Informationsteknologi

Introduction to recursive identification algorithms

Torsten Söderström

Division of Systems and Control
Department of Information Technology
Uppsala University, Sweden

Contents

Introduction

Consider the problem of estimating a constant signal from noisy measurements.

Extension to dynamic systems will be treated later.

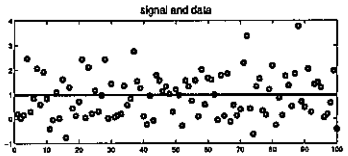
Introduction, cont'd

Model

$$y(t) = \theta + e(t)$$

Natural estimate

$$\hat{\theta}(t) = \frac{1}{t}\sum_{s=1}^{t} y(s)$$



Making the estimate recursive

Aim: Rewrite the estimate so that $\hat{\theta}(t)$ is updated from $\hat{\theta}(t-1)$ when the new measurement $y(t)$ is available.

$$\begin{aligned}\hat{\theta}(t) &= \frac{1}{t}\left[y(t) + \sum_{s=1}^{t-1} y(s)\right] \\ &= \frac{1}{t}\left[y(t) + (t-1)\hat{\theta}(t-1)\right] \\ &= \hat{\theta}(t-1) + \frac{1}{t}\left[y(t) - \hat{\theta}(t-1)\right]\end{aligned}$$

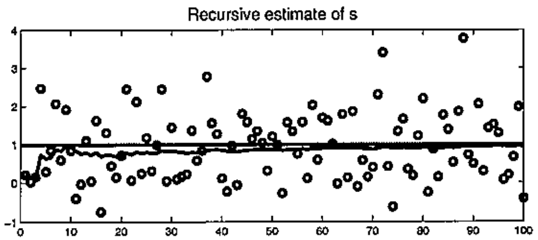
This equation captures the essence of recursive algorithms (also for dynamic systems)!

Interpretation

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{1}{t}\left[y(t) - \hat{\theta}(t-1)\right]$$

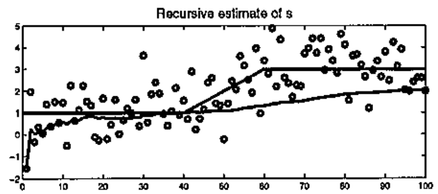
$\hat{\theta}(t)$	new estimate
$\hat{\theta}(t-1)$	old estimate
$1/t$	gain sequence
$y(t)$	new measurement
$\hat{\theta}(t-1)$	prediction of $y(t)$ based on available data at time $t-1$ and model parameter $\hat{\theta}(t-1)$
$y(t) - \hat{\theta}(t-1)$	prediction error

Simulations



Time-varying signals

Assume $\theta(t)$ changes with time.



After a change, give less weight to old data. How?

Approach for time-varying case

Time-invariant case

$$V_t(\theta) = \sum_{s=1}^{t}[y(s) - \theta]^2$$

Weighted criterion for the time-varying case

$$V_t(\theta) = \sum_{s=1}^{t}\beta(t,s)[y(s) - \theta]^2$$

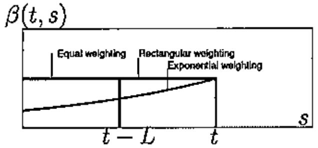
Weights $\beta(t,s)$ should fulfil

- $\beta(t,t) = 1$
- $\beta(t,s) \ll 1$ when $s \ll t$

Approach for time-varying case

Weighting patterns

$$V_t(\theta) = \sum_{s=1}^{t} \beta(t,s)[y(s) - \theta]^2$$



Approaches for time-varying case

- Exponential forgetting
  $V_t^{(e)}(\theta) = \Sigma_{s=1}^t \lambda^{t-s}[y(s) - \theta]^2 \quad \lambda < 1$
- Rectangular forgetting
  $V_t^{(r)}(\theta) = \Sigma_{s=t-L+1}^t [y(s) - \theta]^2$
- Observer
  $\hat\theta(t+1) = \hat\theta(t) + K[y(t) - \hat\theta(t)]$

Exponential forgetting

Criterion

$$\begin{aligned} V_t^{(e)}(\theta) &= \sum_{s=1}^{t} \lambda^{t-s}[y(s)-\theta]^2 \\ &= \theta^2 \sum_{s=1}^{t} \lambda^{t-s} - 2\theta \sum_{s=1}^{t} \lambda^{t-s} y(s) + \text{const.} \end{aligned}$$

Estimate

$$\begin{aligned} \hat\theta(t) &= \frac{\Sigma_{s=1}^t \lambda^{t-s} y(s)}{\Sigma_{s=1}^t \lambda^{t-s}} = \\ &= \frac{1}{\Sigma_{s=1}^t \lambda^{t-s}} \left[y(t) + \lambda \sum_{s=1}^{t-1} \lambda^{t-1-s} y(s) \right] \end{aligned}$$

