Optimal Designs for Gaussian Process Models via Spectral Decomposition

Ofir Harari

Department of Statistics & Actuarial Sciences, Simon Fraser University

September 2014

Outline

SFU

Outline

SFU

The Gaussian Process Model

- Suppose that we observe $\boldsymbol{y} = [y(\boldsymbol{x}_1), \ldots, y(\boldsymbol{x}_n)]^{\mathsf{T}}$, where

$$y(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^{\mathsf{T}} \beta + Z(\boldsymbol{x}) \quad .$$

- Here $\boldsymbol{f}(\boldsymbol{x})$ is a vector of regression functions evaluated at $\boldsymbol{x}$ and $Z(\boldsymbol{x})$ is a zero mean stationary Gaussian process with marginal variance σ^2 and correlation function $R(\cdot;\theta)$.
- An extremely popular modeling approach in spatial statistics, geostatistics, computer experiments and more.
- Sometimes measurement error (i.e. *a nugget term*) is added.

SFU

The Gaussian Process Model (cont.)

Notations:

$\boldsymbol{R}_{ij} = R(\mathbf{x}_i - \mathbf{x}_j)$ - the correlation matrix at the design $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$.

$\boldsymbol{y} = \left[y(\boldsymbol{x}_1), \ldots, y(\boldsymbol{x}_n)\right]^{\mathsf{T}}$ - the vecor of observatios at $\mathcal{D}$.

$\boldsymbol{r}(\mathbf{x}) = \left[R(\mathbf{x} - \mathbf{x}_1), \ldots, R(\mathbf{x} - \mathbf{x}_n)\right]^{\mathsf{T}}$ - the vector of correlations at site $\mathbf{x}$.

$\boldsymbol{F}_{ij} = f_j(\boldsymbol{x}_i)$ - the design matrix at $\mathcal{D}$.

SFU

The Gaussian Process Model (cont.)

Universal Kriging

Assuming $\pi(\beta) \propto 1$

$$\widehat{y}(\boldsymbol{x}) = \mathbb{E}\left\{y(\boldsymbol{x})\middle|\boldsymbol{y}\right\} = \boldsymbol{f}^{\mathsf{T}}(\boldsymbol{x})\,\widehat{\beta} + \boldsymbol{r}^{\mathsf{T}}(\boldsymbol{x})\,\boldsymbol{R}^{-1}\left(\boldsymbol{y} - \boldsymbol{F}\widehat{\beta}\right) \quad , \tag{1}$$

is the minimizer of the mean squared prediction error (MSPE), which will then be

$$\mathbb{E}\left[\left\{\widehat{y}(\boldsymbol{x}) - y(\boldsymbol{x})\right\}^2\middle|\boldsymbol{y}\right] = \operatorname{var}\left\{y(\boldsymbol{x})\middle|\boldsymbol{y}\right\}$$

$$= \sigma^2\left\{1 - \left[\boldsymbol{f}^{\mathsf{T}}(\boldsymbol{x}), \boldsymbol{r}^{\mathsf{T}}(\boldsymbol{x})\right]\begin{bmatrix}\boldsymbol{0} & \boldsymbol{F}^{\mathsf{T}} \\ \boldsymbol{F} & \boldsymbol{R}\end{bmatrix}^{-1}\begin{bmatrix}\boldsymbol{f}(\boldsymbol{x}) \\ \boldsymbol{r}(\boldsymbol{x})\end{bmatrix}\right\} \quad .$$

- If the assumption of Gaussianity is dropped, (1) would still be the BLUP for $y(\boldsymbol{x})$.

SFU

Spectral Decomposition of a GP

Mercer's Theorem

If $\mathcal{X}$ is compact and R (the covariance function of a GP) is continuous in $\mathcal{X}^2$,

$$\text{then} \qquad R(\boldsymbol{x}, \boldsymbol{y}) = \sum_{i=1}^{\infty} \lambda_i \varphi_i(\boldsymbol{x}) \varphi_i(\boldsymbol{y}) \quad ,$$

where the φ_i's and the λ_i's are the solutions of the homogeneuos Fredholm integral equation of the second kind

$$\int_{\mathcal{X}} R(\boldsymbol{x}, \boldsymbol{y}) \varphi_i(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} = \lambda_i \varphi_i(\boldsymbol{x})$$

and

$$\int_{\mathcal{X}} \varphi_i(\boldsymbol{x}) \varphi_j(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \delta_{ij} \quad .$$

SFU

Spectral Decomposition of a GP (cont.)

The Karhunen-Loève Decomposition

Under the aforementioned conditions

$$Z(\boldsymbol{x}) = \sum_{i=1}^{\infty} \alpha_i \varphi_i(\boldsymbol{x}) \quad \text{(in the MSE sense)},$$

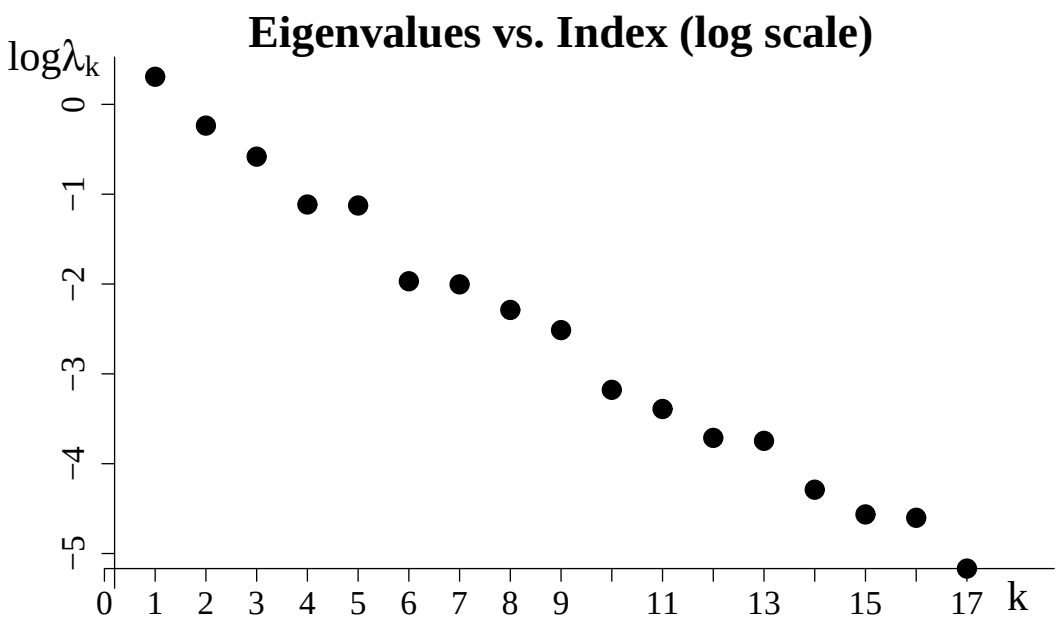
where the α_i's are independent $\mathcal{N}(0, \lambda_i)$ random variables.

- The best approximation (in the MSE sense) of $Z(\boldsymbol{x})$ is the truncated series in which the eigenvalues are arranged in decreasing order.
- For piecewise analytic $R(\boldsymbol{x}, \boldsymbol{y})$, $\lambda_k \leq c_1 \exp\left(-c_2 k^{1/d}\right)$ for some constants c_1 and c_2 (Frauenfelder *et al.* 2005).
- Numerical solutions involve Galerkin-type methods.

