $q = 2$ in $\Psi_{b,B,q}(\cdot)$

➠ $\text{Eff}_{mM}(X_n)$ as a function of n

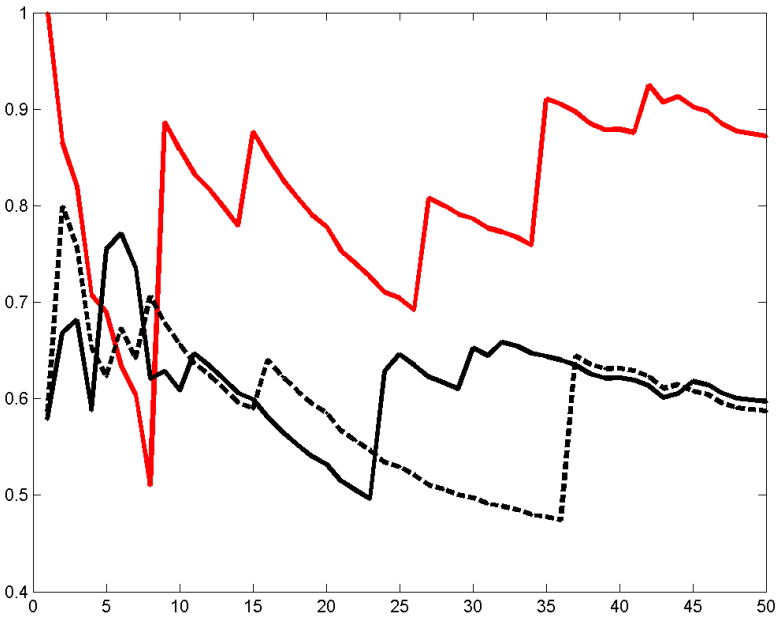
$\text{Eff}_{mM}(X_n)$: $\Psi_{b,B,q}(\cdot)$ —,
Halton LDS —, Sobol' LDS - -



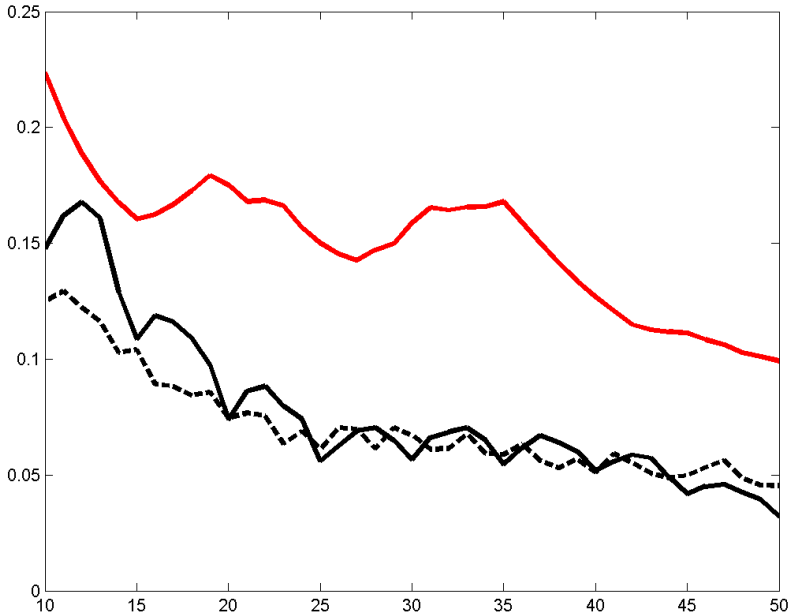
Ex: $\mathscr{X} = [0,1]^3$, $\mathscr{X}_Q =$ grid with $Q = 11^3 = 1331$ points
$n_{\min} = 15$, $n_{\max} = 50$, $q = 2$ in $\Psi_{b,B,q}(\cdot)$

