CH5230: System Identification

Arun K. Tangirala

Department of Chemical Engineering, IIT Madras
January - April 2011
Lecture Set 1

Lectures:Mon, Thu, Fri Extra class: Monday evenings

Predictions

One of the primary uses of a model is prediction. We first learn how to build a predictor given a model, particularly the LTI description

$$y[k] = G(q)u[k] + v[k] = G(q)u[k] + H(q)e[k]$$

- Clearly, prediction requires knowledge of the past/present in addition to the model
  - The model tell us how the future evolves as a function of the past/present
- The "quality" (accuracy, precision) of the prediction clearly depends on
  - "Quality" of the model, prediction horizon (how far ahead we wish to predict) and uncertainty levels (variance of v[k])
- *One of the foremost uses of a prediction expression is in the construction of a prediction error, which can be then used in estimating the model*
  - For e.g., minimization of squared prediction error is commonly used to estimate G(q)

One-step ahead predictions

We begin with one-step ahead predictions. Predictions are often denoted by a hat, for e.g., $\hat{y}[k|k-1]$ should be understood as prediction of $y[k]$, given all the information up to k-1

- In making predictions, we assume inputs are known accurately.
- Given a model (i.e., for a fixed model), the one-step ahead prediction

$$\hat{y}[k|k-1] = G(q)u[k] + \hat{v}[k|k-1]$$

- The task therefore is to build a predictor for v[k]
  - In practice, we do not have information on v[k−1], but rather have measurements y[k−1] & inputs u[k−1]
- We first learn now how to predict a stochastic signal given its past/ present.

Prediction - fundamental approach

- The fundamental problem in forecasting is that of approximating a (random) variable (that is to be predicted) given another (random) variable
- To obtain the best forecast, we study an important result that allows us to break up a RV Y into two components, one that depends on another RV X and another that is orthogonal (or uncorrelated) to the first.

Decomposition Property: Any random variable y can be expressed as

$$y = E(y|x) + \epsilon$$

where ϵ is a random variable satisfying
(i) $E(\epsilon|x) = 0$ and (ii) $E(h(x)\epsilon) = 0$ where $h(.)$ is any function of x

Proof:
i)

$$E(\epsilon|x) = E(y - E(y|x)) = E(y|x) - E(y|x) = 0$$

ii)

$$E(\epsilon h(x)) = E(h(x)E(\epsilon|x)) = 0$$

Best prediction: Conditional Expectation

The first condition on residual says it has zero conditional mean and the second term states that the residual lies in an orthogonal space to that of the explanatory variable
Any variable can be decomposed into two components - (i) the conditional expectation and (ii) an orthogonal error term

- What is the best approximation of a RV *Y* given another RV *X*?
- By best, we mean the one that minimizes the mean square error

Prediction Property: Let $h(x)$ be any function of x. Then

$$E(y|x) = \min_{m(x)} E(y - m(x))^2$$

The conditional expectation E(Y | X) gives the minimum mean square error approximation of Y given another RV X

Example 1: Prediction for MA(1) process

Consider the following example of an MA(1) process: $v[k] = e[k] + c_1 e[k\text{-}1]$

- The one-step ahead predictor is

$$\hat{v}[k+1|k] = E(v[k+1]|k) = c_1 e[k|k]$$

  - The quantity e[k|k] is not known, but has to be estimated using the observations

$$e[k|k] = (1 + c_1 q^{-1})^{-1} v[k] = \sum_{j=-\infty}^{0} (-c_1)^{-j} v[k-j]$$

- Clearly, if the "prediction" of e[*k*|*k*] to be stable, the magnitude of d_1 should be less than unity
  - Stated otherwise, the inverse of the noise model should be stable
  - This is an ideal predictor. In practice, truncated predictors have to be used.
- ***If the model is unknown, the predictor is non-linear in the unknowns (c_1 and e[k])!***
  - This is always the case with Moving Average models

