Design of Computer Experiments — (2) with model —

Luc Pronzato

Université Côte d'Azur, CNRS, I3S, France

Plan I

Objectives (same as part (1))

Computer experiments: based on simulations

➤ Usually, $\mathbf{x} \in \mathbb{R}^d$ ➠ observation $Y(\mathbf{x})$ (physical experiment)
➤ here, numerical simulation: $Y(\mathbf{x}) = f(\mathbf{x})$, observation = evaluation of an unknown function $f(\cdot)$
(no measurement error)

Objectives (same as part (1))

Computer experiments: based on simulations

➤ Usually, $\mathbf{x} \in \mathbb{R}^d$ ➠ observation $Y(\mathbf{x})$ (physical experiment)
➤ here, numerical simulation: $Y(\mathbf{x}) = f(\mathbf{x})$, observation = evaluation of an unknown function $f(\cdot)$
(no measurement error)

from pairs $(\mathbf{x}_i, f(\mathbf{x}_i))$, $i = 1, 2, \ldots, n$

- optimization: find $\mathbf{x}^* = \arg\max_{\mathbf{x} \in \mathscr{X}} f(\mathbf{x})$
- inversion: construct $\{\mathbf{x} \in \mathscr{X} : f(\mathbf{x}) = T\}$
- estimation of a probability of failure: $\text{Prob}\{f(\mathbf{x}) > C\}$ when $\mathbf{x} \sim$ probability density $\varphi(\cdot)$
- sensitivity analysis
- **approximation/interpolation of $f(\cdot)$** by a predictor $\eta_n(\cdot)$, to be constructed

1 Optimal design for Gaussian process models & kriging

1.1 Gaussian processes and kriging

Model for $f(\cdot)$: Gaussian process

$f(\mathbf{x}) = \mathbf{r}^\top(\mathbf{x})\beta + Z(\mathbf{x})$, with
$\mathbf{r}(\mathbf{x})$ a vector of known functions of $\mathbf{x}$ (the trend)
$Z(\mathbf{x})$ = realization of a random process (random field), second-order stationary, typically supposed to be Gaussian)
$\mathrm{E}\{Z(\mathbf{x})\} = 0$, $\mathrm{E}\{Z(\mathbf{x})Z(\mathbf{x}')\} = \sigma^2\, C(\mathbf{x} - \mathbf{x}'; \theta)$

1 Optimal design for Gaussian process models & kriging

1.1 Gaussian processes and kriging

Model for $f(\cdot)$: Gaussian process

$f(\mathbf{x}) = \mathbf{r}^\top(\mathbf{x})\beta + Z(\mathbf{x})$, with
$\mathbf{r}(\mathbf{x})$ a vector of known functions of $\mathbf{x}$ (the trend)
$Z(\mathbf{x})$ = realization of a random process (random field), second-order stationary, typically supposed to be Gaussian)
$\mathrm{E}\{Z(\mathbf{x})\} = 0$, $\mathrm{E}\{Z(\mathbf{x})Z(\mathbf{x}')\} = \sigma^2\, C(\mathbf{x} - \mathbf{x}'; \theta)$

Computer experiments

Following (Sacks et al., 1989), choose $C(\delta; \theta)$ continuous at $\delta = 0$, $C(0; \theta) = 1$

➠ 2 repetitions at the same $\mathbf{x}$ yield the same $f(\mathbf{x})$
(no measurement error)

Objective = interpolation (or extrapolation): build a predictor $\eta_n(\mathbf{x})$ based on a single realization of $Z(\cdot)$

much different from prediction of other realizations of $Z(\cdot)$ (➠ simply estimate β)

Objective = interpolation (or extrapolation): build a predictor $\eta_n(\mathbf{x})$ based on a single realization of $Z(\cdot)$
much different from prediction of other realizations of $Z(\cdot)$ (➠ simply estimate β)

ordinary kriging
(expression for universal kriging with trend $\mathbf{r}^\top(\mathbf{x})\beta$, $\beta \in \mathbb{R}^p$, $p > 1$, are slightly more complicated):
$f(\mathbf{x}) = \beta + Z(\mathbf{x}) \rightarrow \eta_n(\mathbf{x}) = \eta_n[f](\mathbf{x})$

BLUP (Best Linear Unbiased Predictor) at $\mathbf{x}$: $\boxed{\eta_n(\mathbf{x}) = \mathbf{v}_n^\top(\mathbf{x})\mathbf{y}_n}$ with

- $\mathbf{y}_n = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))^\top$
- $\mathbf{v}_n(x)$ minimizes $\mathrm{E}\{(\mathbf{v}_n^\top \mathbf{y}_n - [\beta + Z(\mathbf{x})])^2\}$
- with the constraint $\mathrm{E}\{\mathbf{v}_n^\top \mathbf{y}_n\} = \beta \sum_{i=1}^n \{\mathbf{v}_n\}_i = \mathrm{E}\{f(\mathbf{x})\} = \beta$, i.e., $\sum_{i=1}^n \{\mathbf{v}_n\}_i = 1$

Prediction: $\boxed{\eta_n(\mathbf{x}) = \hat{\beta}^n + \mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n\mathbf{1})}$

MSE (Mean-Squared Error) proportional to

$$\boxed{\rho_n(x) = \left(1 - \left[\mathbf{c}_n^\top(x)\ 1\right]\begin{bmatrix}\mathbf{C}_n & \mathbf{1}\\ \mathbf{1}^\top & 0\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{c}_n(x)\\ 1\end{bmatrix}\right)}$$

[with $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j);\theta)$, $\{\mathbf{c}_n(x)\}_i = C((X_i - x);\theta)$, $\hat{\beta}^n = (\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{y}_n)/(\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{1})$ (WLS) and $\mathbf{1} = (1,\ldots,1)^\top$]

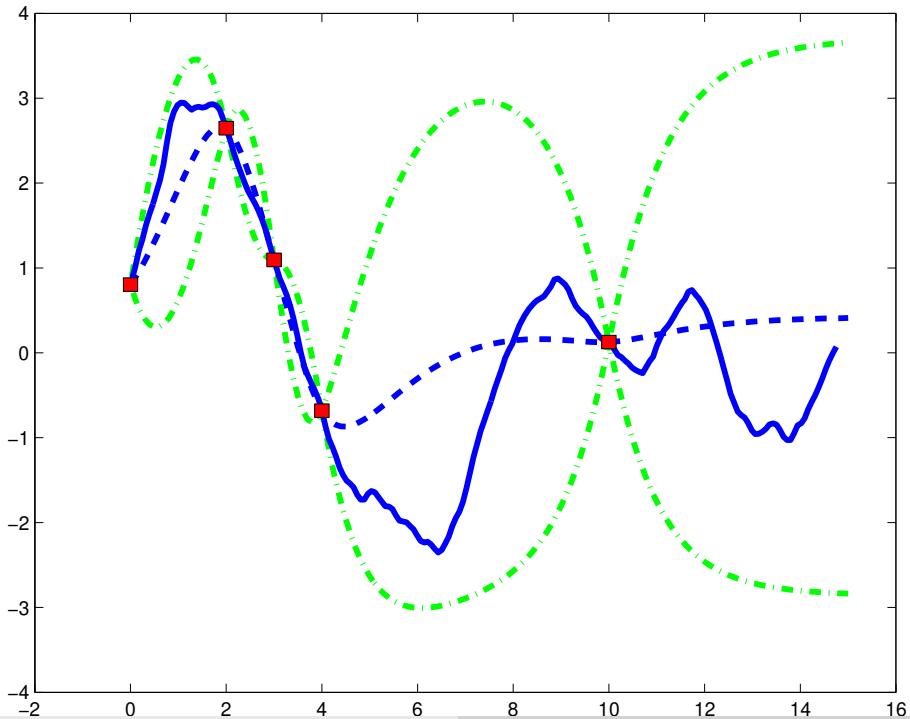
Prediction: $\eta_n(\mathbf{x}) = \hat{\beta}^n + \mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n\mathbf{1})$

MSE (Mean-Squared Error) proportional to

$$\rho_n(x) = \left(1 - \left[\mathbf{c}_n^\top(x)\ 1\right] \begin{bmatrix} \mathbf{C}_n & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_n(x) \\ 1 \end{bmatrix}\right)$$

[with $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j);\theta)$, $\{\mathbf{c}_n(x)\}_i = C((X_i - x);\theta)$, $\hat{\beta}^n = (\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{y}_n)/(\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{1})$ (WLS) and $\mathbf{1} = (1,\ldots,1)^\top$]

Ex. with $d = 1$, $n = 5$
(note that $\rho_n(\mathbf{x}_i) = 0$
— no measurement error)



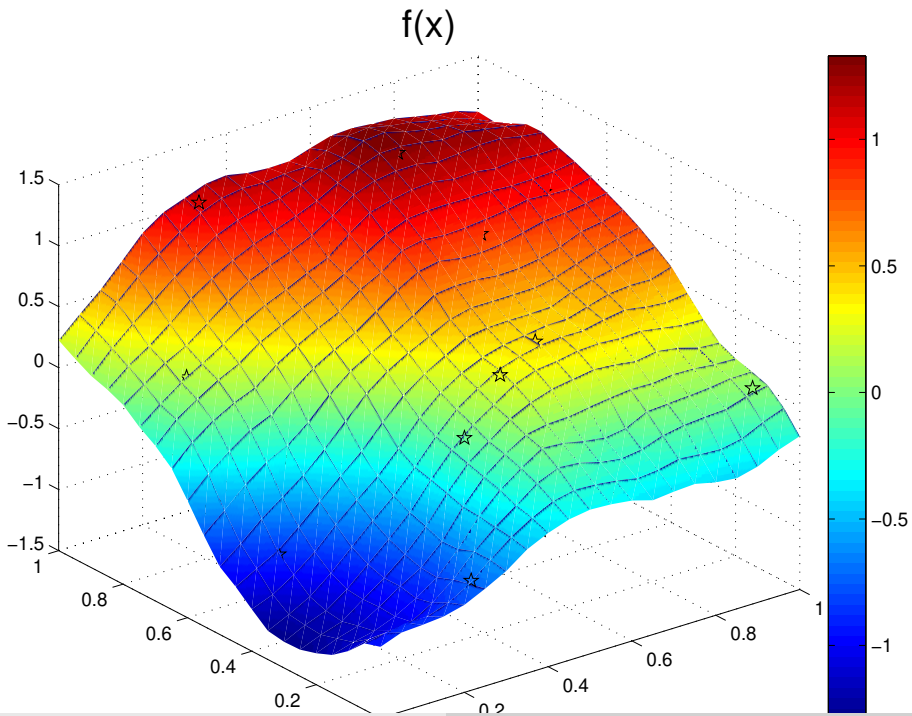
Prediction: $\boxed{\eta_n(\mathbf{x}) = \hat{\beta}^n + \mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n\mathbf{1})}$

MSE (Mean-Squared Error) proportional to

$$\rho_n(x) = \left(1 - \left[\mathbf{c}_n^\top(x)\ 1\right] \begin{bmatrix} \mathbf{C}_n & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_n(x) \\ 1 \end{bmatrix}\right)$$

[with $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j);\theta)$, $\{\mathbf{c}_n(x)\}_i = C((X_i - x);\theta)$, $\hat{\beta}^n = (\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{y}_n)/(\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{1})$ (WLS) and $\mathbf{1} = (1,\ldots,1)^\top$]

Ex. with $d = 2$, $n = 20$
($\mathbf{X}_n$ = random Lh)



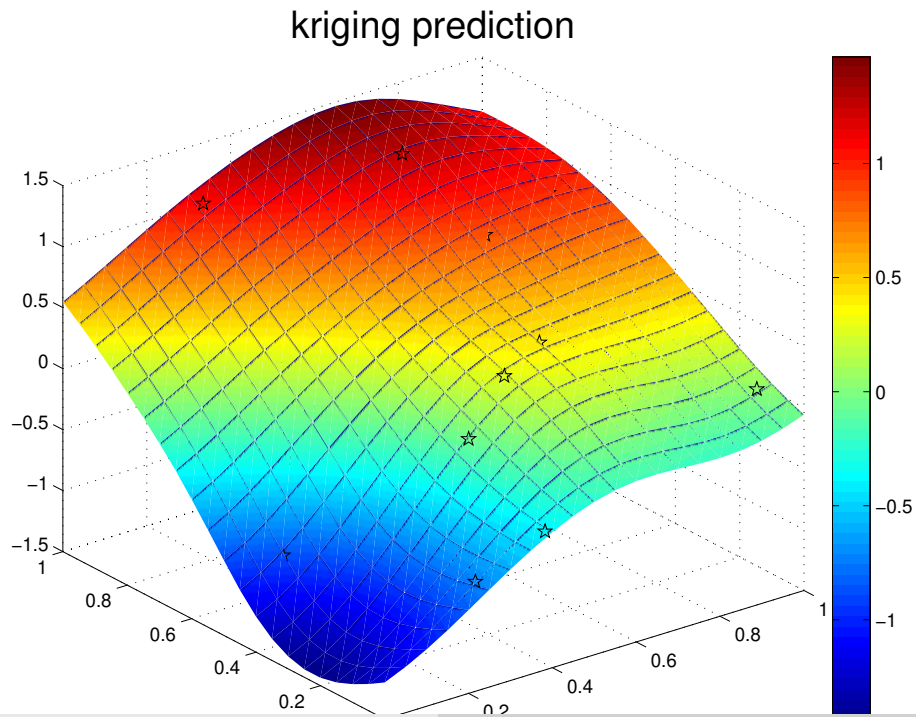
Prediction: $\boxed{\eta_n(\mathbf{x}) = \hat{\beta}^n + \mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n\mathbf{1})}$

MSE (Mean-Squared Error) proportional to

$$\rho_n(x) = \left(1 - \left[\mathbf{c}_n^\top(x)\ 1\right] \begin{bmatrix} \mathbf{C}_n & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_n(x) \\ 1 \end{bmatrix}\right)$$

[with $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j);\theta)$, $\{\mathbf{c}_n(x)\}_i = C((X_i - x);\theta)$, $\hat{\beta}^n = (\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{y}_n)/(\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{1})$ (WLS) and $\mathbf{1} = (1,\ldots,1)^\top$]

Ex. with $d = 2$, $n = 20$
($\mathbf{X}_n$ = random Lh)



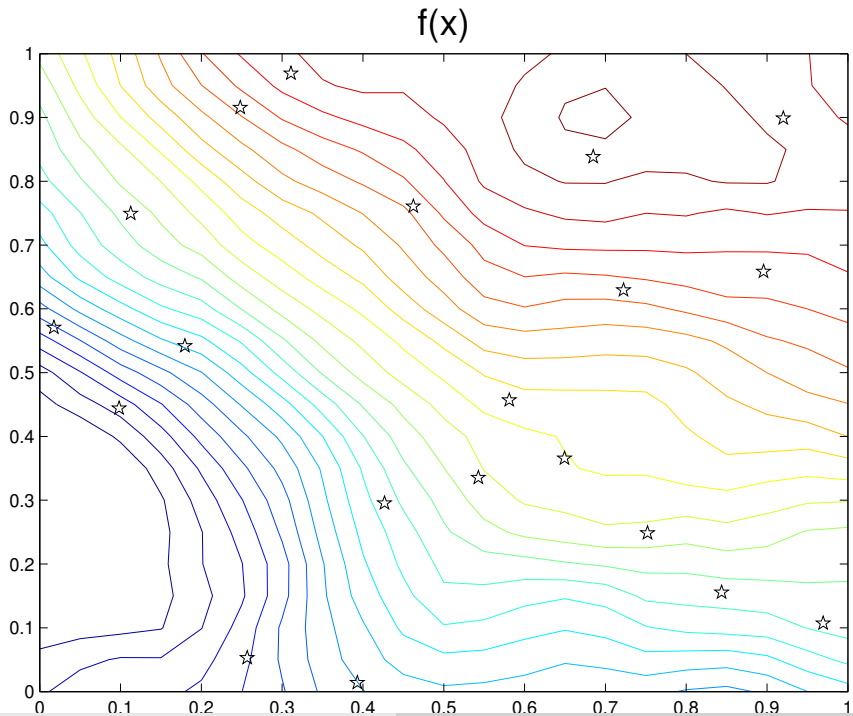
Prediction: $\boxed{\eta_n(\mathbf{x}) = \hat{\beta}^n + \mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n\mathbf{1})}$

MSE (Mean-Squared Error) proportional to

$$\boxed{\rho_n(x) = \left(1 - \left[\mathbf{c}_n^\top(x)\ 1\right] \begin{bmatrix} \mathbf{C}_n & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_n(x) \\ 1 \end{bmatrix}\right)}$$

[with $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j);\theta)$, $\{\mathbf{c}_n(x)\}_i = C((X_i - x);\theta)$, $\hat{\beta}^n = (\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{y}_n)/(\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{1})$ (WLS) and $\mathbf{1} = (1,\ldots,1)^\top$]

Ex. with $d = 2$, $n = 20$
($\mathbf{X}_n$ = random Lh)



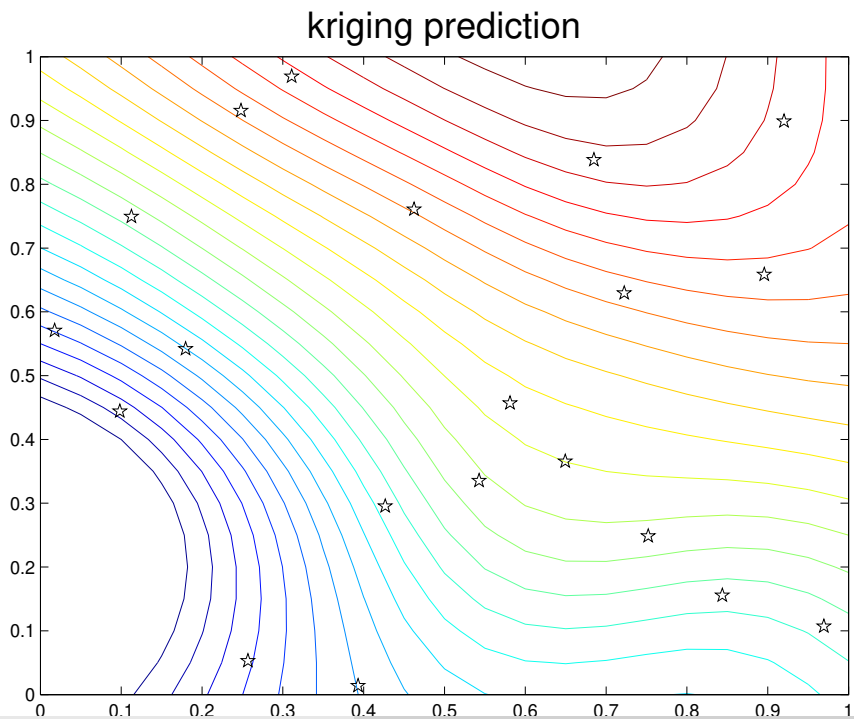
Prediction: $\boxed{\eta_n(\mathbf{x}) = \hat{\beta}^n + \mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}(\mathbf{y}_n - \hat{\beta}^n\mathbf{1})}$

MSE (Mean-Squared Error) proportional to

$$\rho_n(x) = \left(1 - \left[\mathbf{c}_n^\top(x)\ 1\right] \begin{bmatrix} \mathbf{C}_n & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_n(x) \\ 1 \end{bmatrix}\right)$$

[with $\{\mathbf{C}_n\}_{i,j} = C((X_i - X_j);\theta)$, $\{\mathbf{c}_n(x)\}_i = C((X_i - x);\theta)$, $\hat{\beta}^n = (\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{y}_n)/(\mathbf{1}^\top\mathbf{C}_n^{-1}\mathbf{1})$ (WLS) and $\mathbf{1} = (1,\ldots,1)^\top$]

Ex. with $d = 2$, $n = 20$
($\mathbf{X}_n$ = random Lh)



1.2 Criteria based on MSE

A natural idea: minimize $\rho_n(\mathbf{x})$ for all $\mathbf{x}$

In practice:

- minimize $\text{MMSE}(\mathbf{X}_n) = \max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})$
- minimize $\text{IMSE}(\mathbf{X}_n) = \int_{\mathscr{X}} \rho_n(\mathbf{x})\mathrm{d}\mu(\mathbf{x})$, with $\mu(\cdot)$ some measure of interest over $\mathscr{X}$

1.2 Criteria based on MSE

A natural idea: minimize $\rho_n(\mathbf{x})$ for all $\mathbf{x}$

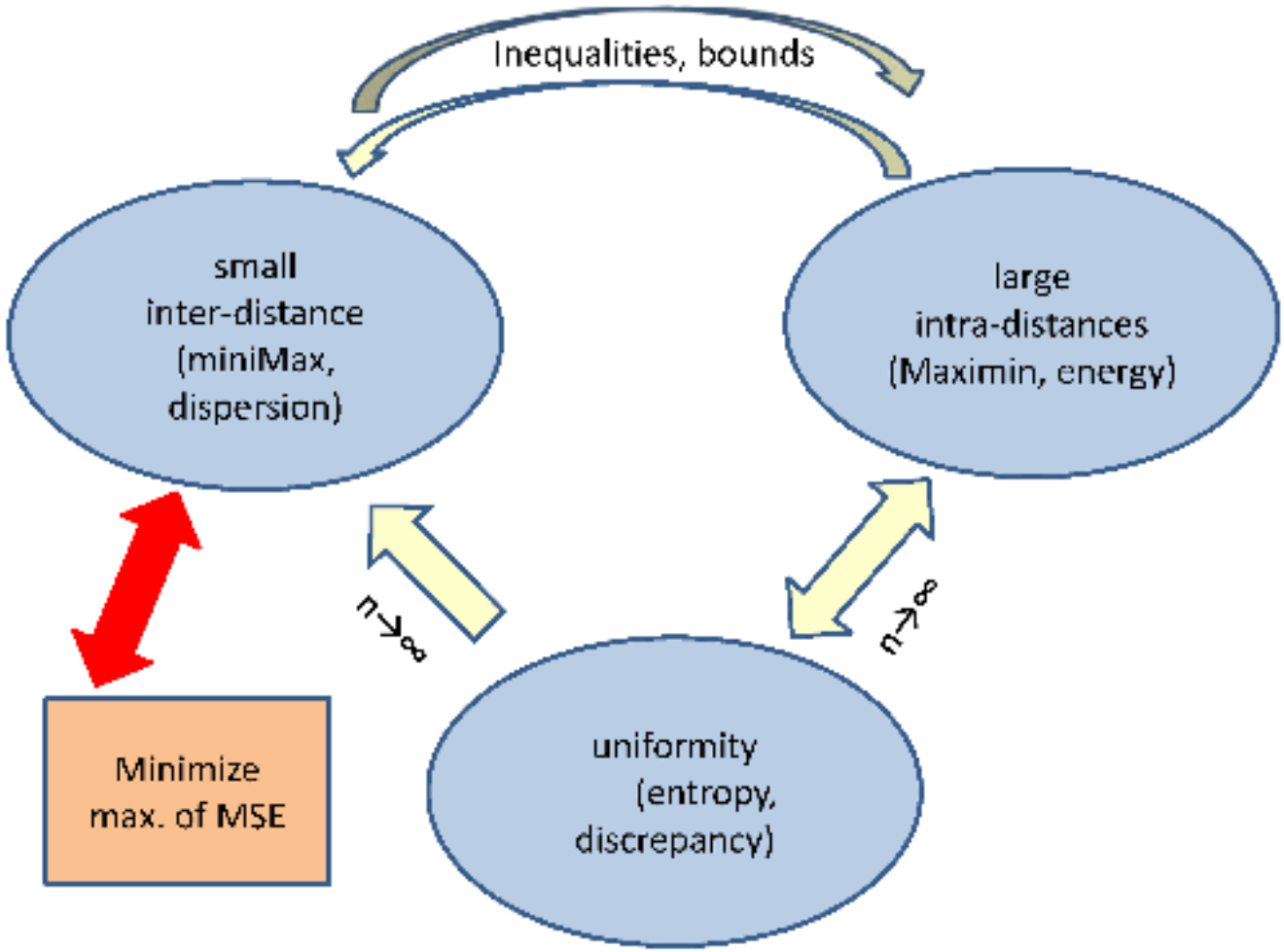
In practice:

- minimize $\text{MMSE}(\mathbf{X}_n) = \max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})$
- minimize $\text{IMSE}(\mathbf{X}_n) = \int_{\mathscr{X}} \rho_n(\mathbf{x})\mathrm{d}\mu(\mathbf{x})$, with $\mu(\cdot)$ some measure of interest over $\mathscr{X}$

Optimal designs are typically space-filling:

Johnson et al. (1990): if $C(\mathbf{x}-\mathbf{x}') = c(\|\mathbf{x}-\mathbf{x}'\|)$ with $c(\cdot)$ decreasing, then $\mathbf{X}_n^*$ optimal for $\Phi_{mM}(\cdot)$ (miniMax optimal) tends to be optimal for $\text{MMSE}(\mathbf{X}_n)$ with covariance $C_a(\mathbf{x}-\mathbf{x}') = [C(\mathbf{x}-\mathbf{x}')]^a$ when $a \to \infty$

➠ no point $\mathbf{x}_i$ on the boundary of $\mathscr{X}$


Inequalities, bounds
small
inter-distance
(miniMax,
dispersion)
large
intra-distances
(Maximin, energy)
$n \to \infty$
$n \to \infty$
Minimize
max. of MSE
uniformity
(entropy,
discrepancy)

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

Calculation of MMSE($\mathbf{X}_n$):

Compute $\rho_n(\mathbf{x}^{(k)})$ for a finite Q-points set $\mathscr{X}_Q = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(Q)}\}$
(e.g., first Q point of a LDS in $\mathscr{X}$),
then MMSE($\mathbf{X}_n$) $\simeq \max_k \rho_n(\mathbf{x}^{(k)})$,
to be minimized, for instance by simulated annealing

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

Calculation of MMSE($\mathbf{X}_n$):

Compute $\rho_n(\mathbf{x}^{(k)})$ for a finite Q-points set $\mathscr{X}_Q = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(Q)}\}$
(e.g., first Q point of a LDS in $\mathscr{X}$),
then MMSE($\mathbf{X}_n$) $\simeq \max_k \rho_n(\mathbf{x}^{(k)})$,
to be minimized, for instance by simulated annealing

Calculation of IMSE($\mathbf{X}_n$) $= \int_{\mathscr{X}} \rho_n(\mathbf{x}) \mathrm{d}\mu(\mathbf{x})$ (Gauthier and P. 2014, 2016) :

Without trend ($\mathbf{r}(\mathbf{x}) = \mathbf{0}\ \forall \mathbf{x}$) ➠ $\rho_n(\mathbf{x}) = 1 - \mathbf{c}_n^\top(\mathbf{x}) \mathbf{C}_n^{-1} \mathbf{c}_n(\mathbf{x})$,
where $\{\mathbf{c}_n(\mathbf{x})\}_i = C(\mathbf{x} - \mathbf{x}_i)$, $\{\mathbf{C}_n\}_{ij} = C(\mathbf{x}_i - \mathbf{x}_j)$

$$\begin{aligned} \text{IMSE}(\mathbf{X}_n) &= 1 - \text{trace}\left[\mathbf{C}_n^{-1} \int_{\mathscr{X}} \mathbf{c}_n(\mathbf{x}) \mathbf{c}_n^\top(\mathbf{x}) \mathrm{d}\mu(\mathbf{x})\right] \\ &= 1 - \text{trace}\left[\mathbf{C}_n^{-1} \Sigma_n\right] \end{aligned}$$

Calculation for a finite Q-points set $\mathscr{X}_Q = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(Q)}\}$:

$$\begin{aligned} \text{IMSE}(\mathbf{X}_n) \simeq \widehat{\text{IMSE}}(\mathbf{X}_n) &= \sum_{k=1}^{Q} w_k \, \rho_n(\mathbf{x}^{(k)}) \\ &= 1 - \text{trace}\left[\mathbf{C}_n^{-1} \widehat{\Sigma}_n\right] \end{aligned}$$

with $\sum_{k=1}^{Q} w_k = 1$ ($w_k = 1/Q$ when μ if uniform)
and $\widehat{\Sigma}_n = \sum_{k=1}^{Q} w_k \mathbf{c}_n(\mathbf{x}^{(k)}) \mathbf{c}_n^\top(\mathbf{x}^{(k)})$

Calculation for a finite Q-points set $\mathscr{X}_Q = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(Q)}\}$:

$$\begin{aligned} \text{IMSE}(\mathbf{X}_n) \simeq \widehat{\text{IMSE}}(\mathbf{X}_n) &= \sum_{k=1}^{Q} w_k \, \rho_n(\mathbf{x}^{(k)}) \\ &= 1 - \text{trace}\left[\mathbf{C}_n^{-1} \widehat{\Sigma}_n\right] \end{aligned}$$

with $\sum_{k=1}^{Q} w_k = 1$ ($w_k = 1/Q$ when μ if uniform)
and $\widehat{\Sigma}_n = \sum_{k=1}^{Q} w_k \mathbf{c}_n(\mathbf{x}^{(k)}) \mathbf{c}_n^\top(\mathbf{x}^{(k)})$

If, moreover, $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathscr{X}_Q$, with $\mathbf{x}_i = \mathbf{x}^{(k_i)}$, $i = 1, \ldots, n$,

then $\widehat{\Sigma}_n = \{\mathbf{QWQ}\}_{\mathbb{J}_n \mathbb{J}_n}$
with $\{\mathbf{Q}\}_{k\ell} = C(\mathbf{x}^{(k)} - \mathbf{x}^{(\ell)})$, $\mathbf{W} = \text{diag}\{w_1, \ldots, w_Q\}$
and $\mathbb{J}_n = \{k_1, \ldots, k_n\}$

➠ $\text{IMSE}(\mathbf{X}_n) \simeq 1 - \text{trace}\left[\mathbf{Q}_{\mathbb{J}_n \mathbb{J}_n}{}^{-1} \{\mathbf{QWQ}\}_{\mathbb{J}_n \mathbb{J}_n}\right]$
not expensive to compute once $\mathbf{Q}$ and $\mathbf{QWQ}$ have been calculated

(a bit more complicated with a trend $\mathbf{r}^\top(\mathbf{x})\beta$)

