System Identification, Lecture 11

Kristiaan Pelckmans (IT/UU, 2338)

Course code: 1RT880, Report code: 61800 - Spring 2012
F, FRI Uppsala University, Information Technology

16 Mai 2012

Overview Part II

1. State Space Systems.
2. Subspace Identification.
3. Further Topics.
4. Identification of Nonlinear Models.
5. Wider View.

Overview Identification of Nonlinear Models

1. Taxonomy.

2. Nonlinear Dynamic Models.

3. Nonlinear Approximation.

Nonlinear Dynamic Models

Definition 1. [Linear Superposition Principle (LSP)] *Let* $\mathcal{S} : \mathbb{R}^n \to \mathbb{R}^n$ *be a system. It satisfies the linear superposition principle iff for any inputs* $\mathbf{u}, \mathbf{u}' \in \mathbb{R}^n$*, one has that*

$$\mathcal{S}(\mathbf{u}) + \mathcal{S}(\mathbf{u}') = \mathcal{S}(\mathbf{u} + \mathbf{u}')$$

Deviations from LSP ('nonlinear'):

- Time-varying.
- Dynamics depending on inputs (operation regime).
- Saturation, Quantization, Hysteresis, Threshold effects, Limit Cycles.

Nonlinear Models

LTI:

$$\begin{cases} \mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t \\ \mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t \end{cases}$$

Parameter Varying:

$$\begin{cases} \mathbf{x}_{t+1} = \mathbf{A}_t\mathbf{x}_t + \mathbf{B}_t\mathbf{u}_t \\ \mathbf{y}_t = \mathbf{C}_t\mathbf{x}_t + \mathbf{D}_t\mathbf{u}_t \end{cases}$$

Bilinear:

$$\begin{cases} \mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}(\mathbf{x}_t \times \mathbf{u}_t) \\ \mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t \end{cases}$$

Nonlinear:

$$\begin{cases} \mathbf{x}_{t+1} = g(\mathbf{x}_t, \mathbf{u}_t) \\ \mathbf{y}_t = h(\mathbf{x}_t, \mathbf{u}_t) \end{cases}$$

Different places to put noise in.

Block-Structured Nonlinear Models

Compromise between Flexibility and Insight.

- Hammerstein models.
- Wiener models.
- Hammerstein-Wiener Models.
- Wiener - Hammerstein Models.
- Volterra Models.

Predictor Models

Optimal Predictor (PEM) for LTI:

- In general model

$$y_{t+1} = H(q^{-1}, \theta_0)u_{t+1} + G(q^{-1}, \theta_0)e_{t+1}$$

where $H(q^{-1}, \theta_0) = 1 + h_1 q^{-1} + \dots$ and $G(q^{-1}, \theta_0) = 1 + g_1 q^{-1} + \dots$.

- Rewrite as optimal predictor

$$\hat{y}_{t+1|t} = L_1(q^{-1}, \theta_0)u_{t+1} + L_2(q^{-1}, \theta_0)y_{t+1}$$

where $L_2(1, \cdot) = 0$.

- Optimal predictor

$$\hat{y}_{t+1|t} = \big(G^{-1}(q^{-1}, \theta_0)H(q^{-1}, \theta_0)\big)u_{t+1} + \big(1 - G^{-1}(q^{-1}, \theta_0)\big)y_{t+}$$

But this does not work in general:

- H, G?
- Monic G?
- Invertible?
- Convolution?

$\Rightarrow$ no general optimal predictor corresponding to nonlinear models.

Trick: formulate *predictor model*:

$$y_{t|t-1} = f_\theta(z_t)$$

with

$$z_t = (u_t, \ldots, u_{t-n}, y_{t-1}, \ldots, y_{t-d'})$$

But description of dynamics?

Model estimation:

$$\min_{\theta \in \Theta} \sum_{t=k'}^{n} \ell \left(y_t - f_\theta(z_t) \right)$$

where

- $'f'_\theta$ is a nonlinear function with unknowns θ.
- $\Theta = \{\theta\}$ set of plausible values.
- Loss function $\ell : \mathbb{R} \to \mathbb{R}$.

