Perspectives on System Identification

Lennart Ljung
Linköping University, Sweden

The Problem
Flight tests with Gripen at high alpha
Person in Magnet camera, stabilizing a pendulum by thinking "right"-"left"

The Confusion

This Talk
Two objectives:
• Place System Identification on the global map. Who are our neighbours in this part of the universe?
• Discuss some open areas in System Identification.

The core

\[ \text{Model } \theta \rightarrow \text{Data } X \rightarrow \text{Complexity (Flexibility) } C \]

\[ \text{Information } I \rightarrow \text{Error } E \]

\[ \text{Estimation } \hat{\theta} = \nabla \log \mathcal{L}(\theta | X) \text{ (Learning = Generation)} \]

\[ \text{Model fit } \mathcal{F}[\theta, X] \]
Estimation

Squeeze out the relevant information in data

But NOT MORE!

All data contain information and misinformation ("Signal and noise")

So need to meet the data with a prejudice!

Estimation Prejudices

- Nature is Simple!
  - Occam's razor
  - God is subtle, but He is not malicious (Einstein)
  - So, conceptually:
    \[ \hat{m} = \text{arg min} \{ \text{Fit} + \text{Complexity Penalty} \} \]
    \[ m \in M \]
    - Ex: Akaike:
      \[ \hat{m} = \text{arg min} \log \sum f(i, \theta) + 2 \dim \theta \]
    - Regularization:
      \[ \hat{m} = \text{arg min} \sum f(i, \delta) + \| \delta \| \]

Estimation and Validation

Fit to estimation data \( Z^N \) \( \sim \mathcal{N} \); Number of data points!
\[ F(\hat{m}, Z^N) \rightarrow \text{The empirical risk} \]

Now try your model on a fresh data set: Validation data \( Zv \)
\[ f(\hat{m}, Zv) \approx \hat{F}(m, Z^N) + f(C(M), N) \]
\( f \) is a function of the complexity, so the more flexible the model set the more the expected fit to validation data deteriorates. (Exact formulations: Akaike's FPE (AIC), Vapnik's learning/generalization results, Holdemarker average...)

So don't be impressed by a good fit to data in a flexible model set!

Bias and Variance

\( S \) = True system
\( \hat{m} \) = Estimate
\( m^* = E \hat{m} \)
\( m \in \mathcal{M} \); Typically \( m^* \) is the model closest to \( S \) in \( \mathcal{M} \).
\[ E[|S - \hat{m}|^2 - |S - m^*|^2 + E[|\hat{m} - m^*|^2] \]

MSE  =  BIAS (B) + VARIANCE (V)

Error  =  Systematic + Random

As \( C(M) \) increases, \( B \) decreases and \( V \) increases

This bias/variance tradeoff is at the heart of estimation!

Note that the \( C \) that minimizes the MSE typically has a \( K \) / \( \tilde{K} \)

Information Contents in Data and the CR Inequality

The value of information in data depends on prior knowledge. Let \( f \) be the probability density function be \( f_f(y | x) \). The (Fisher) Information Matrix is
\[ I \equiv E\left[ \frac{\partial^2 \log f_f(y | x, \theta)}{\partial \theta \partial \theta'} \right] \]

The Cramér-Rao inequality tells us that
\[ \text{var} \theta > I^{-1} \]
for any (unbiased) estimator \( \theta \) of the parameter.

\( I \) is thus a prime quantity for Experiment Design.

The Communities Around the Core I

- Statistics: The mother area
  - EM algorithm for ML estimation
  - Resampling techniques (bootstrap...)
  - Regularization: LARS, Lasso...
- Statistical learning theory
  - Convex formulations, SVM (support vector machines)
  - VC-dimensions
- Machine learning
  - Grown out of artificial intelligence: Logical trees, Self-organizing maps.
  - More and more influence from statistics, Gaussian Proc., HMM, Bayesian nets
The Communities Around the Core I

- **Manifold learning**
  - Observed data belongs to a high-dimensional space
  - The action takes place on a lower dimensional manifold: Find that!
- **Chemometrics**
  - High-dimensional data spaces (Many process variables)
  - Find linear low dimensional subspaces that capture the essential state: PCA, PLS (Partial Least Squares),...
- **Econometrics**
  - Volatility Clustering
  - Common roots for variations

The Communities Around the Core II

- **Data mining**
  - Sort through large data bases looking for information: ANN, NN, Trees, SVD…
  - Google, Business, Finance…
- **Artificial neural networks**
  - Origin: Rosenblatt’s perceptron
  - Flexible parametrization of hyper-surfaces
- **Fitting ODE coefficients to data**
  - No statistical framework: Just link ODE/DAE solvers to optimizers
- **System Identification**
  - Experiment design
  - Dualities between time- and frequency domains

System Identification – Past and Present

- **Two basic avenues, both laid out in the 1960’s**
  - Statistical route: ML etc: Åström-Bohlin 1965
    - Prediction error framework: postulate predictor and apply curve-fitting
  - Realization based techniques: Ho-Kalman 1966
    - Construct/estimate states from data and apply LS (Subspace methods).

