

# Correlated Optimum Design with Parametrized Covariance Function: Justification of the Fisher Information Matrix and of the Method of Virtual Noise



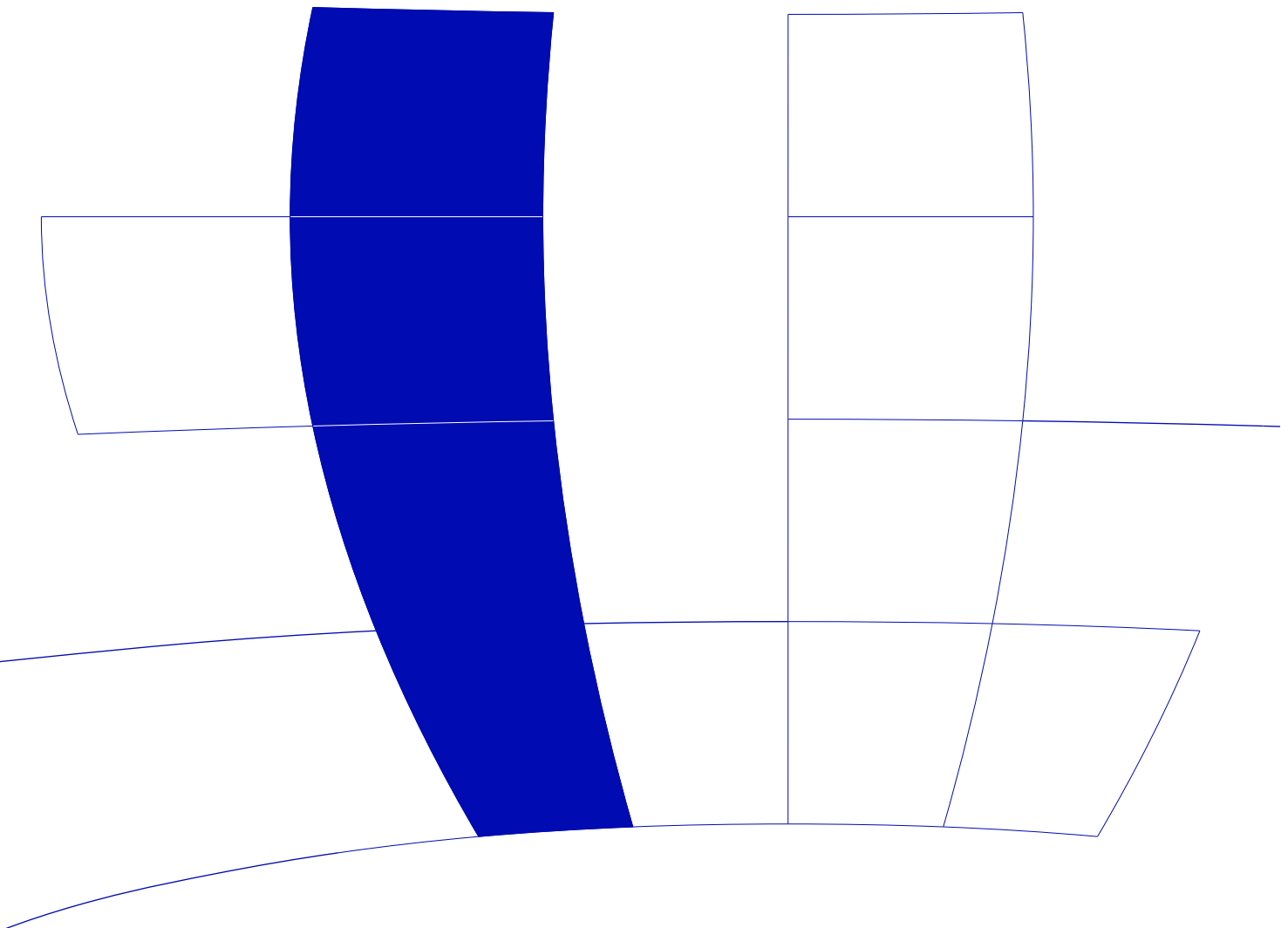
**Andrej Pázman**

Department of Statistics and Mathematics  
Wirtschaftsuniversität Wien

**Research Report Series**

Report 5  
June 2004

<http://statistik.wu-wien.ac.at/>



**Correlated optimum design with parametrized covariance  
function:  
Justification of the use of the Fisher information matrix and of  
the method of virtual noise**

*Andrej Pázman, Comenius University, Bratislava,  
pazman@fmph.uniba.sk*

*Abstract*

*We consider observations of a random field (or a random process), which is modeled by a nonlinear regression with a parametrized mean (or trend) and a parametrized covariance function. In the first part we show that under the assumption that the errors are normal with small variances, even when the number of observations is small, the ML estimators of both parameters are approximately unbiased, uncorrelated, with variances given by the inverse of the Fisher information matrix. In the second part we are extending the result of [4] to the case of parametrized covariance function, namely we prove that the optimum designs with and without the presence of the virtual noise are identical. This in principle justify the use the method of virtual noise as a computational device also in this case.*

## **1. Introduction**

Consider a regression model of the form

$$y(x_i) = \eta(\theta, x_i) + \varepsilon(x_i) \tag{1}$$

with the points  $x_1, \dots, x_N$  (=the design) taken from a set  $\mathcal{X}$  (= the design space). The vector parameter  $\theta = (\theta_1, \dots, \theta_p)^T \in \Theta$  is unknown, and  $\eta(\cdot)$  is a known function. The model is supposed to be without systematic errors (i.e.  $E(\varepsilon(x_i)) = 0$ ), and the variance-covariance structure of the observed variables  $y(x_i)$

$$Cov(y(x_i), y(x_j)) = C(x_i, x_j, \beta)$$

depends on another unknown vector parameter  $\beta = (\beta_1, \dots, \beta_q)^T \in B$ .