Perspectives:

1. Algorithmic.
2. Representation.
3. Convergence.
4. Inference.

$\to$ Model selection.

Choice of Models

Bias - variance decomposition:

$$\mathbb{E}\|f_* - f_{\hat{\theta}}\|^2 = \|f_* - f_{\theta_*}\|^2 + \mathbb{E}\|f_{\theta_*} - f_{\hat{\theta}}\|^2$$

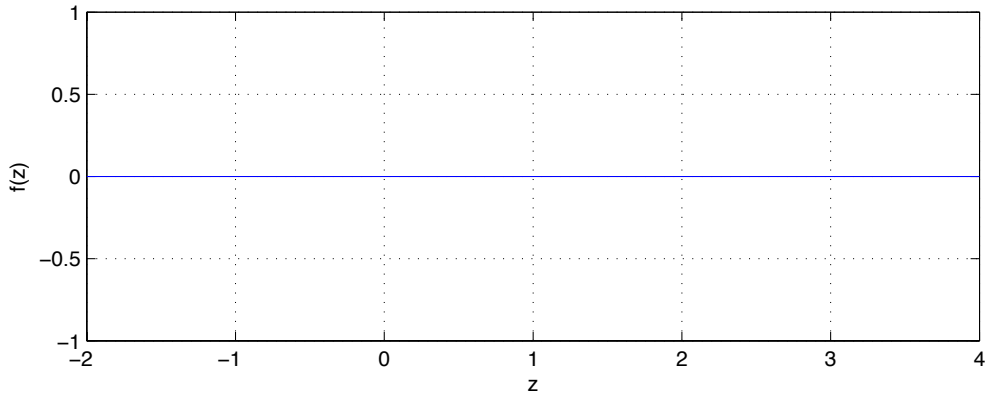
where:

- f_{θ_*} equals best one could do in $\{f_\theta : \theta \in \Theta\}$.
- $\|f_* - f_{\theta_*}\|^2$ equals bias^2, proportional to 'form' Θ.
- $\mathbb{E}\|f_{\theta_*} - f_{\hat{\theta}}\|^2$ variance proportional to size Θ.

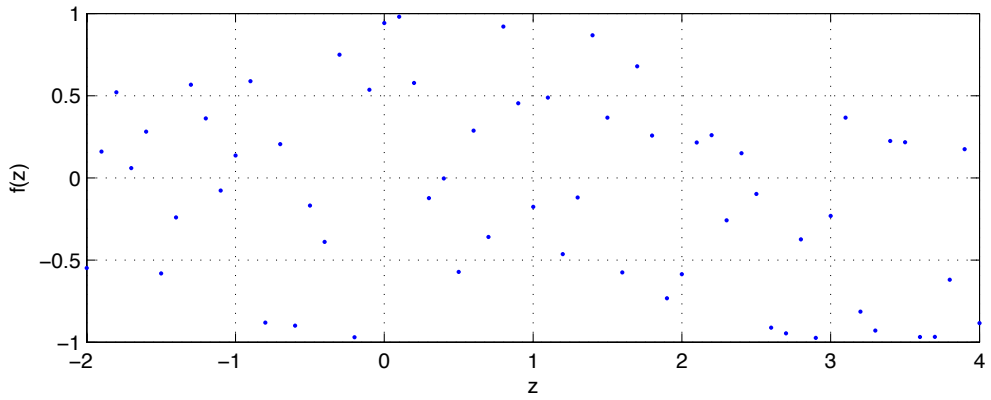
So choose a f_θ and Θ which can be expected to trade bias and variance optimally. Extrema:

- Constant function $f(z) = 0$.
- Lookup table with infinite number of entries $\theta = \{(z, y)\}$.

