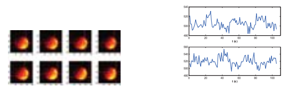
Approaches to Identification of *Nonlinear systems*

Lennart Ljung

Division of Automatic Control
Linköping University
Sweden

The Problem

Brain Activity (fMRI):



Pulp Buffer Vessel:

Forest Crane:

This Presentation ...

... aims at

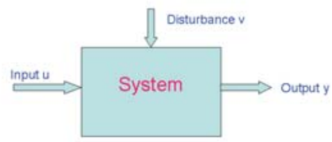
- an explanation of the essence of the problem of non-linear identification
- a color-coded overview of typical parametric approaches
- some information of on-going research on a new, non-parametric approach

Dynamic Systems

A dynamic system has an output response y that depends on (all) previous values of an input signal u. It is also typically affected by a disturbance signal v.
So the output at time t can be written as

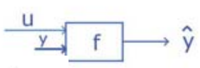
$$y(t) = g(u^t, v^t)$$

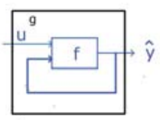
where superscript denotes the signal's values from the remote past up to the indicated time.
The input signal u is known (measured), while the disturbance v is unmeasured.



More Formalized Questions

Think discrete time data sequences:
$Z^t = [u(1), u(2), ..., u(t), y(1), y(2), ..., y(t)]$
We need to get hold of a "simulation function" $y(t) = g(u^t)$ and/or a prediction function $\hat{y}(t|t-1) = \tilde{f}(Z^{t-1})$



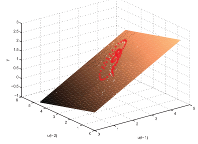


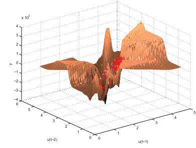
Note that $f \Rightarrow g$!

The Predictor

The predictor function $\hat{y}(t|t-1) = \tilde{f}(Z^{t-1}) = f(\varphi(Z^{t-1}))$ is what we try to estimate from data.

Two basic cases for f: (special case $\varphi(t) = [u(t-1), u(t-2)]$)

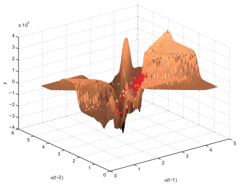




Linear: $\hat{y}(t) = a_1 u(t-1) + a_2 u(t-2)$ Nonlinear: $\hat{y}(t) = f(u(t-1), u(t-2))$

The observations Z^t are points in this space.

A Quick Classification of Non-Linear Models



How to describe the surface ?

- Parametric: $\hat{f}(\varphi) = f(\varphi, \hat{\theta})$ $\quad \hat{\theta} = \arg\min \sum \|y(t) - f(\varphi(t), \theta)\|^2$
- Nonparametric: Form the surface by smoothing the observations $y(t)$

Parametric Models: A Palette from White to Black:

White – Off-white – Smoke-grey – Steel-grey –Slate-grey – Black

White to Off-White: Physical Modeling

Perform physical modeling (e.g. in MODELICA) and denote (unknown) physical parameters by θ. Collect the model equations as

$$\dot{x}(t) = f(x(t), u(t), \theta)$$
$$y(t) = h(x(t), u(t), \theta)$$

(or in DAE, Differential Algebraic Equations, form.) For each parameter θ this defines a simulated (predicted) output $\hat{y}(t|\theta)$ which is the parameterized function

$$\hat{y}(t|\theta) = \tilde{f}(Z^{t-1}, \theta)$$

in somewhat implicit form. θ is then found by optimizing the fit to observations.
The approach is conceptually simple, but could be very demanding in practice.