Exponential forgetting, cont'd

$$\begin{aligned} \hat\theta(t) &= \frac{1}{\Sigma_{s=1}^t \lambda^{t-s}} \left[y(t) + \lambda \sum_{s=1}^{t-1} \lambda^{t-1-s} y(s) \right] \\ &= \frac{1-\lambda}{1-\lambda^t} \left[y(t) + \lambda \frac{1-\lambda^{t-1}}{1-\lambda} \hat\theta(t-1) \right] \\ &= \frac{1-\lambda}{1-\lambda^t} y(t) + \frac{\lambda - \lambda^t}{1-\lambda} \hat\theta(t-1) \\ &= \hat\theta(t-1) + \frac{1-\lambda}{1-\lambda^t} y(t) + \frac{-1+\lambda}{1-\lambda^t} \hat\theta(t-1) \\ &= \hat\theta(t-1) + \frac{1-\lambda}{1-\lambda^t} [y(t) - \hat\theta(t-1)] \end{aligned}$$

Rectangular forgetting

Criterion

$$\begin{aligned} V_t^{(r)}(\theta) &= \sum_{s=t-L+1}^{t} [y(s)-\theta]^2 \\ &= \theta^2 \times L - 2\theta \sum_{s=t-L+1}^{t} y(s) + \text{const.} \end{aligned}$$

Estimate

$$\hat\theta(t) = \frac{1}{L} \sum_{s=t-L+1}^{t} y(s)$$

Rectangular forgetting, cont'd

Estimate

$$\begin{aligned} \hat\theta(t) &= \frac{1}{L} \sum_{s=t-L+1}^{t} y(s) \\ &= \frac{1}{L} \left[y(t) - y(t-L) + \sum_{s=t-L}^{t-1} y(s) \right] \\ &= \frac{1}{L} \left[y(t) - y(t-L) + L\hat\theta(t-1) \right] \\ &= \hat\theta(t-1) + \frac{1}{L} [y(t) - y(t-L)] \end{aligned}$$

Observer-based estimator

Model

$$\begin{aligned} \theta(t+1) &= \theta(t) \\ y(t) &= \theta(t) + e(t) \end{aligned}$$

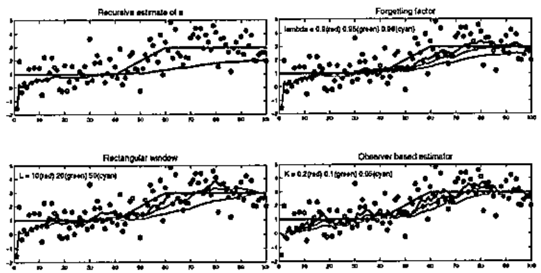
Observer

$$\hat\theta(t+1) = \hat\theta(t) + K[y(t) - \hat\theta(t)]$$

Observer error: $\tilde\theta(t) = \hat\theta(t) - \theta(t)$

$$\tilde\theta(t+1) = (1-K)\tilde\theta(t) + Ke(t)$$

Simulations



Analysis

Examine

1. Noise sensitivity (Assume $y(t) = \theta + e(t)$)
   $\tilde\theta(t) = \hat\theta(t) - \theta$
   Measure: $E\tilde\theta^2(t)$
2. Tracking capability
   $$\hat\theta(t) = \alpha\hat\theta(t-1) + (1-\alpha)\left[y(t) - \hat\theta(t-1)\right]$$
   Measure (dominating pole location): α
   $(0 \leq \alpha \leq 1)$

Analysis: time-invariant case

$$
\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + \frac{1}{t}\left[y(t) - \hat{\theta}(t-1)\right] \\
\tilde{\theta}(t) &= \hat{\theta}(t) - \theta \\
\tilde{\theta}(t) &= \tilde{\theta}(t-1) + \frac{1}{t}\left[e(t) - \tilde{\theta}(t-1)\right] \\
t\tilde{\theta}(t) &= (t-1)\tilde{\theta}(t-1) + e(t)
\end{aligned}
$$

Analysis: time-invariant case, cont'd

$$
\begin{aligned}
t\tilde{\theta}(t) &= (t-1)\tilde{\theta}(t-1) + e(t) \\
P(t) &\triangleq E[t\tilde{\theta}(t)]^2 \\
\Rightarrow P(t) &= P(t-1) + \sigma^2 \\
\Rightarrow P(t) &= t\sigma^2 \\
\Rightarrow E\tilde{\theta}^2(t) &= P(t)/t^2 = \sigma^2/t
\end{aligned}
$$

Hence consistency: $\hat{\theta}(t) \to \theta$ as $t \to \infty$.