Constant $f(z) = 0$:



Lookup table with 60 entries $\{(-2, 0.1), \ldots, (4, 0.15)\}$:



General Approximators: Basis Functions

Abstract into 1D: $f_* : \mathbb{R} \to \mathbb{R}$ approximated by $f_\theta : \mathbb{R} \to \mathbb{R}$. Given set $\{\phi_i : \mathbb{R} \to \mathbb{R}\}$, assume the function

$$f_\theta(x) = \sum_{i=1}^{m} \theta_i \phi_i(x)$$

- Linear in the parameters θ! So

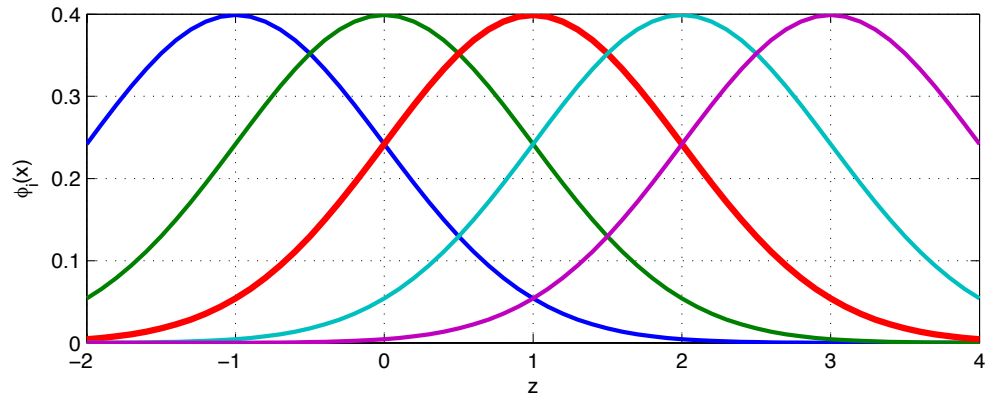
$$\hat{\theta} = \operatorname*{argmin}_{\theta \in \mathbb{R}^m} \sum_{t=k'}^{n} \left(y_t - f_\theta(z_t)\right)^2$$

Analytical solution as

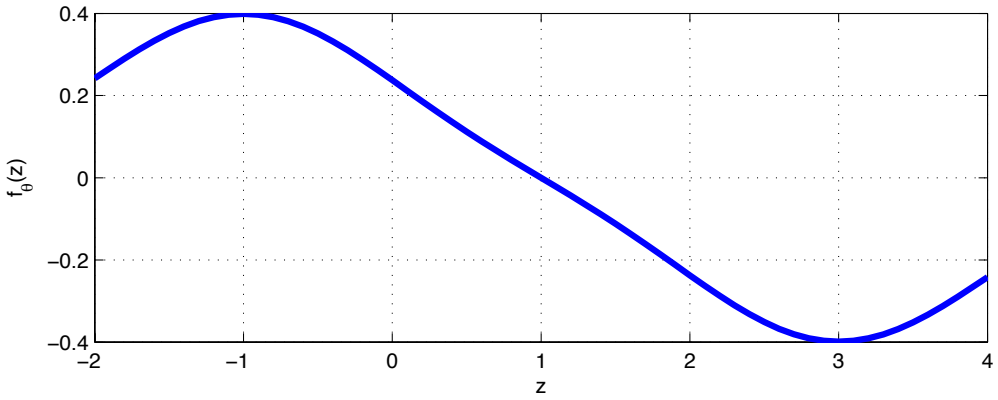
$$\mathbf{R}_m \hat{\theta} = \mathbf{r}_m$$

with covariance matrix $\mathbf{R}_{m,ij} = \sum_{t=1}^{m} \phi_i(z_t)\phi_j(z_t)$ and $\mathbf{r}_{m,i} = \sum_{t=1}^{m} \phi_i(z_t) y_t$.

5 basis functions:



A function f_θ with parameter vector $(1, 0, 0, 0, -1)$:



1. If $m = o(n)$?

2. Choice of Basis Functions.

3. Linear in parameters.

Too much freedom: Regularization

Parameteric methods:

$$\min_{\theta \in \mathbb{R}^d} \sum_t \ell(y_t - f_\theta(z_t))$$

When $d \to n$, too high variance (or $\mathbf{R}_d$ ill-conditioned). Better

$$\min_{\theta \in \Theta} \sum_t \ell(y_t - f_\theta(z_t)) + \gamma c(\theta)$$

where

- $c(\theta)$ measures complexity.
- $\gamma > 0$ regularization trade-off.
- If two models give equivalent fit, choose the least complex one.
- $c(\theta) \to c'(f_\theta)$.