The design  $x_1, \dots, x_N$  is good if it gives precise estimators of the parameters. If the useful information is only in the parameter  $\theta$ , this precision is measured by some optimality criterion expressed as a function of the mean

squares error matrix of their estimator  $\hat{\theta}$

$$MSE_{\theta}(\hat{\theta}) = E_{\theta} \left[ (\hat{\theta} - \theta) (\hat{\theta} - \theta)^T \right]$$

Similarly for  $\beta$ , or both  $\theta$  and  $\beta$  if all these parameters are important. So the first problem when optimizing designs is to express the mean squares error matrix of the parameter estimators.

In the particular case of a linear model,

$$\eta(\theta, x_i) = f^T(x_i) \theta$$

with  $\Theta = R^p$ , and with covariances and variances not depending on  $\beta$ , the minimum variance unbiased estimator of  $\theta$  is the weighted least squares estimator

$$\begin{aligned} \hat{\theta} &= \arg \min_{\theta \in R^p} [y - \eta(\theta)]^T C^{-1} [y - \eta(\theta)] \\ &= M^{-1} F^T y \end{aligned}$$

where  $y$  is the vector of observed variables,  $\eta(\theta) = (\eta(\theta, x_1), \dots, \eta(\theta, x_N))^T$ ,  $C$  is a given  $N \times N$  matrix with entries  $Cov(y(x_i), y(x_j))$ ,  $F^T = (f(x_1), \dots, f(x_N))$ , and

$$M = F^T C^{-1} F$$

is the information matrix. In this case the MSE of  $\hat{\theta}$  is equal to its variance matrix  $Var(\hat{\theta})$  and

$$Var(\hat{\theta}) = M^{-1}.$$

So the optimality criterion is usually expressed as a function of the information matrix  $M$ , in a form

$$\Phi(M)$$

(e.g.  $\Phi(M) = -\ln \det(M)$  for the D-optimality criterion,  $\Phi(M) = tr(M^{-1})$  for the A-optimality criterion, etc.), and it does not depend on  $\theta$ .

In the **nonlinear** regression model with **uncorrelated observations** it is standard to base the optimality criteria again on the information matrix. This is justified by the fact, that in the uncorrelated case **replications** of observations are allowed, and **asymptotically** (for large numbers of replications), under some regularity conditions, the maximum likelihood estimators are asymptotically normally distributed, unbiased, and with the variance matrix equal to the inverse of the Fisher information matrix.

This argumentation can not be used in case of correlated observations, where replication as a rule are not allowed, and asymptotic approximations are not justified. Nevertheless, as we shall show, **in case that the errors are normally distributed with sufficiently small variances, one can even for small samples consider the maximum likelihood estimator having the same mean and variance as in the asymptotic case.** That means that the mean square error matrix is approximately equal to the inverse of the information matrix. This however does not mean that the estimator is approximately normal, as we see below.

## 2. The Fisher information matrix.

For a fixed design consider the nonlinear regression model in the vector form

$$\begin{aligned} y &= \eta(\theta) + \varepsilon \\ \varepsilon &\sim \mathcal{N}(0, C(\beta)) \end{aligned} \quad (2)$$

We suppose that the design is such that the mapping  $\theta \in \Theta \rightarrow \eta(\theta) \in R^N$  is one-to-one and the  $N \times N$  covariance matrix  $C(\beta)$  with entries  $C(x_i, x_j, \beta)$  is nonsingular. Suppose also that  $\bar{\theta}$  and  $\bar{\beta}$ , the true values of  $\theta$  and  $\beta$ , are points of the interiors  $\text{int}(\Theta)$ , resp.  $\text{int}(B)$  of  $\Theta$  resp. of  $B$ . We consider the MLE

$$(\hat{\theta}, \hat{\beta})^T = (\hat{\theta}(y), \hat{\beta}(y))^T = \arg \max_{\theta \in \Theta, \beta \in B} \ln f(y | \theta, \beta)$$

where

$$-\ln f(y | \theta, \beta) = \frac{1}{2} \left\{ [y - \eta(\theta)]^T C^{-1}(\beta) [y - \eta(\theta)] + \ln \det(C(\beta)) + N \ln(2\pi) \right\} \quad (3)$$

By taking the derivatives (see Appendix) one obtains that the Fisher information matrix of model (2) is equal to

$$\begin{aligned} M(\theta, \beta) &= E_{\theta, \beta} \left\{ - \begin{pmatrix} \frac{\partial^2 \ln f(y|\theta, \beta)}{\partial \theta \partial \theta^T} & \frac{\partial^2 \ln f(y|\theta, \beta)}{\partial \theta \partial \beta^T} \\ \frac{\partial^2 \ln f(y|\theta, \beta)}{\partial \beta \partial \theta^T} & \frac{\partial^2 \ln f(y|\theta, \beta)}{\partial \beta \partial \beta^T} \end{pmatrix} \right\} \\ &= \begin{pmatrix} \frac{\partial \eta^T(\theta)}{\partial \theta} C^{-1}(\beta) \frac{\partial \eta(\theta)}{\partial \theta^T} & 0 \\ 0 & \frac{1}{2} \text{tr} \left\{ C^{-1}(\beta) \frac{\partial C(\beta)}{\partial \beta} C^{-1}(\beta) \frac{\partial C(\beta)}{\partial \beta^T} \right\} \end{pmatrix} \quad (4) \end{aligned}$$

