

# A convex approach to optimum design of experiments with correlated observations

Werner G. Müller

joint work with Andrej Pázman and Markus Hainy

mODa13, Southampton, UK, July 13<sup>th</sup> 2023



# Introduction and motivation

- ▶ Theory for optimal design for classical (non-)linear regression with uncorrelated errors has well developed into textbook status: Atkinson et al. (2007), Morris (2010), Goos and Jones (2011), etc.
- ▶ However, for the correlated errors setting the literature is only scattered: Sacks and Ylvisaker (1966) , Näther (1985), Fedorov (1996), series of papers started with Zhigljavsky et al. (2010), latest Dette et al. (2017).
- ▶ Brandnew: Ucinski (2023)!
- ▶ Entirely different suggestion was given in Pázman and Müller (1998) and fully developed in Müller and Pázman (2003), nonconvex and thus no Kiefer-Wolfowitz-type equivalence theorem available.
- ▶ We have fixed that here.

# Setup and problem description

Observations to be generated from the model

$$y(x) = f^T(x)\theta + \varepsilon(x); \quad x \in \mathcal{X}, \quad (1)$$

where  $\mathcal{X} = \{x_1, \dots, x_N\}$  is a finite design space containing  $N$  points and  $\theta$  is the unknown vector of parameters with  $\dim(\theta) = p$ .

Further assumptions:  $E\{\varepsilon(x)\} = 0$ , and is correlated, defined through the  $N \times N$  known matrix  $C$  with elements

$$C_{ij} = \text{cov}\{\varepsilon(x_i), \varepsilon(x_j)\}; \quad i, j = 1, \dots, N.$$

Design criteria  $\Phi$  are usually increasing functions of the information matrix  $M_\xi = \sum_{x \in \mathcal{X}} f(x)f^T(x)\xi(x)$  and it is the purpose of optimum design to find

$$\xi^* = \arg \max_{\xi} \Phi \{M_\xi\}. \quad (2)$$

## Classical theory, i.e. the case $C = \sigma^2 I$

goes back to Kiefer (1959).

Assuming  $\xi$  to be approximated by a probability measure, then (2) can be tackled by (standard) multivariate calculus.

So for any concave criterion  $\Phi$  and assuming the existence (and particular form) of a directional derivative

$$\lim_{\alpha \rightarrow 0} \frac{\Phi[(1 - \alpha)M_\xi + \alpha M_{\xi'}] - \Phi[M_\xi]}{\alpha} = C - \int_{\mathcal{X}} \phi(x, \xi) \dot{\xi} dx.$$

We then have equivalent to (2):

$$\xi^* = \arg \min_{\xi} \max_x \phi(x, \xi). \quad (3)$$

Also note that by construction  $M_\xi$  is additive in  $x$ !

# Virtual noise

Define a convex set of restricted probability measures on  $\mathcal{X}$

$$\Xi = \left\{ \xi : \sum_{x \in \mathcal{X}} \xi(x) = 1, \forall x \in \mathcal{X} \ 0 \leq \xi(x) \leq 1/n \right\},$$

where  $n$  is the number of desired points of support of an optimum exact design  $\xi^*$ . Instead of (1), consider the potentially perturbed observations

$$y(x) = f^T(x)\theta + \varepsilon(x) + w_\xi(x); \quad x \in \mathcal{X}, \quad (4)$$

where the variance of the supplementary virtual noise  $w_\xi$ , independent of  $\varepsilon$ , is fixed as

$$\text{var} \{w_\xi(x)\} = \sigma_\xi^2(x) = \kappa \frac{1/n - \xi(x)}{\xi(x)},$$

whereas  $\text{cov} \{w_\xi(x), w_\xi(x')\} = 0$  when  $x \neq x'$ .

# Interpretation

- ▶ If  $\xi(x) = 0$  i.e.  $\sigma_{\xi}^2 = \infty$  there is no observation at the point  $x$ ,
- ▶ if  $\xi(x) = 1/n$  i.e.  $\sigma_{\xi}^2 = 0$  the observation at the point  $x$  is not disturbed at all by the virtual noise.
  
- ▶ It follows that an exact  $k$ -point design is represented by such a  $\xi$  that  $\xi(x) = 1/n$  in  $k$  points and  $\xi = 0$  in all other points of  $\mathcal{X}$ .
- ▶ By construction  $\Xi$  contains all  $n$ -point designs but no  $k$ -point designs with  $k \neq n$ .