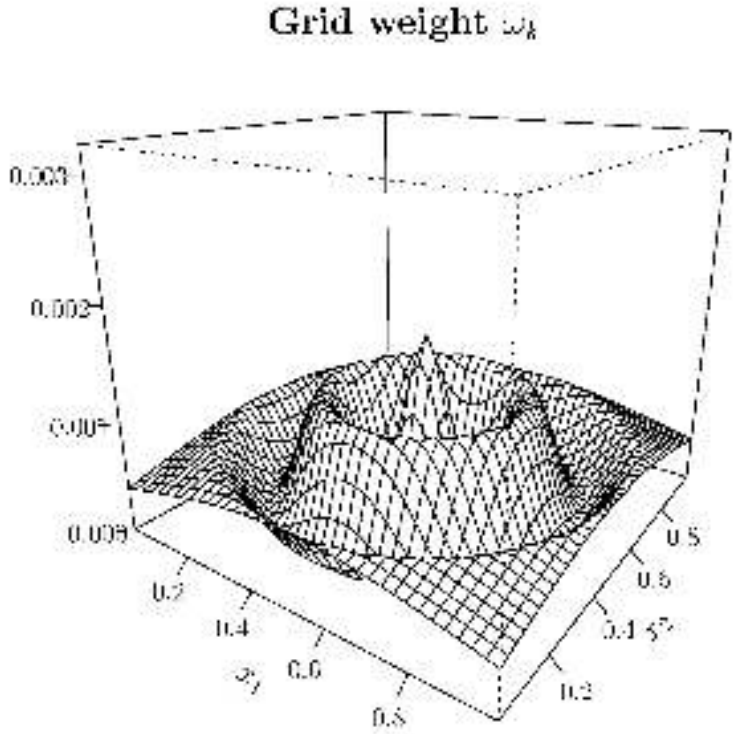
Minimization not obvious (for instance, by simulated annealing),
see § 3.3 for another approach

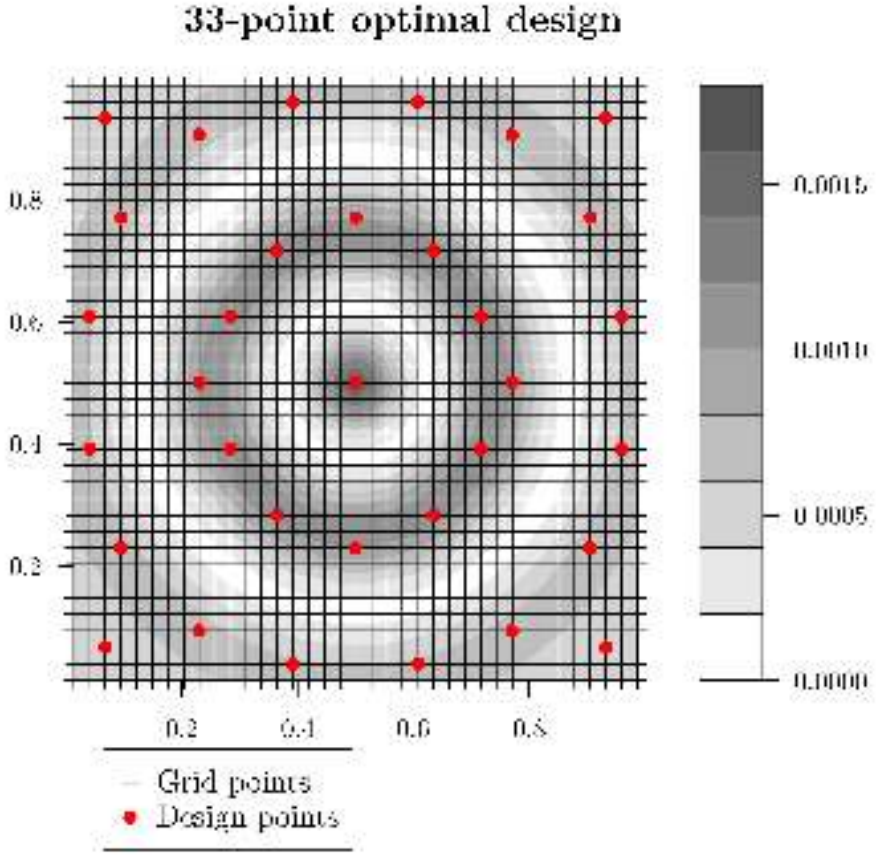
Ex. of IMSE-optimal design (Gauthier & P., 2014, 2016):

$\mathscr{X}$ = regular grid with $37^2 = 1\,369$ points

$C(\mathbf{x} - \mathbf{x}') = C_1(\{\mathbf{x}\}_1 - \{\mathbf{x}'\}_1) \times C_2(\{\mathbf{x}\}_2 - \{\mathbf{x}'\}_2)$,

$C_i(x - x') = (1 + 25/\sqrt{3}|x - x'|) \exp[-25/\sqrt{3}|x - x'|]$ (Matérn 3/2)





(measure of interest μ)

Sequential construction of an optimal design:

For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\boxed{\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})}$$

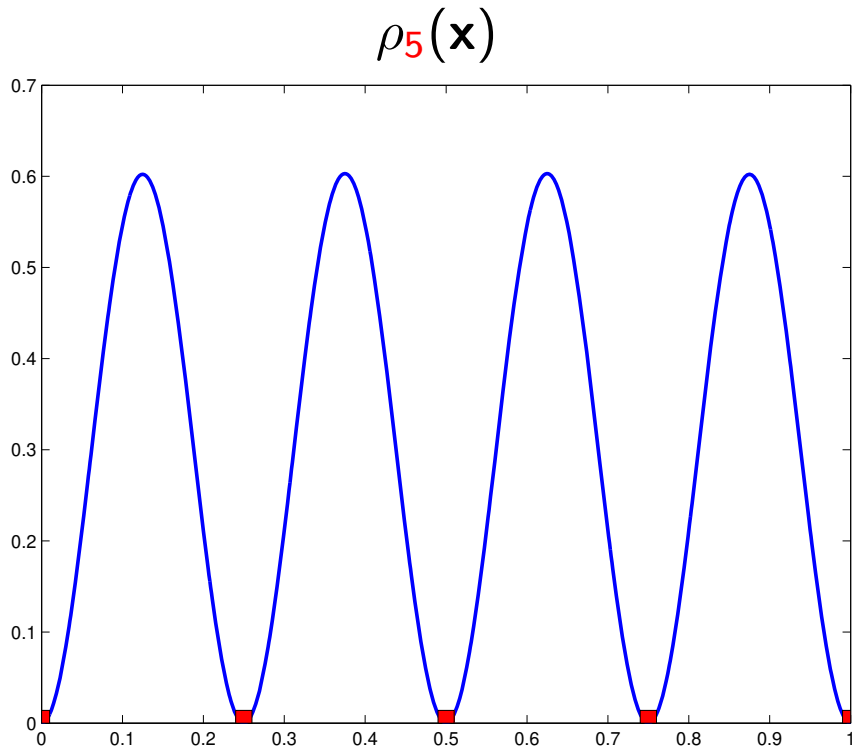
Sequential construction of an optimal design:

For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\boxed{\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})}$$

For MMSE : do not choose $\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$!

➠ take instead $\boxed{\mathbf{x}^*_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$



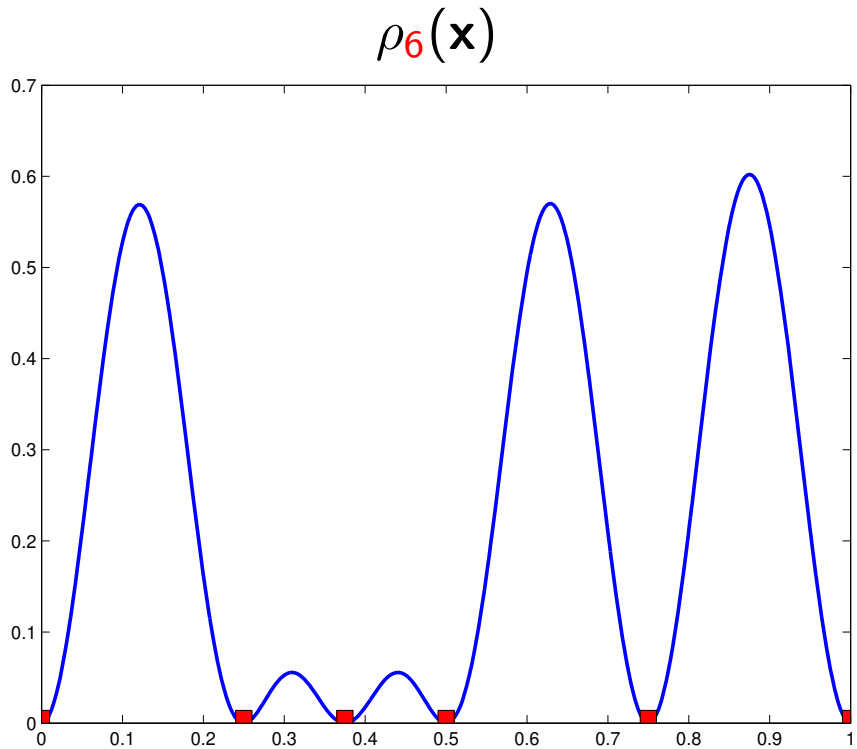
Sequential construction of an optimal design:

For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\boxed{\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})}$$

For MMSE : do not choose $\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$!

➠ take instead $\boxed{\mathbf{x}^*_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$



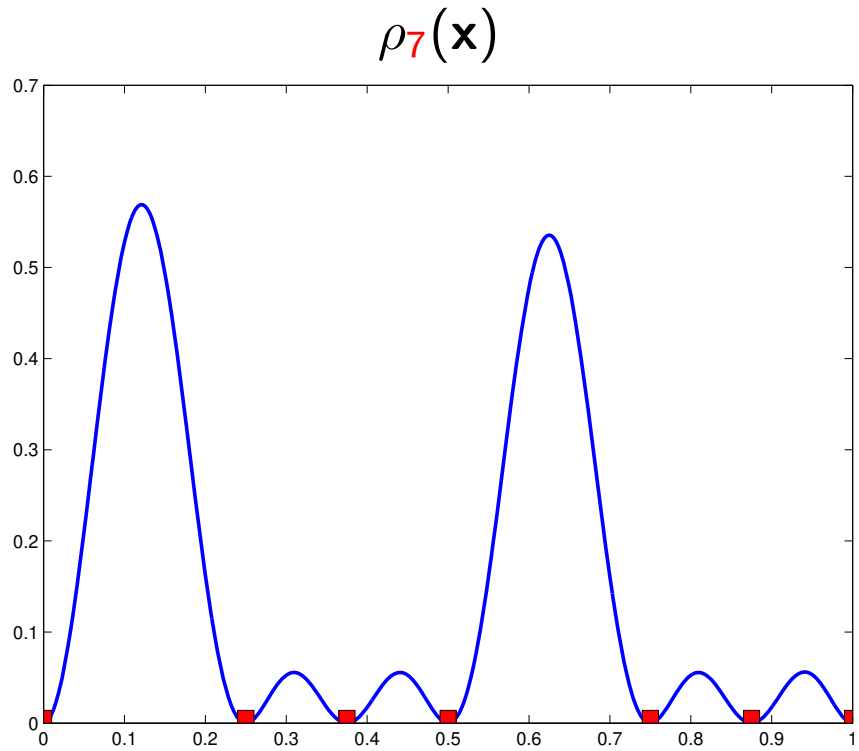
Sequential construction of an optimal design:

For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\boxed{\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})}$$

For MMSE : do not choose $\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$!

➠ take instead $\boxed{\mathbf{x}^*_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$



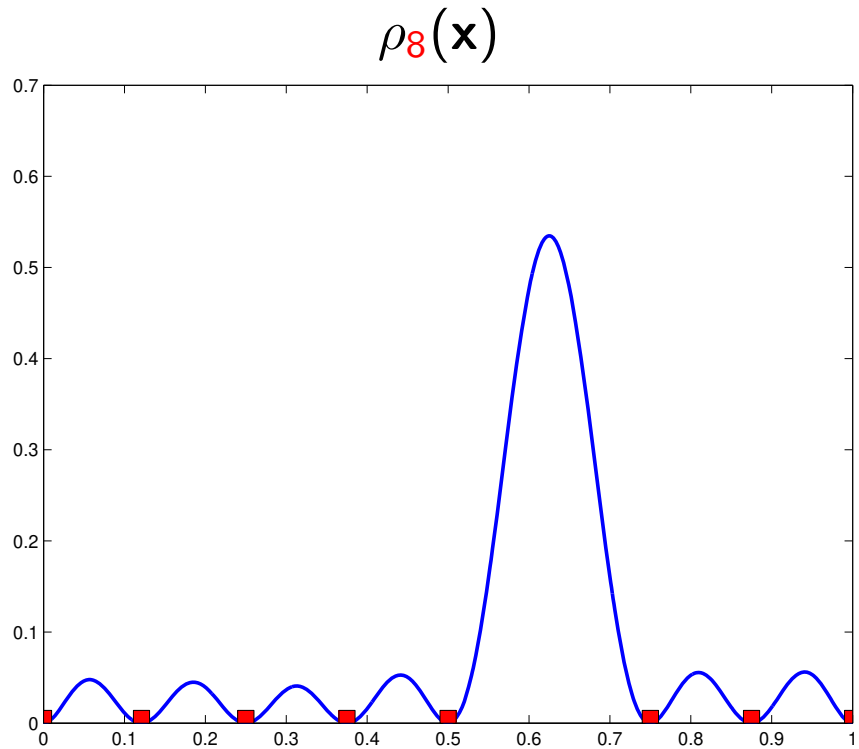
Sequential construction of an optimal design:

For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\boxed{\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})}$$

For MMSE : do not choose $\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$!

➠ take instead $\boxed{\mathbf{x}^*_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$



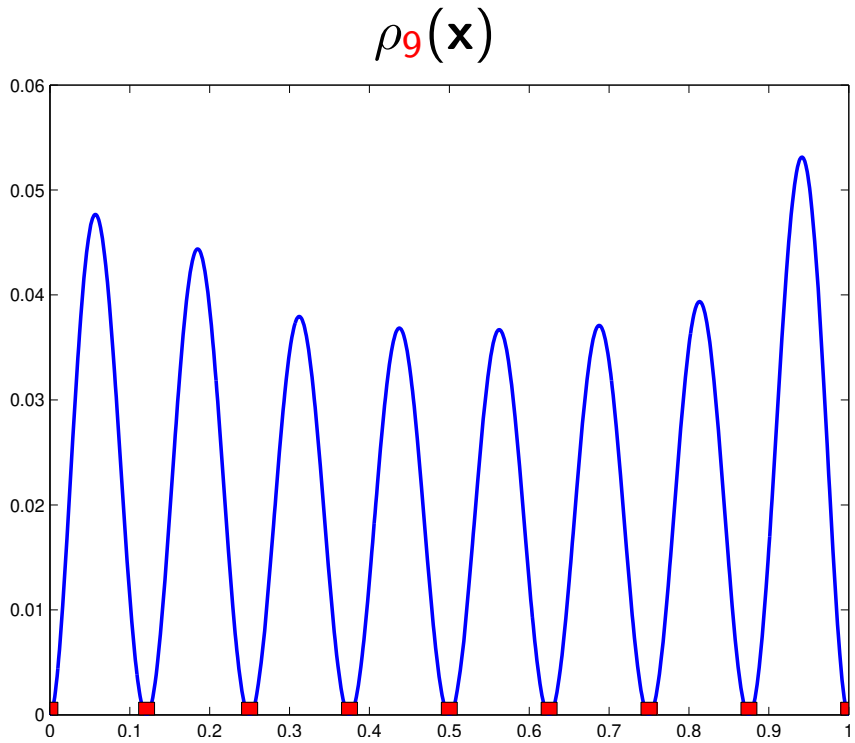
Sequential construction of an optimal design:

For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\boxed{\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})}$$

For MMSE : do not choose $\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$!

➠ take instead $\boxed{\mathbf{x}^*_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$



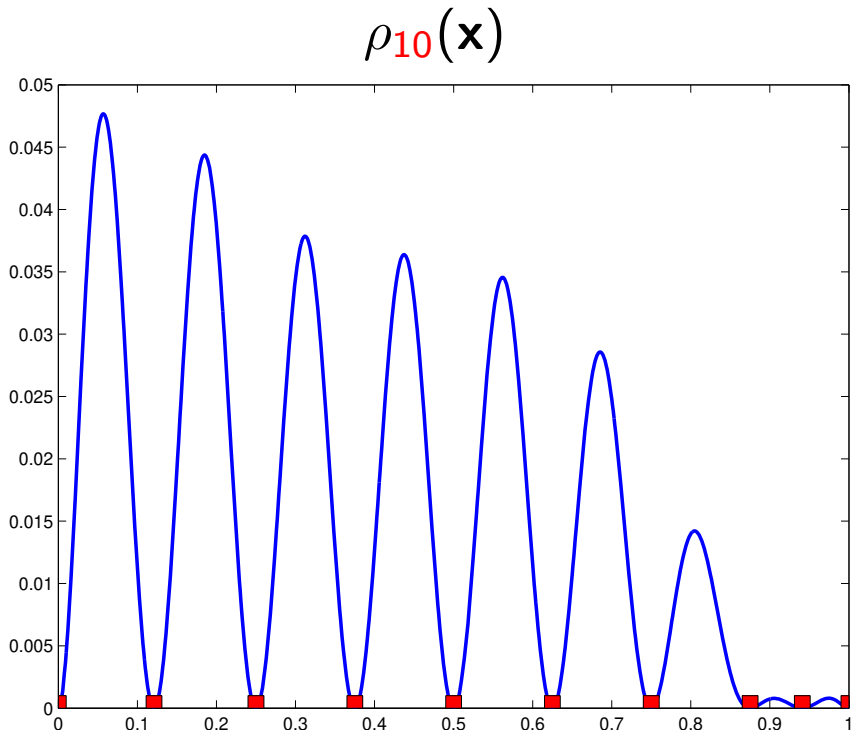
Sequential construction of an optimal design:

For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\boxed{\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})}$$

For MMSE : do not choose $\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$!

➠ take instead $\boxed{\mathbf{x}^*_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$



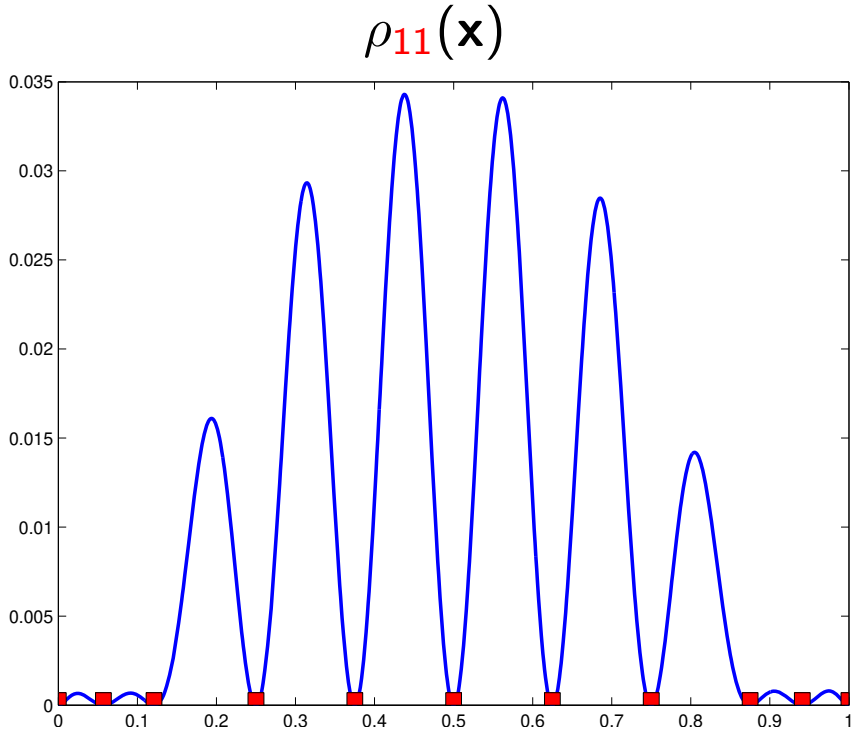
Sequential construction of an optimal design:

For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\boxed{\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})}$$

For MMSE : do not choose $\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$!

➠ take instead $\boxed{\mathbf{x}^*_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$



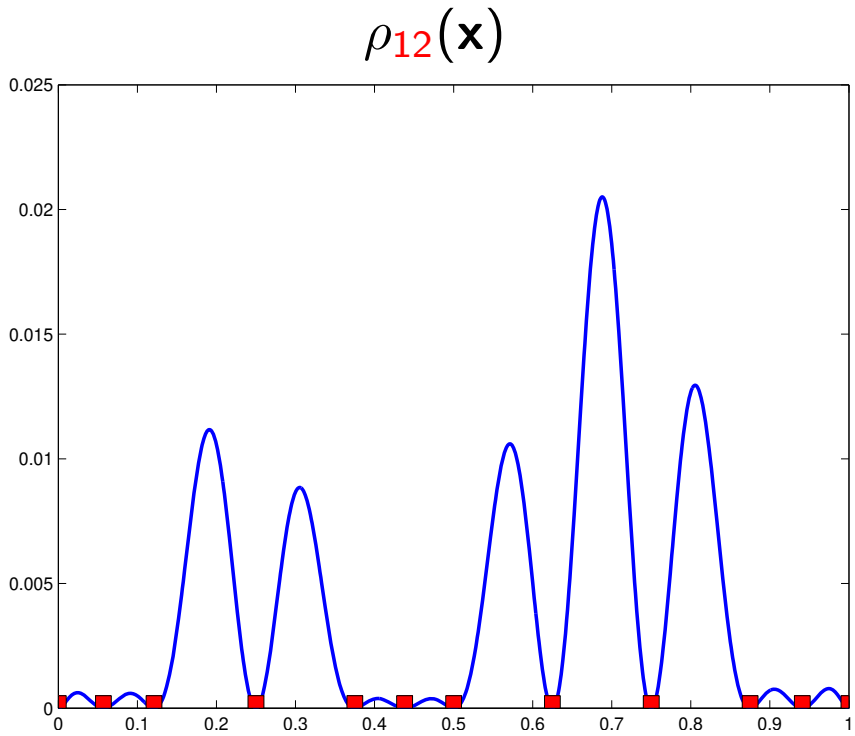
Sequential construction of an optimal design:

For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\boxed{\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})}$$

For MMSE : do not choose $\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$!

➠ take instead $\boxed{\mathbf{x}^*_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$



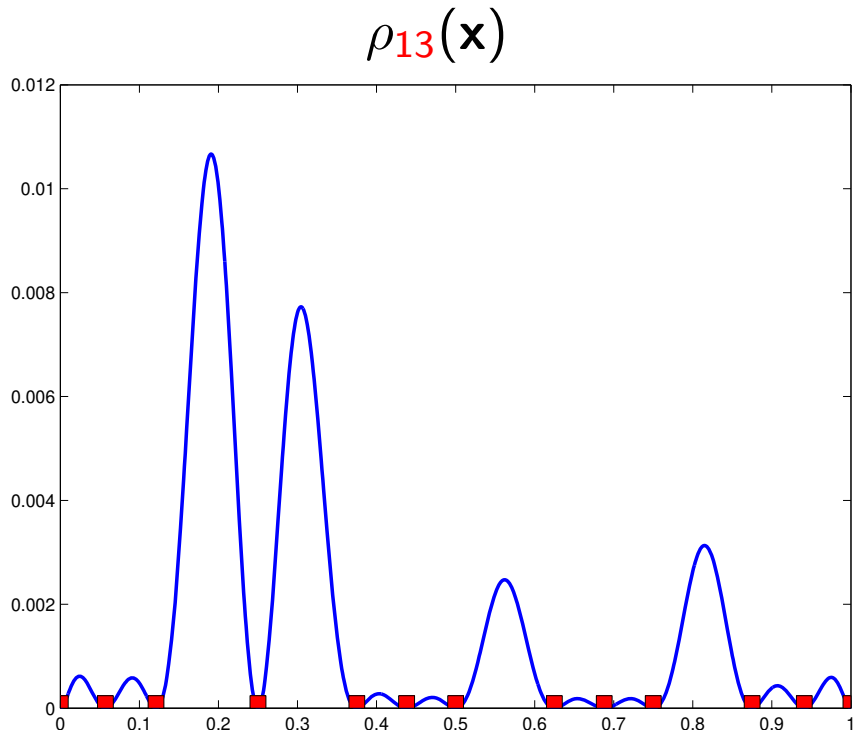
Sequential construction of an optimal design:

For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\boxed{\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})}$$

For MMSE : do not choose $\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$!

➠ take instead $\boxed{\mathbf{x}^*_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$



Sequential construction of an optimal design:

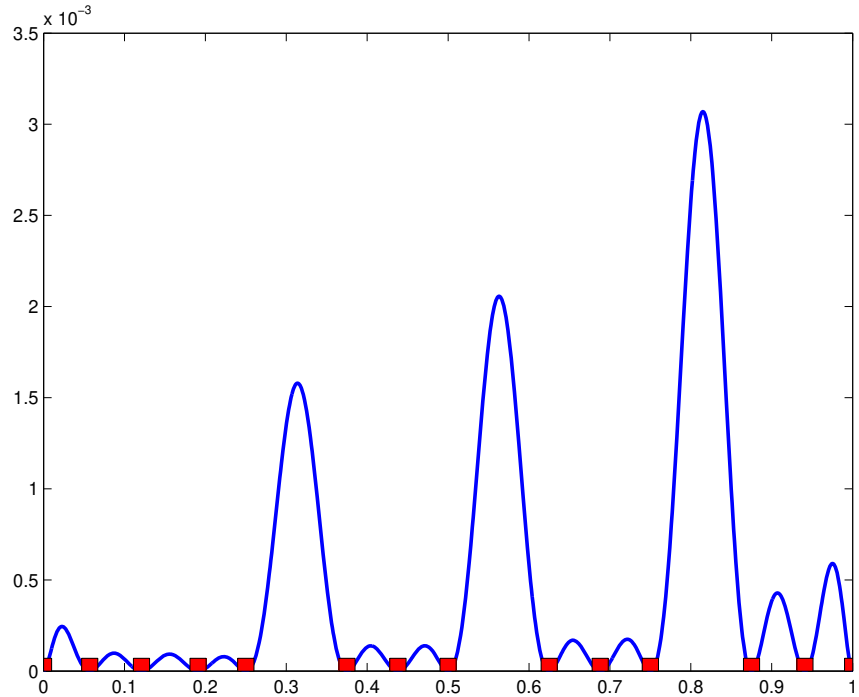
For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\boxed{\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})}$$

For MMSE : do not choose $\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$!

➠ take instead $\boxed{\mathbf{x}^*_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$

$\rho_{14}(\mathbf{x})$



Sequential construction of an optimal design:

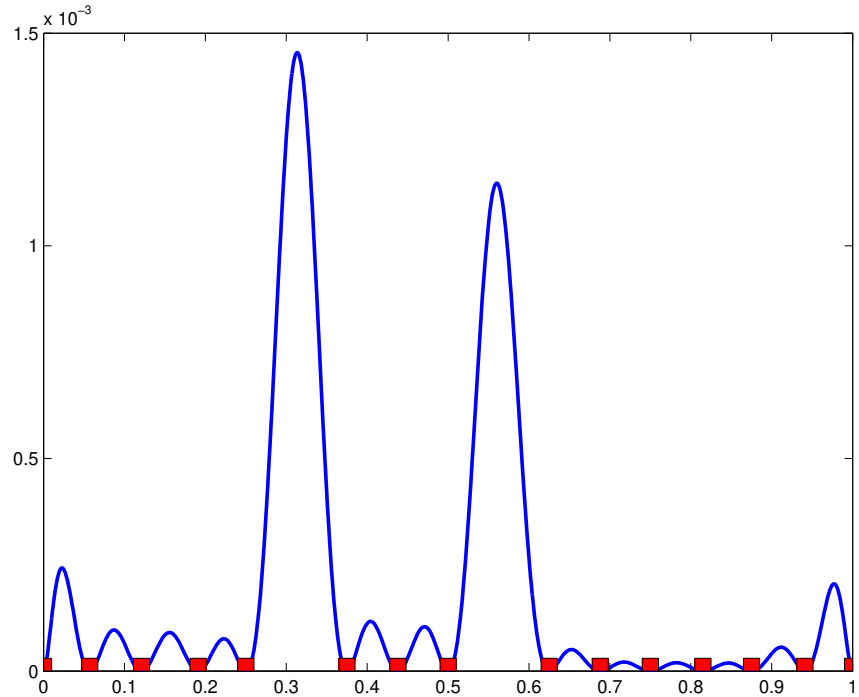
For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\boxed{\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})}$$

For MMSE : do not choose $\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$!

➠ take instead $\boxed{\mathbf{x}^*_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$

$\rho_{15}(\mathbf{x})$



Sequential construction of an optimal design:

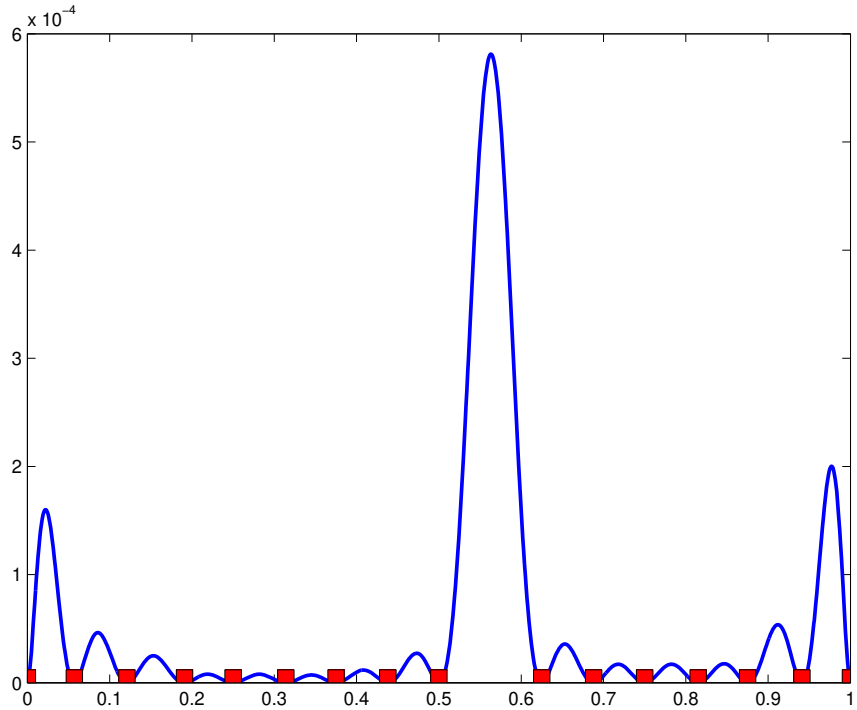
For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})$$

For MMSE : do not choose $\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$!