SFU

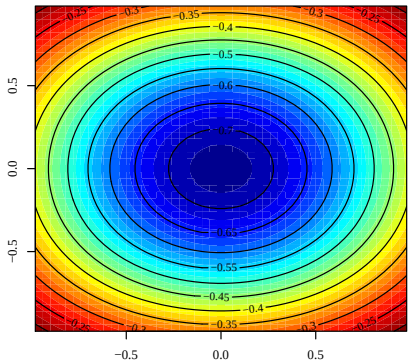
A Worked Example

$$R(\boldsymbol{x}, \boldsymbol{w}) = \exp\left\{-(x_1 - w_1)^2 - 2(x_2 - w_2)^2\right\} \quad ,$$

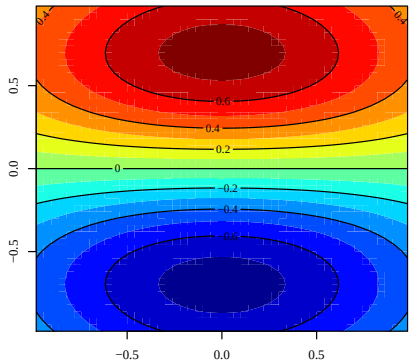


A Worked Example (cont.)

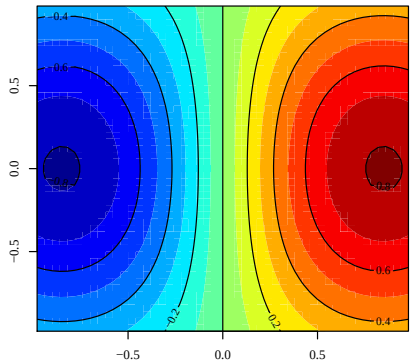
Eigenfunction #1 , λ = 1.3597



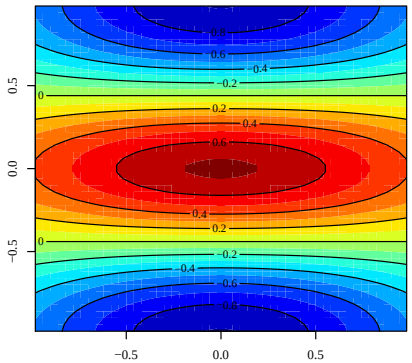
Eigenfunction #2 , λ = 0.7893



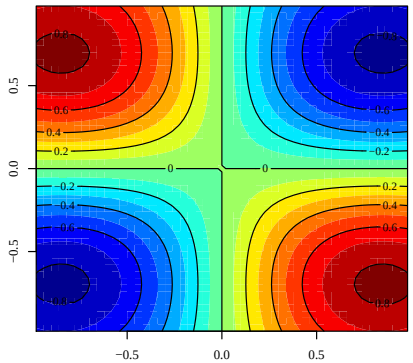
Eigenfunction #3 , λ = 0.5588



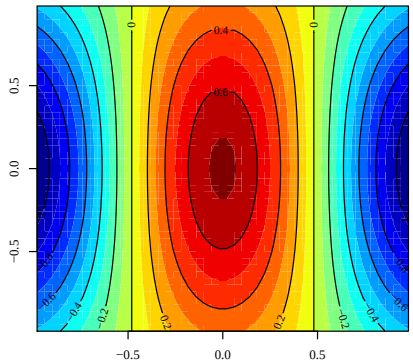
Eigenfunction #4 , λ = 0.328



Eigenfunction #5 , λ = 0.3244

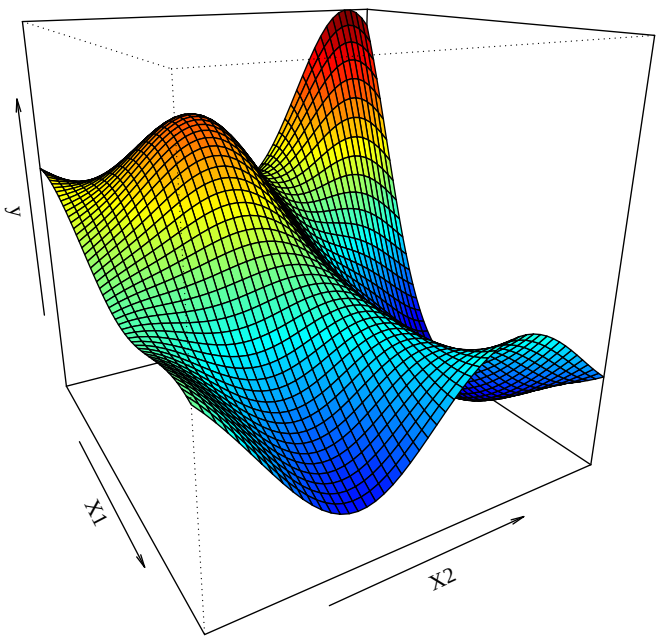


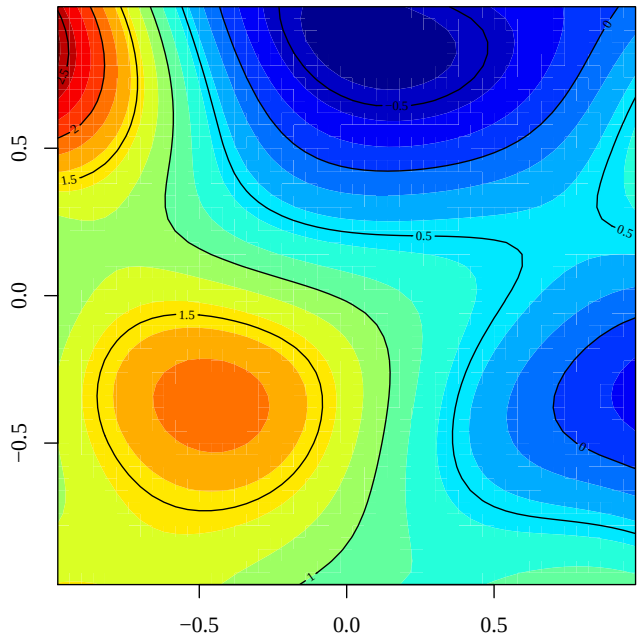
Eigenfunction #6 , λ = 0.1396



A Worked Example (cont.)

A single realization based on the first 17 K-L terms ($\approx$ 99.5% of the process' energy)





Outline

SFU

Minimum IMSPE Designs

Integrated Mean Squared Prediction Error

Sacks *et al.* (1989b) suggested a suitable design for computer experiments should minimize

$$\frac{\mathcal{J}\left(\mathcal{D},\widehat{y}\right)}{\sigma^2} = \frac{1}{\sigma^2}\int_{\mathcal{X}} \mathbb{E}\left[\left\{\widehat{y}(\boldsymbol{x}) - y(\boldsymbol{x})\right\}^2 \middle| \boldsymbol{y}\right] \mathrm{d}\boldsymbol{x}$$

$$= \mathrm{vol}\left(\mathcal{X}\right) - \mathrm{tr}\left\{ \begin{bmatrix} \mathbf{0} & \boldsymbol{F}^{\mathsf{T}} \\ \boldsymbol{F} & \boldsymbol{R} \end{bmatrix}^{-1} \int_{\mathcal{X}} \begin{bmatrix} \boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{f}^{\mathsf{T}}\left(\boldsymbol{x}\right) & \boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{r}^{\mathsf{T}}\left(\boldsymbol{x}\right) \\ \boldsymbol{r}\left(\boldsymbol{x}\right)\boldsymbol{f}^{\mathsf{T}}\left(\boldsymbol{x}\right) & \boldsymbol{r}\left(\boldsymbol{x}\right)\boldsymbol{r}^{\mathsf{T}}\left(\boldsymbol{x}\right) \end{bmatrix} \mathrm{d}\boldsymbol{x} \right\}$$

where $\mathrm{vol}(\mathcal{X})$ is the volume of $\mathcal{X}$.

SFU

Application of the K-L Decomposition

Consider the standard Gaussian process

$$y(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^{\mathsf{T}} \beta + Z(\boldsymbol{x}) \quad ,$$

where β is (for now) a known vector.