### 3. The regression model represented by an exponential family.

For further use we need to represent the regression model by an exponential family. We can write

$$\begin{aligned} \ln f(y | \theta, \beta) &= y^T C^{-1}(\beta) \eta(\theta) - \frac{1}{2} \text{tr} \{ y y^T C^{-1}(\beta) \} \\ &\quad - \frac{1}{2} \eta^T(\theta) C^{-1}(\beta) \eta(\theta) - \frac{1}{2} \ln \det [C(\beta)] - \frac{N}{2} \ln(2\pi) \end{aligned}$$

Let us denote

$$\begin{aligned} t(y) &= \begin{pmatrix} t_1(y) \\ t_2(y) \end{pmatrix} = \begin{pmatrix} y \\ \text{vec}(y y^T) \end{pmatrix} \\ \gamma_1(\theta, \beta) &= C^{-1}(\beta) \eta(\theta) \\ \gamma_2(\theta, \beta) &= -\frac{1}{2} \text{vec}[C^{-1}(\beta)] \\ \gamma(\theta, \beta) &= \begin{pmatrix} \gamma_1(\theta, \beta) \\ \gamma_2(\theta, \beta) \end{pmatrix} \end{aligned}$$

Notice that the mapping

$$C \rightarrow \gamma_2 = -\frac{1}{2} \text{vec}[C^{-1}]$$

is one-to-one. So we can define a function

$$\kappa(\gamma) = \kappa(\gamma_1, \gamma_2) = \frac{1}{2} \ln \det(C) + \frac{1}{2} \gamma_1^T C \gamma_1 + \frac{N}{2} \ln(2\pi)$$

With these notations we obtain

$$f(y | \theta, \beta) = \exp \left\{ t^T(y) \gamma(\theta, \beta) - \kappa[\gamma(\theta, \beta)] \right\}$$

This means that the family

$$\{f(y | \theta, \beta) : \theta \in \Theta, \beta \in B\}$$

is an exponential family where  $t(y)$  is a sufficient statistics, and  $\gamma(\theta, \beta)$  is the canonical function (cf. [1]). Important here are the following known relations: the mean and the variance of the sufficient statistics are equal to

$$E_{\theta, \beta}[t(y)] \equiv \mu(\theta, \beta) = \left[ \frac{\partial \kappa(\gamma)}{\partial \gamma} \right]_{\gamma=\gamma(\theta, \beta)} \quad (5)$$

$$\text{Var}_{\theta, \beta}[t(y)] = \left[ \frac{\partial^2 \kappa(\gamma)}{\partial \gamma \partial \gamma^T} \right]_{\gamma=\gamma(\theta, \beta)} \quad (6)$$

and the Fisher information matrix (for the parameters  $(\theta, \beta)$ ) can be expressed in the form

$$\begin{aligned} M(\theta, \beta) &= \begin{pmatrix} \frac{\partial \mu^T(\theta, \beta)}{\partial \theta} \frac{\partial \gamma(\theta, \beta)}{\partial \theta^T} & \frac{\partial \mu^T(\theta, \beta)}{\partial \theta} \frac{\partial \gamma(\theta, \beta)}{\partial \beta^T} \\ \frac{\partial \mu^T(\theta, \beta)}{\partial \beta} \frac{\partial \gamma(\theta, \beta)}{\partial \theta^T} & \frac{\partial \mu^T(\theta, \beta)}{\partial \beta} \frac{\partial \gamma(\theta, \beta)}{\partial \beta^T} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\partial \gamma^T(\theta, \beta)}{\partial \theta} \\ \frac{\partial \gamma^T(\theta, \beta)}{\partial \beta} \end{pmatrix} \left[ \frac{\partial^2 \kappa(\gamma)}{\partial \gamma \partial \gamma^T} \right]_{\gamma=\gamma(\theta, \beta)} \begin{pmatrix} \frac{\partial \gamma(\theta, \beta)}{\partial \theta^T} & \frac{\partial \gamma(\theta, \beta)}{\partial \beta^T} \end{pmatrix} \end{aligned} \quad (7)$$

#### 4. Approximation of the ML estimator in exponential families when the variances of the components of the sufficient statistics are small.

In an exponential family the estimator of the parameters can be expressed as a function of the sufficient statistics  $t$

$$\begin{pmatrix} \hat{\theta} \\ \hat{\beta} \end{pmatrix} = \arg \max_{\theta \in \Theta, \beta \in B} \left\{ t^T \gamma(\theta, \beta) - \kappa[\gamma(\theta, \beta)] \right\} \quad (8)$$

The domain where this estimator is defined is equal to

$$T = \left\{ t = \begin{pmatrix} y \\ \text{vec}(yy^T) \end{pmatrix} : y \in R^N \right\}$$

Define

$$T^* = \left\{ t = \begin{pmatrix} y \\ \text{vec}(Z) \end{pmatrix} : y \in R^N, Z \in R^{N \times N} \text{ and positive semidefinite} \right\}$$

Denote by  $\begin{pmatrix} \theta^*(t) \\ \beta^*(t) \end{pmatrix}$  the extensions of  $\begin{pmatrix} \hat{\theta}(t) \\ \hat{\beta}(t) \end{pmatrix}$  from  $T$  to the set  $T^*$

$$\begin{aligned} \begin{pmatrix} \theta^*(t) \\ \beta^*(t) \end{pmatrix} &= \arg \max_{\theta \in \Theta, \beta \in B} \left\{ y^T C^{-1}(\beta) \eta(\theta) - \frac{1}{2} \text{tr} [Z C^{-1}(\beta)] \right. \\ &\quad \left. - \frac{1}{2} \eta^T(\theta) C^{-1}(\beta) \eta(\theta) - \frac{1}{2} \ln \det [C(\beta)] \right\} \end{aligned} \quad (9)$$