The Palette

- White: Known model
- Off-white: Careful Physical Modeling
- Smoke-grey
- Steel-grey
- Slate-grey
- Black

Smoke-Grey: Semi-physical Models

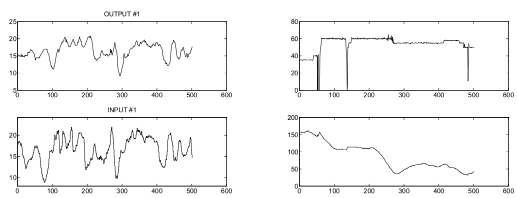
Apply non-linear transformations to the measured data, so that the transformed data stand a better chance to describe the system in a linear relationship.

"Rules: Only high-school physics and max 10 minutes"

Toy Example: Immersion heater: Input: voltage to the heater. Output: temperature of the fluid. . . ..
. . .. Square the voltage! Sense morale: No excuse for not thinking over the basic physical facts!

Another example: . . .

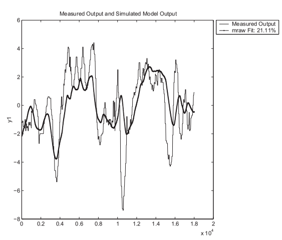
Buffer Vessel Dynamics



κ-number of outflow,
κ-number of inflow,

flow
volume

Model Based on Raw Data



Dashed line: κ-number after the vessel, actual measurements.
Solid line: Simulated κ-number using the input only and a process model estimated using the first 200 data points. $G(s) = \frac{0.818}{1+676s}e^{-480s}$

Now it's time to

Think: ....

No mixing ("Plug flow"): The vessel is then just a pure time delay for the pulp flow: Delay time: Vessel Volume/Pulp Flow (dimension time.)

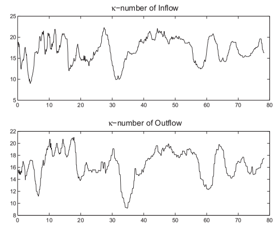
Perfect mixing in tank: A text-book first order system with gain=1 and time constant = Volume/Flow

So if Volume and Flow are changing, we have a time-varying system (or non-linear!)

The natural time variable is really Volume/Flow, (which we have measured). Let us re-sample the observed data according to this natural time variable.

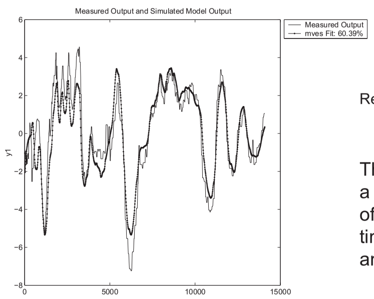
Re-sample Data

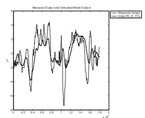

```
z = [y,u]; pf = flow./level;
t = 1:length(z)
newt = interp1([cumsum(pf),t],[pf(1):sum(pf)]');
newz = interp1([t,z], newt);
```



Semi-physical Model

$$G(s) = \frac{0.8116}{1 + 110.28s} e^{-369.58s}$$





Recall Linear model.

The semi-physical model gives a sufficiently good description of the buffer, to allow proper time-marking of the pulp before and after.

The Palette

- White: Known model
- Off-white: Careful Physical Modeling
- Smoke-grey. Semi-physical modeling (Could be used more!)
- Steel-grey
- Slate-grey
- Black

Steel-Grey: Composite Local Models

Non-linear systems are often handled by linearization around a working point. The idea behind Composite Local (Local Linear) Models is to deal with the nonlinearities by selecting or averaging over relevant linearized models.

Example: Tank with inflow u and free outflow y and level h: (Bernoulli's) equations:

$$\dot{h} = -\sqrt{h} + u$$
$$y = \sqrt{h}$$

Linearize around level h^* with corresponding flows $u^* = y^* = \sqrt{h^*}$:

$$\dot{h} = -\frac{1}{2\sqrt{h^*}}(h - h^*) + (u - u^*)$$
$$y = y^* + \frac{1}{2\sqrt{h^*}}(h - h^*)$$

Tank Example, ctd

Sampled data model around level h^* (Sampling time T_s):

$$y(t) = \gamma(h^*) + \alpha(h^*)y(t - T_s) + \beta(h^*)u(t - T_s) = \theta^T(h^*)\varphi(t)$$