Analysis: exponential forgetting

$$
\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + \gamma(t)\left[y(t) - \hat{\theta}(t-1)\right] \\
\tilde{\theta}(t) &= \hat{\theta}(t) - \theta \\
\tilde{\theta}(t) &= \tilde{\theta}(t-1) + \gamma(t)\left[e(t) - \tilde{\theta}(t-1)\right]
\end{aligned}
$$

Gain sequence

$$
\gamma(t) = \frac{1-\lambda}{1-\lambda^t} \to 1-\lambda \quad \text{if } t \to \infty
$$

Analysis: exponential forgetting, cont'd

$$
\begin{aligned}
\tilde{\theta}(t) &= \tilde{\theta}(t-1) + (1-\lambda)\left[e(t) - \tilde{\theta}(t-1)\right] \\
\tilde{\theta}(t) &= \lambda\tilde{\theta}(t-1) + (1-\lambda)e(t)
\end{aligned}
$$

AR(1) process!

$$
E\tilde{\theta}^2(t) = \frac{(1-\lambda)^2\sigma^2}{1-\lambda^2} = \sigma^2\frac{1-\lambda}{1+\lambda}
$$

Dominating pole = λ.

Analysis: rectangular forgetting

$$
\begin{aligned}
\hat{\theta}(t) &= \frac{1}{L}\sum_{s=t-L+1}^{t} y(s) \\
&= \theta + \frac{1}{L}\sum_{s=t-L+1}^{t} e(s) \\
&= \theta + \tilde{\theta}(t) \\
E\tilde{\theta}^2(t) &= \frac{\sigma^2}{L}
\end{aligned}
$$

Analysis: Observer

$$
\tilde{\theta}(t) = (1-K)\tilde{\theta}(t-1) + Ke(t)
$$

AR(1) process!

$$
E\tilde{\theta}^2(t) = \frac{K^2\sigma^2}{1-(1-K)^2} = \sigma^2\frac{K}{2-K}
$$

Dominating pole $= 1-K$

Noise sensitivity

1. Time-invariant case: $E\tilde{\theta}^2(t) = \sigma^2/t$
2. Time-varying case (when $t \to \infty$):
   - Exponential forgetting: $E\tilde{\theta}^2(t) = \sigma^2\frac{1-\lambda}{1+\lambda}$
   - Rectangular forgetting: $E\tilde{\theta}^2(t) = \sigma^2/L$
   - Observer: $E\tilde{\theta}^2(t) = \sigma^2\frac{K}{2-K}$

Inherent tradeoff

- Decrease noise sensitivity [$E\tilde{\theta}^2(t)$ small]
  [$\lambda \approx 1$; L large; $K \approx 0$]
  long memory

Inherent tradeoff

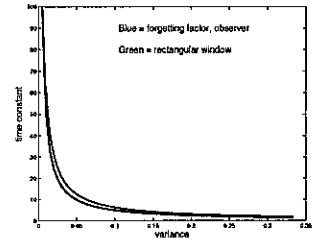
- Decrease noise sensitivity [$E\tilde{\theta}^2(t)$ small]
  [$\lambda \approx 1$; L large; $K \approx 0$]
  long memory
- Fast tracking capability [pole ≈ 0]
  [$\lambda \ll 1$; L small; $K \approx 1$]
  short memory

Inherent tradeoff

- Decrease noise sensitivity [$E\tilde{\theta}^2(t)$ small]
  $[\lambda \approx 1;\ L$ large; $K \approx 0]$
  long memory
- Fast tracking capability [pole ≈ 0]
  $[\lambda \ll 1;\ L$ small; $K \approx 1]$
  short memory

Inherent tradeoff, cont'd

Time constant for step change in $\theta(t)$ vs estimation error variance
(Time in sampling intervals to reach 63 % of final value)



Contents

- Introduction
- Derivation of algorithms
- Tracking time-varying dynamics
- Summary

Approaches for recursive identification

Apply recursive identification to linear dynamic models

- Modification of off-line methods
- Nonlinear filtering
- Stochastic approximation
- Pseudolinear regression and model reference techniques

Approach 1: Modify off-line algorithms

Least squares method, linear regression model

$$y(t) = \varphi^T(t)\theta + v(t)$$

The regressor vector $\varphi(t)$ depends on data available at time t.