➠ $\text{Eff}_{mM}(X_n)$ as a function of n

$\text{Eff}_{mM}(X_n)$: $\Psi_{b,B,q}(\cdot)$ —,
Halton LDS —, Sobol' LDS - -

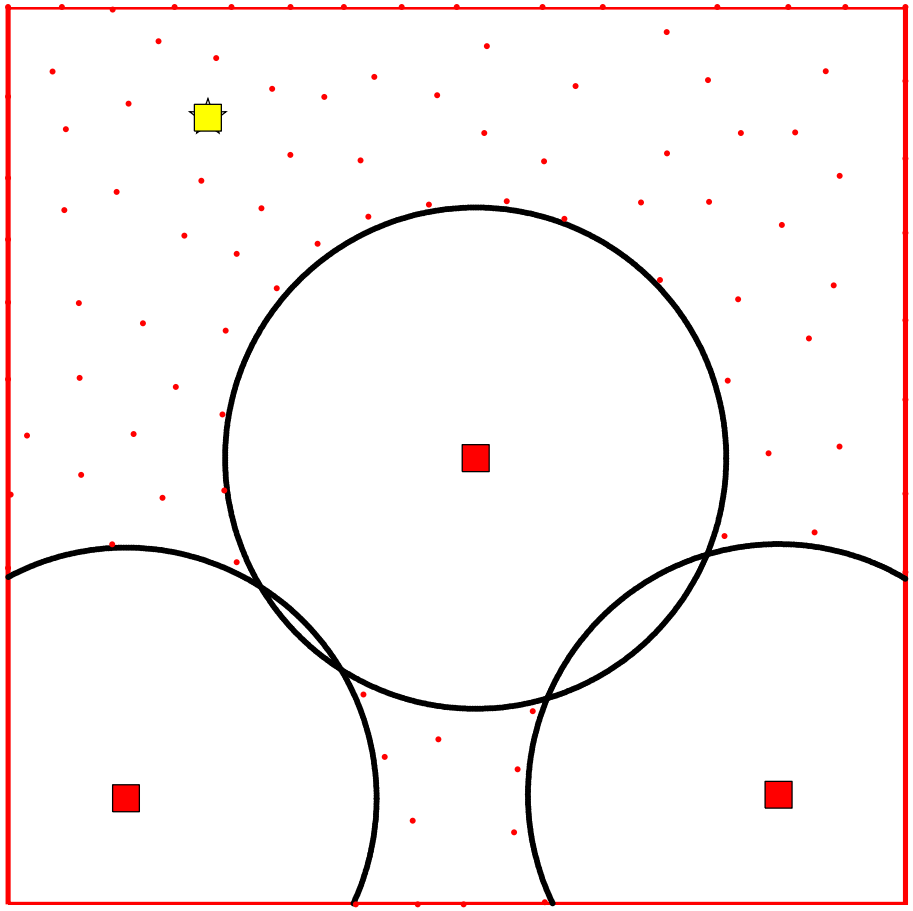


Centered L_2 discrepancies



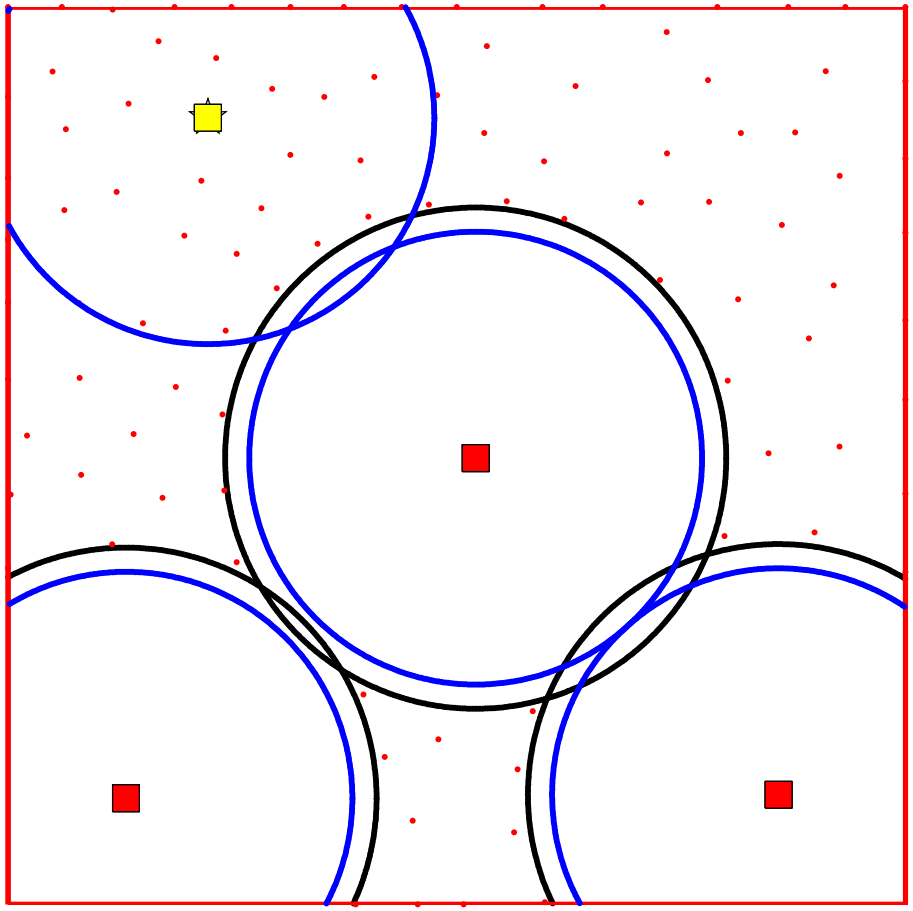
Large d ($d > 3$, say): we cannot use a regular grid $\mathscr{X}_Q$

→ adaptive grid with MCMC: illustration for $d = 2$ ($Q \approx n_{\max} d$)



Large d ($d > 3$, say): we cannot use a regular grid $\mathscr{X}_Q$

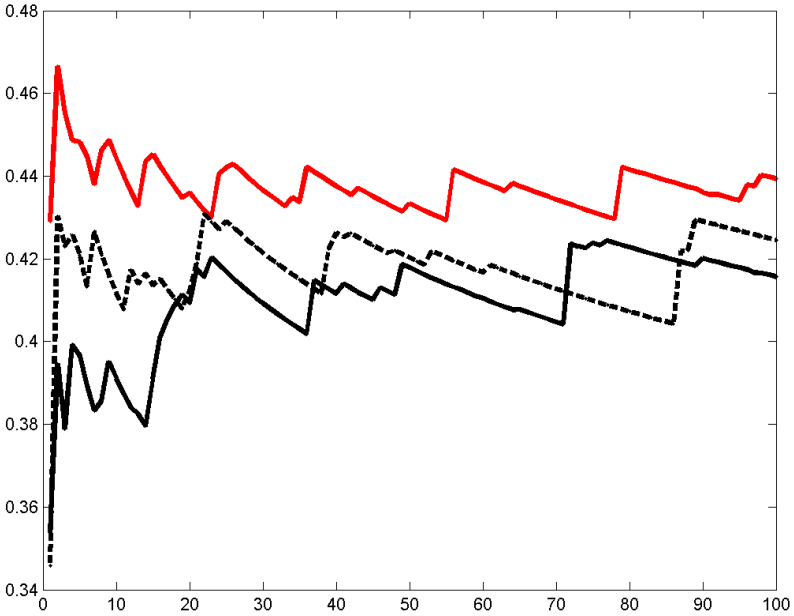
→ adaptive grid with MCMC: illustration for $d = 2$ ($Q \approx n_{\max} d$)



