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
Motivation - Identifiability

• Estimation of parameters in physical/first principles model structure

\[ \hat{y} = h(\theta, u; x_0) \]

• Number of parameters can be very large (e.g. after spatially discretizing \( \text{pde} \)) 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 \textit{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 \( \hat{y} = h(\theta, u; x_0) \)

• **Locally identifiable** in \( \theta_m \) for given \( u \) and \( x_0 \) if in neighbourhood of \( \theta_m \):
  \[
  \{ h(u, \theta_1; x_0) = h(u, \theta_2; x_0) \} \Rightarrow \theta_1 = \theta_2
  \]

  [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

\[ V(\theta) := \frac{1}{2} \epsilon(\theta)^T P_v^{-1} \epsilon(\theta), \quad \epsilon(\theta) = y - \hat{y} = y - h(\theta, u; x_0), \]

- Hessian given by

\[ \frac{\partial^2 V(\theta)}{\partial \theta^2} = \frac{\partial \hat{y}^T}{\partial \theta} P_v^{-1} \left( \frac{\partial \hat{y}^T}{\partial \theta} \right)^T + S \]

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

- With quadratic approximation of cost function around \( \hat{\theta} \): Hessian given by

\[ \frac{\partial \hat{y}^T}{\partial \theta} P_v^{-1} \left( \frac{\partial \hat{y}^T}{\partial \theta} \right)^T \]
Testing local identifiability in identification

- Rank test on Hessian through SVD

\[
\frac{\partial \hat{y}^T}{\partial \theta} P_v^{-\frac{1}{2}} \bigg|_{\theta = \hat{\theta}} = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^T \\ V_2^T \end{bmatrix}
\]

- If \( \Sigma_2 = 0 \) then lack of local identifiability

- SVD can be used to reparameterize the model structure through

\[
\theta = U_1 \rho, \quad \text{dim}(\rho) << \text{dim}(\theta)
\]

in order to achieve local identifiability in \( \rho \)

- Columns of \( U_1 \) are basis functions of the identifiable parameter space
Testing local identifiability in identification

\[ \frac{\partial y^T}{\partial \theta} P_v^{-\frac{1}{2}} \bigg|_{\theta=\hat{\theta}} = \left[ \begin{array}{c} U_1 \\ U_2 \end{array} \right] \left[ \begin{array}{cc} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{array} \right] \left[ \begin{array}{c} V_1^T \\ V_2^T \end{array} \right] \]

• What if $\Sigma_2 \neq 0$ but contains (many) small singular values?

No lack of identifiability, but possibly very poor variance properties.
Approximating the identifiable parameter space

Asymptotic variance analysis:  
\[
\text{cov}(\hat{\theta}) = J^{-1} = \left( \mathbb{E} \left[ \frac{\partial^2 V(\theta)}{\partial \theta^2} \bigg| \hat{\theta} \right] \right)^{-1}
\]
with \( J = \text{Fisher Information Matrix} \).

- Sample estimate of parameter variance:

\[
cov(\hat{\theta}) = \left\{ \begin{array}{ll}
\left[ \begin{array}{cc}
U_1 & U_2 \\
0 & \Sigma_2^{-2}
\end{array} \right] & \left[ \begin{array}{c}
\Sigma_1^{-2} \\
V_1^T
\end{array} \right] & \text{for } \Sigma_2 > 0 \\
\infty & V_2^T & \text{for } \Sigma_2 = 0
\end{array} \right.
\]

\[
cov(U_1 \hat{\rho}) = U_1 \Sigma_1^{-2} U_1^T
\]

\[\text{cov}(\hat{\theta}) > \text{cov}(U_1 \hat{\rho}) \quad \text{if } \Sigma_2 > 0\]
Approximating the identifiable parameter space

\[ \text{cov}(\hat{\theta}) > \text{cov}(U_1\hat{\rho}) \quad \text{if} \quad \Sigma_2 > 0 \]

• 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:

\[ \Sigma_2 = 0, \quad \Sigma_2 \neq 0 \]

is not influenced by a scaling of parameter values:

\[ \hat{\theta} = \Gamma \hat{\theta}_1, \quad \Gamma = \text{diag}(s_1, \ldots s_n) \]

• However for \( \Sigma_2 \neq 0 \), scaling will influence the numerical values of \( \Sigma_1, \Sigma_2 \) 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

\[ \text{cov}(\Gamma^{-1}_\hat{\theta} \hat{\theta}), \quad \text{e.g. } \Gamma_\hat{\theta} = \text{diag}(|\hat{\theta}_1| \ldots |\hat{\theta}_q|) \]

• Motivates analysis of scaled Hessian

\[ \Gamma_\hat{\theta} \frac{\partial^2 V(\theta)}{\partial \theta^2} \bigg|_{\hat{\theta}} \]

• Essential information in

\[ \Gamma_\hat{\theta} \frac{\partial h(\theta)^T}{\partial \theta} P_v^{-\frac{1}{2}} = \left[ \begin{array}{cc} U_1 & U_2 \end{array} \right] \left[ \begin{array}{cc} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{array} \right] \left[ \begin{array}{c} V_1^T \\ V_2^T \end{array} \right] \]

• 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:
\[ \hat{y}(t, \theta) = \alpha u(t - 1) + \beta u(t - 2), \quad \theta := [\alpha \quad \beta]^T. \]
\[ \psi(t, \theta_0) := \frac{\partial \hat{y}(t, \theta)}{\partial \theta} = \begin{bmatrix} u(t - 1) \\ u(t - 2) \end{bmatrix} \]

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:

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

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)