OPTIMUM DESIGNS FOR ILL-CONDITIONED MODELS:

K-OPTIMALITY AND STABLE PARAMETERIZATIONS

Anthony Atkinson, July 2023

Joint work with Belmiro Duarte, Instituto Politécnico de Coimbra, Portugal

CI AND NONLINEAR LEAST SQUARES

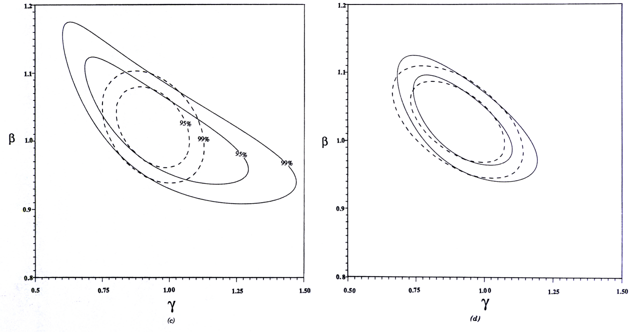


Figure: Continuous line - SS contour, dashed line linearized model. Two different designs for MM model from Seber and Wild (2003, p.114)

P. 113 has more curved SS contours from worse designs, but omits the dashed lines promised in the figure caption.

PROBLEMS WITH NONLINEAR LEAST SQUARES

1. Some NLLS problems give SS contours which are twisting narrow valleys in parameter space.
2. Numerical problems - convergence of algorithms
3. Inferential problems: calculation of SS contours and content of such regions.
4. Reparameterize to obtain approximately elliptical contours, which will be close to those from linearized model

STABLE PARAMETERIZATION

1. *Parametric transformation* — maps the original vector of parameters $\boldsymbol{\theta}$ into $\boldsymbol{\vartheta} \in \mathbb{R}^{n_\theta}$: $\mathcal{T} : \boldsymbol{\theta} \mapsto \boldsymbol{\vartheta}$.
2. $\mathcal{T}$ might include linear and nonlinear functions.
3. *Stable parameter vector* — a set of parameters that, after transformation is, for a given design and response model, less intercorrelated than the estimates of the original model structure.
4. Measured by the conditioning of the dispersion matrices, or graphically by the orientation and eccentricity of the likelihood contours.

STABLE PARAMETERIZATION: MICHAELIS-MENTEN 1

1. How to find a stable parameterization? Ross uses responses at equally spaced values in (one-dimensional) x.
2. Ross (1990, Cap.2) gives an example with the Michaelis-Menten rate model
$$\mathbb{E}(y) = \frac{\theta_2\, x}{\theta_3 + x}. \tag{1}$$
Using data at $x = [1, \ldots, 7]$, let $\vartheta_1 = \hat{y}(x = 3)$ and $\vartheta_2 = \hat{y}(x = 6)$. Then
$$\mathbb{E}(y) = \frac{\vartheta_1 \vartheta_2\, x}{(2\vartheta_1 - \vartheta_2)x + 6(\vartheta_2 - \vartheta_1)}. \tag{2}$$
3. **NOTE!** This parametric transformation does not change the relationship between y and x. (It may change the estimated relationship)

STABLE PARAMETERIZATION: MICHAELIS-MENTEN 2

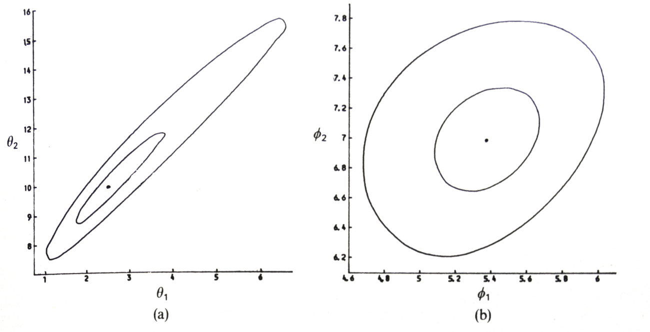


Figure: Left: original parameterization (1); right: stable parameterization (2). 60% and 95% SS based CI. Ross, p.15

STABLE PARAMETERIZATION: LINEAR REGRESSION

$$\mathbb{E}(y) = \beta_0 + \beta_1 x,$$

Suppose x values in a small interval around 10^6; then $\hat{\beta}_0$ and $\hat{\beta}_1$ highly correlated. But $\mathcal{T} : (\beta_0, \beta_1) \mapsto (\alpha - \beta_1 \bar{x}, \beta_1)$ gives the model

$$\mathbb{E}(y) = \alpha + \beta_1 (x - \bar{x}),$$

for which $\hat{\alpha}$ and $\hat{\beta}_1$ are uncorrelated.