Example 2: Prediction of an AR(1) process

Consider now an AR(1) process: $v[k] = -d_1 v[k\text{-}1] + e[k]$

- The one-step ahead prediction of $v[k+1]$ is then,

$$\begin{aligned}\hat{v}[k+1|k] =& E(v[k+1]|k) \\ =& -d_1 v[k] + E(e[k+1]|k) \\ =& -d_1 v[k]\end{aligned}$$

- The second term is zero by virtue of definition of white-noise
- **If the model is unknown, the predictor is linear in the unknown**
  - This is always the case with auto-regressive models
  - Minimization of squared prediction errors with linear predictors provide unique solutions
  - Thus, auto-regressive models hold an edge over moving average models in identification

General expression for prediction

We now develop a general expression for prediction. This general expression involves the **inverse of noise model**.

- Consider the causal impulse response model for v[k]

$$v[k] = \sum_{n=0}^{\infty} h[n] e[k-n] \quad \text{with} \quad h[0] = 1$$

  so that

$$H(q^{-1}) = \sum_{n=0}^{\infty} h[n] q^{-n}$$

- Then, for one-step ahead prediction, we re-write

$$v[k] = e[k] + \sum_{n=1}^{\infty} h[n] e[k-n]$$

  so that the second-term on the RHS is (theoretically) known at *k*-1
  - The best prediction of the first term is zero given any amount of past

One-step ahead prediction

- The best one-step ahead prediction is

$$\begin{aligned}\hat{v}[k|k-1] &= \sum_{n=1}^{\infty} h[n]e[k-n] \\ &= \sum_{n=0}^{\infty} h[n]q^{-n}e[k-n] - e[k] \\ &= (H(q^{-1})-1)e[k]\end{aligned}$$

  - The quantities e[k−1], e[k−2], ... are unknown, but their effects are "felt" in v[k−1], v[k−2], ... It is a good practice to re-write the predictions in terms of known quantities.
  - The following result is useful. If,

$e[k] = \sum_{n=0}^{\infty} \tilde{h}[n]v[k-n]$ then, $\left\{\tilde{h}[n]\right\}$ are the IR coefficients of $H^{-1}(z)$

which means we may write,

$$e[k] = H^{-1}(q)v[k]$$

One-step ahead prediction

- Thus,

$$\hat{v}[k|k-1] = (1 - H^{-1}(q))v[k]$$

  - The RHS of the expression may look a bit awkward, but by recalling that the leading coefficient of the noise model is unity.

- We substitute this expression in our prediction for *y*[*k*]. Thus,

$$\hat{y}[k|k-1] = G(q)u[k] + (1 - H^{-1}(q))v[k]$$

  - The quantity v[k] is unknown, but "felt" in y[k] and can be recovered as

$$v[k] = y[k] - G(q)u[k]$$

- Therefore, we write the expressions for predictions and prediction errors

$$\boxed{\hat{y}[k|k-1] = H^{-1}(q)G(q)u[k] + (1 - H^{-1}(q))y[k]}$$

$$\boxed{\epsilon[k|k-1] = y[k] - \hat{y}[k|k-1] = H^{-1}(q)(y[k] - G(q)u[k])}$$

Predictors for parametric models

Using the general expression for one-step ahead predictions, we can develop the predictors (and the errors) for different parametric models

FIR	$\hat{y}[k\|k-1] = B(q)u[k] \quad \text{since} \quad (H(q) = 1)$
ARX	$\hat{y}[k\|k-1] = B(q)u[k] + (1 - A(q))y[k]$
ARMAX	$\hat{y}[k\|k-1] = B(q)u[k] + \left(1 - \frac{A(q)}{C(q)}\right) y[k]$
OE	$\hat{y}[k\|k-1] = G(q)u[k]$

- The FIR model is both a non-parametric as well as a parametric model
- Both the OE and FIR model predictions do not involve any output measurements

Difference equation forms

For implementation purposes, it is best to re-write the predictors in difference equation forms