➠ take instead $\mathbf{x}^*_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})$

$\rho_{16}(\mathbf{x})$



Sequential construction of an optimal design:

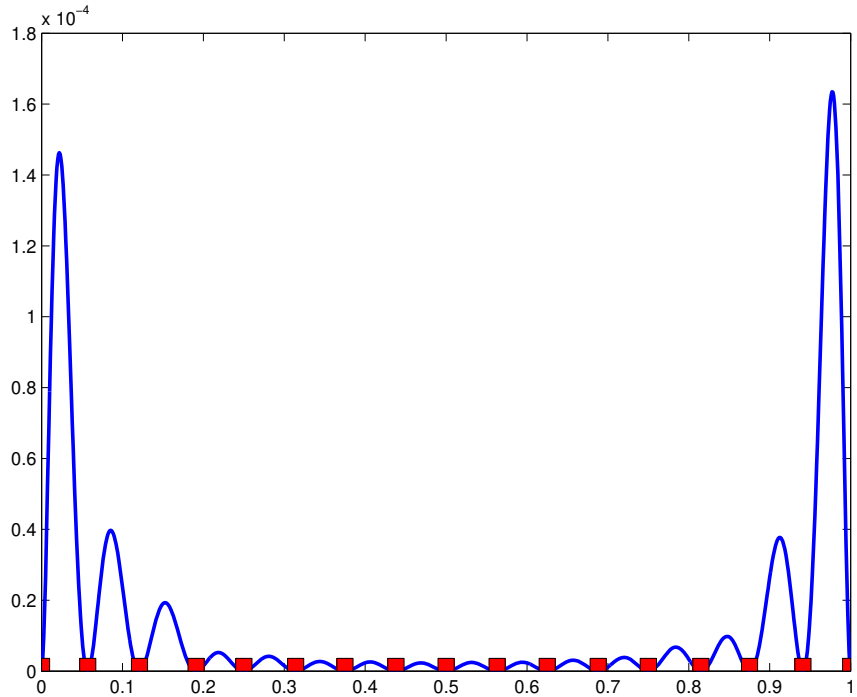
For IMSE: nothing special, at step $n+1$, $\mathbf{X}_{n+1} = \{\mathbf{X}_n, \mathbf{x}_{n+1}\}$ with

$$\boxed{\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{IMSE}(\{\mathbf{X}_n, \mathbf{x}\})}$$

For MMSE : do not choose $\mathbf{x}^*_{n+1} = \arg\min_{\mathbf{x}\in\mathscr{X}} \text{MMSE}(\{\mathbf{X}_n, \mathbf{x}\})$!

➠ take instead $\boxed{\mathbf{x}^*_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$

$\rho_{17}(\mathbf{x})$



1.3 **Maximum Entropy Sampling** (Shewry and Wynn, 1987)

Without trend ($\mathbf{r}(\mathbf{x}) = \mathbf{0}\ \forall \mathbf{x}$) :

- $\mathbf{z}_Q \triangleq$ vector with components $Z(\mathbf{x}^{(k)})$, $\mathbf{x}^{(k)} \in \mathscr{X}_Q$
- $\mathbf{z}_n \triangleq$ vector with components $Z(\mathbf{x}_i)$, $i = 1, \ldots, n$ (observations)
- $H_1(\mathbf{z}) \triangleq -\int \varphi(\mathbf{z}) \log[\varphi(\mathbf{z})]\, \mathrm{d}\mathbf{z}$ Shannon entropy of $\varphi(\mathbf{z})$
  $= $ measure of "dispersion"

  $H_1(\mathbf{z}_1|\mathbf{z}_2) \triangleq$ conditional entropy of $\mathbf{z}_1$ given $\mathbf{z}_2$
  $= \int \left[-\int \varphi(\mathbf{z}_1|\mathbf{z}_2) \log[\varphi(\mathbf{z}_1|\mathbf{z}_2)]\, \mathrm{d}\mathbf{z}_1 \right] \varphi(\mathbf{z}_2)\, \mathrm{d}(\mathbf{z}_2)$

1.3 **Maximum Entropy Sampling** (Shewry and Wynn, 1987)

Without trend ($\mathbf{r}(\mathbf{x}) = \mathbf{0}\ \forall \mathbf{x}$) :

- $\mathbf{z}_Q \triangleq$ vector with components $Z(\mathbf{x}^{(k)})$, $\mathbf{x}^{(k)} \in \mathscr{X}_Q$
- $\mathbf{z}_n \triangleq$ vector with components $Z(\mathbf{x}_i)$, $i = 1, \ldots, n$ (observations)
- $H_1(\mathbf{z}) \triangleq -\int \varphi(\mathbf{z}) \log[\varphi(\mathbf{z})]\, \mathrm{d}\mathbf{z}$ Shannon entropy of $\varphi(\mathbf{z})$
  $= $ measure of "dispersion"

  $H_1(\mathbf{z}_1|\mathbf{z}_2) \triangleq$ conditional entropy of $\mathbf{z}_1$ given $\mathbf{z}_2$
  $= \int \left[-\int \varphi(\mathbf{z}_1|\mathbf{z}_2) \log[\varphi(\mathbf{z}_1|\mathbf{z}_2)]\, \mathrm{d}\mathbf{z}_1 \right] \varphi(\mathbf{z}_2)\, \mathrm{d}(\mathbf{z}_2)$

We get $\boxed{H_1(\mathbf{y}_Q) = H_1(\mathbf{y}_n) + \mathrm{E}\{H_1(\mathbf{y}_Q|\mathbf{y}_n)\}}$

1.3 **Maximum Entropy Sampling** (Shewry and Wynn, 1987)

Without trend ($\mathbf{r}(\mathbf{x}) = \mathbf{0}\ \forall \mathbf{x}$) :

- $\mathbf{z}_Q \triangleq$ vector with components $Z(\mathbf{x}^{(k)})$, $\mathbf{x}^{(k)} \in \mathscr{X}_Q$
- $\mathbf{z}_n \triangleq$ vector with components $Z(\mathbf{x}_i)$, $i = 1, \ldots, n$ (observations)
- $H_1(\mathbf{z}) \triangleq -\int \varphi(\mathbf{z}) \log[\varphi(\mathbf{z})]\, \mathrm{d}\mathbf{z}$ Shannon entropy of $\varphi(\mathbf{z})$
  $= $ measure of "dispersion"

  $H_1(\mathbf{z}_1|\mathbf{z}_2) \triangleq$ conditional entropy of $\mathbf{z}_1$ given $\mathbf{z}_2$
  $= \int \left[-\int \varphi(\mathbf{z}_1|\mathbf{z}_2) \log[\varphi(\mathbf{z}_1|\mathbf{z}_2)]\, \mathrm{d}\mathbf{z}_1 \right] \varphi(\mathbf{z}_2)\, \mathrm{d}(\mathbf{z}_2)$

We get $\boxed{\underbrace{H_1(\mathbf{y}_Q)}_{=\text{constant}} = H_1(\mathbf{y}_n) + \underbrace{\mathrm{E}\{H_1(\mathbf{y}_Q|\mathbf{y}_n)\}}_{\text{to be minimized}}}$

1.3 **Maximum Entropy Sampling** (Shewry and Wynn, 1987)

Without trend ($\mathbf{r}(\mathbf{x}) = \mathbf{0}\ \forall \mathbf{x}$) :

- $\mathbf{z}_Q \triangleq$ vector with components $Z(\mathbf{x}^{(k)})$, $\mathbf{x}^{(k)} \in \mathscr{X}_Q$
- $\mathbf{z}_n \triangleq$ vector with components $Z(\mathbf{x}_i)$, $i = 1, \ldots, n$ (observations)
- $H_1(\mathbf{z}) \triangleq -\int \varphi(\mathbf{z}) \log[\varphi(\mathbf{z})]\, \mathrm{d}\mathbf{z}$ Shannon entropy of $\varphi(\mathbf{z})$
  $= $ measure of "dispersion"

  $H_1(\mathbf{z}_1|\mathbf{z}_2) \triangleq$ conditional entropy of $\mathbf{z}_1$ given $\mathbf{z}_2$
  $= \int \left[-\int \varphi(\mathbf{z}_1|\mathbf{z}_2) \log[\varphi(\mathbf{z}_1|\mathbf{z}_2)]\, \mathrm{d}\mathbf{z}_1\right] \varphi(\mathbf{z}_2)\, \mathrm{d}(\mathbf{z}_2)$

We get $\boxed{\underbrace{H_1(\mathbf{y}_Q)}_{=\text{constant}} = H_1(\mathbf{y}_n) + \underbrace{\mathrm{E}\{H_1(\mathbf{y}_Q|\mathbf{y}_n)\}}_{\text{to be minimized}}}$

Minimize $\mathrm{E}\{H_1(\mathbf{y}_Q|\mathbf{y}_n)\}$ w.r.t. $\mathbf{X}_n \Leftrightarrow$ maximize $H_1(\mathbf{y}_n)$

$Z(\mathbf{x})$ is Gaussian ➠ $\boxed{\text{maximize } \det[\mathbf{C}_n]}$

= intra-distances criterion

Sequential construction of an optimal design:

$$\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \det[\mathbf{C}_{n+1}] = \arg\max_{\mathbf{x}\in\mathscr{X}} \underbrace{\det\begin{bmatrix} \mathbf{C}_n & \mathbf{c}_n(\mathbf{x}) \\ \mathbf{c}_n^\top(\mathbf{x}) & 1 \end{bmatrix}}_{=\det[\mathbf{C}_n]\underbrace{(1-\mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}\mathbf{c}_n(\mathbf{x}))}_{=\rho_n(\mathbf{x})}}$$

$$\Rightarrow \boxed{\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$$

Sequential construction of an optimal design:

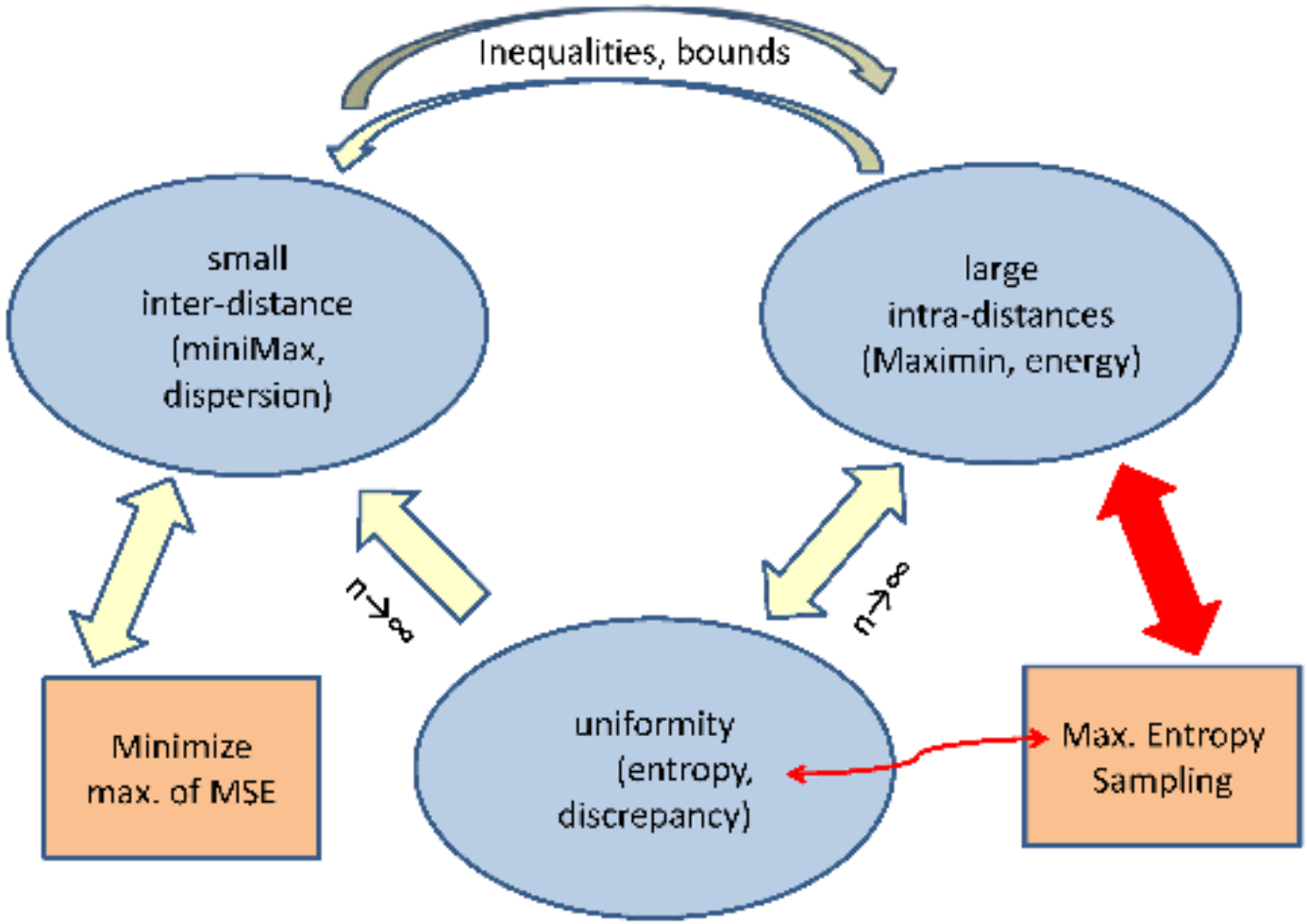
$$\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \det[\mathbf{C}_{n+1}] = \arg\max_{\mathbf{x}\in\mathscr{X}} \underbrace{\det\begin{bmatrix} \mathbf{C}_n & \mathbf{c}_n(\mathbf{x}) \\ \mathbf{c}_n^\top(\mathbf{x}) & 1 \end{bmatrix}}_{=\det[\mathbf{C}_n]\underbrace{(1-\mathbf{c}_n^\top(\mathbf{x})\mathbf{C}_n^{-1}\mathbf{c}_n(\mathbf{x}))}_{=\rho_n(\mathbf{x})}}$$

$$\Rightarrow \boxed{\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} \rho_n(\mathbf{x})}$$

The designs obtained are typically space-filling:

Johnson et al. (1990) : if $C(\mathbf{x}-\mathbf{x}') = c(\|\mathbf{x}-\mathbf{x}'\|)$ with $c(\cdot)$ decreasing, then $\mathbf{X}_n^*$ optimal for $\Phi_{Mm}(\cdot)$ (Maximin optimal) tends to be optimal for $\det[\mathbf{C}_n]$ with covariance $C_a(\mathbf{x}-\mathbf{x}') = [C(\mathbf{x}-\mathbf{x}')]^a$ when $a \to \infty$

$\Rightarrow$ there are design points $\mathbf{x}_i$ on the boundary of $\mathscr{X}$


Inequalities, bounds
small
inter-distance
(miniMax,
dispersion)
large
intra-distances
(Maximin, energy)
$n \to \infty$
$n \to \infty$
Minimize
max. of MSE
uniformity
(entropy,
discrepancy)
Max. Entropy
Sampling

2 Optimal design for linear regression

2.1 Linear regression

Observations $\boxed{y_i = y(\mathbf{x}_i) = \mathbf{r}^\top(\mathbf{x}_i)\gamma + \varepsilon_i}$, $\gamma \in \mathbb{R}^p$

with (ε_i) i.i.d., $\mathrm{E}\{\varepsilon_i\} = 0$, $\mathrm{var}\{\varepsilon_i\} = \sigma^2 \ \forall i$

Estimation of γ by Least-Squares (LS)

$$\hat{\gamma}_n = (\mathbf{R}_n^\top \mathbf{R}_n)^{-1}\mathbf{R}_n^\top \mathbf{y}_n, \text{ with } \mathbf{y}_n = (y_1, \ldots, y_n)^\top \text{ and } \mathbf{R}_n = \begin{pmatrix} \mathbf{r}^\top(\mathbf{x}_1) \\ \vdots \\ \mathbf{r}^\top(\mathbf{x}_n) \end{pmatrix}$$

$\mathrm{E}\{\hat{\gamma}_n\} = \gamma$ (unbiased)

2 Optimal design for linear regression

2.1 Linear regression

Observations $\boxed{y_i = y(\mathbf{x}_i) = \mathbf{r}^\top(\mathbf{x}_i)\gamma + \varepsilon_i}$, $\gamma \in \mathbb{R}^p$

with (ε_i) i.i.d., $\mathrm{E}\{\varepsilon_i\} = 0$, $\mathrm{var}\{\varepsilon_i\} = \sigma^2 \ \forall i$

Estimation of γ by Least-Squares (LS)

$$\hat{\gamma}_n = (\mathbf{R}_n^\top \mathbf{R}_n)^{-1}\mathbf{R}_n^\top \mathbf{y}_n, \text{ with } \mathbf{y}_n = (y_1, \ldots, y_n)^\top \text{ and } \mathbf{R}_n = \begin{pmatrix} \mathbf{r}^\top(\mathbf{x}_1) \\ \vdots \\ \mathbf{r}^\top(\mathbf{x}_n) \end{pmatrix}$$

$\mathrm{E}\{\hat{\gamma}_n\} = \gamma$ (unbiased)

$$\text{Covariance} = \mathrm{cov}(\hat{\gamma}_n) = \sigma^2(\mathbf{R}_n^\top \mathbf{R}_n)^{-1} = \frac{\sigma^2}{n}\left[\underbrace{\frac{1}{n}\sum_{i=1}^{n}\mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)}_{\mathbf{M}_n}\right]^{-1}$$

$\text{cov}(\hat{\gamma}_n) = \frac{\sigma^2}{n}\mathbf{M}_n^{-1}$, with

$$\boxed{\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)} \in \mathbb{R}^{p\times p}$$

$=$ information matrix (per observation)

$\text{cov}(\hat{\gamma}_n) = \frac{\sigma^2}{n}\mathbf{M}_n^{-1}$, with

$$\boxed{\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)} \in \mathbb{R}^{p\times p}$$

$$= \text{information matrix (per observation)}$$

Optimal design $\mathbf{X}_n^*$: maximizes a scalar function $\Phi(\cdot)$ of $\mathbf{M}_n$ (with $\Phi(\cdot)$ Loewner increasing)

- *E*-optimality: maximize $\lambda_{\min}(\mathbf{M}_n)$
  (minimize longest axis of confidence ellipsoids for γ)

$\mathrm{cov}(\hat{\gamma}_n) = \frac{\sigma^2}{n}\mathbf{M}_n^{-1}$, with

$$\boxed{\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)} \in \mathbb{R}^{p\times p}$$

$$= \text{information matrix (per observation)}$$

Optimal design $\mathbf{X}_n^*$: maximizes a scalar function $\Phi(\cdot)$ of $\mathbf{M}_n$ (with $\Phi(\cdot)$ Loewner increasing)

- *E*-optimality: maximize $\lambda_{\min}(\mathbf{M}_n)$
  (minimize longest axis of confidence ellipsoids for γ)
- *A*-optimality: maximize $-\mathrm{trace}[\mathbf{M}_n^{-1}] \Leftrightarrow$ maximize $1/\mathrm{trace}[\mathbf{M}_n^{-1}]$
  (minimize sum of squared lengths of axes of confidence ellipsoids for γ)

$\mathrm{cov}(\hat{\gamma}_n) = \frac{\sigma^2}{n}\mathbf{M}_n^{-1}$, with

$$\boxed{\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)} \in \mathbb{R}^{p\times p}$$

$$= \text{information matrix (per observation)}$$

Optimal design $\mathbf{X}_n^*$: maximizes a scalar function $\Phi(\cdot)$ of $\mathbf{M}_n$ (with $\Phi(\cdot)$ Loewner increasing)

- *E*-optimality: maximize $\lambda_{\min}(\mathbf{M}_n)$
  (minimize longest axis of confidence ellipsoids for γ)
- *A*-optimality: maximize $-\mathrm{trace}[\mathbf{M}_n^{-1}] \Leftrightarrow$ maximize $1/\mathrm{trace}[\mathbf{M}_n^{-1}]$
  (minimize sum of squared lengths of axes of confidence ellipsoids for γ)
- more generally, *L*-optimality: maximize $-\mathrm{trace}[\mathbf{L}\mathbf{M}_n^{-1}]$
  (we only consider the case $\mathbf{L}$ symmetric positive definite)

- *D*-optimality: maximize $\log\det \mathbf{M}_n$
  (minimize volume of confidence ellipsoids for γ)
  Very much used:
  - a *D*-optimal design is invariant by reparametrization:

$$\det \mathbf{M}'_n(\beta(\gamma)) = \det \mathbf{M}_n(\gamma) \det^{-2} \left(\frac{\partial \beta}{\partial \gamma^\top} \right)$$

  - ➠ often leads to repeat the same experimental conditions (replications)
    (we assumed i.i.d. errors $\varepsilon_i \Rightarrow$ several observations at the same $\mathbf{x}_i$ carry information)

⏩ Tensor-product models

2.2 Exact design

n observations at $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{x}_i \in \mathbb{R}^d$

Maximize $\Phi(\mathbf{M}_n)$ w.r.t. $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathbb{R}^{n \times d}$

with $\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)$

2.2 Exact design

n observations at $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{x}_i \in \mathbb{R}^d$

Maximize $\Phi(\mathbf{M}_n)$ w.r.t. $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathbb{R}^{n \times d}$

with $\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)$

- If problem dimension $n \times d$ is not too big
  - "standard" algorithm (but careful with constraints and local optimas!)

2.2 Exact design

n observations at $\mathbf{x}_1, \ldots, \mathbf{x}_n$, $\mathbf{x}_i \in \mathbb{R}^d$

Maximize $\Phi(\mathbf{M}_n)$ w.r.t. $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\} \in \mathbb{R}^{n \times d}$
with $\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)$

- If problem dimension $n \times d$ is not too big
  - ➠ "standard" algorithm (but careful with constraints and local optimas!)
- Otherwise, ➠ specific algorithm

Exchange method: at step k, exchange **one** support point $\mathbf{x}_j$ with a better one $\mathbf{x}^*$ in $\mathscr{X}$ (better for $\Phi(\cdot)$)

$$\mathbf{X}_n^k = (\mathbf{x}_1, \ldots, \boxed{\begin{array}{c}\mathbf{x}_j \\ \updownarrow \\ \mathbf{x}^*\end{array}}, \ldots, \mathbf{x}_n)$$

⏩ Approximate design

Fedorov (1972) algorithm:

At each iteration k, consider all n possible exchanges successively, each time starting from $\mathbf{X}_n^k$, retain the «best» one among these $n \rightarrow \mathbf{X}_n^{k+1}$ ➠ $\mathbf{X}_n^{k+1}$

$$\mathbf{X}_n^k = \left(\begin{array}{ccccc} \mathbf{x}_1 & , \ldots , & \mathbf{x}_j & , \ldots , & \mathbf{x}_n \\ \updownarrow & & \updownarrow & & \updownarrow \\ \mathbf{x}_1^* & & \mathbf{x}_j^* & & \mathbf{x}_n^* \end{array} \right)$$

Fedorov (1972) algorithm:

At each iteration k, consider all n possible exchanges successively, each time starting from $\mathbf{X}_n^k$, retain the «best» one among these $n \rightarrow \mathbf{X}_n^{k+1}$ ➠ $\mathbf{X}_n^{k+1}$

$$\mathbf{X}_n^k = \left(\begin{array}{ccccc} \mathbf{x}_1 & , \ldots , & \mathbf{x}_j & , \ldots , & \mathbf{x}_n \\ \updownarrow & & \updownarrow & & \updownarrow \\ \mathbf{x}_1^* & & \mathbf{x}_j^* & & \mathbf{x}_n^* \end{array} \right)$$

One iteration $\rightarrow$ n optimizations of dimension d followed by ranking n criterion values

DETMAX algorithm Mitchell (1974):

If one additional observation were allowed: optimal choice

$$\mathbf{X}_{n+1}^{k+} = (\mathbf{x}_1, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_n, \mathbf{x}_{n+1}^*)$$

Then, remove one support point to return to a n-points design:

➔ consider all $n+1$ possible cancellations,
retain the less penalizing in the sense of $\Phi(\cdot)$

DETMAX algorithm Mitchell (1974):

If one additional observation were allowed: optimal choice

$$\mathbf{X}_{n+1}^{k+} = (\mathbf{x}_1, \ldots, \mathbf{x}_j, \ldots, \mathbf{x}_n, \mathbf{x}_{n+1}^*)$$

Then, remove one support point to return to a n-points design:

➔ consider all $n+1$ possible cancellations,
retain the less penalizing in the sense of $\Phi(\cdot)$

➔ globally, exchange some $\mathbf{x}_j$ with $\mathbf{x}_{n+1}^*$

[= excursion of length 1, longer excursions are possible...]

One iteration $\rightarrow$ 1 optimization of dimension d followed by ranking $n+1$ criterion values

- DETMAX has simpler iterations than Fedorov, but usually requires more iterations

- DETMAX has simpler iterations than Fedorov, but usually requires more iterations
- dead ends are possible:
  - DETMAX: the point to be removed is $\mathbf{x}_{n+1}$
  - Fedorov: no possible improvement when optimizing **one** $\mathbf{x}_i$ at a time

- DETMAX has simpler iterations than Fedorov, but usually requires more iterations
- dead ends are possible:
  - DETMAX: the point to be removed is $\mathbf{x}_{n+1}$
  - Fedorov: no possible improvement when optimizing **one** $\mathbf{x}_i$ at a time
- ▲ both give local optima only ▲

- DETMAX has simpler iterations than Fedorov, but usually requires more iterations
- dead ends are possible:
  - DETMAX: the point to be removed is $\mathbf{x}_{n+1}$
  - Fedorov: no possible improvement when optimizing **one** $\mathbf{x}_i$ at a time
- ▲ both give local optima only ▲
- Other methods:
  - Branch and bound: guaranteed convergence, but complicated [Welch 1982]
  - Rounding an optimal design measure (support points $\mathbf{x}_i$ and associated weights w_i^*, $i = 1, \ldots, m$, presented next in § 2.3):
    choose n integers r_i (r_i= nb. of replications of observations at x_i) such that $\sum_{i=1}^m r_i = n$ and $r_i/n \approx w_i^*$
    (e.g., maximize $\min_{i=1,\ldots,m} r_i/w_i^*$ = Adams apportionment, see [Pukelsheim & Reider 1992])

2.3 Approximate design theory

(Chernoff, 1953; Kiefer and Wolfowitz, 1960; Fedorov, 1972; Silvey, 1980; Pukelsheim, 1993) ...

$$\boxed{\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)}$$

(the additive form is essential — related to the independence of observations)

2.3 Approximate design theory

(Chernoff, 1953; Kiefer and Wolfowitz, 1960; Fedorov, 1972; Silvey, 1980; Pukelsheim, 1993) ...

$$\boxed{\mathbf{M}_n = \mathbf{M}(\mathbf{X}_n) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)}$$

(the additive form is essential — related to the independence of observations)

If several $\mathbf{x}_i$ coincide (repetitions), with only $m < n$ different $\mathbf{x}_i$

$$\boxed{\mathbf{M}(\mathbf{X}_n) = \sum_{i=1}^{m} \frac{r_i}{n}\mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)}$$

$$\begin{aligned} \frac{r_i}{n} &= \text{proportion of observations collected at } \mathbf{x}_i \\ &= \text{«percentage of experimental effort» at } \mathbf{x}_i \\ &= \text{weight } w_i \text{ of support point } \mathbf{x}_i \end{aligned}$$

$$\mathbf{M}(\mathbf{X}_n) = \sum_{i=1}^{m} w_i \mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)$$

- design $\mathbf{X}_n \Leftrightarrow \left\{ \begin{array}{ccc} \mathbf{x}_1 & \cdots & \mathbf{x}_m \\ w_1 & \cdots & w_m \end{array} \right\}$ with $\sum_{i=1}^{m} w_i = 1$
- normalized discrete distribution on the $\mathbf{x}_i$,

with constraints $w_i = r_i/n$

$\mathbf{M}(\mathbf{X}_n) = \sum_{i=1}^{m} w_i \mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)$

- design $\mathbf{X}_n \Leftrightarrow \left\{ \begin{array}{ccc} \mathbf{x}_1 & \cdots & \mathbf{x}_m \\ w_1 & \cdots & w_m \end{array} \right\}$ with $\sum_{i=1}^{m} w_i = 1$
- normalized discrete distribution on the $\mathbf{x}_i$,

  with constraints $w_i = r_i/n$
- Release the constraints: only enforce $w_i \geq 0$ et $\sum_{i=1}^{m} w_i = 1$
- ξ = discrete probability measure on $\mathcal{X}$

  support points $\mathbf{x}_i$ and associated weights w_i

  = "approximate design"

$\mathbf{M}(\mathbf{X}_n) = \sum_{i=1}^m w_i \mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)$

- design $\mathbf{X}_n \Leftrightarrow \left\{ \begin{array}{ccc} \mathbf{x}_1 & \cdots & \mathbf{x}_m \\ w_1 & \cdots & w_m \end{array} \right\}$ with $\sum_{i=1}^m w_i = 1$
- normalized discrete distribution on the $\mathbf{x}_i$,

  with constraints $w_i = r_i/n$
- Release the constraints: only enforce $w_i \geq 0$ et $\sum_{i=1}^m w_i = 1$
- ξ = discrete probability measure on $\mathscr{X}$

  support points $\mathbf{x}_i$ and associated weights w_i

  = "approximate design"

More general expression: ξ = any probability measure on $\mathscr{X}$

$$\mathbf{M}(\xi) = \int_{\mathscr{X}} \mathbf{r}(\mathbf{x})\mathbf{r}^\top(\mathbf{x})\, \xi(\mathrm{d}\mathbf{x}) \text{ with } \int_{\mathscr{X}} \xi(\mathrm{d}\mathbf{x}) = 1$$

$\mathbf{M}(\xi) \in$ convex closure of $\mathcal{M}=$ set of rank 1 matrices

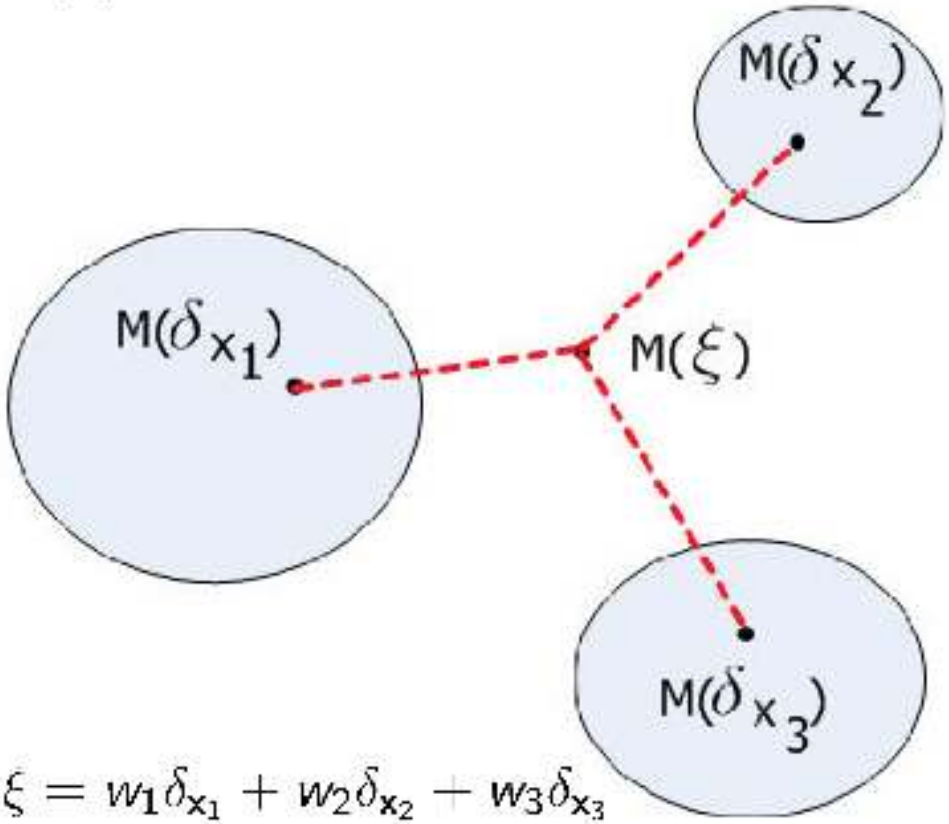
$$\mathbf{M}(\delta_x) = \mathbf{r}(\mathbf{x})\mathbf{r}^\top(\mathbf{x})$$

$\mathbf{M}(\xi)$ is symmetric $p \times p \Rightarrow \in q$-dimensional space, $q = \frac{p(p+1)}{2}$

$\mathbf{M}(\xi) \in$ convex closure of $\mathcal{M}=$ set of rank 1 matrices

$$\mathbf{M}(\delta_x) = \mathbf{r}(\mathbf{x})\mathbf{r}^\top(\mathbf{x})$$

$\mathbf{M}(\xi)$ is symmetric $p \times p \Rightarrow \in q$-dimensional space, $q = \frac{p(p+1)}{2}$



$$\xi = w_1\delta_{\mathbf{x}_1} + w_2\delta_{\mathbf{x}_2} + w_3\delta_{\mathbf{x}_3}$$

(3 points are enough for $q = 2$)

Caratheodory Theorem:

$\mathbf{M}(\xi)$ can be written as the linear combination of at most $q+1$ elements of $\mathcal{M}$:

$\mathbf{M}(\xi) = \sum_{i=1}^{m} w_i \, \mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)\,, \quad m \leq \frac{p(p+1)}{2} + 1$

$\Rightarrow$ consider discrete probability measures with $\frac{p(p+1)}{2} + 1$ support points at most

(true in particular for the optimum design!)