K-OPTIMAL DESIGNS

1. Instead of means, for example, at 3 and 6, use optimal design to find design points and weights for a given model (not data).
2. The K–optimality criterion minimizes the condition number of the FIM, defined as the ratio

$$\gamma[\mathcal{M}(\xi|\mathbf{x},\boldsymbol{\theta})] = \lambda_{\mathsf{max}}[\mathcal{M}(\xi|\mathbf{x}\boldsymbol{\theta})]/\lambda_{\mathsf{min}}[\mathcal{M}(\xi|\mathbf{x},\boldsymbol{\theta})],$$

   of extreme eigenvalues of the FIM. The design problem is

$$\xi_K = \arg\min_{\xi\in\Xi} \frac{\lambda_{\mathsf{max}}[\mathcal{M}(\xi|\mathbf{x},\mathbf{p})]}{\lambda_{\mathsf{min}}[\mathcal{M}(\xi|\mathbf{x},\mathbf{p})]}. \quad (3)$$

3. The formulation for K–optimal designs was first considered in Ye and Zhou (2013)

PARAMETRIC TRANSFORMATIONS TO STABLE PARAMETERS

1. Techniques to find parametric transformations to vectors of stable parameters
2. Michaelis-Menten example deceptively simple since can solve algebraic equations
3. Reparameterization uses a family of symbolic curves constructed to pass through a set of previously chosen points in the domain of the regressors.
4. To find the set of stable parameters use the support points of the K–optimal design Two distinct approaches to the numerical problem

PARAMETRIC TRANSFORMATIONS TO STABLE PARAMETERS: APPROACH 1

1. Let $\mathbf{x}_s$, of size n_s, be the support points of the K–optimal design. Find $\mathcal{T}$, from a set of algebraic (symbolic) equations where the vector of stable parameters represent the response surface at the support points, i.e.

$$\vartheta_i = f(x_i, \boldsymbol{\theta}), \quad i \in [n_s]. \tag{4}$$

2. A rearrangement of (4) allows expressing $\boldsymbol{\theta}$ as a function of $\boldsymbol{\vartheta}$

$$\boldsymbol{\theta} = \mathbf{g}(\mathbf{x}_s, \boldsymbol{\vartheta}) \tag{5}$$

   where $\mathbf{g}(\bullet)$ is a vector of functions forming the parametric transformation $\mathcal{T}$

3. In some cases, the analytic expression of (5) can be obtained through symbolic computation tools, such as `Mathematica`®.
4. If the $f(x, \boldsymbol{\theta})$ do not involve trigonometric or transcendental terms the solution can be exactly expressed in closed form (MM). Otherwise, an implicit definition or the explicit form of some approximation need to be considered.

PARAMETRIC TRANSFORMATIONS TO STABLE PARAMETERS: APPROACH 2

1. An alternative approach approximates $f(x, \boldsymbol{\theta})$ by a first-order Taylor expansion with respect to $\boldsymbol{\theta}$. Then (4) becomes

$$\boldsymbol{\vartheta} \simeq \mathbf{f}_0 + J(\mathbf{x}, \mathbf{p}_0)\,(\boldsymbol{\theta} - \mathbf{p}_0) \tag{6}$$

   where the vector $\mathbf{f}_0 = \{f(x_i, \mathbf{p}_0\}$ contains the model predictions using the original vector of parameters, $\mathbf{p}_0$. Let $J(\mathbf{x}, \boldsymbol{\theta})$ be the $n_s \times n_\theta$ Jacobian matrix

$$J(\mathbf{x}, \mathbf{p}_0) = \left\{ \left(\frac{\partial f(x_i, \boldsymbol{\theta})}{\partial \theta_j} \right) \bigg|_{\boldsymbol{\theta}=\mathbf{p}_0}, \quad j \in [n_\theta] \right\} i \in n_s.$$

   formed by $1 \times n_\theta$ vectors containing the derivatives of the model with respect to to the parameters at support point i:

2. The solution for $\boldsymbol{\theta}$ is therefore

$$\boldsymbol{\theta} = \mathbf{p}_0 + [J(\mathbf{x}, \mathbf{p}_0)]^{-1}\,(\boldsymbol{\vartheta} - \mathbf{f}_0)$$

   Consequently, $\mathbf{g}(\mathbf{x}_s, \boldsymbol{\vartheta}) = \mathbf{p}_0 + [J(\mathbf{x}, \mathbf{p}_0)]^{-1}\,(\boldsymbol{\vartheta} - \mathbf{f}_0)$.

PERFORMANCE INDICATORS 1

1. Purpose: to compare the effect of model reparameterization on model fitting using various design criteria.
2. 1 The maximum of the absolute cross-correlation among the parameter estimates (ϱ). We first construct an approximation to the correlation matrix using the covariance matrix $C(\hat{\boldsymbol{\theta}})$, given by the linearised model from the LS algorithm at convergence with estimates $\hat{\boldsymbol{\theta}}$. Let $B(\hat{\boldsymbol{\theta}})$ be a diagonal (square) matrix of size n_θ containing the square roots of the diagonal elements of $C(\hat{\boldsymbol{\theta}})$, i.e. $B_{i,i} = \sqrt{C_{i,i}},\ i \in [n_\theta]$ and $B_{i,j} = 0,\ i,j \in [n_\theta],\ i \neq j$. The correlation matrix of the parameter estimates is then given by

$$R(\hat{\boldsymbol{\theta}}) = B^{-1}(\hat{\boldsymbol{\theta}})\ C(\hat{\boldsymbol{\theta}})\ B^{-1}(\hat{\boldsymbol{\theta}}).$$

   and

$$\varrho = \max_{\substack{i,j \in [n_\theta] \\ j \neq i}} |R_{i,j}|.$$

3. Both for θ and ϑ.

PERFORMANCE INDICATORS 2

1. **2** The condition number of the sensitivity matrix (represented by κ).
2. $\kappa = \lambda_{\max}[C(\hat{\boldsymbol{\theta}})]/\lambda_{\min}[C(\hat{\boldsymbol{\theta}})]$, the ratio of the maximum and minimum eigenvalues of the covariance matrix. A lower condition number is an indication of reduced parametric collinearity. Belsley *et al.* (2005) suggest the arbitrary numbers of 0.90 as the cut-off value for the maximum absolute parametric correlation, and 20 as the threshold for the condition number.
3. In addition use the graphical representation of the 95% confidence ellipsoids for pairs of parameters using elliptical CI. More eccentric ellipsoids in general denote higher parameter collinearity. When the eccentricity is 1.0 the confidence region is circular and the parameters are uncorrelated. Contrarily, when the axes of the ellipses are not the coordinate axes, as the eccentricity tends to $+\infty$ the confidence region tends to a line, and the parameters are highly correlated.

EXAMPLE 1: EXPONENTIAL MODEL 1

$$\mathbb{E}(y) = \theta_1 + \theta_2 \exp(-\theta_3 \, x) \tag{7}$$

with design region $\mathcal{X} = [0, 10]$, discretized with $\Delta x = 0.1$ and $\mathbf{p}_0 \equiv (1.0,\ 1.0,\ 0.1)^\mathsf{T}$.

Table: Exponential model (7): optimal designs, $\mathbf{X} = [0, 10]$, $\Delta x = 0.1$, and $\mathbf{p}_0 = (1.0,\ 1.0,\ 0.1)^\mathsf{T}$.