- The Kriging predictor would now be $\widehat{y}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^{\mathsf{T}} \beta + \boldsymbol{r}^{\mathsf{T}}(\boldsymbol{x}) \boldsymbol{R}^{-1} (\boldsymbol{y} - \boldsymbol{F}\beta)$.

Proposition

For any compact $\mathcal{X}$ and continuous covariance kernel R,

$$\frac{\mathcal{J}\left(\mathcal{D}, \widehat{y}\right)}{\sigma^2} = \mathrm{vol}(\mathcal{X}) - \sum_{k=1}^{\infty} \lambda_k^2 \phi_k^{\mathsf{T}} \boldsymbol{R}^{-1} \phi_k \quad , \tag{2}$$

where $\phi_p = [\varphi_p(\boldsymbol{x}_1), \ldots, \varphi_p(\boldsymbol{x}_n)]^{\mathsf{T}}$.

SFU

Application of the K-L Decomposition (cont.)

Theorem

1. *For any design* $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$,

$$\mathcal{J}\left(\mathcal{D}, \widehat{y}\right) \geq \sigma_z^2 \sum_{k=n+1}^{\infty} \lambda_k \tag{3}$$

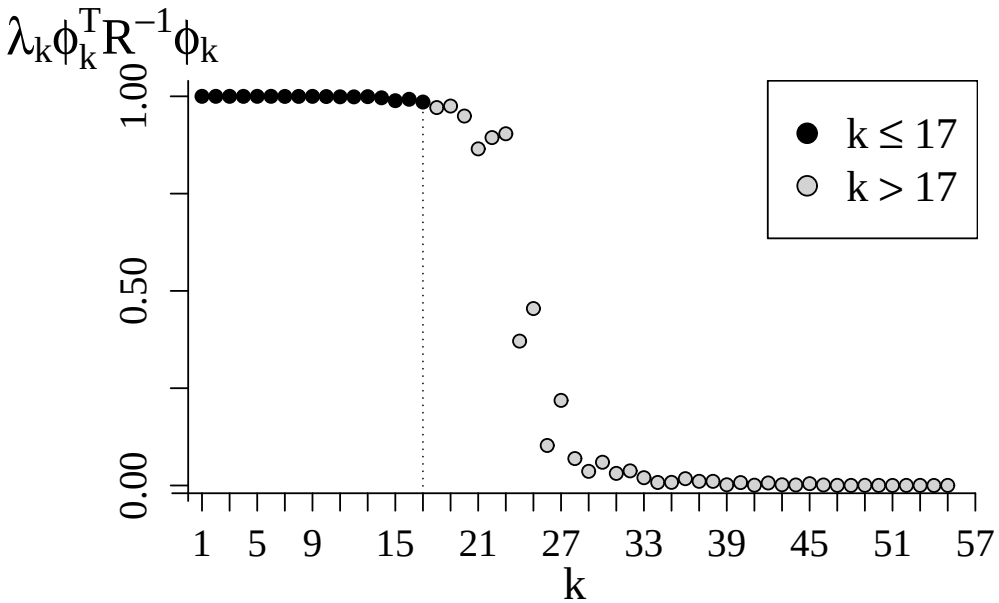
2. *The lower bound (3) will be achieved if* $\mathcal{D}$ *satisfies*

$$\lambda_k \phi_k^{\mathsf{T}} \boldsymbol{R}^{-1} \phi_k = \begin{cases} 1 & k \leq n \\ 0 & k > n \end{cases}$$

3. *If such an ideal* $\mathcal{D}$ *exists, the prediction* $\widehat{y}(\boldsymbol{x})$ *at any* $\boldsymbol{x} \in \mathcal{X}$ *would be identical to the prediction based on the finite dimensional Bayesian linear regression model* $y(\boldsymbol{x}) = \alpha_1 \varphi_1(\boldsymbol{x}) + \cdots + \alpha_n \varphi_n(\boldsymbol{x})$ *where* $\alpha_j \sim \mathcal{N}\left(0, \lambda_j\right)$ *are independent.*

Application of the K-L Decomposition (cont.)

- By definition, IMSPE-optimal designs are prediction-oriented. What the Theorem tells us is that in some sense, they are also useful for variable selection (w.r.t the K-L basis functions).
- In reality, such ideal designs do not exist - but IMSPE-optimal designs get pretty close:



Approximate Minimum IMSPE Designs

Approximate IMSPE

$$\frac{\widetilde{\mathcal{J}}_M\left(\mathcal{D},\widehat{y}\right)}{\sigma^2} = \mathrm{vol}(\mathcal{X}) - \sum_{k=1}^{M} \lambda_k^2 \phi_k^{\mathsf{T}} \boldsymbol{R}^{-1} \phi_k = \mathrm{vol}(\mathcal{X}) - \mathrm{tr}\left\{\boldsymbol{\Lambda}^2 \boldsymbol{\Phi}^{\mathsf{T}} \boldsymbol{R}^{-1} \boldsymbol{\Phi}\right\} \quad ,$$

where

$$\boldsymbol{\Lambda} = \mathrm{diag}\left(\lambda_1, \ldots, \lambda_M\right) \quad \text{and} \quad \boldsymbol{\Phi}_{ij} = \varphi_j\left(\boldsymbol{x}_i\right) \quad , \quad 1 \leq i \leq n \quad , \quad 1 \leq j \leq M \quad .$$

Approximate Minimum IMSPE Designs

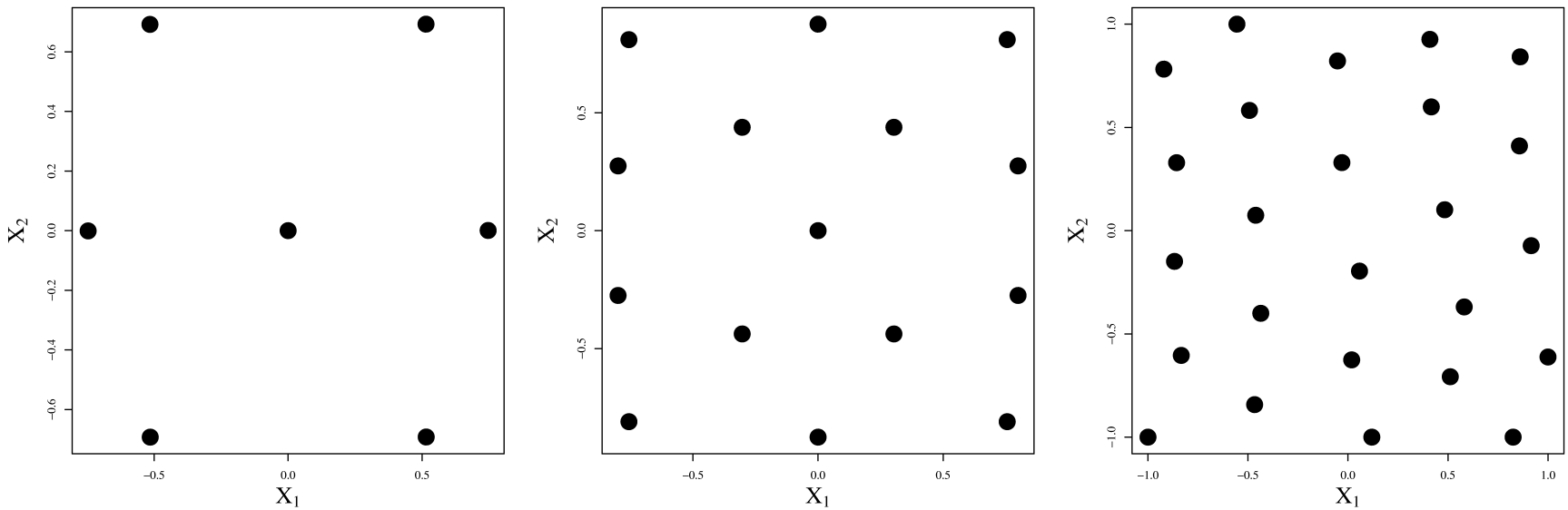
$$\mathcal{D}^* = \underset{\mathcal{D}\subset\mathcal{X}}{\mathrm{argmin}}\, \widetilde{\mathcal{J}}_M\left(\mathcal{D},\widehat{y}\right) = \underset{\mathcal{D}\subset\mathcal{X}}{\mathrm{argmax}}\, \mathrm{tr}\left\{\boldsymbol{\Lambda}^2 \boldsymbol{\Phi}^{\mathsf{T}} \boldsymbol{R}^{-1} \boldsymbol{\Phi}\right\}$$

SFU

A Worked Example (cont.)