[Even better: for many criteria $\Phi(\cdot)$, if ξ^* is optimal (maximizes $\Phi[\mathbf{M}(\xi)]$) then $\mathbf{M}(\xi^*)$ is on the boundary of the convex closure of $\mathcal{M}$ and $\frac{p(p+1)}{2}$ support points are enough]

Caratheodory Theorem:

$\mathbf{M}(\xi)$ can be written as the linear combination of at most $q+1$ elements of $\mathcal{M}$:
$\mathbf{M}(\xi) = \sum_{i=1}^{m} w_i \, \mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)\,, \quad m \leq \frac{p(p+1)}{2} + 1$

$\Rightarrow$ consider discrete probability measures with $\frac{p(p+1)}{2} + 1$ support points at most
(true in particular for the optimum design!)

[Even better: for many criteria $\Phi(\cdot)$, if ξ^* is optimal (maximizes $\Phi[\mathbf{M}(\xi)]$) then $\mathbf{M}(\xi^*)$ is on the boundary of the convex closure of $\mathcal{M}$ and $\frac{p(p+1)}{2}$ support points are enough]

Suppose we found an optimal $\xi^* = \sum_{i=1}^{m} w_i^* \delta_{\mathbf{x}_i}$

☞ for a given n, choose the r_i so that $\frac{r_i}{n} \simeq w_i^*$ optimum
$\rightarrow$ *rounding of an approximate design*

Caratheodory Theorem:

$\mathbf{M}(\xi)$ can be written as the linear combination of at most $q+1$ elements of $\mathcal{M}$:
$\mathbf{M}(\xi) = \sum_{i=1}^{m} w_i \, \mathbf{r}(\mathbf{x}_i)\mathbf{r}^\top(\mathbf{x}_i)\,, \quad m \leq \frac{p(p+1)}{2} + 1$

$\Rightarrow$ consider discrete probability measures with $\frac{p(p+1)}{2} + 1$ support points at most
(true in particular for the optimum design!)

[Even better: for many criteria $\Phi(\cdot)$, if ξ^* is optimal (maximizes $\Phi[\mathbf{M}(\xi)]$) then $\mathbf{M}(\xi^*)$ is on the boundary of the convex closure of $\mathcal{M}$ and $\frac{p(p+1)}{2}$ support points are enough]

Suppose we found an optimal $\xi^* = \sum_{i=1}^{m} w_i^* \delta_{\mathbf{x}_i}$

☞ for a given n, choose the r_i so that $\frac{r_i}{n} \simeq w_i^*$ optimum
$\rightarrow$ *rounding of an approximate design*

Why design measures are interesting?
How does it simplify the optimization problem?

➠ Maximize $\Phi(\cdot)$ concave w.r.t. $\mathbf{M}(\xi)$ in a convex set

Ex: D-optimality: $\forall\, \mathbf{M}_1 \succ \mathbf{O}$, $\mathbf{M}_2 \succeq \mathbf{O}$, with $\mathbf{M}_2 \not\propto \mathbf{M}_2$, $\forall \alpha$, $0 < \alpha < 1$,

$\log\det[(1-\alpha)\mathbf{M}_1 + \alpha\mathbf{M}_2] > (1-\alpha)\log\det\mathbf{M}_1 + \alpha\log\det\mathbf{M}_2$

$\Rightarrow \log\det[\cdot]$ is (strictly) concave

convex set + concave criterion $\Rightarrow$ one unique optimum!

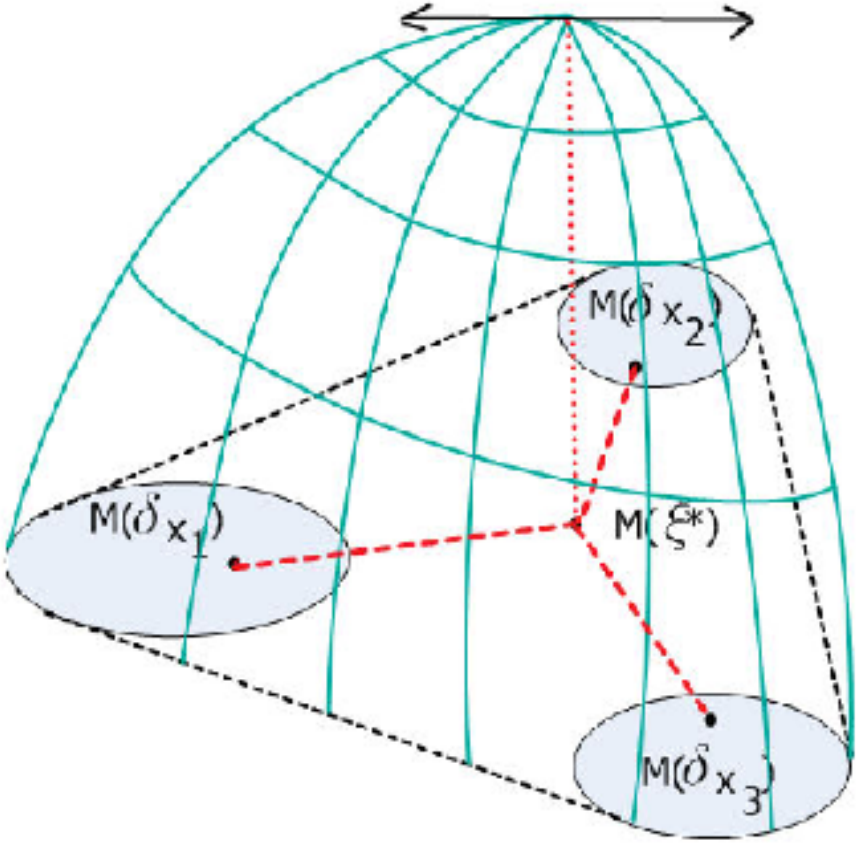
➠ Maximize $\Phi(\cdot)$ concave w.r.t. $\mathbf{M}(\xi)$ in a convex set

Ex: D-optimality: $\forall\, \mathbf{M}_1 \succ \mathbf{O}$, $\mathbf{M}_2 \succeq \mathbf{O}$, with $\mathbf{M}_2 \not\propto \mathbf{M}_2$, $\forall \alpha$, $0 < \alpha < 1$,

$\log\det[(1-\alpha)\mathbf{M}_1 + \alpha\mathbf{M}_2] > (1-\alpha)\log\det\mathbf{M}_1 + \alpha\log\det\mathbf{M}_2$

$\Rightarrow \log\det[\cdot]$ is (strictly) concave

convex set + concave criterion $\Rightarrow$ one unique optimum!



ξ^* is optimal $\Leftrightarrow$ directional derivative ≤ 0 in all directions

$\Longrightarrow$ "Equivalence Theorem" [Kiefer & Wolfowitz 1960]

Ξ = set of probability measures on $\mathscr{X}$, $\Phi(\cdot)$ concave, $\phi(\xi) = \Phi[\mathbf{M}(\xi)]$

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

$= $ directional derivative of $\phi(\cdot)$ at ξ in direction ν

Equivalence Theorem: $\boxed{\xi^* \text{ maximizes } \phi(\xi) \Leftrightarrow \max_{\nu\in\Xi} F_\phi(\xi^*;\nu) \leq 0}$

$\Longrightarrow$ "Equivalence Theorem" [Kiefer & Wolfowitz 1960]

Ξ = set of probability measures on $\mathcal{X}$, $\Phi(\cdot)$ concave, $\phi(\xi) = \Phi[\mathbf{M}(\xi)]$

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

$= $ directional derivative of $\phi(\cdot)$ at ξ in direction ν

Equivalence Theorem: $\boxed{\xi^* \text{ maximizes } \phi(\xi) \Leftrightarrow \max_{\nu\in\Xi} F_\phi(\xi^*;\nu) \leq 0}$

➔ Takes a simple form when $\Phi(\cdot)$ is differentiable

$$\boxed{\xi^* \text{ maximizes } \phi(\xi) \Leftrightarrow \max_{\mathbf{x}\in\mathcal{X}} F_\phi(\xi^*;\delta_{\mathbf{x}}) \leq 0}$$

☞ Check optimality of ξ^* by plotting $F_\phi(\xi^*;\delta_{\mathbf{x}})$

Ex: *D*-optimal design

- ξ_D^* maximizes $\log\det[\mathbf{M}(\xi)]$ w.r.t. $\xi \in \Xi$
- $\Leftrightarrow \max_{\mathbf{x}\in\mathscr{X}} d(\xi_D^*, \mathbf{x}) \leq p$
- $\Leftrightarrow \xi_D^*$ minimizes $\max_{\mathbf{x}\in\mathscr{X}} d(\xi, \mathbf{x})$ w.r.t. $\xi \in \Xi$

where $d(\xi, \mathbf{x}) = \mathbf{r}^\top(\mathbf{x})\mathbf{M}^{-1}(\xi)\mathbf{r}(\mathbf{x})$

Moreover, $d(\xi_D^*, \mathbf{x}_i) = p = \dim(\theta)$ for any $\mathbf{x}_i$ = support point of ξ_D^*

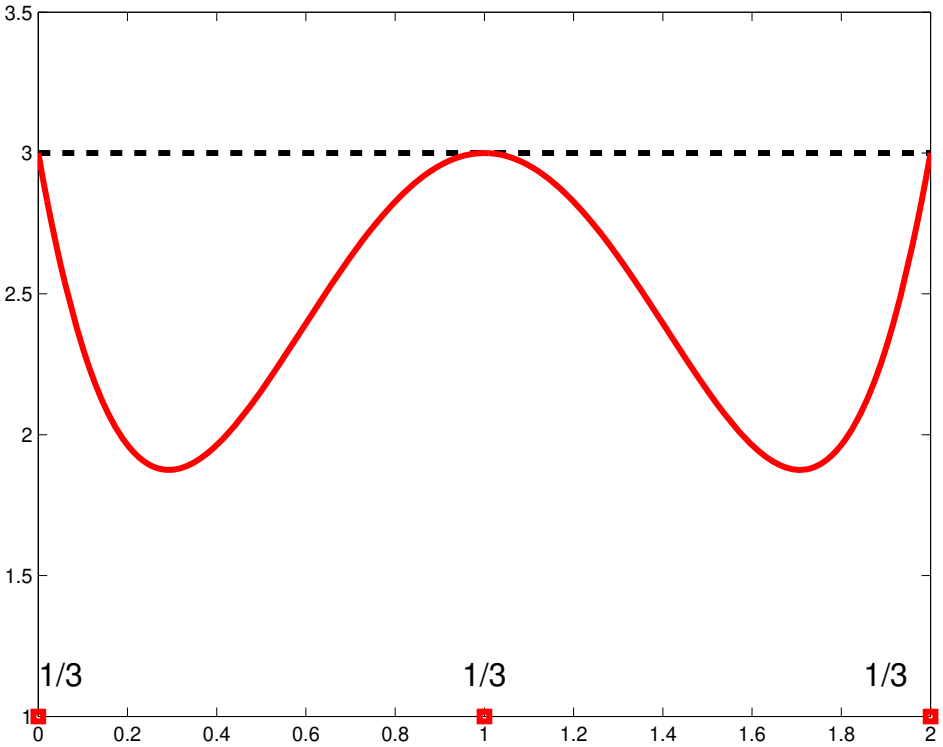
Ex. : $\mathbf{r}(x) = (1 \; x \; x^2)^\top$ $(p = 3)$ i.i.d. erreurs, $\mathscr{X} = [0, 2]$

➠ $d(\xi, x)$ as a function of x

Ex. : $\mathbf{r}(x) = (1 \; x \; x^2)^\top$ ($p = 3$) i.i.d. erreurs, $\mathscr{X} = [0, 2]$

➠ $d(\xi, x)$ as a function of x

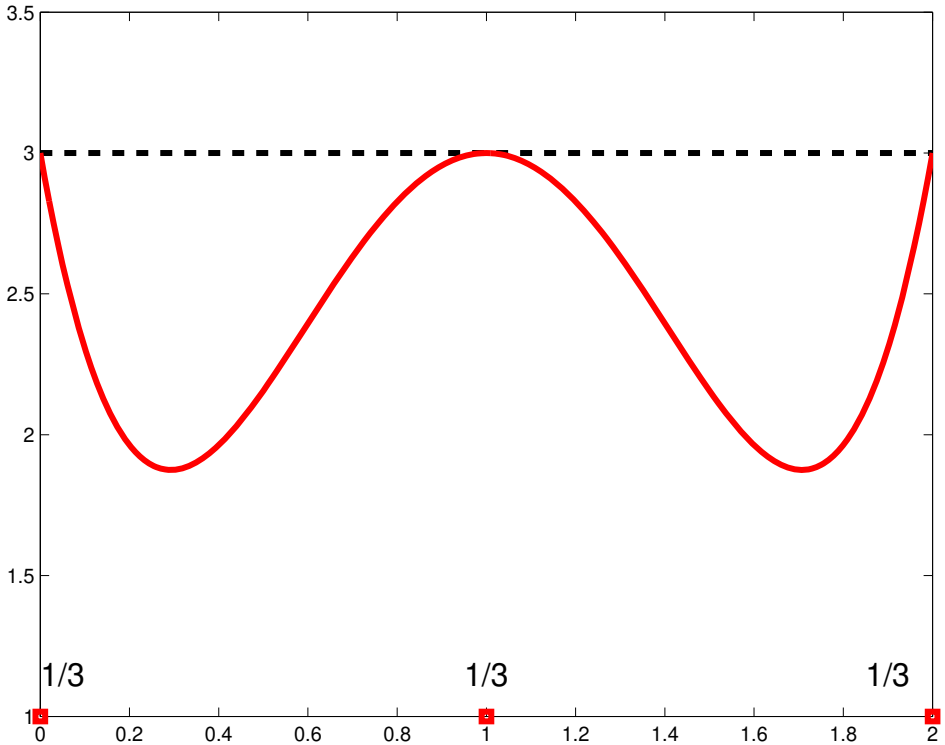
$$\xi_D^* = \left\{ \begin{array}{ccc} 0 & 1 & 2 \\ 1/3 & 1/3 & 1/3 \end{array} \right\}$$

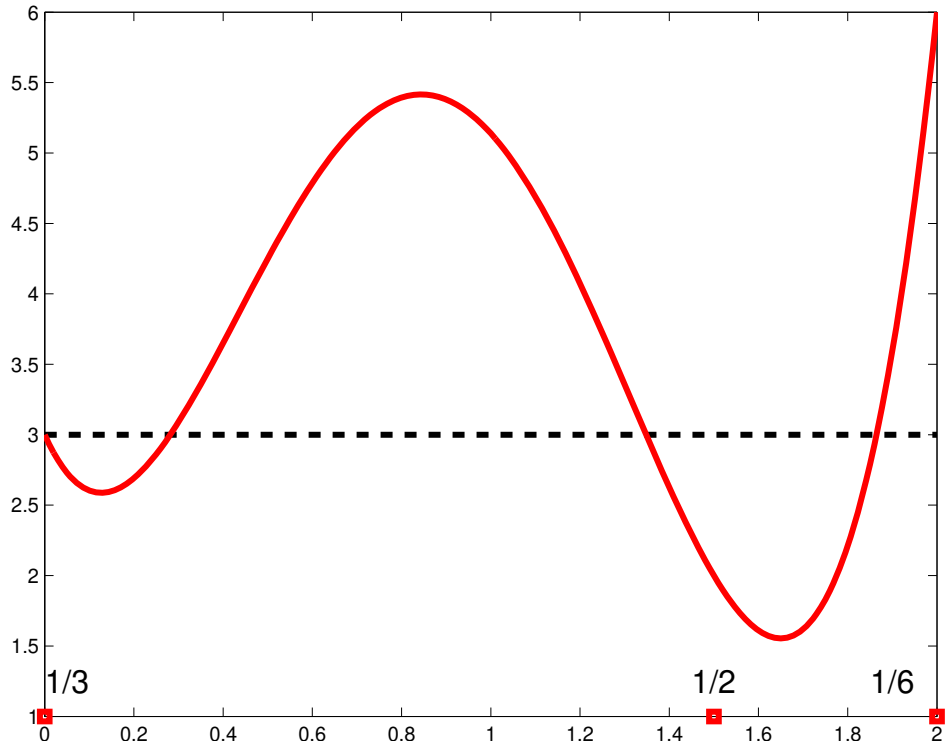


Ex. : $\mathbf{r}(x) = (1 \; x \; x^2)^\top$ $(p = 3)$ i.i.d. erreurs, $\mathscr{X} = [0, 2]$

➠ $d(\xi, x)$ as a function of x

$$\xi_D^* = \left\{ \begin{array}{ccc} 0 & 1 & 2 \\ 1/3 & 1/3 & 1/3 \end{array} \right\} \qquad \xi = \left\{ \begin{array}{ccc} 0 & 1.5 & 2 \\ 1/3 & 1/2 & 1/6 \end{array} \right\}$$





KW Eq. Th. relates optimality in γ space (parameters)
to optimality in y space (observations)

$$n\text{var}[\mathbf{r}^\top(\mathbf{x})\hat{\gamma}^n)] = \sigma^2\,\mathbf{r}^\top(\mathbf{x})\mathbf{M}^{-1}(\xi)\mathbf{r}(\mathbf{x}) = \sigma^2\,d(\xi,\mathbf{x}) \text{ (i.i.d. errors)}$$

$$D\text{-optimality} \Leftrightarrow G\text{-optimality}$$

➠ ξ_D^* minimizes the maximum value of prediction variance over $\mathscr{X}$

KW Eq. Th. relates optimality in γ space (parameters) to optimality in y space (observations)

$$n\text{var}[\mathbf{r}^\top(\mathbf{x})\hat{\gamma}^n)] = \sigma^2\,\mathbf{r}^\top(\mathbf{x})\mathbf{M}^{-1}(\xi)\mathbf{r}(\mathbf{x}) = \sigma^2\,d(\xi,\mathbf{x}) \text{ (i.i.d. errors)}$$

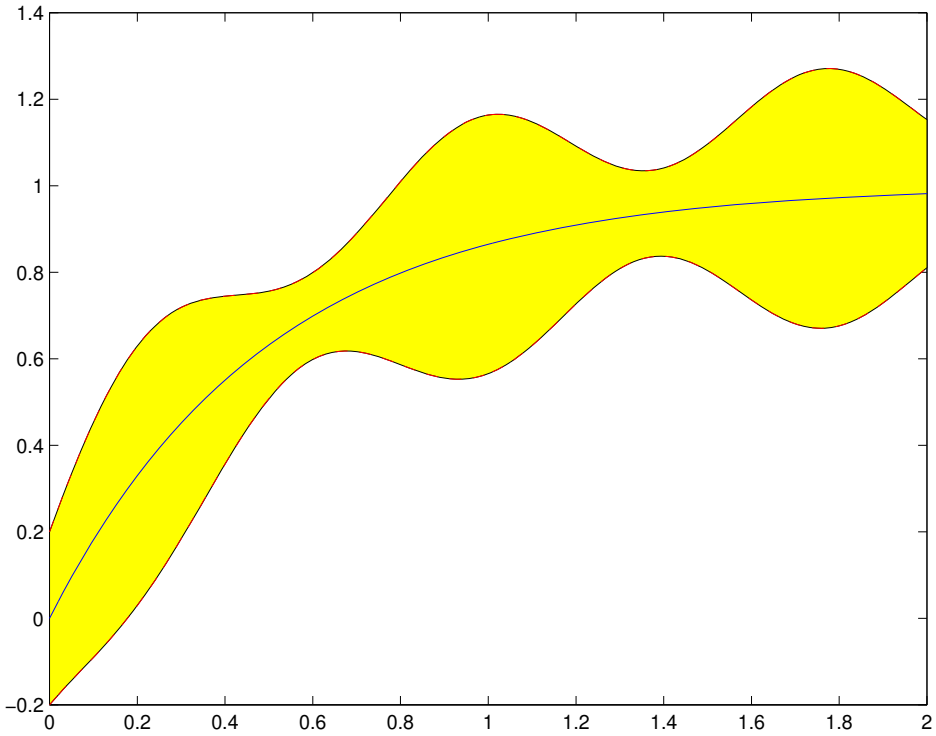
D-optimality $\Leftrightarrow$ G-optimality

➠ ξ_D^* minimizes the maximum value of prediction variance over $\mathscr{X}$

$\eta(\mathbf{x},\hat{\gamma}^n)$

$\eta(\mathbf{x},\hat{\gamma}^n) \pm$ 2 standard deviations

➠ put next observation where $d(\xi,\mathbf{x})$ is large



⏩ Tensor-product models

Construction of an optimal design measure

Central idea (▲ for a differentiable $\Phi(\cdot)$ ▲): use steepest-ascent direction

Fedorov–Wynn :

- 1 : Choose ξ^1 non degenerate ($\det \mathbf{M}(\xi^1) > 0$)
- 2 : Compute $\mathbf{x}_k^* = \arg\max_{\mathscr{X}} F_\phi(\xi^k; \delta_{\mathbf{x}})$
  If $F_\phi(\xi^k; \delta_{\mathbf{x}_k^+}) < \epsilon$, stop: ξ^k is ϵ-optimal
- 3 : $\xi^{k+1} = (1 - \alpha_k)\xi^k + \alpha_k \delta_{\mathbf{x}_k^*}$ (delta measure at $\mathbf{x}_k^*$)
  [*Vertex Direction*]
  $k \to k+1$, return to step 2

Construction of an optimal design measure

Central idea (▲ for a differentiable $\Phi(\cdot)$ ▲): use steepest-ascent direction

Fedorov–Wynn :

- 1 : Choose ξ^1 non degenerate ($\det \mathbf{M}(\xi^1) > 0$)
- 2 : Compute $\mathbf{x}_k^* = \arg\max_{\mathscr{X}} F_\phi(\xi^k; \delta_{\mathbf{x}})$
  If $F_\phi(\xi^k; \delta_{\mathbf{x}_k^+}) < \epsilon$, stop: ξ^k is ϵ-optimal
- 3 : $\xi^{k+1} = (1 - \alpha_k)\xi^k + \alpha_k \delta_{\mathbf{x}_k^*}$ (delta measure at $\mathbf{x}_k^*$)
  [*Vertex Direction*]
  $k \to k+1$, return to step 2

Step size α_k?

➠ $\alpha_k = \arg\max \phi(\xi^{k+1})$
$= \frac{d(\xi^k, \mathbf{x}_k^*) - p}{p[d(\xi^k, \mathbf{x}_k^*) - 1]}$ for *D*-optimality (Fedorov, 1972)
$\to$ monotone convergence

➠ $\alpha_k > 0$, $\lim_{k\to\infty} \alpha_k = 0$, $\sum_{i=1}^{\infty} \alpha_k = \infty$
((Wynn, 1970) for *D*-optimality)

Remarks:

- Sequential design, one $\mathbf{x}_i$ at at a time enters $\mathbf{M}(\mathbf{X})$:

$$\mathbf{M}(\mathbf{X}_{k+1}) = \frac{k}{k+1}\mathbf{M}(\mathbf{X}_k) + \frac{1}{k+1}\mathbf{r}(\mathbf{x}_{k+1})\mathbf{r}^\top(\mathbf{x}_{k+1})$$

with $\mathbf{x}_{k+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} F_\phi(\xi^k; \delta_\mathbf{x})$

$\Leftrightarrow$ Wynn's algorithm with $\alpha_k = \frac{1}{k+1}$

Remarks:

- Sequential design, one $\mathbf{x}_i$ at at a time enters $\mathbf{M}(\mathbf{X})$:

$$\mathbf{M}(\mathbf{X}_{k+1}) = \frac{k}{k+1}\mathbf{M}(\mathbf{X}_k) + \frac{1}{k+1}\mathbf{r}(\mathbf{x}_{k+1})\mathbf{r}^\top(\mathbf{x}_{k+1})$$

with $\mathbf{x}_{k+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} F_\phi(\xi^k; \delta_\mathbf{x})$

$\Leftrightarrow$ Wynn's algorithm with $\alpha_k = \frac{1}{k+1}$

- **Guaranteed convergence to the optimum**

Remarks:

- Sequential design, one $\mathbf{x}_i$ at at a time enters $\mathbf{M}(\mathbf{X})$:

$$\mathbf{M}(\mathbf{X}_{k+1}) = \frac{k}{k+1}\,\mathbf{M}(\mathbf{X}_k) + \frac{1}{k+1}\mathbf{r}(\mathbf{x}_{k+1})\mathbf{r}^\top(\mathbf{x}_{k+1})$$

  with $\mathbf{x}_{k+1} = \arg\max_{\mathbf{x}\in\mathscr{X}} F_\phi(\xi^k; \delta_\mathbf{x})$

  $\Leftrightarrow$ Wynn's algorithm with $\alpha_k = \frac{1}{k+1}$

- **Guaranteed convergence to the optimum**
- There exist faster methods:
  - remove support points from ξ^k ($\approx$ allow α_k to be < 0) (Atwood, 1973; Böhning, 1985, 1986)
  - combine with gradient projection (or a second-order method) (Wu, 1978)
  - use a multiplicative algorithm (Titterington, 1976; Torsney, 1983, 2009; Yu, 2010) (for A and D optimality, far from the optimum)
  - combine different methods (Yu, 2011)
  - Still an active topic — especially for non differentiable $\Phi(\cdot)\ldots$

2.4 Tensor-product models

D-optimality (also true for *A*-optimality under some conditions (Schwabe, 1996))

$[\mathbf{r}^{(k)}(x)]^\top \theta^{(k)} \triangleq \sum_{i=1}^{d_k} \theta_i^{(k)} x^i$ polynomial with degree d_k, $\dim(\theta^{(k)}) = p_k = 1 + d_k$

Global model for $\mathbf{x} = (\{\mathbf{x}\}_1, \{\mathbf{x}\}_2, \ldots, \{\mathbf{x}\}_d)^\top$:

$\mathbf{r}^\top(\mathbf{x})\gamma = \prod_{k=1}^{d} [\mathbf{r}^{(k)}(x)]^\top \theta^{(k)}$,

total degree $\sum_{k=1}^{d} d_k$, $\dim(\gamma) = \prod_{k=1}^{d} p_k$

2.4 Tensor-product models

D-optimality (also true for *A*-optimality under some conditions (Schwabe, 1996))

$[\mathbf{r}^{(k)}(x)]^\top \theta^{(k)} \triangleq \sum_{i=1}^{d_k} \theta_i^{(k)} x^i$ polynomial with degree d_k, $\dim(\theta^{(k)}) = p_k = 1 + d_k$

Global model for $\mathbf{x} = (\{\mathbf{x}\}_1, \{\mathbf{x}\}_2, \ldots, \{\mathbf{x}\}_d)^\top$:

$\mathbf{r}^\top(\mathbf{x})\gamma = \prod_{k=1}^{d} [\mathbf{r}^{(k)}(x)]^\top \theta^{(k)}$,

total degree $\sum_{k=1}^{d} d_k$, $\dim(\gamma) = \prod_{k=1}^{d} p_k$

Example:

$$\begin{aligned}
\mathbf{r}^\top(\mathbf{x})\gamma &= (\theta_0^{(1)} + \theta_1^{(1)}\{\mathbf{x}\}_1 + \theta_2^{(1)}\{\mathbf{x}\}_1^2) \times (\theta_0^{(2)} + \theta_1^{(2)}\{\mathbf{x}\}_2 + \theta_2^{(2)}\{\mathbf{x}\}_2^2) \\
&= \gamma_0 + \gamma_1\{\mathbf{x}\}_1 + \gamma_2\{\mathbf{x}\}_2 + \gamma_{12}\{\mathbf{x}\}_1\{\mathbf{x}\}_2 + \gamma_{11}\{\mathbf{x}\}_1{}^2 + \gamma_{22}\{\mathbf{x}\}_2{}^2 \\
&\quad + \gamma_{112}\{\mathbf{x}\}_1{}^2\{\mathbf{x}\}_2 + \gamma_{122}\{\mathbf{x}\}_1\{\mathbf{x}\}_2{}^2 + \gamma_{1122}\{\mathbf{x}\}_1{}^2\{\mathbf{x}\}_2{}^2
\end{aligned}$$

2.4 Tensor-product models

D-optimality (also true for *A*-optimality under some conditions (Schwabe, 1996))

$[\mathbf{r}^{(k)}(x)]^\top \theta^{(k)} \triangleq \sum_{i=1}^{d_k} \theta_i^{(k)} x^i$ polynomial with degree d_k, $\dim(\theta^{(k)}) = p_k = 1 + d_k$

Global model for $\mathbf{x} = (\{\mathbf{x}\}_1, \{\mathbf{x}\}_2, \ldots, \{\mathbf{x}\}_d)^\top$:

$\mathbf{r}^\top(\mathbf{x})\gamma = \prod_{k=1}^{d} [\mathbf{r}^{(k)}(x)]^\top \theta^{(k)}$,

total degree $\sum_{k=1}^{d} d_k$, $\dim(\gamma) = \prod_{k=1}^{d} p_k$

Example:

$$\begin{aligned}
\mathbf{r}^\top(\mathbf{x})\gamma &= (\theta_0^{(1)} + \theta_1^{(1)}\{\mathbf{x}\}_1 + \theta_2^{(1)}\{\mathbf{x}\}_1^2) \times (\theta_0^{(2)} + \theta_1^{(2)}\{\mathbf{x}\}_2 + \theta_2^{(2)}\{\mathbf{x}\}_2^2) \\
&= \gamma_0 + \gamma_1\{\mathbf{x}\}_1 + \gamma_2\{\mathbf{x}\}_2 + \gamma_{12}\{\mathbf{x}\}_1\{\mathbf{x}\}_2 + \gamma_{11}\{\mathbf{x}\}_1{}^2 + \gamma_{22}\{\mathbf{x}\}_2{}^2 \\
&\quad + \gamma_{112}\{\mathbf{x}\}_1{}^2\{\mathbf{x}\}_2 + \gamma_{122}\{\mathbf{x}\}_1\{\mathbf{x}\}_2{}^2 + \gamma_{1122}\{\mathbf{x}\}_1{}^2\{\mathbf{x}\}_2{}^2
\end{aligned}$$

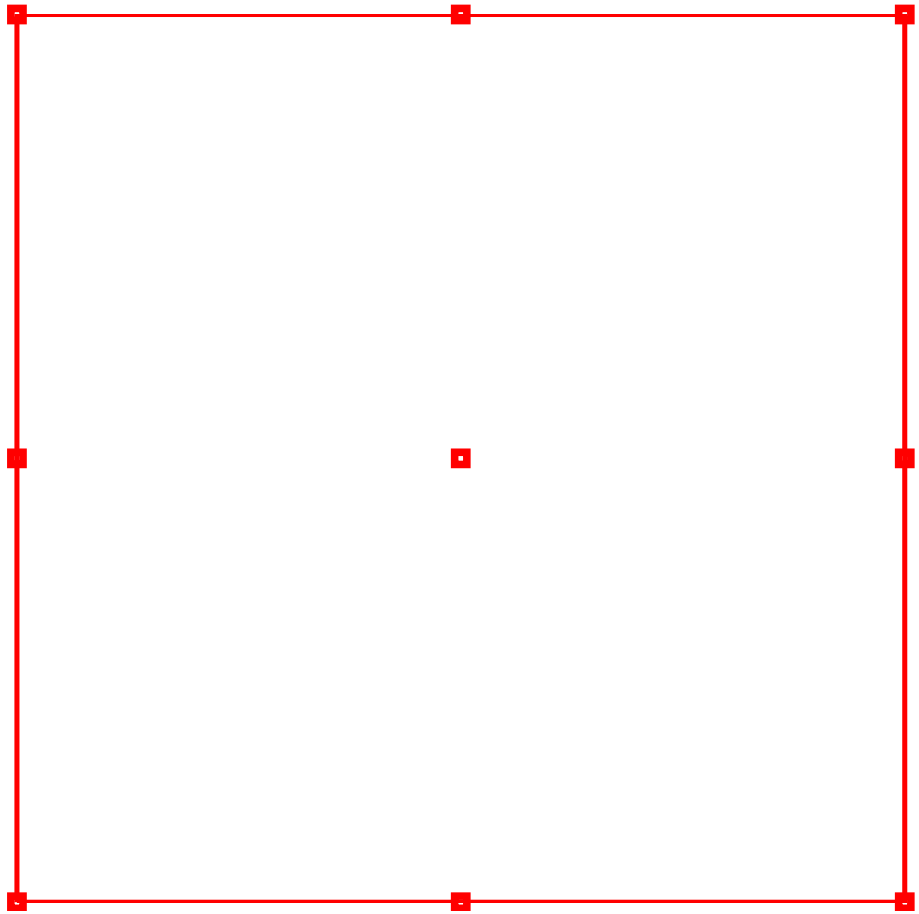
D-optimal design (approximate theory) = tensor product of d one-dimensional *D*-optimal designs

(true for any type of model, not only polynomials)

Polynomial of degree k: D-optimal design supported on $k+1$ points,
(on $[-1,1]$, roots of $(1-t^2)P_k'(t)$, with $P_k(t) \triangleq k$-th Legendre polynomial)
all with equal weight $1/(k+1)$

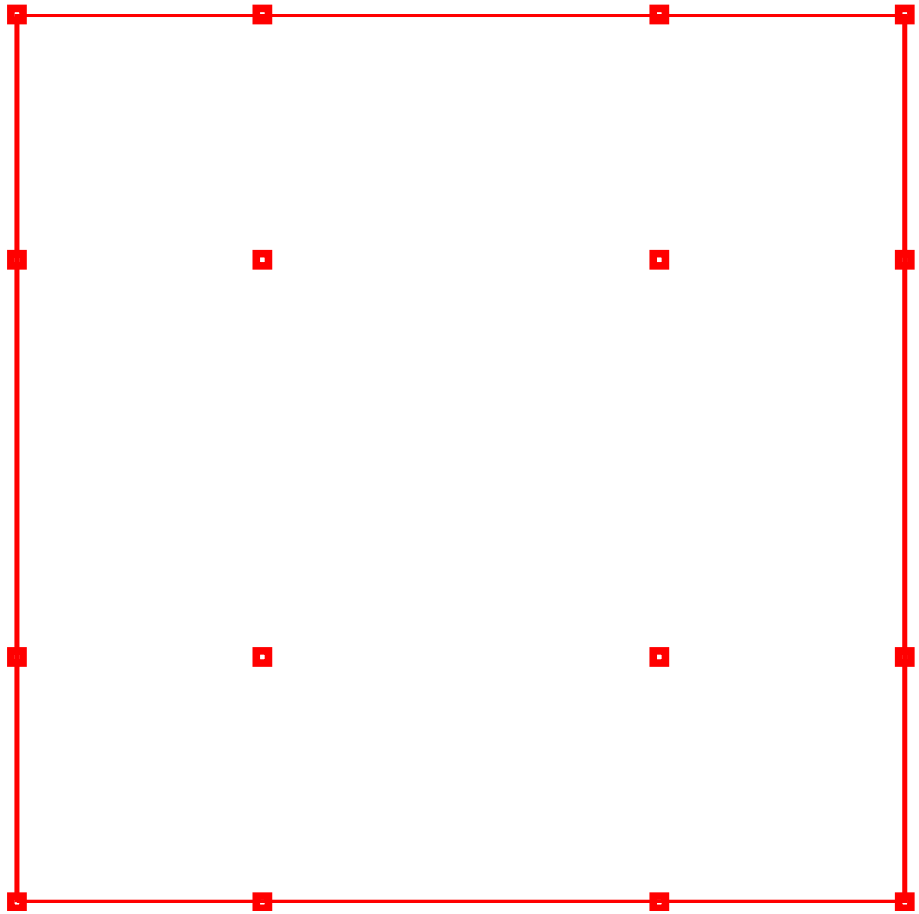
Polynomial of degree k: D-optimal design supported on $k+1$ points,
(on $[-1,1]$, roots of $(1-t^2)P_k'(t)$, with $P_k(t) \triangleq k$-th Legendre polynomial)
all with equal weight $1/(k+1)$

dimension 2, $d_1 = d_2 = 2$
9 points, weights $=1/9$



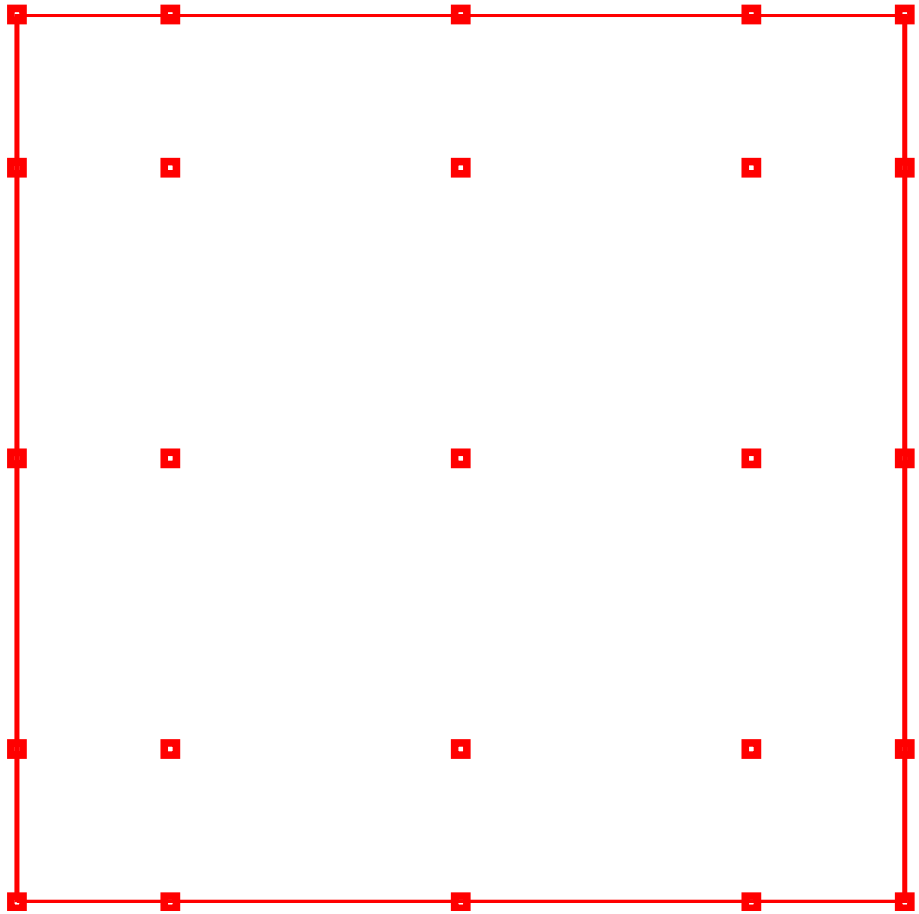
Polynomial of degree k: D-optimal design supported on $k+1$ points,
(on $[-1,1]$, roots of $(1-t^2)P_k'(t)$, with $P_k(t) \triangleq k$-th Legendre polynomial)
all with equal weight $1/(k+1)$

dimension 2, $d_1 = d_2 = 3$
16 points, weights $=1/16$



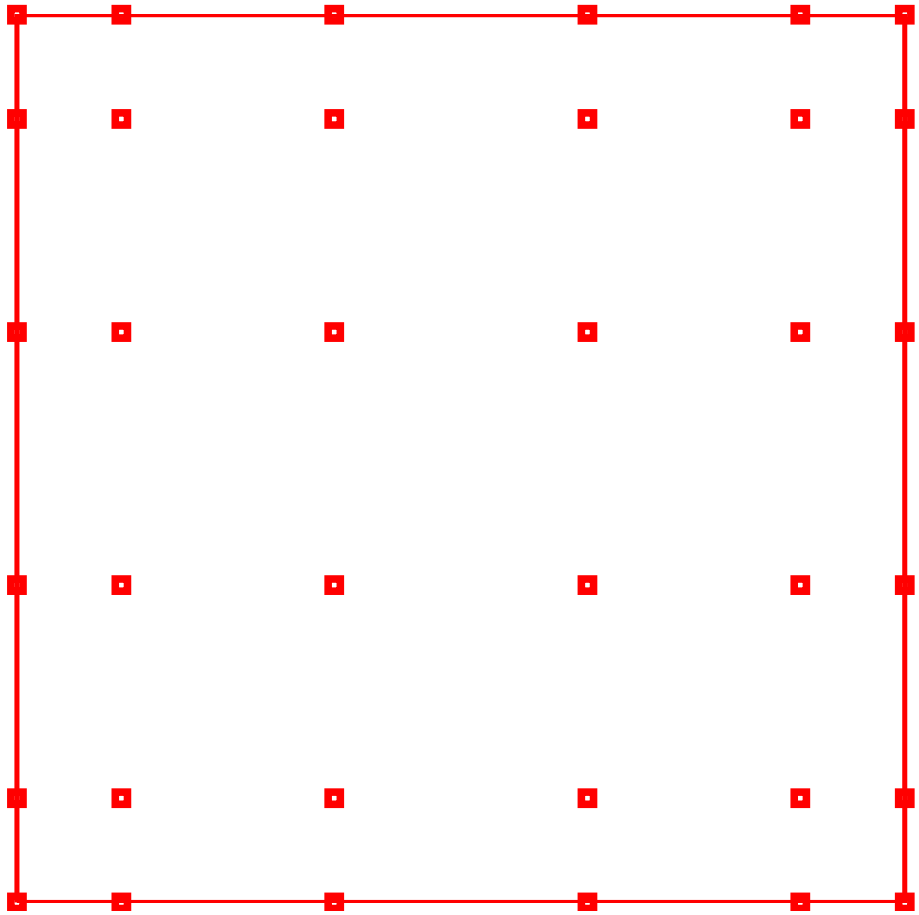
Polynomial of degree k: D-optimal design supported on $k+1$ points,
(on $[-1,1]$, roots of $(1-t^2)P_k'(t)$, with $P_k(t) \triangleq k$-th Legendre polynomial)
all with equal weight $1/(k+1)$

dimension 2, $d_1 = d_2 = 4$
25 points, weights $=1/25$



Polynomial of degree k: D-optimal design supported on $k+1$ points,
(on $[-1,1]$, roots of $(1-t^2)P_k'(t)$, with $P_k(t) \triangleq k$-th Legendre polynomial)
all with equal weight $1/(k+1)$

dimension 2, $d_1 = d_2 = 5$
36 points, weights $=1/36$



Sum of polynomials?

$$\mathbf{r}^\top(\mathbf{x})\gamma = \sum_{k=1}^{d} [\mathbf{r}^{(k)}(x)]^\top \theta^{(k)},$$

total degree $\max_{k=1}^{d} d_k$, $\dim(\gamma) = (\sum_{k=1}^{d} p_k) - 1 = \sum_{k=1}^{d} d_k + d - 1$

Sum of polynomials?

$$\mathbf{r}^\top(\mathbf{x})\gamma = \sum_{k=1}^{d} [\mathbf{r}^{(k)}(x)]^\top \theta^{(k)},$$

total degree $\max_{k=1}^{d} d_k$, $\dim(\gamma) = (\sum_{k=1}^{d} p_k) - 1 = \sum_{k=1}^{d} d_k + d - 1$

Example:

$$\begin{aligned} \mathbf{r}^\top(\mathbf{x})\gamma &= (\theta_0^{(1)} + \theta_1^{(1)}\{\mathbf{x}\}_1 + \theta_2^{(1)}\{\mathbf{x}\}_1^2) + (\theta_0^{(2)} + \theta_1^{(2)}\{\mathbf{x}\}_2 + \theta_2^{(2)}\{\mathbf{x}\}_2^2) \\ &= \gamma_0 + \gamma_1\{\mathbf{x}\}_1 + \gamma_2\{\mathbf{x}\}_2 + \gamma_{11}\{\mathbf{x}\}_1{}^2 + \gamma_{22}\{\mathbf{x}\}_2{}^2 \end{aligned}$$

(no interaction term)

Again, *D*-optimal design (approximate theory) = tensor product of d one-dimensional *D*-optimal designs (Schwabe, 1996)

Sum of polynomials?

$$\mathbf{r}^\top(\mathbf{x})\gamma = \sum_{k=1}^{d} [\mathbf{r}^{(k)}(x)]^\top \theta^{(k)},$$

total degree $\max_{k=1}^{d} d_k$, $\dim(\gamma) = (\sum_{k=1}^{d} p_k) - 1 = \sum_{k=1}^{d} d_k + d - 1$

Example:

$$\begin{aligned} \mathbf{r}^\top(\mathbf{x})\gamma &= (\theta_0^{(1)} + \theta_1^{(1)}\{\mathbf{x}\}_1 + \theta_2^{(1)}\{\mathbf{x}\}_1^2) + (\theta_0^{(2)} + \theta_1^{(2)}\{\mathbf{x}\}_2 + \theta_2^{(2)}\{\mathbf{x}\}_2^2) \\ &= \gamma_0 + \gamma_1\{\mathbf{x}\}_1 + \gamma_2\{\mathbf{x}\}_2 + \gamma_{11}\{\mathbf{x}\}_1{}^2 + \gamma_{22}\{\mathbf{x}\}_2{}^2 \end{aligned}$$

(no interaction term)

Again, *D*-optimal design (approximate theory) = tensor product of d one-dimensional *D*-optimal designs (Schwabe, 1996)

Difficult to apply in big dimension:

d polynomials of degree k ➠ $(k+1)^d$ support points!

but a general lesson, and possible extension towards Gaussian process models and kriging

2.5 Consequences for space-filling design

D-optimality + polynomials ➠ more points close to the boundary as degree increases

Erdös-Turan theorem: roots r of orthonormal polynomials on $[0,1]$ are asymptotically distributed with the arcsine law, with density $\varphi_0(r) = \frac{1}{\pi\sqrt{r\,(1-r)}}$

➠ **Should we put more points close to the boundary ?**

2.5 Consequences for space-filling design

D-optimality + polynomials ➠ more points close to the boundary as degree increases

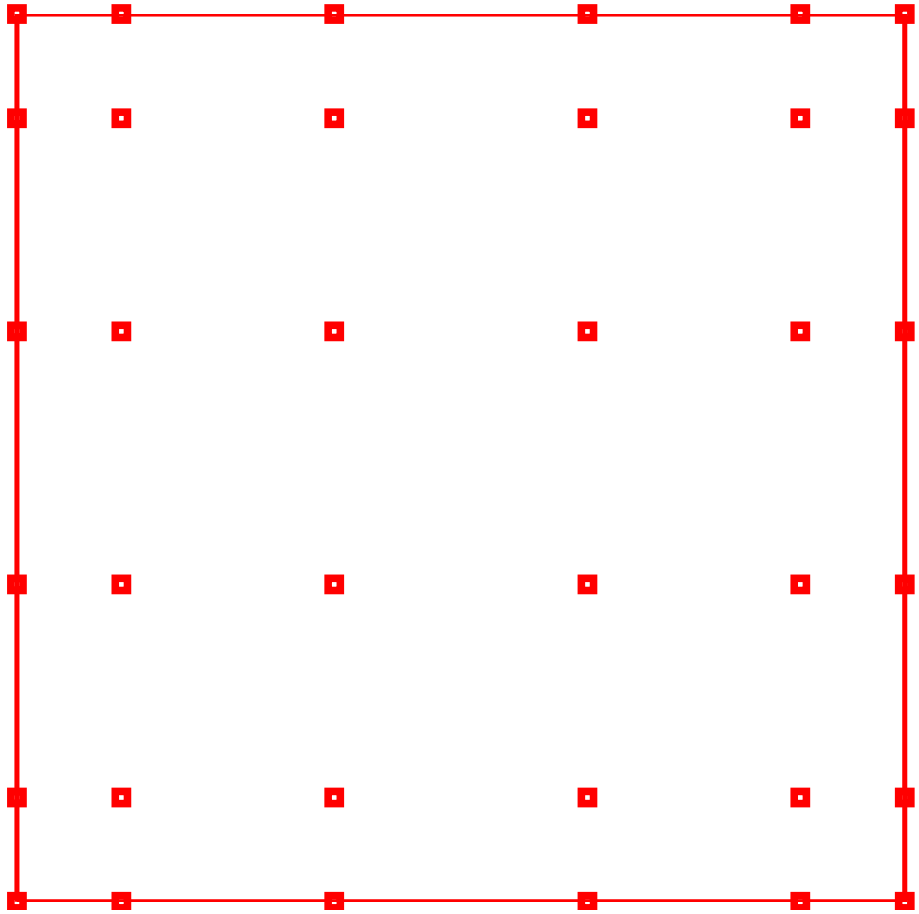
Erdös-Turan theorem: roots r of orthonormal polynomials on $[0,1]$ are asymptotically distributed with the arcsine law, with density $\varphi_0(r) = \frac{1}{\pi\sqrt{r\,(1-r)}}$

➠ **Should we put more points close to the boundary ?**

In order to counter the boundary effect Dette and Pepelyshev (2010):

- Take a "standard" space-filling design (e.g., Maximin, Lh Maximin, LDS),
- for each $j = 1, \ldots, d$, transform j-th coordinates $\{\mathbf{x}_i\}_j$ with $T : x \mapsto z = T(x) = \frac{1+\cos(\pi\, x)}{2}$ ($x \sim$ uniform $\rightarrow z \sim$ arcsine),
- Use the transformed design $\mathbf{Z}_n = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$

dimension 2, 5-degree polynomials:
D-optimal design has 36 points, weights $=1/36$



dimension 2, $n = 36$

Transformed (arcsine) Maximin-optimal design

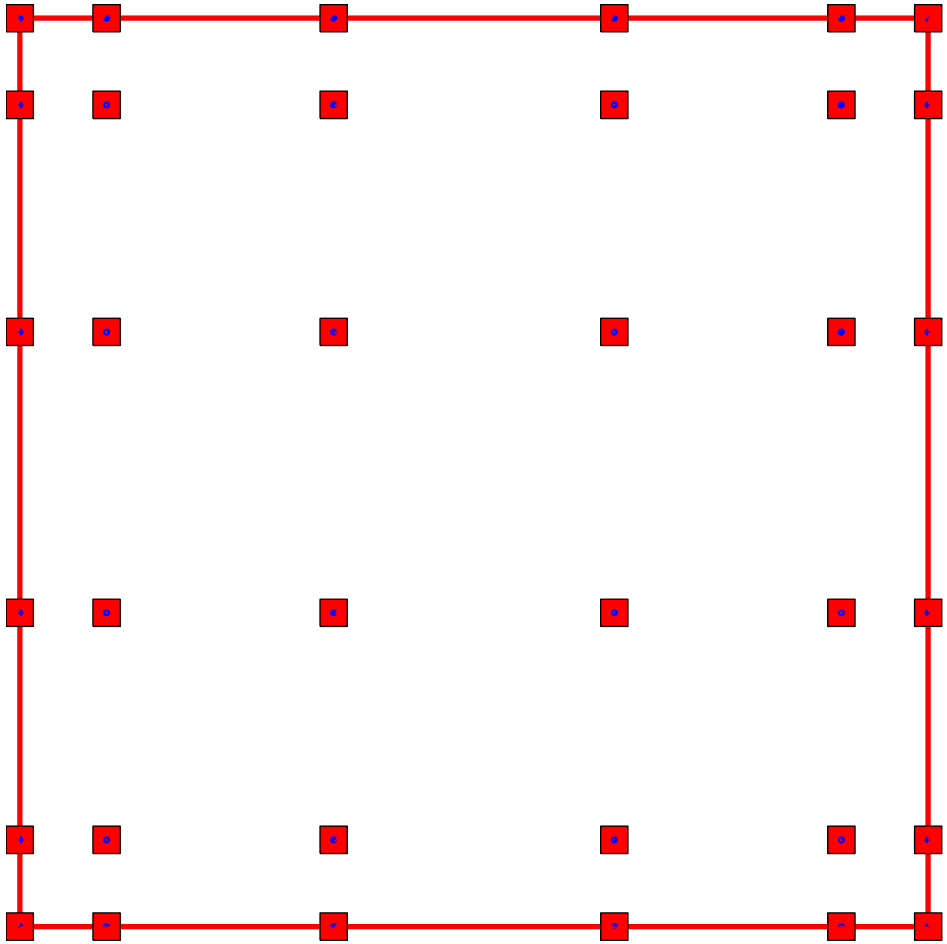


Illustration of boundary effect: $d = 1$, $n = 11$ observations in $[0, 1]$, ordinary kriging with covariance $C(t) = \exp(-50\, t^2)$ ➠ plot of $\rho_n(x)$

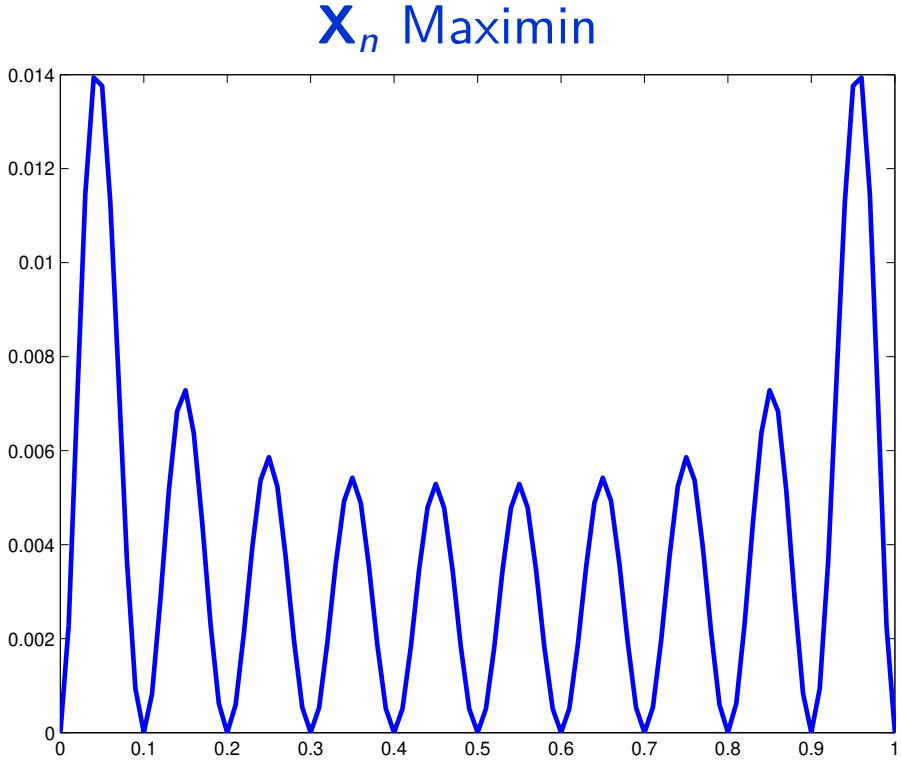
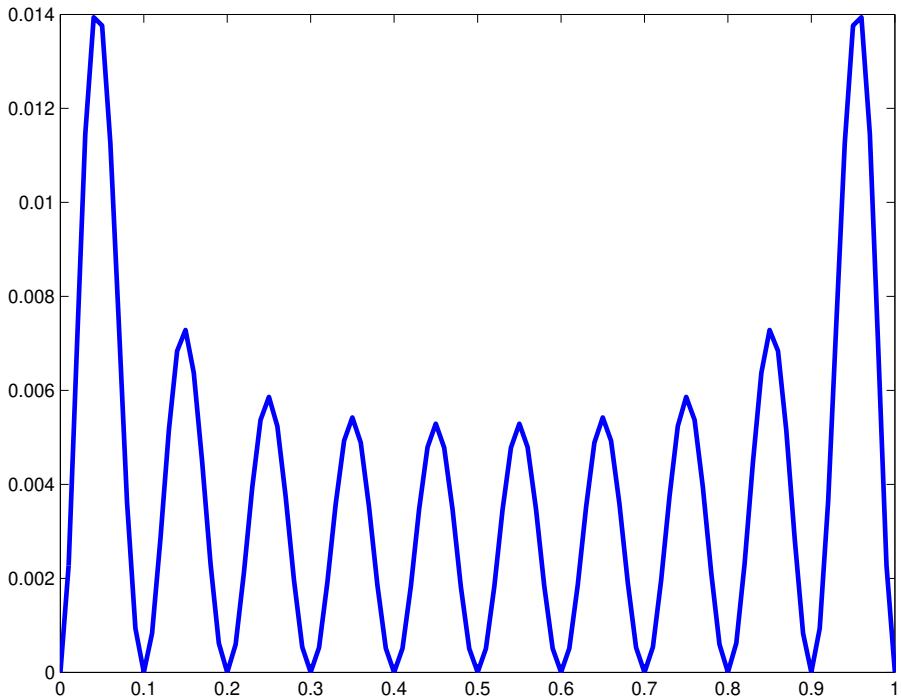
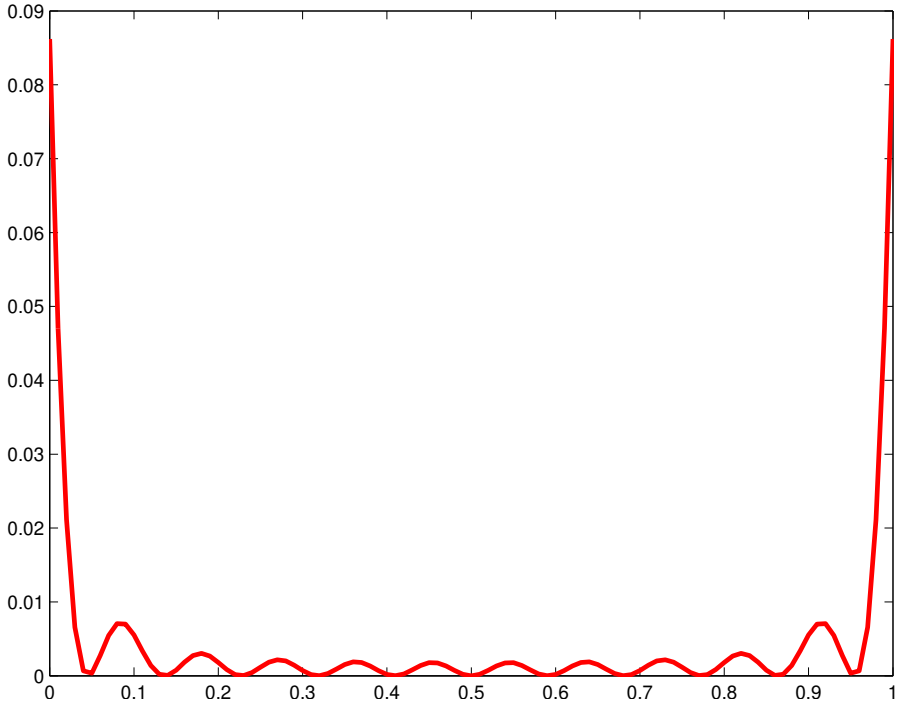


Illustration of boundary effect: $d = 1$, $n = 11$ observations in $[0, 1]$, ordinary kriging with covariance $C(t) = \exp(-50\, t^2)$ ➠ plot of $\rho_n(x)$