Least squares estimate at time t:

$$\left[\sum_{s=1}^{t} \varphi(s)\varphi^T(s)\right]\hat{\theta}(t) = \left[\sum_{s=1}^{t} \varphi(s)y(s)\right]$$

Example of linear regression model

ARX (autoregressive with exogenous input) model

$$A(q^{-1})y(t) = B(q^{-1})u(t) + v(t)$$

$$\begin{aligned} &y(t) + a_1 y(t-1) + \ldots + a_n y(t-n) \\ &= b_1 u(t-1) + \ldots + b_n u(t-n) + v(t) \end{aligned}$$

Then

$$\begin{aligned} \varphi(t) &= (-y(t-1) \ldots u(t-n))^T \\ \theta &= (a_1 \ldots a_n\, b_1 \ldots b_n)^T \end{aligned}$$

Derivation of recursive least squares (RLS)

Set

$$R(t) = \sum_{s=1}^{t} \varphi(s)\varphi^T(s), \quad P(t) = R^{-1}(t)$$

$$\begin{aligned} R(t) &= R(t-1) + \varphi(t)\varphi^T(t) \\ &\Rightarrow \text{[matrix inversion lemma]} \\ P(t) &= P(t-1) - \frac{P(t-1)\varphi(t)\varphi^T(t)P(t-1)}{1 + \varphi^T(t)P(t-1)\varphi(t)} \end{aligned}$$

Updating the parameter estimates

$$\begin{aligned} \hat{\theta}(t) &= P(t)\left[\varphi(t)y(t) + \sum_{s=1}^{t-1} \varphi(s)y(s)\right] \\ &= P(t)\left[\varphi(t)y(t) + \left(R(t) - \varphi(t)\varphi^T(t)\right)\hat{\theta}(t-1)\right] \\ &= \hat{\theta}(t-1) + P(t)\varphi(t)\left[y(t) - \varphi^T(t)\hat{\theta}(t-1)\right] \\ &= \hat{\theta}(t-1) + K(t)\varepsilon(t) \end{aligned}$$

Prediction error

$$\varepsilon(t) = y(t) - \varphi^T(t)\hat{\theta}(t-1)$$

Estimator gain

$$\begin{aligned} K(t) &= P(t)\varphi(t) \\ &= P(t-1)\varphi(t) - \frac{P(t-1)\varphi(t)\varphi^T(t)P(t-1)}{1 + \varphi^T(t)P(t-1)\varphi(t)}\varphi(t) \\ &= P(t-1)\varphi(t)\left(1 - \frac{\varphi^T(t)P(t-1)\varphi(t)}{1 + \varphi^T(t)P(t-1)\varphi(t)}\right) \\ &= P(t-1)\varphi(t)\frac{1}{1 + \varphi^T(t)P(t-1)\varphi(t)} \\ P(t) &= P(t-1) - K(t)\varphi^T(t)P(t-1) \end{aligned}$$

Comments on RLS

- No approximations used!
- An instrumental variable estimator can be written in a similar way [Substitute $\varphi(t)$ by $z(t)$; keep $\varphi^T(t)$].
- Recursive (or on-line) algorithms RLS and RIV have the same convergence properties ($t \to \infty$) as the batch (or off-line) algorithms.
- How to set the initial values $P(0)$, $\hat{\theta}(0)$?

Recursive prediction error method

Prediction error criterion

$$V_t(\theta) = \frac{1}{2}\sum_{s=1}^{t} \varepsilon^2(s,\theta)$$

The model structure specifies how the predition error $\varepsilon(t,\theta)$ is to be computed from available input and output data.
An **exact** recursive minimizer of $V_t(\theta)$ does not exist in general.
Important exception: LS, for linear regression model.
In the general case: Make approximations, that hold without errors for the LS case.

Example: ARMAX model

$$A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})e(t)$$

Parameter vector

$$\theta = (a_1 \dots a_n \; b_1 \dots b_n \; c_1 \dots c_n)^T$$

Prediction errors (for one step optimal prediction)

$$\varepsilon(t,\theta) = \frac{A(q^{-1})}{C(q^{-1})}y(t) - \frac{B(q^{-1})}{C(q^{-1})}u(t)$$

Prediction errors

Prediction errors in general obtained by IIR filters!

For linear models, the prediction errors are

- linear in the data $y(\cdot)$, $u(\cdot)$.
- generally nonlinear in the parameters θ. (RLS is an exception!)

RPEM, cont'd

Taylor series expansion around $\hat{\theta}(t-1)$:

$$\begin{aligned} V_t(\theta) &= V_t(\hat{\theta}(t-1)) + V_t'(\hat{\theta}(t-1))[\theta - \hat{\theta}(t-1)] \\ &+ \frac{1}{2}[\theta - \hat{\theta}(t-1)]^T V_t''(\hat{\theta}(t-1))[\theta - \hat{\theta}(t-1)] \end{aligned}$$

Minimize!