Size 7, 15 and 25 approximate Minimum IMSPE designs for

$$R(\boldsymbol{x}, \boldsymbol{w}) = \exp\left\{-(x_1 - w_1)^2 - 2(x_2 - w_2)^2\right\}$$



Controlling the Relative Truncation Error

Proposition

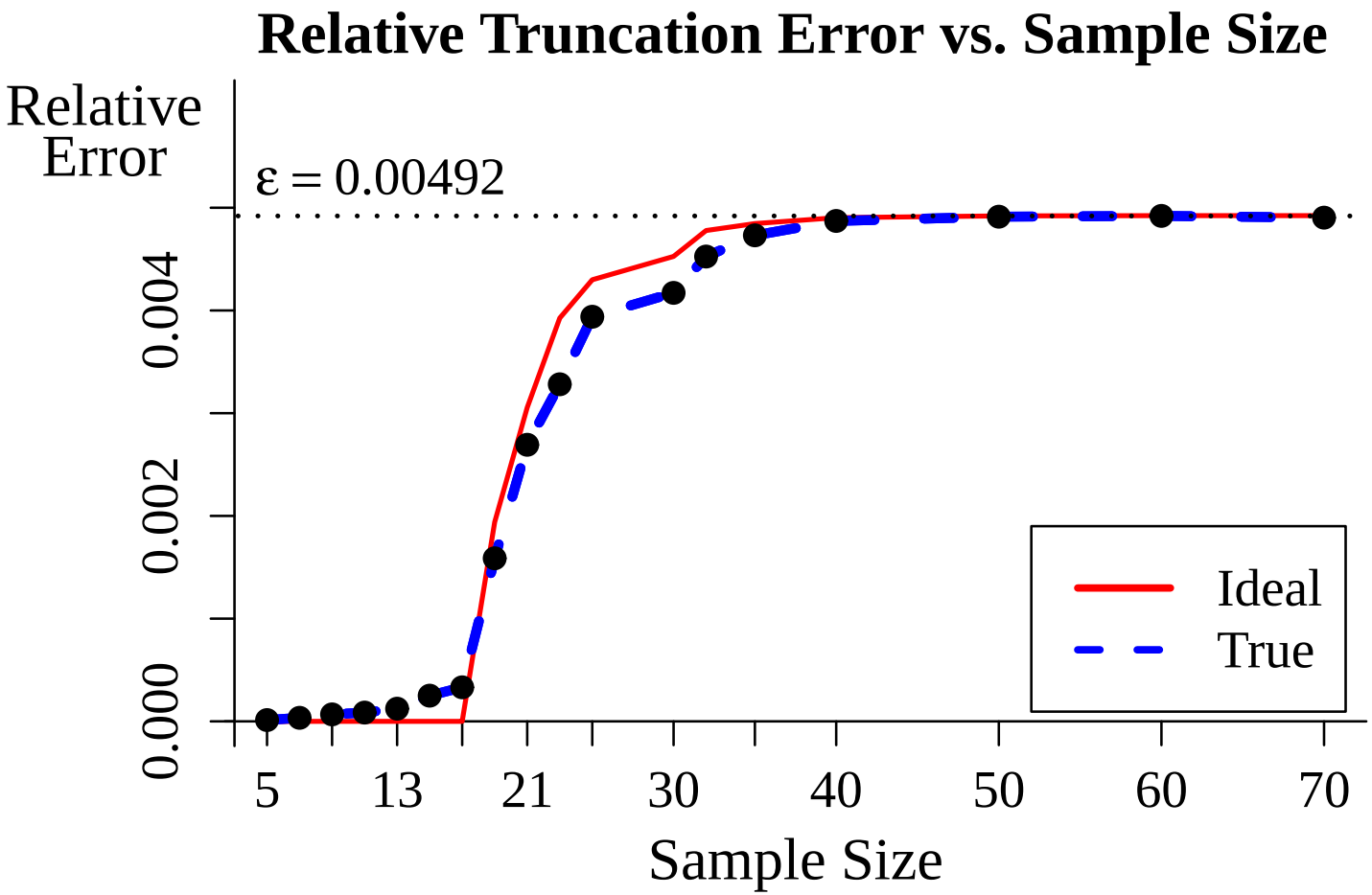
Under some conditions (which are easily met),

$$r_M = \frac{\sum_{k=M+1}^{\infty} \lambda_k^2 \phi_k^{\mathsf{T}} \boldsymbol{R}^{-1} \phi_k}{\sum_{j=1}^{\infty} \lambda_j^2 \phi_j^{\mathsf{T}} \boldsymbol{R}^{-1} \phi_j} \leq \frac{\sum_{k=M+1}^{\infty} \lambda_k}{\sum_{j=1}^{\infty} \lambda_j} \quad .$$

- ◇ By conserving enough of the process' energy, we are guaranteed to have a design as close to optimal as we wish.

Controlling the Relative Truncation Error (cont.)

Relative truncation error vs. sample size for $M = 17$ ($\Longleftrightarrow$ 0.492% energy loss):



Adaptive Designs

- Suppose now that we have the opportunity (or we are forced) to run a multi-stage experiment: n_1 runs at stage 1, n_2 runs at stage 2 and so on.
- We can use that to learn/re-estimate the essential parameters and plug-in the new estimates, to hopefully improve the design from one stage to another.
- Such designs are called “adaptive” or “batch-sequential” designs.

SFU

Adaptive Designs (cont.)

Denote

$$\boldsymbol{R}_{\text{aug}} = \left[\begin{array}{c|c} \boldsymbol{R}^{\text{new}}_{n_2 \times n_2} & \boldsymbol{R}^{\text{cross}}_{n_2 \times n_1} \\ \hline \boldsymbol{R}^{\top\,\text{cross}}_{n_1 \times n_2} & \boldsymbol{R}^{\text{old}}_{n_1 \times n_1} \end{array}\right],$$

where $\boldsymbol{R}_{\text{aug}}$, $\boldsymbol{R}^{\text{new}}$, $\boldsymbol{R}^{\text{old}}$ and $\boldsymbol{R}^{\text{cross}}$ are the augmented correlation matrix, the matrix of correlations within the new inputs, the matrix of correlations within the original design and the cross correlation matrix, and

$$\boldsymbol{\Phi}_{\text{aug}} = \left[\begin{array}{c} \boldsymbol{\Phi}^{\text{new}}_{n_2 \times M} \\ \hline \boldsymbol{\Phi}^{\text{old}}_{n_1 \times M} \end{array}\right].$$

- Parameters may be re-estimated in between batches.

Adaptive Designs (cont.)