Ex: $\mathscr{X} = [0,1]^{10}$, $\mathscr{X}_Q =$ adaptive grid with $Q = 1000$ points
$n_{\min} = 30$, $n_{\max} = 100$, $q = 2$ in $\Psi_{b,B,q}(\cdot)$

➠ $\underline{\text{Eff}}_{mM}(X_n) = \dfrac{R_n^*}{\Phi_{mM}(X_n)}$ as a function of n

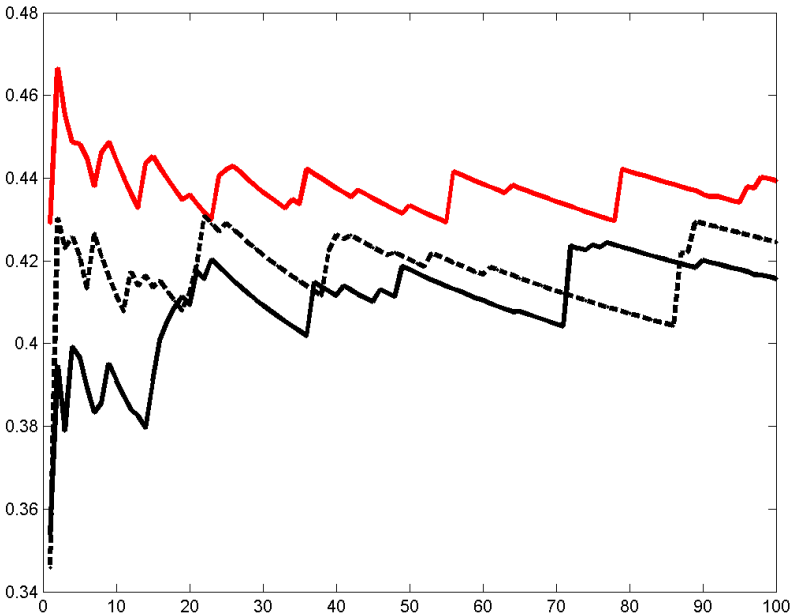
$\underline{\text{Eff}}_{mM}(X_n)$: $\Psi_{b,B,q}(\cdot)$ —,
Halton LDS —, Sobol' LDS - -



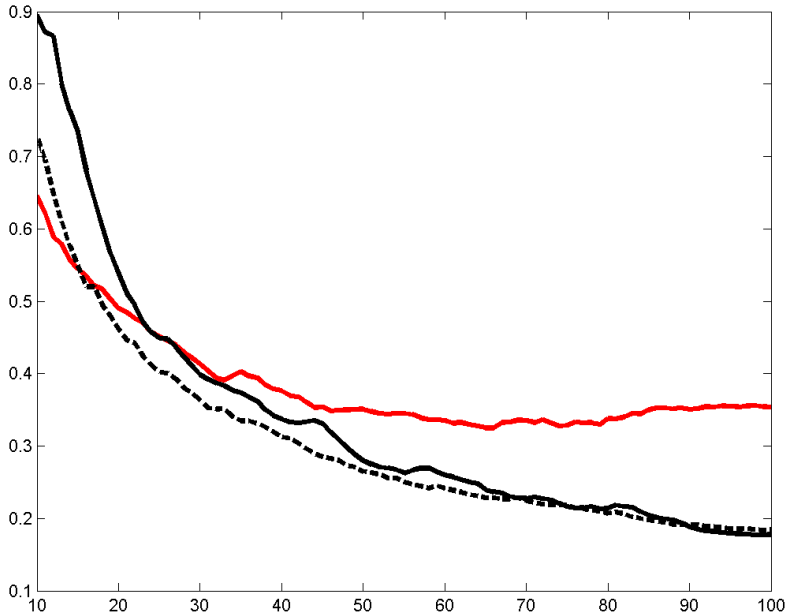
Ex: $\mathscr{X} = [0,1]^{10}$, $\mathscr{X}_Q =$ adaptive grid with $Q = 1000$ points
$n_{\min} = 30$, $n_{\max} = 100$, $q = 2$ in $\Psi_{b,B,q}(\cdot)$

➠ $\underline{\text{Eff}_{mM}}(X_n) = \dfrac{R_n^*}{\Phi_{mM}(X_n)}$ as a function of n

$\underline{\text{Eff}_{mM}}(X_n)$: $\Psi_{b,B,q}(\cdot)$ —,
Halton LDS —, Sobol' LDS - -



Centered L_2 discrepancies



B) L_q relaxation

Approximate $\mathscr{X}$ by $\mathscr{X}_Q$ with Q elements $\mathbf{s}_k$, $k = 1, \ldots, q$, $q > 0$, minimize

$$\Phi_{q,Q}(X_n) \triangleq \left[\frac{1}{Q} \sum_{k=1}^{Q} \left(\frac{1}{n} \sum_{i=1}^{n} \|\mathbf{s}_k - \mathbf{x}_i\|^{-q} \right)^{-1} \right]^{1/q}$$

B) L_q relaxation

Approximate $\mathscr{X}$ by $\mathscr{X}_Q$ with Q elements $\mathbf{s}_k$, $k = 1, \ldots, q$, $q > 0$, minimize

$$\Phi_{q,Q}(X_n) \triangleq \left[\frac{1}{Q} \sum_{k=1}^{Q} \left(\frac{1}{n} \sum_{i=1}^{n} \|\mathbf{s}_k - \mathbf{x}_i\|^{-q} \right)^{-1} \right]^{1/q}$$