Notice that this is just a mapping, not an estimator. The idea is to approximate it by a Taylor expansion around the point

$$\bar{\mu} = \mu(\bar{\theta}, \bar{\beta}) = \begin{pmatrix} E_{\bar{\theta}, \bar{\beta}}(y) \\ \text{vec} [E_{\bar{\theta}, \bar{\beta}}(yy^T)] \end{pmatrix} = \begin{pmatrix} \eta(\bar{\theta}) \\ \text{vec} [C(\bar{\beta}) + \eta(\bar{\theta}) \eta^T(\bar{\theta})] \end{pmatrix}$$

So we have

$$\begin{aligned} \theta^*(t) &= \theta^*[\bar{\mu}] + \frac{\partial \theta^*(t)}{\partial t^T} \Big|_{t=\bar{\mu}} (t - \bar{\mu}) \\ &+ \frac{1}{2} (t - \bar{\mu})^T \frac{\partial^2 \theta^*(t)}{\partial t \partial t^T} \Big|_{t=q} (t - \bar{\mu}) \end{aligned}$$

and similarly for  $\beta^*(t)$ . Here  $q$  is a point between  $t$  and  $\bar{\mu}$ . Since  $\theta^*(t)$  is just an extension of  $\hat{\theta}(t)$  we can write

$$\hat{\theta}(t) = \theta^*[\bar{\mu}] + \frac{\partial \theta^*(t)}{\partial t^T} \Big|_{t=\bar{\mu}} (t - \bar{\mu}) : t \in T$$

We neglected the quadratic term since the variances of the components of the statistics  $t$  are small when the variances of  $y(x_i)$  are small (see Appendix). Similarly

$$\hat{\beta}(t) = \beta^*[\bar{\mu}] + \frac{\partial \beta^*(t)}{\partial t^T} \Big|_{t=\bar{\mu}} (t - \bar{\mu}) : t \in T$$

These are the approximation that we shall use.

#### 4.1 The expressions for $\theta^*[\bar{\mu}]$ and $\beta^*[\bar{\mu}]$ .

From (9) we obtain

$$\begin{aligned} \begin{pmatrix} \theta^*[\bar{\mu}] \\ \beta^*[\bar{\mu}] \end{pmatrix} &= \arg \max_{\theta \in \Theta, \beta \in B} \left\{ \eta^T(\bar{\theta}) C^{-1}(\beta) \eta(\theta) - \frac{1}{2} \text{tr} \left[ (C(\bar{\beta}) + \eta(\bar{\theta}) \eta^T(\bar{\theta})) C^{-1}(\beta) \right] \right. \\ &\quad \left. - \frac{1}{2} \eta^T(\theta) C^{-1}(\beta) \eta(\theta) - \frac{1}{2} \ln \det [C(\beta)] \right\} \\ &= \arg \min_{\theta \in \Theta, \beta \in B} \left\{ \frac{1}{2} [\eta(\theta) - \eta(\bar{\theta})]^T C^{-1}(\beta) [\eta(\theta) - \eta(\bar{\theta})] \right. \\ &\quad \left. + \frac{1}{2} \text{tr} [C(\bar{\beta}) C^{-1}(\beta)] + \frac{1}{2} \ln \det [C(\beta)] \right\} \end{aligned}$$

Hence

$$\begin{aligned} \theta^*[\bar{\mu}] &= \bar{\theta} \\ \beta^*[\bar{\mu}] &= \arg \min_{\beta \in B} \left\{ \ln \det [C(\beta)] + \text{tr} [C(\bar{\beta}) C^{-1}(\beta)] \right\} = \bar{\beta} \end{aligned}$$

To prove the last equality we write in an abbreviated notation ( $C = C(\beta)$ ,  $\bar{C} = C(\bar{\beta})$ ):

$$\frac{\partial}{\partial \beta} \left\{ \ln \det [C] + \text{tr} [\bar{C} C^{-1}] \right\} = \text{tr} C^{-1} [C - \bar{C}] C^{-1} \frac{\partial C}{\partial \beta}$$

which is zero if  $\beta = \bar{\beta}$ , and

$$\begin{aligned} \frac{\partial^2}{\partial\beta\partial\beta^T} \left\{ \ln \det [C] + \text{tr} [\bar{C}C^{-1}] \right\}_{\beta=\bar{\beta}} &= \text{tr} \left[ C^{-1} \frac{\partial C}{\partial\beta} C^{-1} \frac{\partial C}{\partial\beta^T} \right]_{C=\bar{C}} \\ &= \text{tr} \left[ C^{-1/2} \frac{\partial C}{\partial\beta} C^{-1} \frac{\partial C}{\partial\beta^T} C^{-1/2} \right]_{C=\bar{C}} \end{aligned}$$

which is positive definite since  $C^{-1}$  is positive definite.

#### 4.2. The expressions for $\frac{\partial\theta^*(t)}{\partial t^T} \Big|_{t=\bar{\mu}}$ and $\frac{\partial\beta^*(t)}{\partial t^T} \Big|_{t=\bar{\mu}}$

By taking the derivatives with respect to  $\theta$  and  $\beta$  we obtain from (9) and using (5) the normal equation for  $\theta^*, \beta^*$

$$G(\theta, \beta, t)_{\theta=\theta^*(t), \beta=\beta^*(t)} = 0$$

where

$$G(\theta, \beta, t) = \begin{pmatrix} \frac{\partial\gamma^T(\theta, \beta)}{\partial\theta} \\ \frac{\partial\gamma^T(\theta, \beta)}{\partial\beta} \end{pmatrix} (t - \mu(\theta, \beta))$$