Using basic block matrix properties, we look to maximize

$$\text{tr}\left\{\Lambda^2 \Phi_{\text{aug}}^{\mathsf{T}} \boldsymbol{R}_{\text{aug}}^{-1} \Phi_{\text{aug}}\right\} =$$

$$= \text{tr}\left[\Lambda^2 \left\{\left(\Phi_{\text{new}}^{\mathsf{T}} - \Phi_{\text{old}}^{\mathsf{T}} \boldsymbol{R}_{\text{old}}^{-1} \boldsymbol{R}_{\text{cross}}^{\mathsf{T}}\right) \boldsymbol{Q}_1^{-1} \Phi_{\text{new}} + \left(\Phi_{\text{old}}^{\mathsf{T}} - \Phi_{\text{new}}^{\mathsf{T}} \boldsymbol{R}_{\text{new}}^{-1} \boldsymbol{R}_{\text{cross}}\right) \boldsymbol{Q}_2^{-1} \Phi_{\text{old}}\right\}\right] ,$$

where

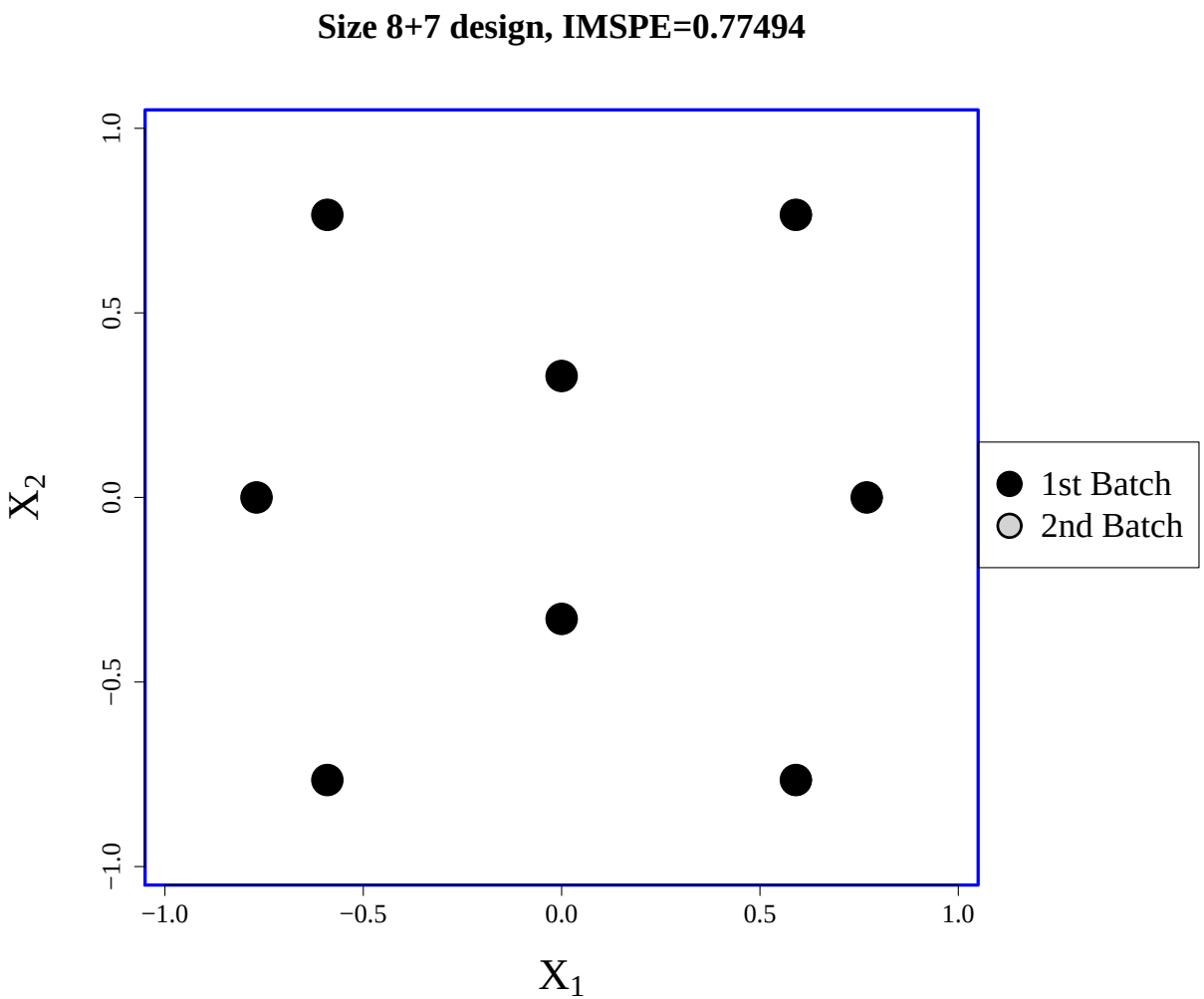
$$\boldsymbol{Q}_1 = \boldsymbol{R}_{\text{new}} - \boldsymbol{R}_{\text{cross}}^{\mathsf{T}} \boldsymbol{R}_{\text{old}}^{-1} \boldsymbol{R}_{\text{cross}} \quad \text{and} \quad \boldsymbol{Q}_2 = \boldsymbol{R}_{\text{old}} - \boldsymbol{R}_{\text{cross}} \boldsymbol{R}_{\text{new}}^{-1} \boldsymbol{R}_{\text{cross}}^{\mathsf{T}} ,$$

and $\boldsymbol{R}_{\text{old}}$ and Φ_{old} remain unchanged.

SFU

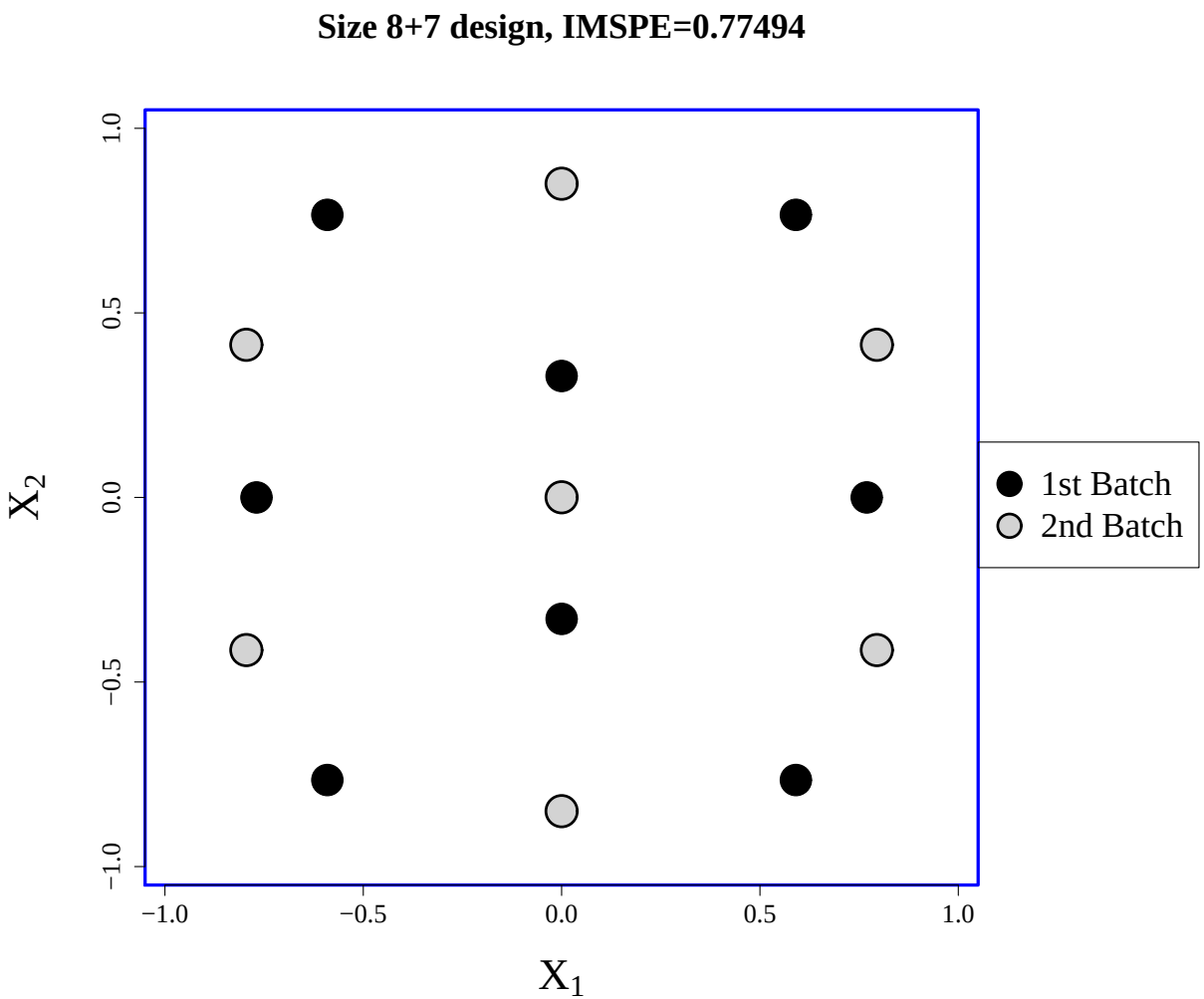
Adaptive Designs (cont.)