For any X_n, $\Phi_{q,Q}(X_n) \to \Phi_{mM}(X_n; \mathscr{X}_Q)$, $q \to \infty$

where $\Phi_{mM}(X_n; \mathscr{X}_Q) = \max_{\mathbf{x} \in \mathscr{X}_Q} d(\mathbf{x}, X_n)$

Efficiency:

If $X^*_{n,q}$ minimizes $\Phi_{q,Q}(\cdot)$, then

$$\text{Eff}_{mM}(X^*_{n,q}; \mathscr{X}_Q) \geq (nQ)^{-1/q}$$

Efficiency:

If $X^*_{n,q}$ minimizes $\Phi_{q,Q}(\cdot)$, then

$$\text{Eff}_{mM}(X^*_{n,q}; \mathscr{X}_Q) \geq (nQ)^{-1/q}$$

- $\Phi_{q,Q}(\cdot)$ is non-increasing
- $\Psi(\cdot) = \frac{1}{n}\Phi^q_{q,Q}(\cdot)$ is supermodular

[ongoing joint work with João Rendas (CNRS, I3S, UCA) & Céline Helbert (École Centrale Lyon)]

4) Measures minimizing regularized dispersion

— joint work with Anatoly Zhigljavsky (LP & AZ, 2017)

For a n-point design, L_q relaxation:

$$\Phi_{q,Q}(X_n) \triangleq \left[\frac{1}{Q} \sum_{k=1}^{Q} \left(\frac{1}{n} \sum_{i=1}^{n} \|\mathbf{s}_k - \mathbf{x}_i\|^{-q} \right)^{-1} \right]^{1/q}, \quad q > 0$$

For a design measure ξ, integral version:

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

with μ uniform prob. measure on $\mathscr{X}$ $(\mu(\mathscr{X}) = 1)$

4) Measures minimizing regularized dispersion

— joint work with Anatoly Zhigljavsky (LP & AZ, 2017)

For a n-point design, L_q relaxation:

$$\Phi_{q,Q}(X_n) \triangleq \left[\frac{1}{Q} \sum_{k=1}^{Q} \left(\frac{1}{n} \sum_{i=1}^{n} \|\mathbf{s}_k - \mathbf{x}_i\|^{-q} \right)^{-1} \right]^{1/q}, \quad q > 0$$

For a design measure ξ, integral version:

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

with μ uniform prob. measure on $\mathscr{X}$ $(\mu(\mathscr{X}) = 1)$

Th 1: $\phi_q^q(\cdot)$, $q > 0$, is convex, and is strictly convex when $0 < q < d$

$\underline{q \geq d}$

- $\phi_q(\xi) > 0$ for any discrete measure ξ
- $\phi_q(\xi) = 0$ for any ξ equivalent to the Lebesgue measure on $\mathscr{X}$

. . . not very interesting

$q \geq d$

- $\phi_q(\xi) > 0$ for any discrete measure ξ
- $\phi_q(\xi) = 0$ for any ξ equivalent to the Lebesgue measure on $\mathscr{X}$

. . . not very interesting

$0 < q < d$

(Strict) convexity of $\phi_q^q(\cdot)$ ➞ "equivalence theorem"

Th 2: $\xi^{q,*}$ minimizes $\phi_q(\cdot)$ iff $\forall \mathbf{y} \in \mathscr{X}$, $d(\xi^{q,*}, \mathbf{y}) \leq \phi_q^q(\xi^{q,*})$

where $d(\xi, \mathbf{y}) = \int_{\mathscr{X}} \left\{ \|\mathbf{y} - \mathbf{x}\|^{-q} \left[\int_{\mathscr{X}} \|\mathbf{x} - \mathbf{z}\|^{-q} \xi(\mathrm{d}\mathbf{z}) \right]^{-2} \right\} \mu(\mathrm{d}\mathbf{x})$

$=$ directional derivative of $\phi_q^q(\cdot)$ at ξ in the direction of $\delta_{\mathbf{y}}$

$\xi^{q,*}$ is unique and $d(\xi^{q,*}, \mathbf{y}) = \phi_q^q(\xi^{q,*})$ for $\xi^{q,*}$-almost all $\mathbf{y} \in \mathscr{X}$

Two distinct situations

$0 < q \leq d - 2$

$\xi^{q,*}$ may be singular

Ex: $\mathscr{X} = \mathscr{B}_d(0,1)$; $\xi^{q,*} = \delta_0$ is optimal

Two distinct situations

$0 < q \leq d - 2$

$\xi^{q,*}$ may be singular

Ex: $\mathscr{X} = \mathscr{B}_d(0,1)$; $\xi^{q,*} = \delta_0$ is optimal

$\max\{0, d-2\} < q < d$

Th 3: $\xi^{q,*}$ does not possess atoms in the interior of $\mathscr{X}$

$\rightarrow$ Minimization of $\Phi_{q,Q}(X_n)$: take $q > d - 2$ to be space-filling

Construction of $\xi^{q,*}$?