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \left[V_t''(\hat{\theta}(t-1))\right]^{-1}\left[V_t'(\hat{\theta}(t-1))\right]^T$$

RPEM, cont'd

Negative gradient

$$\psi(t,\theta) \triangleq \left[-\frac{\partial}{\partial\theta}\varepsilon(t,\theta)\right]^T$$

Recursion of V_t implies

$$\begin{aligned} V_t'(\theta) &= -\sum_s \varepsilon(s,\theta)\psi(s,\theta) \\ &= V_{t-1}'(\theta) - \varepsilon(t,\theta)\psi(t,\theta) \\ V_t''(\theta) &= \sum \psi(s,\theta)\psi^T(s,\theta) + \sum \varepsilon(s,\theta)\varepsilon''(s,\theta) \\ &= V_{t-1}''(\theta) + \psi(t,\theta)\psi^T(t,\theta) + \varepsilon(t,\theta)\varepsilon''(t,\theta) \end{aligned}$$

RPEM cont'd

Approximations

- Taylor series of $V_t(\theta)$ up to second order terms
- $V_{t-1}'(\hat{\theta}(t-1)) = 0$ [assume $\hat{\theta}(t-1) = \theta$]
- $\varepsilon\varepsilon''$ negligible compared to $\psi\psi^T$
- $V_t''(\hat{\theta}(t)) = V_t''(\hat{\theta}(t-1))$

RPEM derivation, cont'd

Set

$$\begin{aligned} R(t) &= \sum_{s=1}^{t} \psi(s,\hat{\theta}(t-1))\psi^T(s,\hat{\theta}(t-1)) \\ &= R(t-1) + \psi(t,\hat{\theta}(t-1))\psi^T(t,\hat{\theta}(t-1)) \\ \hat{\theta}(t) &= \hat{\theta}(t-1) + R^{-1}(t)\psi(t,\hat{\theta}(t-1))\varepsilon(t,\hat{\theta}(t-1)) \end{aligned}$$

How to compute
$\psi(s,\hat{\theta}(t-1))$, $\varepsilon(s,\hat{\theta}(t-1))$, $s \le t$?

Computing prediction errors

Find approximation (depends on the model parameterization)

$$\varepsilon(t,\hat{\theta}(t-1)) \sim \varepsilon(t), \quad \psi(t,\hat{\theta}(t-1)) \sim \psi(t)$$

Set $P(t) = R^{-1}(t)$. Algorithm (cf RLS!)

$$\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) + K(t)\varepsilon(t) \\ P(t) &= P(t-1) - \frac{P(t-1)\psi(t)\psi^T(t)P(t-1)}{1 + \psi^T(t)P(t-1)\psi(t)} \\ K(t) &= P(t)\psi(t) = \frac{P(t-1)\psi(t)}{1 + \psi^T(t)P(t-1)\psi(t)} \end{aligned}$$

Approximate prediction errors

Principle: Use recent parameter estimates
Example: MA(1) process

$$
\begin{aligned}
y(t) &= e(t) + ce(t-1), \quad \theta = c \\
\varepsilon(t,\theta) &= \frac{1}{1+cq^{-1}} y(t) \\
\psi(t,\theta) &= \frac{-1}{(1+cq^{-1})^2} y(t-1) = \frac{-1}{1+cq^{-1}} \varepsilon(t-1,\theta)
\end{aligned}
$$

Approximation

$$
\begin{aligned}
\varepsilon(t) &= -\hat{c}(t-1)\varepsilon(t-1) + y(t) \\
\psi(t) &= -\hat{c}(t-1)\psi(t-1) - \varepsilon(t-1)
\end{aligned}
$$

Approaches for recursive identification

- Modification of off-line methods
- Nonlinear filtering
- Stochastic approximation
- Pseudolinear regression and model reference techniques

Approach 2: Nonlinear filtering

Regard θ as a random variable.

Find the conditional pdf $p(\theta|Z^t)$,
$Z^t = \{y(1), u(1), \dots y(t), u(t)\}$.

Choose

$$\hat{\theta}(t) = E(\theta|Z^t)$$

Nonlinear filtering, linear regresssion

Model

$$y(t) = \varphi^T(t)\theta + e(t), \quad e(t) \sim N(0, r_2)$$

Here, $\varphi(t)$ depends on Z^{t-1}.
A priori distribution $\theta \sim N(\theta_0, P_0)$.
Use Bayes' formula

$$
\begin{aligned}
p(\theta|Y^t) &= p(\theta|Y^{t-1}, y(t)) = \frac{p(y(t)|\theta, Y^{t-1})p(\theta|Y^{t-1})}{p(y(t)|Y^{t-1})} \\
p(\theta|Y^0) &\sim N(\theta_0, P_0)
\end{aligned}
$$

Nonlinear filtering, cont'd

Gaussian case

$$p(y(t)|\theta, Y^{t-1}) \sim N(\varphi^T(t)\theta, r_2(t))$$

Assume

$$p(\theta|Y^{t-1}) \sim N(\hat{\theta}(t-1)|P(t-1))$$

Use induction to show

$$p(\theta|Y^t) \sim N(\hat{\theta}(t)|P(t))$$

How are $\hat{\theta}(t)$ and $P(t)$ updated?