Adding 7 New Runs to an Existing 8 Run Design (λ = 0.1):



Adaptive Designs (cont.)

Adding 7 New Runs to an Existing 8 Run Design ($\lambda = 0.1$):



Approximate IMSPE Criterion for the Universal Kriging Model

Expanding the covariance kernel, using Mercer's Theorem, and truncating after the first M terms yields

$$\widetilde{\mathcal{H}}_M\left(\mathcal{D},\widehat{y}\right) = \widetilde{\mathcal{J}}_M\left(\mathcal{D},\widehat{y}\right)$$

$$+\mathrm{tr}\left\{\left(\boldsymbol{F}^{\mathsf{T}}\boldsymbol{R}^{-1}\mathbf{F}\right)^{-1}\left[\int_{\mathcal{X}}\boldsymbol{f}\left(\boldsymbol{x}\right)\boldsymbol{f}^{\mathsf{T}}\left(\boldsymbol{x}\right)\mathrm{d}\boldsymbol{x} - 2\boldsymbol{A}^{\mathsf{T}}\int_{\mathcal{X}}\phi\left(\boldsymbol{x}\right)\boldsymbol{f}^{\mathsf{T}}\left(\boldsymbol{x}\right)\mathrm{d}\boldsymbol{x} + \boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\right]\right\}, \tag{4}$$

where $\widetilde{\mathcal{J}}_M$ is the criterion previously derived for the simple Kriging model, $\boldsymbol{A} = \boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathsf{T}}\boldsymbol{R}^{-1}\mathbf{F}$ and $\phi\left(\boldsymbol{x}\right) = \left[\varphi_1\left(\boldsymbol{x}\right),\ldots,\varphi_M\left(\boldsymbol{x}\right)\right]^{\mathsf{T}}$.

SFU

Approximate IMSPE Criterion for the Universal Kriging Model

Expression (4) is somewhat simplified by substituting $\boldsymbol{f}(\boldsymbol{x}) = 1$ and $\boldsymbol{F} = \mathbf{1}_n$ to

$$\widetilde{\mathcal{H}}_M\left(\mathcal{D}, \widehat{y}\right) = \widetilde{\mathcal{J}}_M\left(\mathcal{D}, \widehat{y}\right) + \frac{1}{\mathbf{1}_n^{\mathsf{T}} \boldsymbol{R}^{-1} \mathbf{1}_n} \left\{ \mathrm{vol}\left(\mathcal{X}\right) - 2\boldsymbol{a}^{\mathsf{T}}\gamma + \boldsymbol{a}^{\mathsf{T}}\boldsymbol{a} \right\} \quad ,$$

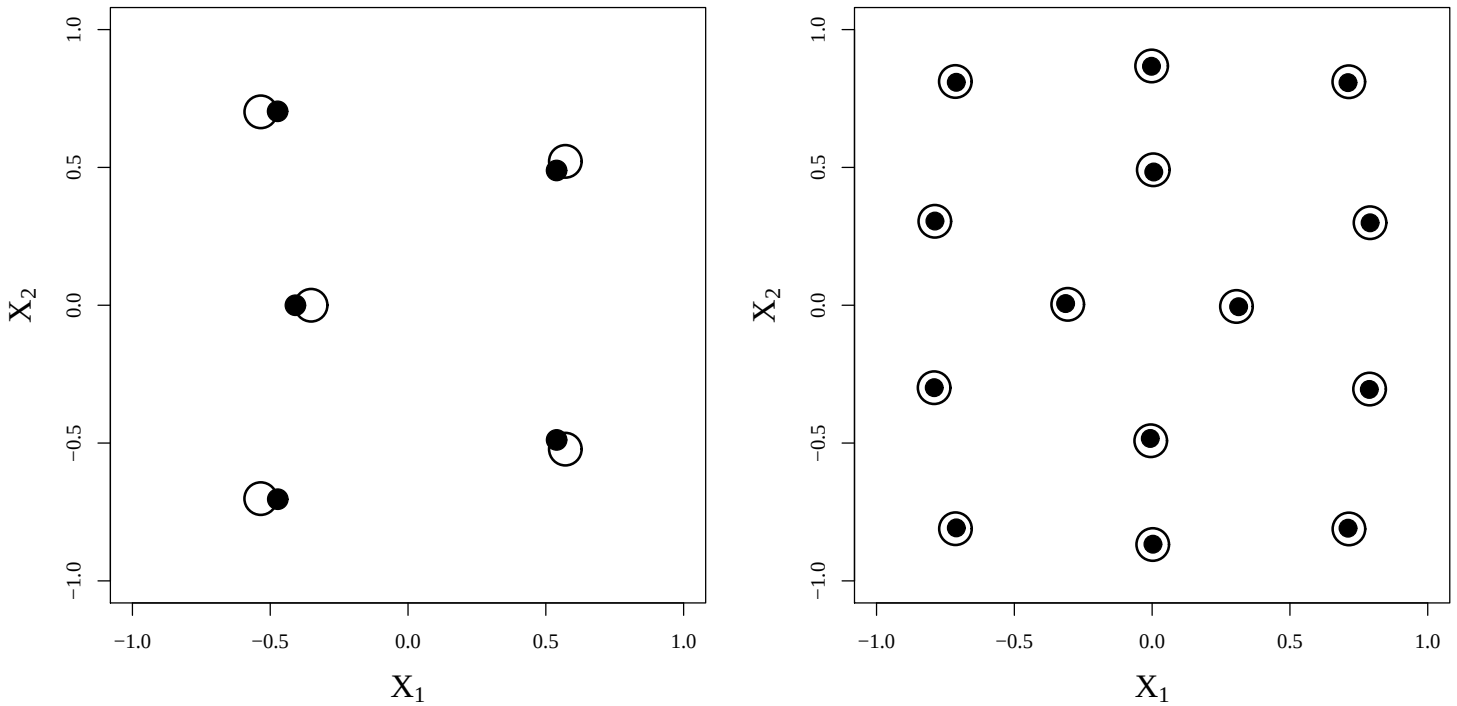
where

$$\boldsymbol{a} = \boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathsf{T}}\boldsymbol{R}^{-1}\mathbf{1}_n \quad \text{and} \quad \gamma = \left[\int_{\mathcal{X}} \varphi_1(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \ldots, \int_{\mathcal{X}} \varphi_M(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \right]^{\mathsf{T}} \quad .$$