Discretize $\mathscr{X}$ (again): replace μ by $\mu_Q = \frac{1}{Q}\sum_{k=1}^{Q} \delta_{\mathbf{s}_k}$ (grid or LDS)

$$\phi_q^q(\xi; \mu_Q) = \text{trace}[\mathbf{M}^{-1}(\xi)]$$

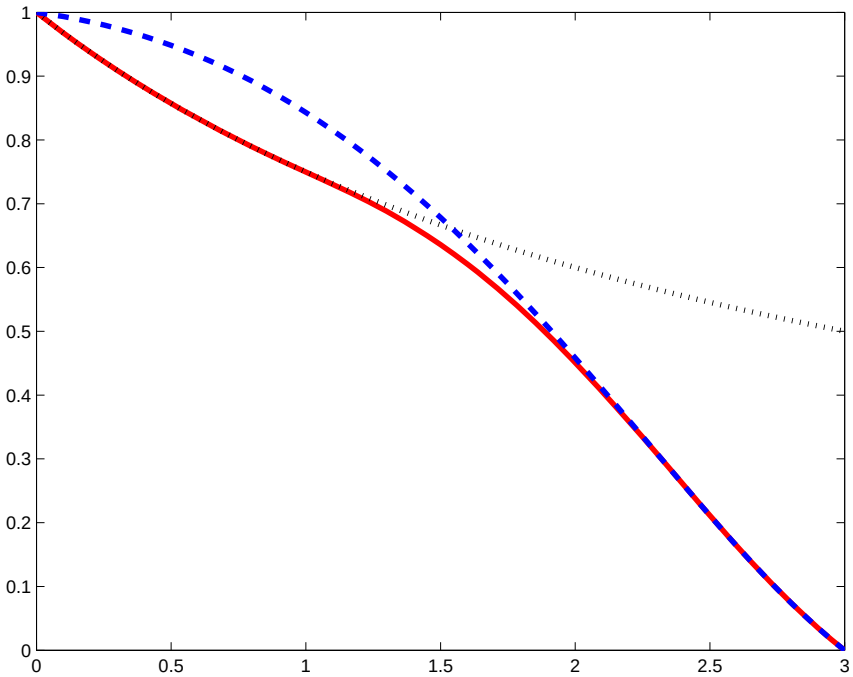
with $\mathbf{M}(\xi) = \int_{\mathscr{X}} \text{diag}\{Q\,\|\mathbf{x} - \mathbf{s}_k\|^{-q},\ k = 1, \ldots, Q\}\,\xi(\text{d}\mathbf{x})$ ($Q \times Q$-dimensional)

$\rightarrow$ an A-optimal design problem: multiplicative, or vertex-direction, algorithm

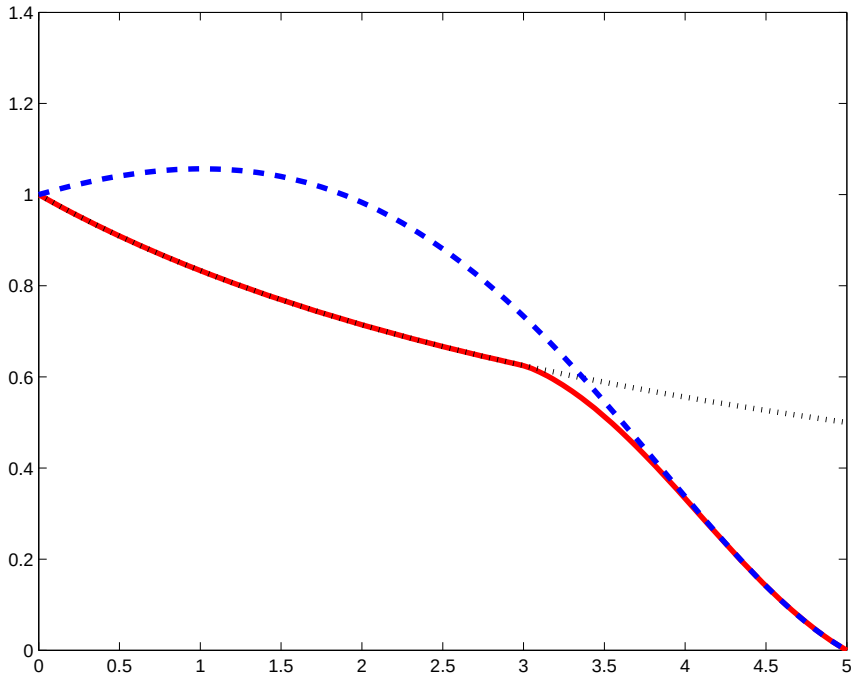
Ex: $\mathscr{X} = \mathscr{B}_d(0,1)$, make use of symmetry
(only consider distributions of the radii)

$\phi_q^q(\xi)$ function of q for $\xi = \delta_0$ (. . .), $\xi = \mu$ (- - -) and $\xi = \xi^{q,*}$ (—)

