Modeling and Learning for Dynamical Systems

Lecture 8

Martin Enqvist

LINKÖPING UNIVERSITY

Choice of Model Order and Structure – Model Validation and Regularization

Basic Model Validation Principles

- It is not possible to give a general statement about what is a good cost function value or model fit since these values depend on the noise level even for a perfect model
- Increasing the model order or using a more flexible model structure result in a lower value of the cost function (or a higher model fit) if evaluated on *estimation data* since more parameters are available to describe the input-output relation. Problem: **overfit** (some parameters are used to describe the particular noise realization in the estimation dataset and will just give errors when the model is used for a new dataset)
- There are approaches for distinguishing the relevant model fit from overfit by modifying the cost function such that it penalizes the use of many parameters (these methods have names like FPE, AIC, BIC, MDL)
- However, the best approach is always to evaluate the model fit on a separate validation dataset: **cross-validation** (provided that this dataset is available or can be obtained by splitting the original dataset)

Validation Approaches

- Cross-validation: Evaluate model fit on validation data
- Compare parametric and non-parametric models in time and frequency domain
- Look at the auto-correlation of the model residual $\varepsilon(t)$ (shows the quality of the noise model) and the cross-correlation between $\varepsilon(t)$ and $u(t-\tau)$ (nonzero values for $\tau \geq 0$ indicate undermodeling and nonzero values for $\tau < 0$ indicate the presence of feedback)
- Look at parameter uncertainties (to see if there are unneccessary parameters)
- Look at the zeros and poles (to see if there are possible pole-zero cancellations)

Regularization

- Besides the problem of overfitting, the use of a higher order or more flexible model structure will usually result in a higher variance of the model estimate.
- The use of cross-validation will penalize also this, but the result could be undermodeling (the estimation dataset might contain too little information to estimate an accurate model of the same complexity as the true system)
- **Regularization** can be a useful tool if some prior knowledge about the system is available. In this case, the following problem is solved:

$$\hat{\theta}_N = \arg\min_{\theta} V_N(\theta)$$

$$\text{where } V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N} l(\varepsilon(t,\theta)) + h_N(\theta)$$

Here, the function $h_N(\theta)$ is selected based on different assumptions about the properties of the true system. For example, the smoothness properties of the true impulse response can be expressed using particular functions, known as *kernels*, and several options are available.

Some Regularization Kernels

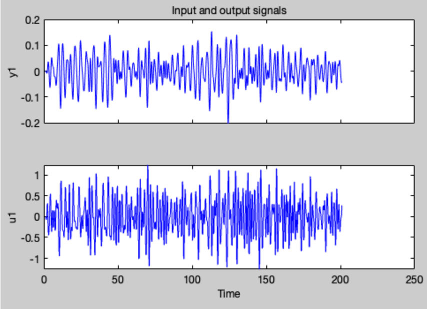
Some common regularization kernels:

- TC: Tuned and correlated kernel
- SE: Squared exponential kernel
- SS: Stable spline kernel
- HF: High frequency stable spline kernel
- DI: Diagonal kernel
- DC: Diagonal and correlated kernel

(The detailed descriptions are beyond the scope of this course, but the practical advice is to simply try several of them)

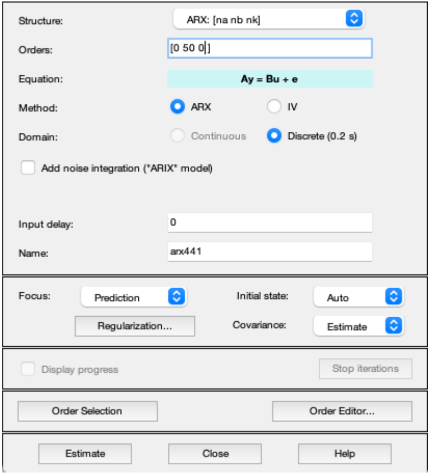
Matlab Regularization Example

Consider a dataset with a narrow-banded input signal (which provides less information about the system than for example a white-noise input)



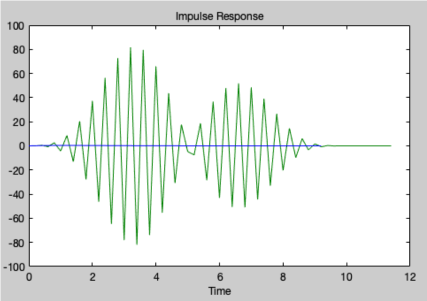
Matlab Regularization Example...

We will estimate FIR models with 50 parameters



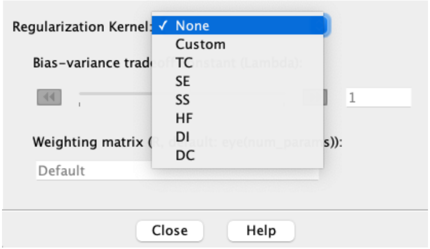
Matlab Regularization Example...

The LS estimate (green) is far from the true impulse response (blue)



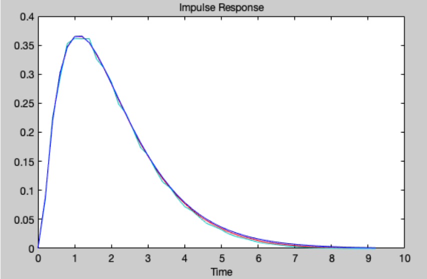
Matlab Regularization Example. . .

Let us try to use regularization



Matlab Regularization Example...

The regularized FIR estimates with DC (red) and TC (cyan) kernels are very similar to the true impulse response (blue)



Statistical Properties of PEM Model Estimates

Bias and Variance

Two types of errors in a model:

- **Bias** errors: Systematic errors, for example due to an unsuitable model structure or feedback effects.
- **Variance** errors: Random model errors due to the effects of noise. Can typically be reduced by using more data for the estimation.

Convergence and Bias

Key result 1:

$$\hat{\theta}_N \to \theta^* = \arg\min_{\theta} \bar{V}(\theta) \text{ w.p.1 as } N \to \infty$$

where $\bar{V}(\theta) = \mathrm{E}(\varepsilon^2(t,\theta))$.

Key result 2 (linear system and linear model with fixed noise model H_*):

$$\hat{\theta}_N \to \theta^* = \arg\min_{\theta} \int_{-\pi}^{\pi} |G_0(e^{i\omega}) - G(e^{i\omega},\theta)|^2 \frac{\Phi_u(\omega)}{|H_*(e^{i\omega})|^2}\, d\omega \text{ w.p.1 as } N \to \infty$$

These results can be used to analyze the convergence and bias properties of the PEM parameter estimates

Variance

Key result 3 (unbiased estimator where the residual for $\theta = \theta_0$ is white noise with variance λ):

$$P_N = \mathrm{E}((\hat{\theta} - \theta_0)(\hat{\theta} - \theta_0)^T) \approx \frac{1}{N} \lambda \bar{R}^{-1}$$

where

$$\bar{R} = \mathrm{E}(\psi(t, \theta_0)\psi^T(t, \theta_0))$$

$$\psi(t, \theta) = \frac{d}{d\theta} \hat{y}(t|\theta)$$

This result can be used to estimate and analyze the variance of the PEM parameter estimates.

Summary

- Overfit
- Cross-validation
- Validation approaches
- Regularization
- Three key results about the convergence, bias and variance of prediction-error method (PEM) parameter estimates

www.liu.se

LINKÖPING
UNIVERSITY