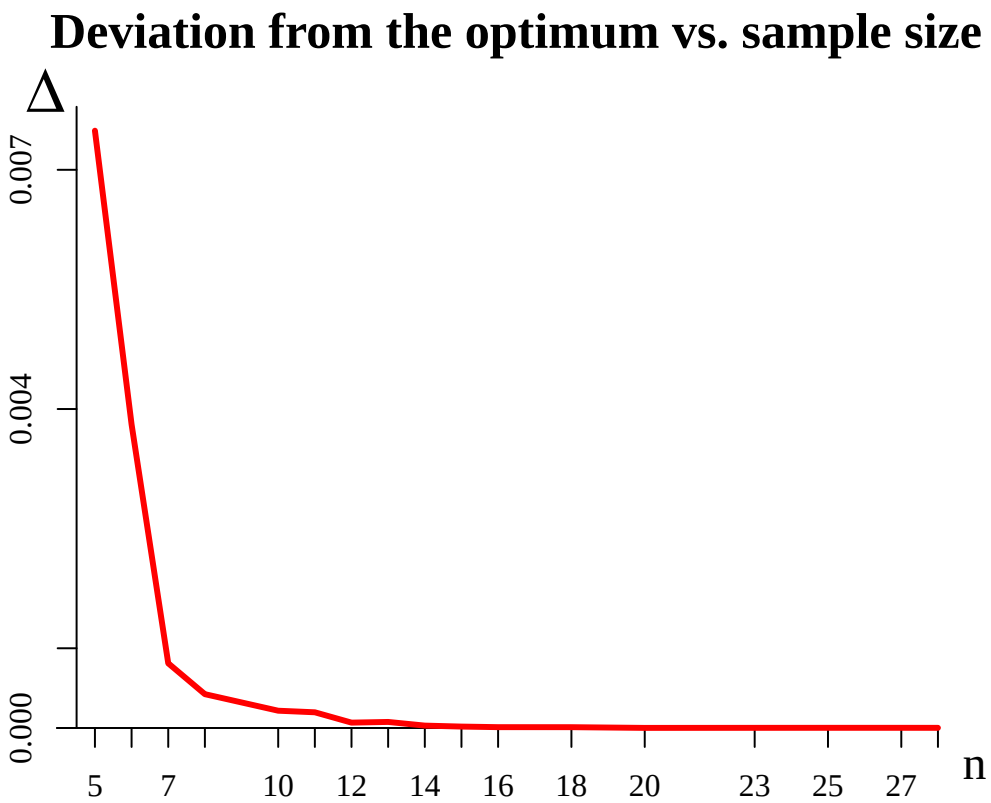
- The vector of integrals γ only needs to be evaluated once.
- Almost identical designs to those obtained for the simple Kriging model, for a reasonably large n.

SFU

With and Without Estimating the Intercept



With and Without Estimating the Intercept (cont.)



Outline

SFU

Bayesian A-optimal Designs

- In classical DOE, a design is called A-optimal if it minimizes $\text{tr}\left\{\text{var}\left(\widehat{\boldsymbol{\theta}}\right)\right\}$, where $\widehat{\boldsymbol{\theta}}$ is the vector of estimators for the model parameters.
- For our model

$$y(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^{\mathsf{T}}\beta + Z(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^{\mathsf{T}}\beta + \sum_{k=1}^{\infty}\alpha_k\varphi_k(\boldsymbol{x}) \quad , \quad \alpha_k \sim \mathcal{N}(0,\lambda_i) \quad ,$$

  we may consider the Bayesian A-optimality criterion

$$\mathcal{Q}\left(\mathcal{D},\widehat{y}\right) = \sum_{k=1}^{\infty}\text{var}\left\{\alpha_k\middle|\boldsymbol{y}\right\} \quad .$$

- When β is a known vector, the IMSPE and the A-optimality criteria coincide.

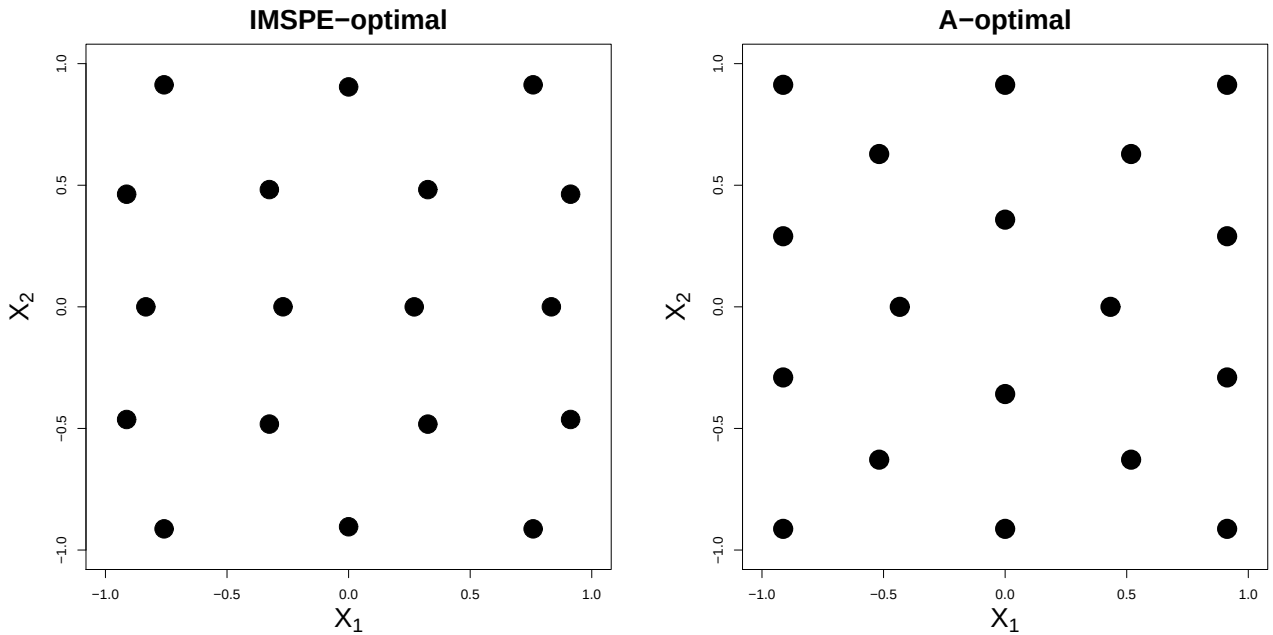
Bayesian A-optimal Designs (cont.)