$d = 3$



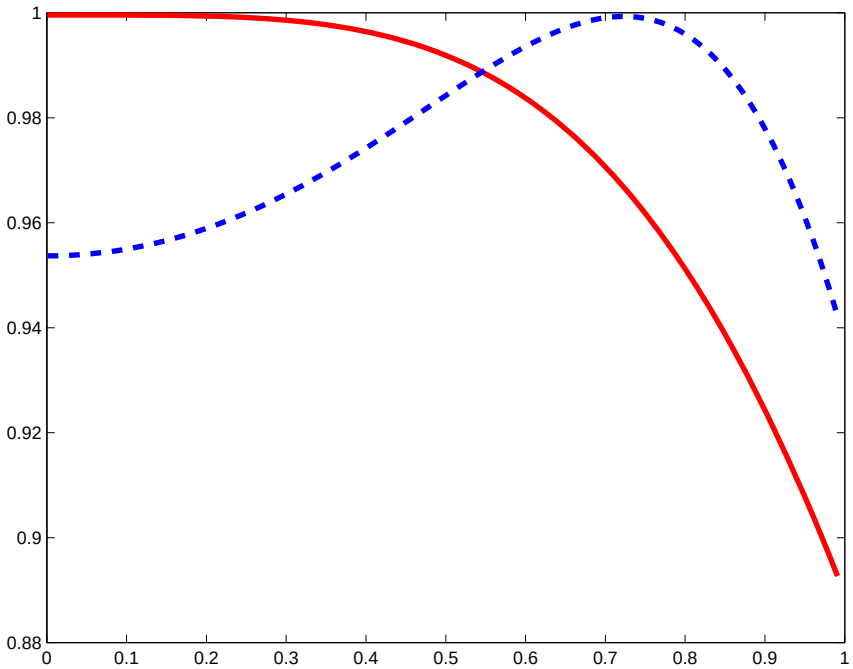
$d = 5$



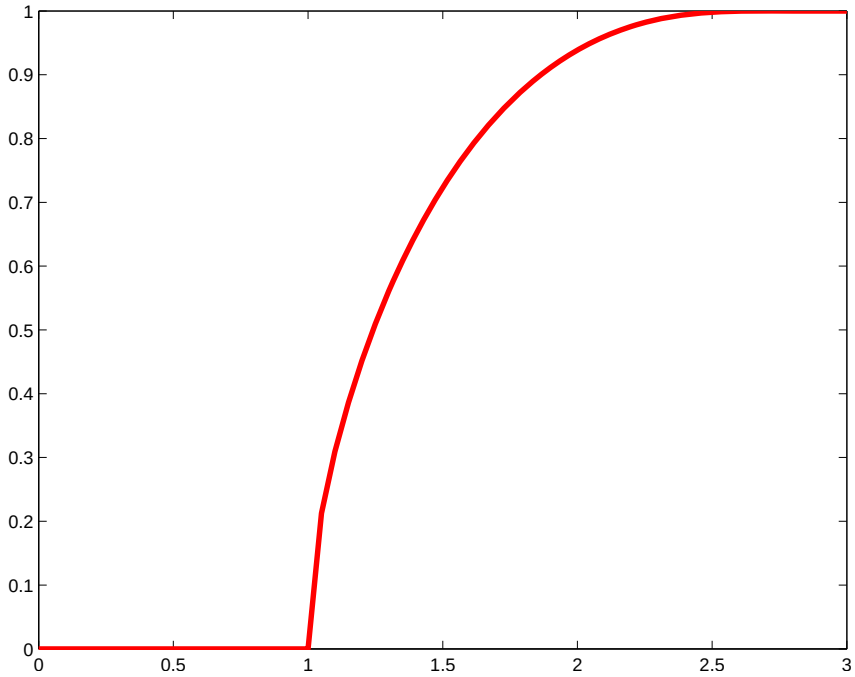
$\mu^{(r)}$ uniform on $\mathscr{B}_d(0, r)$, $d = 3$

efficiency $\frac{\phi_q^q(\xi^{q,*})}{\phi_q^q(\mu^{(r)})}$ of $\mu^{(r)}$ function of r