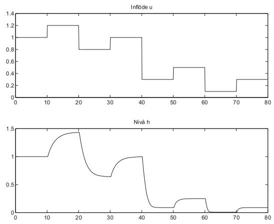
An ARX-model with level-dependent parameters. Now compute linearized model for d different levels, $h_1, h_2, \ldots, h_d$. Total model: select or average over these local models

$$\hat{y}(t) = \sum_{k=1}^{d} w_k(h, h_k)\theta^T(h_k)\varphi(t)$$

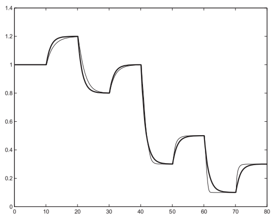
Choices of weights w_k : ....

Data and Linear Model

Measured data:



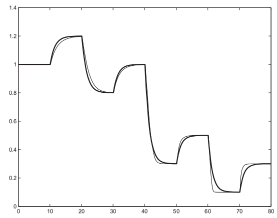
Linear Model ($d = 1$)



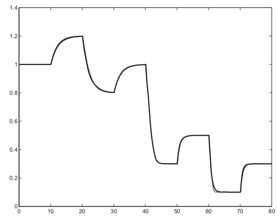
Thick line: Model. Thin: Measured.

Local Linear Models

Two levels (models) (d=2)



Five levels (models) ($d = 5$)



Composite Local Models: General Comments

Let the measured working point variable (tank level in example) be denoted by $\rho(t)$ (sometimes called regime variable). If the regime variable is partitioned into d values ρ_k, and model output according to value ρ_k is $\hat{y}^{(k)}(t)$ the predicted output will be

$$\hat{y}(t) = \sum_{k=1}^{d} w_k(\rho(t), \rho_k)\hat{y}^{(k)}(t)$$

If the prediction $\hat{y}^{(k)}(t)$ corresponding to ρ_k is linear in the parameters, $\hat{y}^{(k)}(t) = \varphi^T(t)\theta^{(k)}$, and the weights w are fixed, the whole model will be a linear regression.

Important connections to active research areas

- LPV (Linear Parameter-Varying) Models
- Hybrid Models ($\approx w(\cdot,\cdot)$ is estimated too.)

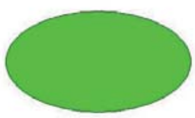
The Palette

- White: Known model
- Off-white: Physical Modeling
- Smoke-grey: Semi-physical modeling
- Steel-grey: Composite (local) models (Most common NL model in industry(?))
- Slate-grey
- Black

Slate Grey: Block-oriented Models

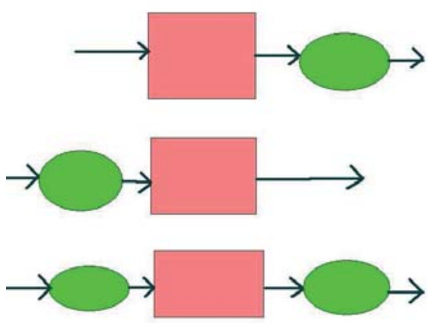
Building Blocks:

Linear Dynamic System
$G(s)$



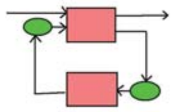
Nonlinear static function
$f(u)$

Common Models



Other Combinations 25

Active Research Field:



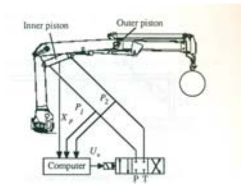
With the linear blocks parameterized as a linear dynamic system and the static blocks parameterized as a function ("curve"), this gives a parameterization of the output as

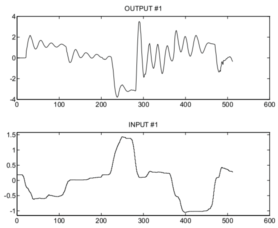
$$\hat{y}(t|\theta) = g(Z^{t-1}, \theta)$$

and the general approach of model fitting can be applied.
However, in this contexts many algorithmic variants have been suggested.