The functions  $\theta^*(t)$ ,  $\beta^*(t)$  are defined implicitly by the normal equation. From the implicit function theorem (cf. [5]) we obtain

$$\begin{pmatrix} \frac{\partial\theta^*(t)}{\partial t^T} \\ \frac{\partial\beta^*(t)}{\partial t^T} \end{pmatrix}_{t=\bar{\mu}} = - \left[ \begin{pmatrix} \frac{\partial G}{\partial\theta^T} & \frac{\partial G}{\partial\beta^T} \end{pmatrix}^{-1} \frac{\partial G}{\partial t^T} \right]_{\bar{\theta}, \bar{\beta}, \bar{\mu}}$$

From the definition of  $G$  we have

$$\begin{aligned} \begin{pmatrix} \frac{\partial G}{\partial\theta^T} & \frac{\partial G}{\partial\beta^T} \end{pmatrix}_{\bar{\theta}, \bar{\beta}, \bar{\mu}} &= - \begin{pmatrix} \frac{\partial\gamma^T}{\partial\theta} \\ \frac{\partial\gamma^T}{\partial\beta} \end{pmatrix}_{\bar{\theta}, \bar{\beta}} \begin{pmatrix} \frac{\partial\mu}{\partial\theta^T} & \frac{\partial\mu}{\partial\beta^T} \end{pmatrix}_{\bar{\theta}, \bar{\beta}} \\ &= -M(\bar{\theta}, \bar{\beta}) \\ \frac{\partial G}{\partial t^T} \Big|_{\bar{\theta}, \bar{\beta}, \bar{\mu}} &= \begin{pmatrix} \frac{\partial\gamma^T}{\partial\theta} \\ \frac{\partial\gamma^T}{\partial\beta} \end{pmatrix}_{\bar{\theta}, \bar{\beta}} \end{aligned}$$

So

$$\begin{pmatrix} \frac{\partial\theta^*(t)}{\partial t^T} \\ \frac{\partial\beta^*(t)}{\partial t^T} \end{pmatrix}_{t=\bar{\mu}} = M^{-1}(\bar{\theta}, \bar{\beta}) \begin{pmatrix} \frac{\partial\gamma^T}{\partial\theta} \\ \frac{\partial\gamma^T}{\partial\beta} \end{pmatrix}_{\bar{\theta}, \bar{\beta}}$$



## 5. The approximation of the MSE

Summarizing the results we obtain that in case of small variances of  $y(x_i)$  the approximate expressions for the MLE are

$$\begin{pmatrix} \hat{\theta} \\ \hat{\beta} \end{pmatrix} \doteq \begin{pmatrix} \bar{\theta} \\ \bar{\beta} \end{pmatrix} + M^{-1}(\bar{\theta}, \bar{\beta}) \begin{pmatrix} \frac{\partial \gamma^T}{\partial \theta} \\ \frac{\partial \gamma^T}{\partial \beta} \end{pmatrix}_{\bar{\theta}, \bar{\beta}} (t - \bar{\mu}) \quad (10)$$

This gives

$$E_{\bar{\theta}, \bar{\beta}} \left[ \begin{pmatrix} \hat{\theta} \\ \hat{\beta} \end{pmatrix} \right] \doteq \begin{pmatrix} \bar{\theta} \\ \bar{\beta} \end{pmatrix}$$

$$\begin{aligned} Var_{\bar{\theta}, \bar{\beta}} \left[ \begin{pmatrix} \hat{\theta} \\ \hat{\beta} \end{pmatrix} \right] &\doteq M^{-1}(\bar{\theta}, \bar{\beta}) \begin{pmatrix} \frac{\partial \gamma^T}{\partial \theta} \\ \frac{\partial \gamma^T}{\partial \beta} \end{pmatrix}_{\bar{\theta}, \bar{\beta}} Var_{\bar{\theta}, \bar{\beta}}(t) \begin{pmatrix} \frac{\partial \gamma}{\partial \theta^T} & \frac{\partial \gamma}{\partial \beta^T} \end{pmatrix} M^{-1}(\bar{\theta}, \bar{\beta}) \\ &= M^{-1}(\bar{\theta}, \bar{\beta}) \end{aligned}$$

where we used (5) and (7). Hence within this approximation the estimators  $\hat{\theta}$  and  $\hat{\beta}$  are unbiased, uncorrelated with variances

$$\begin{aligned} Var_{\bar{\theta}, \bar{\beta}}(\hat{\theta}) &= \left[ \frac{\partial \eta^T(\theta)}{\partial \theta} C^{-1}(\beta) \frac{\partial \eta(\theta)}{\partial \theta^T} \right]_{\bar{\theta}, \bar{\beta}}^{-1} \\ \left\{ [Var_{\bar{\theta}, \bar{\beta}}(\hat{\beta})]^{-1} \right\}_{ij} &= \frac{1}{2} tr \left\{ C^{-1}(\beta) \frac{\partial C(\beta)}{\partial \beta_i} C^{-1}(\beta) \frac{\partial C(\beta)}{\partial \beta_j} \right\}_{\bar{\beta}} \end{aligned}$$

Notice that this does not mean that  $\hat{\beta}$  is approximately normally distributed, since, although  $\hat{\beta}$  is expressed as a linear function of  $t$ , by definition of  $t$  it is a quadratic function of the observed variables  $y(x_i)$ .

## 6. The method of virtual noise.

The method of virtual noise for computing optimum designs in linear models with known covariance function has been introduced in [3,4]. It consists in changing the discrete optimization problem to a continuous one. The corresponding algorithms have been presented in [2].

Here we shall show that the method can be extended to nonlinear models with parametrized covariance functions.