General Approximators: Artificial Neural Networks

Model:

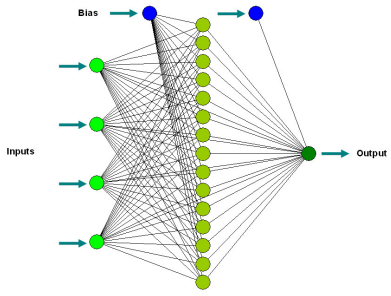
$$f_\theta(x) = \sum_{i=1}^{m} \theta_i \phi_i(x;\theta)$$

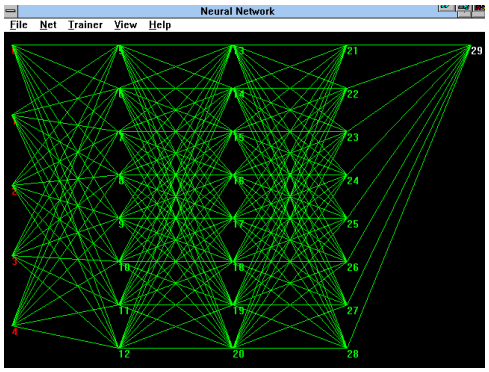
and

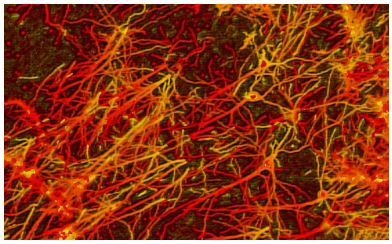
$$\phi_i(x;\theta) = \sum_{i=1}^{m} \theta'_i \phi'_i(x;\theta)$$

and ...

Graphical representation:







But:

- Algorithmic (backpropagation).
- Representation and Design.
- Optimality?

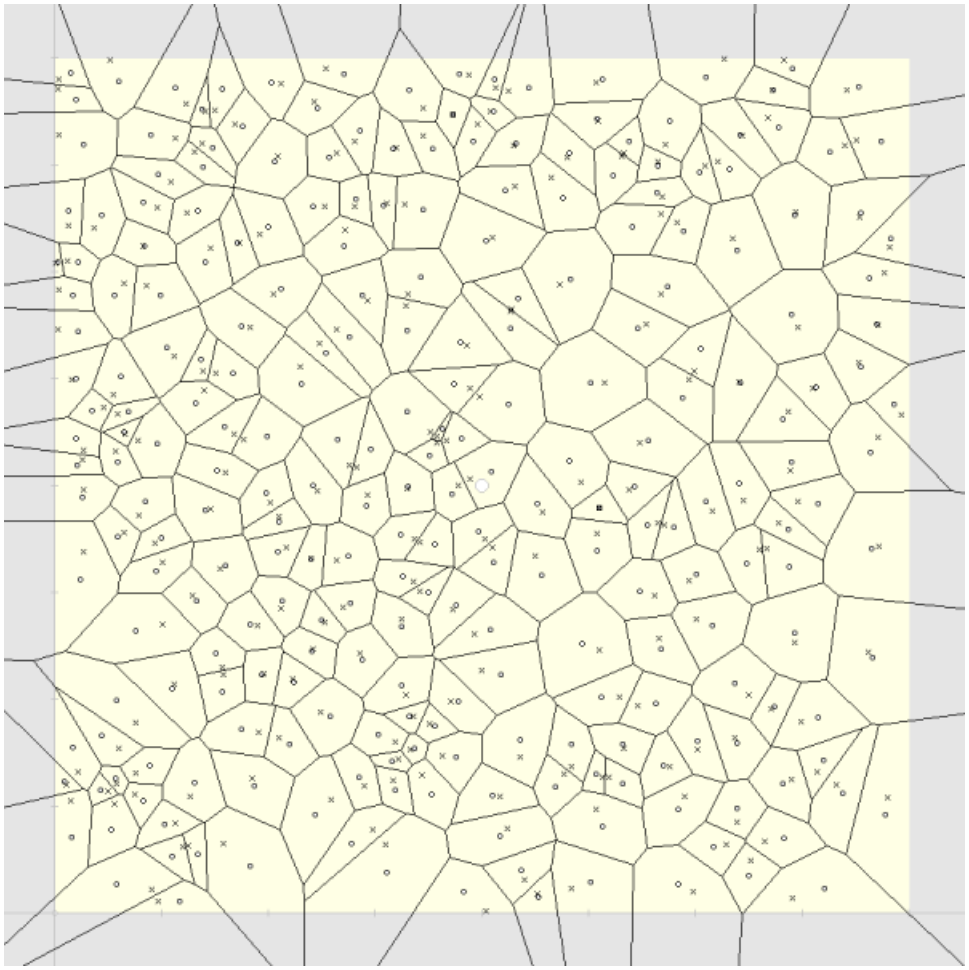
In 90s (70s): when starting from proper optimality function