Nonlinear filtering algorithm

Using properties of conditional Gaussian pdf's:

$$\Rightarrow$$

$$
\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + K(t)[y(t) - \varphi^T(t)\hat{\theta}(t-1)] \\
K(t) &= \frac{P(t-1)\varphi(t)}{r_2 + \varphi^T(t)P(t-1)\varphi(t)} \\
P(t) &= P(t-1) - \frac{P(t-1)\varphi(t)\varphi^T(t)P(t-1)}{r_2 + \varphi^T(t)P(t-1)\varphi(t)}
\end{aligned}
$$

[This is RLS, with scaling if $r_2 \neq 1$]

Kalman filter interpretation

Linear regression model

$$
\begin{aligned}
\theta(t+1) &= \theta(t) \quad [+v(t)] \\
y(t) &= \varphi^T(t)\theta + e(t)
\end{aligned}
$$

Find

$$\hat{x}(t+1) = E[x(t+1)|Y^t] = \hat{x}(t+1|t)$$

Kalman filter, cont'd

$$
\begin{aligned}
\hat{x}(t+1) &= \hat{x}(t) + K(t)[y(t) - \varphi^T(t)\hat{x}(t-1)] \\
K(t) &= \frac{\overline{P}(t)\varphi(t)}{r_2 + \varphi^T(t)\overline{P}(t)\varphi(t)} \\
\overline{P}(t) &= \overline{P}(t-1) + R_1 - \frac{\overline{P}(t)\varphi(t)\varphi^T(t)\overline{P}(t)}{r_2 + \varphi^T(t)\overline{P}(t)\varphi(t)}
\end{aligned}
$$

(Set $\overline{P}(t) = P(t-1)$.)

Note: Interpretation of $\hat{\theta}(0), P(0)$.

Nonlinear filtering, general case

State-space model

$$
\begin{aligned}
x(t+1) &= F(\theta)x(t) + G(\theta)u(t) + w(t) \\
y(t) &= H(\theta)x(t) + e(t) \\
& \mathrm{cov}(w(t)) = R_1(\theta), \ \mathrm{cov}(e(t)) = R_2(\theta)
\end{aligned}
$$

Form an extended state space model using

$$\overline{x}(t) = \begin{pmatrix} x(t) \\ \theta \end{pmatrix}$$

and use, say, an extended Kalman filter for estimating $\overline{x}(t)$.

Approaches for recursive identification

- Modification of off-line methods
- Nonlinear filtering
- Stochastic approximation
- Pseudolinear regression and model reference techniques

Approach 3: Stochastic approximation

Problem: Solve

$$0 = f(x) = EQ(x, e(t))$$

w r t x; $e(t)$ is unknown; $Q(x, e)$ can be observed for arbitrary x.
Example: Minimize

$$V(\theta) = \frac{1}{2}E[y(t) - \varphi^T(t)\theta]^2$$

$$0 = E\varphi(t)[y(t) - \varphi^T(t)\theta]$$

Here, $x = \theta$, $e(t) = \begin{pmatrix} y(t) & \varphi(t) \end{pmatrix}$.

Stochastic approximation algorithm

$$\hat{x}(t) = \hat{x}(t-1) + \gamma(t)Q(\hat{x}(t-1), e(t))$$

Steplength $\gamma(t)$ satisfies

$$\gamma(t) > 0, \quad \gamma(t) \to 0, t \to \infty$$

Usual conditions

$$\sum_{t=1}^{\infty} \gamma(t) = \infty, \quad \sum_{t=1}^{\infty} \gamma^2(t) < \infty$$

Example: $\gamma(t) = 1/t$

Stochastic approximation example

Solve

$$E[e(t) - x] = 0$$

[Hence, $x = Ee(t)$.] Algorithm with $\gamma(t) = 1/t$:

$$\begin{aligned}
\hat{x}(t) &= \hat{x}(t-1) + \frac{1}{t}[e(t) - \hat{x}(t-1)] \\
&= \frac{t-1}{t}\hat{x}(t-1) + \frac{1}{t}e(t) \\
t\hat{x}(t) &= (t-1)\hat{x}(t-1) + e(t) \\
t\hat{x}(t) &= \sum_{s=1}^{t} e(s) \Rightarrow \hat{x}(t) = \frac{1}{t}\sum_{s=1}^{t} e(s)
\end{aligned}$$

SA algorithm, linear regresion

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \gamma(t)\varphi(t)[y(t) - \varphi^T(t)\hat{\theta}(t-1)]$$

LMS (least mean squares)

Some gradient algorithms:

- $\gamma(t) = \gamma_o$ (fixed)
- $\gamma(t) = \frac{\gamma_o}{|\varphi(t)|^2}$ (normalized)
- $\gamma(t) = \frac{\gamma_o}{\sum_{s=1}^{t} |\varphi(s)|^2}$ (decreasing)

SA algorithm, Newton direction

Set

$$V'' = E\varphi(t)\varphi^T(t)$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \gamma(t)\widehat{V''}^{-1}Q(\hat{\theta}(t-1), e(t))$$

Update V'' $\Rightarrow$

Algorithm

$$\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + \gamma(t)R^{-1}(t)\varphi(t)[y(t) - \varphi^T(t)\hat{\theta}(t-1)] \\
R(t) &= R(t-1) + \gamma(t)[\varphi(t)\varphi^T(t) - R(t-1)]
\end{aligned}$$

Approaches for recursive identification

- Modification of off-line methods
- Nonlinear filtering
- Stochastic approximation
- Pseudolinear regression and model reference techniques

Approach 4: Pseudolinear regression (PLR)

Idea: Treat $\varepsilon(t)$ as a second input, known at times $s \leq t+1$. Apply standard RLS.

The idea is applicable to the general linear model

$$A(q^{-1})y(t) = \frac{B(q^{-1})}{F(q^{-1})}u(t) + \frac{C(q^{-1})}{D(q^{-1})}e(t)$$

(and special cases thereof).

PLR for ARMA models

Model

$$A(q^{-1})y(t) = C(q^{-1})e(t)$$

$$\begin{aligned}
y(t) &= \varphi_o^T(t)\theta + e(t) \\
\theta &= \begin{pmatrix} a_1 & \dots & a_n & c_1 & \dots & c_n \end{pmatrix}^T \\
\varphi_o^T(t) &= \begin{pmatrix} -y(t-1) & \dots & e(t-n) \end{pmatrix}
\end{aligned}$$

Define model regressor

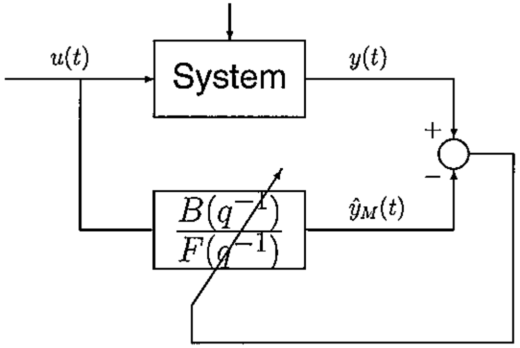
$$\varphi(t) = \begin{pmatrix} -y(t-1) & \dots & \varepsilon(t-n) \end{pmatrix}$$

PLR algorithm

Note $\varphi(t)$ is formed from Y^{t-1} and $\hat{\theta}(t-1),\ \hat{\theta}(t-2), \dots$

$$\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + K(t)\varepsilon(t)\\
\varepsilon(t) &= y(t) - \varphi^T(t)\hat{\theta}(t-1)\\
K(t) &= \frac{P(t-1)\varphi(t)}{1+\varphi^T(t)P(t-1)\varphi(t)}\\
P(t) &= P(t-1) - \frac{P(t-1)\varphi(t)\varphi^T(t)P(t-1)}{1+\varphi^T(t)P(t-1)\varphi(t)}
\end{aligned}$$

Model reference technique



Model reference technique, cont'd

Find unknown output $\hat{y}_M(t)$.

$$y(t) \approx \frac{B(q^{-1})}{F(q^{-1})}u(t)$$

Rewrite model

$$\begin{aligned}
\hat{y}_M(t) &= \varphi^T(t)\hat{\theta}(t-1)\\
&= \begin{pmatrix} -\hat{y}_M(t-1) & \dots & u(t-n) \end{pmatrix} \begin{pmatrix} \hat{f}_1(t-1) \\ \vdots \\ \hat{b}_n(t-1) \end{pmatrix}
\end{aligned}$$

Model reference algorithm

$$\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + P(t)\varphi(t)\left[y(t) - \hat{y}_M(t)\right]\\
P(t) &= \left[\sum_{s=1}^{t} \varphi(s)\varphi^T(s)\right]^{-1}
\end{aligned}$$

This can be interpreted as a recursive output error method.

The algorithm has better convergence than RLS (often no bias).

Contents

- Introduction
- Derivation of algorithms
- Tracking time-varying dynamics
- Summary

Tracking time-varying dynamics

Basic algorithm

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t)\varepsilon(t)$$

When the true parameter vector is time-varying, attempt to achieve $\hat{\theta}(t) \approx \theta_o(t)$.

