Recursive Parameter Estimation with Constraints: a Discussion Document

B. Erik Ydstie, Carnegie Mellon University

Abstract

This handouts describes the constrained least squares estimation algorithm and modifications needed to allow a trade-off between an old (seed model) and new data.

1 Constrained Parameter Estimation

We consider the problem of estimating parameters of a system that can be written on the transfer function form

$$y(t) = \frac{A(q^{-1})}{B(q^{-1})} u(t-d) + e(t) \tag{1}$$

where $d \geq 1$ is the delay,

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n} \tag{2}$$

$$B(q^{-1}) = b_1 q^{-1} + \cdots + b_m q^{-m} \tag{3}$$

are polynomials in the backward shift operator q^{-1} and $e(t)$ is the external disturbance that represent model errors, bias and noise.

Many different methods have been proposed to represent the system including the series expansion method based on the use of orthogonal functions. In the case of Laguerre functions we write

$$y(t) = \sum_{i=1}^{n_L} g_i L_i(q) u(t) + e(t) \tag{4}$$

where

$$L_i(q) = \frac{\sqrt{(1-a^2)T_s}}{q-a} \left(\frac{1-aq}{q-a} \right)$$

are the Laguerre functions. The parameter a represents the dominant time constant. If $a = 0$ then we get the FIR model

$$y(t) = \sum_{i=1}^{n_L} u(t-i) + e(t) \tag{5}$$

Many possibilities exist. iLS uses a combination of methods.

At the heart of most linear estimation methods we find the linear regression model

$$\begin{aligned} \theta^T &= (-a_1, -a_2, ..., -a_n, b_1, b_2, ..., b_m) \\ \phi(t)^T &= (x(t), x(t-1), ..., x(t-n+1), u(t), u(t-1), ..., u(t-m+1) \end{aligned}$$

where $x(t)$ represents an estimate of the output of the transfer function. We can then write the prediction model

$$y(t) = \phi(t-1)^T \theta$$

We can then define the prediction error for a given set of parameters θ so that

$$y_f(t) = \phi(t-1)^T \theta^* + e_f(t) \tag{6}$$

We assume here that θ^* represents a good representation of the plant model and subscript f implies that the data has been filtered. Filtering has the effect of removing bias and high frequency noise. For the FIR case filtering is often achieved by using the filter

$$F(q^{-1}) = \frac{1}{1-q^{-1}} \frac{1-f}{1-fq^{-1}}$$

Where $0 \leq f < 1$ is the discrete time constant for the low pass filter. With $f = 0$ we get

$$y(t) = y(t-1) + \sum_{i=1}^{n_L} (1-q^{-1}) u(t-i) + e_f(t)$$

Many possibilities exist. In addition to filtering, it is also useful to normalize the data and introduce functions of the type

$$y = \frac{y_m - \text{mean}(y_m)}{y_m^{\max} - y_m^{\min}}, \quad y = \frac{u_m - \text{mean}(u_m)}{u_m^{\max} - u_m^{\min}}$$

where y_m refers to the measured value, $\text{mean}(y_m)$ to the mean value and $y_m^{\max}, y_m^{\min}$ to the maximum and minimum values respectively.

We assume that a good model fit is achieved by minimizing the prediction error while keeping the parameter estimates constrained to a compact set Θ, which we called the *admissible set* . The admissible set may simply consist of upper and lower bounds for each parameter or it may be more complex as we will see later. For a given data set we therefore find

$$\theta(t) = \text{argmin}_{\theta \in \Theta} J(\theta, t) \tag{7}$$

where the objective function given by

$$J(\theta, t) = \sum_{i=1}^{t} w_0(i)(y(i) - \varphi(i-1)^T \theta)^2 + (\theta - \theta_0(t))^T F_0(t)(\theta - \theta_0(t)) \tag{8}$$

Some parameters have been introduced

1. The weighting function $0 \leq w_0(i) < \infty$ assigns weight to data point i. If $w_0(i) = 0$ then the data point is discarded.