$\mathbf{X}_n$ Maximin



$\mathbf{X}_n$ miniMax



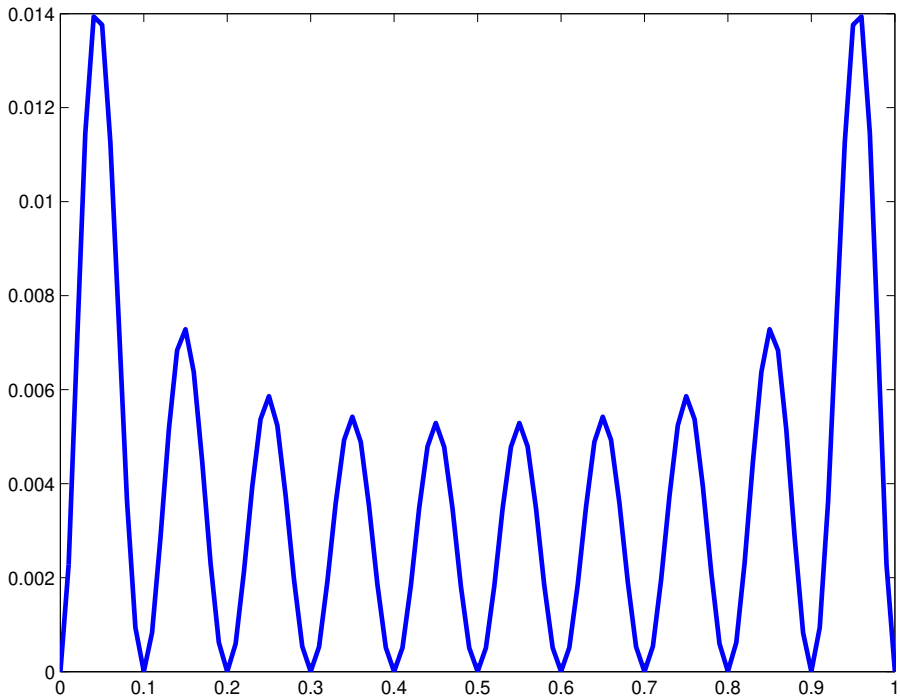
Uniform distribution of design points

$\Rightarrow$ prediction near boundaries relies on less points

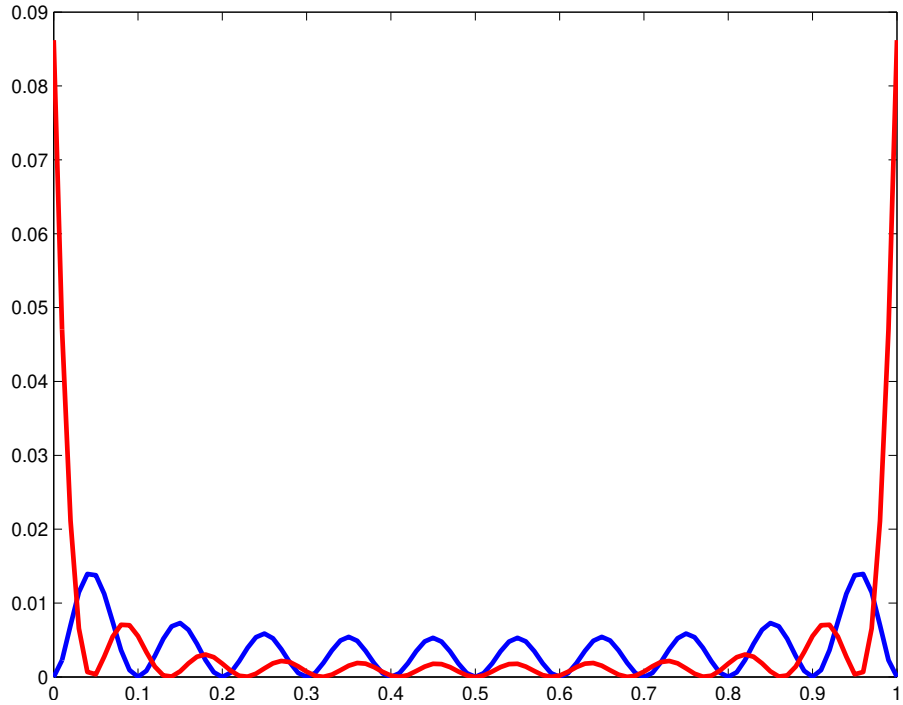
$\Rightarrow$ precision is worse close to boundaries

Illustration of boundary effect: $d = 1$, $n = 11$ observations in $[0, 1]$, ordinary kriging with covariance $C(t) = \exp(-50\, t^2)$ ➠ plot of $\rho_n(x)$

$\mathbf{X}_n$ Maximin



Maximin and miniMax

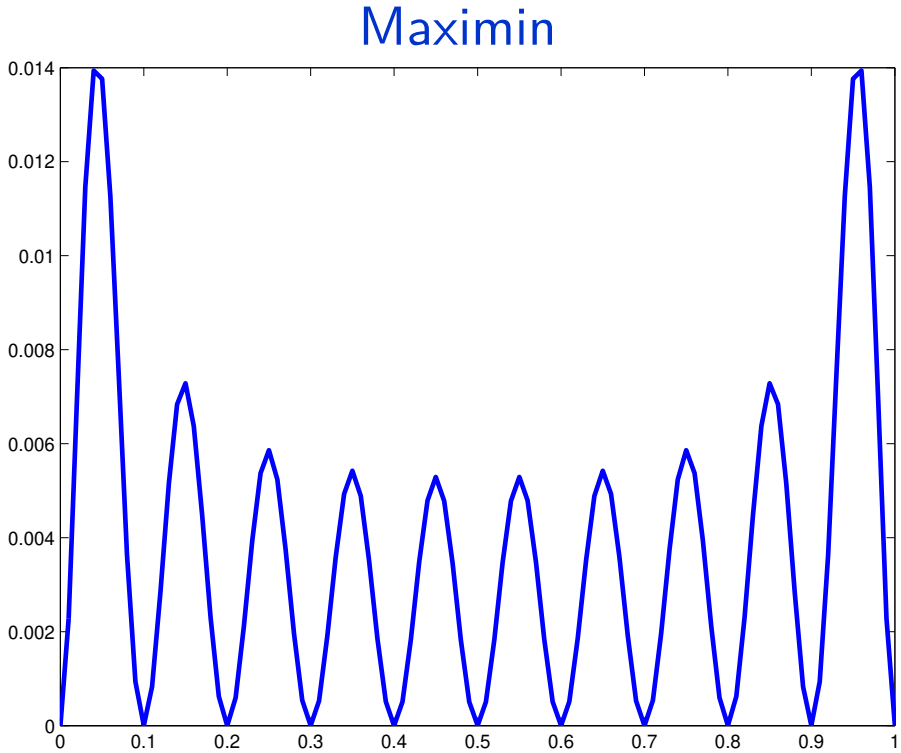


Uniform distribution of design points

$\Rightarrow$ prediction near boundaries relies on less points

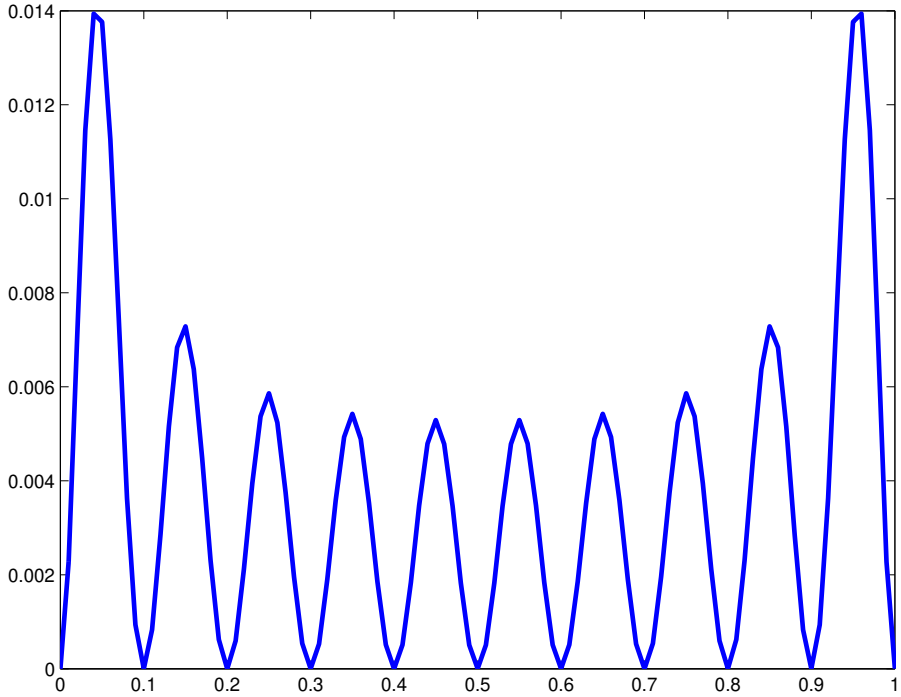
$\Rightarrow$ precision is worse close to boundaries

... But transformation $T : x \mapsto z = T(x) = \frac{1+\cos(\pi\, x)}{2}$ may be too "strong"

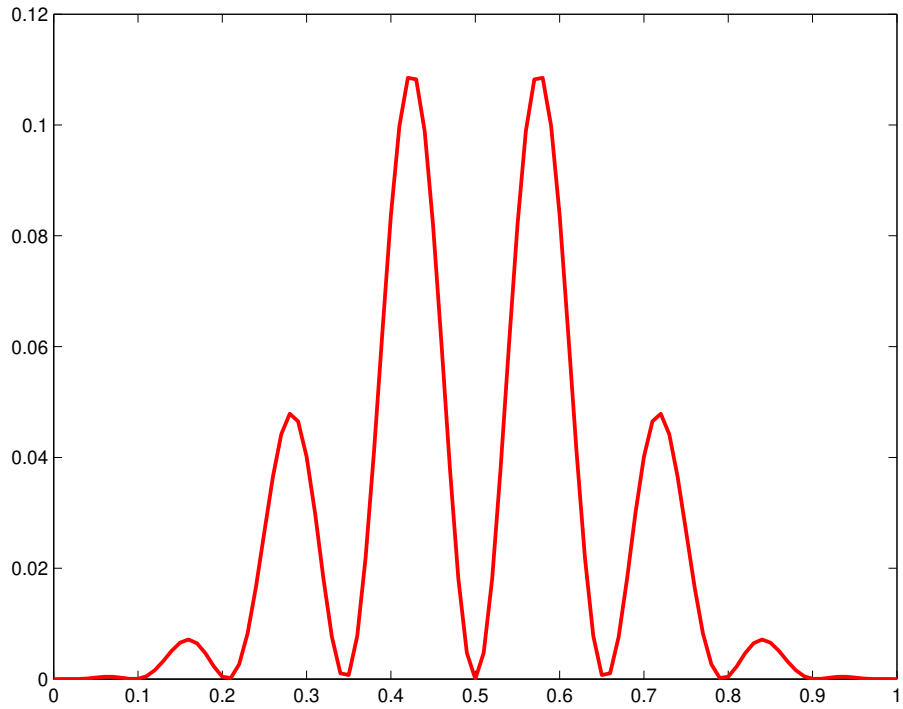


... But transformation $T : x \mapsto z = T(x) = \frac{1+\cos(\pi x)}{2}$ may be too "strong"

Maximin

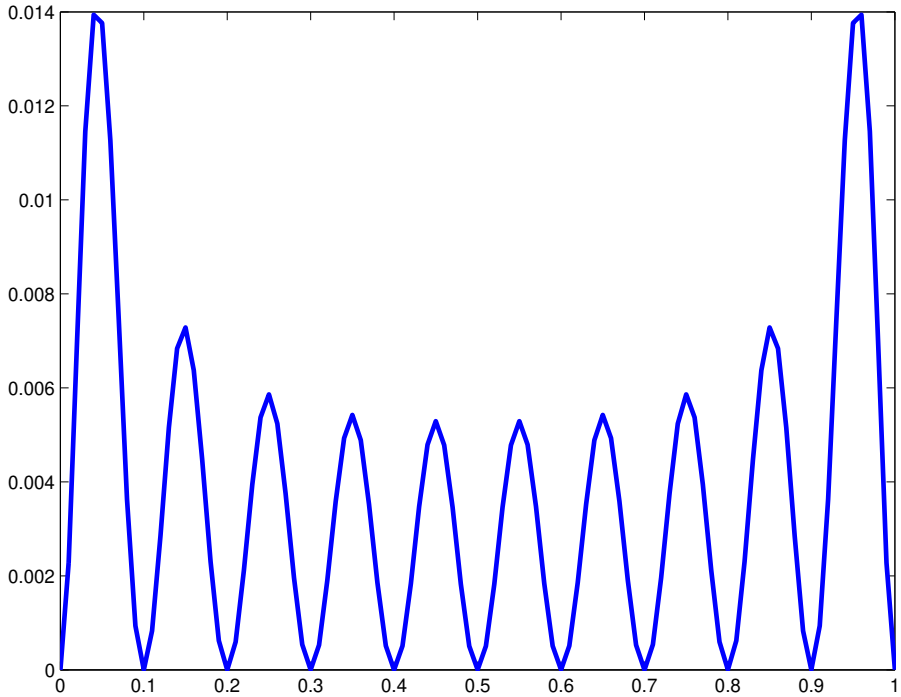


Transformed (arcsine) Maximin

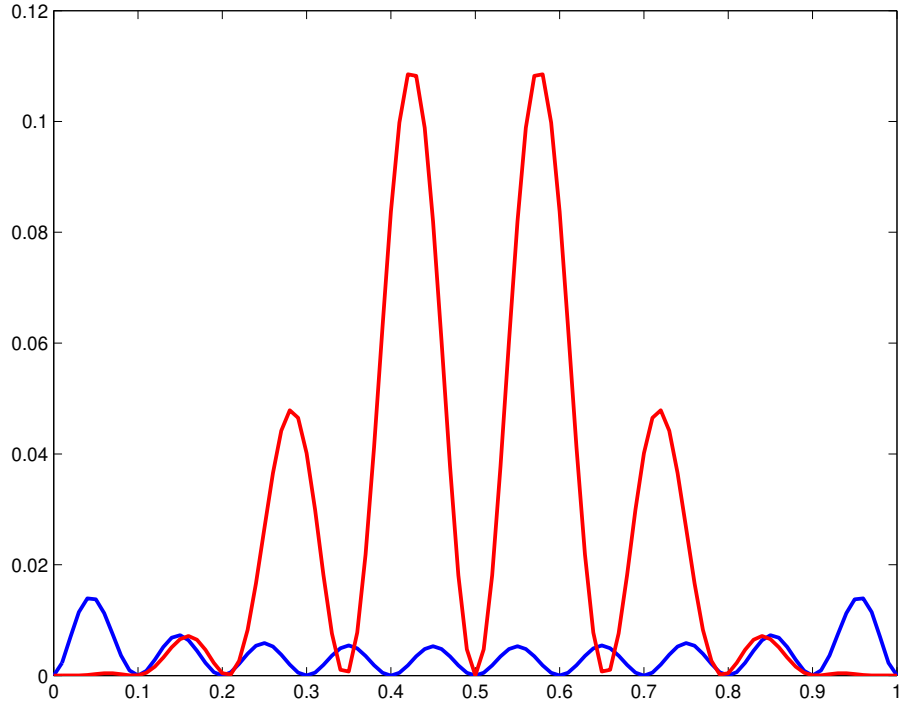


... But transformation $T : x \mapsto z = T(x) = \frac{1+\cos(\pi x)}{2}$ may be too "strong"

Maximin



Maximin + transformed Maximin



Arcsine distribution: maximizes $\tilde{\Phi}_{[0]}(\xi) = \exp\left[\int_0^1 \int_0^1 \log\|x-y\|\,\xi(dx)\,\xi(dy)\right]$

(continuous version of $\overline{\phi}_{[0]}(\mathbf{X}) = \exp\left[\sum_{i<j} \mu_{ij}\,\log(d_{ij})\right]$, see § I-1.6)

Arcsine distribution: maximizes $\tilde{\Phi}_{[0]}(\xi) = \exp\left[\int_0^1 \int_0^1 \log\|x-y\|\,\xi(dx)\,\xi(dy)\right]$

(continuous version of $\overline{\phi}_{[0]}(\mathbf{X}) = \exp\left[\sum_{i<j} \mu_{ij}\,\log(d_{ij})\right]$, see § I-1.6)

➤ Maximization of

$$\tilde{\Phi}_{[q]}(\xi) = \left[\int_0^1 \int_0^1 \|x-y\|^{-q}\,\xi(dx)\,\xi(dy)\right]^{-1/q}, \quad 0 < q < 1$$

(continuous version of $\overline{\phi}_{[q]}(\mathbf{X}) = \left[\sum_{i<j} \mu_{ij}\,d_{ij}^{-q}\right]^{-1/q}$, see § I-1.6) yields a measure ξ with density $\varphi_q(x) = \frac{x^{(q-1)/2}(1-x)^{(q-1)/2}}{B(\frac{q+1}{2},\frac{q+1}{2})}$ (Beta distribution) (Zhigljavsky et al., 2010)

(tends to arcsine when $q \to 0$, to uniform when $q \to 1$)

In order to counter the boundary effect (Dette and Pepelyshev, 2010) :

- Take a "standard" space-filling design (e.g., Maximin, Lh Maximin, LDS),
- for each $j = 1, \ldots, d$, transform j-th coordinates $\{\mathbf{x}_i\}_j$ with $T : x \mapsto z = T(x)$ such that $x = \int_0^z \varphi_q(t)\,\mathrm{d}t$ ($x \sim$ uniform $\rightarrow$ $z \sim \varphi_q$),
- Use the transformed design $\mathbf{Z}_n = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$

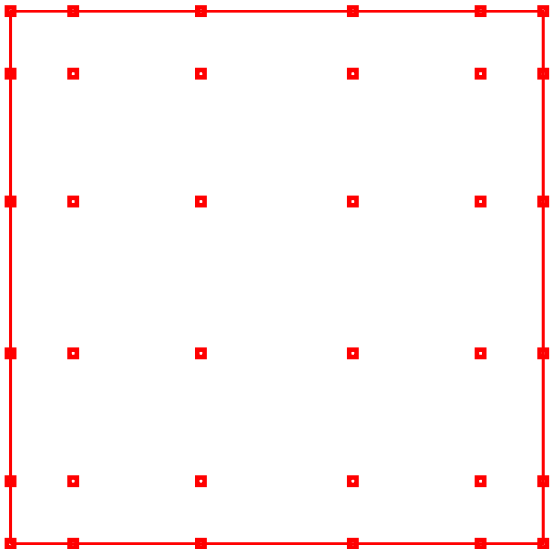
In order to counter the boundary effect (Dette and Pepelyshev, 2010) :

- Take a "standard" space-filling design (e.g., Maximin, Lh Maximin, LDS),
- for each $j = 1, \ldots, d$, transform j-th coordinates $\{\mathbf{x}_i\}_j$ with $T : x \mapsto z = T(x)$ such that $x = \int_0^z \varphi_q(t)\,\mathrm{d}t$ ($x \sim$ uniform $\rightarrow$ $z \sim \varphi_q$),
- Use the transformed design $\mathbf{Z}_n = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$

dimension 2, 5-degree polynomials

D-optimal design:

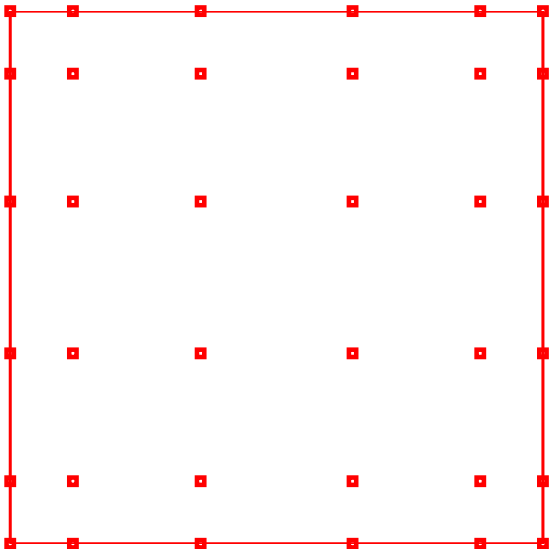
36 points, weights $=1/36$



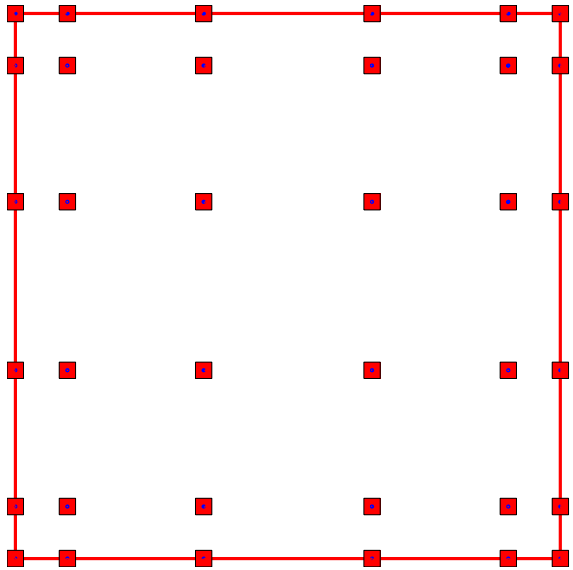
In order to counter the boundary effect (Dette and Pepelyshev, 2010) :

- Take a "standard" space-filling design (e.g., Maximin, Lh Maximin, LDS),
- for each $j = 1, \ldots, d$, transform j-th coordinates $\{\mathbf{x}_i\}_j$ with $T : x \mapsto z = T(x)$ such that $x = \int_0^z \varphi_q(t)\,\mathrm{d}t$ ($x \sim$ uniform $\rightarrow$ $z \sim \varphi_q$),
- Use the transformed design $\mathbf{Z}_n = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$

dimension 2, 5-degree polynomials
D-optimal design:
36 points, weights $=1/36$



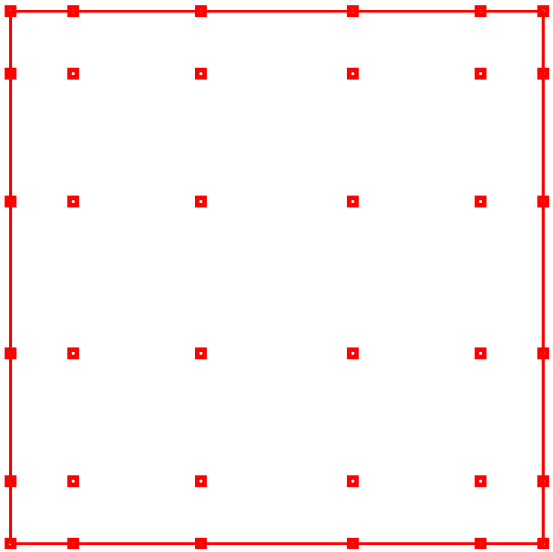
dimension 2, $n = 36$
Transformed (arcsine)
Maximin-optimal design



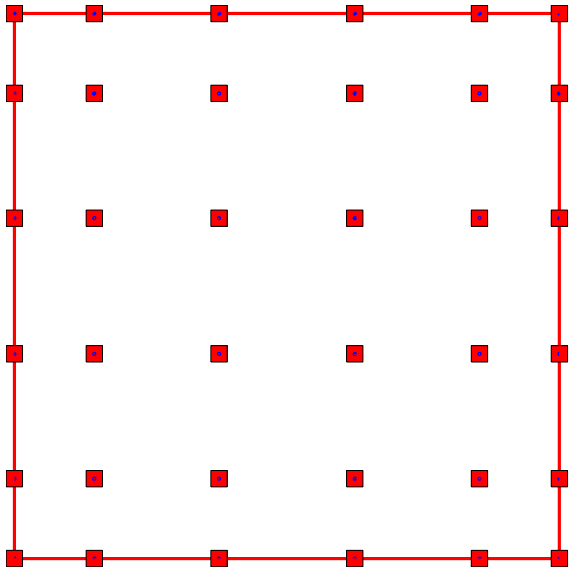
In order to counter the boundary effect (Dette and Pepelyshev, 2010) :

- Take a "standard" space-filling design (e.g., Maximin, Lh Maximin, LDS),
- for each $j = 1, \ldots, d$, transform j-th coordinates $\{\mathbf{x}_i\}_j$ with $T : x \mapsto z = T(x)$ such that $x = \int_0^z \varphi_q(t)\,\mathrm{d}t$ ($x \sim$ uniform $\rightarrow$ $z \sim \varphi_q$),
- Use the transformed design $\mathbf{Z}_n = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$

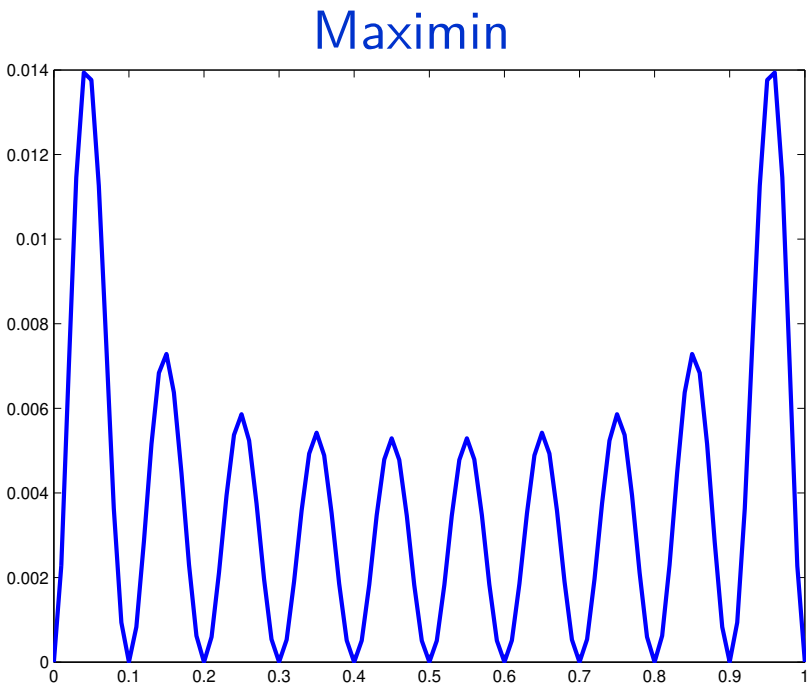
dimension 2, 5-degree polynomials
D-optimal design:
36 points, weights $=1/36$



dimension 2, $n = 36$
Transformed (Beta, $q = 0.4$)
Maximin-optimal design

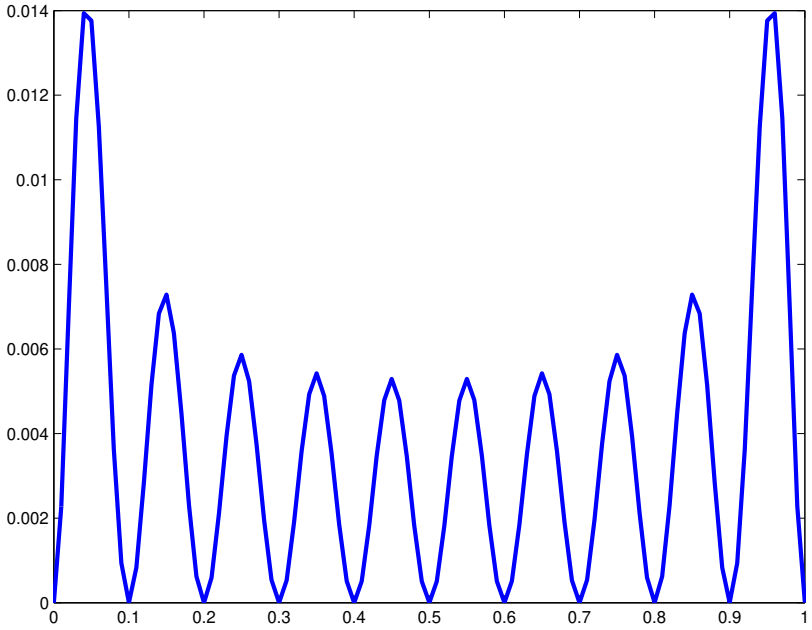


. . . For a suitably chosen Beta-transformation ($q = 0.84$)

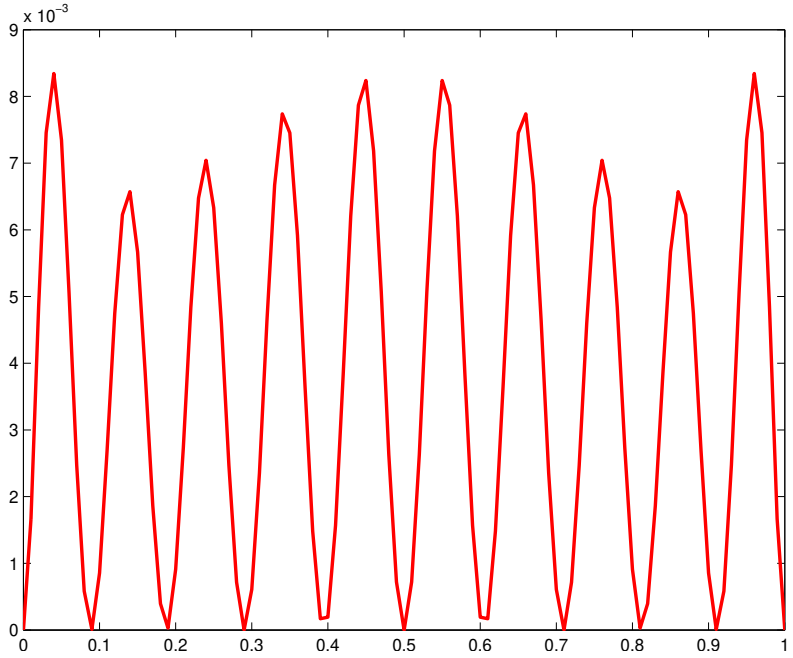


. . . For a suitably chosen Beta-transformation ($q = 0.84$)

Maximin

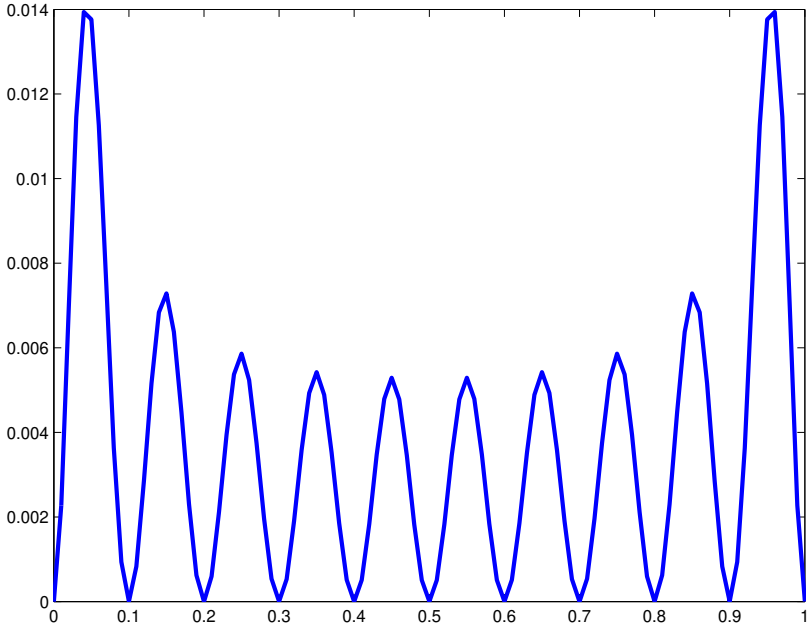


Transformed (Beta) Maximin

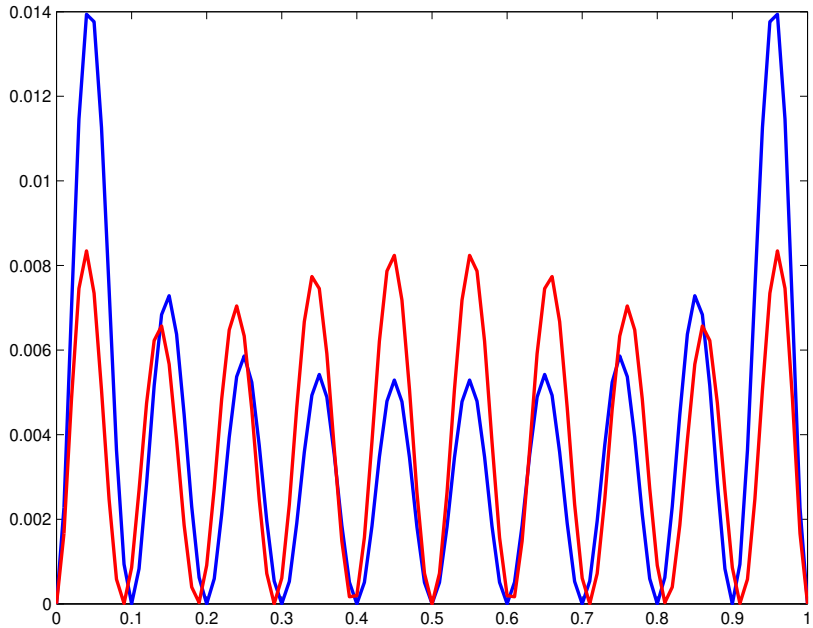


. . . For a suitably chosen Beta-transformation ($q = 0.84$)

Maximin



Maximin + transformed Maximin



Choice of a suitable q?