Prevent $K(t)$ from decreasing to 0: For example,

- Modify model
- Modify criterion
- Modify gain sequence $\gamma(t)$

Tracking, Bayes' approach

Modified model (drift, random walk)

$$\begin{aligned}
\theta(t+1) &= \theta(t) + v(t), \mathrm{cov}(v(t)) = R_1\\
y(t) &= \varphi^T(t)\theta(t) + e(t),\ \ \mathrm{cov}(e(t)) = R_2
\end{aligned}$$

Algorithm (Kalman filter!)

$$\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + K(t)\varepsilon(t)\\
\varepsilon(t) &= y(t) - \varphi^T(t)\hat{\theta}(t-1)\\
K(t) &= \frac{P(t-1)\varphi(t)}{r_2+\varphi^T(t)P(t-1)\varphi(t)}\\
P(t) &= P(t-1) + R_1 - \frac{P(t-1)\varphi(t)\varphi^T(t)P(t-1)}{r_2+\varphi^T(t)P(t-1)\varphi(t)}
\end{aligned}$$

Tracking, modified off-line approach

Modified criterion

$$V_t(\theta) = \frac{1}{2}\sum_{s=1}^{t} \lambda^{t-s}\varepsilon^2(s,\theta)$$

with forgetting factor $\lambda < 1$. Algorithm

$$\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + K(t)\varepsilon(t)\\
\varepsilon(t) &= y(t) - \varphi^T(t)\hat{\theta}(t-1)\\
K(t) &= \frac{P(t-1)\varphi(t)}{\lambda+\varphi^T(t)P(t-1)\varphi(t)}\\
P(t) &= \left[P(t-1) - \frac{P(t-1)\varphi(t)\varphi^T(t)P(t-1)}{\lambda+\varphi^T(t)P(t-1)\varphi(t)}\right]\frac{1}{\lambda}
\end{aligned}$$

Using a forgetting factor, cont'd

Modified criterion

$$V_t(\theta) = \frac{1}{2}\sum_{s=1}^{t} \lambda^{t-s}\varepsilon^2(s,\theta)$$

Assume $\lambda \approx 1$.
Memory, time constant T?

$$\lambda^T = e^{T\log(\lambda)} \approx e^{T(\lambda-1)} = e^{-1}$$

gives

$$T = \frac{1}{1-\lambda}$$

Tracking, SA algorithm

Choose gain sequence such that $\gamma(t) \not\to 0$.

$$
\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + \gamma(t)R^{-1}(t)\varphi(t)\varepsilon(t) \\
R(t) &= R(t-1) + \gamma(t)[\varphi(t)\varphi^T(t) - R(t-1)]
\end{aligned}
$$

Compare with

$$
\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + P(t)\varphi(t)\varepsilon(t) \\
P(t) &= \left[P(t-1) - \frac{P(t-1)\varphi(t)\varphi^T(t)P(t-1)}{\lambda(t) + \varphi^T(t)P(t-1)\varphi(t)}\right]\frac{1}{\lambda(t)}
\end{aligned}
$$

Are the two algorithms identical?

Tracking, SA algorithms

The two algorithms (SA-based, and RPEM based, resp.) are identical, if

$$
\lambda(t) = \frac{1-\gamma(t)}{\gamma(t)}\gamma(t-1)
$$

Examples:

1. $\gamma(t) = \gamma_o \Rightarrow \lambda(t) = \lambda = 1 - \gamma_o$
2. $\gamma(t) = \frac{1}{t} \Rightarrow \lambda(t) = 1$

Contents

- Introduction
- Derivation of algorithms
- Tracking time-varying dynamics
- Summary

Summary

There is only one recursive identification algorithm,

$$
\begin{aligned}
\hat{\theta}(t) &= \hat{\theta}(t-1) + P(t)\varphi(t)\varepsilon(t) \\
P(t) &= \left[P(t-1) - \frac{P(t-1)\varphi(t)\varphi^T(t)P(t-1)}{\lambda(t) + \varphi^T(t)P(t-1)\varphi(t)}\right]\frac{1}{\lambda(t)}
\end{aligned}
$$

(+ a part depending on the model structure)

A large number of varieties, due to various user choices.

User choices in recursive algorithms

- Model structure/model parameterization
- Input signal [experimental condition]
- Criterion function
- Gain sequence $\gamma(t)$ [or forgetting factor $\lambda(t)$]
- Search direction [gradient vs. Newton]
- Initial conditions [$\theta(0)$, $P(0)$]
- [Optional: instrumental variable variant]
- [Optional: PLR variant]

Further reading

- L. Ljung and T. Söderström: Theory and Practice of Recursive Identification. MIT Press, Cambridge, USA, 1983.
- T. Söderström, L. Ljung and I. Gustavsson: A theoretical analysis of recursive identification methods. *Automatica,* vol 14, pp 231-244, 1978.