$q = 0.5$, $q = 1.5$

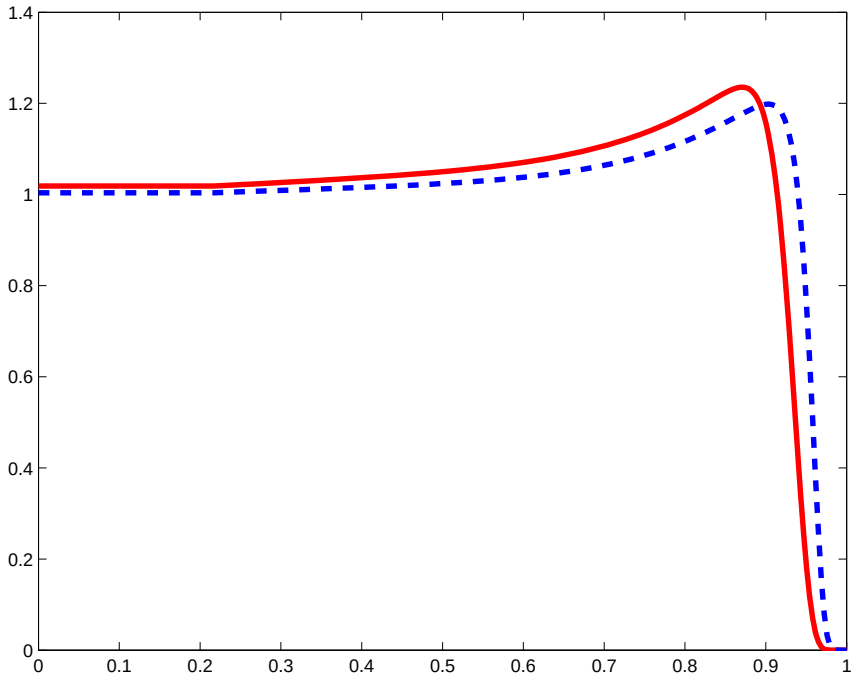


optimal r function of q

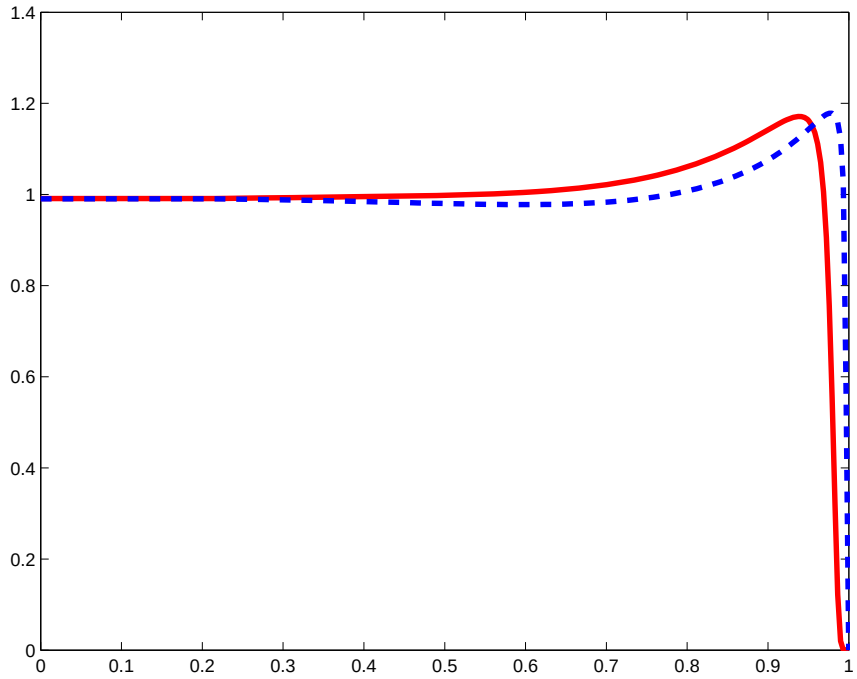


$d = 3$, optimal density of radii for $\xi^{q,*}$ (with respect to $\varphi(r) = dr^{d-1}$)

$q = 2$, $q = 2.1$

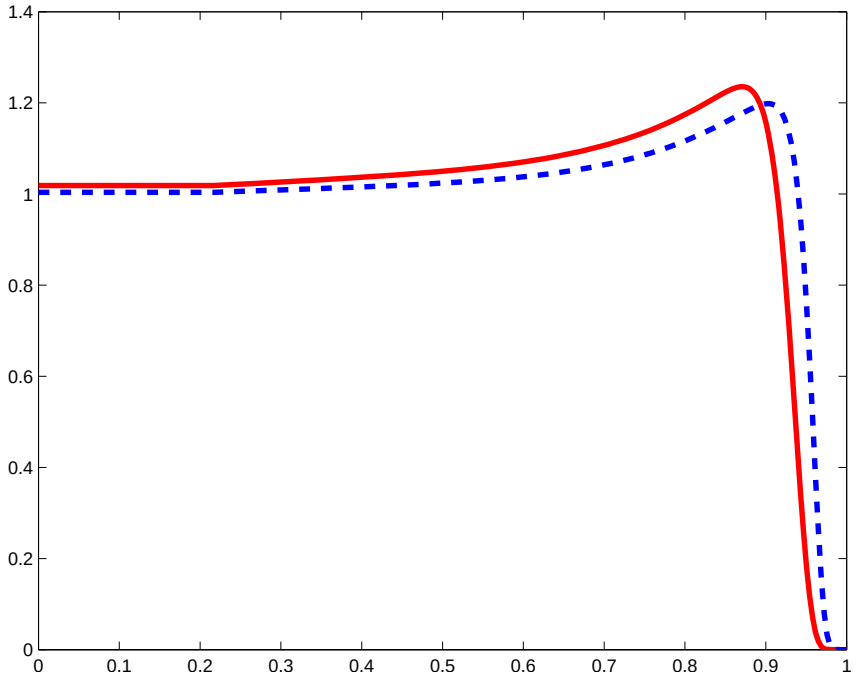


$q = 2.25$, $q = 2.5$

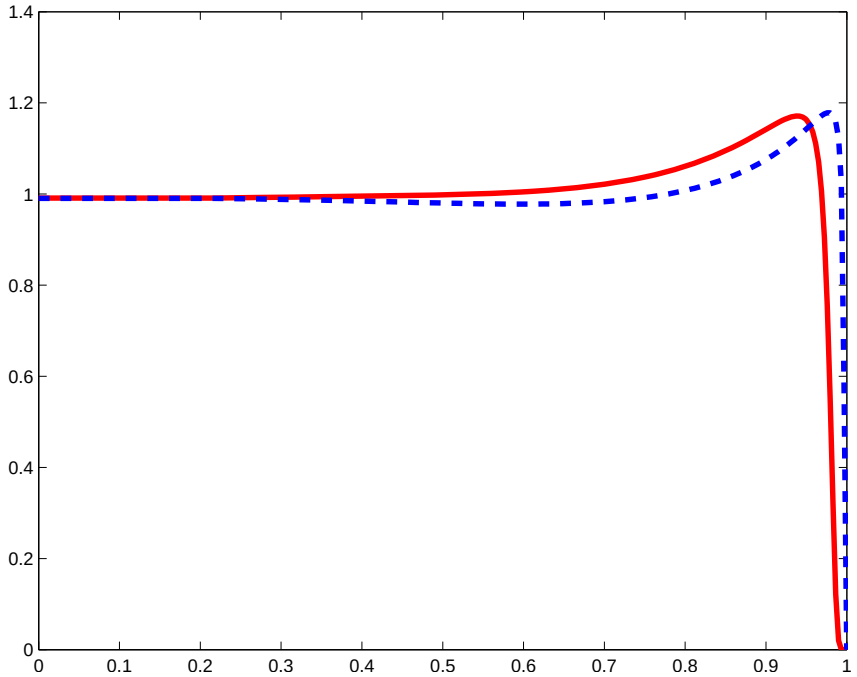


$d = 3$, optimal density of radii for $\xi^{q,*}$ (with respect to $\varphi(r) = dr^{d-1}$)

