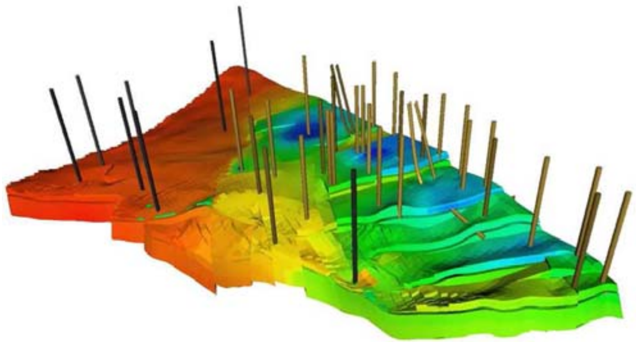
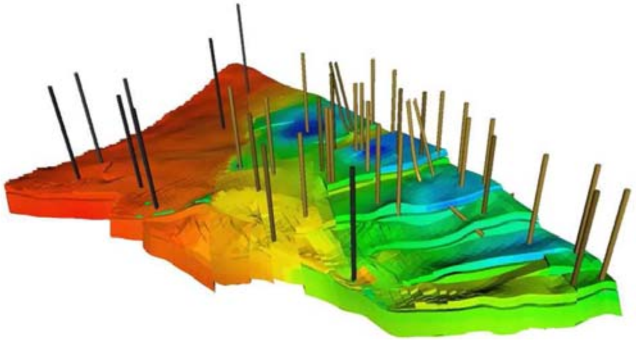


Identifiability: from qualitative analysis to model structure approximation

Jorn F.M. Van Doren, Sippe G. Douma, Paul M.J. Van den Hof,
Jan Dirk Jansen, Okko H. Bosgra

Delft University of Technology – Systems and Control,Geotechnology
Shell International Exploration and Production

Motivation - Identifiability



- Estimation of parameters in physical/first principles model structure

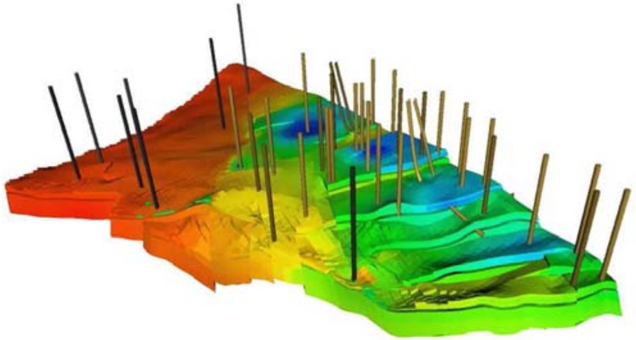
$$\hat{\mathbf{y}} = \mathbf{h}(\boldsymbol{\theta}, \mathbf{u}; \mathbf{x}_0)$$

- Number of parameters can be very large
  (e.g. after spatially discretizing *pde*)
  and too large for reliable estimation

- Physical structure (e.g. nonlinear dynamics) is considered important in view of reliable long-term model predictions

- How to reduce the model structure in terms of its ***parameter space?***

Motivation - Identifiability



- In case of linear dynamics, black model structures can be used as intermediates to extract all information from the data
- Notion of *identifiability* is instrumental in analyzing model structure properties
- Identifiability mostly considered in a yes/no setting: qualitative rather than quantitative [Bellman and Åström (1970), Grewal and Glover (1976)]
- Approach: *quantitative* analysis of appropriate parameter space, maintaining physical parameter interpretation

Contents

- Identifiability
- Testing local identifiability
- Model structure approximation
- Effect of parameter scaling
- Example
- Discussion

Identifiability

- Consider nonlinear model structure $\hat{\mathbf{y}} = \mathbf{h}(\boldsymbol{\theta}, \mathbf{u}; \mathbf{x}_0)$

- Locally identifiable in $\boldsymbol{\theta}_m$ for given $\mathbf{u}$ and $\mathbf{x}_0$ if in neighbourhood of $\boldsymbol{\theta}_m$:

$$\{\mathbf{h}(\mathbf{u}, \boldsymbol{\theta}_1; \mathbf{x}_0) = \mathbf{h}(\mathbf{u}, \boldsymbol{\theta}_2; \mathbf{x}_0)\} \Rightarrow \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$$

  [Grewal and Glover 1976]

- Structural identifiability in similar way on the basis of transfer functions (rather than output), for the linear dynamics case.
  [Bellman and Åström (1970)]

Testing local identifiability in identification

- In PE framework, identification criterion

$$V(\boldsymbol{\theta}) := \frac{1}{2}\boldsymbol{\epsilon}(\boldsymbol{\theta})^T \mathbf{P}_v^{-1}\boldsymbol{\epsilon}(\boldsymbol{\theta}), \quad \boldsymbol{\epsilon}(\boldsymbol{\theta}) = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{h}(\boldsymbol{\theta}, \mathbf{u}; \mathbf{x}_0),$$