Instead of the original process

$$\begin{aligned} y(x_i) &= \eta(\theta, x_i) + \varepsilon(x_i) \\ E[\varepsilon(x_i)] &= 0, \text{Cov}[\varepsilon(x_i), \varepsilon(x_j)] = C(x_i, x_j, \beta) \end{aligned}$$

we consider a "virtual process

$$\begin{aligned} y^*(x_i) &= \eta(\theta, x_i) + \varepsilon(x_i) + \varepsilon_\xi(x_i) \\ E[\varepsilon_\xi(x_i)] &= 0, \text{Cov}[\varepsilon(x_i), \varepsilon_\xi(x_j)] = 0 \\ \text{Var}[\varepsilon_\xi(x_i)] &= \gamma \ln \frac{\xi_{\max}}{\xi(x)}, \text{ and } \text{Cov}[\varepsilon_\xi(x_i), \varepsilon_\xi(x_j)] = 0 \text{ for } i \neq j \end{aligned}$$

Here  $\varepsilon_\xi(x)$  is a virtual white noise which is used just for computational purposes. The parameter  $\gamma > 0$  is a tuning parameter (for numerical computations) and  $\xi$  is any probability measure defined on the design space  $\mathcal{X}$  and having a finite support equal to

$$S_\xi = \{x \in \mathcal{X} : \xi(x) > 0\}$$

Finally,  $\xi_{\max}$  means

$$\xi_{\max} = \max_{x \in \mathcal{X}} \xi(x)$$

In analogy to classical design theory we call the measure  $\xi$  "a design measure" although it has a different interpretation (cf. [3]).

Let  $A = \{x_1, x_2, \dots, x_N\}$  be a fixed set of design points (an "exact" design). Denote by  $C(A, \beta)$  the  $N \times N$  matrix with entries  $C(x_i, x_j, \beta)$ ;  $x_i, x_j \in A$ , and let us suppose that it is nonsingular. Let  $\xi$  be a design measure supported by  $A$ , i.e.  $S_\xi = A$ . Denote by  $V(\xi)$  the  $N \times N$  matrix with entries

$$\begin{aligned} \{V(\xi)\}_{ij} &= \gamma \ln \frac{\xi_{\max}}{\xi(x_i)} \text{ if } i = j \\ &= 0 \quad \text{if } i \neq j \end{aligned}$$

By  $M(\theta, \beta, A)$  we denote the information matrix defined in (4). It corresponds to the real process (1) when the design  $A$  is used. Denote by  $M^*(\theta, \beta, \xi)$  the Fisher information matrix corresponding to the vector of virtual observations  $(y^*(x_1), \dots, y^*(x_N))^T$ . Since the covariance matrix of this vector is equal to  $C(A, \beta) + V(\xi)$ , we have, in the same way as in (4)

$$M^*(\theta, \beta, \xi) = \begin{pmatrix} M_I & M_{II} \\ M_{II}^T & M_{III} \end{pmatrix}$$

where

$$\begin{aligned} M_I &= \frac{\partial \eta^T(\theta)}{\partial \theta} [C(A, \beta) + V(\xi)]^{-1} \frac{\partial \eta(\theta)}{\partial \theta^T} \\ M_{II} &= 0 \\ \{M_{III}\}_{ij} &= \frac{1}{2} \text{tr} \left\{ [C(A, \beta) + V(\xi)]^{-1} \frac{\partial C(A, \beta)}{\partial \beta_i} [C(A, \beta) + V(\xi)]^{-1} \frac{\partial C(A, \beta)}{\partial \beta_j} \right\} \end{aligned}$$

**Theorem 1.**

If  $A = S_\xi$ , then for any  $\theta, \beta$  and for any value of the tuning parameter  $\gamma$  we have

$$M^*(\theta, \beta, \xi) \leq M(\theta, \beta, A)$$

in the Lowner ordering of matrices. If the design measure  $\xi$  is uniform, i.e. if

$$\xi(x) = \xi_A(x) \equiv \frac{1}{\#A}; \quad x \in A$$

then the equality

$$M^*(\theta, \beta, \xi) = M(\theta, \beta, A)$$

holds. Conversely, if the matrix  $M(\theta, \beta, A)$  is nonsingular, and the equality holds, then  $\xi(x) = \xi_A(x)$ .

**Remark.** The theorem is a generalization (and slight extension) of Properties 2 and 3 in [4] to the case of parametrized covariance.

**Proof.**

We have to prove that for any  $u \in R^p$ ,  $v \in R^q$ ,  $u \neq 0$ ,  $v \neq 0$

$$\begin{aligned} & u^T \frac{\partial \eta^T(\theta)}{\partial \theta} [C(A, \beta) + V(\xi)]^{-1} \frac{\partial \eta(\theta)}{\partial \theta^T} u \\ & \leq u^T \frac{\partial \eta^T(\theta)}{\partial \theta} [C(A, \beta)]^{-1} \frac{\partial \eta(\theta)}{\partial \theta^T} u \\ & \quad \sum_{ij} v_i v_j \text{tr} \left\{ [C(A, \beta) + V(\xi)]^{-1} \frac{\partial C(A, \beta)}{\partial \beta_i} [C(A, \beta) + V(\xi)]^{-1} \frac{\partial C(A, \beta)}{\partial \beta_j} \right\} \\ & \leq \sum_{ij} v_i v_j \text{tr} \left\{ [C(A, \beta)]^{-1} \frac{\partial C(A, \beta)}{\partial \beta_i} [C(A, \beta)]^{-1} \frac{\partial C(A, \beta)}{\partial \beta_j} \right\} \end{aligned}$$

The first inequality is evident, since  $V(\xi) \geq 0$  implies  $C(A, \beta) + V(\xi) \geq C(A, \beta)$ , hence  $[C(A, \beta) + V(\xi)]^{-1} \leq [C(A, \beta)]^{-1}$  in the Lowner ordering.

The equality holds if and only if  $C(A, \beta) + V(\xi) = C(A, \beta)$  since the matrix  $\frac{\partial \eta^T(\theta)}{\partial \theta}$  is full rank. That means  $V(\xi) = 0$ , hence  $\xi$  is uniform.