- FIR:
$$\hat{y}[k|k-1] = b_1u[k-1] + \cdots + b_{n_b}u[k-n_b]$$
- ARX:
$$\begin{aligned}\hat{y}[k|k-1] = &- a_1y[k-1] + \cdots - a_{n_a}y[k-n_a] \\ &+ b_1u[k-1] + \cdots + b_{n_b}u[k-n_b]\end{aligned}$$
- ARMAX: (1,1,1)
$$\begin{aligned}\hat{y}[k|k-1] = &- c_1\hat{y}[k-1] + b_1u[k-1] + b_1c_1u[k-2] \\ &+ (c_1 - a_1)y[k-1]\end{aligned}$$
- OE:
$$\begin{aligned}\hat{y}[k|k-1] = &- a_1\hat{y}[k-1] + \cdots - a_{n_a}\hat{y}[k-n_a] \\ &+ b_1u[k-1] + \cdots + b_{n_b}u[k-n_b]\end{aligned}$$

Theoretical one-step ahead prediction error

We observe now that the theoretical one-step ahead prediction error is none other than the white-noise

- First, recall the definition of prediction error

$$\begin{aligned}\epsilon[k|k-1] &= y[k] - \hat{y}[k|k-1] \\ &= H^{-1}G(q)u[k] - H^{-1}(q)y[k] \\ &= e[k]\end{aligned}$$

- Thus, we have an alternate definition of the white-noise e[k] - it is the one-step ahead prediction error
  - For this reason, e[k] is also known as the **innovation** (the new quantity at k)
- In practice, however, the one-step ahead prediction error is never identical to the white-noise sequence because of modelling errors
  - The G and H that generate y[k] will be different from the estimated G and H

Linear regression form

It is useful to re-write the predictions in a linear regression form, particularly, in estimating the parameters of a model. The linear regression form is

$$\hat{y}[k] = \psi^T[k]\theta$$

where $\boldsymbol{\psi}$[k] is the regressor and θ is the parameter vector

- **FIR model:**

$$\begin{aligned}\psi[k] &= \begin{bmatrix} u[k-1] & u[k-2] & \cdots & u[k-n_b] \end{bmatrix}^T \\ \theta &= \begin{bmatrix} b_1 & b_2 & \cdots & b_{n_b} \end{bmatrix}^T\end{aligned}$$

- **ARX model:**

$$\begin{aligned}\psi[k] &= \begin{bmatrix} -y[k-1] & \cdots & -y[k-n_a] & u[k-1] & \cdots & u[k-n_b] \end{bmatrix}^T \\ \theta &= \begin{bmatrix} a_1 & \cdots & a_{n_a} & b_1 & \cdots & b_{n_b} \end{bmatrix}^T\end{aligned}$$

Pseudo-linear regressor form

All models except FIR and ARX models produce predictors that are non-linear in unknowns. However, we may re-write them in a pseudo-linear in unknowns

- Consider the ARMAX case. The predictor can then be obtained by applying the conditional expectation result,

$$\begin{aligned}\hat{y}[k|k-1] = &- a_1 y[k-1] + \cdots - a_{n_a} y[k-n_a] \\ &+ b_1 u[k-1] + \cdots + b_{n_b} u[k-n_b] \\ &+ c_1 e[k-1] + \cdots + c_{n_c} e[k-n_c]\end{aligned}$$

  - From a model estimation perspective, if the past innovations e[k–j]'s were known, then the predictor is linear in unknowns. Then we could form the pseudo–linear regressor

$$\psi[k] = \begin{bmatrix} -y[k-1] & \cdots & -y[k-n_a] & u[k-1] & \cdots & u[k-n_b] & e[k-1] & \cdots & e[k-n_c] \end{bmatrix}^T$$

$$y[k] = \psi^T[k,\theta]\theta$$

  - The advantage is then we can use techniques for estimating linear regression models in an iterative way. **Observe that the regressor is now a function of the parameter vector**

Multi-step prediction

We may be interested in predicting p-steps ahead in many situations. For e.g., in predictive control strategies one usually decides the control moves by predicting the state of the process p-steps ahead in time.