- Hessian given by

$$\frac{\partial^2 V(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} = \frac{\partial \hat{\mathbf{y}}^T}{\partial \boldsymbol{\theta}} \mathbf{P}_v^{-1} \left(\frac{\partial \hat{\mathbf{y}}^T}{\partial \boldsymbol{\theta}}\right)^T + \mathbf{S}$$

- Local identifiability test in $\hat{\boldsymbol{\theta}} = \arg\min V(\boldsymbol{\theta})$: Hessian > 0

- With quadratic approximation of cost function around $\hat{\boldsymbol{\theta}}$: Hessian given by $\frac{\partial \hat{\mathbf{y}}^T}{\partial \boldsymbol{\theta}} \mathbf{P}_v^{-1} \left(\frac{\partial \hat{\mathbf{y}}^T}{\partial \boldsymbol{\theta}}\right)^T$

Testing local identifiability in identification

- Rank test on Hessian through SVD

$$\left.\frac{\partial \hat{\mathbf{y}}^T}{\partial \boldsymbol{\theta}} \mathbf{P}_v^{-\frac{1}{2}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix}$$

- If $\boldsymbol{\Sigma}_2 = 0$ then lack of local identifiability

- SVD can be used to reparameterize the model structure through

$$\boldsymbol{\theta} = \mathbf{U}_1 \boldsymbol{\rho}, \quad \dim(\boldsymbol{\rho}) << \dim(\underline{\theta})$$

  in order to achieve local identifiability in $\boldsymbol{\rho}$

- Columns of $\mathbf{U}_1$ are basis functions of the identifiable parameter space

Testing local identifiability in identification

$$\frac{\partial \hat{\mathbf{y}}^T}{\partial \boldsymbol{\theta}} \mathbf{P}_v^{-\frac{1}{2}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix}$$

- What if $\boldsymbol{\Sigma}_2 \neq 0$ but contains (many) small singular values ?

No lack of identifiability, but possibly very poor variance properties

Approximating the identifiable parameter space

Asymptotic variance analysis: $\operatorname{cov}(\hat{\boldsymbol{\theta}}) = J^{-1} = \left(\mathbb{E}\left[\frac{\partial^2 V(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\bigg|_{\hat{\boldsymbol{\theta}}}\right]\right)^{-1}$

with J = Fisher Information Matrix.

- Sample estimate of parameter variance:

$$cov(\hat{\boldsymbol{\theta}}) = \begin{cases} \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1^{-2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2^{-2} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} & \text{for } \boldsymbol{\Sigma}_2 > 0 \\ \infty & \text{for } \boldsymbol{\Sigma}_2 = 0 \end{cases}$$

$$cov(\boldsymbol{U_1}\hat{\boldsymbol{\rho}}) = \mathbf{U_1}\boldsymbol{\Sigma}_1^{-2}\mathbf{U_1}^T$$

$$cov(\hat{\boldsymbol{\theta}}) > cov(\boldsymbol{U_1}\hat{\boldsymbol{\rho}}) \qquad \text{if } \boldsymbol{\Sigma}_2 > 0$$

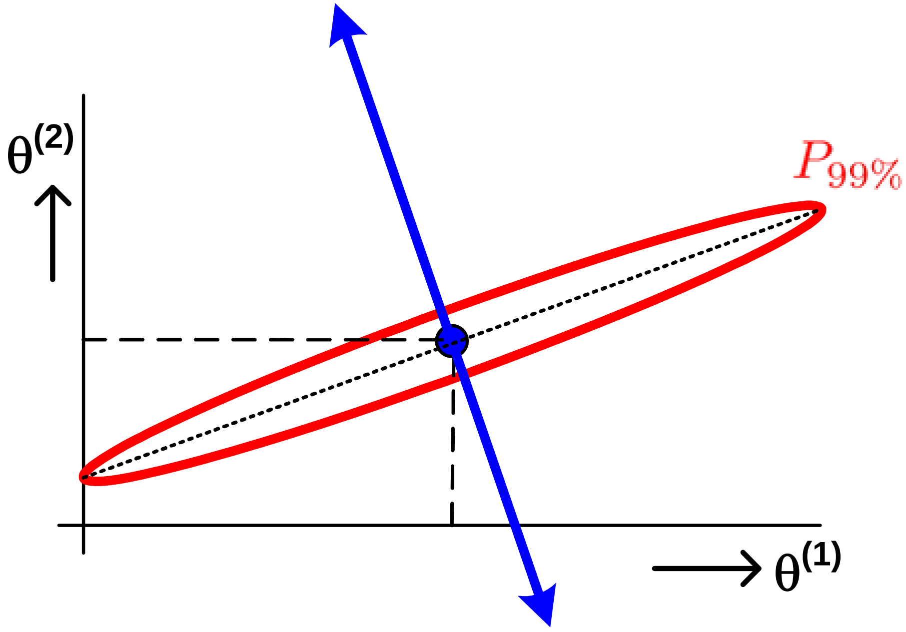
Approximating the identifiable parameter space

$$cov(\hat{\boldsymbol{\theta}}) > cov(\boldsymbol{U}_1\hat{\boldsymbol{\rho}}) \qquad \text{if } \boldsymbol{\Sigma}_2 > 0$$