2. The matrix $0 \leq F_0(t) < \infty$ is introduced to bias the estimated parameters towards some pre-set values $\theta_0(t)$. This model may for example correspond to a "seed model" and $P(0) = F_0^{-1}$ provides an estimate of the covariance matrix of the seed model. This approach is referred to as ridge regression and it serves to regularize the optimization problem if it is ill-conditioned.

3. The constraint set may simply be upper and lower bounds so that

$$\theta_{i,\min} \leq \theta_i \leq \theta_{i,\max}$$

we may also introduce constraints on the steady state gain of the transfer function so that

$$K_{p,\min} \leq K_p \leq K_{p,\max}$$

where

$$K_p \frac{\sum b_i}{1 + \sum a_i}$$

This becomes a nonlinear constraint in the general case, but it is linear for Laguerre and FIR models. We can also introduce more general linear and nonlinear equality and inequality constraints.

We now define the data vectors

$$\begin{aligned} Y(t)^T &= (y(1), y(2), ...y(t)) \\ \Phi(t-1) &= (\phi(0), \phi(1), \phi(2), ..., \phi(t-1)) \\ E(t) &= (e(1), e(2), e(3), ..., e(t)) \end{aligned}$$

We can then write the model on the vector form

$$Y(t) = \Phi(t-1)^T \theta + E(t)$$

The objective function can then be written

$$J(\theta, t) = (Y(t) - \Phi(t-1)^T \theta)^T W_0 (Y(t) - \Phi(t-1)^T \theta) + (\theta - \theta_0)^T F_0 (\theta - \theta_0) \tag{9}$$

By differentiating with respect to θ we get the normal equations

$$-\Phi^T W (Y(t) - \Phi(t-1)^T \theta) - \theta_0(t)^T F_0 (\theta - \theta_0(t))$$

Or

$$(\Phi^T W \Phi(t-1)^T + F_0)\theta \Phi^T W Y(t) + F_0 \theta_0 \tag{10}$$

If there are no constraints then we can solve explicitly for the parameters and set

$$\theta(t) = (\Phi^T W \Phi(t-1)^T + F_0)^{-1} (\Phi^T W Y(t) + F_0 \theta_0)$$

This is rarely done in practice. Instead we solve the linear problem $eq : normal$ using Gauss elimination. With constraints we solve the Linear Qudratic Problem **??** with the inequality and equality constraints as needed to obtain a robust solution.

The parameter estimation problem posed above is not computationally tractable for large t since the numerical complexity grows as the number of data points grows. However, it is easy to reformulate the problem to reduce computational cost significantly while retaining the possibility of using growing data records. Towards this aim we assume that t is quite large

Following the methodology outlined in the course notes we get the following algorithm for the problem outlined above.

Constrained Least Squares Estimation:

Initialization: Define the admissible set Θ, choose $0 < \sigma < 1, r(0), \theta_0(0), F_0(0) > 0$.

Set $t = 1$ and execute the following algorithm:

Step 1: Update the variance estimate $r(t)$ (this must be done whether the parameters are changed or not) and then the switch $\Delta(t)$.

Step 2: If $\Delta(t) = 1$ update the forgetting factor $\lambda(t)$ and set

$$\begin{aligned} F(t) &= \lambda(t)F(t-1) + \frac{1}{r(t)}\varphi(t-1)\varphi(t-1)^T \\ W(t) &= \lambda(t)W(t-1) + \frac{1}{r(t)}\varphi(t-1)y(t) \end{aligned}$$

Choose $F_0(t) \leq F_0(t-1)$ so that

$$F_0(0) \leq F(t) + F_0(t) \tag{11}$$

Solve the following optimization problem using a nonlinear optimization routine such as IPOPT or fmincon in Matlab

$$\theta(t) = \mathrm{argmin}_{\theta\in\Theta} \left\{-2(W(t) + Q(t)\theta(0))^T\theta + \theta^T(F(t) + Q(t))\theta\right\}$$

Step 3: If $\Delta(t) = 0$ set

$$\begin{aligned} F(t) &= F(t-1) \\ W(t) &= W(t-1) \\ \theta(t) &= \theta(t-1) \end{aligned}$$

Step 4: Set $t = t + 1$ and go to Step 1.