Optimality Criterion	Design	Optimum	$\varrho^\dagger$	$\kappa^\dagger$	$\varrho^\ddagger$	$\kappa^\ddagger$
D–	$\begin{pmatrix} 0.0000 & 4.1999 & 10.0000 \\ 0.3334 & 0.3333 & 0.3333 \end{pmatrix}$	0.2051	0.9952	4.4778E+3	0.6192	9.9220
A–	$\begin{pmatrix} 0.0000 & 4.200 & 10.0000 \\ 0.1567 & 0.4826 & 0.3608 \end{pmatrix}$	506.064	0.9896	4.8957E+3	0.7835	20.4597
E–	$\begin{pmatrix} 0.0000 & 4.2000 & 10.0000 \\ 0.1042 & 0.6520 & 0.2438 \end{pmatrix}$	0.0019	0.9896	5.0800E+3	0.8065	20.7862
K–	$\begin{pmatrix} 0.0000 & 2.5000 & 10.0000 \\ 0.5458 & 0.3399 & 0.1143 \end{pmatrix}$	3408.26	0.9996	2.8100E+3	0.4071	3.3525

† — based on original model; ‡ — based on parameterized model.

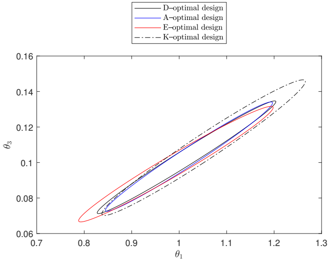
EXAMPLE 1: EXPONENTIAL MODEL 2

1. ρ: maximum correlation ≤ 0.9
2. κ: condition number < 20.

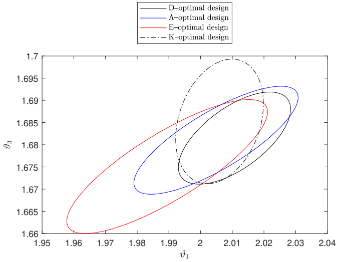
Optimality Criterion	Design	Optimum	$\varrho^\dagger$	$\kappa^\dagger$	$\varrho^\ddagger$	$\kappa^\ddagger$
D–	$\begin{pmatrix} 0.0000 & 4.1999 & 10.0000 \\ 0.3334 & 0.3333 & 0.3333 \end{pmatrix}$	0.2051	0.9952	4.4778E+3	0.6192	9.9220
A–	$\begin{pmatrix} 0.0000 & 4.200 & 10.0000 \\ 0.1567 & 0.4826 & 0.3608 \end{pmatrix}$	506.064	0.9896	4.8957E+3	0.7835	20.4597
E–	$\begin{pmatrix} 0.0000 & 4.2000 & 10.0000 \\ 0.1042 & 0.6520 & 0.2438 \end{pmatrix}$	0.0019	0.9896	5.0800E+3	0.8065	20.7862
K–	$\begin{pmatrix} 0.0000 & 2.5000 & 10.0000 \\ 0.5458 & 0.3399 & 0.1143 \end{pmatrix}$	3408.26	0.9996	2.8100E+3	0.4071	3.3525

† — based on original model; ‡ — based on reparameterized model.

EXAMPLE 1: EXPONENTIAL MODEL 3



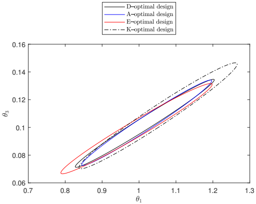
(a)



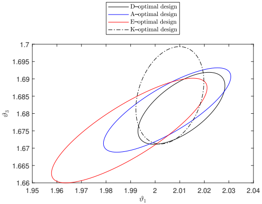
(b)

Figure: Exponential model (7): 95% confidence ellipses for parameters obtained with the optimal designs in Table 1 for $\hat{\theta}_3$ vs. $\hat{\theta}_1$: (a) considering the original parameters; (b) after reparameterization.

EXAMPLE 1: EXPONENTIAL MODEL 4



(a)



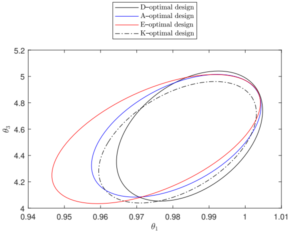
(b)

The ellipses in (b) are calculated by simulation from the optimum designs for the various criteria. Data from all simulations are then fitted by the model with stable parameters. K-optimality here is not fully orthogonal due to approximations in solving $\boldsymbol{\theta} = \mathbf{g}(\mathbf{x}_s, \boldsymbol{\vartheta})$.