Example: Hydraulic Crane Data 26

These are data from a forest harvest machine:

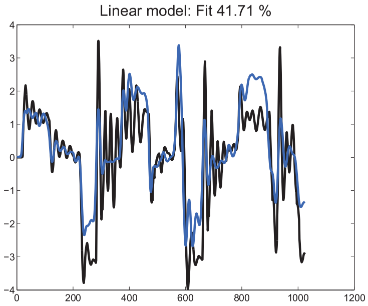




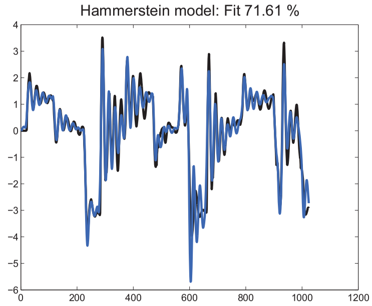
Input: Hydraulic Pressure.
Output: Tip Position

Linear Model 27

Black: Measured Output
Blue: Model Simulated Output



Hammerstein Model of the Hydraulic Crane 28



The Hammerstein Model gives a good fit. The extra flexibility offered by the input nonlinearity is quite useful, (even though no direct physical explanation is obvious.)

The Palette 29

- White: Known model
- Off-white: Physical Modeling
- Smoke-grey: Semi-physical modeling
- Steel-grey: Composite (local) models
- Slate-grey: Block-oriented models
- Black

Black: Basis-Function Expansion 30

$$\hat{y}(t|\theta) = \tilde{f}(Z^{t-1}, \theta) = f(\varphi(t), \theta)$$

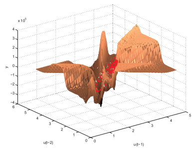
$\varphi(t) = \varphi(Z^{t-1})$ "state" of fixed dimension

$$f(\varphi, \theta) = \sum_{k=1}^{d} \alpha_k \kappa(\beta_k(\varphi - \gamma_k)), \quad \theta = \{\alpha_k, \beta_k, \gamma_k\} \quad \kappa: \text{ unit function}$$

Intuitive picture: Think of a scalar φ and let $\kappa(z)$ be a unit pulse for $0 \le z \le 1$. Then $\kappa(\beta(\varphi - \gamma))$ is a pulse of width $1/\beta$ starting in $\varphi = \gamma$. The sum above is then a piecewise constant function, capable of approximation "any" function arbitrary well for large enough d.

- The whole ANN (Artificial Neural Network), neuro-fuzzy, LS-SVM (Least Squares Support Vector Machines), etc business

Non-Parametric Methods



Form the surface by smoothing over the observation points in the space!

- Even Blacker!
- Huge literature – Mostly in the statistical community and now also in machine learning
- Only one aspect will be discussed here: Semi-supervised Regression

Semi-supervised Regression: The Problem

Given a standard regression problem:

$$y(t) = f(\varphi(t)) + \text{noise}$$

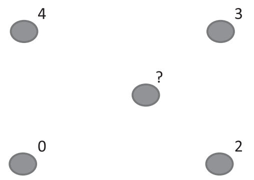
f unknown, $y(t)$ and $\varphi(t)$ observed for $t = 1, \dots, N_L$. Find f!

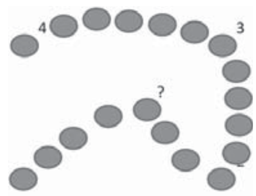
Or rather, for any given φ^* find a good value of $f(\varphi^*)$ ("Model on Demand", "Just in time model")

Extra feature: We have several measurements of $\varphi(t), t = N_L + 1, \dots, N_L + N_U$ without corresponding values of $y(t)$ (N_U "unlabeled observations")

Information in Unlabeled Regressors?