To prove the second inequality, use the notation

$$Q = \sum_i v_i \frac{\partial C(A, \beta)}{\partial \beta_i}$$

We have to prove that

$$\begin{aligned} & \text{tr} \left\{ [C(A, \beta) + V(\xi)]^{-1} Q [C(A, \beta) + V(\xi)]^{-1} Q \right\} \\ & \leq \text{tr} \left\{ [C(A, \beta)]^{-1} Q [C(A, \beta)]^{-1} Q \right\} \end{aligned}$$

By the same arguments as above, we obtain first

$$Q [C(A, \beta) + V(\xi)]^{-1} Q \leq Q [C(A, \beta)]^{-1} Q$$

with equality for every  $v \in R^q$  if and only if  $\xi$  is uniform. Now, a basic result of matrix algebra states that there is a nonsingular  $N \times N$  matrix  $U$ , such that

$$\begin{aligned} U^T C(A, \beta) U &= I \\ U^T V(\xi) U &= \Lambda = \text{diag} \{ \lambda_1, \dots, \lambda_N \} \end{aligned}$$

with  $\lambda_1, \dots, \lambda_N$  non-negative numbers. Hence

$$\begin{aligned} & \text{tr} \left\{ [C(A, \beta) + V(\xi)]^{-1} Q [C(A, \beta) + V(\xi)]^{-1} Q \right\} \\ &= \text{tr} \left\{ U [I + \Lambda]^{-1} U^T Q [C(A, \beta) + V(\xi)]^{-1} Q \right\} \\ &= \sum_i \frac{1}{1 + \lambda_i} \left\{ U^T Q [C(A, \beta) + V(\xi)]^{-1} Q U \right\}_{ii} \\ &\leq \sum_i \left\{ U^T Q [C(A, \beta)]^{-1} Q U \right\}_{ii} \\ &= \text{tr} \left\{ [C(A, \beta)]^{-1} Q [C(A, \beta)]^{-1} Q \right\} \end{aligned}$$

with equality for all  $v \in R^q$  if and only if  $V(\xi) = 0$ , i.e.  $\xi$  is uniform.

**qed**

An optimality criterion  $\Phi(\cdot)$  is antimonotone, if

$$M \geq \bar{M} \Rightarrow \Phi(M) \leq \Phi(\bar{M})$$

Notice that all standard optimality criteria have this property. We have the following theorem, which is a generalization of the result in [4].

**Theorem 2a.**

Let  $\Phi$  be an antimonotone criterion such that

$$\Phi(M) = \infty \quad \text{for any singular p.s.d. matrix } M \quad (11)$$

Then for every  $\gamma > 0$  we have that all solutions of the minimization problem

$$\min_{\xi, \#S_\xi=N} \Phi(M^*(\theta, \beta, \xi)) \quad (12)$$

are uniform design measures supported by  $N$  points. Each of these supports is a solution of the problem

$$\min_{A \subset \mathcal{X}, \#A=N} \Phi(M(\theta, \beta, A)) \quad (13)$$

and each solution of (13) is a support of such an uniform design measure.

**Proof.** Since (11) is supposed, each solution of these minimization problems is attained at a nonsingular information matrix. Theorem 1 implies that the solution of the minimization problems must be an uniform design measure, and since for any uniform measure  $\xi$  we have  $M^*(\theta, \beta, \xi) = M(\theta, \beta, S_\xi)$ , the solutions of both problems coincide in the above sense. **qed**

If we omit the assumption (11) we obtain a direct generalization of Theorem 1 in [4].

**Theorem 2b.**

Suppose that

$$\min_{A \subset \mathcal{X}: \#A=N} \Phi[M(\theta, \beta, A)] < \infty$$

Then for every  $\gamma > 0$

a) If  $\xi^*$  is a solution of (12) and if  $S = S_{\xi^*}$ , then the design measure  $\xi_S$ , which is uniform on  $S$ , is also a solution of (12), and  $S$  is a solution of (13).

b) If  $A^*$  is a solution of (13), then the design measure  $\xi_{A^*}$ , which is uniform on  $A^*$ , is a solution of (12).

**Proof.** If  $\xi^*$  is a solution of (12) and if  $S = S_{\xi^*}$ , then for any design  $A$  with  $\#A = N$  we have

$$\begin{aligned} \Phi[M(\theta, \beta, S)] &= \Phi[M^*(\theta, \beta, \xi_S)] \leq \Phi[M^*(\theta, \beta, \xi^*)] \\ &\leq \Phi[M^*(\theta, \beta, \xi_{A^*})] \leq \Phi[M(\theta, \beta, A^*)] \end{aligned}$$

where we used the first part of Theorem 1. So the design measure  $\xi_S$  is a solution of (12). Conversely, suppose that  $A^*$  is a solution of (13). Then for every  $\xi$  we have by the first part of Theorem 1

$$\begin{aligned}\Phi [M^* (\theta, \beta, \xi)] &\geq \Phi [M (\theta, \beta, S_\xi)] \geq \Phi [M (\theta, \beta, A^*)] \\ &= \Phi [M^* (\theta, \beta, \xi_{A^*})]\end{aligned}$$

so  $\xi_{A^*}$  is a solution of (12). **qed**

A generalization of Property 1 in [4] to the case of parametrized covariance is in the following statement.