$q = 2$, $q = 2.1$



$q = 2.25$, $q = 2.5$



Minimization of $\Phi_{q,Q}(X_n)$:

take $q > d - 2$ to be space-filling,

no point near the border of $\mathscr{X}$

Conclusions

- Several methods to evaluate $\Phi_{mM}(X_n)$ (MCMC if $d \geq 5$)
- d small: optimization by a variant of Lloyd's method with Chebyshev centers (requires Voronoï tessellation or a fixed finite set approximation $\mathscr{X}_Q$)
- d large: optimization by a stochastic gradient
  (without any evaluation of $\Phi_{mM}(X_n)$)

Conclusions

- Several methods to evaluate $\Phi_{mM}(X_n)$ (MCMC if $d \geq 5$)
- d small: optimization by a variant of Lloyd's method with Chebyshev centers (requires Voronoï tessellation or a fixed finite set approximation $\mathscr{X}_Q$)
- d large: optimization by a stochastic gradient
  (without any evaluation of $\Phi_{mM}(X_n)$)
- Greedy methods based on submodular alternatives to dispersion can generate nested designs with reasonably good minimax efficiency (better than LDS, also without any evaluation of $\Phi_{mM}(X_n)$)

Conclusions

- Several methods to evaluate $\Phi_{mM}(X_n)$ (MCMC if $d \geq 5$)
- d small: optimization by a variant of Lloyd's method with Chebyshev centers (requires Voronoï tessellation or a fixed finite set approximation $\mathscr{X}_Q$)
- d large: optimization by a stochastic gradient
  (without any evaluation of $\Phi_{mM}(X_n)$)
- Greedy methods based on submodular alternatives to dispersion can generate nested designs with reasonably good minimax efficiency (better than LDS, also without any evaluation of $\Phi_{mM}(X_n)$)
  - Use an adaptive grid $\mathscr{X}_Q$ (MCMC) if d is large

Conclusions

- Several methods to evaluate $\Phi_{mM}(X_n)$ (MCMC if $d \geq 5$)
- d small: optimization by a variant of Lloyd's method with Chebyshev centers (requires Voronoï tessellation or a fixed finite set approximation $\mathscr{X}_Q$)
- d large: optimization by a stochastic gradient
  (without any evaluation of $\Phi_{mM}(X_n)$)
- Greedy methods based on submodular alternatives to dispersion can generate nested designs with reasonably good minimax efficiency (better than LDS, also without any evaluation of $\Phi_{mM}(X_n)$)
  - Use an adaptive grid $\mathscr{X}_Q$ (MCMC) if d is large
  - Consider projections on lower dimensional subspaces?

Conclusions

- Several methods to evaluate $\Phi_{mM}(X_n)$ (MCMC if $d \geq 5$)
- d small: optimization by a variant of Lloyd's method with Chebyshev centers (requires Voronoï tessellation or a fixed finite set approximation $\mathscr{X}_Q$)
- d large: optimization by a stochastic gradient
  (without any evaluation of $\Phi_{mM}(X_n)$)
- Greedy methods based on submodular alternatives to dispersion can generate nested designs with reasonably good minimax efficiency (better than LDS, also without any evaluation of $\Phi_{mM}(X_n)$)
  - Use an adaptive grid $\mathscr{X}_Q$ (MCMC) if d is large
  - Consider projections on lower dimensional subspaces?
  - Which submodular alternative is best?
- What about very large d ($d > 20$ say)? Random designs may be useful... (Janson, 1986, 1987)

References I

Auffray, Y., Barbillon, P., Marin, J.-M., 2012. Maximin design on non hypercube domains and kernel interpolation. Statistics and Computing 22 (3), 703–712.

Cardot, H., Cénac, P., Monnez, J.-M., 2012. A fast and recursive algorithm for clustering large datasets. Comput. Statist. Data Anal. 56 (6), 1434–1449.

Du, Q., Faber, V., Gunzburger, M., 1999. Centroidal Voronoi tessellations: applications and algorithms. SIAM Review 41 (4), 637–676.

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

Janson, S., 1986. Random coverings in several dimensions. Acta Mathematica 156 (1), 83–118.

Janson, S., 1987. Maximal spacings in several dimensions. The Annals of Probability 15 (1), 274–280.

Lekivetz, R., Jones, B., 2015. Fast flexible space-filling designs for nonrectangular regions. Quality and Reliability Engineering International 31 (5), 829–837.

Lloyd, S., 1982. Least squares quantization in PCM. IEEE Transactions on Information Theory 28 (2), 129–137.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. In: LeCam, L., Neyman, J. (Eds.), Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1. pp. 281–297.

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

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

Niederreiter, H., 1992. Random Number Generation and Quasi-Monte Carlo Methods. SIAM, Philadelphia.

Pronzato, L., 2017a. Minimax and maximin space-filling designs: some properties and methods for construction. Journal de la Société Française de Statistique 158 (1), 7–36.

Pronzato, L., 2017b. On the elimination of inessential points in the smallest enclosing ball problem. Optimization Methods and SoftwareTo appear.

Pronzato, L., Zhigljavsky, A., 2017. Measures minimizing regularized dispersion. Submitted.

Schaback, R., 1995. Error estimates and condition numbers for radial basis function interpolation. Advances in Computational Mathematics 3 (3), 251–264.

Vazquez, E., Bect, J., 2011. Sequential search based on kriging: convergence analysis of some algorithms. Proc. 58th World Statistics Congress of the ISI, August 21-26, Dublin, Ireland, arXiv preprint arXiv:1111.3866v1.

Yildirim, E., 2008. Two algorithms for the minimum enclosing ball problem. SIAM Journal on Optimization 19 (3), 1368–1391.