Yes, if the regressors live in a confined region, like a manifold:





(Un-/Semi-)Supervised

- Supervised: All Regressors labeled $(y(t), \varphi(t))$ [Standard Regression problems]
- Unsupervised: No labels known. [Clustering, Classification]
- Semi-supervised: Some labels known $\{y(t), \varphi(t), t = 1, \dots, N_L\}$. Some additional regressors known without labels $\{\varphi(t), t = N_L + 1, \dots, N_L + N_U\}$
- Estimation problem: still to "predict" $y^* = f(\varphi^*)$ for any given φ^*
  - "predict": .....

The Suggested Method: WDMR (With H. Ohlsson)

WDMR: Weight Determination by Manifold Regression
(Manifold) Smoothness Assumption: $f(\varphi_1)$ and $f(\varphi_2)$ close if φ_1 and φ_2 are close (in a relevant metric).

Problem: associate $\varphi(i)$ with good values $\hat{f}_t = f(\varphi(t))$ for all regressors, both labeled and unlabeled.
Take care of two kinds of information:

- $f(\varphi(t)) \approx y(t)$ for the measured labels
- $f(\varphi(t)) \approx f(\varphi(j))$ if $\varphi(t) \approx \varphi(j)$.

Formalize the second information using a *kernel* $K(\cdot, \cdot)$:

$$\hat{f}_t \approx \sum_{t=1}^{N_L+N_U} K_{t,j}\hat{f}_j; \qquad K_{t,j} = K(\varphi(t), \varphi(j))$$

WDMR, cont'd

Now, weigh together the two sources of information

$$\lambda \sum_{i=1}^{N_L+N_U} (\hat{f}_i - \sum_{j=1}^{N_L+N_U} K_{i,j}\hat{f}_j)^2 + (1-\lambda)\sum_{t=1}^{N_L}(y(t) - \hat{f}_t)^2 \qquad (*)$$

$$K_{i,j} = K(\varphi(i), \varphi(j))$$

Pick a ("regularization parameter") λ that balances the fit to measured labels and the smoothness prior. Minimize w.r.t. $\hat{f}_t, t = 1, \dots, N_L + N_U$.

That gives the estimated function value $\hat{f}(\varphi)$ for any regressor φ you include among the unlabeled regressors.

Note that $(*)$ is quadratic in $\hat{f}_t$, so the solution is easy to obtain.
Choice of kernel $K(\cdot, \cdot)$: Many possibilities ...

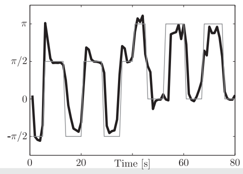
Example 1: fMRI Signals

The patient in the magnet camera is moving his eye focus left - right - up - down. 128 voxels in the visual cortex are monitored by fMRI, giving a vector $\varphi(t) \in R^{128}$ sampled every two seconds. The output $y(t)$ is the viewing angle $0, \pi, \pi/2, -\pi/2$.

fMRI Signals, cont'd

80 (labeled) samples were collected as estimation data. The model is a mapping from the φ-space R^{128} to the scalar space of viewing angle $[\pm\pi]$. For the mapping in this large space. 80 measurements may seem to be very few. But the brain activity in the 128-dimensional space is really triggered by a one-dimensional stimulus - the scalar eye movement. It can thus be argued that the regressors in R^{128} are really confined to a one-dimensional manifold. To test the method, also 40 validation data were collected. (i.e. $N_L = 80, N_U = 40$)

To the right we show the predicted y-values ($[-\pi/2 \quad \pi]$), (thick line) for these unlabeled validation measurements together with the corresponding true angles (thin line).



Conclusions

- The world of non-linear identification is rich and complex.
- Parametric methods may be color-coded in several shades of grey
- Non-parametric methods are gaining importance with inspiration from statistics and "machine learning". They certainly have relevance for system identification
- A "semi-supervised" method WDMR was suggested for non-linear, non-parametric regression, which shows promising results for several examples of different characters (but needs more understanding regarding the potential for system identification)