Proposition

If $\pi(\beta) \propto 1$,

$$\mathcal{Q}\left(\mathcal{D}, \widehat{y}\right) = \mathrm{vol}\left(\mathcal{X}\right) - \mathrm{tr}\left\{ \begin{bmatrix} \mathbf{0} & \boldsymbol{F}^{\mathsf{T}} \\ \boldsymbol{F} & \boldsymbol{R} \end{bmatrix}^{-1} \int_{\mathcal{X}} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{r}(\boldsymbol{x})\,\boldsymbol{r}^{\mathsf{T}}(\boldsymbol{x}) \end{bmatrix} \mathrm{d}\boldsymbol{x} \right\} .$$

(and hence we don't even need to derive the Mercer expansion)



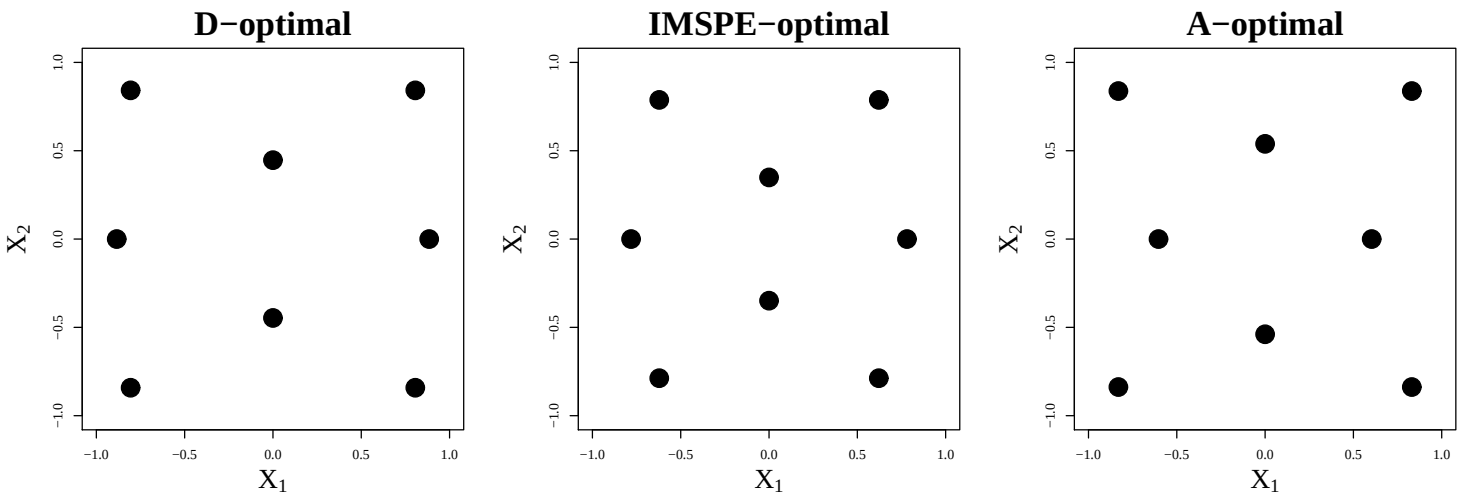
Bayesian D-optimal Designs

- In classical DOE, a design is called D-optimal if it minimizes the determinant of $\mathrm{var}\left(\widehat{\boldsymbol{\theta}}\right)$ (typically the inverse Fisher information).
- In our framework, an approximate Bayesian D-optimal design would be

$$\mathcal{D}^* = \underset{\mathcal{D}}{\operatorname{argmin}} \det\left(\mathrm{var}\left\{\alpha_1, \ldots, \alpha_M \middle| \boldsymbol{y}\right\}\right)$$

Where

$$\mathrm{var}\left\{\alpha_1, \ldots, \alpha_M \middle| \boldsymbol{y}\right\} = \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{\Phi}^{\mathsf{T}}\left[\boldsymbol{R}^{-1} - \boldsymbol{R}^{-1}\boldsymbol{F}\left(\boldsymbol{F}^{\mathsf{T}}\boldsymbol{R}^{-1}\boldsymbol{F}\right)^{-1}\boldsymbol{F}^{\mathsf{T}}\boldsymbol{R}^{-1}\right]\boldsymbol{\Phi}\boldsymbol{\Lambda} \quad .$$



Outline

SFU

Few comments before we go

- If $Z(\boldsymbol{x})$ is a zero mean non-Gaussian, weakly-stationary process with the same correlation structure
  - the Kriging predictor will no longer be the posterior mean, but will still be the BLUP.
  - The MSPE will no longer be the posterior variance, but all the results still hold.
- Inclusion of measurement error (i.e. *Nugget Term*)

$$y(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})^{\mathsf{T}}\beta + Z(\boldsymbol{x}) + \varepsilon(\boldsymbol{x}) \quad , \quad \varepsilon \sim \mathcal{N}\left(\boldsymbol{0}, \tau^2 \boldsymbol{I}\right) \quad , \quad \lambda = \tau^2/\sigma^2$$

will simply result in replacing $\boldsymbol{R}$ with $\boldsymbol{R} + \lambda \boldsymbol{I}$ everywhere, but the theory will cease to apply.

  - Great computational benefits for small values of λ.

SFU

Thank you!

SFU

References I

1. Andrianakis, Ioannis, & Challenor, Peter G. 2012. The effect of the nugget on Gaussian process emulators of computer models. *Computational Statistics and Data Analysis*, **56**(12), 4215–4228.
2. Atkinson, A.C., Doney, A.N., & Tobias, R.D. 2007. *Optimum Experimental Designs, with SAS*. Oxford Statistical Science Series.
3. Bursztyn, D., & Steinberg, D. M. 2004. Data analytic tools for understanding random field regression models. *Technometrics*, **46**(4), 411–420.
4. Cressie, Noel A. C. 1993. *Statistics for Spatial Data*. Wiley-Interscience.
5. Frauenfelder, Philipp, Schwab, Christoph, & Todor, Radu Alexandru. 2005. Finite Elements for Elliptic Problems with Stochastic Coefficients. *Computer Methods in Applied Mechanics and Engineering*, **194**(2), 205–228.
6. Guenther, Ronald B., & Lee, John W. 1996. *Partial Differential Equations of Mathematical Physics and Integral Equations*. Courier Dover Publications.
7. Jones, Brad, & Johnson, Rachel T. 2009. The Design and Analysis of the Gaussian Process Model. *Quality and Reliability Engineering International*, **25**(5), 515–524.
8. Journel, A., & Rossi, M. 1989. When Do We Need a Trend Model in Kriging? *Mathematical Geology*, **21**(7), 715–739.

SFU

References II

9. Linkletter, Crystal, Bingham, Derek, Hengartner, Nicholas, Higdon, David, & Ye, Kenny Q. 2006. Variable Selection for Gaussian Process Models in Computer Experiments. *Technometrics*, **48**(4), 478–490.

10. Loeppky, Jason L., Sacks, Jerome, & Welch, William J. 2009. Choosing the Sample Size of a Computer Experiment: A Practical Guide. *Technometrics*, **51**(4), 366–376.

11. Loeppky, Jason L., Moore, Leslie M., & Williams, Brian J. 2010. Batch Sequential Designs for Computer Experiments. *Journal of Statistical Planning and Inference*, **140**(6), 1452–1464.

12. Picheny, Victor, Ginsbourger, David, Roustant, Olivier, Haftka, Raphael T., & Kim, Nam H. 2010. Adaptive Designs of Experiments for Accurate Approximation of a Target Region. *Journal of Mechanical Design*, **132**(7), 071008.

13. Rasmussen, Carl Edward, & Williams, Chris. 2006. *Gaussian Processes for Machine Learning*. The MIT Press.

14. Sacks, Jerome, Welch, William J., Mitchell, Toby J., & Wynn, Henry P. 1989a. Design and Analysis of Computer Experiments. *Statistical Science*, **4**(4), 409–423.

15. Sacks, Jerome, Schiller, Susannah B., & Welch, William J. 1989b. Designs for Computer Experiments. *Technometrics*, **31**(1), 41–47.

SFU

References III

16. Santner, Thomas J., Williams, Brian J., & Notz, William I. 2003. *The Design and Analysis of Computer Experiments*. Springer.

17. Shewry, Michael C., & Wynn, Henry P. 1987. Maximum Entropy Sampling. *Journal of Applied Statistics*, **14**(2), 165–170.

18. Singhee, Amith, Singhal, Sonia, & Rutenbar, Rob A. 2008. Exploiting Correlation Kernels for Efficient Handling of Intra-Die Spatial Correlation, with Application to Statistical Timing. *Pages 856–861 of: Proceedings of the conference on Design, Automation and Test in Europe*.

19. Wahba, Grace. 1990. *Spline Models for Observational Data*. Society for Industrial and Applied Mathematics.

SFU