$$\min_{f \in \mathcal{F}} \sum_t (y_t - f(x_t))^2 + \gamma \|f\|_H$$

then *optimal* representation (network):

$$f(x) = \sum_{i=1}^{n} \alpha_i K(x_i, x)$$

General Approximators: Nearest Neighbor rules



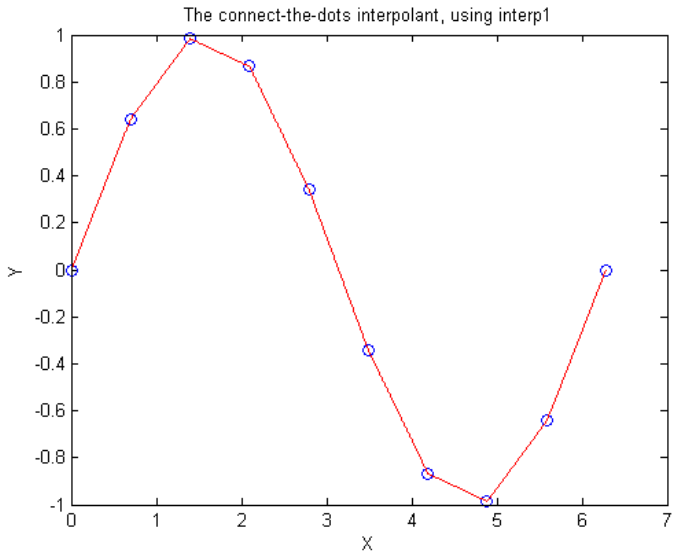
$$f_\theta(x) = y_i I(x \in S_i)$$

General Approximators: Piecewise Linear Systems

Divide domain in disjunct regions $\{S_i\}$

$$f_\theta(x) = \sum_{i=1}^{m} I(x \in S_i)(\theta_i x + b_i)$$

such that joined at knots.



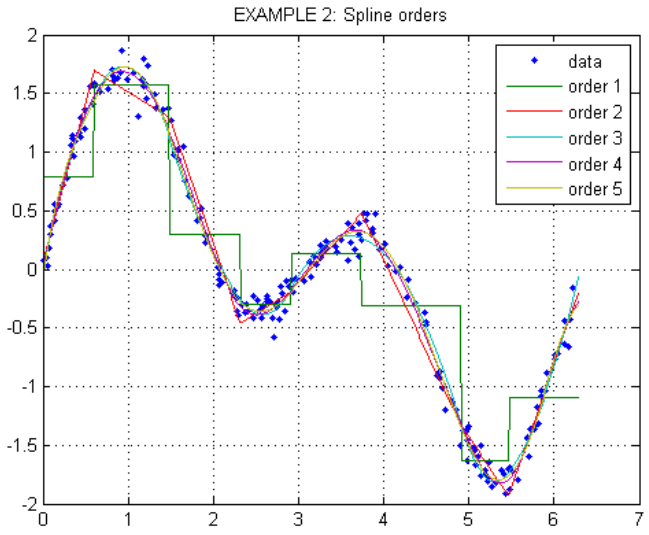
But where to put the knots?