- To build the p-step ahead predictor, first observe

$$\begin{aligned}v[k] &= \sum_{n=0}^{p-1} h[n]e[k-n] + \sum_{n=p}^{\infty} h[n]e[k-n] \\ &= \bar{H}_p(q)e[k] + \tilde{H}_p(q)e[k]\end{aligned}$$

  so that

$$\begin{aligned}H(q) &= \bar{H}_p(q) + \tilde{H}_p(q) \\ \hat{v}[k|k-p] &= \sum_{n=p}^{\infty} h[n]e[k-n] \\ &= \tilde{H}_p(q)e[k] \\ &= \tilde{H}_p(q)H^{-1}(q)v[k]\end{aligned}$$

p-Step ahead predictor

- The p-step ahead predictor for $y[k]$ is now obtained as

$$\begin{aligned} y[k|k-p] &= G(q)u[k] + \hat{v}[k|k-p] \\ &= G(q)u[k] + \tilde{H}_p(q)H^{-1}(q)v[k] \\ &= (1 - \tilde{H}_p(q)H^{-1}(q))G(q)u[k] + \tilde{H}_p(q)H^{-1}(q)y[k] \\ &= \bar{H}_p(q)H^{-1}(q)G(q)u[k] + (1 - \bar{H}_p(q)H^{-1}(q))y[k] \end{aligned}$$

- Define

$$W_p(q) = \bar{H}_p(q)H^{-1}(q)$$

so that

$$\hat{y}[k|k-p] = W_p(q)G(q)u[k] + (1 - W_p(q))y[k]$$

- **Thus, the *p*-step ahead prediction is equivalent to one-step ahead prediction with a noise model $W_p(q)$**
  - Verify that p = 1 produces one-step ahead predictor (i.e., $W_1(q) = H^{-1}(q)$)

Infinite step-ahead predictions

- We now make an important observation with regards to the infinite step-ahead prediction ($p \to \infty$)

$$\text{As } p \to \infty,\ W_p(q) \to 1 \text{ since } \bar{H}_p(q) \to 1 \text{ as } p \to \infty$$

- Therefore, the infinite step-ahead prediction is

$$\hat{y}[k|-\infty] = G(q)u[k]$$

which is **identical** to the **one-step ahead prediction** of an **OE model**

  - Thus, minimizing the one-step ahead prediction error of an OE model is the same as minimizing infinite-step ahead prediction errors
  - Observe that the no stochastic terms enter the infinite step-ahead prediction

- *Infinite-step ahead prediction is therefore a one-step ahead prediction while treating the noise term as unpredictable,*
  i.e., $v[k] = e[k] \Rightarrow H = 1$!

Simulation

- The infinite-step ahead prediction is to be considered as the simulation of the process
  - Based purely on inputs. No measurements, i.e., "feedback" from the plant are involved.
- For any model, the acid test is its ability to simulate rather than simply deliver good one- or *p*-step ahead predictions
- The OE model by its inherent structure always yields infinite step-ahead predictions
- ***Therefore, estimation of G(q) by fitting an OE structure to the process and minimization of one-step ahead prediction errors should be expected to deliver "best" estimates of the plant model G(q).***
- *MATLAB commands*:
  `predict, pe, compare` (*default option: infinite-step ahead prediction*)

Summary

- Predictions are central to the end-use of models and their estimation
  - The predictor is entirely governed by the model structure and the given information
  - **Predictions are only stable if the noise model is invertible**
- Prediction expressions can be cast into linear regression forms
  - Certain model structures, namely, FIR and ARX produce predictors that are linear in model parameters
  - Other structures produce non-linear regressors
- The theoretical one-step ahead prediction error is the white-noise sequence, also known as **innovations**
- The one-step ahead prediction of an OE model = infinite-step ahead prediction
- Finally, stating the model of a process is equivalent to
  - Providing the predictor expression
  - Describing the prediction error (for e.g., white-noise)