➠ Optimize a precision criterion based on $\rho_n(\mathbf{x})$
(depends on covariance $C(\cdot)$)

▲ requires $\mathscr{X}$ = hypercube ▲

▲ if d is big, many points are on the boundary (or at the vertices!) of $\mathscr{X}$ ▲

3 Optimal design for Bayesian prediction

3.1 Karhunen-Loève decomposition of a Gaussian process

Model without trend: $f(\mathbf{x}) = Z(\mathbf{x})$, Gaussian process

$\mathrm{E}\{Z(\mathbf{x})\} = 0$, $\mathrm{E}\{Z(\mathbf{x})Z(\mathbf{x}')\} = C(\mathbf{x}, \mathbf{x}')$ $(= C(\mathbf{x} - \mathbf{x}')$ if stationary)

$$\mathrm{IMSE}_\mu(\mathbf{X}_n) \triangleq \int_{\mathscr{X}} \mathrm{E}\left\{[Z(\mathbf{x}) - \mathrm{E}\{Z(\mathbf{x})|\mathbf{y}_n\}]^2\right\} \mathrm{d}\mu(\mathbf{x})$$

3 Optimal design for Bayesian prediction

3.1 Karhunen-Loève decomposition of a Gaussian process

Model without trend: $f(\mathbf{x}) = Z(\mathbf{x})$, Gaussian process
$\mathrm{E}\{Z(\mathbf{x})\} = 0$, $\mathrm{E}\{Z(\mathbf{x})Z(\mathbf{x}')\} = C(\mathbf{x}, \mathbf{x}')$ $(= C(\mathbf{x} - \mathbf{x}')$ if stationary)

$$\boxed{\mathrm{IMSE}_\mu(\mathbf{X}_n) \triangleq \int_{\mathscr{X}} \mathrm{E}\left\{[Z(\mathbf{x}) - \mathrm{E}\{Z(\mathbf{x})|\mathbf{y}_n\}]^2\right\} \mathrm{d}\mu(\mathbf{x})}$$

The integral operator T_μ defined by
$\forall f \in L^2(\mathscr{X}, \mu),\ \forall \mathbf{x} \in \mathscr{X},\quad T_\mu[f](\mathbf{x}) = \int_{\mathscr{X}} f(\mathbf{x}')K(\mathbf{x}, \mathbf{x}')\mathrm{d}\mu(\mathbf{x}')$
is diagonalisable:

eigenvalues λ_i, $i = 1, 2, 3 \ldots$ (in $\searrow$ order)
associated eigenfunctions $\varphi_i(\cdot)$ (extended over $\mathscr{X}$), with
$\int_{\mathscr{X}} \varphi_i(\mathbf{x})\varphi_j(\mathbf{x})\mathrm{d}\mu(\mathbf{x}) = \delta_{ij}$

$$\boxed{Z'(\mathbf{x}) \triangleq P_{\mathbb{H}_\mu}[Z_{\mathbf{x}}] = \sum_i \zeta_i \sqrt{\lambda_i} \varphi_i(\mathbf{x})}$$

with all ζ_i i.i.d. $\mathcal{N}(0,1)$

$P_{\mathbb{H}_\mu}$ = projection $\perp$ on the space "which contributes to IMSE$_\mu$"

$Z'(\mathbf{x}) = \sum_i \gamma_i \varphi_i(\mathbf{x})$ where the r.v. γ_i are independent $\mathcal{N}(0, \lambda_i)$

$$\boxed{Z'(\mathbf{x}) \triangleq P_{\mathbb{H}_\mu}[Z_{\mathbf{x}}] = \sum_i \zeta_i \sqrt{\lambda_i} \varphi_i(\mathbf{x})}$$

with all ζ_i i.i.d. $\mathcal{N}(0,1)$

$P_{\mathbb{H}_\mu}$ = projection $\perp$ on the space "which contributes to IMSE$_\mu$"

$Z'(\mathbf{x}) = \sum_i \gamma_i \varphi_i(\mathbf{x})$ where the r.v. γ_i are independent $\mathcal{N}(0, \lambda_i)$

For a given truncation level m,

$$\begin{aligned} Z'(\mathbf{x}) &= \sum_{i=1}^{m} \gamma_i \varphi_i(\mathbf{x}) + \sum_{i>m} \gamma_i \varphi_i(\mathbf{x}) \\ &\simeq \boxed{Z''(\mathbf{x}) = \sum_{i=1}^{m} \gamma_i \varphi_i(\mathbf{x}) + \varepsilon(\mathbf{x})} \end{aligned}$$

with $\mathsf{E}\{\varepsilon(\mathbf{x}_i)\} = 0$, $\mathsf{E}\{\varepsilon(\mathbf{x}_i)\varepsilon(\mathbf{x}_j)\} = \sigma^2 \delta_{ij}$ et $\sigma^2 = \sum_{i>m} \lambda_i$

$$\boxed{Z''(\mathbf{x}_i) = \phi^\top(\mathbf{x}_i)\gamma + \varepsilon_i}$$

= linear regression model (as in § 2.1)
(with eigenfunctions $\varphi_i(\cdot)$, $i = 1, \ldots, m$, instead of polynomials)

(Fedorov, 1996) ➠ construct an optimal design for this model

3.2 Bayesian prediction for $Z''(\mathbf{x}_i) = \phi^\top(\mathbf{x}_i)\gamma + \varepsilon_i$

LS estimation:

$$\hat{\gamma}_n = (\Phi_n^\top \Phi_n)^{-1}\Phi_n^\top \mathbf{y}_n, \text{ with } \mathbf{y}_n = (y_1, \ldots, y_n)^\top \text{ and } \Phi_n = \begin{pmatrix} \phi^\top(\mathbf{x}_1) \\ \vdots \\ \phi^\top(\mathbf{x}_n) \end{pmatrix}$$

$$\text{cov}(\hat{\gamma}_n) = \sigma^2(\Phi_n^\top \Phi_n)^{-1} = \frac{\sigma^2}{n}\left[\underbrace{\frac{1}{n}\sum_{i=1}^{n}\phi(\mathbf{x}_i)\phi^\top(\mathbf{x}_i)}_{\mathbf{M}_n}\right]^{-1}$$

3.2 Bayesian prediction for $Z''(\mathbf{x}_i) = \phi^\top(\mathbf{x}_i)\gamma + \varepsilon_i$

LS estimation:

$$\hat{\gamma}_n = (\Phi_n^\top \Phi_n)^{-1}\Phi_n^\top \mathbf{y}_n, \text{ with } \mathbf{y}_n = (y_1, \ldots, y_n)^\top \text{ and } \Phi_n = \begin{pmatrix} \phi^\top(\mathbf{x}_1) \\ \vdots \\ \phi^\top(\mathbf{x}_n) \end{pmatrix}$$

$$\text{cov}(\hat{\gamma}_n) = \sigma^2(\Phi_n^\top \Phi_n)^{-1} = \frac{\sigma^2}{n}\left[\underbrace{\frac{1}{n}\sum_{i=1}^{n}\phi(\mathbf{x}_i)\phi^\top(\mathbf{x}_i)}_{\mathbf{M}_n}\right]^{-1}$$

Prediction at $\mathbf{x}$: $\eta_n(\mathbf{x}) = \phi^\top(\mathbf{x})\hat{\gamma}_n$

$$\text{IMSE}(\mathbf{X}_n) = \int_{\mathscr{X}} \phi^\top(\mathbf{x})\text{cov}(\hat{\gamma}_n)\phi(\mathbf{x})\,\mathrm{d}\mu(\mathbf{x}) = \frac{\sigma^2}{n}\text{trace}[\mathbf{M}_n^{-1}]$$

$= A$-optimality criterion

(requires $n \geq m$ to have a full rank $\mathbf{M}_n$)

Bayesian estimation: prior distribution $\mathcal{N}(\mathbf{0}, \Lambda_m)$ for γ,
with $\Lambda_m = \mathrm{diag}\{\lambda_1, \ldots, \lambda_m\}$

$$\hat{\gamma}_n = [\Phi_n^\top \Phi_n/\sigma^2 + \Lambda_m^{-1}]^{-1}[\Phi_n^\top \mathbf{y}_n/\sigma^2]$$

$$\mathrm{cov}(\hat{\gamma}_n) = [\Phi_n^\top \Phi_n/\sigma^2 + \Lambda_m^{-1}]^{-1} = \frac{\sigma^2}{n} \underbrace{\left[\underbrace{\frac{1}{n}\sum_{i=1}^{n} \phi(\mathbf{x}_i)\phi^\top(\mathbf{x}_i)}_{\mathbf{M}_n} + \frac{\sigma^2}{n}\Lambda_m^{-1} \right]^{-1}}_{\mathbf{M}_B(\mathbf{x}_n)}$$

Bayesian estimation: prior distribution $\mathcal{N}(\mathbf{0}, \Lambda_m)$ for γ,
with $\Lambda_m = \mathrm{diag}\{\lambda_1, \ldots, \lambda_m\}$

$$\hat{\gamma}_n = [\Phi_n^\top \Phi_n/\sigma^2 + \Lambda_m^{-1}]^{-1}[\Phi_n^\top \mathbf{y}_n/\sigma^2]$$

$$\mathrm{cov}(\hat{\gamma}_n) = [\Phi_n^\top \Phi_n/\sigma^2 + \Lambda_m^{-1}]^{-1} = \frac{\sigma^2}{n} \underbrace{\left[\underbrace{\frac{1}{n}\sum_{i=1}^{n} \phi(\mathbf{x}_i)\phi^\top(\mathbf{x}_i)}_{\mathbf{M}_n} + \frac{\sigma^2}{n}\Lambda_m^{-1} \right]^{-1}}_{\mathbf{M}_B(\mathbf{x}_n)}$$

Prediction at $\mathbf{x}$: $\eta_n(\mathbf{x}) = \phi^\top(\mathbf{x})\hat{\gamma}_n$

$$\mathrm{IMSE}(\mathbf{X}_n) = \int_{\mathcal{X}} \phi^\top(\mathbf{x})\mathrm{cov}(\hat{\gamma}_n)\phi(\mathbf{x})\,\mathrm{d}\mu(\mathbf{x}) = \frac{\sigma^2}{n}\mathrm{trace}[\mathbf{M}_B^{-1}(\mathbf{X}_n)]$$

$= A$-optimality criterion applied to $\mathbf{M}_B(\mathbf{X}_n)$

($\mathbf{M}_B(\mathbf{X}_n)$ has full rank for any m!)

3.3 IMSE-optimal design

All the machinery of optimal design for parametric models is available (Pilz, 1983)

Exact design: Spöck and Pilz (2010) for prediction of spatial random fields (but no guaranteed convergence to the optimum)

3.3 IMSE-optimal design

All the machinery of optimal design for parametric models is available (Pilz, 1983)

Exact design: Spöck and Pilz (2010) for prediction of spatial random fields (but no guaranteed convergence to the optimum)

Approximate design: minimize $\Psi(\xi) = \text{trace}[\mathbf{M}_B^{-1}(\xi)]$, with ξ a probability measure over $\mathscr{X}$ and

$$\mathbf{M}_B(\xi) = \int \phi(\mathbf{x})\phi^\top(\mathbf{x})\,\xi(\mathrm{d}\mathbf{x}) + \frac{\sigma^2}{n}\Lambda_m^{-1}$$

➤ guaranteed convergence towards an optimal ξ^*, with N^* support points

In practice: eigen-decomposition

- use a finite Q-point set $\mathscr{X}_Q = \{\mathbf{x}^{(k)}, \ldots, \mathbf{x}^{(Q)}\}$
- diagonalize $\mathbf{QW}$ where

$$\{\mathbf{Q}\}_{k\ell} = C(\mathbf{x}^{(k)}, \mathbf{x}^{(\ell)}),\ \mathbf{W} = \mathrm{diag}\{w_1, \ldots, w_Q\}$$

($w_k = 1/Q$ when μ uniform)

$$\rightarrow \mathbf{QW} = \mathbf{P}\Lambda\mathbf{P}^{-1} \text{ with } \mathbf{P}^\top\mathbf{WP} = \mathbf{I}_Q \text{ and}$$

$$\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_Q\} \text{ with } \lambda_1 \geq \cdots \geq \lambda_Q$$

In practice: eigen-decomposition

- use a finite Q-point set $\mathscr{X}_Q = \{\mathbf{x}^{(k)}, \ldots, \mathbf{x}^{(Q)}\}$
- diagonalize $\mathbf{QW}$ where

$$\{\mathbf{Q}\}_{k\ell} = C(\mathbf{x}^{(k)}, \mathbf{x}^{(\ell)}),\ \mathbf{W} = \text{diag}\{w_1, \ldots, w_Q\}$$

($w_k = 1/Q$ when μ uniform)

$$\rightarrow \mathbf{QW} = \mathbf{P}\Lambda\mathbf{P}^{-1} \text{ with } \mathbf{P}^\top\mathbf{WP} = \mathbf{I}_Q \text{ and}$$

$$\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_Q\} \text{ with } \lambda_1 \geq \cdots \geq \lambda_Q$$

$$\mathbf{M}_B(\xi) = \sum_{k=1}^{m} p_k \phi(\mathbf{x}_k)\phi^\top(\mathbf{x}_k) + \frac{\sum_{i=m+1}^{Q} \lambda_i}{n} \text{diag}\{\lambda_1^{-1}, \ldots, \lambda_m^{-1}\},\ m < Q$$

where $p_k = \xi\{\mathbf{x}_k\}$ and $\{\phi(\mathbf{x}_k)\}_j = \varphi_j(\mathbf{x}_k) = \mathbf{P}_{kj}$, $k = 1, \ldots, m$, $j = 1, \ldots, m$

In practice: eigen-decomposition

➤ use a finite Q-point set $\mathscr{X}_Q = \{\mathbf{x}^{(k)}, \ldots, \mathbf{x}^{(Q)}\}$

➤ diagonalize $\mathbf{QW}$ where

$\{\mathbf{Q}\}_{k\ell} = C(\mathbf{x}^{(k)}, \mathbf{x}^{(\ell)})$, $\mathbf{W} = \mathrm{diag}\{w_1, \ldots, w_Q\}$

($w_k = 1/Q$ when μ uniform)

➔ $\mathbf{QW} = \mathbf{P}\Lambda\mathbf{P}^{-1}$ with $\mathbf{P}^\top\mathbf{WP} = \mathbf{I}_Q$ and

$\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_Q\}$ with $\lambda_1 \geq \cdots \geq \lambda_Q$

$$\mathbf{M}_B(\xi) = \sum_{k=1}^{m} p_k \phi(\mathbf{x}_k)\phi^\top(\mathbf{x}_k) + \frac{\sum_{i=m+1}^{Q} \lambda_i}{n} \mathrm{diag}\{\lambda_1^{-1}, \ldots, \lambda_m^{-1}\}, \quad m < Q$$

where $p_k = \xi\{\mathbf{x}_k\}$ and $\{\phi(\mathbf{x}_k)\}_j = \varphi_j(\mathbf{x}_k) = \mathbf{P}_{kj}$, $k = 1, \ldots, m$, $j = 1, \ldots, m$

➤ minimization of $\mathrm{trace}[\mathbf{M}_B^{-1}(\xi)]$ ➔ ξ^*

$p_k^* = 0$ for many k, but some $p_k^* > 0$ are very small,

there may exist clusters of points, etc.

➔ aggregate support points of ξ^*

➔ remove some points (transfer their weights on others, optimally)

(Gauthier & P., 2016)

The number N of points is not totally controlled, but 2 tuning parameters are available: m (truncation level) and n (take $m \approx n \approx N$)

N points ➠ initialization for optimization of the true IMSE($\mathbf{X}_N$) by any standard algorithm (optimal points remain in the convex hull of $\mathscr{X}$)

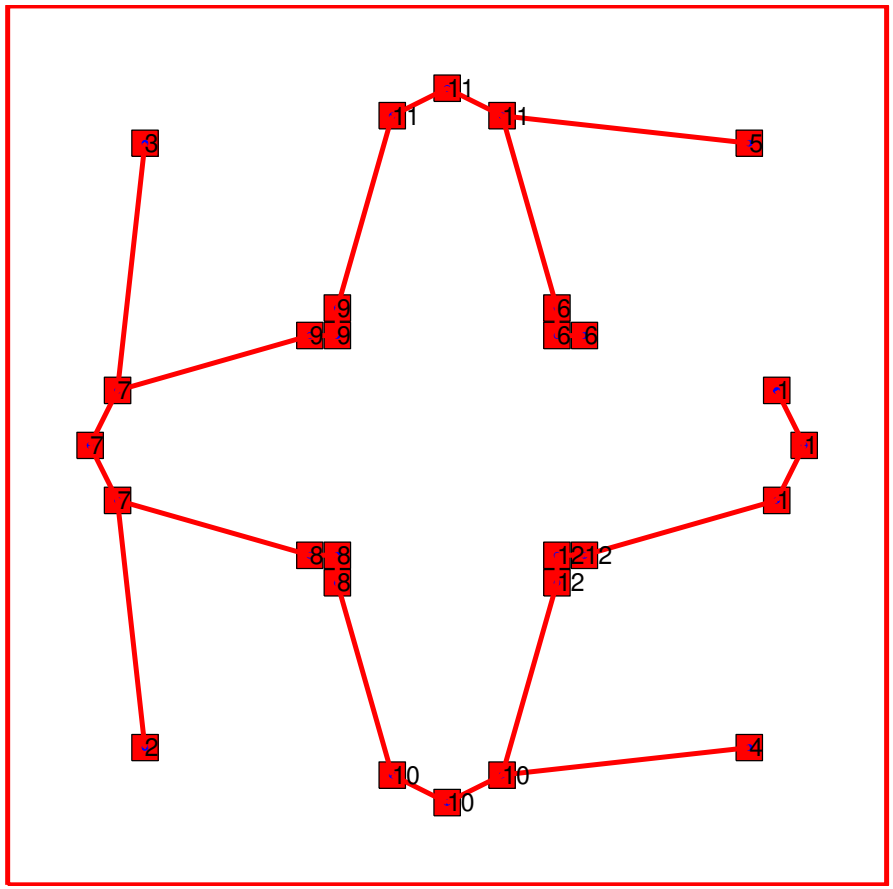
Example :

$d = 2$, $C(\mathbf{x}, \mathbf{x}') = (1 + 10\|\mathbf{x} - \mathbf{x}'\|)\exp(-10\|\mathbf{x} - \mathbf{x}'\|)$ (Matérn 3/2)

$\mathscr{X}_Q$ = regular grid with $Q = 33 \times 33 = 1\,089$ points, $\sigma^2 = \sum_{i>m} \lambda_i$

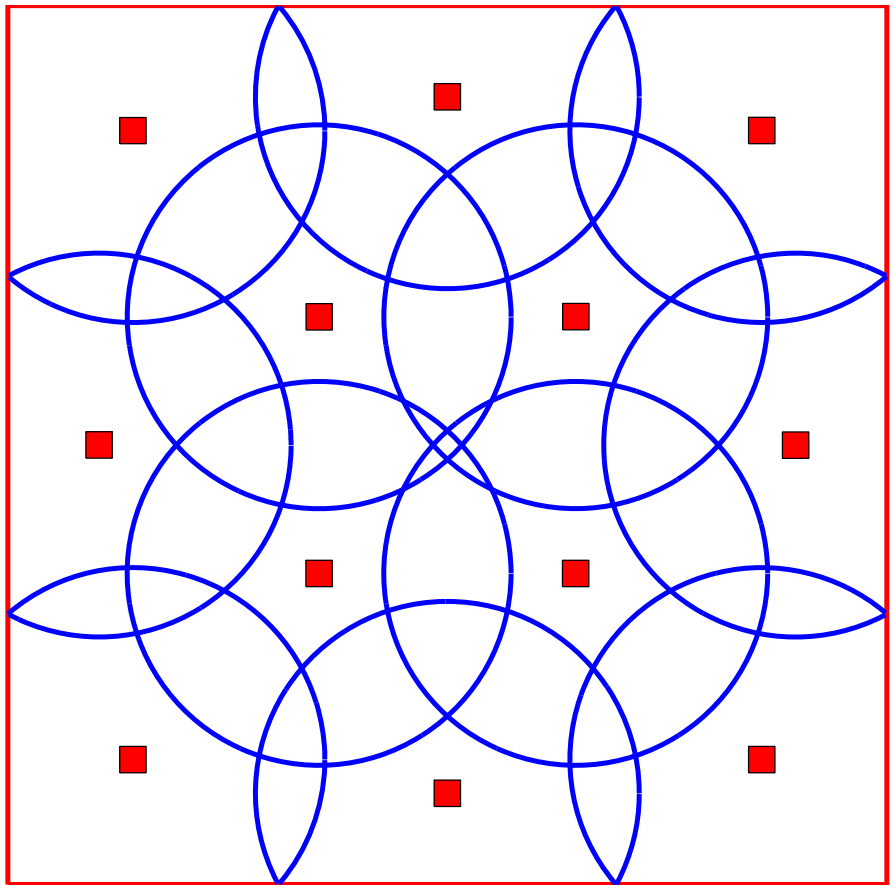
$m = n = 10$

aggregation of support of ξ^*



$m = n = 10$

➠ $N = 12$

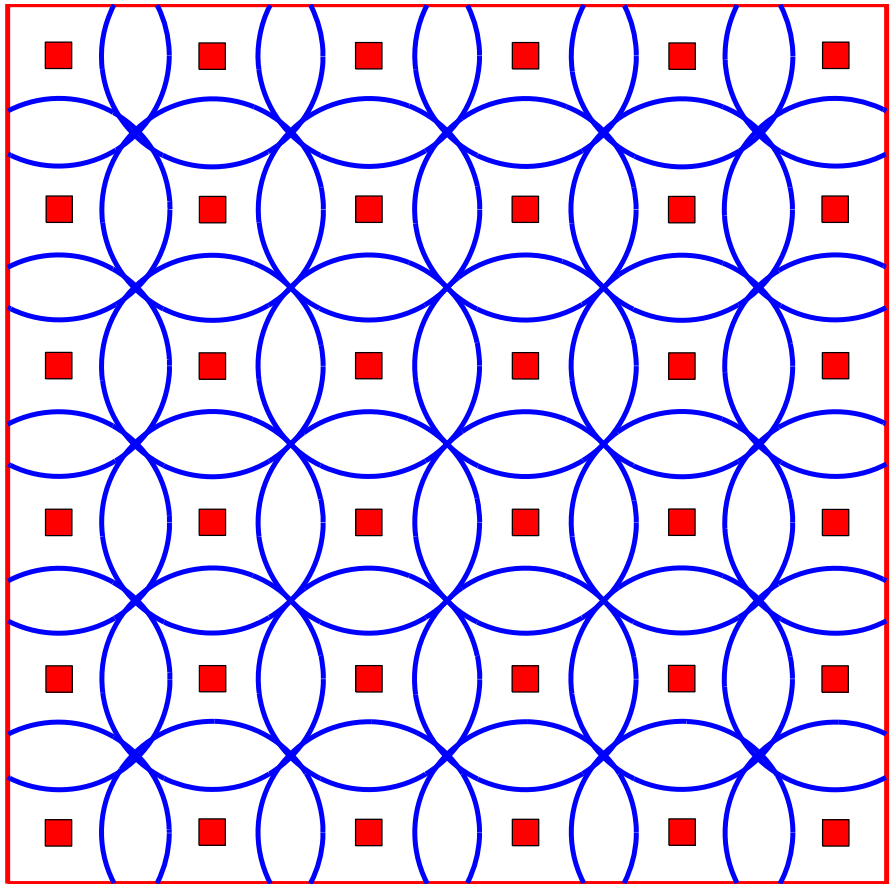


Example :

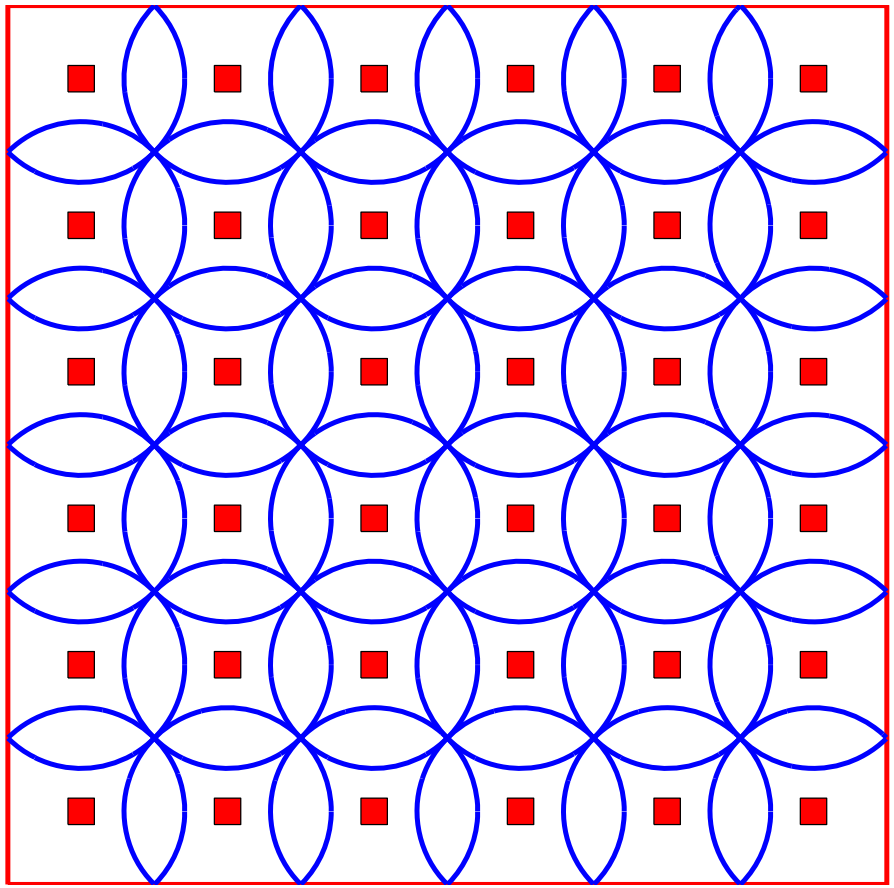
$d = 2$, $C(\mathbf{x}, \mathbf{x}') = (1 + 10\|\mathbf{x} - \mathbf{x}'\|)\exp(-10\|\mathbf{x} - \mathbf{x}'\|)$ (Matérn 3/2)

$\mathscr{X}_Q$ = regular grid with $Q = 33 \times 33 = 1\,089$ points, $\sigma^2 = \sum_{i>m} \lambda_i$

$m = 30$, $n = 10$ ➠ $N = 36$



$\mathbf{X}_{36}$ miniMax optimal



Can be used for any $\mathscr{X}$ if d not too big: use a finite $\mathscr{X}_Q$ given by first Q points of a LDS in $\mathscr{X}$

Can be used for any $\mathscr{X}$ if d not too big: use a finite $\mathscr{X}_Q$ given by first Q points of a LDS in $\mathscr{X}$

With trend, $f(\mathbf{x}) = Z(\mathbf{x}) + \mathbf{r}^\top(\mathbf{x})\beta$?

Same thing (Spöck and Pilz, 2010), with slightly more complicated expressions

Can be used for any $\mathscr{X}$ if d not too big: use a finite $\mathscr{X}_Q$ given by first Q points of a LDS in $\mathscr{X}$

With trend, $f(\mathbf{x}) = Z(\mathbf{x}) + \mathbf{r}^\top(\mathbf{x})\beta$?