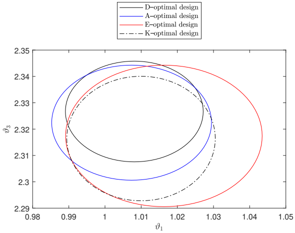
EXAMPLE 2: GENERALIZED MICHAELIS-MENTEN MODEL 1

$$\mathbb{E}(y) = \theta_1 + \frac{\theta_2\, x}{\theta_3 + x}, \tag{8}$$

is a rational polynomial function and admits exact solutions by applying purely algebraic manipulation techniques.



(c)



(d)

Michaelis-Menten model (8): 95% confidence ellipses for parameters obtained with optimal designs for $\hat{\theta}_3$ vs. $\hat{\theta}_1$: (a) considering the original parameters; (b) after reparameterization.

EXAMPLE 2: GENERALIZED MICHAELIS-MENTEN MODEL 2

The stable transformation vector that gives exactly zero correlation in the preceding figure is found from the support points of the K-optimum design. Since the model is a rational polynomial function, these can be found to be

$$\boldsymbol{\theta} = \left(\vartheta_1, \quad \frac{77.0\left(\vartheta_1 \vartheta_2 + \vartheta_1 \vartheta_3 - \vartheta_2 \vartheta_3 - {\vartheta_1}^2\right)}{77.0\,\vartheta_1 - 100.0\,\vartheta_2 + 23.0\,\vartheta_3}, \quad \frac{230.0\,\vartheta_2 - 230.0\,\vartheta_3}{77.0\,\vartheta_1 - 100.0\,\vartheta_2 + 23.0\,\vartheta_3} \right)^{\mathsf{T}}. \qquad (9)$$

MORE POINTS

1. One purpose of near orthogonality is that nearly independent inferences can be made about the parameters of the model. But here the ϑ do not have any clear physical meaning
2. Point is to obtain good parameter estimates. Then models can be compared in terms of Residual SS. Nested Models or BIC and relatives for non-nested models. Ross's book is called 'Nonlinear Estimation'
3. Examples from the laboratory. Look at SS contours and compare with first-order approximations.
4. Are compound designs important? Difference between D-and K-optimum designs in exponential model example.
5. I am grateful to Dr B. Bogacka for her help in the preparation of this talk.
6. Details, including computing, in Duarte *et al.* (2023).

REFERENCES

Belsley, D. A., Kuh, E., and Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.

Duarte, B., Atkinson, A. C., and Oliveira, N. M. C. (2023). Optimum design for ill-conditioned models: K-optimality and stable parameterizations. *Chemometrics and Intelligent Laboratory Systems*, **239**. doi:10.1016/j.chemolab.2023.104874.

Ross, G. J. S. (1990). *Nonlinear Estimation*. Springer Verlag, New York, NY. Springer Series in Statistics.

Seber, G. A. F. and Wild, C. J. (2003). *Nonlinear Regression*. John Wiley & Sons, New York.

Ye, J. J. and Zhou, J. (2013). Minimizing the condition number to construct design points for polynomial regression models. *SIAM Journal on Optimization*, **23**(1), 666–686.

Belsley, D. A., Kuh, E., and Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.

Duarte, B., Atkinson, A. C., and Oliveira, N. M. C. (2023). Optimum design for ill-conditioned models: K-optimality and stable parameterizations. *Chemometrics and Intelligent Laboratory Systems*, **239**. doi:10.1016/j.chemolab.2023.104874.

Ross, G. J. S. (1990). *Nonlinear Estimation*. Springer Verlag, New York, NY. Springer Series in Statistics.

Seber, G. A. F. and Wild, C. J. (2003). *Nonlinear Regression*. John Wiley & Sons, New York.

Ye, J. J. and Zhou, J. (2013). Minimizing the condition number to construct design points for polynomial regression models. *SIAM Journal on Optimization*, **23**(1), 666–686.