# Concavity theorem

The information matrix of  $\theta$  in model (4) is then given by

$$M(\xi) = F^T \{C + W(\xi)\}^{-1} F, \quad (5)$$

where  $C$  is the  $N \times N$  covariance matrix defined after (1), and  $F$  is the  $N \times p$  matrix  $F_{i,j} = f_j(x_i)$ ,  $x_i \in \mathcal{X}$ ,  $i = 1, \dots, N$ ,  $j = 1, \dots, p$ . We now have the following:

If  $\kappa \leq \lambda_{\min}(C)$ , the minimal eigenvalue of the matrix  $C$ , and if  $\Phi(M)$  is any optimality criterion expressed as a concave increasing function of the matrix  $M$ , then the mapping

$$\xi \in \Xi \rightarrow \Phi\{M(\xi)\}$$

is concave as well, with  $M(\xi)$  defined in (5).

# Equivalence theorem

Denote now for any global criterion having a gradient  $\nabla_M \Phi(M)$  for any nonsingular  $M$

$$\begin{aligned}G(\xi) &= F [\nabla_M \Phi \{M(\xi)\}] F^T, \\T(\xi) &= \left[ (C - \kappa I) \operatorname{diag} \{\xi(\cdot)\} + \frac{\kappa}{n} I \right]^{-1}, \\h(x, \xi) &= \{T(\xi)\}_{x, \cdot} G(\xi) \{T^T(\xi)\}_{\cdot, x}.\end{aligned}$$

*A design measure  $\bar{\xi}$  with  $\operatorname{supp}(\bar{\xi}) = \mathcal{X}$  is  $\Phi$ -optimal (within our theory) if and only if for every  $n$ -tuple  $z_1, \dots, z_n$  of points from  $\mathcal{X}$  we have*

$$\frac{1}{n} \sum_{i=1}^n h(z_i, \bar{\xi}) \leq \sum_{x \in \mathcal{X}} \bar{\xi}(x) h(x, \bar{\xi}). \quad (6)$$



## A new upper bound for given $n$

Now for some  $\delta > 0$ , we have that (6) is equivalent to the inequality

$$\Phi \{ M(\bar{\xi}) \} \geq \max_{\xi} \Phi \{ M(\xi) \} - \frac{\kappa\delta}{n},$$

which provides a rather close upper bound for the performance of all exact  $n$ -point designs.

We can thus calibrate the performance of all existing designs/methods by this bound!

# A design algorithm

Any concave and positive criterion can be written in a form

$$\Phi \{M(\xi)\} = \min_{\mu \in \Xi} \left\{ a(\mu) + \sum_{i=1}^N b_i(\mu) \xi(x_i) \right\}.$$

Here,  $a(\mu)$  and  $b_i(\mu)$  can be obtained either by using the Taylor formula of  $\Phi \{M(\xi)\}$  at  $\mu$ , or using some algebraic tricks as in Burclová and Pázman (2016). At the  $k$ th step we start with a set  $\Xi_k$ . Consider the following finite set of constraints, which are linear in the variables  $t \in \mathbb{R}$  and  $\xi(x)$ ;  $x \in \mathcal{X}$ ,

$$t \geq 0 \quad t \leq a(\mu) + \sum_{i=1}^N b_i(\mu) \xi(x_i); \quad \mu \in \Xi_k$$
$$\xi(x_i) \geq 0; \quad \xi(x_i) \leq \frac{1}{n}; \quad i = 1, \dots, N, \quad \sum_{i=1}^N \xi(x_i) = 1,$$

and a standard linear program which maximizes  $t$  under these constraints.

## Example 1: a classic one-parameter model

This example has originally been considered by Sacks and Ylvisaker (1966). Dette et al. (2016) use it to illustrate the efficiency of their method. It utilizes D-optimality and is a one-parameter model given by

$$f(x) = 1 + 0.5 \sin(2\pi x), \quad x \in [1, 2],$$
$$\text{cov} \{ \varepsilon(x), \varepsilon(x') \} = \begin{cases} x^2 x' & x \leq x' \\ x(x')^2 & x > x', \end{cases} \quad \lambda_{\min}(C) = 0.00276.$$

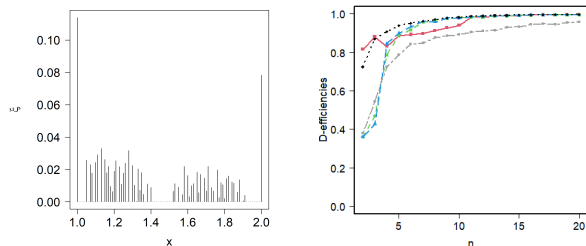


Figure 1: Our measure (left) and efficiencies (based on our bound) (right).

