Dynamic Systems

A dynamic system has an output response that depends on both past and present input values. It is also typically affected by a disturbance signal \( u \). The input signal \( u \) and the disturbance signal \( v \) are known (measured), while the disturbance \( v \) is unmeasured.

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

Two basic cases for \( f \):
- Linear: \( \hat{y}(t + 1) = f(Z^t) = y(t) + k_1 u(t) + k_2 u(t-1) + k_3 u(t-2) \)
- Non-linear: \( \hat{y}(t + 1) = f(Z^t) = y(t) + k_1 u(t) + k_2 u(t-1) + k_3 u(t-2) - Z(t-1) \)

The observations, \( Z \), are points in this space.

Approaches to Identification of Non-linear Systems

- Pulp Buffer Vessel
- Brain Activity (fMRI)
- Forest Crane

This Presentation...

This discrete-time data sequence:

\[ y(t + 1) = f(Z^t) = y(t) + k_1 u(t) + k_2 u(t-1) + k_3 u(t-2) - Z(t-1) \]

Some formalization of ongoing research on a new, non-parametric approach.
A Quick Classification of Non-Linear Models

How to describe the surface
- Parametric: \( f(\varphi) = f(\varphi, \theta) \) \( \theta = \arg \min \| y(t) - f(\varphi(t), \theta) \|^2 \)
- Nonparametric: Form the surface by smoothing the observations \( y(t) \)

Parametric Models: A Palette from White to Black:
- White: Known model
- Off-white: Careful Physical Modeling
- Black: Transformer used to test the system in a linear relationship.

White to Off-White: Physical Modeling

Perform physical modeling (e.g., in Modelica) and denote \( \theta \) unknown physical parameters by \( \theta \). Collect the model equations as

\[
\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), u(t), \theta) \\
y(t) = \mathbf{h}(\mathbf{x}(t), u(t), \theta)
\]

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

\[
\hat{y}(t|\theta) = \mathbf{f}(\mathbf{z}^{-1}, \theta)
\]

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

White to Off-White: Physical Modeling

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

- Dashed line: \( \lambda \)-number after the vessel, actual measurements.
- Solid line: Simulated \( \lambda \)-number using the input only and a process model estimated using the first 200 data points. \( Q = \frac{\lambda}{1 + \exp(-10 \lambda)} \)}
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);

The Palette

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

Semi-physical Model

\[ G(s) = \frac{0.8116}{1 + 110.28s} e^{-30.59s} \]

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.

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:

\[ \begin{align*}
\dot{h} &= -\sqrt{h} + u \\
y &= \sqrt{h}
\end{align*} \]

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

\[ \begin{align*}
\dot{h} &= -\frac{1}{2\sqrt{h^*}} (h-h^*) + (u-u^*) \\
y &= y^* - \frac{1}{2\sqrt{h^*}} (h-h^*)
\end{align*} \]

Tank Example, ctd

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

\[ y(t) = \gamma(h^*) + \alpha(h^*)y(t-T_s) + \beta(h^*)u(t-T_s) = \theta^T(h^*)z(t) \]

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

\[ \dot{y}(t) = \sum_{k=1}^{n} w_k(h, h_k) \theta^T(h_k)z(t) \]

Choices of weights \( w_k : \ldots \)
Data and Linear Model

Measured data: Linear Model ($d=1$)

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 $\rho_k$, and model output according to value $\rho_k$ 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 $\rho_k$ is linear in the parameters, $\hat{y}_k(t) = \phi^T(t) \theta_k$, 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 ($w(\cdot, \cdot)$ 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

Wiener

Hammerstein

Hammerstein-Wiener
Other Combinations

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

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

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

Linear Model

Black: Measured Output
Blue: Model Simulated Output

Hammerstein Model of the Hydraulic Crane

Hammerstein model: Fit 71.61%

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

- 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

\[
\hat{y} (\theta) = f(\hat{Z}^{t-1}, \theta) = f(\varphi(t), \theta) \\
\varphi(t) = \varphi(\hat{Z}^{t-1}) \quad \text{“state” of fixed dimension} \\
f(\varphi, \theta) = \sum_{k=1}^{d} a_k \kappa(\beta_k(\varphi - \gamma_k)), \quad \theta = \{a_k, \beta_k, \gamma_k\} \quad \kappa: \text{unit function}
\]

Intuitive picture: Think of a scalar \( \varphi \) and let \( \kappa(z) \) be a unit pulse for \( 0 \leq z \leq 1 \). 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 approximating "any" function arbitrarily 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

Information in Unlabeled Regressors?

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

\[ \phi \]

Problem: associate two kinds of information:
- \( f(\phi(t)) \approx y(t) \) for the measured labels
- \( f(\phi(t)) \approx f(\phi(j)) \) if \( \phi(t) \approx \phi(j) \).

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

\[ \hat{y}_i := \sum_{j=1}^{N_L+N_U} K_{i,j} \hat{y}_j \quad K_{i,j} = K(\phi(t), \phi(j)) \]

Semi-supervised Regression: The Problem

Given a standard regression problem:

\[ y(t) = f(\phi(t)) + \text{noise} \]

\( f \) unknown, \( y(t) \) and \( \phi(t) \) observed for \( t = 1, \ldots, N_L \). Find \( f \)

Or rather, for any given \( \phi^* \) find a good value of \( f(\phi^*) \) (“Model on Demand”, “Just in time model”)

Extra feature: We have several measurements of \( \phi(t) \), \( t = N_L + 1, \ldots, N_L + N_U \) without corresponding values of \( y(t) \) (\( N_U \) “unlabeled observations”)

(Un-/Semi-)Supervised

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

The Suggested Method: WDMR (With H. Ohlsson)

WDMR: Weight Determination by Manifold Regression

(Manifold) Smoothness Assumption: \( f(\phi_1) \) and \( f(\phi_2) \) close if \( \phi_1 \) and \( \phi_2 \) are close (in a relevant metric).

Problem: associate \( \phi(t) \) with good values \( \hat{y}_i = f(\phi(t)) \) for all regressors, both labeled and unlabeled.

Take care of two kinds of information:
- \( f(\phi(t)) \approx y(t) \) for the measured labels
- \( f(\phi(t)) \approx f(\phi(j)) \) if \( \phi(t) \approx \phi(j) \).

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

\[ \hat{y}_i := \sum_{j=1}^{N_L+N_U} K_{i,j} \hat{y}_j \quad K_{i,j} = K(\phi(t), \phi(j)) \]

WDNR, cont’d

Now, weigh together the two sources of information

\[ \lambda \sum_{i=1}^{N_L+N_U} \sum_{j=1}^{N_L+N_U} K_{i,j} \hat{y}_j^2 + (1-\lambda) \sum_{i=1}^{N_L} (y(t) - \hat{y}_i)^2 \quad (\star) \]

\[ K_{i,j} = K(\phi(t), \phi(j)) \]

Pick a (“regularization parameter”) \( \lambda \) that balances the fit to measured labels and the smoothness prior. Minimize w.r.t. \( \hat{y}_i, t = 1, \ldots, N_L + N_U \).

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

Note that (\( \star \)) is quadratic in \( \hat{y}_i \), 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 $\varphi(t) \in \mathbb{R}^{128}$ 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 $\varphi$-space $\mathbb{R}^{128}$ to the scalar space of viewing angle $[-\pi, \pi]$. 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 $\mathbb{R}^{128}$ 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 ($[-\pi/2, \pi]$), (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).