  **Past and Present:**
  - Useful model structures
  - Adapt and adopt core’s fundamentals
  - Experiment Design …
    - … with intended model use in mind (“identification for control”)

Model Reduction

- System Identification is really “System Approximation” and therefore closely related to Model Reduction.

  Model Reduction is a separate area with an extensive literature (“another satellite”), which can be more seriously linked to the system identification field.

  - **Linear systems - linear models**
    - Divide, conquer and reunite (outputs)?
  - **Non-linear systems – linear models**
    - Understand the linear approximation: Is it good for control?
  - **Nonlinear systems -- nonlinear reduced models**
    - Much work remains

System Identification - Future: Open Areas

- **Spend more time with our neighbours!**
  - Report from a visit later on
- **Model reduction and system identification**
- **Issues in identification of nonlinear systems**
- **Meet demands from industry**
- **Convexification**
  - Formulate the estimation task as a convex optimization problem

Linear Systems - Linear Models

- **Divide – Conquer – Reunite!**

  Helicopter data: 1 pulse input, 8 outputs (only 3 shown here)
  State Space model of order 20 wanted.
  First fit all 8 outputs at the same time:
  Next fit 8 SISO models of order 12, one for each output:
Linear Systems - Linear Models
Divide – Conquer – Reunite!

Now, concatenate the 8 SISO models, reduce the 96th order model to order 20, and run some more iterations.

```matlab
mm = [m1;...;m8]; mr = balred(mm,20); model = pem(zd,mr); compare(zd,model)
```

Linear Models from Nonlinear Systems

System 
\( y(t) = w(t) + 0.01e^t u(t) \)
\( y \) non-Gaussian \( \| x(t) \| < 5 \) (Mastic Parry et al.)

Mode: 
\( y(t) = 5u(t) + u(t) \), \( u = e^{\sin(\omega t)} \)

Red: Amplitude Bode plot for estimated model
Blue: Model without \( 0.01e^t u(t) \)
Output discrepancy \( \leq 1\% \)

Model reduction from nonlinear to linear could be surprising!

Nonlinear Systems

- A user’s guide to nonlinear model structures suitable for identification and control
- Unstable nonlinear systems, stabilized by unknown regulator
- Stability handle on NL blackbox models

Industrial Demands

- Data mining in large historical process data bases ("K,M,G,T,P")
  - All process variables, sampled at 1 Hz for 100 years = 0.1 PByte
  - A serious integration of physical modeling and identification (not just parameter optimization in simulation software)

Industrial Demands: Simple Models

- Simple Models/Experiments for certain aspects of complex systems
- Use input that enhances the aspects, ...
- ... and also conceals irrelevant features
  - Steady state gain for arbitrary systems
  - Use constant input!
  - Nyquist curve at phase crossover
  - Use relay feedback experiments
  - But more can be done ...
  - ... Hjalmarsson et al: "Cost of Complexity".

An Example of a Specific Aspect

- Estimate a non-minimum-phase zero in complex systems (without estimating the whole system) – For control limitations.
- A NMP zero at \( a \) for an arbitrary system can be estimated by using the input

\[
\frac{c}{z^{-1} + a}
\]

Example: 100 complex systems, all with a zero at 2, are estimated as 2nd order

FIR models 
\( y(t) = a_0 u(t) + b_0 u(t-1) \)
**System Identification - Future: Open Areas**

- Spend more time with our neighbours!
- Report from a visit later on
- Model reduction and system identification
- Issues in identification of nonlinear systems
- Meet demands from industry
- **Convexification**
  - Formulate the estimation task as a convex optimization problem

---

**Example:**

Michaelis – Menten kinetics

\[
\dot{y} = \frac{\theta y}{\theta + y} - y + u
\]

\[y_0, (k) - g(k) + r(k)\]

Are Local Minima an Inherent feature of a model structure?

\[
\sum_{i=1}^{n} (y_i - \Phi_i(k) \theta)'^2
\]

Convexification I

**Convexification II**

Manifold Learning

1. \(X: \text{Original regressors}\)
2. \(g(x): \text{Nonlinear, nonparametric recoordination}\)
3. \(Z: \text{New regressor, possibly of lower dimension}\)
4. \(h(z): \text{Simple convex map}\)
5. \(Y: \text{Goal variable (output)}\)

---

**Narendra-Li’s Example**

Use LLE for the nonlinear recoordination of regressors and unfold simple maps. Compare with standard ANN (Hennia et al. 1996).

Simulate estimation and validation data \(x_t, y_t\) in the ANN structure and build a NN ARX model.

**Conclusions**

- System identification is a mature subject...
- Same age as IFAC, with the longest running symposium series
- ...and much progress has allowed important industrial applications...
- ...but it still has an exciting and bright future!