# Example 1: 4-point comparative performance (quantiles)

Table 1: Optimal designs and D-efficiencies for Example 1

	$x_1$	$x_2$	$x_3$	$x_4$	D-eff
Q-VN	1.10	1.23	1.40	1.76	0.8316
Q-VN+EP	1.00	1.21	1.58	2.00	0.7865
Q-DPZ+EP	1.00	1.28	1.69	2.00	0.8455
R-UNIF					0.6955
BKSF	1.19	1.67	1.79	2.00	0.9075
EXS	1.22	1.66	1.79	2.00	0.9158

## Example 2: a multiparameter case

taken from Section 3.6 of Dette et al. (2016). The specifications of this four-parameter model are

$$\begin{aligned}f^T(x) &= (1, x, x^2, x^3), \quad x \in [1, 2], \\ \text{cov}\{\varepsilon(x), \varepsilon(x')\} &= \min(x, x'), \quad \lambda_{\min}(C) = 0.0025.\end{aligned}$$

We again consider the D-criterion, this is one of the rare cases in which our method does even slightly better than the BKS-algorithm.

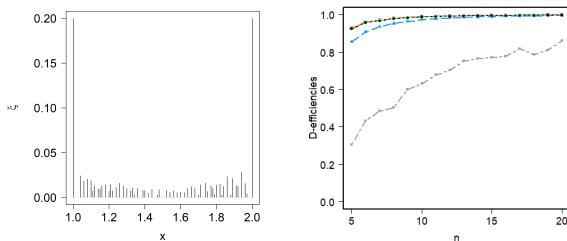


Figure 2: Our measure (left) and efficiencies versus sample size (right).

## Example 1: 5-point comparative performance (quantiles)

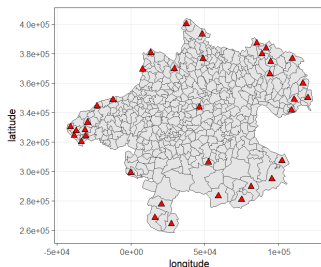
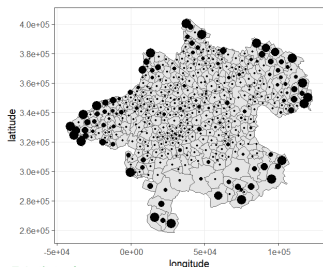
Table 2: Optimal designs and D-efficiencies for Example 2

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	D-eff
Q-VN	1.00	1.16	1.52	1.84	2.00	0.9251
Q-VN+EP	1.00	1.20	1.52	1.82	2.00	0.9300
Q-DPZ+EP	1.00	1.14	1.33	1.60	2.00	0.8554
R-UNIF						0.3208
BKSF	1.00	1.16	1.46	1.83	2.00	0.9270
EXS	1.00	1.21	1.61	1.84	2.00	0.9308

## Example 3: a real application

inspired by the example used throughout Mateu and Müller (2012) with rainfall measurements gathered from 37 weather stations in the Austrian state of Upper Austria during the years 1994 – 2009. The response function is assumed to be a plane and we assume that the parameters of the kernel function are known and set them to the kriging estimates. The model we consider is therefore

$$f^T(x) = (1, x_1, x_2), \quad x = (x_1, x_2)^T,$$
$$k(x, x') = 1756.65 \cdot \exp(\|x - x'\|_2 / 40792.35).$$



# Thank you for your attention!




BROWSE ▾

RESOURCES ▾

ABOUT ▾

[Home](#) > [Journals](#)

 Open Access

 Sprache auswählen ▾

[Translator Disclaimer](#)

2022

## A convex approach to optimum design of experiments with correlated observations

[Andrej Pázman](#), [Markus Hainy](#), [Werner G. Müller](#)

[Author Affiliations +](#)

Electron. J. Statist. 16(2): 5659-5691 (2022). DOI: 10.1214/22-EJS2071

The paper can be found at <https://projecteuclid.org/journals/electronic-journal-of-statistics/volume-16/issue-2/A-convex-approach-to-optimum-design-of-experiments-with-correlated-observations/10.1214/22-EJS2071.full>