General Approximators: Splines

Divide domain in disjunct regions $\{S_i\}$

$$f_\theta(x) = \sum_{i=1}^{m} I(x \in S_i)(\theta_i x^d + \cdots + b_i)$$

such that joined and differentiable at knots.



But where to put the knots?

- Interpolation $\rightarrow$ B-Splines (numerical).

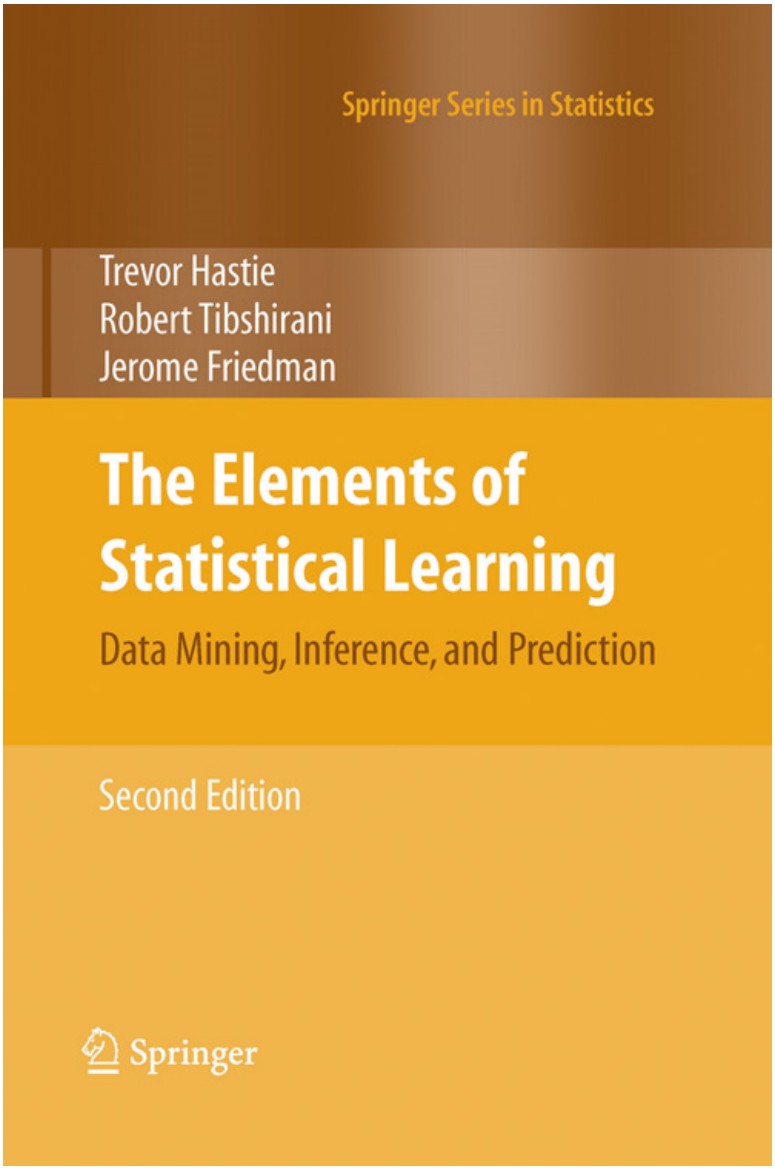
- Noise $\rightarrow$ Smoothing Splines (Bayesian).

General Approximators: Nonparametric Techniques

- Parameter set Θ to function set $\{f\}$.
- No explicit form.
- Algorithmic construction.
- Semi-parameteric $f(z) = f_\theta(z) + g(z)$.
- Fitting noise.
- Prediction and generalization.
- High-dimensional problems.

General Approximators

The Elements of Statistical Learning, Hastie et al. 2002,2009.



Conclusions

To remember

- Nonlinear Dynamic Models.
- Regularization.
- Toolbox.
- Optimization.