Same thing (Spöck and Pilz, 2010), with slightly more complicated expressions

More advanced: avoid mixing eigenfunctions $\varphi_i(\cdot)$ with trend components $\{\mathbf{r}\}_j(\cdot)$ (Gauthier & P., 2016)

⏩ Conclusions part (2)

4 Beyond space filling

Optimal design for kriging: there is a hidden difficulty
the value of θ in covariance $C(\cdot;\theta)$ is unknown

➠ use the same observations to estimate θ and then construct $\eta_n(x)$ with $\hat{\theta}^n$ estimated (plug-in method)

4 Beyond space filling

Optimal design for kriging: there is a hidden difficulty
the value of θ in covariance $C(\cdot;\theta)$ is unknown

➠ use the same observations to estimate θ and then construct $\eta_n(x)$ with $\hat{\theta}^n$ estimated (plug-in method)

➠ In particular, we may use $\hat{\theta}^n$ = Maximum Likelihood Estimator (MLE) ($Z(\mathbf{x})$ is supposed to be Gaussian)

➠ there is a corrective term (Harville and Jeske, 1992; Abt, 1999) :

$$\hat{\rho}_n(\mathbf{x};\theta) = \rho_n(\mathbf{x};\theta) + \mathrm{trace}\{\mathbf{M}_\theta^{-1} \frac{\partial \mathbf{v}_n(\mathbf{x};\theta)}{\partial \theta} \mathbf{C}_n \frac{\partial \mathbf{v}_n(\mathbf{x};\theta)}{\partial \theta^\top}\}$$

(= Empirical Kriging (EK) variance)

avec :

$\mathbf{v}_n(\mathbf{x};\theta)$ such that $\eta_n(\mathbf{x}) = \mathbf{v}_n^\top(\mathbf{x};\theta)\mathbf{y}_n$

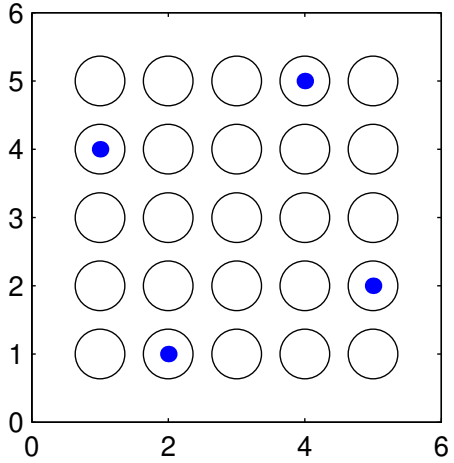
$\mathbf{M}_\theta$ = Fisher Information Matrix (FIM) for θ

Example (Zimmerman, 2006):

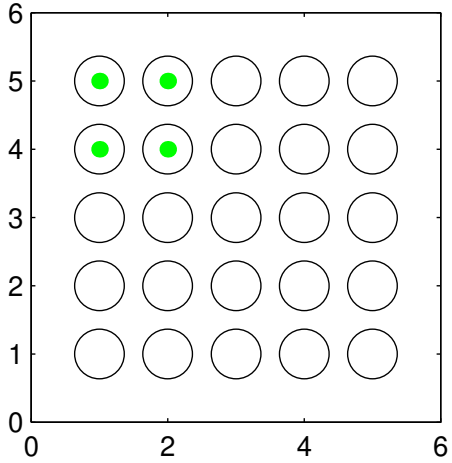
$\mathrm{E}\{Z(\mathbf{x})Z(\mathbf{x}')\} = \theta^{\|\mathbf{x}-\mathbf{x}'\|}$, $\theta = 0.3$

$\mathscr{X}$ = regular grid 5×5, optimal designs for:

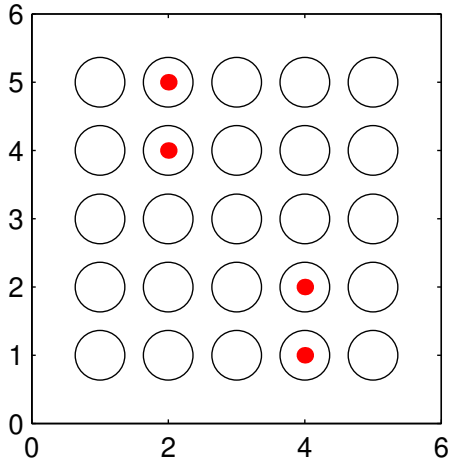
prediction for θ known



estimation of θ



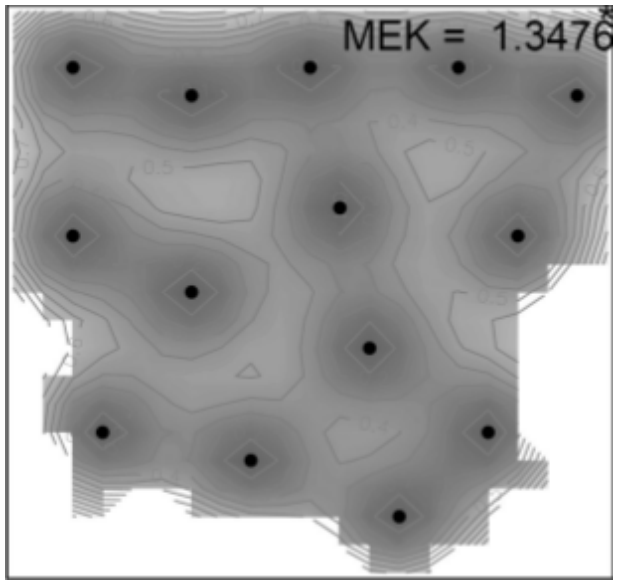
prediction with θ estimated



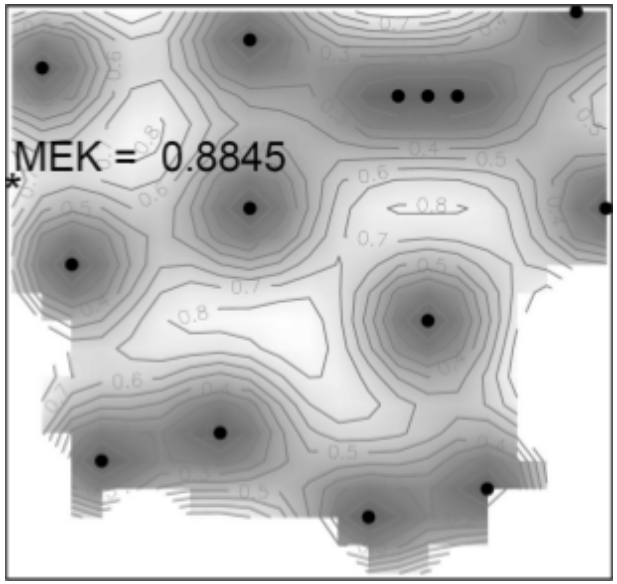
prediction with known θ:
$\mathbf{X}_4$ minimizes $\max_{\mathbf{x}\in\mathscr{X}} \rho_4(\mathbf{x})$
estimation of θ:
$\mathbf{X}_4$ maximizes $\det \mathbf{M}_\theta$
prediction with θ estimated:
$\mathbf{X}_4$ minimizes MEK=$\max_{\mathbf{x}\in\mathscr{X}} \hat{\rho}_4(\mathbf{x};\theta)$

Example (Müller et al., 2015):
NH4 concentration in north sea (collaboration with MUMM, Belgium) — simulated data, kriging with Matérn 3/2 kernel

$\hat{\rho}_{14}(\mathbf{x};\theta)$ for miniMax-optimal design $\mathbf{X}^*_{mM,n=14}$



$\hat{\rho}_{14}(\mathbf{x};\theta)$ for $\mathbf{X}^*_{14}$ minimizing MEK$=\max_{\mathbf{x}\in\mathscr{X}} \hat{\rho}_{14}(\mathbf{x};\theta)$



Choosing $\mathbf{X}_n$ that minimizes MEK$=\max_{\mathbf{x}\in\mathscr{X}} \hat{\rho}_n(\mathbf{x};\theta)$ is difficult

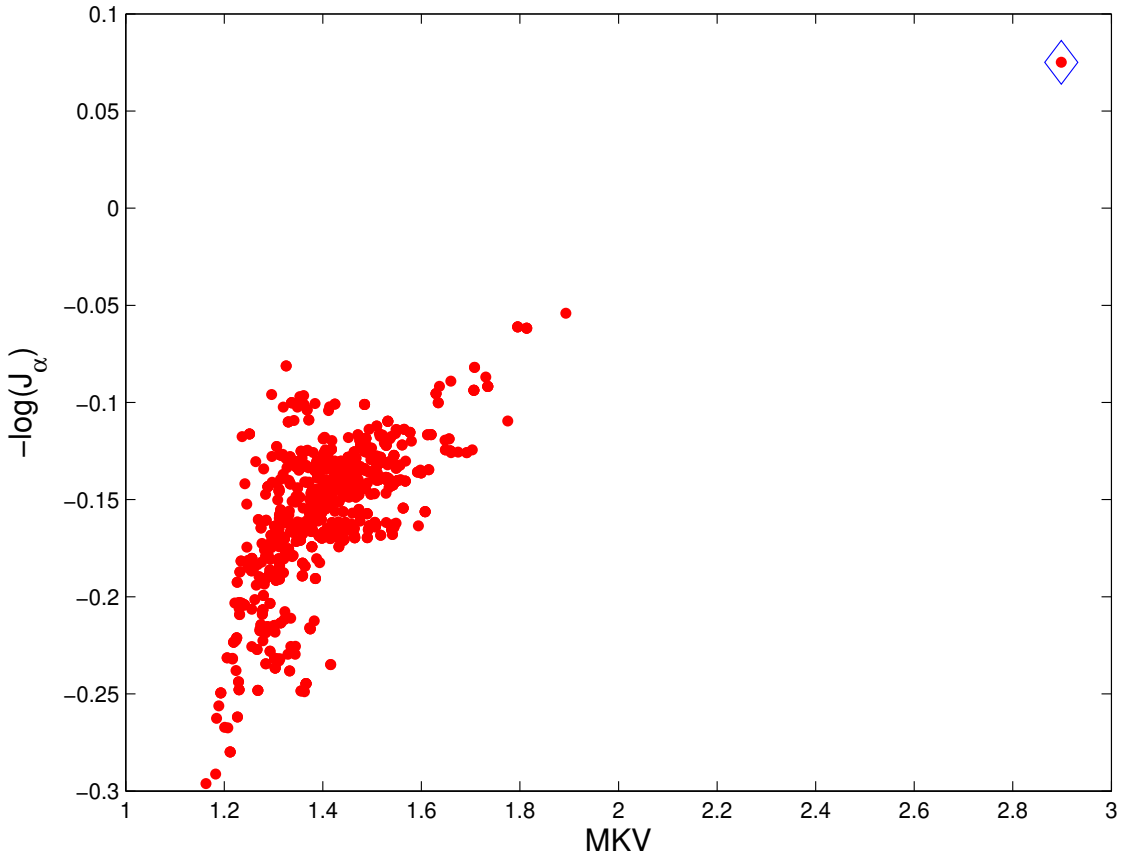
Choosing $\mathbf{X}_n$ that minimizes MEK$=\max_{\mathbf{x}\in\mathscr{X}} \hat{\rho}_n(\mathbf{x};\theta)$ is difficult

➠ Use a compromise criterion between space filling and "aggregation of points"

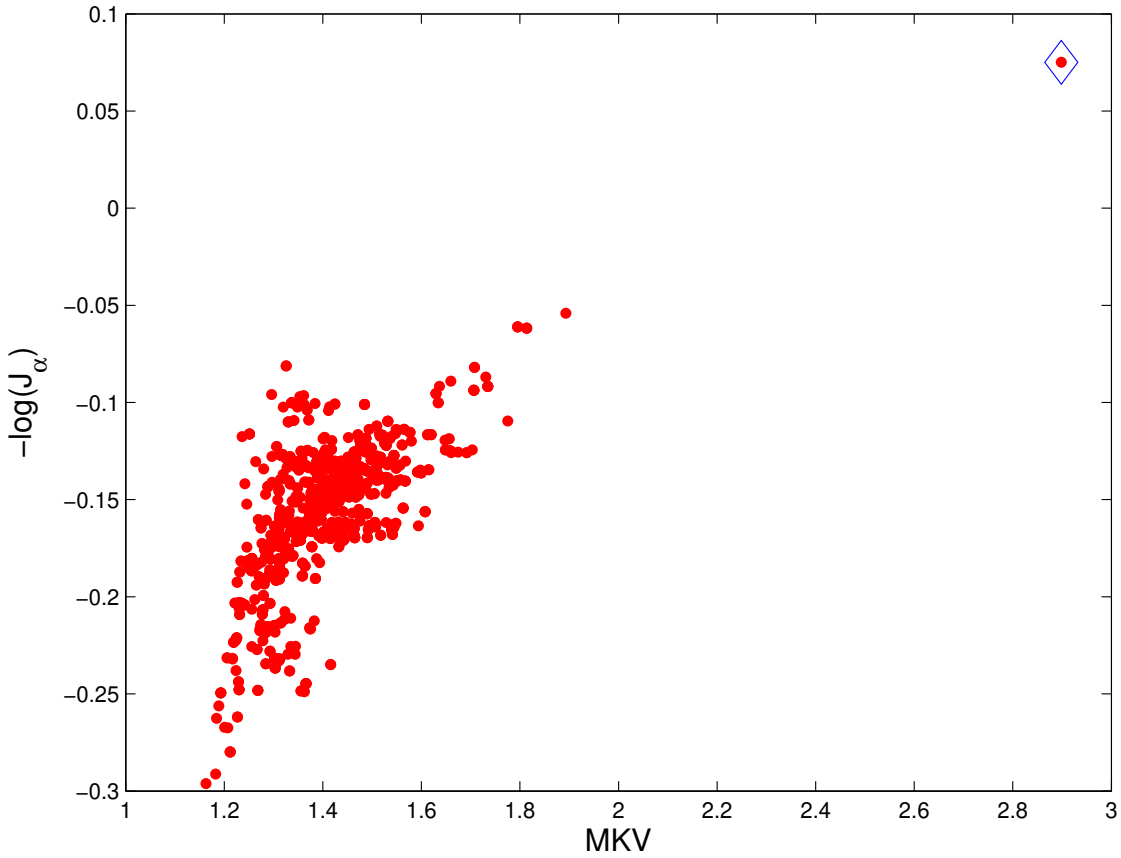
for instance, take $\mathbf{X}_n$ that maximizes $\gamma \log\det(\mathbf{M}_\beta) + (1-\gamma)\log\det(\mathbf{M}_\theta)$ (Müller et al., 2011, 2015), with

- $\mathbf{M}_\beta$= FIM for trend parameters β (maximization $\rightarrow$ space filling design)
- $\mathbf{M}_\theta$= FIM for correlation parameters θ (maximization $\rightarrow$ aggregation)

Example: $n = 7$, $d = 2$, $C(\mathbf{x} - \mathbf{x}'; \theta) = \exp(-\theta \|\mathbf{x} - \mathbf{x}'\|)$, $\theta = 0.7$,
1000 Lh (999 random + $\diamondsuit$ for a Maximin optimal design)
MKV=$\max_{\mathbf{x} \in \mathscr{X}} \hat{\rho}_n(\mathbf{x}; \theta)$, $J_\alpha = \det^\alpha(\mathbf{M}_\beta) + \det^{1-\alpha}(\mathbf{M}_\theta)$ ($\alpha = 0.8$)



Example: $n = 7$, $d = 2$, $C(\mathbf{x} - \mathbf{x}'; \theta) = \exp(-\theta \|\mathbf{x} - \mathbf{x}'\|)$, $\theta = 0.7$,
1000 Lh (999 random + $\diamondsuit$ for a Maximin optimal design)
MKV=$\max_{\mathbf{x} \in \mathscr{X}} \hat{\rho}_n(\mathbf{x}; \theta)$, $J_\alpha = \det^\alpha(\mathbf{M}_\beta) + \det^{1-\alpha}(\mathbf{M}_\theta)$ $(\alpha = 0.8)$



However, the effet of corrective term in
$\hat{\rho}_n(\mathbf{x}; \theta) = \rho_n(\mathbf{x}; \theta) + \mathrm{trace}\{\mathbf{M}_\theta^{-1} \frac{\partial \mathbf{v}_n(\mathbf{x};\theta)}{\partial \theta} \mathbf{C}_n \frac{\partial \mathbf{v}_n(\mathbf{x};\theta)}{\partial \theta^\top}\}$
quickly vanishes as n increases

5 Conclusions part (2) — with model

➤ Design criteria relying on a Gaussian-process model (entropy, MMSE, IMSE) depend on the chosen covariance (and on θ in $C(\cdot, \theta)$)

- ➠ expectation w.r.t. θ (Joseph et al., 2015) $\rightarrow$ entropy
- ➠ worst case w.r.t. θ (Spöck and Pilz, 2010) $\rightarrow$ IMSE
- ➠ However, the model is often just a tool to generate a space-filling design: the value of θ is not critical (choose a small enough correlation length to spread the points in $\mathscr{X}$)

5 Conclusions part (2) — with model

➤ Design criteria relying on a Gaussian-process model (entropy, MMSE, IMSE) depend on the chosen covariance (and on θ in $C(\cdot, \theta)$)

- ➠ expectation w.r.t. θ (Joseph et al., 2015) $\rightarrow$ entropy
- ➠ worst case w.r.t. θ (Spöck and Pilz, 2010) $\rightarrow$ IMSE
- ➠ However, the model is often just a tool to generate a space-filling design: the value of θ is not critical (choose a small enough correlation length to spread the points in $\mathscr{X}$)

➤ Put more points near the boundaries than in the central part of $\mathscr{X}$ (but be careful ➠ in high dimension almost all the volume is near the boundaries!)

➤ Put a few points close to each other to help estimation of θ

5 Conclusions part (2) — with model

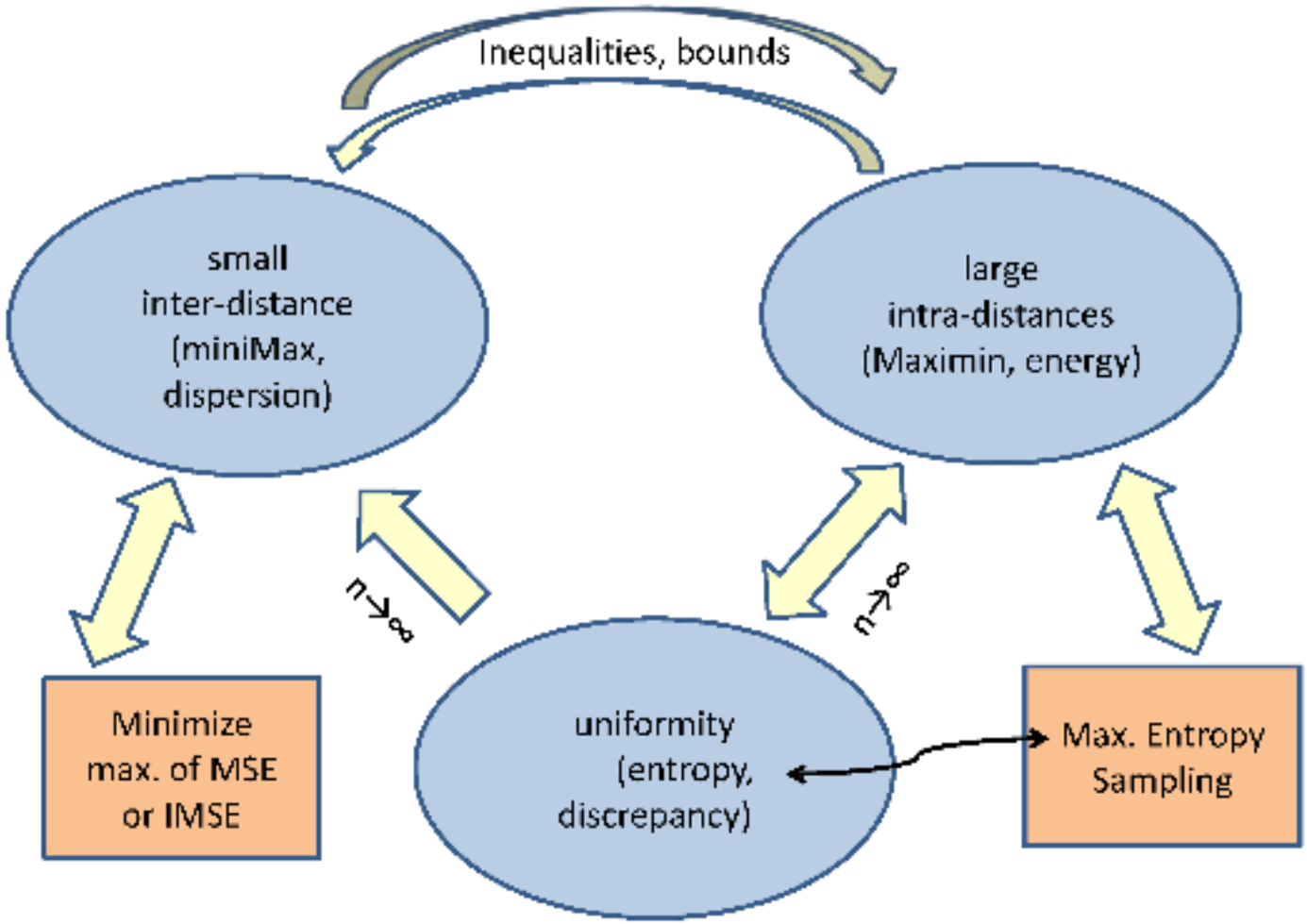
➤ Design criteria relying on a Gaussian-process model (entropy, MMSE, IMSE) depend on the chosen covariance (and on θ in $C(\cdot, \theta)$)

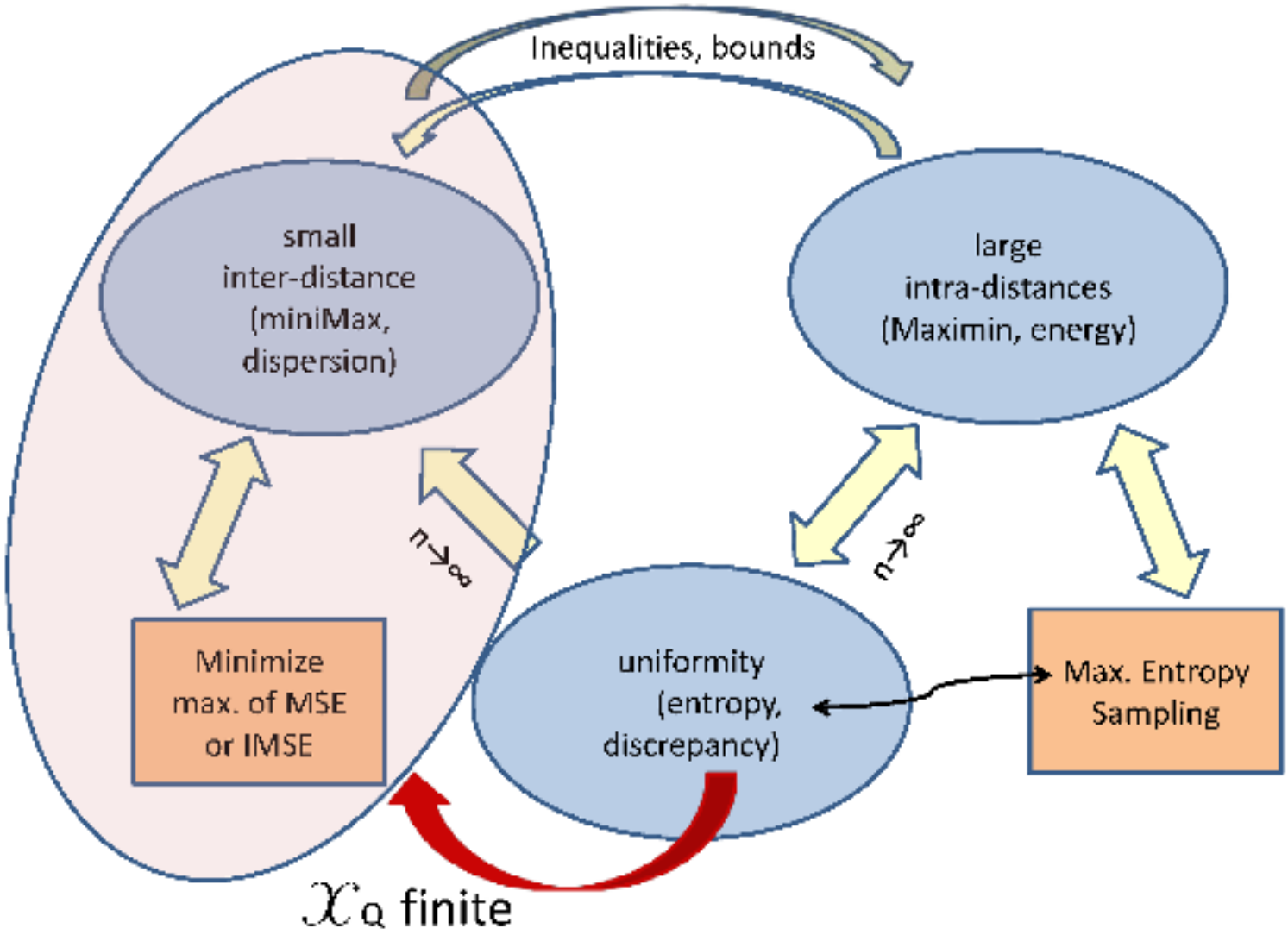
- ➠ expectation w.r.t. θ (Joseph et al., 2015) $\rightarrow$ entropy
- ➠ worst case w.r.t. θ (Spöck and Pilz, 2010) $\rightarrow$ IMSE
- ➠ However, the model is often just a tool to generate a space-filling design: the value of θ is not critical (choose a small enough correlation length to spread the points in $\mathscr{X}$)

➤ Put more points near the boundaries than in the central part of $\mathscr{X}$ (but be careful ➠ in high dimension almost all the volume is near the boundaries!)

➤ Put a few points close to each other to help estimation of θ

Test several methods (none is perfect) by comparing values of different criteria


Inequalities, bounds
small
inter-distance
(miniMax,
dispersion)
large
intra-distances
(Maximin, energy)
$n \to \infty$
$n \to \infty$
Minimize
max. of MSE
or IMSE
uniformity
(entropy,
discrepancy)
Max. Entropy
Sampling


Inequalities, bounds
small
inter-distance
(miniMax,
dispersion)
large
intra-distances
(Maximin, energy)
$n \to \infty$
$n \to \infty$
Minimize
max. of MSE
or IMSE
uniformity
(entropy,
discrepancy)
Max. Entropy
Sampling
$\mathcal{X}_Q$ finite

Références I

Abt, M., 1999. Estimating the prediction mean squared error in gaussian stochastic processes with exponential correlation structure. Scandinavian Journal of Statistics 26 (4), 563–578.

Atwood, C., 1973. Sequences converging to *D*-optimal designs of experiments. Annals of Statistics 1 (2), 342–352.

Böhning, D., 1985. Numerical estimation of a probability measure. Journal of Statistical Planning and Inference 11, 57–69.

Böhning, D., 1986. A vertex-exchange-method in *D*-optimal design theory. Metrika 33, 337–347.

Chernoff, H., 1953. Locally optimal designs for estimating parameters. Annals of Math. Stat. 24, 586–602.

Dette, H., Pepelyshev, A., 2010. Generalized latin hypercube design for computer experiments. Technometrics 52 (4), 421–429.

Fedorov, V., 1972. Theory of Optimal Experiments. Academic Press, New York.

Fedorov, V., 1996. Design of spatial experiments: model fitting and prediction. In: Gosh, S., Rao, C. (Eds.), Handbook of Statistics, vol. 13. Elsevier, Amsterdam, Ch. 16, pp. 515–553.

Gauthier, B., Pronzato, L., 2014. Spectral approximation of the IMSE criterion for optimal designs in kernel-based interpolation models. SIAM/ASA J. Uncertainty Quantification 2, 805–825, DOI 10.1137/130928534.

Gauthier, B., Pronzato, L., 2016. Approximation of IMSE-optimal designs via quadrature rules and spectral decomposition. Communications in Statistics – Simulation and Computation 45 (5), 1600–1612.

Gauthier, B., Pronzato, L., 2017. Convex relaxation for IMSE optimal design in random field models. Computational Statistics and Data Analysis 113, 375–394.

Références II

Harville, D., Jeske, D., 1992. Mean squared error of estimation or prediction under a general linear model. Journal of the American Statistical Association 87 (419), 724–731.

Johnson, M., Moore, L., Ylvisaker, D., 1990. Minimax and maximin distance designs. Journal of Statistical Planning and Inference 26, 131–148.

Joseph, V., Gul, E., Ba, S., 2015. Maximum projection designs for computer experiments. Biometrika 102 (2), 371–380.

Kiefer, J., Wolfowitz, J., 1960. The equivalence of two extremum problems. Canadian Journal of Mathematics 12, 363–366.

Mitchell, T., 1974. An algorithm for the construction of "D-optimal" experimental designs. Technometrics 16, 203–210.

Müller, W., Pronzato, L., Rendas, J., Waldl, H., 2015. Efficient prediction designs for random fields. Applied Stochastic Models in Business and Industry 31 (2), 178–194 (+ discussion pages 195–203).

Müller, W., Pronzato, L., Waldl, H., 2011. Beyond space-filling: An illustrative case. Procedia Environmental Sciences 7, 14–19, dOI 10.1016/j.proenv.2011.07.004.

Pilz, J., 1983. Bayesian Estimation and Experimental Design in Linear Regression Models. Vol. 55. Teubner-Texte zur Mathematik, Leipzig, (also Wiley, New York, 1991).

Pukelsheim, F., 1993. Optimal Experimental Design. Wiley, New York.

Pukelsheim, F., Reider, S., 1992. Efficient rounding of approximate designs. Biometrika 79 (4), 763–770.

Sacks, J., Welch, W., Mitchell, T., Wynn, H., 1989. Design and analysis of computer experiments. Statistical Science 4 (4), 409–435.

Schwabe, R., 1996. Optimum Designs for Multi-Factor Models. Springer, New York.

Références III

Shewry, M., Wynn, H., 1987. Maximum entropy sampling. Applied Statistics 14, 165–170.

Silvey, S., 1980. Optimal Design. Chapman & Hall, London.

Spöck, G., Pilz, J., 2010. Spatial sampling design and covariance-robust minimax prediction based on convex design ideas. Stochastic Environmental Research and Risk Assessment 24 (3), 463–482.

Titterington, D., 1976. Algorithms for computing *D*-optimal designs on a finite design space. In: Proc. of the 1976 Conference on Information Science and Systems. Dept. of Electronic Engineering, John Hopkins University, Baltimore, pp. 213–216.

Torsney, B., 1983. A moment inequality and monotonicity of an algorithm. In: Kortanek, K., Fiacco, A. (Eds.), Proc. Int. Symp. on Semi-infinite Programming and Applications. Springer, Heidelberg, pp. 249–260.

Torsney, B., 2009. W-iterations and ripples therefrom. In: Pronzato, L., Zhigljavsky, A. (Eds.), Optimal Design and Related Areas in Optimization and Statistics. Springer, Ch. 1, pp. 1–12.

Welch, W., 1982. Branch-and-bound search for experimental designs based on *D*-optimality and other criteria. Technometrics 24 (1), 41–28.

Wu, C., 1978. Some algorithmic aspects of the theory of optimal designs. Annals of Statistics 6 (6), 1286–1301.

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

Yu, Y., 2010. Strict monotonicity and convergence rate of Titterington's algorithm for computing *D*-optimal designs. Comput. Statist. Data Anal. 54, 1419–1425.

Yu, Y., 2011. *D*-optimal designs via a cocktail algorithm. Stat. Comput. 21, 475–481.

Zhigljavsky, A., Dette, H., Pepelyshev, A., 2010. A new approach to optimal design for linear models with correlated observations. Journal of the American Statistical Association 105 (491), 1093–1103.

Zimmerman, D., 2006. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. Environmetrics 17 (6), 635–652.