**Theorem 3.**

Let  $\xi_n$ ,  $n = 1, 2, \dots$  be a sequence of design measures having the same support  $S = S_{\xi_n}$  such that

$$\lim_{n \rightarrow \infty} \xi_n (x) = \mu (x), \quad x \in \mathcal{X}$$

where  $\mu$  is a design measure. Then for each  $\theta, \beta$ , and each  $\gamma > 0$

$$\lim_{n \rightarrow \infty} M^* (\theta, \beta, \xi_n) = M^* (\theta, \beta, \mu)$$

**Proof.** The support of  $\mu$  is either equal to  $S$ , or it is a subset of  $S$ . In the first case the statement is evident, for the second case it is proved in [4], in the proof of Property 1, that for fixed  $\beta$

$$\lim_{n \rightarrow \infty} [C (S, \beta) + V (\xi_n)]^{-1} = \begin{pmatrix} [C (S_\mu, \beta) + V (\mu)]^{-1} & 0 \\ 0 & 0 \end{pmatrix}$$

Hence the proof of the theorem follows directly from the definition of  $M^* (\theta, \beta, \xi)$ . **qed**

**References.**

[1] L.D.Brown (1986). Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory. Vol. 9 of Institute of Mathematical Statistics Lecture Notes - Monograph Series. Institute of Mathematical Statistics, Hayward, CA.

[2] W.G.Müller and A.Pázman (1999) An algorithm for computation of optimum designs under a given covariance structure. *Comput. Statist.* **14**, 197-211.

[3] A.Pázman and W.G.Müller (1998) A new interpretation of design measures. In: Atkinson et al (Eds.) Model Oriented Data Analysis 5, Physica Verlag, Heidelberg, p. 239-246.

[4] A.Pázman and W.G.Müller (2001). Optimum design of experiments subject to correlated errors. Stat. & Prob. Letters **52**, 29-34.

[5] M.Spivak (1965) Calculus on Manifolds, p. 41. W.A.Benjamin, Inc., Menlo Park, California.

[6] C.F.Wu (1981) Asymptotic theory of nonlinear least squares estimation. Annals of Statistics **9**, 501-513.

## Appendix

A1)

From (3) we obtain the first order derivatives

$$-\frac{\partial \ln f(y | \theta, \beta)}{\partial \theta_i} = -[y - \eta(\theta)]^T C^{-1}(\beta) \frac{\partial \eta(\theta)}{\partial \theta_i}; \quad i = 1, \dots, p$$

$$-\frac{\partial \ln f(y | \theta, \beta)}{\partial \beta_k} = -\frac{1}{2} [y - \eta(\theta)]^T C^{-1}(\beta) \frac{\partial C(\beta)}{\partial \beta_k} C^{-1}(\beta) [y - \eta(\theta)] + \frac{1}{2} \text{tr} \left\{ C^{-1}(\beta) \frac{\partial C(\beta)}{\partial \beta_k} \right\}; \quad k = 1, \dots, q$$

The second order derivatives are

$$\begin{aligned} -\frac{\partial^2 \ln f(y | \theta, \beta)}{\partial \theta_i \partial \theta_j} &= \frac{\partial \eta^T(\theta)}{\partial \theta_i} C^{-1}(\beta) \frac{\partial \eta(\theta)}{\partial \theta_j} - [y - \eta(\theta)]^T C^{-1}(\beta) \frac{\partial^2 \eta(\theta)}{\partial \theta_i \partial \theta_j} \\ -\frac{\partial^2 \ln f(y | \theta, \beta)}{\partial \beta_k \partial \beta_l} &= [y - \eta(\theta)]^T C^{-1}(\beta) \frac{\partial C(\beta)}{\partial \beta_k} C^{-1}(\beta) \frac{\partial C(\beta)}{\partial \beta_l} C^{-1}(\beta) [y - \eta(\theta)] \\ &\quad - \frac{1}{2} [y - \eta(\theta)]^T C^{-1}(\beta) \frac{\partial^2 C(\beta)}{\partial \beta_k \partial \beta_l} C^{-1}(\beta) [y - \eta(\theta)] \\ &\quad + \frac{1}{2} \text{tr} \left\{ C^{-1}(\beta) \frac{\partial^2 C(\beta)}{\partial \beta_k \partial \beta_l} \right\} \\ &\quad - \frac{1}{2} \text{tr} \left\{ C^{-1}(\beta) \frac{\partial C(\beta)}{\partial \beta_k} C^{-1}(\beta) \frac{\partial C(\beta)}{\partial \beta_l} \right\} \\ -\frac{\partial^2 \ln f(y | \theta, \beta)}{\partial \theta_i \partial \beta_l} &= [y - \eta(\theta)]^T C^{-1}(\beta) \frac{\partial C(\beta)}{\partial \beta_l} C^{-1}(\beta) \frac{\partial \eta(\theta)}{\partial \theta_i} \end{aligned}$$

A2)

When the variances of the observed variables  $y(x_i)$  are small, then the variances of all components of  $t(y)$  are small as well. Indeed, we have just to consider the components of  $t_2(y)$ . In an abbreviated notation we obtain

$$\begin{aligned} \text{Var} [y_i y_j] &= E [y_i y_j - C_{ij} - \eta_i \eta_j]^2 \\ &= E [\varepsilon_i \varepsilon_j + \varepsilon_i \eta_j + \varepsilon_j \eta_i - C_{ij}]^2 \\ &= E [\varepsilon_i \varepsilon_j]^2 + C_{ii} \eta_j^2 + C_{jj} \eta_i^2 + C_{ij} \eta_i \eta_j \end{aligned}$$

and by the Schwarz inequality we have  $E^2 [\varepsilon_i^2 \varepsilon_j^2] \leq E [\varepsilon_i^4] E [\varepsilon_j^4] = 9C_{ii}^2 C_{jj}^2$ ,  $|C_{ij}|^2 \leq C_{ii} C_{jj}$ . So  $\text{Var} [y_i y_j]$  tends to 0 when  $C_{ii}$  and  $C_{jj}$  tend to 0.