- Discarding singular values that are small reduces the variance of the resulting parameter estimate
- Particularly important in situations of (very) large numbers of small s.v.'s
- Model structure approximation (local)
- Quantified notion of identifiability – related to parameter variance

Approximating the identifiable parameter space

- Interpretation:
  Remove the parameter directions that are poorly identifiable (have large variance)



- This is different from removing the (separate) parameters for which the value 0 lies within the confidence bound [Hjalmarsson, 2005]

Effect of parameter scaling/units

- In physical model structures there is a freedom of choice in parameter units (cm,km .....)

- Choice of units should not influence the choice of approximate model structure

- The yes/no question on identiability:

$$\mathbf{\Sigma_2} = 0, \quad \mathbf{\Sigma_2} \neq 0$$

  is not influenced by a scaling of parameter values:

$$\hat{\boldsymbol{\theta}} = \mathbf{\Gamma}\hat{\boldsymbol{\theta}}_1, \quad \mathbf{\Gamma} = \mathrm{diag}(s_1, \cdots s_n)$$

- However for $\mathbf{\Sigma_2} \neq 0$, scaling will influence the numerical values of $\mathbf{\Sigma_1}, \mathbf{\Sigma_2}$ and therefore also the choice of the identifiable parameter space

Effect of parameter scaling/units

- Possible remedy: use relative parameter variance rather than absolute variance as a measure for model structure approximation

$$\mathrm{cov}(\mathbf{\Gamma}_{\hat{\theta}}^{-1}\hat{\boldsymbol{\theta}}), \quad \text{e.g. } \mathbf{\Gamma}_{\hat{\theta}} = \mathrm{diag}\left(\begin{array}{ccc} |\hat{\boldsymbol{\theta}}_1| & \ldots & |\hat{\boldsymbol{\theta}}_q| \end{array} \right)$$

- Motivates analysis of scaled Hessian $\mathbf{\Gamma}_{\hat{\theta}} \left. \frac{\partial^2 V(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right|_{\hat{\boldsymbol{\theta}}} \mathbf{\Gamma}_{\hat{\theta}}$
- Essential information in

$$\mathbf{\Gamma}_{\hat{\theta}} \frac{\partial \mathbf{h}(\boldsymbol{\theta})^T}{\partial \boldsymbol{\theta}} \mathbf{P}_v^{-\frac{1}{2}} = \left[\begin{array}{cc} \mathbf{U}_1 & \mathbf{U}_2 \end{array} \right] \left[\begin{array}{cc} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma_2} \end{array} \right] \left[\begin{array}{c} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{array} \right]$$

- Model structure approximated: parameters identifiable and physically interpretable!

Toy example

Second order system: $y(t) = \alpha_0 u(t-1) + \beta_0 u(t-2)$

$$\alpha_0 = 10^6; \quad \beta_0 = 10^{-6}$$

Second order FIR model:

$$\hat{y}(t,\theta) = \alpha u(t-1) + \beta u(t-2), \quad \theta := [\alpha \;\; \beta]^T.$$

$$\psi(t,\theta_0) := \frac{\partial \hat{y}(t,\theta)}{\partial \theta} = \begin{bmatrix} u(t-1) \\ u(t-2) \end{bmatrix}$$

Fisher information matrix: $J = N \begin{bmatrix} R_u(0) & R_u(1) \\ R_u(1) & R_u(0) \end{bmatrix} = N \cdot I$

unit variance white noise input

→ No indications for reducing model structure

Toy example

After parameter scaling:

Scaled Fisher information matrix:

$$\begin{aligned} \tilde{J} &= N \begin{bmatrix} \alpha_0 & 0 \\ 0 & \beta_0 \end{bmatrix} \begin{bmatrix} R_u(0) & R_u(1) \\ R_u(1) & R_u(0) \end{bmatrix} \begin{bmatrix} \alpha_0 & 0 \\ 0 & \beta_0 \end{bmatrix} \\ &= N \begin{bmatrix} 10^{12} & 0 \\ 0 & 10^{-12} \end{bmatrix} \end{aligned}$$

➡ Indication for reducing model structure by one parameter

Discussion

- Notion of identifiability quantified
- Model structure approximated to achieve identifiability, while retaining interpretation of physical parameters
- Analysis can only be done locally linearized
- Similar results for structural identifiability
- Established relation with
  - Controllability and observability properties
  - Gradient/Hessian approximation in iterative optimization algorithms (Gauss-Newton and steepest descent)

THE VALUE OF SMARTNESS

ISAPP